

ENZYMATIC GLYCOSYLATION OF STEVIOL GLYCOSIDES AND OTHER COMPOUNDS WITH GLUCOSE-1-PHOSPHATE

[0001] The present application claims priority to US Prov. Pat. Appln. Ser. No. 62/350,450, filed June 15, 2016, hereby incorporated by reference in its entirety for all purposes.

FIELD OF THE INVENTION

[0002] The present invention provides glycosyltransferase (GT) enzymes, polypeptides having GT activity, and polynucleotides encoding these enzymes, as well as vectors and host cells comprising these polynucleotides and polypeptides. The present invention also provides methods of using these GT enzymes to generate products with β -glucose linkages.

REFERENCE TO A "SEQUENCE LISTING," A TABLE, OR A COMPUTER PROGRAM LISTING APPENDIX SUBMITTED AS AN ASCII TEXT FILE

[0003] The official copy of the Sequence Listing is submitted concurrently with the specification as an ASCII formatted text file via EFS-Web, with a file name of "CX8-155USP1_ST25.txt", a creation date of June 14, 2016, and a size of 98,304 bytes. The Sequence Listing filed via EFS-Web is part of the specification and is incorporated in its entirety by reference herein.

BACKGROUND OF THE INVENTION

[0004] Glycosyltransferases (GT) are enzymes that post-translationally transfer glycosyl residues from an activated nucleotide sugar to monomeric and polymeric acceptor molecules (e.g., other sugars, proteins, lipids, and other organic substrates). These glycosylated molecules are involved in various metabolic pathways and processes. The transfer of a glucosyl moiety can alter the acceptor's bioactivity, solubility, and transport properties within cells.

SUMMARY OF THE INVENTION

[0005] The present invention provides glycosyltransferase enzymes, polypeptides having GT activity, and polynucleotides encoding these enzymes, as well as vectors and host cells comprising these polynucleotides and polypeptides. The present invention also provides methods of using these GT enzymes to generate products with β -glucose linkages.

[0006] The present invention provides methods for glycosylation of a substrate to produce a beta-glycosylated product, comprising the steps of: providing at least one glycosyl group donor, a least one glycosyl group acceptor, and at least one glycosyltransferase enzyme; contacting the glycosyl group donor and glycosyl group acceptor with the glycosyltransferase enzyme under conditions such that the glycosyl group acceptor is glycosylated to produce at least one product having beta-glucose linkages. In some embodiments of the methods, the glycosyl group donor is a glycosylphosphate. In some

additional embodiments of the methods, the glycosyl group donor is glucose-1-phosphate. In some further embodiments of the methods, the glycosyl group acceptor is selected from glycosyl, alkoxy, carboxy, aminocarbonyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carboxyalkyl, aminoalkyl, haloalkyl, alkylthioalkyl, heterocycloalkyl, heteroaryl, and heteroarylalkyl groups. In some yet additional embodiments of the methods, the product having beta-glucose linkages is a steviol glycoside. In some further embodiments of the methods, the glycosyl group acceptor is stevioside, said glycosyl group donor is alpha-glucose-1-phosphate, and said product having beta-glucose linkages is rebaudioside A. In some additional embodiments of the methods, the glycosyltransferase is selected from the polypeptides set forth in SEQ ID NOS:2, 4, 6, 8, 10, 12, and 14.

[0007] The present invention also provides methods for production of glucose-1-phosphate, comprising the steps of: providing a phosphorylase, inorganic phosphate, and a disaccharide, trisaccharide, or oligosaccharide substrate of the phosphorylase; contacting said phosphorylase, inorganic phosphate, and saccharide under conditions such that said saccharide is cleaved to produce a monosaccharide and glucose-1-phosphate. In some embodiments of the methods, this method is combined with the previously described method. In some additional embodiments of the methods, the glucosyl group acceptor is selected from glycosyl, alkoxy, carboxys, aminocarbonyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carboxyalkyl, aminoalkyl, haloalkyl, alkylthioalkyl, heterocycloalkyl, heteroaryl, and heteroarylalkyl groups. In still some additional embodiments of the methods, the phosphorylase is sucrose phosphorylase, said saccharide is sucrose, said monosaccharide produced is sucrose, and said glucose-1-phosphate produced is α -glucose-1-phosphate. In some further embodiments of the methods, the phosphorylase comprises a polypeptide sequence selected from SEQ ID NOS:16, 18, 20, 22, 24,26, 28, 30, and 32.

DESCRIPTION OF THE DRAWINGS

[0008] Figure 1 provides an enzymatic reaction scheme in which a glycosyltransferase catalyzes the transfer of a glycosyl group from a glycosylphosphate, for example glucose-1-phosphate, to an acceptor, for example R-OH, where R is any glycosyl, alkoxy, carboxy, aminocarbonyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carboxyalkyl, aminoalkyl, haloalkyl, alkylthioalkyl, heterocycloalkyl, heteroaryl, or heteroarylalkyl group.

[0009] Figure 2 provides a schematic of an embodiment of the invention in which the enzymatic reaction described in Figure 1 is applied to the substrate stevioside and the glycosylphosphate is α -glucose-1-phosphate and catalyzes the formation of a β -glucose linkage, to produce the product rebaudioside A.

[0010] Figure 3 provides a schematic of an embodiment of the invention in which two enzymatic reactions are paired for in situ generation of glucose-1-phosphate. In one reaction, sucrose phosphorylase uses inorganic phosphate to cleave sucrose, affording α -glucose-1-phosphate and fructose, and in the other reaction the glycosyl group from α -glucose-1-phosphate is transferred to an

acceptor. R is any glycosyl, alkoxy, carboxy, aminocarbonyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carboxyalkyl, aminoalkyl, haloalkyl, alkylthioalkyl, heterocycloalkyl, heteroaryl, or heteroarylalkyl group.

DESCRIPTION OF THE INVENTION

[0011] The present invention provides glycosyltransferase (GT) enzymes, polypeptides having GT activity, and polynucleotides encoding these enzymes, as well as vectors and host cells comprising these polynucleotides and polypeptides. The present invention also provides methods of using these GT enzymes to generate products with β -glucose linkages.

[0012] Unless defined otherwise, all technical and scientific terms used herein generally have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. Generally, the nomenclature used herein and the laboratory procedures of cell culture, molecular genetics, microbiology, organic chemistry, analytical chemistry and nucleic acid chemistry described below are those well-known and commonly employed in the art. Such techniques are well-known and described in numerous texts and reference works well known to those of skill in the art. Standard techniques, or modifications thereof, are used for chemical syntheses and chemical analyses. All patents, patent applications, articles and publications mentioned herein, both *supra* and *infra*, are hereby expressly incorporated herein by reference.

[0013] Although any suitable methods and materials similar or equivalent to those described herein find use in the practice of the present invention, some methods and materials are described herein. It is to be understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary, depending upon the context they are used by those of skill in the art. Accordingly, the terms defined immediately below are more fully described by reference to the invention as a whole.

[0014] It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of the present invention. The section headings used herein are for organizational purposes only and not to be construed as limiting the subject matter described. Numeric ranges are inclusive of the numbers defining the range. Thus, every numerical range disclosed herein is intended to encompass every narrower numerical range that falls within such broader numerical range, as if such narrower numerical ranges were all expressly written herein. It is also intended that every maximum (or minimum) numerical limitation disclosed herein includes every lower (or higher) numerical limitation, as if such lower (or higher) numerical limitations were expressly written herein.

Abbreviations

[0015] The abbreviations used for the genetically encoded amino acids are conventional and are as follows:

Amino Acid	Three-Letter Abbreviation	One-Letter Abbreviation
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartate	Asp	D
Cysteine	Cys	C
Glutamate	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

[0016] When the three-letter abbreviations are used, unless specifically preceded by an “L” or a “D” or clear from the context in which the abbreviation is used, the amino acid may be in either the L- or D-configuration about α -carbon (C_α). For example, whereas “Ala” designates alanine without specifying the configuration about the α -carbon, “D-Ala” and “L-Ala” designate D-alanine and L-alanine, respectively. When the one-letter abbreviations are used, upper case letters designate amino acids in the L-configuration about the α -carbon and lower case letters designate amino acids in the D-configuration about the α -carbon. For example, “A” designates L-alanine and “a” designates D-alanine. When polypeptide sequences are presented as a string of one-letter or three-letter abbreviations (or mixtures thereof), the sequences are presented in the amino (N) to carboxy (C) direction in accordance with common convention.

[0017] The abbreviations used for the genetically encoding nucleosides are conventional and are as follows: adenosine (A); guanosine (G); cytidine (C); thymidine (T); and uridine (U). Unless specifically delineated, the abbreviated nucleosides may be either ribonucleosides or 2'-deoxyribonucleosides. The nucleosides may be specified as being either ribonucleosides or 2'-deoxyribonucleosides on an individual basis or on an aggregate basis. When nucleic acid sequences are presented as a string of one-letter abbreviations, the sequences are presented in the 5' to 3' direction in accordance with common convention, and the phosphates are not indicated.

Definitions

[0018] In reference to the present invention, the technical and scientific terms used in the descriptions herein will have the meanings commonly understood by one of ordinary skill in the art, unless specifically defined otherwise. Accordingly, the following terms are intended to have the following meanings.

[0019] As used herein, the singular forms "a", "an" and "the" include plural referents unless the context clearly indicates otherwise. Thus, for example, reference to "a polypeptide" includes more than one polypeptide.

[0020] Similarly, "comprise," "comprises," "comprising" "include," "includes," and "including" are interchangeable and not intended to be limiting. Thus, as used herein, the term "comprising" and its cognates are used in their inclusive sense (*i.e.*, equivalent to the term "including" and its corresponding cognates).

[0021] It is to be further understood that where descriptions of various embodiments use the term "comprising," those skilled in the art would understand that in some specific instances, an embodiment can be alternatively described using language "consisting essentially of" or "consisting of."

[0022] The term "about" means an acceptable error for a particular value. In some instances "about" means within 0.05%, 0.5%, 1.0%, or 2.0%, of a given value range. In some instances, "about" means within 1, 2, 3, or 4 standard deviations of a given value.

[0023] "EC" number refers to the Enzyme Nomenclature of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB). The IUBMB biochemical classification is a numerical classification system for enzymes based on the chemical reactions they catalyze.

[0024] "ATCC" refers to the American Type Culture Collection whose biorepository collection includes genes and strains.

[0025] "NCBI" refers to National Center for Biological Information and the sequence databases provided therein.

[0026] "Protein," "polypeptide," and "peptide" are used interchangeably herein to denote a polymer of at least two amino acids covalently linked by an amide bond, regardless of length or post-

translational modification (e.g., glycosylation or phosphorylation). Included within this definition are D- and L-amino acids, and mixtures of D- and L-amino acids, as well as polymers comprising D- and L-amino acids, and mixtures of D- and L-amino acids.

[0027] “Amino acids” are referred to herein by either their commonly known three-letter symbols or by the one-letter symbols recommended by IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single letter codes.

[0028] As used herein, “polynucleotide” and “nucleic acid” refer to two or more nucleosides that are covalently linked together. The polynucleotide may be wholly comprised of ribonucleotides (i.e., RNA), wholly comprised of 2’ deoxyribonucleotides (i.e., DNA), or comprised of mixtures of ribo- and 2’ deoxyribonucleotides. While the nucleosides will typically be linked together via standard phosphodiester linkages, the polynucleotides may include one or more non-standard linkages. The polynucleotide may be single-stranded or double-stranded, or may include both single-stranded regions and double-stranded regions. Moreover, while a polynucleotide will typically be composed of the naturally occurring encoding nucleobases (i.e., adenine, guanine, uracil, thymine and cytosine), it may include one or more modified and/or synthetic nucleobases, such as, for example, inosine, xanthine, hypoxanthine, etc. In some embodiments, such modified or synthetic nucleobases are nucleobases encoding amino acid sequences

[0029] “Coding sequence” refers to that portion of a nucleic acid (e.g., a gene) that encodes an amino acid sequence of a protein.

[0030] As used herein, the terms “biocatalysis,” “biocatalytic,” “biotransformation,” and “biosynthesis” refer to the use of enzymes to perform chemical reactions on organic compounds.

[0031] “Glycosyltransferase” refers to a polypeptide having an enzymatic capability of transferring glycosyl residues from an activated sugar, for example a nucleotide diphosphate sugar or a phosphosugar, to monomeric and polymeric acceptor molecules.

[0032] “Phosphorylase” refers to a polypeptide having an enzymatic capability of cleaving glycosidic bonds using inorganic phosphate, releasing a phosphoglycoside and monomeric or polymeric product. In the reverse direction, a phosphorylase may act as a glycosyltransferase by transferring a glycosyl residue from a phosphoglycoside, for example glucose-1-phosphate, to monomeric and polymeric acceptor.

[0033] As used herein, “glycosylation” refers to the formation of a glycosidic linkage between a glycosyl residue and an acceptor molecule.

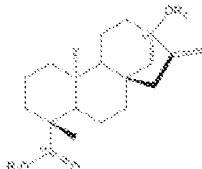
[0034] As used herein, “glucosylation” refers to the formation of a glycosidic linkage between a glucose residue and an acceptor molecule.

[0035] As used herein, “glycosyl” refers to an organic group that is a univalent free radical or substituent structure obtained by removing the hemiacetal hydroxyl group from the cyclic form of a monosaccharide, lower oligosaccharide or oligosaccharide derivative. Glycosyl groups react with inorganic acids (e.g., phosphoric acid) to form esters (e.g., glucose 1-phosphate).

[0036] As used herein, “glycoside” refers to a molecule in which a carbohydrate (e.g., sugar) is bound to another functional group by a glycosidic bond. Glycosides can be hydrolyzed to produce a sugar and a non-sugar (i.e., aglycone) component.

[0037] As used herein, the term “steviol glycoside” refers to a glycoside of steviol, including but not limited to, naturally occurring steviol glycosides (e.g., stevioside, steviolmonoside, steviolbioside, rubusoside, dulcoside B, dulcoside A, rebaudioside B, rebaudioside G, rebaudioside C, rebaudioside F, rebaudioside A, rebaudioside I, rebaudioside E, rebaudioside H, rebaudioside L, rebaudioside K, rebaudioside J, rebaudioside M (also known as rebaudioside X), rebaudioside D, rebaudioside N, rebaudioside O), and synthetic steviol glycosides (e.g., enzymatically glucosylated steviol glycosides), and combinations thereof. The chemical structures of steviol and its glycosides are below (See, WO 2013/176738).

Chemical Structures of Steviol and Its Glycosides

		
Steviol	H	H
Steviolmonoside	H	Glc β 1-
Steviol monoglucosyl ester	Glc β 1-	H
Rubusoside	Glc β 1-	Glc β 1-
Steviolbioside	H	Glc β (1-2) Glc β 1-
Dulcoside A	Glc β 1-	Rha α (1-2) Glc β 1-
Stevioside	Glc β 1-	Glc β (1-2) Glc β 1-
Rebaudioside B	H	Glc β (1-2)[Glc β (1-3)] Glc β 1-
Rebaudioside C	Glc β 1-	Rha α (1-2)[Glc β (1-3)] Glc β 1-

Rebaudioside A	Glc β 1-	Glc β (1-2)[Glc β (1-3)] Glc β 1-
Rebaudioside D	Glc β (1-2) Glc β 1-	Glc β (1-2)[Glc β (1-3)] Glc β 1-
Rebaudioside M	Glc β (1-2)[Glc β (1-3)] Glc β 1-	Glc β (1-2)[Glc β (1-3)] Glc β 1-
(Glc=glucose, Rha=rhamnose)		

[0038] As used herein, “wild-type” and “naturally-occurring” refer to the form found in nature. For example a wild-type polypeptide or polynucleotide sequence is a sequence present in an organism that can be isolated from a source in nature and which has not been intentionally modified by human manipulation.

[0039] As used herein, “recombinant,” “engineered,” and “non-naturally occurring” when used with reference to a cell, nucleic acid, or polypeptide, refer to a material, or a material corresponding to the natural or native form of the material, that has been modified in a manner that would not otherwise exist in nature. In some embodiments, the cell, nucleic acid or polypeptide is identical a naturally occurring cell, nucleic acid or polypeptide, but is produced or derived from synthetic materials and/or by manipulation using recombinant techniques. Non-limiting examples include, among others, recombinant cells expressing genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise expressed at a different level.

[0040] The term “percent (%) sequence identity” is used herein to refer to comparisons among polynucleotides or polypeptides, and are 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 for optimal alignment of the two sequences. The percentage may be 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 result by 100 to yield the percentage of sequence identity. Alternatively, the percentage may be calculated by determining the number of positions at which either the identical nucleic acid base or amino acid residue occurs in both sequences or a nucleic acid base or amino acid residue is aligned with a gap 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 result by 100 to yield the percentage of sequence identity. Those of skill in the art appreciate that there are many established algorithms available to align two sequences. Optimal alignment of sequences for comparison can be conducted by any suitable method, including, but not limited to the local homology algorithm of Smith and

Waterman (Smith and Waterman, *Adv. Appl. Math.*, 2:482 [1981]), by the homology alignment algorithm of Needleman and Wunsch (Needleman and Wunsch, *J. Mol. Biol.*, 48:443 [1970]), by the search for similarity method of Pearson and Lipman (Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444 [1988]), by computerized implementations of these algorithms (e.g., GAP, BESTFIT, FASTA, and TFASTA in the GCG Wisconsin Software Package), or by visual inspection, as known in the art. Examples of algorithms that are suitable for determining percent sequence identity and sequence similarity include, but are not limited to the BLAST and BLAST 2.0 algorithms, which are described by Altschul et al. (See Altschul et al., *J. Mol. Biol.*, 215: 403-410 [1990]; and Altschul et al., *Nucl. Acids Res.*, 3389-3402 [1977], respectively). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information website. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length *W* in the query sequence, which either match or satisfy some positive-valued threshold score *T* when aligned with a word of the same length in a database sequence. *T* is referred to as, the neighborhood word score threshold (See, Altschul et al., *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters *M* (reward score for a pair of matching residues; always >0) and *N* (penalty score for mismatching residues; always <0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity *X* from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters *W*, *T*, and *X* determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (*W*) of 11, an expectation (*E*) of 10, *M*=5, *N*=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (*W*) of 3, an expectation (*E*) of 10, and the BLOSUM62 scoring matrix (See, Henikoff and Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 [1989]). Exemplary determination of sequence alignment and % sequence identity can employ the BESTFIT or GAP programs in the GCG Wisconsin Software package (Accelrys, Madison WI), using default parameters provided.

[0041] As used herein, “reference sequence” refers to a defined sequence used as a basis for a sequence and/or activity comparison. A reference sequence may be a subset of a larger sequence, for example, a segment of a full-length gene or polypeptide sequence. Generally, a reference sequence is at least 20 nucleotide or amino acid residues in length, at least 25 residues in length, at least 50 residues in length, at least 100 residues in length or the full length of the nucleic acid or polypeptide. Since two polynucleotides or polypeptides may each (1) comprise a sequence (i.e., a portion of the complete sequence) that is similar between the two sequences, and (2) may further comprise a

sequence that is divergent between the two sequences, sequence comparisons between two (or more) polynucleotides or polypeptides are typically performed by comparing sequences of the two polynucleotides or polypeptides over a “comparison window” to identify and compare local regions of sequence similarity. In some embodiments, a “reference sequence” can be based on a primary amino acid sequence, where the reference sequence is a sequence that can have one or more changes in the primary sequence.

[0042] As used herein, “comparison window” refers to a conceptual segment of at least about 20 contiguous nucleotide positions or amino acid residues wherein a sequence may be compared to a reference sequence of at least 20 contiguous nucleotides or amino acids and wherein the portion of the sequence in the comparison window may comprise additions or deletions (i.e., gaps) of 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The comparison window can be longer than 20 contiguous residues, and includes, optionally 30, 40, 50, 100, or longer windows.

[0043] As used herein, “corresponding to,” “reference to,” and “relative to” when used in the context of the numbering of a given amino acid or polynucleotide sequence refer to the numbering of the residues of a specified reference sequence when the given amino acid or polynucleotide sequence is compared to the reference sequence. In other words, the residue number or residue position of a given polymer is designated with respect to the reference sequence rather than by the actual numerical position of the residue within the given amino acid or polynucleotide sequence. For example, a given amino acid sequence, such as that of an engineered glycosyltransferase, can be aligned to a reference sequence by introducing gaps to optimize residue matches between the two sequences. In these cases, although the gaps are present, the numbering of the residue in the given amino acid or polynucleotide sequence is made with respect to the reference sequence to which it has been aligned.

[0044] As used herein, “substantial identity” refers to a polynucleotide or polypeptide sequence that has at least 80 percent sequence identity, at least 85 percent identity, at least between 89 to 95 percent sequence identity, or more usually, at least 99 percent sequence identity as compared to a reference sequence over a comparison window of at least 20 residue positions, frequently over a window of at least 30-50 residues, wherein the percentage of sequence identity is calculated by comparing the reference sequence to a sequence that includes deletions or additions which total 20 percent or less of the reference sequence over the window of comparison. In some specific embodiments applied to polypeptides, the term “substantial identity” means that two polypeptide sequences, when optimally aligned, such as by the programs GAP or BESTFIT using default gap weights, share at least 80 percent sequence identity, preferably at least 89 percent sequence identity, at least 95 percent sequence identity or more (e.g., 99 percent sequence identity). In some embodiments, residue positions that are not identical in sequences being compared differ by conservative amino acid substitutions.

[0045] As used herein, “amino acid difference” and “residue difference” refer to a difference in the amino acid residue at a position of a polypeptide sequence relative to the amino acid residue at a corresponding position in a reference sequence. The positions of amino acid differences generally are referred to herein as “X_n,” where n refers to the corresponding position in the reference sequence upon which the residue difference is based. For example, a “residue difference at position X₉₃ as compared to SEQ ID NO:4” refers to a difference of the amino acid residue at the polypeptide position corresponding to position 93 of SEQ ID NO:4. Thus, if the reference polypeptide of SEQ ID NO:4 has a serine at position 93, then a “residue difference at position X₉₃ as compared to SEQ ID NO:4” an amino acid substitution of any residue other than serine at the position of the polypeptide corresponding to position 93 of SEQ ID NO:4. In most instances herein, the specific amino acid residue difference at a position is indicated as “X_nY” where “X_n” specified the corresponding position as described above, and “Y” is the single letter identifier of the amino acid found in the engineered polypeptide (i.e., the different residue than in the reference polypeptide). In some instances, the present invention also provides specific amino acid differences denoted by the conventional notation “A_nB”, where A is the single letter identifier of the residue in the reference sequence, “n” is the number of the residue position in the reference sequence, and B is the single letter identifier of the residue substitution in the sequence of the engineered polypeptide. In some instances, a polypeptide of the present invention can include one or more amino acid residue differences relative to a reference sequence, which is indicated by a list of the specified positions where residue differences are present relative to the reference sequence. In some embodiments, where more than one amino acid can be used in a specific residue position of a polypeptide, the various amino acid residues that can be used are separated by a “/” (e.g., X₃₀₇H/X₃₀₇P or X₃₀₇H/P). The slash may also be used to indicate multiple substitutions within a given variant (i.e., there is more than one substitution present in a given sequence). In some embodiments, the present invention includes engineered polypeptide sequences comprising one or more amino acid differences comprising conservative or non-conservative amino acid substitutions. In some additional embodiments, the present invention provides engineered polypeptide sequences comprising both conservative and non-conservative amino acid substitutions.

[0046] As used herein, “conservative amino acid substitution” refers to a substitution of a residue with a different residue having a similar side chain, and thus typically involves substitution of the amino acid in the polypeptide with amino acids within the same or similar defined class of amino acids. By way of example and not limitation, an amino acid with an aliphatic side chain is substituted with another aliphatic amino acid (e.g., alanine, valine, leucine, and isoleucine); an amino acid with an hydroxyl side chain is substituted with another amino acid with an hydroxyl side chain (e.g., serine and threonine); an amino acids having aromatic side chains is substituted with another amino acid having an aromatic side chain (e.g., phenylalanine, tyrosine, tryptophan, and histidine); an amino acid with a basic side chain is substituted with another amino acid with a basis side chain (e.g., lysine and

arginine); an amino acid with an acidic side chain is substituted with another amino acid with an acidic side chain (e.g., aspartic acid or glutamic acid); and/or a hydrophobic or hydrophilic amino acid is replaced with another hydrophobic or hydrophilic amino acid, respectively.

[0047] As used herein, “non-conservative substitution” refers to substitution of an amino acid in the polypeptide with an amino acid with significantly differing side chain properties. Non-conservative substitutions may use amino acids between, rather than within, the defined groups and affects (a) the structure of the peptide backbone in the area of the substitution (e.g., proline for glycine) (b) the charge or hydrophobicity, or (c) the bulk of the side chain. By way of example and not limitation, an exemplary non-conservative substitution can be an acidic amino acid substituted with a basic or aliphatic amino acid; an aromatic amino acid substituted with a small amino acid; and a hydrophilic amino acid substituted with a hydrophobic amino acid.

[0048] As used herein, “deletion” refers to modification to the polypeptide by removal of one or more amino acids from the reference polypeptide. Deletions can comprise removal of 1 or more amino acids, 2 or more amino acids, 5 or more amino acids, 10 or more amino acids, 15 or more amino acids, or 20 or more amino acids, up to 10% of the total number of amino acids, or up to 20% of the total number of amino acids making up the reference enzyme while retaining enzymatic activity and/or retaining the improved properties of an engineered glycosyltransferase enzyme. Deletions can be directed to the internal portions and/or terminal portions of the polypeptide. In various embodiments, the deletion can comprise a continuous segment or can be discontinuous.

[0049] As used herein, “insertion” refers to modification to the polypeptide by addition of one or more amino acids from the reference polypeptide. Insertions can be in the internal portions of the polypeptide, or to the carboxy or amino terminus. Insertions as used herein include fusion proteins as is known in the art. The insertion can be a contiguous segment of amino acids or separated by one or more of the amino acids in the naturally occurring polypeptide.

[0050] A “functional fragment” and “biologically active fragment” are used interchangeably herein to refer to a polypeptide that has an amino-terminal and/or carboxy-terminal deletion(s) and/or internal deletions, but where the remaining amino acid sequence is identical to the corresponding positions in the sequence to which it is being compared (e.g., a full-length engineered glycosyltransferase of the present invention) and that retains substantially all of the activity of the full-length polypeptide.

[0051] As used herein, “isolated polypeptide” refers to a polypeptide which is substantially separated from other contaminants that naturally accompany it (e.g., protein, lipids, and polynucleotides). The term embraces polypeptides which have been removed or purified from their naturally-occurring environment or expression system (e.g., within a host cell or via in vitro synthesis). The recombinant glycosyltransferase polypeptides may be present within a cell, present in the cellular medium, or prepared in various forms, such as lysates or isolated preparations. As such, in some embodiments, the recombinant glycosyltransferase polypeptides can be an isolated polypeptide.

[0052] As used herein, “substantially pure polypeptide” refers to a composition in which the polypeptide species is the predominant species present (i.e., on a molar or weight basis it is more abundant than any other individual macromolecular species in the composition), and is generally a substantially purified composition when the object species comprises at least about 50 percent of the macromolecular species present by mole or % weight. However, in some embodiments, the composition comprising glycosyltransferase comprises glycosyltransferase that is less than 50% pure (e.g., about 10%, about 20%, about 30%, about 40%, or about 50%) Generally, a substantially pure glycosyltransferase composition comprises about 60% or more, about 70% or more, about 80% or more, about 90% or more, about 95% or more, and about 98% or more of all macromolecular species by mole or % weight present in the composition. In some embodiments, the object species is purified to essential homogeneity (i.e., contaminant species cannot be detected in the composition by conventional detection methods) wherein the composition consists essentially of a single macromolecular species. Solvent species, small molecules (<500 Daltons), and elemental ion species are not considered macromolecular species. In some embodiments, the isolated recombinant glycosyltransferase polypeptides are substantially pure polypeptide compositions.

[0053] As used herein, “improved enzyme property” refers to at least one improved property of an enzyme. In some embodiments, the present invention provides engineered glycosyltransferase polypeptides that exhibit an improvement in any enzyme property as compared to a reference glycosyltransferase polypeptide and/or a wild-type glycosyltransferase polypeptide, and/or another engineered glycosyltransferase polypeptide. Thus, the level of “improvement” can be determined and compared between various glycosyltransferase polypeptides, including wild-type, as well as engineered glycosyltransferases. Improved properties include, but are not limited, to such properties as increased protein expression, increased thermoactivity, increased thermostability, increased pH activity, increased stability, increased enzymatic activity, increased substrate specificity or affinity, increased specific activity, increased resistance to substrate or end-product inhibition, increased chemical stability, improved chemoselectivity, improved solvent stability, increased tolerance to acidic pH, increased tolerance to proteolytic activity (i.e., reduced sensitivity to proteolysis), reduced aggregation, increased solubility, and altered temperature profile.

[0054] As used herein, “increased enzymatic activity” and “enhanced catalytic activity” refer to an improved property of the engineered glycosyltransferase polypeptides, which can be represented by an increase in specific activity (e.g., product produced/time/weight protein) or an increase in percent conversion of the substrate to the product (e.g., percent conversion of starting amount of substrate to product in a specified time period using a specified amount of glycosyltransferase) as compared to the reference glycosyltransferase enzyme. Exemplary methods to determine enzyme activity are provided in the Examples. Any property relating to enzyme activity may be affected, including the classical enzyme properties of K_m , V_{max} or k_{cat} , changes of which can lead to increased enzymatic activity. Improvements in enzyme activity can be from about 1.1 fold the enzymatic activity of the

corresponding wild-type enzyme, to as much as 2-fold, 5-fold, 10-fold, 20-fold, 25-fold, 50-fold, 75-fold, 100-fold, 150-fold, 200-fold or more enzymatic activity than the naturally occurring glycosyltransferase or another engineered glycosyltransferase from which the glycosyltransferase polypeptides were derived.

