

PRODUCTION OF STEVIOL GLYCOSIDES THROUGH WHOLE CELL
BIOTRANSFORMATION

PRIORITY

This application claims priority to, and the benefit of, US Provisional Application
5 No. 62/698,617 filed July 16, 2018, which is hereby incorporated by reference in its entirety.

BACKGROUND

Glycosyltransferases of small molecules are encoded by a large multigene family in the plant kingdom. These enzymes transfer sugars from nucleotide sugars to a wide range of secondary metabolites, thereby altering the physical and chemical properties of the acceptor
10 molecule. For example, steviol glycosides are a class of compounds found in the leaves of *Stevia rebaudiana* Bertoni, a perennial shrub of the *Asteraceae* (*Compositae*) family native to certain regions of South America. They are characterized structurally by a single base, steviol, differing by the presence of carbohydrate residues at positions C13 and C19. They accumulate in stevia leaves, composing approximately 10% to 20% of the total dry weight.
15 On a dry weight basis, the four major glycosides found in the leaves of *Stevia* typically include stevioside (9.1%), rebaudioside A (3.8%), rebaudioside C (0.6-1.0%) and dulcoside A (0.3%). Other steviol glycosides are present at small or trace amounts, including rebaudioside B, D, E, F, G, H, I, J, K, L, M and O, dulcoside B, steviolbioside and rubusoside.

20 The minor glycosylation product rebaudioside M (RebM) is estimated to be about 200-350 times more potent than sucrose, and is described as possessing a clean, sweet taste with a slightly bitter or licorice aftertaste. Prakash I. et al., Development of Next Generation Stevia Sweetener: Rebaudioside M, *Foods* 3(1), 162-175 (2014). While RebM is of great interest to the global food industry, its low prevalence in stevia extract necessitates
25 innovative processes for its synthesis.

There remains a need for economical methods for producing high value glycosides that are minor products of natural plant extract.

SUBSTITUTE SHEET (RULE 26)

SUMMARY OF THE INVENTION

In various aspects and embodiments, the invention provides microbial cells and methods for producing advanced glycosylation products from lower glycosylated intermediates. The microbial cell expresses one or more UDP-dependent glycosyltransferase enzymes for glycosylation of the intermediates. When incubating the microbial strain with a plant extract or fraction thereof comprising the glycoside intermediates, these intermediates are available for further glycosylation by the microbial cells. In various embodiments, the advanced glycosylation products are recovered from the media and/or microbial cells.

The glycoside intermediates can be any glycosylated secondary metabolite, such as those naturally found in plant extracts, including glycosylated terpenoids, flavonoids, cannabinoids, polyketides, stilbenoids, and polyphenols, among others. In some embodiments, the glycosylated intermediate is a glycosylated terpenoid, such as steviol glycoside or mogroside, and may find use as a sweetener. In some embodiments, biosynthesis of the product involves at least two glycosylation reactions of the glycoside intermediate in the microbial cell.

In some embodiments, the glycoside intermediates are from stevia leaf extract. For example, RebM and other advanced glycosylation products may be biosynthesized from steviol glycoside intermediates such as stevioside, steviolbioside, rebaudioside A (RebA), and rebaudioside C (RebC), among others. The microbial cell expresses one or more UDP-dependent glycosyl transferase enzymes in the cytoplasm, for glycosylation of the lower value intermediates. When incubating the microbial strain with a stevia leaf extract or fraction thereof comprising the steviol glycoside intermediates, these intermediates are available to the cells for further glycosylation, and these products can be recovered from the media and/or microbial cells. Accordingly, this process uses advanced intermediates in the stevia leaf extract, namely steviol glycosides having from one to five glycosylations (as well as in some embodiments the aglycone core, steviol). Advanced intermediates from stevia leaf extract are readily available from existing industrial extraction of steviol glycosides.

In various embodiments, the microbial cell expresses at least one, or at least two, or at least three, or at least four UGT enzymes that glycosylate a glycoside intermediate, and may be selected from those listed in Table 1 and those provided herein as SEQ ID NOS: 1-45, as well as derivatives thereof. In various embodiments, the UGT enzymes are independently selected from 1-2' glycosylating UGT enzymes, 1-3' glycosylating UGT enzymes, and O-glycosylating UGT enzymes. In various embodiments, these enzymes are expressed intracellularly, in that they do not contain membrane translocation or secretion signals.

In some embodiments, particularly with respect to glycosylation of steviol glycoside intermediates in stevia leaf extract, the microbial cell expresses four UGT enzymes, such as a 13-O UGT glycosylating enzyme, a 19-O UGT glycosylating enzyme, a 1-2' UGT glycosylating enzyme, and a 1-3' UGT glycosylating enzyme.

In various embodiments, the microbial cell expresses a 1-3' glycosylating UGT enzyme. For example, the 1-3' glycosylating UGT enzyme may be selected from SrUGT76G1, MbUGT1-3, and derivatives thereof (e.g., UGT76G1_L200A or MbUGT1-3_1, MbUGT1-3_1.5, or MbUGT1-3_2).

In these and other embodiments, the microbial cell expresses a 1-2' glycosylating UGT enzyme. For example, the 1-2' glycosylating UGT enzyme may be selected from SrUGT91D2, SrUGT91D1, SrUGT91D2e, OsUGT1-2, MbUGT1,2, MbUGT1,2.2, and derivatives thereof.

In these or other embodiments, the microbial cell expresses a C13 O-glycosylating UGT enzyme. For example, the C13 O-glycosylating UGT enzyme may be selected from SrUGT85C2 and derivatives thereof (e.g., MbUGTC13).

In these or other embodiments, the microbial cell expresses a C19 O-glycosylating UGT enzyme. For example, the C19 O-glycosylating enzyme may be selected from SrUGT74G1, MbUGTc19, and derivatives thereof (e.g., MbUGTc19-2).

Whole cell conversion requires that enzymes which are expressed intracellularly act on externally fed substrate (e.g., glycosylated intermediates) and that the cell provides UDP-

glucose cofactor regeneration. This is in contrast to processes that rely on enzymes that are purified or secreted outside the cell, which requires an exogenous UDP-glucose supply or UDP-glucose precursor or UDP-glucose regeneration mechanism or UDP-glucose regeneration enzyme system. In embodiments of the present invention, catalysis (glycosylation) is carried out with live microbial cells. UDP-glucose cofactor recycling takes place using the native cellular metabolism without requiring additional externally provided enzymes or substrate feeding.

In accordance with the present disclosure, genetic modifications to the microbial cell allow for glycoside intermediates to be available for further glycosylation by intracellularly expressed enzymes. In some embodiments, advanced glycosylated products can be recovered from the medium, as opposed to extracted from lysed cells. In some embodiments, the microbial cell has one or more genetic modifications that increase UDP-glucose availability. In some embodiments, without wishing to be bound by theory, these modifications may also stress the cell for glucose availability, leading to the increased expression of endogenous transporters to import the glycoside intermediates into the cell. In some embodiments, without wishing to be bound by theory, the cells are rendered permeable through genetic modification or media components, allowing passive diffusion of products and substrates.

In various embodiments, the microbial cell has an overexpression of one or more endogenous transporters (e.g., as compared to a parent microbial strain), or in certain embodiments, is modified to express a recombinant and/or engineered transport protein. In some embodiments, the microbial cell expresses one or more additional copies of an endogenous transport protein, or derivative thereof. For example, expression or activity of transport proteins can be modified to increase transport into the cell of steviol glycoside substrates (e.g., one or more of stevioside, steviolbioside, RebA, and/or RebC), while exporting product, such as RebM and/or RebD as well as other advanced glycosylation products. Exemplary transport proteins that can be overexpressed or engineered for altered activity or substrate specificity in the microbial cell include *E. coli* acrAD, xylE, ascF, bglF, chbA, ptsG/crr, wzxE, rfbX, as well as orthologs or derivatives thereof. Other transport proteins include those selected from bacterial or endogenous transport proteins that transport

the desired glycoside intermediate. For example, the transporter may be from the host species, or another bacterial or yeast species, and may be engineered to adjust its affinity for the particular glycoside intermediates or products.

In various embodiments, the method results in at least 40% conversion, or at least 50% conversion, or at least 75% conversion of the glycoside intermediates to desired product (e.g., conversion of stevioside, steviolbioside, and RebA to RebD, RebM, and/or other highly glycosylated rebaudiosides). In the production of RebM, the product profile can strongly favor RebM over RebD.

The method may be performed by batch fermentation, fed-batch fermentation, continuous fermentation, or semi-continuous fermentation. For example, in some embodiments, the method is conducted by batch fermentation with incubation times of less than about 72 hours, or in some embodiments, less than about 48 hours, or less than about 24 hours.

In some aspects, the invention provides methods for making a product comprising an advanced steviol glycoside, such as RebM. The method comprises incorporating the target steviol glycoside into a product, such as a food, beverage, oral care product, sweetener, flavoring agent, or other product. In some aspects, the invention provides methods for making a sweetener composition comprising a plurality of high-intensity sweeteners, such as two or more of a steviol glycoside, a mogroside, sucralose, aspartame, neotame, advantame, acesulfame potassium, saccharin, cyclamate, neohesperidin dihydrochalcone, gnetifolin E, and/or piceatannol 4'-O- β -D-glucopyranoside.

Other aspects and embodiments of the invention will be apparent from the following detailed description.

DESCRIPTION OF THE FIGURES

FIG. 1. Structure of RebM. The steviol scaffold (a diterpenoid) is shown boxed. RebM contains six glycosylations: (1) a C13 O-glycosylation, (2) a C13 1-2' glycosylation, (3) a C13 1-3' glycosylation, (4) a C19 O-glycosylation, (5) a C19 1-2' glycosylation, and (6) a C19 1-3' glycosylation.

FIG. 2. Glycosylation products produced by action of UGT enzymes on steviol and steviol glycoside intermediates.

FIG. 3. Glycosylated products produced by action of UGT enzymes on steviol glycoside intermediates containing rhamnose.

5 FIG. 4. Conversion of steviolbioside to RebM. For five major compounds in the RebM pathway, compound concentrations are shown relative to the total concentration of these compounds in a control strain that lacks the UGT enzymes. Strains were fed 0.2 mM steviolbioside and incubated with substrate for 48 hours before extraction. Strains 3 and 4 show no conversion of steviolbioside, while Strain 5 shows substantial conversion to RebD and RebM.
10

FIG. 5. Conversion of stevioside and RebA to RebM. Bioconversion of two commercially available intermediates found in leaf extract, stevioside and RebA, to RebM, is shown using Strains 1 to 5. Also shown is the bioconversion of the steviol core. The values are reported as the percentage of the total steviol glycoside composition. Conversion of
15 steviol to RebM by all strains shows that all strains are capable of RebM formation. As was the case for steviolbioside, both stevioside and RebA are converted to RebM, but only in Strain 5. For conversion from stevioside and RebA, the ratio of RebD:RebM strongly favors RebM (1:20).

FIG. 6. Bioconversion of steviol glycosides in 96-well plates. For five major
20 compounds in the RebM pathway, compound concentrations are shown with or without UGT enzymes expressed. Strains were fed 0.5 g/L of stevioside/RebA leaf extract and incubated with substrate for 48 hours before sampling.

FIG. 7. Bioconversion of a stevioside/RebA leaf extract at bioreactor scale. For five
25 major compounds in the RebM pathway, compound concentrations in the media were sampled at various time points across the fermentation. Strains were fed 2 g/L of the stevioside/RebA leaf extract and incubated with substrate for a total of 30 hours.

FIG. 8. Genetic modifications to increase the native *E. coli* flux to UDP-glucose. Metabolites: G6P, glucose-6-phosphate; G1P, glucose-1-phosphate; 6PGL, 6-

phosphoglucono-1,5-lactone; 6PG, 6-phosphogluconate; R5P, ribose-5-phosphate; PRPP, 5-phosphoribosyl-1-pyrophosphate; UMP, uridine monophosphate; UDP, uridine diphosphate; UTP, uridine triphosphate; UDP-Glu, UDP-glucose; UDP-Gal, UDP-galactose; UDP-GlcNAc, UDP-N-acetylglucosamine; CTP, cytidine triphosphate; dUDP, deoxyuridine-diphosphate; RebA, rebaudioside A; RebM, rebaudioside M. Enzymes: 1. *glk*, glucokinase; 2. *pgi*, glucose-6-phosphate isomerase; 3. *zwf*, glucose-6-phosphate-1-dehydrogenase; 4. *pgl*, 6-phosphogluconolactonase; 5. *edd*, phosphogluconate dehydratase; 6. *gnd*, 6-phosphogluconate dehydrogenase; 7. *rpiA*, ribose-5-phosphate isomerase or *rpe*, ribulose-phosphate 3-epimerase; 8. *prs*, ribose-phosphate diphosphokinase; 9. *pyrEF*, orotate phosphoribosyltransferase and orotidine-5'phosphate-decarboxylase; 10. *pyrK*, uridylylate kinase; 11. *ndk*, UDP kinase; 12. *murG*, N-acetylglucosaminyl transferase; 13. *nrdABEF*, ribonucleoside-diphosphate reductase; 14. *pyrG*, CTP synthase; 15. *pgm*, phosphoglucomutase; 16. *galU*, UDP-glucose pyrophosphorylase; 17. *galETKM*, UDP-glucose-4-epimerase / galactose-1-phosphate uridylyltransferase / galactokinase / galactose-1-epimerase; 18. *ushA*, 5'-nucleotidase / UDP-sugar hydrolase; 19. *UGT*, UDP-glycosyltransferase. The schematic illustrates the sequential genomic modifications that lead from Strain 1 to Strain 5. Strain 1 is a base *E. coli* strain. Strain 2 has a deletion of 17, and Strain 3 has an addition deletion of 18. Strain 4 was built from Strain 3 by deleting 2. Strain 5 was built from Strain 4 by overexpressing 15 and 16.

20 DETAILED DESCRIPTION OF THE INVENTION

In various aspects and embodiments, the invention provides microbial cells and methods for producing advanced glycosylation products from lower glycosylated intermediates. In various embodiments, the microbial cell expresses one or more UDP-dependent glycosyl transferase enzymes (e.g., intracellularly), for glycosylation of the intermediates. In various embodiments, the microbial cell has one or more genetic modifications that increase the availability of UDP-glucose. When incubating the microbial strain with the glycoside intermediates (e.g., from a plant extract or fraction thereof), these glycoside intermediates are available to the cell for further glycosylation by intracellularly-expressed glycosyltransferase enzymes. In some embodiments, the advanced glycosylation

products can be recovered from the media, or in some embodiments, are recovered from the media and microbial cells.

The glycoside intermediates can be any glycosylated secondary metabolite, such as those naturally found in plant extracts, including glycosylated terpenoids, glycosylated flavonoids, glycosylated cannabinoids, glycosylated polyketides, glycosylated stilbenoids, and glycosylated polyphenols, among others. In some embodiments, the glycoside intermediate is a glycosylated terpenoid, such as steviol glycoside or mogroside. In some embodiments, the glycoside intermediate has one, two, three, four, or five glycosyl groups, including glycosyl groups selected from glucosyl, galactosyl, mannosyl, xylosyl, and rhamnosyl groups, among others. In various embodiments, the glycosylated product has at least four glycosyl groups, or at least five glycosyl groups, or at least six glycosyl groups. In other embodiments, the product has seven, eight, nine, or more glycosyl groups. In some embodiments, biosynthesis of the product involves at least two glycosylations of the intermediate by the microbial cell.

In some embodiments, the glycoside intermediates are from stevia leaf extract. For example, steviol glycosides having five, six, or more glycosylations (such as RebD or RebM) may be biosynthesized from steviol glycoside intermediates such as stevioside, steviolbioside, rebaudioside A, dulcoside A, dulcoside B, rebaudioside C, and rebaudioside F, among others. The microbial cell expresses one or more UDP-dependent glycosyl transferase enzymes for glycosylation of these lower value precursors. When incubating the microbial strain with a stevia leaf extract or fraction thereof comprising the steviol glycoside intermediates, these intermediates are available for further glycosylation to RebD or RebM or other advanced glycosylation product (e.g., RebI), which can be recovered from the media and/or microbial cells.

In various embodiments, the UDP-dependent glycosyl transferase enzymes are expressed “intracellularly”, in that the enzymes do not possess membrane translocation or secretion peptides or domains. Thus, expression of the UGT enzymes takes place in the cytoplasm, and these enzymes are not directed outside the cell via a secretion or transport signal. In various embodiments, the UGT enzymes do not contain membrane anchoring

domains. That is, in various embodiments the UGT enzymes do not comprise a transmembrane domain.

In some embodiments, the process uses advanced intermediates in the stevia leaf extract, namely steviol (the aglycone intermediate) and steviol glycosides having from one to five glycosylations, which are available to the microbial cell for further glycosylation. Advanced intermediates from stevia leaf extract are readily available from existing industrial extraction of steviol glycosides. As shown in Table 2, leaf extract may contain primarily the pathway intermediates stevioside and rebaudioside A (RebA). In various embodiments, the stevia leaf extract is an extraction of steviol glycosides. In some embodiments, the extract comprises one or more of stevioside, steviolbioside, and rebaudioside A, as prominent components. In some embodiments, the extract comprises one or more of dulcoside A, dulcoside B, RebC and/or RebF as prominent components. A prominent component generally makes up at least about 10% of the steviol glycosides in the extract or fraction thereof, but in some embodiments, may make up at least about 20%, or at least about 25%, or at least about 30% of the steviol glycosides in the extract or fraction thereof.

RebM is illustrated in FIG. 1. RebM contains six glycosylations of a steviol core: (1) a C13 O-glycosylation, (2) a C13 1-2' glycosylation, (3) a C13 1-3' glycosylation, (4) a C19 O-glycosylation, (5) a C19 1-2' glycosylation, and (6) a C19 1-3' glycosylation. While various glycosylation products are possible (FIG. 2), RebM can be synthesized from steviol through the action of four UGT enzymes, and UGT glycosylation enzymes are each capable of acting on a number of substrates. For example, both UGT91D2 and OsUGT1-2 are 1-2' glycosylating enzymes that can produce steviolbioside from steviolmonoside (by action at C13), as well as RebD from RebA (by action at C19). Further, UGT76G1 is a 1-3' glycosylating enzyme that can produce RebA from stevioside (by action at C13), as well as RebM from RebD (by action at C19).