[0055] As used herein, “conversion” refers to the enzymatic conversion (or biotransformation) of a substrate(s) to the corresponding product(s). “Percent conversion” refers to the percent of the substrate that is converted to the product within a period of time under specified conditions. Thus, the “enzymatic activity” or “activity” of a glycosyltransferase polypeptide can be expressed as “percent conversion” of the substrate to the product in a specific period of time.

[0056] Enzymes with “generalist properties” (or “generalist enzymes”) refer to enzymes that exhibit improved activity for a wide range of substrates, as compared to a parental sequence. Generalist enzymes do not necessarily demonstrate improved activity for every possible substrate. In some embodiments, the present invention provides glycosyltransferase variants with generalist properties, in that they demonstrate similar or improved activity relative to the parental gene for a wide range of sterically and electronically diverse substrates. In addition, the generalist enzymes provided herein were engineered to be improved across a wide range of diverse molecules to increase the production of metabolites/products.

[0057] The term “stringent hybridization conditions” is used herein to refer to conditions under which nucleic acid hybrids are stable. As known to those of skill in the art, the stability of hybrids is reflected in the melting temperature (T_m) of the hybrids. In general, the stability of a hybrid is a function of ion strength, temperature, G/C content, and the presence of chaotropic agents. The T_m values for polynucleotides can be calculated using known methods for predicting melting temperatures (See e.g., Baldino et al., *Meth. Enzymol.*, 168:761-777 [1989]; Bolton et al., *Proc. Natl. Acad. Sci. USA* 48:1390 [1962]; Bresslauer et al., *Proc. Natl. Acad. Sci. USA* 83:8893-8897 [1986]; Freier et al., *Proc. Natl. Acad. Sci. USA* 83:9373-9377 [1986]; Kierzek et al., *Biochem.*, 25:7840-7846 [1986]; Rychlik et al., 1990, *Nucl. Acids Res.*, 18:6409-6412 [1990] (erratum, *Nucl. Acids Res.*, 19:698 [1991]); Sambrook et al., *supra*); Suggs et al., 1981, in Developmental Biology Using Purified Genes, Brown et al. [eds.], pp. 683-693, Academic Press, Cambridge, MA [1981]; and Wetmur, *Crit. Rev. Biochem. Mol. Biol.*, 26:227-259 [1991]). In some embodiments, the polynucleotide encodes the polypeptide disclosed herein and hybridizes under defined conditions, such as moderately stringent or highly stringent conditions, to the complement of a sequence encoding an engineered glycosyltransferase enzyme of the present invention.

[0058] As used herein, “hybridization stringency” relates to hybridization conditions, such as washing conditions, in the hybridization of nucleic acids. Generally, hybridization reactions are performed under conditions of lower stringency, followed by washes of varying but higher stringency. The term “moderately stringent hybridization” refers to conditions that permit target-DNA to bind a complementary nucleic acid that has about 60% identity, preferably about 75% identity, about 85%

identity to the target DNA, with greater than about 90% identity to target-polynucleotide. Exemplary moderately stringent conditions are conditions equivalent to hybridization in 50% formamide, 5× Denhart's solution, 5×SSPE, 0.2% SDS at 42°C, followed by washing in 0.2×SSPE, 0.2% SDS, at 42°C. “High stringency hybridization” refers generally to conditions that are about 10°C or less from the thermal melting temperature T_m as determined under the solution condition for a defined polynucleotide sequence. In some embodiments, a high stringency condition refers to conditions that permit hybridization of only those nucleic acid sequences that form stable hybrids in 0.018M NaCl at 65°C (i.e., if a hybrid is not stable in 0.018M NaCl at 65°C, it will not be stable under high stringency conditions, as contemplated herein). High stringency conditions can be provided, for example, by hybridization in conditions equivalent to 50% formamide, 5× Denhart's solution, 5×SSPE, 0.2% SDS at 42°C, followed by washing in 0.1×SSPE, and 0.1% SDS at 65°C. Another high stringency condition is hybridizing in conditions equivalent to hybridizing in 5X SSC containing 0.1% (w:v) SDS at 65°C and washing in 0.1x SSC containing 0.1% SDS at 65°C. Other high stringency hybridization conditions, as well as moderately stringent conditions, are described in the references cited above.

[0059] As used herein, “codon optimized” refers to changes in the codons of the polynucleotide encoding a protein to those preferentially used in a particular organism such that the encoded protein is efficiently expressed in the organism of interest. Although the genetic code is degenerate in that most amino acids are represented by several codons, called “synonyms” or “synonymous” codons, it is well known that codon usage by particular organisms is nonrandom and biased towards particular codon triplets. This codon usage bias may be higher in reference to a given gene, genes of common function or ancestral origin, highly expressed proteins versus low copy number proteins, and the aggregate protein coding regions of an organism's genome. In some embodiments, the polynucleotides encoding the glycosyltransferase enzymes may be codon optimized for optimal production in the host organism selected for expression.

[0060] As used herein, “preferred,” “optimal,” and “high codon usage bias” codons when used alone or in combination refer interchangeably to codons that are used at higher frequency in the protein coding regions than other codons that code for the same amino acid. The preferred codons may be determined in relation to codon usage in a single gene, a set of genes of common function or origin, highly expressed genes, the codon frequency in the aggregate protein coding regions of the whole organism, codon frequency in the aggregate protein coding regions of related organisms, or combinations thereof. Codons whose frequency increases with the level of gene expression are typically optimal codons for expression. A variety of methods are known for determining the codon frequency (e.g., codon usage, relative synonymous codon usage) and codon preference in specific organisms, including multivariate analysis, for example, using cluster analysis or correspondence analysis, and the effective number of codons used in a gene (See e.g., GCG CodonPreference, Genetics Computer Group Wisconsin Package; CodonW, Peden, University of Nottingham;

McInerney, *Bioinform.*, 14:372-73 [1998]; Stenico et al., *Nucl. Acids Res.*, 22:2437-46 [1994]; Wright, *Gene* 87:23-29 [1990]). Codon usage tables are available for many different organisms (See e.g., Wada et al., *Nucl. Acids Res.*, 20:2111-2118 [1992]; Nakamura et al., *Nucl. Acids Res.*, 28:292 [2000]; Duret, et al., *supra*; Henaut and Danchin, in *Escherichia coli and Salmonella*, Neidhardt, et al. (eds.), ASM Press, Washington D.C., p. 2047-2066 [1996]). The data source for obtaining codon usage may rely on any available nucleotide sequence capable of coding for a protein. These data sets include nucleic acid sequences actually known to encode expressed proteins (e.g., complete protein coding sequences-CDS), expressed sequence tags (ESTS), or predicted coding regions of genomic sequences (See e.g., Mount, *Bioinformatics: Sequence and Genome Analysis*, Chapter 8, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. [2001]; Uberbacher, *Meth. Enzymol.*, 266:259-281 [1996]; and Tiwari et al., *Comput. Appl. Biosci.*, 13:263-270 [1997]).

[0061] As used herein, “control sequence” includes all components, which are necessary or advantageous for the expression of a polynucleotide and/or polypeptide of the present invention. Each control sequence may be native or foreign to the nucleic acid sequence encoding the polypeptide. Such control sequences include, but are not limited to, a leader, polyadenylation sequence, propeptide sequence, promoter sequence, signal peptide sequence, initiation sequence and transcription terminator. At a minimum, the control sequences include a promoter, and transcriptional and translational stop signals. The control sequences may be provided with linkers for the purpose of introducing specific restriction sites facilitating ligation of the control sequences with the coding region of the nucleic acid sequence encoding a polypeptide.

[0062] “Operably linked” is defined herein as a configuration in which a control sequence is appropriately placed (i.e., in a functional relationship) at a position relative to a polynucleotide of interest such that the control sequence directs or regulates the expression of the polynucleotide and/or polypeptide of interest.

[0063] “Promoter sequence” refers to a nucleic acid sequence that is recognized by a host cell for expression of a polynucleotide of interest, such as a coding sequence. The promoter sequence contains transcriptional control sequences, which mediate the expression of a polynucleotide of interest. The promoter may be any nucleic acid sequence which shows transcriptional activity in the host cell of choice including mutant, truncated, and hybrid promoters, and may be obtained from genes encoding extracellular or intracellular polypeptides either homologous or heterologous to the host cell.

[0064] The phrase “suitable reaction conditions” refers to those conditions in the enzymatic conversion reaction solution (e.g., ranges of enzyme loading, substrate loading, temperature, pH, buffers, co-solvents, etc.) under which a glycosyltransferase polypeptide of the present invention is capable of converting a substrate to the desired product compound. Some exemplary “suitable reaction conditions” are provided herein.

[0065] As used herein, “loading,” such as in “compound loading” or “enzyme loading” refers to the concentration or amount of a component in a reaction mixture at the start of the reaction.

[0066] As used herein, "substrate" in the context of an enzymatic conversion reaction process refers to the compound or molecule acted on by the glycosyltransferase polypeptide.

[0067] As used herein, "product" in the context of an enzymatic conversion process refers to the compound or molecule resulting from the action of the glycosyltransferase polypeptide on a substrate.

[0068] As used herein the term "culturing" refers to the growing of a population of microbial cells under any suitable conditions (e.g., using a liquid, gel or solid medium).

[0069] Recombinant polypeptides can be produced using any suitable methods known in the art. Genes encoding the wild-type polypeptide of interest can be cloned in vectors, such as plasmids, and expressed in desired hosts, such as *E. coli*, etc. Variants of recombinant polypeptides can be generated by various methods known in the art. Indeed, there is a wide variety of different mutagenesis techniques well known to those skilled in the art. In addition, mutagenesis kits are also available from many commercial molecular biology suppliers. Methods are available to make specific substitutions at defined amino acids (site-directed), specific or random mutations in a localized region of the gene (regio-specific), or random mutagenesis over the entire gene (e.g., saturation mutagenesis).

Numerous suitable methods are known to those in the art to generate enzyme variants, including but not limited to site-directed mutagenesis of single-stranded DNA or double-stranded DNA using PCR, cassette mutagenesis, gene synthesis, error-prone PCR, shuffling, and chemical saturation mutagenesis, or any other suitable method known in the art. Non-limiting examples of methods used for DNA and protein engineering are provided in the following patents: US Pat. No. 6,117,679; US Pat. No. 6,420,175; US Pat. No. 6,376,246; US Pat. No. 6,586,182; US Pat. No. 7,747,391; US Pat. No. 7,747,393; US Pat. No. 7,783,428; and US Pat. No. 8,383,346. After the variants are produced, they can be screened for any desired property (e.g., high or increased activity, or low or reduced activity, increased thermal activity, increased thermal stability, and/or acidic pH stability, etc.). In some embodiments, "recombinant glycosyltransferase polypeptides" (also referred to herein as "engineered glycosyltransferase polypeptides," "variant glycosyltransferase enzymes," and "glycosyltransferase variants") find use.

[0070] As used herein, a "vector" is a DNA construct for introducing a DNA sequence into a cell. In some embodiments, the vector is an expression vector that is operably linked to a suitable control sequence capable of effecting the expression in a suitable host of the polypeptide encoded in the DNA sequence. In some embodiments, an "expression vector" has a promoter sequence operably linked to the DNA sequence (e.g., transgene) to drive expression in a host cell, and in some embodiments, also comprises a transcription terminator sequence.

[0071] As used herein, the term "expression" includes any step involved in the production of the polypeptide including, but not limited to, transcription, post-transcriptional modification, translation, and post-translational modification. In some embodiments, the term also encompasses secretion of the polypeptide from a cell.

[0072] As used herein, the term “produces” refers to the production of proteins and/or other compounds by cells. It is intended that the term encompass any step involved in the production of polypeptides including, but not limited to, transcription, post-transcriptional modification, translation, and post-translational modification. In some embodiments, the term also encompasses secretion of the polypeptide from a cell.

[0073] As used herein, an amino acid or nucleotide sequence (e.g., a promoter sequence, signal peptide, terminator sequence, etc.) is “heterologous” to another sequence with which it is operably linked if the two sequences are not associated in nature. For example a “heterologous polynucleotide” is any polynucleotide that is introduced into a host cell by laboratory techniques, and includes polynucleotides that are removed from a host cell, subjected to laboratory manipulation, and then reintroduced into a host cell.

[0074] As used herein, the terms “host cell” and “host strain” refer to suitable hosts for expression vectors comprising DNA provided herein (e.g., the polynucleotides encoding the glycosyltransferase variants). In some embodiments, the host cells are prokaryotic or eukaryotic cells that have been transformed or transfected with vectors constructed using recombinant DNA techniques as known in the art.

[0075] The term “analogue” means a polypeptide having more than 70% sequence identity but less than 100% sequence identity (e.g., more than 75%, 78%, 80%, 83%, 85%, 88%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity) with a reference polypeptide. In some embodiments, analogues means polypeptides that contain one or more non-naturally occurring amino acid residues including, but not limited, to homoarginine, ornithine and norvaline, as well as naturally occurring amino acids. In some embodiments, analogues also include one or more D-amino acid residues and non-peptide linkages between two or more amino acid residues.

[0076] The term “effective amount” means an amount sufficient to produce the desired result. One of general skill in the art may determine what the effective amount by using routine experimentation.

[0077] The terms “isolated” and “purified” are used to refer to a molecule (e.g., an isolated nucleic acid, polypeptide, etc.) or other component that is removed from at least one other component with which it is naturally associated. The term “purified” does not require absolute purity, rather it is intended as a relative definition.

[0078] “Stereoselectivity” refers to the preferential formation in a chemical or enzymatic reaction of one stereoisomer over another. Stereoselectivity can be partial, where the formation of one stereoisomer is favored over the other, or it may be complete where only one stereoisomer is formed. When the stereoisomers are enantiomers, the stereoselectivity is referred to as enantioselectivity, the fraction (typically reported as a percentage) of one enantiomer in the sum of both. It is commonly alternatively reported in the art (typically as a percentage) as the enantiomeric excess (e.e.) calculated therefrom according to the formula $[\text{major enantiomer} - \text{minor enantiomer}] / [\text{major enantiomer} + \text{minor enantiomer}]$. Where the stereoisomers are diastereoisomers, the stereoselectivity is referred to

as diastereoselectivity, the fraction (typically reported as a percentage) of one diastereomer in a mixture of two diastereomers, commonly alternatively reported as the diastereomeric excess (d.e.). Enantiomeric excess and diastereomeric excess are types of stereomeric excess.

[0079] As used herein, the terms “regioselectivity” and “regioselective reaction” refer to a reaction in which one direction of bond making or breaking occurs preferentially over all other possible directions. Reactions can be completely (100%) regioselective if the discrimination is complete, substantially regioselective (at least 75%), or partially regioselective (x%, wherein the percentage is set dependent upon the reaction of interest), if the product of reaction at one site predominates over the product of reaction at other sites, for example, preferential formation of the product compound (2) (i.e., 2S,3S0-hydroxyproline over the undesired product (2S,5S)-hydroxyproline.

[0080] As used herein, “thermostable” refers to a glycosyltransferase polypeptide that maintains similar activity (more than 60% to 80% for example) after exposure to elevated temperatures (e.g., 40-80°C) for a period of time (e.g., 0.5-24 h) compared to the wild-type enzyme exposed to the same elevated temperature.

[0081] As used herein, “solvent stable” refers to a glycosyltransferase polypeptide that maintains similar activity (more than e.g., 60% to 80%) after exposure to varying concentrations (e.g., 5-99%) of solvent (ethanol, isopropyl alcohol, dimethylsulfoxide (DMSO), tetrahydrofuran, 2-methyltetrahydrofuran, acetone, toluene, butyl acetate, methyl tert-butyl ether, etc.) for a period of time (e.g., 0.5-24 h) compared to the wild-type enzyme exposed to the same concentration of the same solvent.

[0082] As used herein, “thermo- and solvent stable” refers to a glycosyltransferase polypeptide that is both thermostable and solvent stable.

[0083] As used herein, “reductant” refers to a compound or agent capable of converting Fe^{+3} to Fe^{+2} . An exemplary reductant is ascorbic acid, which is generally in the form of L-ascorbic acid.

[0084] “Alkyl” refers to saturated hydrocarbon groups of from 1 to 18 carbon atoms inclusively, either straight chained or branched, more preferably from 1 to 8 carbon atoms inclusively, and most preferably 1 to 6 carbon atoms inclusively. An alkyl with a specified number of carbon atoms is denoted in parenthesis (e.g., (C₁-C₆)alkyl refers to an alkyl of 1 to 6 carbon atoms).

[0085] “Alkenyl” refers to hydrocarbon groups of from 2 to 12 carbon atoms inclusively, either straight or branched containing at least one double bond but optionally containing more than one double bond.

[0086] “Alkynyl” refers to hydrocarbon groups of from 2 to 12 carbon atoms inclusively, either straight or branched containing at least one triple bond but optionally containing more than one triple bond, and additionally optionally containing one or more double bonded moieties.

[0087] “Alkylene” refers to a straight or branched chain divalent hydrocarbon radical having from 1 to 18 carbon atoms inclusively, more preferably from 1 to 8 carbon atoms inclusively, and most preferably 1 to 6 carbon atoms inclusively, optionally substituted with one or more suitable

substituents. Exemplary “alkylenes” include, but are not limited to, methylene, ethylene, propylene, butylene, and the like.

[0088] “Alkenylene” refers to a straight or branched chain divalent hydrocarbon radical having 2 to 12 carbon atoms inclusively and one or more carbon-carbon double bonds, more preferably from 2 to 8 carbon atoms inclusively, and most preferably 2 to 6 carbon atoms inclusively, optionally substituted with one or more suitable substituents.

[0089] “Heteroalkyl,” “heteroalkenyl,” and “heteroalkynyl,” refer respectively, to alkyl, alkenyl and alkynyl as defined herein in which one or more of the carbon atoms are each independently replaced with the same or different heteroatoms or heteroatomic groups. Heteroatoms and/or heteroatomic groups which can replace the carbon atoms include, but are not limited to, -O-, -S-, -S-O-, -NR^Y-, -PH-, -S(O)-, -S(O)2-, -S(O)NR^Y-, -S(O)2NR^Y-, and the like, including combinations thereof, where each R^Y is independently selected from hydrogen, alkyl, cycloalkyl, heterocycloalkyl, aryl, and heteroaryl.

[0090] “Aryl” refers to an unsaturated aromatic carbocyclic group of from 6 to 12 carbon atoms inclusively having a single ring (e.g., phenyl) or multiple condensed rings (e.g., naphthyl or anthryl). Exemplary aryls include phenyl, pyridyl, naphthyl and the like.

[0091] “Arylalkyl” refers to an alkyl substituted with an aryl (i.e., aryl-alkyl- groups), preferably having from 1 to 6 carbon atoms inclusively in the alkyl moiety and from 6 to 12 carbon atoms inclusively in the aryl moiety. Such arylalkyl groups are exemplified by benzyl, phenethyl and the like.

[0092] “Aryloxy” refers to -OR^A groups, where R^A is an aryl group, which can be optionally substituted.

[0093] “Cycloalkyl” refers to cyclic alkyl groups of from 3 to 12 carbon atoms inclusively having a single cyclic ring or multiple condensed rings which can be optionally substituted with from 1 to 3 alkyl groups. Exemplary cycloalkyl groups include, but are not limited to, single ring structures such as cyclopropyl, cyclobutyl, cyclopentyl, cyclooctyl, 1-methylcyclopropyl, 2-methylcyclopentyl, 2-methylcyclooctyl, and the like, or multiple ring structures, including bridged ring systems, such as adamantyl, and the like.

[0094] “Cycloalkylalkyl” refers to an alkyl substituted with a cycloalkyl (i.e., cycloalkyl-alkyl- groups), preferably having from 1 to 6 carbon atoms inclusively in the alkyl moiety and from 3 to 12 carbon atoms inclusively in the cycloalkyl moiety. Such cycloalkylalkyl groups are exemplified by cyclopropylmethyl, cyclohexylethyl and the like.

[0095] “Amino” refers to the group -NH₂. Substituted amino refers to the group -NHR^N, NR^NR^N, and NR^NR^NR^N, where each R^N is independently selected from substituted or unsubstituted alkyl, cycloalkyl, cycloheteroalkyl, alkoxy, aryl, heteroaryl, heteroarylalkyl, acyl, alkoxycarbonyl, sulfanyl, sulfinyl, sulfonyl, and the like. Typical amino groups include, but are limited to, dimethylamino,

diethylamino, trimethylammonium, triethylammonium, methylsulfonylamino, furanyl-oxy-sulfamino, and the like.

[0096] “Aminoalkyl” refers to an alkyl group in which one or more of the hydrogen atoms are replaced with one or more amino groups, including substituted amino groups.

[0097] “Aminocarbonyl” refers to $-C(O)NH_2$. Substituted aminocarbonyl refers to $-C(O)NR^nR^n$, where the amino group NR^nR^n is as defined herein.

[0098] “Oxy” refers to a divalent group $-O-$, which may have various substituents to form different oxy groups, including ethers and esters.

[0099] “Alkoxy” or “alkyloxy” are used interchangeably herein to refer to the group $-OR^c$, wherein R^c is an alkyl group, including optionally substituted alkyl groups.

[0100] “Carboxy” refers to $-COOH$.

[0101] “Carbonyl” refers to $-C(O)-$, which may have a variety of substituents to form different carbonyl groups including acids, acid halides, aldehydes, amides, esters, and ketones.

[0102] “Carboxyalkyl” refers to an alkyl in which one or more of the hydrogen atoms are replaced with one or more carboxy groups.

[0103] “Aminocarbonylalkyl” refers to an alkyl substituted with an aminocarbonyl group, as defined herein.

[0104] “Halogen” or “halo” refers to fluoro, chloro, bromo and iodo.

[0105] “Haloalkyl” refers to an alkyl group in which one or more of the hydrogen atoms are replaced with a halogen. Thus, the term “haloalkyl” is meant to include monohaloalkyls, dihaloalkyls, trihaloalkyls, etc. up to perhaloalkyls. For example, the expression “ $(C_1 - C_2)$ haloalkyl” includes 1-fluoromethyl, difluoromethyl, trifluoromethyl, 1-fluoroethyl, 1,1-difluoroethyl, 1,2-difluoroethyl, 1,1,1 trifluoroethyl, perfluoroethyl, etc.

[0106] “Hydroxy” refers to $-OH$.

[0107] “Hydroxyalkyl” refers to an alkyl group in which one or more of the hydrogen atoms are replaced with one or more hydroxy groups.

[0108] “Thiol” or “sulfanyl” refers to $-SH$. Substituted thiol or sulfanyl refers to $-S-R^n$, where R^n is an alkyl, aryl or other suitable substituent.

[0109] “Alkylthio” refers to $-SR^c$, where R^c is an alkyl, which can be optionally substituted. Typical alkylthio group include, but are not limited to, methylthio, ethylthio, n-propylthio, and the like.

[0110] “Alkylthioalkyl” refers to an alkyl substituted with an alkylthio group, $-SR^c$, where R^c is an alkyl, which can be optionally substituted.

[0111] “Sulfonyl” refers to $-SO_2-$. Substituted sulfonyl refers to $-SO_2-R^n$, where R^n is an alkyl, aryl or other suitable substituent.

[0112] "Alkylsulfonyl" refers to $-\text{SO}_2-\text{R}^c$, where R^c is an alkyl, which can be optionally substituted. Typical alkylsulfonyl groups include, but are not limited to, methylsulfonyl, ethylsulfonyl, n-propylsulfonyl, and the like.

[0113] "Alkylsulfonylalkyl" refers to an alkyl substituted with an alkylsulfonyl group, $-\text{SO}_2-\text{R}^c$, where R^c is an alkyl, which can be optionally substituted.

[0114] "Heteroaryl" refers to an aromatic heterocyclic group of from 1 to 10 carbon atoms inclusively and 1 to 4 heteroatoms inclusively selected from oxygen, nitrogen and sulfur within the ring. Such heteroaryl groups can have a single ring (e.g., pyridyl or furyl) or multiple condensed rings (e.g., indolizinyl or benzothienyl).

[0115] "Heteroarylalkyl" refers to an alkyl substituted with a heteroaryl (i.e., heteroaryl-alkyl-groups), preferably having from 1 to 6 carbon atoms inclusively in the alkyl moiety and from 5 to 12 ring atoms inclusively in the heteroaryl moiety. Such heteroarylalkyl groups are exemplified by pyridylmethyl and the like.

[0116] "Heterocycle", "heterocyclic" and interchangeably "heterocycloalkyl" refer to a saturated or unsaturated group having a single ring or multiple condensed rings, from 2 to 10 carbon ring atoms inclusively and from 1 to 4 hetero ring atoms inclusively selected from nitrogen, sulfur or oxygen within the ring. Such heterocyclic groups can have a single ring (e.g., piperidinyl or tetrahydrofuryl) or multiple condensed rings (e.g., indolinyl, dihydrobenzofuran or quinuclidinyl). Examples of heterocycles include, but are not limited to, furan, thiophene, thiazole, oxazole, pyrrole, imidazole, pyrazole, pyridine, pyrazine, pyrimidine, pyridazine, indolizine, isoindole, indole, indazole, purine, quinolizine, isoquinoline, quinoline, phthalazine, naphthylpyridine, quinoxaline, quinazoline, cinnoline, pteridine, carbazole, carboline, phenanthridine, acridine, phenanthroline, isothiazole, phenazine, isoxazole, phenoxazine, phenothiazine, imidazolidine, imidazoline, piperidine, piperazine, pyrrolidine, indoline and the like.

[0117] "Heterocycloalkylalkyl" refers to an alkyl substituted with a heterocycloalkyl (i.e., heterocycloalkyl-alkyl-groups), preferably having from 1 to 6 carbon atoms inclusively in the alkyl moiety and from 3 to 12 ring atoms inclusively in the heterocycloalkyl moiety.

[0118] "Membered ring" is meant to embrace any cyclic structure. The number preceding the term "membered" denotes the number of skeletal atoms that constitute the ring. Thus, for example, cyclohexyl, pyridine, pyran and thiopyran are 6-membered rings and cyclopentyl, pyrrole, furan, and thiophene are 5-membered rings.

[0119] "Fused bicyclic ring" as used herein refers to both unsubstituted and substituted carbocyclic and/or heterocyclic ring moieties having 5 to 8 atoms in each ring, the rings having 2 common atoms.

[0120] Unless otherwise specified, positions occupied by hydrogen in the foregoing groups can be further substituted with substituents exemplified by, but not limited to, hydroxy, oxo, nitro, methoxy, ethoxy, alkoxy, substituted alkoxy, trifluoromethoxy, haloalkoxy, fluoro, chloro, bromo, iodo, halo,

methyl, ethyl, propyl, butyl, alkyl, alkenyl, alkynyl, substituted alkyl, trifluoromethyl, haloalkyl, hydroxyalkyl, alkoxyalkyl, thio, alkylthio, acyl, carboxy, alkoxycarbonyl, carboxamido, substituted carboxamido, alkylsulfonyl, alkylsulfinyl, alkylsulfonylamino, sulfonamido, substituted sulfonamido, cyano, amino, substituted amino, alkylamino, dialkylamino, aminoalkyl, acylamino, amidino, amidoximo, hydroxamoyl, phenyl, aryl, substituted aryl, aryloxy, arylalkyl, arylalkenyl, arylalkynyl, pyridyl, imidazolyl, heteroaryl, substituted heteroaryl, heteroaryloxy, heteroarylalkyl, heteroarylalkenyl, heteroarylalkynyl, cyclopropyl, cyclobutyl, cyclopentyl, cyclohexyl, cycloalkyl, cycloalkenyl, cycloalkylalkyl, substituted cycloalkyl, cycloalkyloxy, pyrrolidinyl, piperidinyl, morpholino, heterocycle, (heterocycle)oxy, and (heterocycle)alkyl; and preferred heteroatoms are oxygen, nitrogen, and sulfur. It is understood that where open valences exist on these substituents they can be further substituted with alkyl, cycloalkyl, aryl, heteroaryl, and/or heterocycle groups, that where these open valences exist on carbon they can be further substituted by halogen and by oxygen-, nitrogen-, or sulfur-bonded substituents, and where multiple such open valences exist, these groups can be joined to form a ring, either by direct formation of a bond or by formation of bonds to a new heteroatom, preferably oxygen, nitrogen, or sulfur. It is further understood that the above substitutions can be made provided that replacing the hydrogen with the substituent does not introduce unacceptable instability to the molecules of the present invention, and is otherwise chemically reasonable.

[0121] As used herein, “optional” and “optionally” mean that the subsequently described event or circumstance may or may not occur, and that the description includes instances where the event or circumstance occurs and instances in which it does not. One of ordinary skill in the art would understand that with respect to any molecule described as containing one or more optional substituents, only sterically practical and/or synthetically feasible compounds are meant to be included. “Optionally substituted” refers to all subsequent modifiers in a term or series of chemical groups. For example, in the term “optionally substituted arylalkyl”, the “alkyl” portion and the “aryl” portion of the molecule may or may not be substituted, and for the series “optionally substituted alkyl, cycloalkyl, aryl and heteroaryl,” the alkyl, cycloalkyl, aryl, and heteroaryl groups, independently of the others, may or may not be substituted.

Glycosylation

[0122] Glycosylation can alter many properties of natural and synthetic products including stability, pharmacodynamics, solubility, and membrane transport. Many molecules, including many secondary metabolites with antimicrobial, antitumor, natural sweetness properties, etc., comprise non-ribosomal peptide, polyketide, or terpenoid backbones modified with β -glycosidic linkages. Many of the diterpene glycosides extracted from the plant, *Stevia rebaudiana* Bertoni, contain β -linked glucose molecules. Naturally, these molecules are glycosylated *in vivo* using UDP-glucose dependent glycosyl transferase enzymes. However, when used *in vitro*, the UDP-glucose can be prohibitively

expensive and/or unavailable. In the present invention a new reaction scheme (See, Figure 2) is provided, in which two different enzyme classes are used to transfer the glucose moiety from α -glucose-1-phosphate to a sample substrate (e.g., stevioside), to produce one or more β -glucose linked products. In some embodiments, the natural stevioside UDP-glucose dependent glycosyltransferase is shown to have activity, using α -glucose-1-phosphate to form rebaudioside A (See, Figure 1). In the some additional embodiments, a laminaribiose phosphorylase acts in the reverse direction to form a β -glucose linked stevioside compound and inorganic phosphate from α -glucose-1-phosphate and stevioside.

[0123] Thus, glycosylation finds use in the production of natural sweeteners, such as those derived from the sweet herb, *Stevia rebaudiana* Bertoni. As indicated above, this plant produces a number of diterpene glycosides which feature high intensity sweetness and sensory properties superior to those of many other high potency sweeteners. The above-mentioned sweet glycosides, have a common aglycone (i.e., steviol), and differ by the number and type of carbohydrate residues at the C13 and C19 positions. Steviol glycosides differ from each other not only in their molecular structure, but also by their taste properties. Usually, stevioside is reported to be 110-270 times sweeter than sucrose, while rebaudioside A is reported to be between 150 and 320 times sweeter than sucrose, and rebaudioside C is reported to be between 40-60 times sweeter than sucrose. Of these common compounds, rebaudioside A has the least astringent, the least bitter, and the least persistent aftertaste. Thus, it has the most favorable sensory attributes of the major steviol glycosides. However, rebaudioside A only constitutes a minor fraction (2-10%) of total glycosides isolated from *Stevia rebaudiana* Bertoni, with other compounds including stevioside (2-10%) and Rebaudioside C (1-2%) making up the rest.

[0124] Changing the glycosylation pattern of some substrates finds use in either simplifying purification and/or to convert less desirable molecules (e.g., stevioside) to more desirable compounds (e.g., rebaudioside A). In some cases, glycosylation is achieved through chemical synthesis methods. However, these methods typically require multiple synthetic steps with undesirable chemicals and processes and can result in mixed products (e.g., with linkages in incorrect positions and/or with undesired anomeric configurations).

[0125] In contrast, glycosylating enzymes can be active under mild conditions and can confer high positional selectivity and stereospecificity in a single step. Many naturally derived glycosylated metabolites are generated *in vivo* using glycosyltransferase (GT) enzymes which transfer sugar moieties from various sugar nucleotides. When used in *in vitro* processes however, the sugar nucleotide donors that these enzymes require can be prohibitively expensive and may not be available at scale. Therefore, there is a need for an enzyme capable of producing glycosylated molecules using less expensive and/or more convenient sugar donors.

Engineered Glycosyltransferase Polypeptides

[0126] The present invention provides polypeptides having glycosyltransferase activity, polynucleotides encoding the polypeptides, methods of preparing the polypeptides, and methods for using the polypeptides. Where the description relates to polypeptides, it is to be understood that it also describes the polynucleotides encoding the polypeptides.

[0127] The suitable reaction conditions under which the above-described improved properties of the engineered polypeptides carry out the transferase reaction can be determined with respect to concentrations or amounts of polypeptide, substrate, co-substrate, transition metal cofactor, reductant, buffer, co-solvent, pH, conditions including temperature and reaction time, and/or conditions with the polypeptide immobilized on a solid support, as further described below and in the Examples.

[0128] In some embodiments, exemplary engineered polypeptides having glycosyltransferase activity with improved properties, particularly in the conversion of steviol glycosides to further glycosylated steviol glycosides, comprise an amino acid sequence that has one or more residue differences as compared to SEQ ID NO: 2, 4, 6, and/or 8.

[0129] Residue differences at these other residue positions can provide for additional variations in the amino acid sequence without adversely affecting the ability of the polypeptide to carry out the transferase reaction. In some embodiments, the sequence further comprises 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-14, 1-15, 1-16, 1-18, 1-20, 1-22, 1-24, 1-26, 1-30, 1-35, 1-40, 1-45, or 1-50 residue differences at other amino acid residue positions as compared to the SEQ ID NO: 2, 4, 6, and/or 8. In some embodiments, the number of amino acid residue differences as compared to the reference sequence can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 30, 35, 40, 45 or 50 residue positions. In some embodiments, the number of amino acid residue differences as compared to the reference sequence can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 20, 21, 22, 23, 24, or 25 residue positions. The residue difference at these other positions can be conservative changes or non-conservative changes. In some embodiments, the residue differences can comprise conservative substitutions and non-conservative substitutions as compared to the naturally occurring glycosyltransferase polypeptide of SEQ ID NO: 2, 4, 6, and/or 8.

[0130] In some embodiments, the present invention also provides engineered polypeptides that comprise a fragment of any of the engineered glycosyltransferase polypeptides described herein that retains the functional activity and/or improved property of that engineered glycosyltransferase. Accordingly, in some embodiments, the present invention provides a polypeptide fragment capable of the transferase reaction under suitable reaction conditions, wherein the fragment comprises at least about 80%, 90%, 95%, 96%, 97%, 98%, or 99% of a full-length amino acid sequence of a glycosyltransferase polypeptide of the present invention, such as the naturally occurring glycosyltransferase polypeptide of 2, 4, 6, and/or 8.

[0131] In some embodiments, the engineered glycosyltransferase polypeptide can have an amino acid sequence comprising a deletion of any one of the glycosyltransferase polypeptides described

herein, such as the naturally occurring glycosyltransferase polypeptide of SEQ ID NO: 2, 4, 6, and/or 8.

[0132] Thus, for each and every embodiment of the engineered glycosyltransferase polypeptides of the invention, the amino acid sequence can comprise deletions of one or more amino acids, 2 or more amino acids, 3 or more amino acids, 4 or more amino acids, 5 or more amino acids, 6 or more amino acids, 8 or more amino acids, 10 or more amino acids, 15 or more amino acids, or 20 or more amino acids, up to 10% of the total number of amino acids, up to 10% of the total number of amino acids, up to 20% of the total number of amino acids, or up to 30% of the total number of amino acids of the glycosyltransferase polypeptides, where the associated functional activity and/or improved properties of the engineered glycosyltransferase described herein are maintained. In some embodiments, the deletions can comprise 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-15, 1-20, 1-21, 1-22, 1-23, 1-24, 1-25, 1-30, 1-35, 1-40, 1-45, or 1-50 amino acid residues. In some embodiments, the number of deletions can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, or 50 amino acid residues. In some embodiments, the deletions can comprise deletions of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 20, 21, 22, 23, 24, or 25 amino acid residues.

[0133] In some embodiments, the engineered glycosyltransferase polypeptide herein can have an amino acid sequence comprising an insertion as compared to any one of the glycosyltransferase polypeptides described herein, such as the naturally occurring glycosyltransferase polypeptide of SEQ ID NO: 2, 4, 6, and/or 8. Thus, some embodiments of the glycosyltransferase polypeptides of the present invention, the insertions can comprise one or more amino acids, 2 or more amino acids, 3 or more amino acids, 4 or more amino acids, 5 or more amino acids, 6 or more amino acids, 8 or more amino acids, 10 or more amino acids, 15 or more amino acids, 20 or more amino acids, 30 or more amino acids, 40 or more amino acids, or 50 or more amino acids, where the associated functional activity and/or improved properties of the engineered glycosyltransferase described herein is maintained. The insertions can be to amino or carboxy terminus, or internal portions of the glycosyltransferase polypeptide.

[0134] In some embodiments, the amino acid sequence has optionally 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-15, 1-20, 1-21, 1-22, 1-23, 1-24, 1-25, 1-30, 1-35, 1-40, 1-45, or 1-50 amino acid residue deletions, insertions and/or substitutions. In some embodiments, the number of amino acid sequence has optionally 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, or 50 amino acid residue deletions, insertions and/or substitutions. In some embodiments, the amino acid sequence has optionally 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 20, 21, 22, 23, 24, or 25 amino acid residue deletions, insertions and/or substitutions. In some embodiments, the substitutions can be conservative or non-conservative substitutions.

[0135] In the above embodiments, the suitable reaction conditions for the engineered polypeptides are provided in the Examples.

[0136] In some embodiments, the polypeptides of the present invention are fusion polypeptides in which the engineered polypeptides are fused to other polypeptides, such as, by way of example and not limitation, antibody tags (e.g., myc epitope), purification sequences (e.g., His tags for binding to metals), and cell localization signals (e.g., secretion signals). Thus, the engineered polypeptides described herein can be used with or without fusions to other polypeptides.

[0137] It is to be understood that the polypeptides described herein are not restricted to the genetically encoded amino acids. In addition to the genetically encoded amino acids, the polypeptides described herein may be comprised, either in whole or in part, of naturally occurring and/or synthetic non-encoded amino acids. Certain commonly encountered non-encoded amino acids of which the polypeptides described herein may be comprised include, but are not limited to: the D-stereoisomers of the genetically-encoded amino acids; 2,3-diaminopropionic acid (Dpr); α -aminoisobutyric acid (Aib); ϵ -aminohexanoic acid (Aha); δ -aminovaleric acid (Ava); N-methylglycine or sarcosine (MeGly or Sar); ornithine (Orn); citrulline (Cit); t-butylalanine (Bua); t-butylglycine (Bug); N-methylisoleucine (MeIle); phenylglycine (Phg); cyclohexylalanine (Cha); norleucine (Nle); naphthylalanine (Nal); 2-chlorophenylalanine (Ocf); 3-chlorophenylalanine (Mcf); 4-chlorophenylalanine (Pcf); 2-fluorophenylalanine (Off); 3-fluorophenylalanine (Mff); 4-fluorophenylalanine (Pff); 2-bromophenylalanine (Obf); 3-bromophenylalanine (Mbf); 4-bromophenylalanine (Pbf); 2-methylphenylalanine (Omef); 3-methylphenylalanine (Mmf); 4-methylphenylalanine (Pmf); 2-nitrophenylalanine (Onf); 3-nitrophenylalanine (Mnf); 4-nitrophenylalanine (Pnf); 2-cyanophenylalanine (Ocf); 3-cyanophenylalanine (Mcf); 4-cyanophenylalanine (Pcf); 2-trifluoromethylphenylalanine (Otf); 3-trifluoromethylphenylalanine (Mtf); 4-trifluoromethylphenylalanine (Ptf); 4-aminophenylalanine (Paf); 4-iodophenylalanine (Pif); 4-aminomethylphenylalanine (Pamf); 2,4-dichlorophenylalanine (Opef); 3,4-dichlorophenylalanine (Mpcef); 2,4-difluorophenylalanine (Opff); 3,4-difluorophenylalanine (Mpff); pyrid-2-ylalanine (2pAla); pyrid-3-ylalanine (3pAla); pyrid-4-ylalanine (4pAla); naphth-1-ylalanine (1nAla); naphth-2-ylalanine (2nAla); thiazolylalanine (taAla); benzothienylalanine (bAla); thienylalanine (tAla); furylalanine (fAla); homophenylalanine (hPhe); homotyrosine (hTyr); homotryptophan (hTrp); pentafluorophenylalanine (5ff); styrylalanine (sAla); authrylalanine (aAla); 3,3-diphenylalanine (Dfa); 3-amino-5-phenylpentanoic acid (Afp); penicillamine (Pen); 1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid (Tic); β -2-thienylalanine (Thi); methionine sulfoxide (Mso); N(w)-nitroarginine (nArg); homolysine (hLys); phosphonomethylphenylalanine (pmPhe); phosphoserine (pSer); phosphothreonine (pThr); homoaspartic acid (hAsp); homoglutamic acid (hGlu); 1-aminocyclopent-(2 or 3)-ene-4 carboxylic acid; pipecolic acid (PA), azetidine-3-carboxylic acid (ACA); 1-aminocyclopentane-3-carboxylic acid; allylglycine (aGly); propargylglycine (pgGly); homoalanine (hAla); norvaline (nVal); homoleucine (hLeu), homovaline (hVal); homoisoleucine (hIle); homoarginine (hArg); N-acetyl lysine (AcLys); 2,4-diaminobutyric acid (Dbu); 2,3-diaminobutyric

acid (Dab); N-methylvaline (MeVal); homocysteine (hCys); homoserine (hSer); hydroxyproline (Hyp) and homoproline (hPro). Additional non-encoded amino acids of which the polypeptides described herein may be comprised will be apparent to those of skill in the art (See e.g., the various amino acids provided in Fasman, CRC Practical Handbook of Biochemistry and Molecular Biology, CRC Press, Boca Raton, FL, pp. 3-70 [1989], and the references cited therein, all of which are incorporated by reference). These amino acids may be in either the L- or D-configuration.

[0138] Those of skill in the art will recognize that amino acids or residues bearing side chain protecting groups may also comprise the polypeptides described herein. Non-limiting examples of such protected amino acids, which in this case belong to the aromatic category, include (protecting groups listed in parentheses), but are not limited to: Arg(tos), Cys(methylbenzyl), Cys(nitropyridinesulfenyl), Glu(δ -benzylester), Gln(xanthyl), Asn(N- δ -xanthyl), His(bom), His(benzyl), His(tos), Lys(fmoc), Lys(tos), Ser(O-benzyl), Thr (O-benzyl) and Tyr(O-benzyl).

[0139] Non-encoding amino acids that are conformationally constrained of which the polypeptides described herein may be composed include, but are not limited to, N-methyl amino acids (L-configuration); 1-aminocyclopent-(2 or 3)-ene-4-carboxylic acid; pipercolic acid; azetidine-3-carboxylic acid; homoproline (hPro); and 1-aminocyclopentane-3-carboxylic acid.

[0140] In some embodiments, the engineered polypeptides are in various forms, for example, such as an isolated preparation, as a substantially purified enzyme, whole cells transformed with gene(s) encoding the enzyme, and/or as cell extracts and/or lysates of such cells. The enzymes can be lyophilized, spray-dried, precipitated or be in the form of a crude paste, as further discussed below.

[0141] In some embodiments, the engineered polypeptides are provided on a solid support, such as a membrane, resin, solid carrier, or other solid phase material. A solid support can be composed of organic polymers such as polystyrene, polyethylene, polypropylene, polyfluoroethylene, polyethyleneoxy, and polyacrylamide, as well as co-polymers and grafts thereof. A solid support can also be inorganic, such as glass, silica, controlled pore glass (CPG), reverse phase silica or metal, such as gold or platinum. The configuration of a solid support can be in the form of beads, spheres, particles, granules, a gel, a membrane or a surface. Surfaces can be planar, substantially planar, or non-planar. Solid supports can be porous or non-porous, and can have swelling or non-swelling characteristics. A solid support can be configured in the form of a well, depression, or other container, vessel, feature, or location.

[0142] In some embodiments, the engineered polypeptides having glycosyltransferase activity of the present invention can be immobilized on a solid support such that they retain their improved activity, stereoselectivity, and/or other improved properties relative to the reference polypeptide of SEQ ID NO: 2, 4, 6, and/or 8. In such embodiments, the immobilized polypeptides can facilitate the biocatalytic conversion of the substrate compounds or other suitable substrates to the product and after the reaction is complete are easily retained (e.g., by retaining beads on which polypeptide is immobilized) and then reused or recycled in subsequent reactions. Such immobilized enzyme

processes allow for further efficiency and cost reduction. Accordingly, it is further contemplated that any of the methods of using the glycosyltransferase polypeptides of the present invention can be carried out using the same glycosyltransferase polypeptides bound or immobilized on a solid support. [0143] Methods of enzyme immobilization are well-known in the art. The engineered polypeptides can be bound non-covalently or covalently. Various methods for conjugation and immobilization of enzymes to solid supports (e.g., resins, membranes, beads, glass, etc.) are well known in the art (See e.g., Yi et al., *Proc. Biochem.*, 42(5): 895-898 [2007]; Martin et al., *Appl. Microbiol. Biotechnol.*, 76(4): 843-851 [2007]; Koszelewski et al., *J. Mol. Cat. B: Enzymatic*, 63: 39-44 [2010]; Truppo et al., *Org. Proc. Res. Dev.*, published online: [dx.doi.org/10.1021/op200157c](https://doi.org/10.1021/op200157c); Hermanson, *Bioconjugate Techniques*, 2nd ed., Academic Press, Cambridge, MA [2008]; Mateo et al., *Biotechnol. Prog.*, 18(3):629-34 [2002]; and "Bioconjugation Protocols: Strategies and Methods," In *Methods in Molecular Biology*, Niemeyer (ed.), Humana Press, New York, NY [2004]; the disclosures of each which are incorporated by reference herein). Solid supports useful for immobilizing the engineered glycosyltransferases of the present invention include but are not limited to beads or resins comprising polymethacrylate with epoxide functional groups, polymethacrylate with amino epoxide functional groups, styrene/DVB copolymer or polymethacrylate with octadecyl functional groups. Exemplary solid supports useful for immobilizing the engineered glycosyltransferase polypeptides of the present invention include, but are not limited to, chitosan beads, Eupergit C, and SEPABEADs (Mitsubishi), including the following different types of SEPABEAD: EC-EP, EC-HFA/S, EXA252, EXE119 and EXE120.

[0144] In some embodiments, the polypeptides described herein are provided in the form of kits. The enzymes in the kits may be present individually or as a plurality of enzymes. The kits can further include reagents for carrying out the enzymatic reactions, substrates for assessing the activity of enzymes, as well as reagents for detecting the products. The kits can also include reagent dispensers and instructions for use of the kits.

[0145] In some embodiments, the kits of the present invention include arrays comprising a plurality of different glycosyltransferase polypeptides at different addressable position, wherein the different polypeptides are different variants of a reference sequence each having at least one different improved enzyme property. In some embodiments, a plurality of polypeptides immobilized on solid supports are configured on an array at various locations, addressable for robotic delivery of reagents, or by detection methods and/or instruments. The array can be used to test a variety of substrate compounds for conversion by the polypeptides. Such arrays comprising a plurality of engineered polypeptides and methods of their use are known in the art (See e.g., WO2009/008908A2).

Polynucleotides Encoding Engineered Glycosyltransferases, Expression Vectors and Host Cells

[0146] In another aspect, the present invention provides polynucleotides encoding the engineered glycosyltransferase polypeptides described herein. The polynucleotides may be operatively linked to

one or more heterologous regulatory sequences that control gene expression to create a recombinant polynucleotide capable of expressing the polypeptide. Expression constructs containing a heterologous polynucleotide encoding the engineered glycosyltransferase are introduced into appropriate host cells to express the corresponding glycosyltransferase polypeptide.

[0147] As will be apparent to the skilled artisan, availability of a protein sequence and the knowledge of the codons corresponding to the various amino acids provide a description of all the polynucleotides capable of encoding the subject polypeptides. The degeneracy of the genetic code, where the same amino acids are encoded by alternative or synonymous codons, allows an extremely large number of nucleic acids to be made, all of which encode the improved glycosyltransferase enzymes. Thus, having knowledge of a particular amino acid sequence, those skilled in the art could make any number of different nucleic acids by simply modifying the sequence of one or more codons in a way which does not change the amino acid sequence of the protein. In this regard, the present invention specifically contemplates each and every possible variation of polynucleotides that could be made encoding the polypeptides described herein by selecting combinations based on the possible codon choices, and all such variations are to be considered specifically disclosed for any polypeptide described herein, including the amino acid sequences disclosed in the sequence listing incorporated by reference herein as the even-numbered sequences in SEQ ID NOS: 1-32.

[0148] In various embodiments, the codons are preferably selected to fit the host cell in which the protein is being produced. For example, preferred codons used in bacteria are used to express the gene in bacteria; preferred codons used in yeast are used for expression in yeast; and preferred codons used in mammals are used for expression in mammalian cells. In some embodiments, all codons need not be replaced to optimize the codon usage of the glycosyltransferases since the natural sequence will comprise preferred codons and because use of preferred codons may not be required for all amino acid residues. Consequently, codon optimized polynucleotides encoding the glycosyltransferase enzymes may contain preferred codons at about 40%, 50%, 60%, 70%, 80%, or greater than 90% of codon positions of the full length coding region.

[0149] In some embodiments, the polynucleotide comprises a codon optimized nucleotide sequence encoding the naturally occurring glycosyltransferase polypeptide amino acid sequence, as represented by SEQ ID NO: 2, 4, 6, and/or 8. In some embodiments, the polynucleotide has a nucleic acid sequence comprising at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identity to the codon optimized nucleic acid sequences encoding the even-numbered sequences in SEQ ID NOS: 1-32. In some embodiments, the polynucleotide has a nucleic acid sequence comprising at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identity to the codon optimized nucleic acid sequences in the odd-numbered sequences in SEQ ID NOS: 1-32. The codon optimized sequences of the odd-numbered sequences in SEQ ID NOS: 1-32, enhance expression of the encoded, wild-type glycosyltransferase, providing preparations of enzyme capable of the transferase activity described herein. In some embodiments, the codon optimized

polynucleotide sequence enhances expression of the glycosyltransferase by at least 1.2 fold, 1.5 fold or 2 fold or greater as compared to the naturally occurring polynucleotide sequence from *Stevia rebaudiana*, *Streptomyces resistomycificus*, *Streptomyces antibioticus* and/or *Paenibacillus* sp. YM1.

[0150] In some embodiments, the polynucleotides are capable of hybridizing under highly stringent conditions to a reference sequence selected from the odd-numbered sequences in SEQ ID NOS: 1-32, or a complement thereof, and encodes a polypeptide having glycosyltransferase activity.

[0151] In some embodiments, as described above, the polynucleotide encodes an engineered polypeptide having glycosyltransferase activity with improved properties as compared to SEQ ID NO: 2, 4, 6, and/or 8, where the polypeptide comprises an amino acid sequence having at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identity to a reference sequence, and one or more residue differences as compared to a sequence selected from the even-numbered sequences in SEQ ID NOS: 1-32. In some embodiments, the reference amino acid sequence is selected from the even-numbered sequences in SEQ ID NOS: 1-32. In some embodiments, the reference amino acid sequence is SEQ ID NO: 2. In some embodiments, the reference amino acid sequence is SEQ ID NO: 4. In some further embodiments, the reference amino acid sequence is SEQ ID NO: 8.

[0152] In some embodiments, the polynucleotide encodes a glycosyltransferase polypeptide capable of the transferase reaction provided herein, with improved properties as compared to SEQ ID NO: 2, 4, 6, and/or 8, wherein the polypeptide comprises an amino acid sequence having at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 2, 4, 6, and/or 8, and one or more residue differences as compared to SEQ ID NO: 2, 4, 6, and/or 8.

[0153] In some embodiments, the polynucleotide encodes a glycosyltransferase polypeptide capable of the transferase reactions provided herein, with improved properties as compared to SEQ ID NO: 2, 4, 6, and/or 8, wherein the polypeptide comprises an amino acid sequence having at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 2, 4, 6, and/or 8, and one or more residue differences as compared to SEQ ID NO: 2, 4, 6, and/or 8.

[0154] In some embodiments, the polynucleotide encodes a glycosyltransferase polypeptide capable of the transferase reactions provided herein, with improved properties as compared to SEQ ID NO: 2, 4, 6, and/or 8, wherein the polypeptide comprises an amino acid sequence having at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 2, 4, 6, and/or 8, and one or more residue differences as compared to SEQ ID NO: 2, 4, 6, and/or 8.

[0155] In some embodiments, the polynucleotide encoding the engineered glycosyltransferase comprises a polynucleotide sequence selected from the odd-numbered sequences in SEQ ID NOS: 1-32.

[0156] In some embodiments, the polynucleotides are capable of hybridizing under highly stringent conditions to a reference polynucleotide sequence selected from the odd-numbered sequences in SEQ ID NOS: 1-32, or a complement thereof, and encodes a polypeptide having glycosyltransferase activity with one or more of the improved properties described herein. In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes a glycosyltransferase polypeptide comprising an amino acid sequence having at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identity to SEQ ID NO: 2, 4, 6, and/or 8, that has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NO: 2, 4, 6, and/or 8.

[0157] In some embodiments, an isolated polynucleotide encoding any of the engineered glycosyltransferase polypeptides provided herein is manipulated in a variety of ways to provide for expression of the polypeptide. In some embodiments, the polynucleotides encoding the polypeptides are provided as expression vectors where one or more control sequences is present to regulate the expression of the polynucleotides and/or polypeptides. Manipulation of the isolated polynucleotide prior to its insertion into a vector may be desirable or necessary depending on the expression vector. The techniques for modifying polynucleotides and nucleic acid sequences utilizing recombinant DNA methods are well known in the art.

[0158] In some embodiments, the control sequences include among other sequences, promoters, leader sequences, polyadenylation sequences, propeptide sequences, signal peptide sequences, and transcription terminators. As known in the art, suitable promoters can be selected based on the host cells used. For bacterial host cells, suitable promoters for directing transcription of the nucleic acid constructs of the present application, include, but are not limited to the promoters obtained from the *E. coli* lac operon, *Streptomyces coelicolor* agarase gene (*dagA*), *Bacillus subtilis* levansucrase gene (*sacB*), *Bacillus licheniformis* alpha-amylase gene (*amyL*), *Bacillus stearothermophilus* maltogenic amylase gene (*amyM*), *Bacillus amyloliquefaciens* alpha-amylase gene (*amyQ*), *Bacillus licheniformis* penicillinase gene (*penP*), *Bacillus subtilis* *xylA* and *xylB* genes, and prokaryotic beta-lactamase gene (See e.g., Villa-Kamaroff et al., Proc. Natl Acad. Sci. USA 75: 3727-3731 [1978]), as well as the *tac* promoter (See e.g., DeBoer et al., Proc. Natl Acad. Sci. USA 80: 21-25 [1983]). Exemplary promoters for filamentous fungal host cells, include promoters obtained from the genes for *Aspergillus oryzae* TKA amylase, *Rhizomucor miehei* aspartic proteinase, *Aspergillus niger* neutral alpha-amylase, *Aspergillus niger* acid stable alpha-amylase, *Aspergillus niger* or *Aspergillus awamori* glucoamylase (*glaA*), *Rhizomucor miehei* lipase, *Aspergillus oryzae* alkaline protease, *Aspergillus oryzae* triose phosphate isomerase, *Aspergillus nidulans* acetamidase, and *Fusarium oxysporum* trypsin-like protease (See e.g., WO 96/00787), as well as the NA2-tpi promoter (a hybrid of the promoters from the genes for *Aspergillus niger* neutral alpha-amylase and *Aspergillus oryzae* triose phosphate isomerase), and mutant, truncated, and hybrid promoters thereof. Exemplary yeast cell promoters can be from any suitable source. In some embodiments, the genes comprise the genes for

Saccharomyces cerevisiae enolase (ENO-1), *Saccharomyces cerevisiae* galactokinase (GAL1), *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP), and *Saccharomyces cerevisiae* 3-phosphoglycerate kinase. Other useful promoters for yeast host cells are known in the art (See e.g., Romanos et al., *Yeast* 8:423-488 [1992]).

[0159] In some embodiments, the control sequence is a suitable transcription terminator sequence, a sequence recognized by a host cell to terminate transcription. The terminator sequence is operably linked to the 3' terminus of the nucleic acid sequence encoding the polypeptide. Any terminator which is functional in the host cell of choice finds use in the present invention. For example, exemplary transcription terminators for filamentous fungal host cells can be obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Aspergillus niger* alpha-glucosidase, and *Fusarium oxysporum* trypsin-like protease. Exemplary terminators for yeast host cells can be obtained from the genes for *Saccharomyces cerevisiae* enolase, *Saccharomyces cerevisiae* cytochrome C (CYC1), and *Saccharomyces cerevisiae* glyceraldehyde-3-phosphate dehydrogenase. Various other useful terminators for yeast host cells are known in the art (See e.g., Romanos et al., *supra*).

[0160] In some embodiments, the control sequence is a suitable leader sequence, a non-translated region of an mRNA that is important for translation by the host cell. The leader sequence is operably linked to the 5' terminus of the nucleic acid sequence encoding the polypeptide. Any leader sequence that is functional in the host cell of choice may be used. Exemplary leaders for filamentous fungal host cells are obtained from the genes for *Aspergillus oryzae* TAKA amylase and *Aspergillus nidulans* triose phosphate isomerase. Suitable leaders for yeast host cells include, but are not limited to those obtained from the genes for *Saccharomyces cerevisiae* enolase (ENO-1), *Saccharomyces cerevisiae* 3-phosphoglycerate kinase, *Saccharomyces cerevisiae* alpha-factor, and *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP).

[0161] The control sequence may also be a polyadenylation sequence, a sequence operably linked to the 3' terminus of the nucleic acid sequence and which, when transcribed, is recognized by the host cell as a signal to add polyadenosine residues to transcribed mRNA. Any polyadenylation sequence which is functional in the host cell of choice may be used in the present invention. Exemplary polyadenylation sequences for filamentous fungal host cells include, but are not limited to those from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Fusarium oxysporum* trypsin-like protease, and *Aspergillus niger* alpha-glucosidase. Useful polyadenylation sequences for yeast host cells are also known in the art (See e.g., Guo and Sherman, *Mol. Cell. Bio.*, 15:5983-5990 [1995]).