UGT enzymes for glycosylation of steviol and steviol glycosides (including for biosynthesis of RebM) are disclosed in US 2017/0332673, which is hereby incorporated by reference in its entirety. Exemplary UGT enzymes are listed in Table 1, below (referenced patent applications are hereby incorporated by reference in their entirety):

Table 1: Example UGT Enzymes

Type of glycosylation	Enzyme	Gene ID	Protein ID	Description
C13	SrUGT85C2	AY345978.1	AAR06916.1	
	MbUGTC13			US 2017/0332673
C19	SrUGT74G1	AY345982.1	AAR06920.1	
	MbUGTc19	-	-	US 2017/0332673
	MbUGTc19-2			US 2017/0332673
1-2'	SrUGT91D1	AY345980.1	AAR06918.1	
	SrUGT91D2	ACE87855.1	ACE87855.1	
	SrUGT91D2e	-	-	US 2011/038967
	OsUGT1-2	NM_001057542.1	NP_001051007.2	WO 2013/022989
	MbUGT1,2	-	-	US 2017/0332673
	MbUGT1,2.2	-	-	US 2017/0332673
1-3'	SrUGT76G1	FB917645.1	CAX02464.1	
	MbUGT1-3			US 2017/0332673
	76G1 L200A			US 2017/0332673
	MbUGT1-3_1			US 62/866,148
	MbUGT1-3_1.5			This disclosure
	MbUGT1-3_2			US 62/866,148

Amino acid sequences for exemplary UGT enzymes are provided by this disclosure as SEQ ID NOS: 1-17. SEQ ID NO:1 is *Stevia rebaudiana* UGT85C2. SEQ ID NO:13 is SrUGT85C2 with a P215T substitution, and insertion of an Ala at the 2nd position to increase stability. SEQ ID NO:2 is *Stevia rebaudiana* UGT74G1 (with insertion of Ala at the 2nd position). SEQ ID NOS:8 and 12 are circular permutants based on SrUGT74G1 (MbUGTC19 and MbUGTC19-2). SEQ ID NO:3 is *Stevia rebaudiana* UGT76G1. Circular permutants with 1-3' glycosylating activity are disclosed as SEQ ID NO:10 (MbUGT1-3) and SEQ ID NOS:15, 16, and 17 (MbUGT1-3_1, MbUGT1-3_1.5, and MbUGT1-3_2, respectively). UGT76G1 with a L200A substitution, and Ala at position 2, is disclosed as SEQ ID NO:14. *Stevia rebaudiana* UGT91D1, UGT91D2, and UGT91D2e are disclosed as SEQ ID NOS:4, 5, and 6. *Oryza sativa* UGT1-2 is disclosed as SEQ ID NO:7. MbUGT1,2 and MbUGT1,2.2 are circular permutant enzymes with 1-2' glycosylating activity (SEQ ID NOS: 9 and 11).

Additional UGT enzymes are provided as SEQ ID NOS:18 to 46, from species *Siraitia grosvenorii* (monk fruit), *Momordica charantia* (bitter melon), *Cucumis sativa* (Cucumber), *Cucurbita maxima* (squash), *Cucurbita moschata* (squash, pumpkin), *Prunus*

persica (peach), *Theobroma cacao* (cacao), *Corchorus capsularis* (white jute), *Ziziphus jujube* (red date), *Vitis vinifera* (grape vine), *Juglans regia* (walnut), *Hevea brasiliensis* (rubber tree), *Manihot esculenta* (cassava), *Cephalotus follicularis* (pitcher plant), and *Coffea Arabica* (coffee). UGT enzymes can be selected and optionally engineered based on
5 the desired product and available intermediate.

In various embodiments, the microbial cell expresses at least one, or at least two, or at least three, or at least four UGT enzymes. In some embodiments, the UGT enzymes glycosylate a steviol glycoside substrate, including those listed in Table 2. In some
10 embodiments, the microbial cell expresses four UGT enzymes that glycosylate a steviol glycoside intermediate, such as a 13-O UGT glycosylating enzyme, a 19-O UGT glycosylating enzyme, a 1-2' UGT glycosylating enzyme, and a 1-3' UGT glycosylating enzyme. The action of these general classes of UGT enzymes on glycosylated intermediates of the stevia leaf are depicted in FIGS. 2 and 3.

In various embodiments, the microbial cell expresses a 1-3' glycosylating UGT
15 enzyme. For example, the 1-3' glycosylating UGT enzyme may be selected from SrUGT76G1, MbUGT1-3_1, MbUGT1-3_1.5, and MbUGT1-3_2, and derivatives thereof.

In these and other embodiments, the microbial cell expresses a 1-2' glycosylating UGT enzyme. For example, the 1-2' glycosylating UGT enzyme may be selected from SrUGT91D2, SrUGT91D1, SrUGT91D2e, OsUGT1-2, MbUGT1,2, MbUGT1,2.2, and
20 derivatives thereof.

In these or other embodiments, the microbial cell expresses a C13 O-glycosylating UGT enzyme. For example, the C13 O-glycosylating UGT enzyme may be selected from SrUGT85C2 and derivatives thereof (e.g., MbUGTC13).

In these or other embodiments, the microbial cell expresses a C19 O-glycosylating
25 UGT enzyme. For example, the C19 O-glycosylating enzyme may be selected from SrUGT74G1, MbUGTc19, and derivatives thereof (e.g., MbUGTc19-2).

In these or other embodiments, the microbial cell expresses a 1-3' glycosylating UGT enzyme and a 1-2' glycosylating UGT enzyme. In some embodiments, the microbial cell

expresses a 1-3' glycosylating UGT enzyme, a 1-2' glycosylating UGT enzyme, and a C13 O-glycosylating UGT enzyme. In some embodiments, the microbial cell expresses a 1-3' glycosylating UGT enzyme, a 1-2' glycosylating UGT enzyme, a C19 O-glycosylating UGT enzyme, and a C19 O-glycosylating UGT enzyme. In some embodiments, the microbial cell
5 expresses a 1-3' glycosylating UGT enzyme, a 1-2' glycosylating UGT enzyme, a C19 O-glycosylating UGT enzyme, and a C13 O-glycosylating UGT enzyme. In some embodiments, the microbial cell expresses a SrUGT85C2 or derivative thereof (e.g., MbUGTC13), MbUGT1,2.2 or derivative thereof, SrUGT74G1 or derivative thereof (e.g., MbUGTc19 or MbUGTc19-2), and SrUGT76G1 or MbUGT1-3_1 or derivative thereof
10 (e.g., 76G1_L200A, MbUGT1-3_1.5, or MbUGT1-3_2). Without being bound by theory, engineered UGT enzymes may provide for increased carbon flux to RebM (as well as higher glycosylation products), and particularly due to substrate binding pockets that are better able to accommodate larger substrates, without substantial loss of activity on lower glycosylated intermediates. In these embodiments, the UGT enzymes (such as the 1-3' and 1-2'
15 glycosylating UGT enzymes) may have an increased rate of activity (e.g., rate of substrate binding and turnover) with more highly glycosylated steviol substrates such as RebA or RebD.

Derivatives of UGT enzymes generally comprise an amino acid sequence having at least about 50% identity, at least about 60% identity, at least about 70% identity, at least
20 about 80% identity, at least about 85% identity, or at least about 90% identity, or at least about 95% identity, or at least 96%, 97%, 98% or 99% identity to one or more of SEQ ID NOS: 1 to 46. In some embodiments, the derivative has from 1 to 20 or from 1 to 10, or from 1 to 5 amino acid modifications (independently selected from amino acid substitutions, insertions, and deletions), with respect to one of SEQ ID NOS: 1 to 46. In some embodiments,
25 for example with regard to production of RebM and other advanced steviol glycosides (e.g., having at least 5 glycosyl groups) from steviol glycoside intermediates, derivatives of these UGT enzymes comprise an amino acid sequence having at least about 50% identity, at least about 60% identity, at least about 70% identity, at least about 80% identity, at least about 85% identity, or at least about 90% identity, or at least about 95% identity, or at least 96%,
30 97%, 98% or 99% identity to one or more of SEQ ID NOS: 1 to 17. In some embodiments, the derivative has from 1 to 20 or from 1 to 10, or from 1 to 5 amino acid modifications

(independently selected from amino acid substitutions, insertions, and deletions), with respect to one of SEQ ID NOS: 1 to 17.

In some embodiments, the microbial cell expresses a 1-3' UGT enzyme comprising an amino acid sequence that is at least about 75% identical to the amino acid sequence of SEQ ID NO:15, SEQ ID NO:16, or SEQ ID NO:17. In some embodiments, the 1-3' UGT comprises an amino acid sequence that is at least about 80% identical to SEQ ID NO:15, 16, or 17. In some embodiments, the amino acid sequence of the 1-3' UGT enzyme is at least about 85% identical to SEQ ID NO:15, 16, or 17, or at least about 90% identical to SEQ ID NO:15, 16, or 17 or at least about 95% identical to SEQ ID NO:15, 16, or 17, or at least about 98% identical to SEQ ID NO: 15, 16, or 17. In some embodiments, the amino acid sequence of the 1-3' UGT enzyme comprises the amino acid of SEQ ID NO:15, 16, or 17.