[0162] In some embodiments, the control sequence is a signal peptide coding region that codes for an amino acid sequence linked to the amino terminus of a polypeptide and directs the encoded polypeptide into the cell's secretory pathway. The 5' end of the coding sequence of the nucleic acid sequence may inherently contain a signal peptide coding region naturally linked in translation reading

frame with the segment of the coding region that encodes the secreted polypeptide. Alternatively, the 5' end of the coding sequence may contain a signal peptide coding region that is foreign to the coding sequence. Any signal peptide coding region that directs the expressed polypeptide into the secretory pathway of a host cell of choice finds use for expression of the engineered glycosyltransferase polypeptides provided herein. Effective signal peptide coding regions for bacterial host cells include, but are not limited to the signal peptide coding regions obtained from the genes for *Bacillus* NCIB 11837 maltogenic amylase, *Bacillus stearothermophilus* alpha-amylase, *Bacillus licheniformis* subtilisin, *Bacillus licheniformis* beta-lactamase, *Bacillus stearothermophilus* neutral proteases (nprT, nprS, nprM), and *Bacillus subtilis* prsA. Further signal peptides are known in the art (See e.g., Simonen and Palva, Microbiol. Rev., 57:109-137 [1993]). Effective signal peptide coding regions for filamentous fungal host cells include, but are not limited to the signal peptide coding regions obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* neutral amylase, *Aspergillus niger* glucoamylase, *Rhizomucor miehei* aspartic proteinase, *Humicola insolens* cellulase, and *Humicola lanuginosa* lipase. Useful signal peptides for yeast host cells include, but are not limited to those from the genes for *Saccharomyces cerevisiae* alpha-factor and *Saccharomyces cerevisiae* invertase.

[0163] In some embodiments, the control sequence is a propeptide coding region that codes for an amino acid sequence positioned at the amino terminus of a polypeptide. The resultant polypeptide is referred to as a “proenzyme,” “propeptide,” or “zymogen,” in some cases). A propeptide can be converted to a mature active polypeptide by catalytic or autocatalytic cleavage of the propeptide from the propeptide. The propeptide coding region includes, but is not limited to the genes for *Bacillus subtilis* alkaline protease (aprE), *Bacillus subtilis* neutral protease (nprT), *Saccharomyces cerevisiae* alpha-factor, *Rhizomucor miehei* aspartic proteinase, and *Myceliophthora thermophila* lactase (See e.g., WO 95/33836). Where both signal peptide and propeptide regions are present at the amino terminus of a polypeptide, the propeptide region is positioned next to the amino terminus of a polypeptide and the signal peptide region is positioned next to the amino terminus of the propeptide region.

[0164] In some embodiments, regulatory sequences are also utilized. These sequences facilitate the regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory systems are those which cause the expression of the gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. In prokaryotic host cells, suitable regulatory sequences include, but are not limited to the lac, tac, and trp operator systems. In yeast host cells, suitable regulatory systems include, but are not limited to the ADH2 system or GAL1 system. In filamentous fungi, suitable regulatory sequences include, but are not limited to the TAKA alpha-amylase promoter, *Aspergillus niger* glucoamylase promoter, and *Aspergillus oryzae* glucoamylase promoter.

[0165] In another aspect, the present invention also provides a recombinant expression vector comprising a polynucleotide encoding an engineered glycosyltransferase polypeptide, and one or more expression regulating regions such as a promoter and a terminator, a replication origin, etc., depending on the type of hosts into which they are to be introduced. In some embodiments, the various nucleic acid and control sequences described above are joined together to produce a recombinant expression vector which includes one or more convenient restriction sites to allow for insertion or substitution of the nucleic acid sequence encoding the variant glycosyltransferase polypeptide at such sites. Alternatively, the polynucleotide sequence(s) of the present invention are expressed by inserting the polynucleotide sequence or a nucleic acid construct comprising the polynucleotide sequence into an appropriate vector for expression. In creating the expression vector, the coding sequence is located in the vector so that the coding sequence is operably linked with the appropriate control sequences for expression.

[0166] The recombinant expression vector may be any vector (e.g., a plasmid or virus), that can be conveniently subjected to recombinant DNA procedures and can result in the expression of the variant glycosyltransferase polynucleotide sequence. The choice of the vector will typically depend on the compatibility of the vector with the host cell into which the vector is to be introduced. The vectors may be linear or closed circular plasmids.

[0167] In some embodiments, the expression vector is an autonomously replicating vector (i.e., a vector that exists as an extra-chromosomal entity, the replication of which is independent of chromosomal replication, such as a plasmid, an extra-chromosomal element, a minichromosome, or an artificial chromosome). The vector may contain any means for assuring self-replication. In some alternative embodiments, the vector may be one which, when introduced into the host cell, is integrated into the genome and replicated together with the chromosome(s) into which it has been integrated. Furthermore, a single vector or plasmid or two or more vectors or plasmids which together contain the total DNA to be introduced into the genome of the host cell, or a transposon may be used.

[0168] In some embodiments, the expression vector preferably contains one or more selectable markers, which permit easy selection of transformed cells. A "selectable marker" is a gene the product of which provides for biocide or viral resistance, resistance to heavy metals, prototrophy to auxotrophy, and the like. Examples of bacterial selectable markers include, but are not limited to the *lac* genes from *Bacillus subtilis* or *Bacillus licheniformis*, or markers, which confer antibiotic resistance such as ampicillin, kanamycin, chloramphenicol or tetracycline resistance. Suitable markers for yeast host cells include, but are not limited to ADE2, HIS3, LEU2, LYS2, MET3, TRP1, and URA3. Selectable markers for use in a filamentous fungal host cell include, but are not limited to, *amdS* (acetamidase), *argB* (ornithine carbamoyltransferases), *bar* (phosphinothricin acetyltransferase), *hph* (hygromycin phosphotransferase), *niaD* (nitrate reductase), *pyrG* (orotidine-5'-phosphate decarboxylase), *sC* (sulfate adenylyltransferase), and *trpC* (anthranilate synthase), as well as equivalents thereof. In another aspect, the present invention provides a host cell comprising a polynucleotide

encoding at least one engineered glycosyltransferase polypeptide of the present application, the polynucleotide being operatively linked to one or more control sequences for expression of the engineered glycosyltransferase enzyme(s) in the host cell. Host cells for use in expressing the polypeptides encoded by the expression vectors of the present invention are well known in the art and include but are not limited to, bacterial cells, such as *E. coli*, *Vibrio fluvialis*, *Streptomyces* and *Salmonella typhimurium* cells; fungal cells, such as yeast cells (e.g., *Saccharomyces cerevisiae* and *Pichia pastoris* [ATCC Accession No. 201178]); insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, BHK, 293, and Bowes melanoma cells; and plant cells. Exemplary host cells are *Escherichia coli* strains (e.g., W3110 (Δ fhuA) and BL21).

[0169] Accordingly, in another aspect, the present invention provides methods for producing the engineered glycosyltransferase polypeptides, where the methods comprise culturing a host cell capable of expressing a polynucleotide encoding the engineered glycosyltransferase polypeptide under conditions suitable for expression of the polypeptide. In some embodiments, the methods further comprise the steps of isolating and/or purifying the glycosyltransferase polypeptides, as described herein.

[0170] Appropriate culture media and growth conditions for the above-described host cells are well known in the art. Polynucleotides for expression of the glycosyltransferase polypeptides may be introduced into cells by various methods known in the art. Techniques include, among others, electroporation, biolistic particle bombardment, liposome mediated transfection, calcium chloride transfection, and protoplast fusion.

[0171] The engineered glycosyltransferase with the properties disclosed herein can be obtained by subjecting the polynucleotide encoding the naturally occurring or engineered glycosyltransferase polypeptide to mutagenesis and/or directed evolution methods known in the art, and as described herein. An exemplary directed evolution technique is mutagenesis and/or DNA shuffling (See e.g., Stemmer, Proc. Natl. Acad. Sci. USA 91:10747-10751 [1994]; WO 95/22625; WO 97/0078; WO 97/35966; WO 98/27230; WO 00/42651; WO 01/75767 and U.S. Pat. 6,537,746). Other directed evolution procedures that can be used include, among others, staggered extension process (StEP), in vitro recombination (See e.g., Zhao et al., Nat. Biotechnol., 16:258-261 [1998]), mutagenic PCR (See e.g., Caldwell et al., PCR Methods Appl., 3:S136-S140 [1994]), and cassette mutagenesis (See e.g., Black et al., Proc. Natl. Acad. Sci. USA 93:3525-3529 [1996]).

[0172] For example, mutagenesis and directed evolution methods can be readily applied to polynucleotides to generate variant libraries that can be expressed, screened, and assayed.

Mutagenesis and directed evolution methods are well known in the art (See e.g., US Patent Nos. 5,605,793, 5,830,721, 6,132,970, 6,420,175, 6,277,638, 6,365,408, 6,602,986, 7,288,375, 6,287,861, 6,297,053, 6,576,467, 6,444,468, 5,811,238, 6,117,679, 6,165,793, 6,180,406, 6,291,242, 6,995,017, 6,395,547, 6,506,602, 6,519,065, 6,506,603, 6,413,774, 6,573,098, 6,323,030, 6,344,356, 6,372,497, 7,868,138, 5,834,252, 5,928,905, 6,489,146, 6,096,548, 6,387,702, 6,391,552, 6,358,742, 6,482,647,

6,335,160, 6,653,072, 6,355,484, 6,303,344, 6,319,713, 6,613,514, 6,455,253, 6,579,678, 6,586,182, 6,406,855, 6,946,296, 7,534,564, 7,776,598, 5,837,458, 6,391,640, 6,309,883, 7,105,297, 7,795,030, 6,326,204, 6,251,674, 6,716,631, 6,528,311, 6,287,862, 6,335,198, 6,352,859, 6,379,964, 7,148,054, 7,629,170, 7,620,500, 6,365,377, 6,358,740, 6,406,910, 6,413,745, 6,436,675, 6,961,664, 6,537,746, 7,430,477, 7,873,499, 7,702,464, 7,783,428, 7,747,391, 7,747,393, 7,751,986, 6,376,246, 6,426,224, 6,423,542, 6,479,652, 6,319,714, 6,521,453, 6,368,861, 7,421,347, 7,058,515, 7,024,312, 7,620,502, 7,853,410, 7,957,912, 7,904,249, 8,383,346, 8,504,498, 8,768,871, 8,762,066, 8,849,575, and all related non-US counterparts; Ling et al., *Anal. Biochem.*, 254:157-78 [1997]; Dale et al., *Meth. Mol. Biol.*, 57:369-74 [1996]; Smith, *Ann. Rev. Genet.*, 19:423-462 [1985]; Botstein et al., *Science*, 229:1193-1201 [1985]; Carter, *Biochem. J.*, 237:1-7 [1986]; Kramer et al., *Cell*, 38:879-887 [1984]; Wells et al., *Gene*, 34:315-323 [1985]; Minshull et al., *Curr. Op. Chem. Biol.*, 3:284-290 [1999]; Christians et al., *Nat. Biotechnol.*, 17:259-264 [1999]; Crameri et al., *Nature*, 391:288-291 [1998]; Crameri, et al., *Nat. Biotechnol.*, 15:436-438 [1997]; Zhang et al., *Proc. Nat. Acad. Sci. U.S.A.*, 94:4504-4509 [1997]; Crameri et al., *Nat. Biotechnol.*, 14:315-319 [1996]; Stemmer, *Nature*, 370:389-391 [1994]; Stemmer, *Proc. Nat. Acad. Sci. USA*, 91:10747-10751 [1994]; WO 95/22625; WO 97/0078; WO 97/35966; WO 98/27230; WO 00/42651; WO 01/75767; WO 2009/152336, WO 2009/102901, WO 2009/102899, WO 2011/035105, WO 2013/138339, WO 2013/003290, WO 2014/120819, WO 2014/120821, WO 2015/0134315, and WO 2015/048573, all of which are incorporated herein by reference).

[0173] In some embodiments, the enzyme clones obtained following mutagenesis treatment are screened by subjecting the enzymes to a defined temperature (or other assay conditions, such as testing the enzyme's activity over a broad range of substrates) and measuring the amount of enzyme activity remaining after heat treatments or other assay conditions. Clones containing a polynucleotide encoding a glycosyltransferase polypeptide are then sequenced to identify the nucleotide sequence changes (if any), and used to express the enzyme in a host cell. Measuring enzyme activity from the expression libraries can be performed using any suitable method known in the art (e.g., standard biochemistry techniques, such as HPLC analysis).

[0174] In some embodiments, the clones obtained following mutagenesis treatment can be screened for engineered glycosyltransferases having one or more desired improved enzyme properties (e.g., improved transferase activity). Measuring enzyme activity from the expression libraries can be performed using the standard biochemistry techniques, such as HPLC analysis, LC-MS/MS analysis, and/or derivatization of products (pre or post separation), as known in the art (e.g., using dansyl chloride or OPA; See e.g., Yaegaki et al., *J Chromatogr.* 356(1):163-70 [1986]).

[0175] For engineered polypeptides of known sequence, the polynucleotides encoding the enzyme can be prepared by standard solid-phase methods, according to known synthetic methods. In some embodiments, fragments of up to about 100 bases can be individually synthesized, then joined (e.g., by enzymatic or chemical ligation methods, or polymerase mediated methods) to form any desired

continuous sequence. For example, polynucleotides and oligonucleotides disclosed herein can be prepared by chemical synthesis using the classical phosphoramidite method (See e.g., Beaucage et al., Tetra. Lett., 22:1859-69 [1981]; and Matthes et al., EMBO J., 3:801-05 [1984]), as it is typically practiced in automated synthetic methods. According to the phosphoramidite method, oligonucleotides are synthesized (e.g., in an automatic DNA synthesizer), purified, annealed, ligated and cloned in appropriate vectors.

[0176] Where the sequence of the engineered polypeptide is known, the polynucleotides encoding the enzyme can be prepared by standard solid-phase methods, according to known synthetic methods. In some embodiments, fragments of up to about 100 bases can be individually synthesized, then joined (e.g., by enzymatic or chemical ligation methods, or polymerase mediated methods) to form any desired continuous sequence. For example, polynucleotides and oligonucleotides encoding portions of the glycosyltransferase can be prepared by chemical synthesis as known in the art (e.g., the classical phosphoramidite method of Beaucage et al., Tet. Lett. 22:1859-69 [1981], or the method described by Matthes et al., EMBO J. 3:801-05 [1984]) as typically practiced in automated synthetic methods. According to the phosphoramidite method, oligonucleotides are synthesized (e.g., in an automatic DNA synthesizer), purified, annealed, ligated and cloned in appropriate vectors. In addition, essentially any nucleic acid can be obtained from any of a variety of commercial sources. In some embodiments, additional variations can be created by synthesizing oligonucleotides containing deletions, insertions, and/or substitutions, and combining the oligonucleotides in various permutations to create engineered glycosyltransferases with improved properties.

[0177] Accordingly, in some embodiments, a method for preparing the engineered glycosyltransferases polypeptide comprises: (a) synthesizing a polynucleotide encoding a polypeptide comprising an amino acid sequence having at least about 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more sequence identity to an amino acid sequence selected from SEQ ID NO: 2, 4, 6, and/or 8, and having one or more residue differences as compared to SEQ ID NO: 2, 4, 6, and/or 8; and (b) expressing the glycosyltransferase polypeptide encoded by the polynucleotide.

[0178] Accordingly, in some embodiments, a method for preparing the engineered glycosyltransferases polypeptide comprises: (a) synthesizing a polynucleotide encoding a polypeptide comprising an amino acid sequence having at least about 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more sequence identity to an amino acid sequence selected from SEQ ID NO: 2, 4, 6, and/or 8, and having one or more residue differences as compared to SEQ ID NO: 2, 4, 6, and/or 8; and (b) expressing the glycosyltransferase polypeptide encoded by the polynucleotide.

[0179] In some embodiments of the method, the polynucleotide encodes an engineered glycosyltransferase that has optionally one or several (e.g., up to 3, 4, 5, or up to 10) amino acid residue deletions, insertions and/or substitutions. In some embodiments, the amino acid sequence has

optionally 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-15, 1-20, 1-21, 1-22, 1-23, 1-24, 1-25, 1-30, 1-35, 1-40, 1-45, or 1-50 amino acid residue deletions, insertions and/or substitutions. In some embodiments, the amino acid sequence has optionally 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, or 50 amino acid residue deletions, insertions and/or substitutions. In some embodiments, the amino acid sequence has optionally 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 20, 21, 22, 23, 24, or 25 amino acid residue deletions, insertions and/or substitutions. In some embodiments, the substitutions can be conservative or non-conservative substitutions.

[0180] In some embodiments, any of the engineered glycosyltransferase enzymes expressed in a host cell can be recovered from the cells and/or the culture medium using any one or more of the well known techniques for protein purification, including, among others, lysozyme treatment, sonication, filtration, salting-out, ultra-centrifugation, and chromatography. Suitable solutions for lysing and the high efficiency extraction of proteins from bacteria, such as *E. coli*, are commercially available (e.g., CellLytic BTM, Sigma-Aldrich, St. Louis MO).

[0181] Chromatographic techniques for isolation of the glycosyltransferase polypeptide include, among others, reverse phase chromatography high performance liquid chromatography, ion exchange chromatography, gel electrophoresis, and affinity chromatography. Conditions for purifying a particular enzyme will depend, in part, on factors such as net charge, hydrophobicity, hydrophilicity, molecular weight, molecular shape, etc., and will be apparent to those having skill in the art.

[0182] In some embodiments, affinity techniques may be used to isolate the improved glycosyltransferase enzymes. For affinity chromatography purification, any antibody which specifically binds the glycosyltransferase polypeptide may be used. For the production of antibodies, various host animals, including but not limited to rabbits, mice, rats, etc., may be immunized by injection with a glycosyltransferase polypeptide, or a fragment thereof. The glycosyltransferase polypeptide or fragment may be attached to a suitable carrier, such as BSA, by means of a side chain functional group or linkers attached to a side chain functional group. In some embodiments, the affinity purification can use a specific ligand bound by the glycosyltransferase, such as poly(L-proline) or dye affinity column (See e.g., EP0641862; Stellwagen, "Dye Affinity Chromatography," In Current Protocols in Protein Science, Unit 9.2-9.2.16 [2001]).

Methods for Using Glycosyltransferases to β -Glycosylate Compounds of Interest

[0183] In some embodiments, the methods, processes, and systems provided herein facilitate the conversion of a substrate to a β -glycosylated product. In some embodiments, the substrate stevioside is converted to the β -glucosylated product rebaudioside A. In some embodiments, this glycosylation reaction is catalyzed by an enzyme. In some embodiments, the enzyme is a glycosyltransferase, while in some alternative embodiments the enzyme is a phosphorylase. In some further embodiments, the glycosyltransferase or phosphorylase uses glucose-1-phosphate (e.g., α -glucose-1-phosphate), as a

glucosyl donor. In some additional embodiments, certain moieties of the substrate (e.g., hydroxyl groups), act as a glycosyl (e.g., glucosyl) acceptor. Some non-limiting examples of glycosyltransferases that find use in the present invention include promiscuous bacterial UDP-glucose-dependent glycosyltransferases (e.g., the glycosyltransferase of SEQ NO: 4, 6, or 12), and phosphorylases that possess β -glycosyl cleavage activity to produce α -glucose-1-phosphate (e.g., a laminaribiose phosphorylase such as that of SEQ NO: 8 or 14). In some embodiments, the enzyme contacts the substrate *in vitro*, while in some alternative embodiments, the enzyme contacts the substrate *in vivo*. In still some additional embodiments, the substrate and enzyme are produced and contacted within an engineered host cell.

[0184] The suitable reaction conditions under which the polypeptides carry out the conversion can be determined by those of skill in the art. The Examples provide exemplary reaction conditions, including the concentrations or amounts of polypeptide, substrate, co-substrate (e.g., glucose-1-phosphate), buffer, co-solvent, pH, temperature, and reaction time. For example, in some embodiments, reactions are performed with substrate concentrations up to 5 mM or the solubility limit of the substrate, up to 5 mM cosubstrate, or the solubility limit of the cosubstrate, in 25-100 mM buffer, with 0-20% ethanol, at pH 5-8, at 30-65 °C, for 5 m – 18 h. In some embodiments, reactions may be additionally performed with a co-enzyme and second co-substrate in order to regenerate glucose-1-phosphate *in situ*, which is useful for performing the reaction in a more inexpensive manner. For example, the co-enzyme may be a retaining phosphorylase, such as sucrose phosphorylase (e.g., even numbered SEQ NO: 16-32), and the additional co-substrate may be the appropriate α -linked disaccharide, trisaccharide, or oligosaccharide enzyme substrate, such as sucrose. The co-enzyme may be an inverting phosphorylase, such as cellobiose phosphorylase (E.C. 2.4.1.20), and the additional co-substrate may be the appropriate β -linked disaccharide, trisaccharide, or oligosaccharide enzyme substrate, such as cellobiose. However, it is not intended that the present invention be limited to the specific reactions and conditions as set forth in the Examples. Various features and embodiments of the invention are illustrated in the following representative examples, which are intended to be illustrative, and not limiting.

EXPERIMENTAL

[0185] The following Examples, including experiments and results achieved, are provided for illustrative purposes only and are not to be construed as limiting the present invention.

[0186] In the experimental disclosure below, the following abbreviations apply: M (molar); mM (millimolar), uM and μ M (micromolar); nM (nanomolar); mol (moles); gm and g (gram); mg (milligrams); ug and μ g (micrograms); L and l (liter); ml and mL (milliliter); cm (centimeters); mm (millimeters); um and μ m (micrometers); sec. (seconds); min(s) (minute(s)); h(s) and hr(s) (hour(s)); U (units); MW (molecular weight); rpm (rotations per minute); psi and PSI (pounds per square inch);

°C (degrees Centigrade); RT and rt (room temperature); CAM and cam (chloramphenicol); PMBS (polymyxin B sulfate); IPTG (isopropyl β -D-l -thiogalactopyranoside); UGT (uridine 5'-diphosphoglucose glycosyltransferase); LB (Luria broth); TB (terrific broth); SFP (shake flask powder); CDS (coding sequence); DNA (deoxyribonucleic acid); RNA (ribonucleic acid); *E. coli* W3110 (commonly used laboratory *E. coli* strain, available from the Coli Genetic Stock Center [CGSC], New Haven, CT); HTP (high throughput); HPLC (high performance liquid chromatography); MS (mass spectrometry); FIOPC (fold improvements over positive control); Sigma-Aldrich (Sigma-Aldrich, St. Louis, MO; Difco (Difco Laboratories, BD Diagnostic Systems, Detroit, MI); Microfluidics (Microfluidics, Westwood, MA); ChromaDex (ChromaDex, Inc., Irvine, CA); and Thermotron (Thermotron, Holland, MI).

EXAMPLE 1

Synthesis, Optimization, and Assaying of UGT Enzymes With Glucosylation Activity

[0187] In this Example, methods used in the synthesis, optimization and assaying of UGT enzymes with glucosylation activity are described.

Gene Synthesis and Optimization:

[0188] The polynucleotide sequence (SEQ ID NO: 1) encoding the wild-type *Stevia rebaudiana* polypeptide (SEQ ID NO: 2) reported to glucosylate steviolbioside to rebaudioside B and glucosylate stevioside to rebaudioside A, was codon-optimized and synthesized as the gene of SEQ ID NO: 9. The polynucleotide sequence (SEQ ID NO: 3) encoding the wild-type *Streptomyces resistomycificus* glycosyltransferase polypeptide (SEQ ID NO: 4), is a homolog (with 82% sequence identity) of the wild-type *Streptomyces antibioticus* oleandomycin glycosyltransferase (SEQ ID NO:5) sequence reported to glucosylate oleandomycin and have promiscuous activity with other nucleotide sugar donors and toward other substrates, was similarly codon-optimized and synthesized (SEQ ID NO: 11). These synthetic genes (SEQ ID NOS: 9 and 11) were individually cloned into a pCK110900 vector system (See e.g., US Pat. Appln. Publn. No. 2006/0195947, which is hereby incorporated by reference herein) and subsequently expressed in *E. coli* W3110 (Δ fluA). The *E. coli* strain W3110 expressed the UGT enzymes under the control of the lac promoter.

Production of Shake Flask Powders (SFP):

[0189] A shake-flask procedure was used to generate the glycosyltransferase polypeptide shake flask powders (SFP) for characterization assays used in the biocatalytic processes described herein. Shake flask powder (SFP) preparation of enzymes provides a more purified preparation (e.g., up to >30% of total protein) of the enzyme as compared to the cell lysate used in HTP assays and also allows for the use of more concentrated enzyme solutions. A single colony of *E. coli* containing a plasmid encoding an engineered polypeptide of interest was inoculated into 5 mL Luria Bertani broth containing 30 µg/ml chloramphenicol and 1% glucose. Cells were grown overnight (at least 16 hours) in an incubator at 30°C with shaking at 250 rpm. The culture was diluted into 250 mL Terrific Broth (12 g/L bacto-tryptone, 24 g/L yeast extract, 4 mL/L glycerol, 65 mM potassium phosphate, pH 7.0, 1 mM MgSO₄) containing 30 µg/ml CAM, in a 1 L flask to an optical density of 600 nm (OD₆₀₀) of 0.2 and allowed to grow at 30°C. Expression of the glycosyltransferase gene was induced by addition of IPTG to a final concentration of 1 mM when the OD₆₀₀ of the culture was 0.6 to 0.8. Incubation was then continued overnight (at least 16 hours). Cells were harvested by centrifugation (5000 rpm, 15 min, 4 °C) and the supernatant discarded. The cell pellet was resuspended in two volumes of 25 mM triethanolamine buffer, pH 7.5, and passed through a MICROFLUIDIZER[®] high pressure homogenizer (Microfluidics), with standard *E. coli* lysis settings and maintained at 4°C. Cell debris was removed by centrifugation (10,000 rpm, 45 minutes, 4°C). The cleared lysate supernatant was collected and frozen at -80°C and then lyophilized to produce a dry shake-flask powder of crude polypeptide.

Assay of SFP for Stevioside Glucosylation:

[0190] SFP was reconstituted to provide 20 g/L powder. Then, 50 µL of these stocks were diluted in 200 µL total reaction volume of 50 mM Tris-HCl buffer, pH 7.5, with 3 mM MgSO₄ and 1 mM stevioside (ChromaDex, >94% purity), with or without 5 mM α-glucose-1-phosphate. The reaction was performed at 30 °C in a Thermotron titre-plate shaker with 300 RPM shaking for 16-18h.

HPLC-MS/MS Analysis:

[0191] The reaction described above was quenched with 0.5 volume/volume acetonitrile with 2% formic acid and precipitated by centrifugation. Glycosylated stevioside products were detected in the supernatant by LC-MS/MS with the following instrument and parameters:

Instrument	Agilent HPLC 1200 series, Sciex 4000 QTrap																		
Column	Poroshell 120 EC C18 50 x 3.0 mm, 2.7 µm with Poroshell 120 EC C18 5 x 3.0 mm, 2.7 µm guard column (Agilent Technologies)																		
Mobile phase	Gradient (A: 0.1% formic acid in water; B: 0.1% formic acid in methanol) <table> <tr> <th><u>Time (m)</u></th><th><u>%B</u></th></tr> <tr><td>0</td><td>60</td></tr> <tr><td>0.50</td><td>60</td></tr> <tr><td>1.00</td><td>70</td></tr> <tr><td>4.33</td><td>70</td></tr> <tr><td>5.00</td><td>95</td></tr> <tr><td>5.33</td><td>95</td></tr> <tr><td>5.34</td><td>60</td></tr> <tr><td>6.00</td><td>60</td></tr> </table>	<u>Time (m)</u>	<u>%B</u>	0	60	0.50	60	1.00	70	4.33	70	5.00	95	5.33	95	5.34	60	6.00	60
<u>Time (m)</u>	<u>%B</u>																		
0	60																		
0.50	60																		
1.00	70																		
4.33	70																		
5.00	95																		
5.33	95																		
5.34	60																		
6.00	60																		
Flow rate	0.8 mL/m																		
Run time	6 m																		
Peak retention times	Rebaudioside A: 2.35 m, Product 176: 1.76 m, Product 218: 2.18 m, Product 222: 2.22 m.																		
Column temperature	40 °C																		
Injection volume	10 µL																		
MS detection	Sciex 4000 QTrap; MRM 990/828 (for steviol tetraglycosides, e.g., rebaudioside A), 1152/828 (for steviol pentaglycosides, e.g., rebaudioside D), 1314/828 (steviol hexaglycosides, e.g., rebaudioside M), 828/666 (for steviol triglycosides, e.g., stevioside), 666/504 (steviol diglycosides, e.g., rubusoside)																		
MS conditions	MODE: MRM; CUR: 30 ; IS: 4750; CAD: high; TEM: 550 °C; GS1: 50; GS2: 50; DP: 150; EP: 10; CXP: 14; DT: 50 ms for each transition For the first three transitions: CE: 85For the last two transitions: CE: 60																		

[0192] Novel activity was detected for the polypeptides encoded by SEQ ID NO: 9 and 11. For the glucosyltransferase polypeptide encoded by SEQ ID NO: 9, two products were detected, one co-

eluting with rebaudioside A and one eluting at 2.22 m. These products were detected in the presence of α -glucose-1-phosphate, but not in the absence of α -glucose-1-phosphate. These products were not present in a negative control sample. For the glucosyltransferase polypeptide encoded by SEQ ID NO: 11, two products were detected at retention times 1.76 and 2.18 m. These products were detected in the presence of α -glucose-1-phosphate and at reduced levels in the absence of α -glucose-1-phosphate. These products were not present in a negative control sample. Thus, the present invention provides a novel process for glucosylating steviol glycoside substrates. In addition, the present invention provides the first enzymes in this class, including the wild-type enzymes encoded by SEQ ID NO: 2 and 4, to be used with α -glucose-1-phosphate instead of uridine 5'-diphosphoglucose to glycosylate steviol substrates.

EXAMPLE 2

In situ Formation of Glucose-1-Phosphate

[0193] In this Example experiments to assess the in situ formation of glucose-1-phosphate for UDP-glucose-independent glucosylation of substrates (See, Figure 3) are described. Gene synthesis and optimization, as well as production of shake flask powders are performed as described in Example 1.

Assay of SFP:

[0194] SFP is reconstituted to 20 g/L powder. Then, 20 μ L of SFP from *E. coli* expressing SEQ ID NO: 9 or 11 and 10 μ L of SFP from *E. coli* expressing odd-numbered SEQ ID NO: 15-31 or another disaccharide phosphorylase or a negative control are diluted in 200 μ L total reaction volume of 50 mM Tris-HCl buffer, pH 7.5, with 3 mM MgSO_4 , 0.3 M sucrose (or the disaccharide corresponding to the disaccharide phosphorylase), 5 mM inorganic phosphate, and 1 mM stevioside (ChromaDex, >94% purity), with 0-5 mM α -glucose-1-phosphate. The reaction is performed at 30 °C in a Thermotron titre-plate shaker with 300 RPM shaking for 16-18 h.

HPLC-MS/MS Analysis:

[0195] The reaction is quenched and analyzed by LC-MS/MS as described in Example 1. In situ formation of α -glucose-1-phosphate is demonstrated by increased conversion of stevioside to glucosylated products at retention times 1.76, 2.18, 2.22, and 2.35 m in the presence of SFP from *E. coli* expressing odd-numbered SEQ NO: 15-31 or another disaccharide phosphorylase relative to the same samples in the presence of SFP from *E. coli* expressing a negative control.

EXAMPLE 3

Evolution and Screening of Variants

[0196] This Example describes experiments conducted during the evolution and screening of engineered polypeptides derived from SEQ ID NO: 9 and 11 for improved substrate glucosylation using α -glucose-1-phosphate.

[0197] Directed evolution begins with the polynucleotides of SEQ ID NO: 9 or 11, which encode the polypeptide of SEQ ID NO: 10 or 12, respectively, as the starting “backbone” gene sequence. Libraries of engineered polypeptides are generated using various well-known techniques (e.g., saturation mutagenesis, recombination of previously identified beneficial amino acid differences) and screened using HTP assay and analysis methods that measure the ability of the engineered polypeptides to carry out glucosylation of the desired substrate using glucose-1-phosphate or a recycling system for glucose-1-phosphate consisting of an inexpensive disaccharide, inorganic phosphate, and a wild-type or engineered disaccharide phosphorylase, as described in Examples 1 and 2.

[0198] After screening, the engineered polypeptides showing the most improvement over the starting backbone sequence are used as backbone sequences for the construction of further libraries, and the screening process repeated to evolve the polypeptides for the desired activity. Due to the promiscuity of the enzyme for various substrates, this process can be repeated to develop biocatalysts that use an inexpensive co-substrate for the glucosylation of many substrates of interest.

EXAMPLE 4

Synthesis, Optimization, and Assaying Phosphorylase Enzymes with Glucosylation Activity

[0199] In this Example, experiments conducted to synthesize, optimize and assay phosphorylase enzymes having glucosylation activity are described.

Gene Synthesis and Optimization:

[0200] The polynucleotide sequence (SEQ ID NO: 7) encoding the wild-type *Paenibacillus* sp. YM1 laminaribiose phosphorylase polypeptide (SEQ ID NO: 8), reported to phosphorylyse laminaribiose to release glucose and form α -glucose-1-phosphate, was codon-optimized and synthesized as the gene of SEQ ID NO: 13. The synthetic gene of SEQ ID NO: 13 was cloned into a pCK110900 vector system (See e.g., US Pat. Appln. Publ. No. 20060195947, which is hereby incorporated by reference herein) and subsequently expressed in *E. coli* W3110 (Δ fhuA). The *E. coli* strain W3110 expressed the enzyme under the control of the lac promoter.

Production of High-Throughput (HTP) Lysates:

[0201] *E. coli* cells expressing the polypeptide genes of interest were grown and induced in 96-well plates, pelleted, lysed in 250 μ L lysis buffer (0.5 g/L lysozyme and 0.5 g/L PMBS in 20 mM Tris-HCl buffer, pH 7.5) with low-speed shaking for 2 h on a titre-plate shaker at room temperature. The plates

were then centrifuged at 4000 rpm and 4 °C for 20 m and the cleared lysate supernatant was used in the assay reactions described herein.

Assay for Glucosylation of Stevioside:

[0202] In this assay, 50 μ L cleared lysates were diluted in 200 μ L total reaction volume of 50 mM sodium acetate buffer, pH 5.5, with 1.65 mM α -glucose-1-phosphate and 0.5 mM stevioside (ChromaDex, >94% purity). The reaction was performed at 40 °C in a Thermotron titre-plate shaker with 300 RPM shaking for 18 h.

HPLC-MS/MS Analysis:

[0203] The reaction was quenched and analyzed by LC-MS/MS as described in Example 1. A glucosylated stevioside product with retention time 2.18 m was observed in the presence of the laminaribiose phosphorylase but not for the negative control sample. This result indicates that a laminaribiose phosphorylase may be used with α -glucose-1-phosphate as a novel method for α -glucosidation of stevioside and other substrates. As described in Example 3, α -glucose-1-phosphate can be recycled using an inexpensive disaccharide, inorganic phosphate, and a disaccharide phosphorylase.

SEQUENCE LISTING

<110> CODEXIS, INC.
Vroom, Jonathan
Galanie, Stephanie Sue
Alvizo, Oscar

<120> ENZYMATIC GLYCOSYLATION OF STEVIOL GLYCOSIDES AND OTHER COMPOUNDS WITH
GLUCOSE-1-PHOSPHATE

<130> CX8-155USP1

<140> NOT YET ASSIGNED

<141> 2016-06-15

<160> 32

<210> 1

<211> 1326

<212> DNA

<213> Artificial Sequence

<220>

<223> Wild type UDP-glycosyltransferase 76G1 from Stevia

<400> 1

```
atggaaaata aaacggagac caccgttcgc cggcgccgga gaataatatt attcccggta 60
ccatttcaag gccacattaa cccaattctt cagctagcca atgtgttgta ctctaaagga 120
ttcagtatca ccatctttca caccaacttc aacaaaccca aaacatctaa ttaccctcac 180
ttcactttca gattcatcct cgacaacgac ccacaagacg aacgcatttc caatctaccg 240
actcatggtc cgctcgctgg tatgcggtat ccgattatca acgaacacgg agctgacgaa 300
ttacgacgcg aactggaact gttgatgtta gcttctgaag aagatgaaga ggtatcgtgt 360
ttaatcacgg atgctctttg gtacttcgcg caatctgttg ctgacagtct taacctccga 420
cggcttggtt tgatgacaag cagcttggtt aattttcatg cacatgtttc acttcctcag 480
tttgatgagc ttggttacct cgatcctgat gacaaaaccc gtttggaaga acaagcgagt 540
gggtttccta tgctaaaagt gaaagacatc aagtctgcgt attcgaactg gcaaatactc 600
aaagagatat tagggaagat gataaaacaa acaagagcat cttcaggagt catctggaac 660
tcatttaagg aactcgaaga gtctgagctc gaaactgtta tccgtgagat cccggctcca 720
agtttcttga taccactccc caagcatttg acagcctctt ccagcagctt actagaccac 780
gatcgaaccg tttttcaatg gtttagacaa caaccgcaa gttcgggtact gtatgttagt 840
tttggttagta ctagtgaagt ggatgagaaa gatttcttgg aaatagctcg tgggttggtt 900
gatagcaagc agtcgttttt atgggtgggt cgacctgggt ttgtcaaggg ttcgacgtgg 960
gtcgaaccgt tgccagatgg gttcttgggt gaaagaggac gtattgtgaa atgggttcca 1020
cagcaagaag tgctagctca tggagcaata ggcgcattct ggactcatag cggatggaac 1080
tctacgttgg aaagcgtttg tgaaggtgtt cctatgattt tctcggattt tgggctcgat 1140
caaccgttga atgctagata catgagtgat gttttgaagg taggggtgta tttggaaaat 1200
gggtgggaaa gaggagagat agcaaagca ataagaagag ttatggtgga tgaagaagga 1260
gaatacatta gacagaatgc aagagttttg aaacaaaagg cagatgtttc tttgatgaag 1320
ggtggt
```

<210> 2

<211> 442
 <212> PRT
 <213> Artificial Sequence

<220>

<223> Wild type UDP-glycosyltransferase 76G1 from Stevia

<400> 2

```

Met Glu Asn Lys Thr Glu Thr Thr Val Arg Arg Arg Arg Arg Ile Ile
1      5      10      15
Leu Phe Pro Val Pro Phe Gln Gly His Ile Asn Pro Ile Leu Gln Leu
20     25     30
Ala Asn Val Leu Tyr Ser Lys Gly Phe Ser Ile Thr Ile Phe His Thr
35     40     45
Asn Phe Asn Lys Pro Lys Thr Ser Asn Tyr Pro His Phe Thr Phe Arg
50     55     60
Phe Ile Leu Asp Asn Asp Pro Gln Asp Glu Arg Ile Ser Asn Leu Pro
65     70     75     80
Thr His Gly Pro Leu Ala Gly Met Arg Ile Pro Ile Ile Asn Glu His
85     90     95
Gly Ala Asp Glu Leu Arg Arg Glu Leu Glu Leu Leu Met Leu Ala Ser
100    105    110
Glu Glu Asp Glu Glu Val Ser Cys Leu Ile Thr Asp Ala Leu Trp Tyr
115    120    125
Phe Ala Gln Ser Val Ala Asp Ser Leu Asn Leu Arg Arg Leu Val Leu
130    135    140
Met Thr Ser Ser Leu Phe Asn Phe His Ala His Val Ser Leu Pro Gln
145    150    155    160
Phe Asp Glu Leu Gly Tyr Leu Asp Pro Asp Asp Lys Thr Arg Leu Glu
165    170    175
Glu Gln Ala Ser Gly Phe Pro Met Leu Lys Val Lys Asp Ile Lys Ser
180    185    190
Ala Tyr Ser Asn Trp Gln Ile Leu Lys Glu Ile Leu Gly Lys Met Ile
195    200    205
Lys Gln Thr Arg Ala Ser Ser Gly Val Ile Trp Asn Ser Phe Lys Glu
210    215    220
Leu Glu Glu Ser Glu Leu Glu Thr Val Ile Arg Glu Ile Pro Ala Pro
225    230    235    240
Ser Phe Leu Ile Pro Leu Pro Lys His Leu Thr Ala Ser Ser Ser Ser
245    250    255
Leu Leu Asp His Asp Arg Thr Val Phe Gln Trp Leu Asp Gln Gln Pro
260    265    270
Pro Ser Ser Val Leu Tyr Val Ser Phe Gly Ser Thr Ser Glu Val Asp
275    280    285
Glu Lys Asp Phe Leu Glu Ile Ala Arg Gly Leu Val Asp Ser Lys Gln
290    295    300
Ser Phe Leu Trp Val Val Arg Pro Gly Phe Val Lys Gly Ser Thr Trp
305    310    315    320
Val Glu Pro Leu Pro Asp Gly Phe Leu Gly Glu Arg Gly Arg Ile Val
325    330    335
Lys Trp Val Pro Gln Gln Glu Val Leu Ala His Gly Ala Ile Gly Ala
340    345    350
Phe Trp Thr His Ser Gly Trp Asn Ser Thr Leu Glu Ser Val Cys Glu

```


<400> 4

Met	Thr	Ser	Arg	Ser	His	Ile	Ala	Met	Phe	Ser	Ile	Ala	Ala	His	Gly
1				5					10					15	
His	Val	Asn	Pro	Ser	Leu	Glu	Val	Ile	Arg	Glu	Leu	Val	Ala	Arg	Gly
			20					25					30		
His	Arg	Val	Thr	Tyr	Ala	Ile	Pro	Pro	Val	Phe	Ala	Gly	Lys	Val	Ala
		35					40					45			
Glu	Ala	Gly	Ala	Glu	Val	Lys	Leu	Trp	Asn	Ser	Thr	Leu	Pro	Ser	Pro
	50					55					60				
Asp	Asp	Asp	Pro	Gln	Ala	Trp	Gly	Thr	Thr	Leu	Leu	Asp	Asn	Val	Glu
65				70					75						80
Pro	Phe	Leu	Ala	Asp	Ala	Ile	Gln	Ala	Leu	Pro	Gln	Leu	Ile	Glu	Ala
			85					90						95	
Tyr	Glu	Gly	Asp	Glu	Pro	Asp	Leu	Val	Leu	His	Asp	Ile	Thr	Ser	Tyr
			100					105					110		
Pro	Ala	Arg	Val	Leu	Ala	His	Arg	Trp	Gly	Val	Pro	Ala	Val	Ser	Leu
		115					120					125			
Ser	Pro	Asn	Leu	Val	Ala	Trp	Glu	Gly	Tyr	Glu	Lys	Glu	Val	Ala	Glu
	130					135					140				
Pro	Met	Trp	Ala	Glu	Pro	Lys	Arg	Thr	Glu	Arg	Gly	Arg	Ala	Tyr	Tyr
145				150					155						160
Ala	Arg	Phe	His	Ala	Trp	Leu	Glu	Glu	Asn	Gly	Val	Thr	Leu	His	Pro
			165					170						175	
Asp	Asp	Phe	Val	Gly	Arg	Pro	Ala	Arg	Cys	Leu	Val	Leu	Ile	Pro	Lys
		180					185					190			
Ala	Leu	Gln	Pro	Asn	Ala	Asp	Arg	Val	Asp	Glu	Gly	Val	Tyr	Ser	Phe
		195					200					205			
Val	Gly	Ala	Cys	Gln	Gly	Asp	Arg	Gly	Ala	Glu	Gly	Glu	Trp	Arg	Arg
	210					215					220				
Pro	Ala	Gly	Ala	Glu	Lys	Val	Ala	Leu	Val	Ser	Leu	Gly	Ser	Ser	Phe
225					230					235					240
Thr	Lys	Gln	Pro	Asp	Phe	Tyr	Arg	Glu	Cys	Val	Arg	Ala	Phe	Gly	Glu
			245					250						255	
Leu	Pro	Gly	Trp	His	Leu	Val	Leu	Gln	Ile	Gly	Arg	His	Val	Asp	Pro
		260					265						270		
Ala	Asp	Leu	Gly	Asp	Val	Pro	Glu	Asn	Val	Glu	Val	Arg	Ser	Trp	Val
		275					280					285			
Pro	Gln	Leu	Ala	Ile	Leu	Lys	Glu	Ala	Asp	Leu	Phe	Val	Thr	His	Ala
	290					295					300				
Gly	Ala	Gly	Gly	Ser	Gln	Glu	Gly	Leu	Ala	Thr	Ala	Thr	Pro	Met	Ile
305				310						315					320
Ala	Val	Pro	Gln	Ala	Val	Asp	Gln	Phe	Gly	Asn	Ala	Asp	Met	Leu	Gln
			325					330						335	
Gly	Leu	Gly	Val	Ala	Arg	Arg	Leu	Asp	Thr	Glu	Glu	Ala	Thr	Ala	Glu
		340					345						350		
Ala	Leu	Arg	Ala	Ala	Ala	Leu	Ala	Leu	Val	Asp	Asp	Pro	Glu	Val	Ala
		355					360					365			
Arg	Arg	Ala	Arg	Asp	Ile	Gln	Ala	Gln	Met	Ala	Gly	Glu	Gly	Gly	Thr
	370				375						380				
Arg	Arg	Ala	Ala	Asp	Leu	Ile	Glu	Ala	Glu	Leu	Pro	Val	Arg	Gln	Ala
385					390					395					400

<210> 5

<211> 1248
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Wild type macrolide glycosyltransferase from Streptomyces

<400> 5
 atgaccaccc agaccactcc cgcccacatc gccatgttct ccatcgccgc ccacggccat 60
 gtgaaccca gcctggaggt gatccgtgaa ctctgcgcc gcggccaccg ggtcacgtac 120
 gccattccgc ccgtcttcgc cgacaaggtg gccgccaccg gcgcccggcc cgtcctctac 180
 cactccaccc tgcccggccc cgacgccgac ccggaggcat ggggaagcac cctgctggac 240
 aacgtcgaac cgttcctgaa cgacgcgatc caggcgctcc cgcagctcgc cgatgcctac 300
 gccgacgaca tccccgatct cgtcctgcac gacatcacct cctaccggc ccgcgtcctg 360
 gcccgccgct ggggcgtccc ggcggtctcc ctctccccga acctcgtcgc ctggaaggg 420
 tacgaggagg aggtcgccga gccgatgtgg cgcgaacccc ggcagaccga gcgcggacgg 480
 gcctactacg cccggttcga ggcatggctg aaggagaacg ggatcaccga gcaccggac 540
 acgttcgcca gtcattccgc gcgctccctg gtgctcatcc cgaaggcgct ccagccgcac 600
 gccgaccggg tggacgaaga cgtgtacacc ttctcgggc cctgccaggg agaccgcgc 660
 gaggaaggcg gctggcagcg gcccgccggc gcggagaagg tcgtcctggg gtcgtcggc 720
 tcggcgttca ccaagcagcc cgccttctac cgggagtgcg tgcgcgcctt cggaacctg 780
 cccggctggc acctcgtcct ccagatcggc cggaagggtga ccccgccga actgggggag 840
 ctgccggaca acgtggaggt gcacgactgg gtgccgcagc tcgcgacatc gcgccaggcc 900
 gatctgttcg tcaccacgc gggcgccggc ggcagccagg aggggctggc caccgcgacg 960
 cccatgatcg ccgtaccgca ggccgtcgac cagtccggca acgccgacat gctccaaggg 1020
 ctcggtcgct cccggaagct ggcgaccgag gaggccaccg ccgacctgct ccgcgagacc 1080
 gccctcgctc tgggtggacga cccggaggtc gcgcgccggc tccggcggat ccaggcggag 1140
 atggcccagg agggcggcac ccggcgggcg gccgacctca tcgaggccga actgcccgcg 1200
 cgccacgagc ggcaggagcc ggtgggcgac cgaccaacg gtgggtga 1248

<210> 6
 <211> 415
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Wild type macrolide glycosyltransferase from Streptomyces

<400> 6
 Met Thr Thr Gln Thr Thr Pro Ala His Ile Ala Met Phe Ser Ile Ala
 1 5 10 15
 Ala His Gly His Val Asn Pro Ser Leu Glu Val Ile Arg Glu Leu Val
 20 25 30
 Ala Arg Gly His Arg Val Thr Tyr Ala Ile Pro Pro Val Phe Ala Asp
 35 40 45
 Lys Val Ala Ala Thr Gly Ala Arg Pro Val Leu Tyr His Ser Thr Leu
 50 55 60
 Pro Gly Pro Asp Ala Asp Pro Glu Ala Trp Gly Ser Thr Leu Leu Asp
 65 70 75 80
 Asn Val Glu Pro Phe Leu Asn Asp Ala Ile Gln Ala Leu Pro Gln Leu
 85 90 95
 Ala Asp Ala Tyr Ala Asp Asp Ile Pro Asp Leu Val Leu His Asp Ile


```

ttcatgttta tcgtcaccat cgtgcagaac ggagtggcgt tgtttgtggc tgtgctgac 300
aaccaacggc tgcgcggcga ttatTTTTac cgggccgtat tttttcttcc ggtcgtgctt 360
ggggtggccg ttgttgcgct cgtttggggc atgatgttcg atccgatgag cgggccgggt 420
aacatgctgt atcaatcgct gttcggatat aaagacatgt ttttctccag ctttacgcat 480
gcgttcggct acattatTTT cgttcaaatc tggatgtaca tgggctattc aatgctgatt 540
tttctcgccg gcctacaatc cgtgccaaaa gatctgtacg aagcggggta catcgacggt 600
acgaaccgct ggcaatcggt caagaatatt acgttcccgc tgatcgcccc gtcttttacg 660
gtcaatatgc tgctttcgat catcggggcg atgtccacgt tcgatattat cctcgccacg 720
acggacggca agttcaacac cagaacgatg gcttacgacg tatataagga aacgttccgc 780
ggcagtctgg aaatgggctt gccttcggcg ctgtccgtgg tccagttcct gatcattttg 840
gtgtttgtta tcgtggccgt cagacagacg cgtaagagag aggtggaata ttaa 894

```

<210> 8

<211> 297

<212> PRT

<213> Artificial Sequence

<220>

<223> Wild type laminaribiose phosphorylase from Paenibacillus

<400> 8

```

Met Tyr Pro Phe Gly Lys Gly Phe Ala Arg Tyr Val Pro Phe Leu Leu
1          5          10          15
Leu Leu Ile Pro Leu Ala Leu Tyr Val Val Phe Tyr Phe Gly Pro Ser
20          25          30
Val Leu Thr Val Ile Tyr Ser Phe Thr Asp Val Lys Asn Ile Pro Gly
35          40          45
Ser Ser Met Asn Phe Val Gly Leu Asp Asn Tyr Lys Asp Leu Phe Phe
50          55          60
Ala Gly Asn Ser Gly Glu Arg Trp Val Ser Ile Lys Asn Thr Leu Ile
65          70          75          80
Phe Met Phe Ile Val Thr Ile Val Gln Asn Gly Val Ala Leu Phe Val
85          90          95
Ala Val Leu Ile Asn Gln Arg Leu Arg Gly Asp Tyr Phe Tyr Arg Ala
100         105         110
Val Phe Phe Leu Pro Val Val Leu Gly Val Ala Val Val Ala Leu Val
115         120         125
Trp Gly Met Met Phe Asp Pro Met Ser Gly Pro Val Asn Met Leu Tyr
130         135         140
Gln Ser Leu Phe Gly Tyr Lys Asp Met Phe Phe Ser Ser Phe Thr His
145         150         155         160
Ala Phe Gly Tyr Ile Ile Phe Val Gln Ile Trp Met Tyr Met Gly Tyr
165         170         175
Ser Met Leu Ile Phe Leu Ala Gly Leu Gln Ser Val Pro Lys Asp Leu
180         185         190
Tyr Glu Ala Gly Tyr Ile Asp Gly Thr Asn Arg Trp Gln Ser Phe Lys
195         200         205
Asn Ile Thr Phe Pro Leu Ile Ala Pro Ser Phe Thr Val Asn Met Leu
210         215         220
Leu Ser Ile Ile Gly Ala Met Ser Thr Phe Asp Ile Ile Leu Ala Thr
225         230         235         240
Thr Asp Gly Lys Phe Asn Thr Arg Thr Met Ala Tyr Asp Val Tyr Lys

```

				245					250					255					
Glu	Thr	Phe	Arg	Gly	Ser	Leu	Glu	Met	Gly	Leu	Pro	Ser	Ala	Leu	Ser				
			260					265					270						
Val	Val	Gln	Phe	Leu	Ile	Ile	Leu	Val	Phe	Val	Ile	Val	Ala	Val	Arg				
		275					280					285							
Gln	Thr	Arg	Lys	Arg	Glu	Val	Glu	Tyr											
	290					295													

<210> 9

<211> 1374

<212> DNA

<213> Artificial Sequence

<220>

<223> Variant of UDP-glycosyltransferase 76G1 from Stevia

<400> 9

```

atggagaaca aaaccgaaac caccgtgcgt cgtcgtcgtc gtatcattct gtttccgggt 60
ccgttccagg gccacatcaa cccgattctg caactggcga acgtgctgta cagcaagggt 120
tttagcatca ccattttttca caccaacttc aacaagccga aaaccagcaa ctatccgcac 180
ttcacctttc gtttcatcct ggacaacgat ccgcaggacg agcgtattag caacctgccg 240
acccatggtc cgctggcggg tatgcgtatc ccgatcatta acgagcacgg cgcggatgaa 300
ctgcgtcgtg agctggaaact gctgatgctg gcgagcgagg aagacgagga agtttagctgc 360
ctgattaccg atgcgctgtg gtacttcgcg caaagcgtgg cggacagcct gaacctgcgt 420
cgtctgggtc tgatgaccag cagcctgttt aacttccacg cgcacgtgag cctgccgcag 480
tttgacgagc tgggctacct ggatccggac gataaaaccc gtctggagga acaagcgagc 540
ggtttcccga tgctgaaggt taaagatatc aagagcgcgt atagcaactg gcagatcctg 600
aaagaaattc tgggcaagat gatcaagcag acccgtgcga gcagcgggtg gatttggaac 660
agctttaaag agctggagga aagcgagctg gaaaccgtta tccgtgaaat tccggcgccg 720
agcttcctga ttccgctgcc gaagcacctg accgcgagca gcagcagcct gctggaccac 780
gatcgtaccg tgttccagtg gctggaccag caaccgccga gcagcgtgct gtacgttagc 840
tttggcagca ccagcgaggt ggacgaaaag gatttcctgg agattgcgcg tgggtctggtt 900
gacagcaaac aaagctttct gtgggtgggt cgtccgggct tcgtgaaggg tagcacctgg 960
gttgagccgc tgccggatgg ttttctgggc gaacgtggtc gtatcgtgaa gtgggttccg 1020
caacaagaag tgctggcgca cggcgcgatt ggtgcgttct ggaccacag cggttggaac 1080
agcacctgg agagcgtgtg cgaaggcgtt ccgatgatct ttagcgactt cggctctggat 1140
cagccgtga acgcgcgtta catgagcgat gttctgaaag tgggcgttta tctggagaac 1200
ggctgggagc gtggtgaaat cgcaacgcg attcgtcgtg tgatggttga cgaggaaggt 1260
gagtacatcc gtcagaacgc gcgtgttctg aagcaaaaag cggatgttag cctgatgaag 1320
ggtggcagca gctacgagag cctggaaagc ctggttagct atattagcag cctg 1374

```

<210> 10

<211> 458

<212> PRT

<213> Artificial Sequence

<220>

<223> Variant of UDP-glycosyltransferase 76G1 from Stevia

<400> 10

Met	Glu	Asn	Lys	Thr	Glu	Thr	Thr	Val	Arg	Arg	Arg	Arg	Arg	Ile	Ile
1				5				10						15	

Leu	Phe	Pro	Val	Pro	Phe	Gln	Gly	His	Ile	Asn	Pro	Ile	Leu	Gln	Leu
			20					25					30		
Ala	Asn	Val	Leu	Tyr	Ser	Lys	Gly	Phe	Ser	Ile	Thr	Ile	Phe	His	Thr
		35					40					45			
Asn	Phe	Asn	Lys	Pro	Lys	Thr	Ser	Asn	Tyr	Pro	His	Phe	Thr	Phe	Arg
	50					55					60				
Phe	Ile	Leu	Asp	Asn	Asp	Pro	Gln	Asp	Glu	Arg	Ile	Ser	Asn	Leu	Pro
65					70					75					80
Thr	His	Gly	Pro	Leu	Ala	Gly	Met	Arg	Ile	Pro	Ile	Ile	Asn	Glu	His
				85						90				95	
Gly	Ala	Asp	Glu	Leu	Arg	Arg	Glu	Leu	Glu	Leu	Leu	Met	Leu	Ala	Ser
			100					105					110		
Glu	Glu	Asp	Glu	Glu	Val	Ser	Cys	Leu	Ile	Thr	Asp	Ala	Leu	Trp	Tyr
		115					120					125			
Phe	Ala	Gln	Ser	Val	Ala	Asp	Ser	Leu	Asn	Leu	Arg	Arg	Leu	Val	Leu
	130					135					140				
Met	Thr	Ser	Ser	Leu	Phe	Asn	Phe	His	Ala	His	Val	Ser	Leu	Pro	Gln
145					150					155					160
Phe	Asp	Glu	Leu	Gly	Tyr	Leu	Asp	Pro	Asp	Asp	Lys	Thr	Arg	Leu	Glu
				165					170					175	
Glu	Gln	Ala	Ser	Gly	Phe	Pro	Met	Leu	Lys	Val	Lys	Asp	Ile	Lys	Ser
			180					185					190		
Ala	Tyr	Ser	Asn	Trp	Gln	Ile	Leu	Lys	Glu	Ile	Leu	Gly	Lys	Met	Ile
		195					200					205			
Lys	Gln	Thr	Arg	Ala	Ser	Ser	Gly	Val	Ile	Trp	Asn	Ser	Phe	Lys	Glu
	210					215					220				
Leu	Glu	Glu	Ser	Glu	Leu	Glu	Thr	Val	Ile	Arg	Glu	Ile	Pro	Ala	Pro
225					230					235					240
Ser	Phe	Leu	Ile	Pro	Leu	Pro	Lys	His	Leu	Thr	Ala	Ser	Ser	Ser	Ser
				245					250					255	
Leu	Leu	Asp	His	Asp	Arg	Thr	Val	Phe	Gln	Trp	Leu	Asp	Gln	Gln	Pro
			260					265					270		
Pro	Ser	Ser	Val	Leu	Tyr	Val	Ser	Phe	Gly	Ser	Thr	Ser	Glu	Val	Asp
		275					280					285			
Glu	Lys	Asp	Phe	Leu	Glu	Ile	Ala	Arg	Gly	Leu	Val	Asp	Ser	Lys	Gln
	290					295					300				
Ser	Phe	Leu	Trp	Val	Val	Arg	Pro	Gly	Phe	Val	Lys	Gly	Ser	Thr	Trp
305					310					315					320
Val	Glu	Pro	Leu	Pro	Asp	Gly	Phe	Leu	Gly	Glu	Arg	Gly	Arg	Ile	Val
				325					330					335	
Lys	Trp	Val	Pro	Gln	Gln	Glu	Val	Leu	Ala	His	Gly	Ala	Ile	Gly	Ala
			340					345					350		
Phe	Trp	Thr	His	Ser	Gly	Trp	Asn	Ser	Thr	Leu	Glu	Ser	Val	Cys	Glu
		355					360					365			
Gly	Val	Pro	Met	Ile	P										