For example, the amino acid sequence may have from 1 to 20 amino acid modifications independently selected from substitutions, deletions, and insertions, with respect to the amino acid sequence SEQ ID NO:15, 16, or 17. In some embodiments, the amino acid sequence has from 1 to 10 amino acid modifications (e.g., from 1 to 5) independently selected from substitutions, deletions, and insertions, with respect to the amino acid sequence of SEQ ID NO:15, 16, or 17. Amino acid modifications to the amino acid sequence of SEQ ID NO:15, 16, or 17 can be guided by available enzyme structures and construction of homology models. Exemplary structures are described in, e.g., Li, et al., Crystal Structure of Medicago truncatula UGT85H2—insights into the Structural Basis of a Multifunctional (iso) Flavonoid Glycosyltransferase, *J. of Mol. Biol.* 370.5 (2007): 951-963. Publicly available crystal structures (e.g., PDB entry: 2PQ6) may be used to inform amino acid modifications. For example, one or more amino acid modifications can be made to the active site or in the vicinity of the active site to improve the binding of substrate, and/or to improve reaction geometries of these substrates with catalytic side chains.

In some embodiments, the 1-3' UGT enzyme comprises an amino acid substitution at positions corresponding to positions 29, 200, 357, and 414 of SEQ ID NO:3 (*Stevia rebaudiana* UGT76G1). Substitutions at these positions, which are included in the enzyme of SEQ ID NOS:15 and 17 (positions 183, 354, 54, and 111, respectfully, in SEQ ID NO:15)

can provide dramatic improvements in activity. In some embodiments, the identity of amino acids at positions corresponding to positions 183, 354, 54, and 111 of SEQ ID NO:15, allows for further modification at other positions. For example, in some embodiments, the 1-3' UGT enzyme comprises an amino acid sequence that is at least about 60% identical to the amino acid sequence of SEQ ID NO:15 or 17, wherein the UGT enzyme comprises: a glycine (G) or threonine (T) at the position corresponding to position 54 of SEQ ID NO:15; a leucine (L) or isoleucine (I) at the position corresponding to position 111 of SEQ ID NO:15; a methionine (M) or leucine (L) at the position corresponding to position 183 of SEQ ID NO:15; and an alanine (A), or glycine (G), or serine (S) at the position corresponding to position 354 of SEQ ID NO:15. In some embodiments, the 1-3' UGT enzyme comprises a methionine (M) at the position corresponding to position 183 of SEQ ID NO:15. In some embodiments, the 1-3' UGT enzyme comprises a glycine (G) at the position corresponding to position 54 of SEQ ID NO:15. In some embodiments, the 1-3' UGT enzyme comprises a leucine (L) at the position corresponding to position 111 of SEQ ID NO:15. In some embodiments, the 1-3' UGT has two or three of a methionine (M) at the position corresponding to position 183 of SEQ ID NO:15, a glycine (G) at the position corresponding to position 54 of SEQ ID NO:15, and a leucine (L) at the position corresponding to position 111 of SEQ ID NO:15. These modifications can provide substantial improvements to the activity of the enzyme.

In some embodiments, the 1-3' UGT enzyme comprises an insertion of from 5 to about 15 amino acids, such as from 6 to 12 amino acids, or about 6 or about 11 amino acids, after the position corresponding to position 155 of SEQ ID NO:15. In some embodiments, the insertion is a flexible and hydrophilic sequence that is predominately Glycine and Serine residues. In some embodiments, the sequence is GSGGSG (SEQ ID NO:47) or GSGGSGGSG (SEQ ID NO:48).

In various embodiments, the 1-3' UGT enzyme shows improved conversion of stevioside to Reb A, and improved conversion of RebD to RebM, as compared to UGT76G1-L200A (SEQ ID NO:14). This improved conversion is exhibited in a bioconversion assay where stevioside or RebD substrate is fed to microbial cells expressing the 1-3' UGT enzyme. Improved conversion can be demonstrated in reactions with cell lysates containing

recombinantly expressed 1-3' UGT, or in vitro reactions with purified or partially purified 1-3' UGT.

Changes to the amino acid sequence of an enzyme can alter its activity or have no measurable effect. Silent changes with no measurable effect are often conservative substitutions and small insertions or deletions on solvent-exposed surfaces that are located away from active sites and substrate-binding sites. In contrast, enzymatic activity is more likely to be affected by non-conservative substitutions, large insertions or deletions, and changes within active sites, substrate-binding sites, and at buried positions important for protein folding or conformation. Changes that alter enzymatic activity may increase or decrease the reaction rate or increase or decrease the affinity or specificity for a particular substrate. For example, changes that increase the size of a substrate-binding site may permit an enzyme to act on larger substrates and changes that position a catalytic amino acid side chain closer to a target site on a substrate may increase the enzymatic rate.

Knowledge of the three-dimensional structure of an enzyme and the location of relevant active sites, substrate-binding sites, and other interaction sites can facilitate the rational design of derivatives and provide mechanistic insight into the phenotype of specific changes. Plant UGTs share a highly conserved secondary and tertiary structure while having relatively low amino acid sequence identity. Osmani et al, Substrate specificity of plant UDP-dependent glycosyltransferases predicted from crystal structures and homology modeling, *Phytochemistry* 70 (2009) 325–347. The sugar acceptor and sugar donor substrates of UGTs are accommodated in a cleft formed between the N- and C-terminal domains. Several regions of the primary sequence contribute to the formation of the substrate binding pocket including structurally conserved domains as well as loop regions differing both with respect to their amino acid sequence and sequence length.

Construction of UGT derivatives can be guided based on homology modeling compared to known structures. For example, based on crystal structure analysis of SrUGT76G1_L200A and an amino acid sequence alignment of SrUGT76G1 to MbUGT3-1_1, it is predicted that the steviol core of stevioside is close (within 4 Å) to the following residues of MbUGT3-1_1: I244, L280, W351, A354, I357, M362, and T438. Further, the

C19 1-2 glycosylation is predicted to be close (within 4 Å) to T438. The steviol core of RebD is predicted to be close (within 4 Å) of the following hydrophobic side chains of MbUGT3-1_1: L239, M242, I244, L280, I353, A354, and I357. The C13 1-2' glycosylation is predicted to be close (within 4 Å) of the following hydrogen bonding side chains of MbUGT3-1_1: S301 and D77. Positioning and amino acid content of the V341-Q352 and K355-A367 helices of MbUGT3-1_1 may be important for catalysis as the mutation corresponding to L200A is in a loop between these helices. Positions L76 and/or D77 of MbUGT3-1_1 may interact with the C13 glycosylation of stevioside.

The amino acid sequence of one or more of the UGT enzymes can optionally include an alanine inserted or substituted at position 2 to decrease turnover in the cell. In various embodiments, one or more UGT enzymes comprise an alanine amino acid residue inserted or substituted at position 2 to provide additional stability *in vivo*.

Identity of amino acid sequences, i.e. the percentage of sequence identity, can be determined via sequence alignments. Such alignments can be carried out with several known algorithms, such as that described by Karlin and Altschul (Karlin & Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90: 5873-5877), with hmalign (HMMER package) or with the CLUSTAL algorithm (Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res.* 22, 4673-80). The grade of sequence identity (sequence matching) may be calculated using e.g. BLAST, BLAT or BlastZ (or BlastX). A similar algorithm is incorporated into the BLASTN and BLASTP programs of Altschul et al (1990) *J. Mol. Biol.* 215: 403-410. BLAST protein alignments may be performed with the BLASTP program, score = 50, word length = 3. To obtain gapped alignments for comparative purposes, Gapped BLAST is utilized as described in Altschul et al (1997) *Nucleic Acids Res.* 25: 3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs are used.

The UGT enzymes may be integrated into the chromosome of the microbial cell, or alternatively, are expressed extrachromosomally. For example, the UGT enzymes may be expressed from a bacterial artificial chromosome (BAC) or yeast artificial chromosome (YAC).

Expression of the UGT enzymes can be tuned for optimal activity, using, for example, gene modules (e.g., operons) or independent expression of the UGT enzymes. For example, expression of the genes or operons can be regulated through selection of promoters, such as inducible or constitutive promoters, with different strengths (e.g., strong, intermediate, or weak). Several non-limiting examples of promoters of different strengths include Trc, T5 and T7. Additionally, expression of genes or operons can be regulated through manipulation of the copy number of the gene or operon in the cell. In some embodiments, the cell expresses a single copy of each UGT enzyme. In some embodiments, expression of genes or operons can be regulated through manipulating the order of the genes within a module, where the genes transcribed first are generally expressed at a higher level. In some embodiments, expression of genes or operons is regulated through integration of one or more genes or operons into the chromosome.

Optimization of UGT expression can also be achieved through selection of appropriate promoters and ribosomal binding sites. In some embodiments, this may include the selection of high-copy number plasmids, or single-, low- or medium-copy number plasmids. The step of transcription termination can also be targeted for regulation of gene expression, through the introduction or elimination of structures such as stem-loops.

Whole cell conversion requires that substrate (e.g., glycoside intermediates) is available to the cell for glycosylation by the expressed enzymes (e.g., the intracellularly expressed enzymes) and preferably product can be extracted from the extracellular media. Whole cell systems have advantages, since the cell provides UDP-glucose cofactor regeneration. This is in contrast to processes that use enzymes from cell lysis or secretion outside the cell, which requires an exogenous UDP-glucose supply or UDP-glucose precursor or UDP-glucose regeneration mechanism or UDP-glucose regeneration enzyme system. In embodiments of the present invention, catalysis (glycosylation) is carried out with live microbial cells, and UDP-glucose cofactor recycling takes place using the cellular metabolism without requiring enzyme feeding or substrate feeding for UDP-glucose regeneration.

US 2017/0332673 describes *E. coli* strains that overexpress MEP pathway enzymes, along with a downstream steviol biosynthesis pathway, and UGT enzymes to drive production of RebM from glucose. However, these strains do not perform biocatalysis of fed steviol glycoside intermediates to RebM, which may be, in part, due to the inability of the host cell to gain access to the steviol glycoside substrate. In accordance with the present disclosure, genetic modifications to the microbial cell allow for glycosylated intermediates to be available for further glycosylation in a whole cell system. Further, product can be recovered from the extracellular media, which facilitates downstream purification and processing of the product.

In some embodiments, the microbial cell has one or more genetic modifications that increase UDP-glucose availability. In some embodiments, without wishing to be bound by theory, these modifications may also stress the cell for glucose availability, leading to the increased expression of endogenous transporters to import steviol glycosides into the cell. Wild-type UDP-glucose levels in exponentially growing *E. coli* is about 2.5 mM (Bennett BD, Kimball EH, Gao M, Osterhout R, Van dien SJ, Rabinowitz JD. Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nat Chem Biol*. 2009;5(8):593-9.). In some embodiments, genetic modifications to the host cell are engineered to increase UDP-glucose, e.g., to at least 5 mM, or at least 10 mM, in exponentially growing cells (e.g., that do not have recombinant expression of UGT enzymes).

In some embodiments, the microbial cell has a deletion, inactivation, or reduced activity or expression of a gene encoding an enzyme that consumes UDP-glucose. For example, the microbial cell may have a deletion, inactivation, or reduced activity of *ushA* (UDP-sugar hydrolase) and/or one or more of *galE*, *galT*, *galK*, and *galM* (which are responsible for UDP-galactose biosynthesis from UDP-glucose), or ortholog thereof in the microbial species. In some embodiments, *galETKM* genes are inactivated, deleted, or substantially reduced in expression. Alternatively or in addition, the microbial cell has a deletion, inactivation, or reduced activity or expression of *E. coli* *otsA* (trehalose-6-phosphate synthase), or ortholog thereof in the microbial species. Alternatively or in addition, the microbial cell has a deletion, inactivation, or reduced activity or expression of

E. coli *ugd* (UDP-glucose 6-dehydrogenase), or ortholog thereof in the microbial species. Reducing or eliminating activity of *otsA* and *ugd* can remove or reduce UDP-glucose sinks to trehalose or UDP-glucuronidate, respectively.

Other UDP-glucose sinks that can be reduced or eliminated include eliminating or
 5 reducing activity or expression of genes responsible for lipid glycosylation and LPS biosynthesis, and genes responsible for glycosylating undecaprenyl-diphosphate (UPP). Genes involved in glycosylating lipids or LPS biosynthesis include *E. coli* *waaG* (lipopolysaccharide glucosyltransferase 1), *E. coli* *waaO* (UDP-D-glucose:(glucosyl)LPS α -1,3-glucosyltransferase)), and *E. coli* *waaJ* (UDP-glucose:(glucosyl)LPS α -1,2-
 10 glucosyltransferase)). Genes responsible for glycosylating undecaprenyl-diphosphate (UPP) include *E. coli* *yfdG* (putative bactoprenol-linked glucose translocase), *E. coli* *yfdH* (bactoprenol glucosyl transferase), *E. coli* *yfdI* (serotype specific glucosyl transferase), and *E. coli* *wcaJ* (undecaprenyl-phosphate glucose phosphotransferase). Deletion, inactivation, or reduction in activity or expression of one or more of these gene products (or corresponding
 15 orthologs in the microbial cell) can increase UDP-glucose availability.

In these or other embodiments, the microbial cell has a deletion, inactivation, or reduced activity or expression of a gene encoding an enzyme that consumes a precursor to UDP-glucose. For example, in some embodiments, the microbial cell has a deletion, inactivation, or reduced activity or expression of *pgi* (glucose-6 phosphate isomerase), or
 20 ortholog thereof in the microbial species of the host cell.

In these or other embodiments, the cell has an overexpression or increased activity of one or more genes encoding an enzyme involved in converting glucose-6-phosphate to UDP-glucose. For example, *pgm* (phosphoglucomutase) and/or *galU* (UTP-glucose-1-phosphate uridylyltransferase) (or ortholog or derivative thereof) can be overexpressed, or
 25 modified to increase enzyme productivity. Alternatively or in addition, *E. coli* *ycjU* (β -phosphoglucomutase), which converts glucose-6-phosphate to glucose-1-phosphate, and *Bifidobacterium bifidum* *ugpA*, which converts glucose-1-phosphate to UDP, or ortholog or derivative of these enzymes, can be overexpressed, or modified to increase enzyme productivity.

Alternatively or in addition, the microbial cell has one or more genetic modifications that increase flux to the pentose phosphate pathway (PPP), such as an overexpression or increased activity of *E. coli* zwf (or homologue or engineered derivative thereof), which is an NADP⁺-dependent glucose-6-phosphate dehydrogenase.

5 Alternatively or in addition, the microbial cell has one or more genetic modifications that increase glucose transport. Such modifications include increased expression or activity of *E. coli* galP (galactose:H⁺-symporter) and *E. coli* glk (glucokinase), or alternatively *Zymomonas mobilis* glf and *E. coli* glk, or homologues, orthologs, or engineered derivatives of these genes.

10 Alternatively or in addition, the microbial cell has one or more genetic modifications that increase UTP production and recycling. Such modifications include increased expression or activity of *E. coli* pyrH (UMP kinase), *E. coli* cmk (cytidylate kinase), *E. coli* adk (adenylate kinase), or *E. coli* ndk (nucleoside diphosphate kinase), or homologs, orthologs, or engineered derivatives of these enzymes.

15 Alternatively or in addition, the microbial cell has one or more genetic modifications that increase UDP production. Such modifications include overexpression or increased activity of one or more of *E. coli* upp (uracil phosphoribosyltransferase), *E. coli* dctA (C4 dicarboxylate/orotate:H⁺-symporter), *E. coli* pyrE (orotate phosphoribosyltransferase), and *E. coli* pyrF (orotidine-5'-phosphate decarboxylase), including homologs, orthologs, or
20 engineered derivatives thereof. For example, in some embodiments, the microbial cell overexpresses or has increased activity of upp, pyrH and cmk, or homolog or engineered derivative thereof. Alternatively, the microbial cell overexpresses or has increased activity of dctA, pyre, pyrH and cmk, or homolog or engineered derivative thereof.

25 Alternatively or in addition, the microbial cell may have one or more genetic modifications to remove or reduce regulation of glucose uptake. For example, the microbial cell may have a deletion, inactivation, or reduced expression of sgrS, which is a small regulatory RNA in *E. coli*.

Alternatively or in addition, the microbial cell may have one or more genetic modifications that reduce dephosphorylation of glucose-1-phosphate. Exemplary modifications include deletion, inactivation, or reduced expression or activity of one or more of *E. coli* agp (glucose-1-phosphatase), *E. coli* yihX (α -D-glucose-1-phosphate phosphatase), *E. coli* ybiV (sugar phosphatase), *E. coli* yidA (sugar phosphatase), *E. coli* yigL (phosphosugar phosphatase), and *E. coli* phoA (alkaline phosphatase), or an ortholog thereof in the microbial cell.

Alternatively or in addition, the microbial cell may have one or more genetic modifications that reduce conversion of glucose-1-phosphate to TDP-glucose. Exemplary modifications include deletion, inactivation, or reduced expression or activity of one or more of *E. coli* rffH (dTDP-glucose pyrophosphorylase) and *E. coli* rfbA (dTDP glucose pyrophosphorylase), or an ortholog thereof in the microbial cell.

Alternatively or in addition, the microbial cell may have one or more genetic modifications that reduce conversion of glucose-1-phosphate to ADP-glucose. Exemplary modifications include deletion, inactivation, or reduced expression or activity of *E. coli* gigC (glucose-1-phosphate adenylyltransferase), or an ortholog thereof in the microbial cell.

In some embodiments, the microbial cell is a bacterial cell comprising the genetic modifications: *ushA* and *galETKM* are deleted, inactivated, or reduced in expression; *pgi* is deleted, inactivated, or reduced in expression; and *pgm* and *galU* are overexpressed or complemented.

In some embodiments, endogenous genes are edited, to either inactivate or reduce enzymatic activity by changing the amino acid sequence of the encoded protein, or to reduce expression through editing of expression control sequences. Editing can modify endogenous promoters, ribosomal binding sequences, or other expression control sequences, and/or in some embodiments modifies trans-acting and/or cis-acting factors in gene regulation. Genome editing can take place using CRISPR/Cas genome editing techniques, or similar techniques employing zinc finger nucleases and TALENs. In some embodiments, the endogenous genes are replaced by homologous recombination.

The microbial cell in various embodiments does not express a recombinant biosynthesis pathway for production of precursors (e.g., comprising one or more plant enzymes). For example, for microbial cells producing RebM (and other advanced glycosylation products) in accordance with embodiments of the invention, do not express a
5 steviol biosynthesis pathway, such as copalyl synthase, kaurene synthase, kaurene oxidase and/or kaurenoic acid hydroxylase, such that production of RebM and other advanced glycosylation products is dependent on feeding steviol glycoside intermediates to the cell.