435 440 445
 Glu Ser Leu Val Ser Tyr Ile Ser Ser Leu
 450 455

<210> 11
 <211> 1203
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Variant of macrolide glycosyltransferase from Streptomyces

<400> 11
 atgaccagtc gttcccatat tgccatgttt agtatcgcg cccatgggtca cgtgaacccg 60
 tccctggaag tcattcgtga actgggtggcg cgtgggtcacc gtgttaccta tgccatccc 120
 ccggttttcg ccggtaaagt tgctgaagcg ggcgccgaag tgaaactgtg gaacagcacg 180
 ctgccgtctc cggatgacga tccgcaggcg tggggtagca cgctgctgga taatgttgaa 240
 ccgtttctgg cagacgctat tcaggctctg ccgcaactga tcgaagcgta tgaaggcgac 300
 gaaccggatc tggtcctgca tgatatcacc tcatatccgg ctctgtttct ggcccatcgc 360
 tgggggtgtcc cggccgtgtc actgtcgccg aatctgggtg cctgggaagg ctacgaaaaa 420
 gaagtcgctg aaccgatgtg ggcggaaccg aaacgtaccg aacgtgggtc cgcctattac 480
 gcacgctttc atgcctggct ggaagaaaac ggtgttacgc tgcattccgga tgattttgtt 540
 ggtcgtccgg caggttgctt ggtgctgatt ccgaaagccc tgcagccgaa tgcagaccgt 600
 gtggatgaag gtgtttatag ctttgtcggc gcctgtcaag gcgatcgcg tgcagagggt 660
 gaatggcgtc gcccggcggg tgcagaaaaa gttgccctgg tctctctggg cagctctttt 720
 accaaacagc cggatttcta ccgtgaatgc gtgcgcgcct tcggtgaatt accgggctgg 780
 catctgggtc tgcaaattgg tcgccatgtg gaccgggccg atctgggtga tgtgccggaa 840
 aacgtggaag ttcgctcatg ggttccgcag ctggctatcc tgaaagaagc ggacctgttt 900
 gtgacgcatg caggtgctgg cggttcgcaa gaaggcctgg ccaccgcaac gccgatgatt 960
 gcagtgccgc aggcagttga tcaattcggg aatgcagaca tgctgcaggg tctgggtgtt 1020
 gcccgtcgcc tggataccga agaagcaacc gctgaagcgc tgcgtgcagc ggcgctggcg 1080
 ctggttgatg atccggaagt tgctcgtcgc gcgcgcgata ttcaggcgca gatggcgggt 1140
 gaaggcggta cccgtcgcgc agcagacctg atcgaagctg aactgccggt gcgtcaggca 1200
 taa 1203

<210> 12
 <211> 400
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Variant of macrolide glycosyltransferase from Streptomyces

<400> 12
 Met Thr Ser Arg Ser His Ile Ala Met Phe Ser Ile Ala Ala His Gly
 1 5 10 15
 His Val Asn Pro Ser Leu Glu Val Ile Arg Glu Leu Val Ala Arg Gly
 20 25 30
 His Arg Val Thr Tyr Ala Ile Pro Pro Val Phe Ala Gly Lys Val Ala
 35 40 45
 Glu Ala Gly Ala Glu Val Lys Leu Trp Asn Ser Thr Leu Pro Ser Pro

50	55	60																	
Asp	Asp	Asp	Pro	Gln	Ala	Trp	Gly	Thr	Thr	Leu	Leu	Asp	Asn	Val	Glu				
65					70					75					80				
Pro	Phe	Leu	Ala	Asp	Ala	Ile	Gln	Ala	Leu	Pro	Gln	Leu	Ile	Glu	Ala				
				85					90					95					
Tyr	Glu	Gly	Asp	Glu	Pro	Asp	Leu	Val	Leu	His	Asp	Ile	Thr	Ser	Tyr				
			100					105					110						
Pro	Ala	Arg	Val	Leu	Ala	His	Arg	Trp	Gly	Val	Pro	Ala	Val	Ser	Leu				
		115					120					125							
Ser	Pro	Asn	Leu	Val	Ala	Trp	Glu	Gly	Tyr	Glu	Lys	Glu	Val	Ala	Glu				
	130					135				140									
Pro	Met	Trp	Ala	Glu	Pro	Lys	Arg	Thr	Glu	Arg	Gly	Arg	Ala	Tyr	Tyr				
145					150				155						160				
Ala	Arg	Phe	His	Ala	Trp	Leu	Glu	Glu	Asn	Gly	Val	Thr	Leu	His	Pro				
				165					170					175					
Asp	Asp	Phe	Val	Gly	Arg	Pro	Ala	Arg	Cys	Leu	Val	Leu	Ile	Pro	Lys				
		180					185					190							
Ala	Leu	Gln	Pro	Asn	Ala	Asp	Arg	Val	Asp	Glu	Gly	Val	Tyr	Ser	Phe				
		195				200					205								
Val	Gly	Ala	Cys	Gln	Gly	Asp	Arg	Gly	Ala	Glu	Gly	Glu	Trp	Arg	Arg				
	210				215					220									
Pro	Ala	Gly	Ala	Glu	Lys	Val	Ala	Leu	Val	Ser	Leu	Gly	Ser	Ser	Phe				
225					230					235					240				
Thr	Lys	Gln	Pro	Asp	Phe	Tyr	Arg	Glu	Cys	Val	Arg	Ala	Phe	Gly	Glu				
			245				250							255					
Leu	Pro	Gly	Trp	His	Leu	Val	Leu	Gln	Ile	Gly	Arg	His	Val	Asp	Pro				
		260					265					270							
Ala	Asp	Leu	Gly	Asp	Val	Pro	Glu	Asn	Val	Glu	Val	Arg	Ser	Trp	Val				
		275					280					285							
Pro	Gln	Leu	Ala	Ile	Leu	Lys	Glu	Ala	Asp	Leu	Phe	Val	Thr	His	Ala				
	290				295					300									
Gly	Ala	Gly	Gly	Ser	Gln	Glu	Gly	Leu	Ala	Thr	Ala	Thr	Pro	Met	Ile				
305				310			315								320				
Ala	Val	Pro	Gln	Ala	Val	Asp	Gln	Phe	Gly	Asn	Ala	Asp	Met	Leu	Gln				
			325				330							335					
Gly	Leu	Gly	Val	Ala	Arg	Arg	Leu	Asp	Thr	Glu	Glu	Ala	Thr	Ala	Glu				
		340					345					350							
Ala	Leu	Arg	Ala	Ala	Ala	Leu	Ala	Leu	Val	Asp	Asp	Pro	Glu	Val	Ala				
	355					360				365									
Arg	Arg	Ala	Arg	Asp	Ile	Gln	Ala	Gln	Met	Ala	Gly	Glu	Gly	Gly	Thr				
	370				375				380										
Arg	Arg	Ala	Ala	Asp	Leu	Ile	Glu	Ala	Glu	Leu	Pro	Val	Arg	Gln	Ala				
385				390					395						400				

<210> 13

<211> 2733

<212> DNA

<213> Artificial Sequence

<220>

<223> Variant of laminaribiose phosphorylase from Paenibacillus

<400> 13

```
atgggccaga agggttggaa atttcagggc gaacaaggcg agttccgtct ggaacaaccg 60
gagcacaaca gctacctgta ttttccgctg gtgaacgaag cgggtatgat gagcgcggtt 120
accccgaacc tgcacggcga gatcaccagc ggtcacaaca ctttcctgat ggaaccgggtg 180
agcgcgggaga gcctgcacaa cagcaaggcg agccgtaact tttgggtgtt cattgaaggt 240
tatggtgctg ggagcgtagg cggcaacagc gcgcgtcaga acgcggcgcg ttttaccggt 300
gaggaagagc gtagcgcggt ggaggcgggt ttcctgtggc acgcggttac ccgtgaaaac 360
gagaaggcgg gtctgaaagc gcgtaccgtg agcttcgtgc cggttaccga cgataagatc 420
gaactgatgc gtgttaccct gaccaacacc ggcaacgcgc cgtgaaact gaccccgacc 480
gcggcgattc cgctgtatgg tcgtagcgcg gacgatctgc gtgatcaccg tcacgtgacc 540
agcctgctgc accgtatctt taccagcgaa tacggcattg aggttcagcc ggcgctgagc 600
ttcgacgaac gtggtcaccg tgtgaacaag gttacctatg gcgtgtttgg tgcggaggcg 660
gggtggcaccg cgccggcggg cttctttccg gttaccgaag atttcattgg cgagggtggc 720
gcgctggact ggccggaagc ggtggttgcg aaccgtgagc cggatgcgca ggccggcacc 780
gcggtggaag gttatgaggc ggttggcgcg ctgcgttttg cgccggtgga gctggcgccg 840
ggcaaaagcg tgagctatgt ggttgcgatg gttatcagcg gcgaccgtat tgatgttggt 900
cgttacgcgg cggattatct ggccggcgggt cgttttgacg cgtgctgga gcagaaccgt 960
gcgtactggc gtgacaagct ggataccgtg cgtttcagca gcggtgatgg cgagcaggac 1020
ctgtggatga aatgggttac cctgcaaccg attctgcgtc gtctgtatgg taacagcttc 1080
ctgccgtacc acgattatgg tcgtggtggc cgtggctggc gtgatctgtg gcaagactgc 1140
ctggcgctga tggatgagga accggcggaa gtgcgtcacc tgctgctgaa caactacgcg 1200
gggtgtgcgta tggacggcag caacgcgacc atcattgggtg cgggcccggg tgaatttgtg 1260
gcggatcgta acaacatccc gcgtgttttg atggaccatg gtgcgtggcc gctgatgacc 1320
accctgctgt acctgcacca gagcggtagc ctggatctgc tgttccagcc gcaaagctat 1380
tttctgatg tgttcgtaa gcgttgccgt gagcgtgatg cgagctggac cccggagcag 1440
ggtaacaaac tgctgaccgc ggacggtcaa atttacgaag gcaccatcct ggagcacatt 1500
ctgctgcaaa acatcgtgcc gttctttaac gttggcgaac acggtaacat taaactggag 1560
ggcgcggaact ggaacgatgg tctggacctg gcgccggaac gtggtgagag cgttgcgttt 1620
accgcgttct acgcgagcaa cctgatggaa ctgagcgagc tgctgctgga gctgcagaag 1680
cgtaccggca aagatagcct ggacatcgcg gaagagatgg cgctgctgct ggataccctg 1740
ggcaagccga tcagctatga cagcattcaa gaaaaacgta gcctgctgga tcgttactat 1800
gacgcggtga cccgcgtgt tagcggcaag aaactgctgc tggatatccg taaggtggcg 1860
gaggatctga agcgtaaagc ggactgggcg gttgcgcacc tgcgtggtag cgaatggatt 1920
cagagcaaag agggctacgc gtggtttaac gttactata acaacgacgg tgaacgtgtg 1980
gagggcgatc acccgacgg tgttcgtatg accctgaccg gtcaagtgtt cgcgatcatg 2040
gggtggcggtg cgaccgatga acagaccgag aagattagcc aagcggtgaa ccgttacctg 2100
aaagacgaac gtatcggtc tctctgaac agccgttttg gtggcattca gcaaaacctg 2160
ggtcgtgcgt tcggctttgc gttcggtcac aaggagaacg gtgcgatgtt cagccacatg 2220
accgtgatgt acgcgaacgc gctgtataaa cgtggctttg tgcaggaagg tttcgagggt 2280
ctggatagca tctaccgtct gagcgcggac tttgaaaaca gccgtatcta cccggcggtt 2340
ccggaatata ttaacgagcg tggccgtggc atgtacacct atctgaccgg tagcgcgagc 2400
tggctgctgc tgaccaact gaccgaagtg tatggcgtaa agggtcgttt cggcgatctg 2460
cgtctggagc cgaaactggt gcaggcgcaa tttgacggca gcggtgaagc ggcggttgag 2520
accctgttcg cgggtcgtat gctgcgtgtg gtttatcgta acccgaggc ggcgagcat 2580
ggtcaatatc gtgtggatag cgttagcctg aacggtcaga gcgttgattg ccaaaacgac 2640
ggcgcggggt gcctgatcgg tcgtagcctg attgaagcgc tgccggcgga tgggtgtcat 2700
gagctgatcg ttaccctggg tcgtaacatt agc 2733
```

<210> 14

<211> 911

<212> PRT

<213> Artificial Sequence

<220>

<223> Variant of laminaribiose phosphorylase from Paenibacillus

<400> 14

```
Met Gly Gln Lys Gly Trp Lys Phe Gln Gly Glu Gln Gly Glu Phe Arg
1      5      10      15
Leu Glu Gln Pro Glu His Asn Ser Tyr Leu Tyr Phe Pro Leu Val Asn
20      25      30
Glu Ala Gly Met Met Ser Ala Val Thr Pro Asn Leu His Gly Glu Ile
35      40      45
Thr Ser Gly His Asn Thr Phe Leu Met Glu Pro Val Ser Ala Glu Ser
50      55      60
Leu His Asn Ser Lys Ala Ser Arg Asn Phe Trp Val Phe Ile Glu Gly
65      70      75      80
Tyr Gly Ala Trp Ser Val Ser Gly Asn Ser Ala Arg Gln Asn Ala Ala
85      90      95
Arg Phe Thr Gly Glu Glu Glu Arg Ser Ala Val Glu Ala Gly Phe Leu
100     105     110
Trp His Ala Val Thr Arg Glu Asn Glu Lys Ala Gly Leu Lys Ala Arg
115     120     125
Thr Val Ser Phe Val Pro Val Thr Asp Asp Lys Ile Glu Leu Met Arg
130     135     140
Val Thr Leu Thr Asn Thr Gly Asn Ala Pro Leu Lys Leu Thr Pro Thr
145     150     155     160
Ala Ala Ile Pro Leu Tyr Gly Arg Ser Ala Asp Asp Leu Arg Asp His
165     170     175
Arg His Val Thr Ser Leu Leu His Arg Ile Phe Thr Ser Glu Tyr Gly
180     185     190
Ile Glu Val Gln Pro Ala Leu Ser Phe Asp Glu Arg Gly His Arg Val
195     200     205
Asn Lys Val Thr Tyr Gly Val Phe Gly Ala Glu Ala Gly Gly Thr Ala
210     215     220
Pro Ala Gly Phe Phe Pro Val Thr Glu Asp Phe Ile Gly Glu Gly Gly
225     230     235     240
Ala Leu Asp Trp Pro Glu Ala Val Val Ala Asn Arg Glu Pro Asp Ala
245     250     255
Gln Ala Gly Thr Ala Val Glu Gly Tyr Glu Ala Val Gly Ala Leu Arg
260     265     270
Phe Ala Pro Val Glu Leu Ala Pro Gly Lys Ser Val Ser Tyr Val Val
275     280     285
Ala Met Val Ile Ser Gly Asp Arg Ile Asp Val Gly Arg Tyr Ala Ala
290     295     300
Asp Tyr Leu Ala Ala Gly Arg Phe Asp Ala Leu Leu Glu Gln Asn Arg
305     310     315     320
```

Ala	Tyr	Trp	Arg	Asp	Lys	Leu	Asp	Thr	Val	Arg	Phe	Ser	Ser	Gly	Asp	325	330	335
Gly	Glu	Gln	Asp	Leu	Trp	Met	Lys	Trp	Val	Thr	Leu	Gln	Pro	Ile	Leu	340	345	350
Arg	Arg	Leu	Tyr	Gly	Asn	Ser	Phe	Leu	Pro	Tyr	His	Asp	Tyr	Gly	Arg	355	360	365
Gly	Gly	Arg	Gly	Trp	Arg	Asp	Leu	Trp	Gln	Asp	Cys	Leu	Ala	Leu	Met	370	375	380
Val	Met	Glu	Pro	Ala	Glu	Val	Arg	His	Leu	Leu	Asn	Asn	Tyr	Ala		385	390	395
Gly	Val	Arg	Met	Asp	Gly	Ser	Asn	Ala	Thr	Ile	Ile	Gly	Ala	Gly	Pro	405	410	415
Gly	Glu	Phe	Val	Ala	Asp	Arg	Asn	Asn	Ile	Pro	Arg	Val	Trp	Met	Asp	420	425	430
His	Gly	Ala	Trp	Pro	Leu	Met	Thr	Thr	Leu	Leu	Tyr	Leu	His	Gln	Ser	435	440	445
Gly	Asp	Leu	Asp	Leu	Leu	Phe	Gln	Pro	Gln	Ser	Tyr	Phe	Arg	Asp	Val	450	455	460
Phe	Val	Lys	Arg	Cys	Arg	Glu	Arg	Asp	Ala	Ser	Trp	Thr	Pro	Glu	Gln	465	470	475
Gly	Asn	Lys	Leu	Leu	Thr	Ala	Asp	Gly	Gln	Ile	Tyr	Glu	Gly	Thr	Ile	485	490	495
Leu	Glu	His	Ile	Leu	Leu	Gln	Asn	Ile	Val	Pro	Phe	Phe	Asn	Val	Gly	500	505	510
Glu	His	Gly	Asn	Ile	Lys	Leu	Glu	Gly	Ala	Asp	Trp	Asn	Asp	Gly	Leu	515	520	525
Asp	Leu	Ala	Pro	Glu	Arg	Gly	Glu	Ser	Val	Ala	Phe	Thr	Ala	Phe	Tyr	530	535	540
Ala	Ser	Asn	Leu	Met	Glu	Leu	Ser	Glu	Leu	Leu	Leu	Glu	Leu	Gln	Lys	545	550	555
Arg	Thr	Gly	Lys	Asp	Ser	Leu	Asp	Ile	Ala	Glu	Glu	Met	Ala	Leu	Leu	565	570	575
Leu	Asp	Thr	Leu	Gly	Lys	Pro	Ile	Ser	Tyr	Asp	Ser	Ile	Gln	Glu	Lys	580	585	590
Arg	Ser	Leu	Leu	Asp	Arg	Tyr	Tyr	Asp	Ala	Val	Thr	Pro	Arg	Val	Ser	595	600	605
Gly	Lys	Leu	Leu	Leu	Asp	Ile	Arg	Lys	Val	Ala	Glu	Asp	Leu	Lys		610	615	620
Arg	Lys	Ala	Asp	Trp	Ala	Val	Ala	His	Leu	Arg	Gly	Ser	Glu	Trp	Ile	625	630	635
Gln	Ser	Lys	Glu	Gly	Tyr	Ala	Trp	Phe	Asn	Gly	Tyr	Tyr	Asn	Asn	Asp	645	650	655
Gly	Glu	Arg	Val	Glu	Gly	Asp	His	Pro	Asp	Gly	Val	Arg	Met	Thr	Leu	660	665	670
Thr	Gly	Gln	Val	Phe	Ala	Ile	Met	Gly	Gly	Val	Ala	Thr	Asp	Glu	Gln	675	680	685
Thr	Glu	Lys	Ile	Ser	Gln	Ala	Val	Asn	Arg	Tyr	Leu	Lys	Asp	Glu	Arg	690	695	700
Ile	Gly	Tyr	Arg	Leu	Asn	Ser	Arg	Phe	Gly	Gly	Ile	Gln	Gln	Asn	Leu	705	710	715
Gly	Arg	Ala	Phe	Gly	Phe	Ala	Phe	Gly	His	Lys	Glu	Asn	Gly	Ala	Met	725	730	735
Phe	Ser	His	Met	Thr	Val	Met	Tyr	Ala	Asn	Ala	Leu	Tyr	Lys	Arg	Gly			

tactatagcg ttgacgaagt gaaagaggaa gttaagcgtc cggtggttgc gaaactgctg 1260
aagctgctga gctggcgtaa caacttcgcg gcgtttgacc tggatggtag cattgatgtg 1320
gaaaccccga gcgacaccac catcaagatt acccgtaaag ataagagcgg cgagaacgtg 1380
gcggttctgg tggcgaacgc ggcggacaaa acctttacca tcaccgcgaa cggcgaggaa 1440
attctggcga acaccgaggc ggacaagcag caactg 1476

<210> 16

<211> 492

<212> PRT

<213> Artificial Sequence

<220>

<223> Sucrose phosphorylase from *Leuconostoc*

<400> 16

Met	Glu	Ile	Gln	Asn	Lys	Ala	Met	Phe	Ile	Thr	Tyr	Ala	Asp	Ser	Leu
1				5				10					15		
Gly	Lys	Asn	Leu	Lys	Asp	Val	Arg	Lys	Val	Leu	Lys	Glu	Asp	Ile	Gly
		20					25					30			
Asp	Ala	Ile	Gly	Gly	Val	His	Leu	Leu	Pro	Phe	Phe	Pro	Ser	Thr	Gly
	35					40					45				
Asp	Arg	Gly	Phe	Ala	Pro	Ser	Asp	Tyr	Thr	Arg	Val	Asp	Ser	Ala	Phe
	50					55				60					
Gly	Asp	Trp	Ser	Asp	Val	Glu	Ala	Leu	Gly	Glu	Glu	Tyr	Tyr	Leu	Met
65				70				75						80	
Phe	Asp	Phe	Met	Ile	Asn	His	Ile	Ser	Arg	Glu	Ser	Val	Met	Tyr	Gln
			85					90					95		
Asp	Phe	Lys	Lys	Asn	His	Asp	Glu	Ser	Lys	Tyr	Lys	Asp	Phe	Phe	Ile
		100					105					110			
Arg	Trp	Glu	Lys	Phe	Trp	Ala	Lys	Ala	Gly	Glu	Asn	Arg	Pro	Thr	Gln
	115					120					125				
Ala	Asp	Val	Asp	Leu	Ile	Tyr	Lys	Arg	Lys	Asp	Lys	Ala	Pro	Thr	Gln
	130					135				140					
Glu	Ile	Thr	Phe	Asp	Asp	Gly	Thr	Thr	Glu	Asn	Leu	Trp	Asn	Thr	Phe
145				150					155					160	
Gly	Asp	Glu	Gln	Ile	Asp	Ile	Asp	Val	Asn	Ser	Ala	Ile	Ala	Lys	Glu
			165					170					175		
Phe	Ile	Lys	Thr	Thr	Leu	Glu	Asp	Met	Val	Lys	His	Gly	Ala	Asn	Leu
		180				185				190					
Ile	Arg	Leu	Asp	Ala	Phe	Ala	Tyr	Ala	Val	Lys	Lys	Val	Asp	Thr	Asn
	195					200				205					
Asp	Phe	Phe	Val	Glu	Pro	Glu	Ile	Trp	Asp	Thr	Leu	Asn	Glu	Val	Arg
	210				215					220					
Glu	Ile	Leu	Thr	Pro	Leu	Lys	Ala	Glu	Ile	Leu	Pro	Glu	Ile	His	Glu
225				230						235				240	
His	Tyr	Ser	Ile	Pro	Lys	Lys	Ile	Asn	Asp	His	Gly	Tyr	Phe	Thr	Tyr
			245					250					255		
Asp	Phe	Ala	Leu	Pro	Met	Thr	Thr	Leu	Tyr	Thr	Leu	Tyr	Ser	Gly	Lys
		260						265					270		
Thr	Asn	Gln	Leu	Ala	Lys	Trp	Leu	Lys	Met	Ser	Pro	Met	Lys	Gln	Phe
	275					280						285			
Thr	Thr	Leu	Asp	Thr	His	Asp	Gly	Ile	Gly	Val	Val	Asp	Ala	Arg	Asp

290	295	300
Val Leu Thr Asp Glu	Glu Ile Asp Tyr Ala Ser	Glu Glu Leu Tyr Lys
305	310	315
Val Gly Ala Asn Val	Lys Lys Thr Tyr Ser Ser	Ala Ser Tyr Asn Asp
325	330	335
Leu Asp Ile Tyr Gln	Ile Asn Ser Thr Tyr Tyr	Ser Ala Leu Gly Asn
340	345	350
Asp Asp Ala Tyr Leu	Leu Ser Arg Ile Phe Gln	Val Phe Ala Pro
355	360	365
Gly Ile Pro Gln Ile	Tyr Tyr Val Gly Leu Leu	Ala Gly Glu Ser Asp
370	375	380
Ile Glu Leu Leu Glu	Ser Ser Lys Glu Gly Arg	Asn Ile Asn Arg His
385	390	395
Tyr Tyr Ser Val Asp	Glu Val Lys Glu Glu Val	Lys Arg Pro Val Val
405	410	415
Ala Lys Leu Leu Lys	Leu Leu Ser Trp Arg Asn	Asn Phe Ala Ala Phe
420	425	430
Asp Leu Asp Gly Ser	Ile Asp Val Glu Thr Pro	Ser Asp Thr Thr Ile
435	440	445
Lys Ile Thr Arg Lys	Asp Lys Ser Gly Glu Asn	Val Ala Val Leu Val
450	455	460
Ala Asn Ala Ala Asp	Lys Thr Phe Thr Ile Thr	Ala Asn Gly Glu Glu
465	470	475
Ile Leu Ala Asn Thr	Glu Ala Asp Lys Gln Gln	Leu
485	490	

<210> 17

<211> 1512

<212> DNA

<213> Artificial Sequence

<220>

<223> Sucrose phosphorylase from Bifidobacterium

<400> 17

```

atgaagaaca aagttcaact gatcacctac gcggaccgtc tgggtgatgg caccatcaag 60
agcatgaccg atattctgcg taccggtttc gacggtgttt atgatggcgt gcacatcctg 120
ccgttcttta ccccgttcga cgggtgcgat gcgggctttg acccgattga tcacaccaag 180
gttgacgagc gtctgggtag ctgggacgat gtggcggaac tgagcaaaac ccacaacatc 240
atggttgatg cgattgtgaa ccacatgagc tgggagagca agcagtttca agacgtttctg 300
gcgaaaggcg aggaaagcga atactatccg atgttcctga ccatgagcag cgtgtttccg 360
aacggtgcga ccgaggaaga cctggcgggt atctatcgtc cgcgtccggg tctgccgttc 420
accactata agtttgcggg taaaacccgt ctggtgtggg ttagcttcac cccgcagcaa 480
gttgatattg acaccgatag cgacaagggc tgggagtacc tgatgagcat ctttgatcaa 540
atggcggcga gccacgttag ctacattcgt ctggacgcgg tgggttatgg cgcgaaagaa 600
gcgggtacca gctgcttcat gacccgaag acctttaa acagctacta taagaaacag 720
gtggagattg cgagcaaagt ggaccgtgtt tatgattttg cgctgccgcc gctgctgctg 780
catgcgctga gcaccggtca tgtggaaccg gttgcgcact ggaccgatat ccgtccgaac 840
aacgcggtga ccgttctgga caccacgat ggtattggcg ttatcgacat tggtagcgac 900
caactggatc gtagcctgaa gggctctggt ccgatgagg acgttgataa cctggtgaac 960
accatccacg cgaacacca cggtgaaagc caagcggcga ccggcgcggc ggcgagcaac 1020

```

```

ctggatctgt accaagttaa cagcacctac tatagcgcgc tgggttgcaa cgaccagcac 1080
tatattgcgg cgcggtgcgg gcagttcttt ctgccgggcg tgccgcaagt ttactatgtg 1140
gggtgcgctgg cgggcaagaa cgacatggag ctgctgcgta aaaccaacaa cggtcgtgat 1200
atcaaccgtc actactatag caccgcgagg attgacgaaa acctgaagcg tccggtgggtt 1260
aaagcgctga acgcgctggc gaagttccgt aacgaactgg acgcgttcga tggtagcttt 1320
agctacacca ccgacgatga caccagcatc agcttcacct ggcgtggcga gaccagccag 1380
gcgaccctga cctttgaacc gaaacgtggg ctgggctgtg acaacaccac cccggtggcg 1440
atgctggagt gggaagacag cgcgggtgat caccgtagcg atgacctgat tgcgaaccgc 1500
ccggtgggtg cg                                     1512

```

<210> 18

<211> 504

<212> PRT

<213> Artificial Sequence

<220>

<223> Sucrose phosphorylase from Bifidobacterium

<400> 18

```

Met Lys Asn Lys Val Gln Leu Ile Thr Tyr Ala Asp Arg Leu Gly Asp
1          5          10          15
Gly Thr Ile Lys Ser Met Thr Asp Ile Leu Arg Thr Arg Phe Asp Gly
20          25          30
Val Tyr Asp Gly Val His Ile Leu Pro Phe Phe Thr Pro Phe Asp Gly
35          40          45
Ala Asp Ala Gly Phe Asp Pro Ile Asp His Thr Lys Val Asp Glu Arg
50          55          60
Leu Gly Ser Trp Asp Asp Val Ala Glu Leu Ser Lys Thr His Asn Ile
65          70          75          80
Met Val Asp Ala Ile Val Asn His Met Ser Trp Glu Ser Lys Gln Phe
85          90          95
Gln Asp Val Leu Ala Lys Gly Glu Glu Ser Glu Tyr Tyr Pro Met Phe
100         105         110
Leu Thr Met Ser Ser Val Phe Pro Asn Gly Ala Thr Glu Glu Asp Leu
115         120         125
Ala Gly Ile Tyr Arg Pro Arg Pro Gly Leu Pro Phe Thr His Tyr Lys
130         135         140
Phe Ala Gly Lys Thr Arg Leu Val Trp Val Ser Phe Thr Pro Gln Gln
145         150         155         160
Val Asp Ile Asp Thr Asp Ser Asp Lys Gly Trp Glu Tyr Leu Met Ser
165         170         175
Ile Phe Asp Gln Met Ala Ala Ser His Val Ser Tyr Ile Arg Leu Asp
180         185         190
Ala Val Gly Tyr Gly Ala Lys Glu Ala Gly Thr Ser Cys Phe Met Thr
195         200         205
Pro Lys Thr Phe Lys Leu Ile Ser Arg Leu Arg Glu Glu Gly Val Lys
210         215         220
Arg Gly Leu Glu Ile Leu Ile Glu Val His Ser Tyr Tyr Lys Lys Gln
225         230         235         240
Val Glu Ile Ala Ser Lys Val Asp Arg Val Tyr Asp Phe Ala Leu Pro
245         250         255
Pro Leu Leu Leu His Ala Leu Ser Thr Gly His Val Glu Pro Val Ala