In various embodiments, the microbial cell is a bacteria selected from *Escherichia* spp., *Bacillus* spp., *Rhodobacter* spp., *Zymomonas* spp., or *Pseudomonas* spp. In some
10 embodiments, the bacterial species is selected from *Escherichia coli*, *Bacillus subtilis*, *Rhodobacter capsulatus*, *Rhodobacter sphaeroides*, *Zymomonas mobilis*, or *Pseudomonas putida*. In some embodiments, the bacterial cell is *E. coli*.

In other embodiments, the cell is a fungal cell such as a yeast cell, such as, for example, *Saccharomyces* spp., *Schizosaccharomyces* spp., *Pichia* spp., *Phaffia* spp.,
15 *Kluyveromyces* spp., *Candida* spp., *Talaromyces* spp., *Brettanomyces* spp., *Pachysolen* spp., *Debaryomyces* spp., *Yarrowia* spp., and industrial polyploid yeast strains. In an embodiment, the yeast may be a species of *Saccharomyces*, *Pichia*, or *Yarrowia*, including *Saccharomyces cerevisiae*, *Pichia pastoris*, and *Yarrowia lipolytica*. In some embodiments, the yeast cell expresses one or more bacterial transporters, or derivatives thereof, that import glycoside
20 intermediates into the cell for further glycosylation.

In various embodiments, the microbial cell has an overexpression of one or more endogenous transporters (e.g., as compared to a parent microbial strain), or in certain embodiments, is modified to express a recombinant and/or engineered transport protein. In some embodiments, the microbial cell expresses one or more additional copies of an
25 endogenous transport protein, or derivative thereof. For example, expression or activity of transport proteins can be modified to increase transport into the cell of steviol glycoside intermediates (e.g., one or more of stevioside, steviolbioside, and RebA, among others), while exporting product, such as RebM and/or RebD, and/or other advanced glycosylated steviol glycosides.

Exemplary transport proteins that can be overexpressed or engineered for altered activity or substrate specificity in the microbial cell include *E. coli* acrAD, xylE, ascF, bglF, chbA, ptsG/crr, wzxE, rfbX, as well as orthologs or derivatives thereof. Derivatives and orthologs of these proteins generally comprise an amino acid sequences having at least about 5 30%, or at least about 40%, or at least about 50% identity, at least about 60% identity, at least about 70% identity, at least about 80% identity, at least about 85% identity, or at least about 90% identity, or at least about 95% identity, or at least 96%, 97%, 98% or 99% identity to the *E. coli* transport protein.

acrAD is an *E. coli* multidrug efflux pump described in Aires JR and Nikaido H.,
10 Aminoglycosides are captured from both periplasm and cytoplasm by the AcrD multidrug efflux transporter of *Escherichia coli*, *J Bacteriol.* 2005; 187(6):1923-9. Further a homolog structure is described in Yu, EW, et al., Structural basis of multiple drug-binding capacity of the AcrB multidrug efflux pump, *Science* 2003; 300(5621):976-80.

xylE is an *E. coli* xylose transporter with homology to glucose transporters. xylE is
15 described in Sumiya M, et al., Molecular genetics of a receptor protein for D-xylose, encoded by the gene xylF, in *Escherichia coli*, *Recept Channels* 1995;3(2):117-28, with structure described by Sun L, et al., Crystal structure of a bacterial homologue of glucose transporters GLUT-1-4, *Nature* 2012;490(7420):361-6.

ascF is described by Hall BG and Xu L, Nucleotide sequence, function, activation,
20 and evolution of the cryptic asc operon of *Escherichia coli* K12, *Mol. Biol. Evol.* 1992;9(4):688-706. The asc operon is considered to play a role in cellobiose metabolism in *E. coli*.

bglF is described by Schnetz K, et al., Identification of catalytic residues in the beta-glucoside permease of *Escherichia coli* by site-specific mutagenesis and demonstration of
25 interdomain cross-reactivity between the beta-glucoside and glucose systems, *J. Biol. Chem.* 1990;265(23):13464-71. A homologous structure is described in Herzberg O., An atomic model for protein-protein phosphoryl group transfer, *J. Biol. Chem.* 1992;276(34):24819-23. bglF catalyzes transport and phosphorylation of beta-glucosides.

chba is described in Keyhani NO, et al., The transport/phosphorylation of N,N'-diacetylchitobiose in *Escherichia coli*. Characterization of phosphor-IIB(Chb) and of a potential transition state analogue in the phosphotransfer reaction between the proteins IIA(Chb) and IIB(Chb). *J. Biol. Chem.* 2000;275(42):33102-9; with structure described in

5 Tang C, et al., Solution structure of enzyme IIA(Chitobiose) from the N,N'-diacetylchitobiose branch of the *Escherichia coli* phosphotransferase system. *J. Biol. Chem.* 2005;280(12):11770-80. The phosphoenolpyruvate-dependent sugar phosphotransferase system (sugar PTS), is a major carbohydrate active transport system that catalyzes the phosphorylation of incoming sugar substrates concomitantly with their translocation across

10 the cell membrane.

ptsG encodes the glucose-specific permease of the phosphotransferase transport system (PTS) and is described in Meins M, et al., Glucose permease of *Escherichia coli*. Purification of the IIGlc subunit and functional characterization of its oligomeric forms. *J. Biol. Chem.* 1988;263(26):12986-93. A structure is described in Cai M, et al., Solution

15 structure of the phosphoryl transfer complex between the signal-transducing protein IIA_{Glucose} and the cytoplasmic domain of the glucose transporter IICB_{Glucose} of the *Escherichia coli* glucose phosphotransferase system. *J. Biol. Chem.* 2003;278(27):25191-206.

wzxE and its role in molecular transport is described by Rick PD, et al., Evidence

20 that the wzxE gene of *Escherichia coli* K-12 encodes a protein involved in the transbilayer movement of a trisaccharide-lipid intermediate in the assembly of enterobacterial common antigen. *J. Biol. Chem.* 2003;278(19):16534-42.

rfbx is a lipopolysaccharide transporter described in Hong Y, et al., Progress in our understanding of wzx flippase for translocation of bacterial membrane lipid-linked

25 oligosaccharide. *J. Bacteriol.* 2018;200(1).

Other transport proteins include those selected from bacterial or endogenous transport proteins that transport the desired glycoside intermediate into the cell, and/or transport the desired product out of the cell. For example, the transporter may be from the host species, or another bacterial or yeast species, and may be engineered to adjust its affinity

for the particular glycoside intermediates or products. For example, the host cell may overexpress a transporter that is at least about 30%, or at least about 40%, or at least about 50% identical to an *E. coli* transporter selected from ampG, araE, araJ, bcr, cynX, emrA, emrB, emrD, emrE, emrK, emrY, entS, exuT, fsr, fucP, galP, garP, glpT, gudP, gudT, hcaT, hsrA, kgtP, lacY, lgoT, lplT, lptA, lptB, lptC, lptD, lptE, lptF, lptG, mdhA, mdtD, mdtG, mdtH, mdtM, mdtL, mhpT, msbA, nanT, narK, narU, nepI, nimT, nupG, proP, setA, setB, setC, shiA, tfaP, tolC, tsgA, uhpT, xapB, xylE, yaaU, yajR, ybjJ, ycaD, ydeA, ydeF, ydfJ, ydhC, ydhP, ydjE, ydjK, ydiM, ydiN, yebQ, ydcO, yegT, yfaV, yfcJ, ygaY, ygcE, ygcS, yhhS, yhjE, yhjX, yidT, yihN, yjhB, and ynfM. In some embodiments, the transporter is at least about 60%, at least about 70%, at least about 80%, at least about 90%, or at least about 95% identical to the *E. coli* transporter.

In some embodiments, the microbial cell expresses a transport protein that is at least 50% identical to a transporter from a eukaryotic cell, such as a yeast, fungus, or plant cell. In some embodiments, the transport protein is an ABC family transporter, and which is optionally of a subclass PDR (pleiotropic drug resistance) transporter, MDR (multidrug resistance) transporter, MFS family (Major Facilitator Superfamily) transporter, or SWEET (aka PQ-loop, Saliva, or MtN3 family) family transporter. In other embodiments, the transport protein is of a family selected from: AAP, SulP, LCT, APC, MOP, ZIP, MPT, VIC, CPA2, ThrE, OPT, Trk, BASS, DMT, MC, AEC, Amt, Nramp, TRP-CC, ACR3, NCS1, PiT, ArsAB, IISp, GUP, MIT, Ctr, and CDF.

In some embodiments, the transporter is an ABC family transport protein (a/k/a ATP-binding cassette transporters), which generally include multiple subunits, one or two of which are transmembrane proteins and one or two of which are membrane-associated ATPases. The ATPase subunits utilize the energy of adenosine triphosphate (ATP) binding and hydrolysis to energize the translocation of various substrates across membranes, either for uptake or for export of the substrate. The ABC family transporter may be of any subclass, including, but not limited to: ABCA, ABCB, ABCC, ABCD, ABCE, ABCF, and ABCG.

In some embodiments, the transport protein is an MFS family transport protein (a/k/a Major Facilitator Superfamily), which are single-polypeptide secondary carriers capable of

transporting small solutes in response to chemiosmotic ion gradients. Compounds transported by MFS transport proteins can include simple sugars, oligosaccharides, inositols, drugs, amino acids, nucleosides, organophosphate esters, Krebs cycle metabolites, and a large variety of organic and inorganic anions and cations. By way of example, MFS transport proteins include XylE (from *E. coli*), QacA (from *S. aureus*), Bmr (of *B. subtilis*), UhpT (from *E. coli*), LacY (from *E. coli*), FucP (from *E. coli*), and ExtU (from *E. coli*).

In some embodiments, the transporter is of SWEET (Sugars Will Eventually be Exported Transporters) family of transport proteins (a/k/a the PQ-loop, Saliva or MtN3 family), which is a family of sugar transporters and a member of the TOG superfamily. Eukaryotic family members of SWEET have 7 transmembrane segments (TMSs) in a 3+1+3 repeat arrangement. By way of example, SWEET transporter proteins include SWEET1, SWEET2, SWEET9, SWEET12, SWEET13, and SWEET14.

In some embodiments, the transport protein is at least 50% identical to a transport protein from *S. cerevisiae*. In some embodiments, the transporter is at least about 60%, at least about 70%, at least about 80%, at least about 90%, or at least about 95% identical to the *S. cerevisiae* transporter. Exemplary *S. cerevisiae* transport proteins include AC1, ADP1, ANT1, AQR1, AQY3, ARN1, ARN2, ARR3, ATG22, ATP4, ATP7, ATP19, ATR1, ATX2, AUS1, AVT3, AVT5, AVT6, AVT7, AZR1, CAF16, CCH1, COT1, CRC1, CTR3, DAL4, DNF1, DNF2, DTR1, DUR3, ECM3, ECM27, ENB1, ERS1, FEN2, FLR1, FSF1, FUR4, GAP1, GET3, GEX2, GGC1, GUP1, HOL1, HCT10, HXT3, HXT5, HXT8, HXT9, HXT11, HXT15, KHA1, ITR1, LEU5, LYP1, MCH1, MCH5, MDL2, MME1, MNR2, MPH2, MPH3, MRS2, MRS3, MTM1, MUP3, NFT1, OAC1, ODC2, OPT1, ORT1, PCA1, PDR1, PDR3, PDR5, PDR8, PDR10, PDR11, PDR12, PDR15, PDR18, PDRI, PDRI 1, PET8, PHO89, PIC2, PMA2, PMC1, PMR1, PRM10, PUT4, QDR1, QDR2, QDR3, RCH1, SAL1, SAM3, SBH2, SEO1, SGE1, SIT1, SLY41, SMF1, SNF3, SNQ2, SPF1, SRP101, SSU1, STE6, STL1, SUL1, TAT2, THI7, THI73, TIM8, TIM13, TOK1, TOM7, TOM70, TPN1, TPO1, TPO2, TPO3, TPO4, TRK2, UGA4, VBA3, VBA5, VCX1, VMA1, VMA3, VMA4, VMA6, VMR1, VPS73, YEA6, YHK8, YIA6, YMC1, YMD8, YOR1, YPK9, YVC1, ZRT1; YBR241C, YBR287W, YDR061W, YDR338C, YFR045W, YGL114W,

YGR125W, YIL166C, YKL050C, YMR253C, YMR279C, YNL095C, YOL075C, YPR003C, and YPR011C.

In some embodiments, the *S. cerevisiae* transport protein is selected from one or more of ADP1, AQR1, ARN1, ARN2, ATR1, AUS1, AZR1, DAL4, DTR1, ENB1, FLR1, GEX2, 5 HOL1, HXT3, HXT8, HXT11, NFT1, PDR1, PDR3, PDR5, PDR8, PDR10, PDR11, PDR12, PDR15, PDR18, QDR1, QDR2, QDR3, SEO1, SGE1, SIT1, SNQ2, SSU1, STE6, THI7, THI73, TIM8, TPN1, TPO1, TPO2, TPO3, TPO4, YHK8, YMD8, YOR1, and YVC1. In some embodiments, *S. cerevisiae* transport protein is selected from one or more of FLR1, PDR1, PDR3, PDR5, PDR10, PDR15, SNQ2, TPO1, and YOR1.

10 In some embodiments, the transporter is at least 50%, at least 60% identical, at least 70% identical, at least 80% identical, or at least 90 or 95% identical to XP_013706116.1 (from *Brassica napus*), NP_001288941.1 (from *Brassica rapa*), NEC1 (from *Petunia hybrida*), and SWEET13 (from *Triticum urartu*).

In various embodiments with regard to biosynthesis of RebM, the method results in 15 at least 40% conversion, or at least 50% conversion, or at least 75% conversion of stevioside, steviolbioside, and RebA to RebM. In some embodiments, the ratio of RebM to RebD is at least 2:1, or at least 4:1, or at least 6:1, or at least 8:1, or at least 9:1, or at least 10:1, or at least 15:1, or at least 20:1.

The method may be performed by batch fermentation, fed-batch fermentation, 20 continuous fermentation, or semi-continuous fermentation. For example, in some embodiments, the method is conducted by batch fermentation or fed-batch fermentation with incubation times of less than about 72 hours, or in some embodiments, less than about 48 hours, or less than about 24 hours.

While the native UGT enzymes are generally plant enzymes (which often have 25 optimal temperatures in the range of 20-24° C) or are derived from plant enzymes, the present disclosure in some embodiments enables production of the glycosylated product at high yield in microbial cells (e.g., bacterial cells such as *E. coli*), with enzyme productivity

at temperatures about 24° C or more, such as from about 24° C to about 37° C, or from about 27° C to about 37° C, or from about 30° C to about 37° C.

In some embodiments, the growth or production phase media may contain one or more detergents in an amount sufficient to enhance cell permeability, without significant impact on growth or viability. Exemplary detergents include Tween 20, Triton X-100, and SDS, among others.

In some embodiments, the process is scalable for large scale production. For example, in some embodiments, the size of the culture is at least about 100 L, at least about 200 L, at least about 500 L, at least about 1,000 L, or at least about 10,000 L.

In various embodiments, methods further include recovering glycosylated product from the cell culture or from cell lysates. In some embodiments, the culture produces at least about 100 mg/L, or at least about 200 mg/L, or at least about 500 mg/L, or at least about 1 g/L, or at least about 2 g/L, or at least about 5 g/L, or at least about 10 g/L, or at least about 20 g/L, or at least about 30 g/L, or at least about 40 g/L, or at least about 50 g/L of the glycosylated product, which in some embodiments is extracted from the culture media.

In some embodiments, the glycosylated products (e.g., RebM) are purified from media components. Thus, in some embodiments, the methods comprise separating growth media from host cells, and isolating the desired glycosylation products (e.g., RebM) from the growth media. In some embodiments, product such as RebM is further extracted from the cellular material.

In some aspects, the invention provides methods for making a product comprising a glycosylated product, such as RebM. The method comprises incorporating the target steviol glycoside (produced according to this disclosure) into a product, such as a food, beverage, oral care product, sweetener, flavoring agent, or other product. Purified steviol glycosides, prepared in accordance with the present invention, may be used in a variety of products including, but not limited to, foods, beverages, texturants (e.g., starches, fibers, gums, fats and fat mimetics, and emulsifiers), pharmaceutical compositions, tobacco products, nutraceutical compositions, oral hygiene compositions, and cosmetic compositions. Non-

limiting examples of flavors for which RebM can be used in combination include lime, lemon, orange, fruit, banana, grape, pear, pineapple, mango, bitter almond, cola, cinnamon, sugar, cotton candy and vanilla flavors. Non-limiting examples of other food ingredients include flavors, acidulants, and amino acids, coloring agents, bulking agents, modified
5 starches, gums, texturizers, preservatives, antioxidants, emulsifiers, stabilizers, thickeners and gelling agents.

In some aspects, the invention provides methods for making a sweetener product comprising a plurality of high-intensity sweeteners, said plurality including two or more of a steviol glycoside, a mogroside, sucralose, aspartame, neotame, advantame, acesulfame
10 potassium, saccharin, cyclamate, neohesperidin dihydrochalcone, gnetifolin E, and/or piceatannol 4'-O- β -D-glucopyranoside. The method may further comprise incorporating the sweetener product into a food, beverage, oral care product, sweetener, flavoring agent, or other product, including those described above.

Target steviol glycoside(s), such as RebM, and sweetener compositions comprising
15 the same, can be used in combination with various physiologically active substances or functional ingredients. Functional ingredients generally are classified into categories such as carotenoids, dietary fiber, fatty acids, saponins, antioxidants, nutraceuticals, flavonoids, isothiocyanates, phenols, plant sterols and stanols (phytosterols and phytostanols); polyols; prebiotics, probiotics; phytoestrogens; soy protein; sulfides/thiols; amino acids; proteins;
20 vitamins; and minerals. Functional ingredients also may be classified based on their health benefits, such as cardiovascular, cholesterol-reducing, and anti-inflammatory.