```



```

gaaaaggcgc gtgcgctggg tattgaggtg ctggttgaaa tccacagcca ctaccgtaaa 720
cagatcga aa ttgcgctgca agtggactgg gtttatgatt ttgcgctgcc gccgctgggtg 780
ctgcatgcgc tgtttgcgag cgacccgcac ccgctggcgc agtggctgag catcagcccg 840
cgtaacgcgg tgaccgttct ggacaccac gatggcatcg gtgtgattga tgttggtgcg 900
gatgcggaag gtaacccggg tctgctgagc ccggcggcga ttgacagcct ggttgagacc 960
atccacagcc gtagccaagg tcaaagccgt gaagcgaccg gtgcggcggc gaacaacctg 1020
gatctgtacc aagtgaactg caccttcctg gatgcgctgg gcggtcgtga accggattat 1080
ctgattgcgc gtgcgctgca gttctttgcg ccgggcatcc cgcaagtgtg ctatgttggt 1140
ctgctgggcg gtaccaacga catggatctg ctgggtcgta gcggtgtggg tcgtgacatt 1200
aaccgtcact actataccga cgcggagatt gatgcggcgc tggcgcgtcc gctggtgcgt 1260
accctgatcg cgctgattcg tctgcgtaac acccaccgg cgttcgcggg tgaatttgat 1320
gtgagcggtc cggcggcgac ccagattcgt ctgcgttggc agcgtcaaga gcactggatc 1380
gaactgcacg ttgacctgag catcccgaag gcgagcatta ccggcaccgg tatccaccg 1440
atcaccattc cgggtgcggc ggatgcgggt gcgccgagc 1479

```

<210> 20

<211> 493

<212> PRT

<213> Artificial Sequence

<220>

<223> Sucrose phosphorylase from uncultured bacterium

<400> 20

```

Met Lys Asn Lys Val Gln Leu Ile Ala Tyr Val Asp Arg Ile Ser Gly
1           5           10           15
Gly Gly Phe Arg Lys Leu His Ala Leu Leu Thr Gly Pro Leu Ala Glu
20           25           30
Ile Phe Gly Gly Ala His Leu Leu Pro Phe Phe Thr Pro Ile Asp Gly
35           40           45
Ala Asp Ala Gly Phe Asp Pro Ser Asp His Thr Gln Val Asp Pro Arg
50           55           60
Leu Gly Thr Trp Asp Asp Val Arg Ile Leu Gly Gly Ala Ile Glu Leu
65           70           75           80
Val Ala Asp Leu Ile Val Asn His Val Ser Ser Ser Ser Pro Gln Phe
85           90           95
Ile Asp Tyr Ser Lys Lys Gly Ser Asp Ser Leu Tyr Ala Gly Met Phe
100          105          110
Leu Thr Tyr Asp Arg Val Phe Pro Glu Gly Ala Arg Glu Ala Asp Ile
115          120          125
Leu Arg Ile Tyr Arg Pro Arg Pro Thr Leu Pro Phe Ser Pro Val Thr
130          135          140
Leu Ser Ser Arg Glu Arg Lys Leu Leu Trp Thr Thr Phe Asn Pro Glu
145          150          155          160
Gln Val Asp Ile Asp Val Arg His Pro Glu Ala Glu Ala Tyr Leu His
165          170          175
Ser Ile Leu Lys Lys Phe Gln Ala Ala Gly Ile Arg Met Ile Arg Leu
180          185          190
Asp Ala Val Gly Tyr Ala Ile Lys Lys Pro Gly Ala Ser Cys Phe Met
195          200          205
Ile Pro Glu Thr Phe Asp Phe Ile Ala Glu Leu Thr Glu Lys Ala Arg
210          215          220

```

Ala	Leu	Gly	Ile	Glu	Val	Leu	Val	Glu	Ile	His	Ser	His	Tyr	Arg	Lys
225					230					235					240
Gln	Ile	Glu	Ile	Ala	Arg	Gln	Val	Asp	Trp	Val	Tyr	Asp	Phe	Ala	Leu
				245					250						255
Pro	Pro	Leu	Val	Leu	His	Ala	Leu	Phe	Ala	Ser	Asp	Pro	His	Pro	Leu
			260						265						270
Ala	Gln	Trp	Leu	Ser	Ile	Ser	Pro	Arg	Asn	Ala	Val	Thr	Val	Leu	Asp
		275						280					285		
Thr	His	Asp	Gly	Ile	Gly	Val	Ile	Asp	Val	Gly	Ala	Asp	Ala	Glu	Gly
	290					295					300				
Asn	Pro	Gly	Leu	Leu	Ser	Pro	Ala	Ala	Ile	Asp	Ser	Leu	Val	Glu	Thr
305					310					315					320
Ile	His	Ser	Arg	Ser	Gln	Gly	Gln	Ser	Arg	Glu	Ala	Thr	Gly	Ala	Ala
				325					330						335
Ala	Asn	Asn	Leu	Asp	Leu	Tyr	Gln	Val	Asn	Cys	Thr	Phe	Leu	Asp	Ala
			340					345						350	
Leu	Gly	Gly	Arg	Glu	Pro	Asp	Tyr	Leu	Ile	Ala	Arg	Ala	Leu	Gln	Phe
		355					360					365			
Phe	Ala	Pro	Gly	Ile	Pro	Gln	Val	Tyr	Tyr	Val	Gly	Leu	Leu	Gly	Gly
		370				375						380			
Thr	Asn	Asp	Met	Asp	Leu	Leu	Gly	Arg	Ser	Gly	Val	Gly	Arg	Asp	Ile
385					390					395					400
Asn	Arg	His	Tyr	Tyr	Thr	Asp	Ala	Glu	Ile	Asp	Ala	Ala	Leu	Ala	Arg
			405						410					415	
Pro	Leu	Val	Arg	Thr	Leu	Ile	Ala	Leu	Ile	Arg	Leu	Arg	Asn	Thr	His
			420					425						430	
Pro	Ala	Phe	Ala	Gly	Glu	Phe	Asp	Val	Ser	Val	Pro	Ala	Ala	Thr	Gln
		435					440					445			
Ile	Arg	Leu	Arg	Trp	Gln	Arg	Gln	Glu	His	Trp	Ile	Glu	Leu	His	Val
	450				455						460				
Asp	Leu	Ser	Ile	Pro	Lys	Ala	Ser	Ile	Thr	Gly	Thr	Gly	Ile	His	Pro
465					470					475					480
Ile	Thr	Ile	Pro	Gly	Ala	Ala	Asp	Ala	Gly	Ala	Pro	Ser			
			485						490						

<210> 21

<211> 1467

<212> DNA

<213> Artificial Sequence

<220>

<223> Sucrose phosphorylase from *Oenococcus*

<400> 21

```

atgccggtga agaacaaagc gatgctgatac acctacagcg atagcatggg taaaaacatc 60
aaggaactgc agtacattct ggacaaatat atcggcgatg cgatttggtg cgttcacctg 120
ctgccgttct ttccgagcac cggatgacgt ggcttcgcgc cgagcgacta taccctgtgtg 180
aaccgggact ttggtgattg ggaggacgtt gaggaactgg gcaagaaata ctatctgatg 240
ttcgacttta tgatcaacca cattagccgt gaaagcatta tgtaccagga cttcaaagag 300
aagaaagatg cgagcagcta taaggacttc tttatccgtt gggaaaaatt ttggccgaag 360
ggctcgtccga ccaaggcgga catcgatctg atttacaagc gtaaagataa ggccgcatc 420
cagggtatta ccttcgcgga cggcagccaa gaacacctgt ggaacacctt tggtgacgag 480
caaatcgata ttaactgtaa aagcaaggtt gcgcaggagt tctttaaaga taccctgcaa 540

```

```

agcatggtga agcacggcgc ggacctgatt cgtctggatg cgttcgcgta cgcgatcaag 600
aaaattgata ccaacgactt ctttatcgaa ccggagattt gggatctgct ggaaagcggt 660
cgtaaaattc tggacccgct gcatgcggag atcctgccgg aaatttacga gcactatacc 720
atccccggcga agattaacga atacggttat ttcacctacg actttgtgct gccgctgggt 780
atcctgtaca ccctgtatag cggcaacccg aaacagctgg cgaaatggct gaagatgagc 840
ccgaagaaac aattcaccac cctggatacc cagcagcgta tcggcgtggg tgatgcgcgt 900
gacatcctga ccgacgagga aattgattac accagcagcg agctgtataa agtgggtgcg 960
aacgttaagc gtacctacag cagcgcggcg tataacaacc tggacatcta ccagattaac 1020
agcacctact atagcgcgct gggtaacgac gataaagcgt atctgctggc gcgtgcatc 1080
cagatttttg cgccgggcat cccgcaaatt tactatgcgg gtctgctggc gggcgaaaac 1140
gacctggatc tgctggaaaa aaccaaagag ggtcgtaca ttaaccgtca ctactatagc 1200
gaggaagagg tggcgaacga ggttcaacgt ccgatcgttg cgtgcctgct gaagctgctg 1260
gcgtggcgta accgtagcgc ggcgttcgac ctgcagggtg atattcaagt tagcgcgacc 1320
gataaaaacg aaatcaagat cattcgtacc agcaccaacg gtcaagacac cgcggagctg 1380
accgcgaacg tggcgtgtaa aacctttacc atcaaagaaa acgacaagat cattctgatt 1440
gaggatcaaa ccgacaccaa ggatattc                                     1467

```

<210> 22

<211> 489

<212> PRT

<213> Artificial Sequence

<220>

<223> Sucrose phosphorylase from *Oenococcus*

<400> 22

```

Met Pro Val Lys Asn Lys Ala Met Leu Ile Thr Tyr Ser Asp Ser Met
1           5           10          15
Gly Lys Asn Ile Lys Glu Leu Gln Tyr Ile Leu Asp Lys Tyr Ile Gly
20          25          30
Asp Ala Ile Gly Gly Val His Leu Leu Pro Phe Phe Pro Ser Thr Gly
35          40          45
Asp Arg Gly Phe Ala Pro Ser Asp Tyr Thr Arg Val Asn Pro Asp Phe
50          55          60
Gly Asp Trp Glu Asp Val Glu Glu Leu Gly Lys Lys Tyr Tyr Leu Met
65          70          75          80
Phe Asp Phe Met Ile Asn His Ile Ser Arg Glu Ser Ile Met Tyr Gln
85          90          95
Asp Phe Lys Glu Lys Lys Asp Ala Ser Ser Tyr Lys Asp Phe Phe Ile
100         105         110
Arg Trp Glu Lys Phe Trp Pro Lys Gly Arg Pro Thr Lys Ala Asp Ile
115         120         125
Asp Leu Ile Tyr Lys Arg Lys Asp Lys Ala Pro Ile Gln Gly Ile Thr
130         135         140
Phe Ala Asp Gly Ser Gln Glu His Leu Trp Asn Thr Phe Gly Asp Glu
145         150         155         160
Gln Ile Asp Ile Asn Val Lys Ser Lys Val Ala Gln Glu Phe Phe Lys
165         170         175
Asp Thr Leu Gln Ser Met Val Lys His Gly Ala Asp Leu Ile Arg Leu
180         185         190
Asp Ala Phe Ala Tyr Ala Ile Lys Lys Ile Asp Thr Asn Asp Phe Phe

```



```

ttcgacttca tgatcaacca catcagccgt gaaagcaaat tctttcagga tttcaagcaa 300
aaccacgagg cgagcccgtg caaggaaatg tttattcgtg tccacgagtt ctttccggaa 360
aaccgtccga ccccgaggga cattgatctg atctataagc gtaaagacaa ggcgccgttc 420
caggaagtga cttttgcgga tgggtaccacc gaacaggttt ggaacacctt cggcgaggaa 480
caaattgacc tggatgtgaa caaagaagtt accaagcaat tcatcaagga caccctgaag 540
gatatggcga gccacgggtg cagcctgatt cgtctggacg cgttcgcgta cgcgatcaag 600
aaactggaca ccaacgattt ctttgtggag ccggaaattt gggacctgct ggatgaggtt 660
cgtgaggaag cggcgaaata caactgcgaa ctgctgccgg agatccacga acactatacc 720
attcagatga agatcgcgga ccacgattac tatgtgtacg acttcgcgct gccgatgac 780
accctgtaca gcctgtatag cggtaaaacc aaccgtctgg cgaactggct gaaaatgagc 840
ccgatgaagc aatttaccac cctggacacc cagcatggta ttggcgtggg tgacgcgaaa 900
gatctgctga ccgacgagga actggattac accagcgagg aactgtataa agtgggcgcg 960
aacgttaaga aaatctacag cagcgcgagc tataacaacc tggacattta ccagatcaac 1020
agcacctact atagcgcgct gggtgacaac gatcgtagct atctgctggc gcgtgcgatt 1080
cagtgtttg cgccgggcat tccgcaaata tactatgtgg gtctgctggc gggcaaaaac 1140
gacctggagc tgctggaaag caccaaagag ggtcgttaaca ttaaccgtca ctactatagc 1200
ctgaccgaaa tcgatgagga agttaaccgt ccggtgggta aggcgctgtt cgacctgctg 1260
cgttttcgta accgtgcggc ggcgtttgat ctggcgggcg agatcaccgt ggaaaccccg 1320
agcgaggaaa ccatcattat caccgtaaa gacgcgaacg gtgatcgtgc ggttctgacc 1380
gcgaacctgc agaccaagga ctttaccgtg acccacaacg gcgaactggg tctgcagcaa 1440
ctggat 1446

```

<210> 24

<211> 482

<212> PRT

<213> Artificial Sequence

<220>

<223> Sucrose phosphorylase from Enterococcus

<400> 24

```

Met Lys Ile Lys Asn Glu Ala Met Leu Ile Thr Tyr Cys Asp Ser Leu
1           5           10          15
Gly Asn Asn Leu Lys Asp Leu Asn Gln Val Leu Asp Gln His Leu Gln
          20          25          30
Gly Val Val Gly Gly Ile His Leu Leu Pro Phe Phe Pro Ser Thr Gly
          35          40          45
Asp Arg Gly Phe Ala Pro Ser Asp Tyr Thr Val Val Asp Ser Ala Phe
          50          55          60
Gly Thr Trp Gln Glu Val Glu Ala Leu Gly Glu Lys Tyr Tyr Leu Met
65          70          75          80
Phe Asp Phe Met Ile Asn His Ile Ser Arg Glu Ser Lys Phe Phe Gln
          85          90          95
Asp Phe Lys Gln Asn His Glu Ala Ser Pro Tyr Lys Glu Met Phe Ile
          100         105         110
Arg Ile His Glu Phe Phe Pro Glu Asn Arg Pro Thr Pro Glu Asp Ile
          115         120         125
Asp Leu Ile Tyr Lys Arg Lys Asp Lys Ala Pro Phe Gln Glu Val Thr
          130         135         140
Phe Ala Asp Gly Thr Thr Glu Gln Val Trp Asn Thr Phe Gly Glu Glu
145         150         155         160
Gln Ile Asp Leu Asp Val Asn Lys Glu Val Thr Lys Gln Phe Ile Lys

```

				165				170					175				
Asp	Thr	Leu	Lys	Asp	Met	Ala	Ser	His	Gly	Cys	Ser	Leu	Ile	Arg	Leu		
			180					185					190				
Asp	Ala	Phe	Ala	Tyr	Ala	Ile	Lys	Lys	Leu	Asp	Thr	Asn	Asp	Phe	Phe		
		195					200					205					
Val	Glu	Pro	Glu	Ile	Trp	Asp	Leu	Leu	Asp	Glu	Val	Arg	Glu	Glu	Ala		
	210					215					220						
Ala	Lys	Tyr	Asn	Cys	Glu	Leu	Leu	Pro	Glu	Ile	His	Glu	His	Tyr	Thr		
225					230					235					240		
Ile	Gln	Met	Lys	Ile	Ala	Asp	His	Asp	Tyr	Tyr	Val	Tyr	Asp	Phe	Ala		
			245					250						255			
Leu	Pro	Met	Ile	Thr	Leu	Tyr	Ser	Leu	Tyr	Ser	Gly	Lys	Thr	Asn	Arg		
			260					265					270				
Leu	Ala	Asn	Trp	Leu	Lys	Met	Ser	Pro	Met	Lys	Gln	Phe	Thr	Thr	Leu		
		275				280						285					
Asp	Thr	His	Asp	Gly	Ile	Gly	Val	Val	Asp	Ala	Lys	Asp	Leu	Leu	Thr		
	290					295					300						
Asp	Glu	Glu	Leu	Asp	Tyr	Thr	Ser	Glu	Glu	Leu	Tyr	Lys	Val	Gly	Ala		
305					310					315					320		
Asn	Val	Lys	Lys	Ile	Tyr	Ser	Ser	Ala	Ser	Tyr	Asn	Asn	Leu	Asp	Ile		
			325					330						335			
Tyr	Gln	Ile	Asn	Ser	Thr	Tyr	Tyr	Ser	Ala	Leu	Gly	Asp	Asn	Asp	Arg		
		340						345					350				
Ser	Tyr	Leu	Leu	Ala	Arg	Ala	Ile	Gln	Cys	Phe	Ala	Pro	Gly	Ile	Pro		
	355					360						365					
Gln	Ile	Tyr	Tyr	Val	Gly	Leu	Leu	Ala	Gly	Lys	Asn	Asp	Leu	Glu	Leu		
	370					375					380						
Leu	Glu	Ser	Thr	Lys	Glu	Gly	Arg	Asn	Ile	Asn	Arg	His	Tyr	Tyr	Ser		
385					390					395					400		
Leu	Thr	Glu	Ile	Asp	Glu	Glu	Val	Asn	Arg	Pro	Val	Val	Lys	Ala	Leu		
			405					410						415			
Phe	Asp	Leu	Leu	Arg	Phe	Arg	Asn	Arg	Ala	Ala	Ala	Phe	Asp	Leu	Ala		
		420					425						430				
Gly	Glu	Ile	Thr	Val	Glu	Thr	Pro	Ser	Glu	Glu	Thr	Ile	Ile	Ile	Thr		
	435					440						445					
Arg	Lys	Asp	Ala	Asn	Gly	Asp	Arg	Ala	Val	Leu	Thr	Ala	Asn	Leu	Gln		
	450				455						460						
Thr	Lys	Asp	Phe	Thr	Val	Thr	His	Asn	Gly	Glu	Leu	Val	Leu	Gln	Gln		
465				470					475					480			
Leu	Asp																

<210> 25

<211> 1440

<212> DNA

<213> Artificial Sequence

<220>

<223> Sucrose phosphorylase from Streptococcus

<400> 25

atgagcatcc agaacaagac catgctgatt acctacagcg atagcctggg taacaacctg 60
aaggacctgt acgagaacct ggaaaaatat ttcggcgatg cggtgggtgg cgttcacctg 120

```

ctgccgttct ttccgagcac cggatgatcgt ggcttcgcgc cggatggatta tgaccagggtt 180
gacccggcgt ttgggtgattg ggaggacgtg aagcgtctgg gcgataagta ctacctgatg 240
ttcgacttca tgatcaacca cattagccgt caaagcaagt actataaaaa ctaccaggag 300
aagcacgacc aaagcgcgta taaagacctg ttcctgaact gggacaagtt ttggccggaa 360
aacgtccga cccaggcgga tgttgacctg atctacaagc gtaaagatcg tgcgccgaaa 420
caagagatta ccttcgcgga cggatccacc gaacacctgt ggaacacctt tggcgaggaa 480
caaatcgatc tggacgtgac caaagagggt accatggact tcatccgtaa aaccattgaa 540
cacctggtta gcaacgggtg cgacctgatt cgtctggatg cgtttgcgta cgcggtgaag 600
aaactggata ccaacgactt ctttgttgag ccgatatct gggacctgct ggataagggtt 660
cgtgatatgg cggcgagcta cggcgcgga ctgctgccgg agattcacga acactatagc 720
atccagttca aaattgcgga tcacgactac tatgtgtacg actttgcgct gccgatgggtt 780
accctgtaca ccctgtatag cagcaagggt gagcgtctgg cgaagtggct gaaaatgagc 840
ccgatgaaac aattcaccac cctggacacc cacgatggta tcggcgtgggt tgacgttaag 900
gatatcctga ccgacgagga aattgattac gcgagcaacg agctgtataa agtgggtgctg 960
aacgttaagc gtaaatacag caccgcgga tataacaacc tggatatcta ccagattaac 1020
agcacctact atagcgcgct gggtgacgat gacagcaagt atttcctggc gcgtctgatc 1080
caggcgtttg cgccgggcat tccgcaagtg tactatgttg gttttctggc gggcaagaac 1140
gacctggagc tgctggaaaa caccaaagaa ggtcgtaaca tcaaccgtca ctactatagc 1200
aacgaggaaa ttgcgagga agtggaaact ccggtgggtta agagcctgct gaaactgttc 1260
agctttcgta acaacagcca agcgttcgat ctggagggca gcatcgaggt ggaaaccccg 1320
gacgaacaca ccatcgttat tacccgtcag aacaaagacc aaaccgcgac cgcggtggcg 1380
cgtatcaacc tggcggtatg tacctaccag gttaccgaga acggccaaga cattgaattt 1440

```

<210> 26

<211> 480

<212> PRT

<213> Artificial Sequence

<220>

<223> Sucrose phosphorylase from Streptococcus

<400> 26

```

Met Ser Ile Gln Asn Lys Thr Met Leu Ile Thr Tyr Ser Asp Ser Leu
1           5           10          15
Gly Asn Asn Leu Lys Asp Leu Tyr Glu Asn Leu Glu Lys Tyr Phe Gly
          20          25          30
Asp Ala Val Gly Gly Val His Leu Leu Pro Phe Phe Pro Ser Thr Gly
          35          40          45
Asp Arg Gly Phe Ala Pro Val Asp Tyr Asp Gln Val Asp Pro Ala Phe
          50          55          60
Gly Asp Trp Glu Asp Val Lys Arg Leu Gly Asp Lys Tyr Tyr Leu Met
65          70          75          80
Phe Asp Phe Met Ile Asn His Ile Ser Arg Gln Ser Lys Tyr Tyr Lys
          85          90          95
Asn Tyr Gln Glu Lys His Asp Gln Ser Ala Tyr Lys Asp Leu Phe Leu
          100         105         110
Asn Trp Asp Lys Phe Trp Pro Glu Asn Arg Pro Thr Gln Ala Asp Val
          115         120         125
Asp Leu Ile Tyr Lys Arg Lys Asp Arg Ala Pro Lys Gln Glu Ile Thr
          130         135         140
Phe Ala Asp Gly Thr Thr Glu His Leu Trp Asn Thr Phe Gly Glu Glu
145         150         155         160

```

Gln	Ile	Asp	Leu	Asp	Val	Thr	Lys	Glu	Val	Thr	Met	Asp	Phe	Ile	Arg
				165					170					175	
Lys	Thr	Ile	Glu	His	Leu	Val	Ser	Asn	Gly	Cys	Asp	Leu	Ile	Arg	Leu
			180					185					190		
Asp	Ala	Phe	Ala	Tyr	Ala	Val	Lys	Lys	Leu	Asp	Thr	Asn	Asp	Phe	Phe
		195					200					205			
Val	Glu	Pro	Asp	Ile	Trp	Asp	Leu	Leu	Asp	Lys	Val	Arg	Asp	Met	Ala
	210				215						220				
Ala	Ser	Tyr	Gly	Ala	Glu	Leu	Leu	Pro	Glu	Ile	His	Glu	His	Tyr	Ser
225				230					235						240
Ile	Gln	Phe	Lys	Ile	Ala	Asp	His	Asp	Tyr	Tyr	Val	Tyr	Asp	Phe	Ala
			245					250						255	
Leu	Pro	Met	Val	Thr	Leu	Tyr	Thr	Leu	Tyr	Ser	Ser	Lys	Val	Glu	Arg
			260					265					270		
Leu	Ala	Lys	Trp	Leu	Lys	Met	Ser	Pro	Met	Lys	Gln	Phe	Thr	Thr	Leu
		275				280						285			
Asp	Thr	His	Asp	Gly	Ile	Gly	Val	Val	Asp	Val	Lys	Asp	Ile	Leu	Thr
	290				295						300				
Asp	Glu	Glu	Ile	Asp	Tyr	Ala	Ser	Asn	Glu	Leu	Tyr	Lys	Val	Gly	Ala
305				310						315					320
Asn	Val	Lys	Arg	Lys	Tyr	Ser	Thr	Ala	Glu	Tyr	Asn	Asn	Leu	Asp	Ile
			325					330						335	
Tyr	Gln	Ile	Asn	Ser	Thr	Tyr	Tyr	Ser	Ala	Leu	Gly	Asp	Asp	Asp	Ser
		340						345					350		
Lys	Tyr	Phe	Leu	Ala	Arg	Leu	Ile	Gln	Ala	Phe	Ala	Pro	Gly	Ile	Pro
		355				360						365			
Gln	Val	Tyr	Tyr	Val	Gly	Phe	Leu	Ala	Gly	Lys	Asn	Asp	Leu	Glu	Leu
	370				375						380				
Leu	Glu	Asn	Thr	Lys	Glu	Gly	Arg	Asn	Ile	Asn	Arg	His	Tyr	Tyr	Ser
385				390						395					400
Asn	Glu	Glu	Ile	Ala	Gln	Glu	Val	Glu	Arg	Pro	Val	Val	Lys	Ser	Leu
			405					410						415	
Leu	Lys	Leu	Phe	Ser	Phe	Arg	Asn	Asn	Ser	Gln	Ala	Phe	Asp	Leu	Glu
		420					425						430		
Gly	Ser	Ile	Glu	Val	Glu	Thr	Pro	Asp	Glu	His	Thr	Ile	Val	Ile	Thr
	435					440						445			
Arg	Gln	Asn	Lys	Asp	Gln	Thr	Ala	Thr	Ala	Val	Ala	Arg	Ile	Asn	Leu
	450				455						460				
Ala	Asp	Gly	Thr	Tyr	Gln	Val	Thr	Glu	Asn	Gly	Gln	Asp	Ile	Glu	Phe
465					470					475					480

<210> 27

<211> 1509

<212> DNA

<213> Artificial Sequence

<220>

<223> Sucrose phosphorylase from Scardovia

<400> 27

atgaagaaca aagttcagct gattacctac gcgcatcgctc tgggtgaagg taccctggcg 60
 agcctgaccg atgtgctgcg taccggtttc gatggcggtt atgaaggtgt gcacattctg 120

```

ccgttcttta ccccgttcga cggcgcggat gcgggttttg acccgatcga tcacaccaag 180
gttgacccgc gtctgggtag ctgggacgat gtggcggagc tgagcaaaac ccacgacatc 240
atggttgatg cgattgtgaa ccacatgagc tggcaaagcg aacagtttca agatgttatg 300
aagaaagggtg aaggcagcga atactatccg atgttcctga ccatgagcgc ggtgtttccg 360
gagggtgcga ccgaggaaga cctggcgggt atttatcgtc cgcgtccggg tctgccgttc 420
accactata cctgggggtgg caagacccgt ctggtttgga ccacctttac cccgcagcaa 480
gtggacatcg ataccgcgag cgacgaaggt tgggcgtacc tgatgagcat cctggatcag 540
atgagcaaaa gccacgttag cttcattcgt ctggacgcgg tgggttatgg cgcgaagcgt 600
gcgaaaacca gctgcttcat gaccccgac acctttgatc tgatcagccg tattaaagag 660
gagagcgcgc gtcgtggcat ggagaccctg atcgaagttc acagctacta taagaaacaa 720
gtggagattg cgaacaaggt ggaccgtgtt tacgattttg cgatcccggc gctgctgctg 780
catgcgctga gcaccggtaa aaccgggtccg ctggcgcact ggctgggtgt gcgtccggtt 840
aacgcggtga acgttctgga caccacgat ggtatcggcg tgattgatgt tggcagcgac 900
cagatggatc gtagcctgaa gggctctggtt ccggacgatg aagttgacca actggtggag 960
accattaaag aaaacagcca tggtgaaagc gcggaagcga ccggtgcggc ggcgagcaac 1020
ctggatctgt accaggttaa ctgcacctac tatagcgcgc tgggctgcga cgatcaaaaa 1080
tacatcgcga gccgtgcggt gcagttctat atgccgggtg tgccgcaagt ttactatgtg 1140
ggcgcgctgg cgggcaagaa cgacatggag ctgctgaaac gtaccaacgt tggctgtgat 1200
atcaaccgtc actactatac cgtggcggag attgacgaaa acctgcgtcg tccggtgggtt 1260
caggcgctga acgcgctggg ccgtttccgt aacaccctgg aggcgtttga cggtgatttc 1320
agctttaccg aaaacgggtg cagcctggcg atgacctgga ccggtcaaag caccagcacc 1380
accctgacct ttgcgccggg caagggtgcg ggtagcagcg gcgtgccggt ttgcaccatc 1440
gcgtggaaaag acgcgagcgg tgaacacgtg accgacgata tcattgcgaa cccgccggcg 1500
gcggcgag                                     1509

```

<210> 28

<211> 503

<212> PRT

<213> Artificial Sequence

<220>

<223> Sucrose phosphorylase from Scardovia

<400> 28

```

Met Lys Asn Lys Val Gln Leu Ile Thr Tyr Ala Asp Arg Leu Gly Glu
1           5           10           15
Gly Thr Leu Ala Ser Leu Thr Asp Val Leu Arg Thr Arg Phe Asp Gly
20           25           30
Val Tyr Glu Gly Val His Ile Leu Pro Phe Phe Thr Pro Phe Asp Gly
35           40           45
Ala Asp Ala Gly Phe Asp Pro Ile Asp His Thr Lys Val Asp Pro Arg
50           55           60
Leu Gly Ser Trp Asp Asp Val Ala Glu Leu Ser Lys Thr His Asp Ile
65           70           75           80
Met Val Asp Ala Ile Val Asn His Met Ser Trp Gln Ser Glu Gln Phe
85           90           95
Gln Asp Val Met Lys Lys Gly Glu Gly Ser Glu Tyr Tyr Pro Met Phe
100          105          110
Leu Thr Met Ser Ala Val Phe Pro Glu Gly Ala Thr Glu Glu Asp Leu
115          120          125
Ala Gly Ile Tyr Arg Pro Arg Pro Gly Leu Pro Phe Thr His Tyr Thr
130          135          140

```

Trp	Gly	Gly	Lys	Thr	Arg	Leu	Val	Trp	Thr	Thr	Phe	Thr	Pro	Gln	Gln
145					150					155					160
Val	Asp	Ile	Asp	Thr	Ala	Ser	Asp	Glu	Gly	Trp	Ala	Tyr	Leu	Met	Ser
				165					170					175	
Ile	Leu	Asp	Gln	Met	Ser	Lys	Ser	His	Val	Ser	Phe	Ile	Arg	Leu	Asp
			180					185					190		
Ala	Val	Gly	Tyr	Gly	Ala	Lys	Arg	Ala	Lys	Thr	Ser	Cys	Phe	Met	Thr
		195					200					205			
Pro	Asp	Thr	Phe	Asp	Leu	Ile	Ser	Arg	Ile	Lys	Glu	Glu	Ser	Ala	Arg
	210					215					220				
Arg	Gly	Met	Glu	Thr	Leu	Ile	Glu	Val	His	Ser	Tyr	Tyr	Lys	Lys	Gln
225					230					235					240
Val	Glu	Ile	Ala	Asn	Lys	Val	Asp	Arg	Val	Tyr	Asp	Phe	Ala	Ile	Pro
				245					250					255	
Ala	Leu	Leu	Leu	His	Ala	Leu	Ser	Thr	Gly	Lys	Thr	Gly	Pro	Leu	Ala
				260				265					270		
His	Trp	Leu	Gly	Val	Arg	Pro	Val	Asn	Ala	Val	Asn	Val	Leu	Asp	Thr
	275					280						285			
His	Asp	Gly	Ile	Gly	Val	Ile	Asp	Val	Gly	Ser	Asp	Gln	Met	Asp	Arg
	290					295					300				
Ser	Leu	Lys	Gly	Leu	Val	Pro	Asp	Asp	Glu	Val	Asp	Gln	Leu	Val	Glu
305					310					315					320
Thr	Ile	Lys	Glu	Asn	Ser	His	Gly	Glu	Ser	Ala	Glu	Ala	Thr	Gly	Ala
				325				330						335	
Ala	Ala	Ser	Asn	Leu	Asp	Leu	Tyr	Gln	Val	Asn	Cys	Thr	Tyr	Tyr	Ser
			340					345					350		
Ala	Leu	Gly	Cys	Asp	Asp	Gln	Lys	Tyr	Ile	Ala	Ser	Arg	Ala	Val	Gln
		355					360					365			
Phe	Tyr	Met	Pro	Gly	Val	Pro	Gln	Val	Tyr	Tyr	Val	Gly	Ala	Leu	Ala
	370					375					380				
Gly	Lys	Asn	Asp	Met	Glu	Leu	Leu	Lys	Arg	Thr	Asn	Val	Gly	Arg	Asp
385					390					395					400
Ile	Asn	Arg	His	Tyr	Tyr	Thr	Val	Ala	Glu	Ile	Asp	Glu	Asn	Leu	Arg
			405						410					415	
Arg	Pro	Val	Val	Gln	Ala	Leu	Asn	Ala	Leu	Gly	Arg	Phe	Arg	Asn	Thr
			420					425					430		
Leu	Glu	Ala	Phe	Asp	Gly	Asp	Phe	Ser	Phe	Thr	Glu	Asn	Gly	Gly	Ser
		435					440					445			
Leu	Ala	Met	Thr	Trp	Thr	Gly	Gln	Ser	Thr	Ser	Thr	Thr	Leu	Thr	Phe
		450				455					460				
Ala	Pro	Gly	Lys	Gly	Ala	Gly	Ser	Ser	Gly	Val	Pro	Val	Cys	Thr	Ile
465					470					475					480
Ala	Trp	Lys	Asp	Ala	Ser	Gly	Glu	His	Val	Thr	Asp	Asp	Ile	Ile	Ala
				485					490					495	
Asn	Pro	Pro	Ala	Ala	Ala	Gln									
			500												

<210> 29

<211> 1506

<212> DNA

<213> Artificial Sequence

<220>

<223> Sucrose phosphorylase from Parascardovia

<400> 29

```
atgaagaaca aagttcagct gatcacctac gcgaaccgtt tcggcgaggg caccattcgt 60
agcctgaccg acgtgctgcg taccctgttt gatggtgttt acgaaggcgt gcacattctg 120
ccgttcttta ccccgttcga cgggtgcggat gcgggctttg acccgattga tcacaccaag 180
gttgatccgc gtctgggtac ctgggacgat atcgcggagc tgagcaaaac ccacgacatc 240
atggttgatg cgattgtgaa ccacatgagc tgggagagca agcagttcca agacgttatg 300
aaaaacggtg aagcgagccc gtactatggc atgttcctga ccatgagcgc ggtgtttccg 360
gatggtgcca ccgaggaaga tctggcgggt atctatcgtc cgcgtccggg tctgccgttc 420
accactata actgggggtgg caagaccctg ctggtttgga ccacctttac cccgcagcaa 480
gtggacatca acaccgatag cattgaaggc tgggactacc tgatgagcat cctggatcag 540
atgagccaaa gccacgttag cttcattcgt ctggacgcgg tgggttatgg cgcgaagcag 600
gcgaaaacca cctgcttcat gaccccgaag acctttgagc tgatcggtcg tattaaagag 660
gagagcgcga aacgtggcat ggagaccctg atcgaagttc acagccacta catgaagcaa 720
gtggaaattg cgggtaaagt ggaccgtgtt tatgattttg cgattccggg tctgctgctg 780
catgcgctga ccaccggtaa aaccggtccg ctggcgcgtt ggatggaagt gcgtccgggt 840
aacgcggtga acgttctgga caccacgat ggtattggcg ttatcgatat tggtagcgac 900
cagatggatc gtagcctgaa gggctctggtg ccggacgagg aagttgatca actggtggag 960
accatcaaag aaaacagcca cggtgaaagc gcggcggcga ccggcgcggc ggcgagcaac 1020
ctggacctgt accaggttaa ctgcacctac tatagcgcgc tgggttgcca cgatcaaaaa 1080
tatattgcgg cgcgtgcggt gcagttcttt atgccgggcg tgccgcaagt ttactatgtg 1140
ggtgcgctgg cgggcgagaa cgacatgcag ctgctgaagg aaaccaacgt tggtcgtgat 1200
atcaaccgtc actactatag cctggaggaa attgacaaaa acctggagcg tccggtggtt 1260
aaagcgctga acgcgctgtg ccgtttccgt aacggtctgg aagcgtttga cggcgatttc 1320
agctttgctc agaacggtga tagcctgacc ttcgcgtgga ccggcaagga gaccgaagcg 1380
accctgacct ttaccccggc gcaaggtgtt ggcaaagatg aggcgccggt gtgcagcctg 1440
cgttggaagg acagcgcggg tgatcacaaa accgatgatc tgattgcgga accgccggcg 1500
gtggcg                                     1506
```

<210> 30

<211> 502

<212> PRT

<213> Artificial Sequence

<220>

<223> Sucrose phosphorylase from Parascardovia

<400> 30

```
Met Lys Asn Lys Val Gln Leu Ile Thr Tyr Ala Asn Arg Phe Gly Glu
1           5           10          15
Gly Thr Ile Arg Ser Leu Thr Asp Val Leu Arg Thr Arg Phe Asp Gly
20          25          30
Val Tyr Glu Gly Val His Ile Leu Pro Phe Phe Thr Pro Phe Asp Gly
35          40          45
Ala Asp Ala Gly Phe Asp Pro Ile Asp His Thr Lys Val Asp Pro Arg
50          55          60
Leu Gly Thr Trp Asp Asp Ile Ala Glu Leu Ser Lys Thr His Asp Ile
65          70          75          80
Met Val Asp Ala Ile Val Asn His Met Ser Trp Glu Ser Lys Gln Phe
85          90          95
```

Gln	Asp	Val	Met	Lys	Asn	Gly	Glu	Ala	Ser	Pro	Tyr	Tyr	Gly	Met	Phe	
			100					105					110			
Leu	Thr	Met	Ser	Ala	Val	Phe	Pro	Asp	Gly	Ala	Thr	Glu	Glu	Asp	Leu	
		115					120					125				
Ala	Gly	Ile	Tyr	Arg	Pro	Arg	Pro	Gly	Leu	Pro	Phe	Thr	His	Tyr	Asn	
	130					135					140					
Trp	Gly	Gly	Lys	Thr	Arg	Leu	Val	Trp	Thr	Thr	Phe	Thr	Pro	Gln	Gln	
145					150					155					160	
Val	Asp	Ile	Asn	Thr	Asp	Ser	Ile	Glu	Gly	Trp	Asp	Tyr	Leu	Met	Ser	
			165					170					175			
Ile	Leu	Asp	Gln	Met	Ser	Gln	Ser	His	Val	Ser	Phe	Ile	Arg	Leu	Asp	
			180					185					190			
Ala	Val	Gly	Tyr	Gly	Ala	Lys	Gln	Ala	Lys	Thr	Thr	Cys	Phe	Met	Thr	
		195					200					205				
Pro	Lys	Thr	Phe	Glu	Leu	Ile	Gly	Arg	Ile	Lys	Glu	Glu	Ser	Ala	Lys	
	210					215					220					
Arg	Gly	Met	Glu	Thr	Leu	Ile	Glu	Val	His	Ser	His	Tyr	Met	Lys	Gln	
225					230				235						240	
Val	Glu	Ile	Ala	Gly	Lys	Val	Asp	Arg	Val	Tyr	Asp	Phe	Ala	Ile	Pro	
			245					250						255		
Gly	Leu	Leu	Leu	His	Ala	Leu	Thr	Thr	Gly	Lys	Thr	Gly	Pro	Leu	Ala	
			260					265					270			
Arg	Trp	Met	Glu	Val	Arg	Pro	Gly	Asn	Ala	Val	Asn	Val	Leu	Asp	Thr	
		275					280					285				
His	Asp	Gly	Ile	Gly	Val	Ile	Asp	Ile	Gly	Ser	Asp	Gln	Met	Asp	Arg	
	290					295					300					
Ser	Leu	Lys	Gly	Leu	Val	Pro	Asp	Glu	Glu	Val	Asp	Gln	Leu	Val	Glu	
305					310					315					320	
Thr	Ile	Lys	Glu	Asn	Ser	His	Gly	Glu	Ser	Ala	Ala	Ala	Thr	Gly	Ala	
				325				330						335		
Ala	Ala	Ser	Asn	Leu	Asp	Leu	Tyr	Gln	Val	Asn	Cys	Thr	Tyr	Tyr	Ser	
			340					345					350			
Ala	Leu	Gly	Cys	Asp	Asp	Gln	Lys	Tyr	Ile	Ala	Ala	Arg	Ala	Val	Gln	
		355					360					365				
Phe	Phe	Met	Pro	Gly	Val	Pro	Gln	Val	Tyr	Tyr	Val	Gly	Ala	Leu	Ala	
						375					380					
Gly	Glu	Asn	Asp	Met	Gln	Leu	Leu	Lys	Glu	Thr	Asn	Val	Gly	Arg	Asp	
385					390				395						400	
Ile	Asn	Arg	His	Tyr	Tyr	Ser	Leu	Glu	Glu	Ile	Asp	Gln	Asn	Leu	Glu	
			405					410						415		
Arg	Pro	Val	Val	Lys	Ala	Leu	Asn	Ala	Leu	Cys	Arg	Phe	Arg	Asn	Gly	
			420					425					430			
Leu	Glu	Ala	Phe	Asp	Gly	Asp	Phe	Ser	Phe	Ala	Gln	Asn	Gly	Asp	Ser	
		435					440						445			
Leu	Thr	Phe	Ala	Trp	Thr	Gly	Lys	Glu	Thr	Glu	Ala	Thr	Leu	Thr	Phe	
	450					455					460					
Thr	Pro	Ala	Gln	Gly	Val	Gly	Lys	Asp	Glu	Ala	Pro	Val	Cys	Ser	Leu	
465					470					475					480	
Arg	Trp	Lys	Asp	Ser	Ala	Gly	Asp	His	Lys	Thr	Asp	Asp	Leu	Ile	Ala	
			485					490						495		
Glu	Pro	Pro	Ala	Val	Ala											
			500													

<210> 31
 <211> 1500
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Sucrose phosphorylase from Alloscardovia

<400> 31
 atgaagaaca aagttcaact gattacctac gcggatcgtc tgggtgacgg caccctgaag 60
 agcatgaccg agaccctgcg taaacacttc gaggggtgttt atgaaggcgt gcacatcctg 120
 ccgttcttta ccccgttcga tgggtgcggac gcgggctttg atccggttga ccacaccaaa 180
 gtggacccgc gtctgggtag ctgggacgat gtggcggaac tgagcaccac ccacgatatc 240
 atggttgaca ccattgtgaa ccacatgagc tgggagagcg aacagttcca agatgttatg 300
 gcgaagggcg aggacagcga atactatccg atgttcctga ccatgagcag catttttccg 360
 gatggtgtga ccgaggaaga cctgaccgcg atctatcgtc cgcgtccggg tctgccgttc 420
 acccactata actggggtgg caaaaccgt ctggtttgga ccacctttac cccgcagcaa 480
 gtggacattg ataccgacag cgagatgggt tggaaactacc tgctgagcat cctggatcag 540
 ctgagccaaa gccacgttag ccaaattcgt ctggacgcgg tgggttatgg cgcgaaggag 600
 aaaaacagca gctgcttcat gaccccgaag acctttaaac tgatcgaacg tattaaggcg 660
 gagggtgaaa aacgtggcct ggagaccctg atcgaagttc acagctacta taagaaacag 720
 gtggagattg cgagcaaggt ggatcgtgtt tacgactttg cgatcccggg tctgctgctg 780
 catgcgctgg aatttggcaa aaccgatgcg ctggcgcaat ggattgacgt tcgtccgaac 840
 aacgcggtga acgttctgga taccacgac ggtatcggcg ttatcgatat tggtagcgat 900
 cagatggacc gtagcctggc gggctctggtg ccggatgagg aagttgacgc gctggttgag 960
 agcatccacc gtaacagcaa gggtgaaagc caggaagcga ccggcgcggc ggcgagcaac 1020
 ctggacctgt accaagttaa ctgcacctac tatgcggcgc tgggtagcga cgatcagaaa 1080
 tatattgcgg cgcgtgcggt gcagttcttt atgccgggcg tgccgcaagt ttactatgtg 1140
 ggtgcgctgg cgggcagcaa cgatatggac ctgctgaagc gtaccaacgt gggctcgtgat 1200
 atcaaccgtc actactatag cgcggcggaa gtggcgagcg aggtggaacg tccggtgggt 1260
 caggcgctga acgcgctggg ccgtttccgt aacaccctga gcgcgttcga tggatgaattt 1320
 agctacagca acgcggacgg cgttctgacc atgacctggg cggatgatgc gaccctgctg 1380
 accctgacct ttgcgccgaa ggcaacagc aacggtgcga gcgtggcgcg tctggagtgg 1440
 accgatgcgg cgggtgaaca tgcgaccgac gatctgatcg cgaaccgcc ggtggttgcg 1500

<210> 32
 <211> 500
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Sucrose phosphorylase from Alloscardovia

<400> 32
 Met Lys Asn Lys Val Gln Leu Ile Thr Tyr Ala Asp Arg Leu Gly Asp
 1 5 10 15
 Gly Thr Leu Lys Ser Met Thr Glu Thr Leu Arg Lys His Phe Glu Gly
 20 25 30
 Val Tyr Glu Gly Val His Ile Leu Pro Phe Phe Thr Pro Phe Asp Gly
 35 40 45
 Ala Asp Ala Gly Phe Asp Pro Val Asp His Thr Lys Val Asp Pro Arg

50	55	60
Leu Gly Ser Trp Asp	Asp Val Ala Glu Leu Ser Thr Thr His Asp Ile	
65	70	75
Met Val Asp Thr Ile	Val Asn His Met Ser Trp Glu Ser Glu Gln Phe	80
	85	90
Gln Asp Val Met Ala Lys Gly Glu Asp Ser Glu Tyr Tyr Pro Met Phe		95
	100	105
Leu Thr Met Ser Ser Ile Phe Pro Asp Gly Val Thr Glu Glu Asp Leu		110
	115	120
Thr Ala Ile Tyr Arg Pro Arg Pro Gly Leu Pro Phe Thr His Tyr Asn		125
	130	135
Trp Gly Gly Lys Thr Arg Leu Val Trp Thr Thr Phe Thr Pro Gln Gln		140
145	150	155
Val Asp Ile Asp Thr Asp Ser Glu Met Gly Trp Asn Tyr Leu Leu Ser		160
	165	170
Ile Leu Asp Gln Leu Ser Gln Ser His Val Ser Gln Ile Arg Leu Asp		175
	180	185
Ala Val Gly Tyr Gly Ala Lys Glu Lys Asn Ser Ser Cys Phe Met Thr		190
	195	200
Pro Lys Thr Phe Lys Leu Ile Glu Arg Ile Lys Ala Glu Gly Glu Lys		205
	210	215
Arg Gly Leu Glu Thr Leu Ile Glu Val His Ser Tyr Tyr Lys Lys Gln		220
225	230	235
Val Glu Ile Ala Ser Lys Val Asp Arg Val Tyr Asp Phe Ala Ile Pro		240
	245	250
Gly Leu Leu Leu His Ala Leu Glu Phe Gly Lys Thr Asp Ala Leu Ala		255
	260	265
Gln Trp Ile Asp Val Arg Pro Asn Asn Ala Val Asn Val Leu Asp Thr		270
	275	280
His Asp Gly Ile Gly Val Ile Asp Ile Gly Ser Asp Gln Met Asp Arg		285
	290	295
Ser Leu Ala Gly Leu Val Pro Asp Glu Glu Val Asp Ala Leu Val Glu		300
305	310	315
Ser Ile His Arg Asn Ser Lys Gly Glu Ser Gln Glu Ala Thr Gly Ala		320
	325	330
Ala Ala Ser Asn Leu Asp Leu Tyr Gln Val Asn Cys Thr Tyr Tyr Ala		335
	340	345
Ala Leu Gly Ser Asp Asp Gln Lys Tyr Ile Ala Ala Arg Ala Val Gln		350
	355	360
Phe Phe Met Pro Gly Val Pro Gln Val Tyr Tyr Val Gly Ala Leu Ala		365
	370	375
Gly Ser Asn Asp Met Asp Leu Leu Lys Arg Thr Asn Val Gly Arg Asp		380
385	390	395
Ile Asn Arg His Tyr Tyr Ser Ala Ala Glu Val Ala Ser Glu Val Glu		400
	405	410
Arg Pro Val Val Gln Ala Leu Asn Ala Leu Gly Arg Phe Arg Asn Thr		415
	420	425
Leu Ser Ala Phe Asp Gly Glu Phe Ser Tyr Ser Asn Ala Asp Gly Val		430
	435	440
Leu Thr Met Thr Trp Ala Asp Asp Ala Thr Arg Ala Thr Leu Thr Phe		445
	450	455
Ala Pro Lys Ala Asn Ser Asn Gly Ala Ser Val Ala Arg Leu Glu Trp		460
465	470	475
		480

Thr	Asp	Ala	Ala	Gly	Glu	His	Ala	Thr	Asp	Asp	Leu	Ile	Ala	Asn	Pro
				485					490					495	
Pro	Val	Val	Ala												
			500												

The Claims defining the invention are as follows:

1. A method for glycosylation of a substrate to produce a beta-glycosylated product, comprising the steps of: providing at least one glycosyl group donor wherein said glycosyl group donor is glucose-1-phosphate, a least one glycosyl group acceptor, and at least one glycosyltransferase enzyme; contacting the glycosyl group donor and glycosyl group acceptor with the glycosyltransferase enzyme under conditions such that the glycosyl group acceptor is glycosylated to produce at least one product having beta-glucose linkages; wherein the conditions are *in vitro* or within an engineered host cell.

2. The method of Claim 1, wherein said glycosyl group acceptor is selected from glycosyl, alkoxy, carboxy, aminocarbonyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carboxyalkyl, aminoalkyl, haloalkyl, alkylthioalkyl, heterocycloalkyl, heteroaryl, and heteroarylalkyl groups.

3. The method of Claim 1 or Claim 2, wherein said product having beta-glucose linkages is a steviol glycoside.

4. The method of any one of Claims 1 to 3, wherein said glycosyl group acceptor is stevioside, said glycosyl group donor is alpha-glucose-1-phosphate, and said product having beta-glucose linkages is rebaudioside A.

5. The method of any one of Claims 1 to 4, wherein said glycosyltransferase is selected from the polypeptides set forth in SEQ ID NOS:2, 4, 6, 8, 10, 12, and 14.

6. A method for production of glucose-1-phosphate, comprising the steps of: providing a phosphorylase, inorganic phosphate, and a disaccharide, trisaccharide, or oligosaccharide substrate of the phosphorylase; contacting said phosphorylase, inorganic phosphate, and saccharide under conditions such that said saccharide is cleaved to produce a monosaccharide and glucose-1-phosphate; wherein the method for production of glucose-1-phosphate is combined with the method of any one of Claims 1 to 5; wherein the conditions are *in vitro* or within an engineered host cell.

7. The method of Claim 6, wherein said glucosyl group acceptor is selected from glycosyl, alkoxy, carboxys, aminocarbonyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carboxyalkyl, aminoalkyl, haloalkyl, alkylthioalkyl, heterocycloalkyl, heteroaryl, and heteroarylalkyl groups.

8. The method of Claim 6 or Claim 7, wherein said phosphorylase is sucrose phosphorylase, said saccharide is sucrose, said monosaccharide produced is sucrose, and said glucose-1-phosphate produced is α -glucose-1-phosphate.

9. The method of any one of Claims 6 to 8, wherein said phosphorylase comprises a polypeptide sequence selected from SEQ ID NOS:16, 18, 20, 22, 24,26, 28, 30, and 32.

ABSTRACT

The present invention provides glycosyl transferase (GT) enzymes, polypeptides having GT activity, and polynucleotides encoding these enzymes, as well as vectors and host cells comprising these polynucleotides and polypeptides. The present invention also provides methods of using these GT enzymes to generate products with β -glucose linkages.

FIGURE 1

1/2

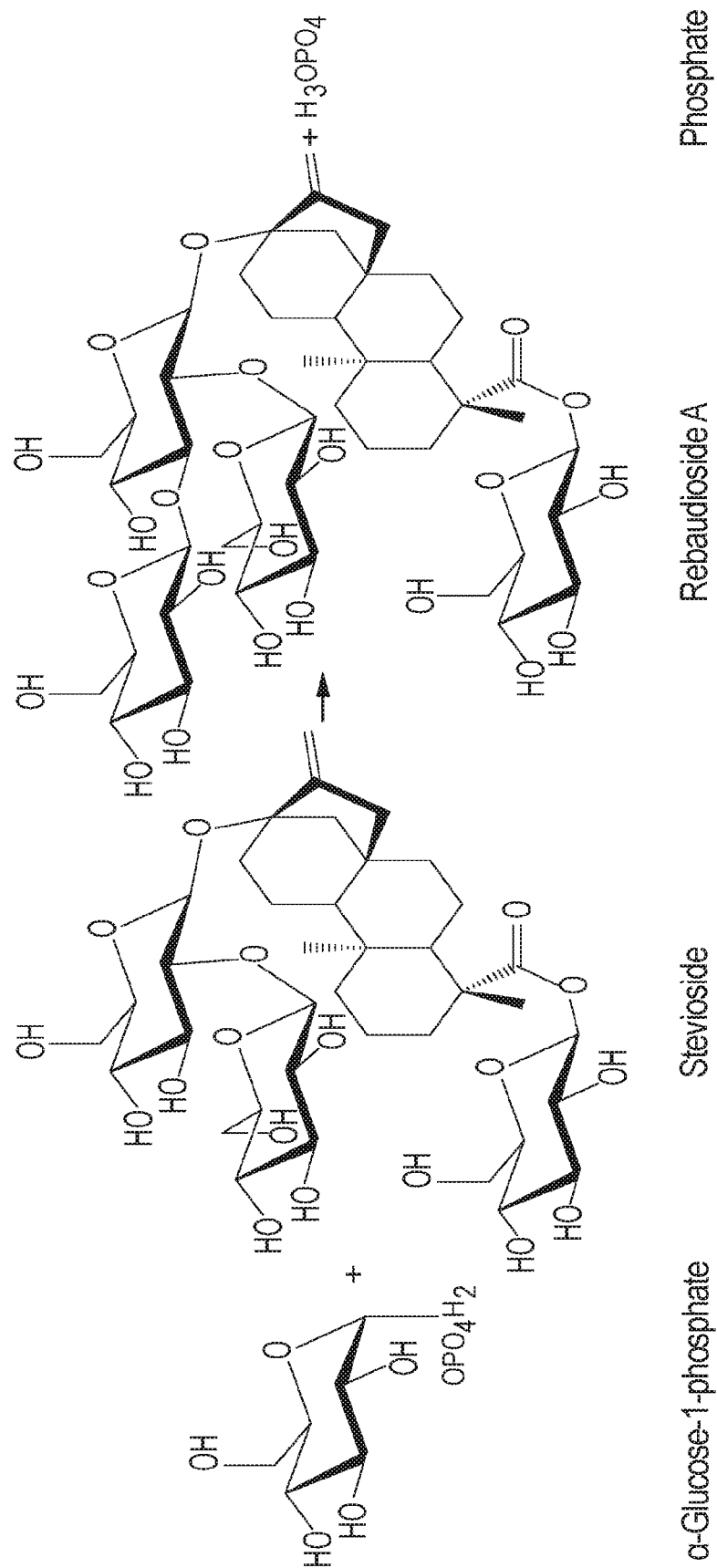


FIG. 1

2/2

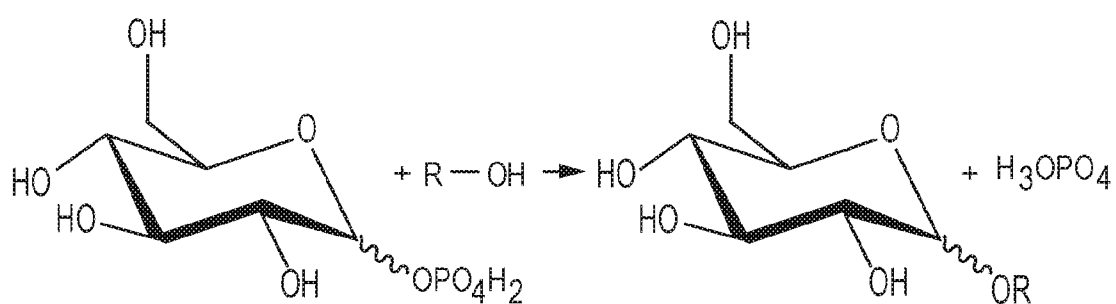


FIG. 2

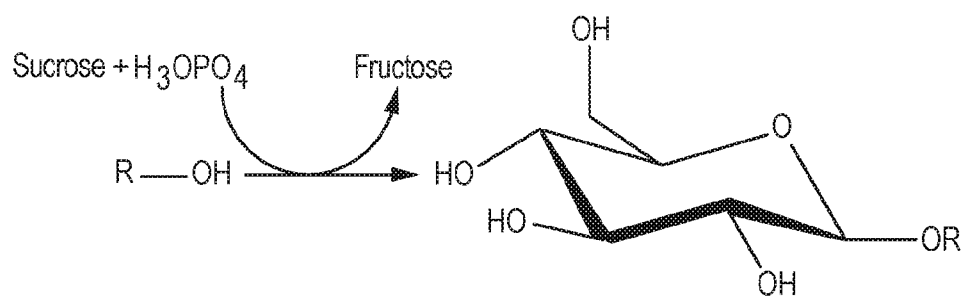


FIG. 3