Further, target steviol glycoside(s), such as RebM, and sweetener compositions obtained according to this invention, may be applied as a high intensity sweetener to produce zero calorie, reduced calorie or diabetic beverages and food products with improved taste
25 characteristics. It may also be used in drinks, foodstuffs, pharmaceuticals, and other products in which sugar cannot be used. In addition, RebM and sweetener compositions can be used as a sweetener not only for drinks, foodstuffs, and other products dedicated for human consumption, but also in animal feed and fodder with improved characteristics.

Examples of products in which target steviol glycoside(s) and sweetener compositions may be used include, but are not limited to, alcoholic beverages such as vodka, wine, beer, liquor, and sake, etc.; natural juices; refreshing drinks; carbonated soft drinks; diet drinks; zero calorie drinks; reduced calorie drinks and foods; yogurt drinks; instant
 5 juices; instant coffee; powdered types of instant beverages; canned products; syrups; fermented soybean paste; soy sauce; vinegar; dressings; mayonnaise; ketchups; curry; soup; instant bouillon; powdered soy sauce; powdered vinegar; types of biscuits; rice biscuit; crackers; bread; chocolates; caramel; candy; chewing gum; jelly; pudding; preserved fruits and vegetables; fresh cream; jam; marmalade; flower paste; powdered milk; ice cream;
 10 sorbet; vegetables and fruits packed in bottles; canned and boiled beans; meat and foods boiled in sweetened sauce; agricultural vegetable food products; seafood; ham; sausage; fish ham; fish sausage; fish paste; deep fried fish products; dried seafood products; frozen food products; preserved seaweed; preserved meat; tobacco; medicinal products; and many others.

15 During the manufacturing of products such as foodstuffs, drinks, pharmaceuticals, cosmetics, table top products, and chewing gum, the conventional methods such as mixing, kneading, dissolution, pickling, permeation, percolation, sprinkling, atomizing, infusing and other methods may be used.

20 Embodiments of the invention are demonstrated in the following, non-limiting examples.

EXAMPLES

Table 2 shows the steviol glycoside content of three batches of stevia leaf extract. Two intermediates, RebA and stevioside, on the pathway to RebM, are the two primary glycosides in the batches.

25 Table 2: Steviol Glycoside Composition of Available Stevia Leaf Extract

%	Batch 1	Batch 2	Batch 3

Rebaudioside A	38.2	10.5	30.3
Stevioside	8.5	9.0	18.4
Rebaudioside C	12.9	4.2	16.6
Rebaudioside B	4.3	7.1	1.2
Rubusoside	5.0	2.2	2.0
Rebaudioside F	2.0	2.7	2.1
Steviolbioside	0.3	3.7	0.3
Rebaudioside D	0.2	2.1	0.9
Dulcoside A	0.9	0.4	0.5

FIG. 4 shows bioconversion of glycosylated steviol glycoside intermediates. In the experiment, 0.2 mM steviolbioside was fed to engineered *E. coli* strains in 96 well plates. Product formation was examined after 48 hours. Whole cell bioconversion of even an early intermediate such as steviolbioside is possible, but for native *E. coli* the glycosylated substrate is not available to the UGT enzymes (Strain 5 shows substantial activity, while Strains 3 and 4 show negligible activity). Details of Strains are described in FIG. 8. The values reported in FIG. 4 are the compound concentration relative to control (the concentration of each compound is divided by the total concentration of these compounds for the empty vector control). The modifications contained in Strain 5 demonstrate bioconversion of steviolbioside to RebM.

Each of Strains 1-5 contained a BAC (bacterial artificial chromosome), that is a single copy plasmid, which expresses four UGT enzymes separately: MbUGTC13, MbUGT1.2_2, MbUGTc19_2, and 76G1_L200A. The control contained the same BAC backbone but without the UGT enzymes.

FIG. 8 shows genetic modifications to increase the native *E. coli* flux to UDP-glucose, a critical substrate for the UGTs in the RebM pathway. Greater than native flux to UDP-glucose may allow for greater RebM formation by increasing the amount of substrate available to the UGTs. The modifications resulting in Strain 5 also result in the ability of the cell to convert fed steviol glycosides to RebM. For Strain 2 and 3, enzymes that consume UDP-glucose (*ushA*, *galETKM*) were deleted. Strain 4 was built from Strain 3 by deleting an enzyme (*pgi*) that consumes a precursor of UDP-glucose, glucose-6-phosphate (G6P). Strain 5 was built from Strain 4 by overexpressing two enzymes that are used to convert G6P to UDP-glucose (*pgm*, *galU*) via glucose-1-phosphate (G1P).

FIG. 5 shows conversion of two commercially available intermediates found in leaf extract, stevioside and RebA, to RebM. Also shown is the bioconversion of the steviol core. The values are reported as the percentage of the total steviol glycoside composition. Conversion of steviol to RebM by all strains shows that all strains are capable of RebM formation. As was the case for steviolbioside, both stevioside and RebA are converted to RebM, but only in Strain 5. It is likely that only Strain 5 makes steviol glycosides available to the UGT enzymes. For conversion from stevioside and RebA, the ratio of RebD:RebM strongly favors RebM (1:20). RebM was secreted into the media.

FIG. 6 shows bioconversion of steviol glycosides in 96-well plates. For these studies, 76G1_L200A was replaced with MbUGT1-3_1.5. For five major compounds in the RebM pathway, compound concentrations are shown with or without UGT enzymes expressed. Strains were fed 0.5 g/L of stevioside/RebA leaf extract and incubated with substrate for 48 hours before sampling. As shown, the cells make almost exclusively RebM, with only small amounts of RebI and RebD.

The strain was used for initial bioreactor experiments. Specifically, the RebM-producing strain was inoculated from a working cell bank in a 50 mL centrifuge tube containing 10 mL of LB (Luria-Bertani) broth with suitable antibiotics. The first seed was cultured for 20 hours in a shaking incubator at 37°C. After 20 hours, the second seed culture was then started by inoculating 100 mL of fermentation medium in a 500 mL Erlenmeyer flask using the first seed. This second seed was cultured for 10 hours before transferring to

the the bioreactors, each containing 170 mL of fermentation medium. Sampling was performed every 10 hours. Relevant metabolites in the medium were quantified using LC-MS-MS and YSI. Cell density was measured via a spectrophotometer at absorbance 600 nm. FIG. 7 shows bioconversion of a stevioside/RebA leaf extract at bioreactor scale. For five
5 major compounds in the RebM pathway, compound concentrations in the media were sampled at various time points across the fermentation. Strains were fed 2 g/L of the stevioside/RebA leaf extract and incubated with substrate for a total of 30 hours. As shown in FIG. 7, stevioside and RebA are lost over time, with the corresponding increase in highly glycosylated products such as RebM, RebI, and RebD.

10 The UGT enzymes that convert pathway intermediates such as steviolbioside, stevioside, and RebA to RebM are expressed in the *E. coli* cytoplasm and thus require intermediates to become available to the UGT enzymes, likely through the action of a transporter or through increased membrane permeability.

15 Steviol glycosides are large molecules that likely are not taken up by native *E. coli* strains. This would explain the negligible conversion of steviolbioside to RebM by Strain 1. Strains 1-4 can take up an earlier, non-glycosylated intermediate (steviol) and convert it to RebM, suggesting the reason for negligible conversion is a lack of uptake of steviol glycosides to the cytoplasm, not inactivity of the UGT enzymes on the pathway intermediates under these conditions.

SEQUENCES

>SrUGT85C2 [*Stevia rebaudiana*]

5 (SEQ ID NO:1)

MDAMATTEKKPHVIFIPFPAQSHIKAMLKLAQLLHHKGLQITFVNTDFIHNQFLESSGPHCLDGAP
GFRFETIPDGVSHSPEASIPRESLLRSIETNFLDRFIDLVTKLDPPTCIIISDGFLSVFTIDA
KLGIPVMMYWTAAACGFMGFYHIHSLIEKGFAPLKDASYLTNGYLDTVIDWVPGMEGIRLKDFPLD
10 WSTDNDKVLMTTEAPQRSHKVVSHHIFHTFDELEPSIIKTLRLRYNHIYTIGPLQLLLDQIPEEK
KQTGITSLHGYSLVKEEPCFQWLQSKEPNSVVYVNFSGSTTVMSLEDMTEFGWGLANSNHYFLWII
RSNLVIGENAVLPPELEEHKKRGFIASWCSQEKVLKHPSVGGFLTHCGWGSTIESLSAGVPMICW
PYSWDQLTNCRYICKEWVGLMGTKVKRDEVKRLVQELMGEGGHKMRNKAKDWKEKARIAIAPNG
15 SSSLNIDKMVKEITVLARN

>SrUGT74G1 [*Stevia rebaudiana*]

20 (SEQ ID NO:2)

MAEQQKIKKSPHVLIIFFPLQGHINPFIQFGKRLISKGVKTTLVTTIHTLNSTLNHSNTTTTTSIEI
QAISDGCDEGGFMSAGESYLETFKQVGSKSLADLIKKLQSEGTTIDAIYDSMTEWVLDVAIEFGI
DGGSFFTQACVVNSLYYHVHKGGLISLPLGETVSVPGFPVLQRWETPLILQNHQIQSPWSQMLFGQ
FANIDQARWVFTNSFYKLEEEVIEWTRKIWNLVIGPTLPSMYLDKRLDDDDKNGFNLYKANHHEC
25 MNWLDDPKESVVYVAFGSLVKHGPEQVEEITRALIDSDVNFLWVIKHKEEGKLPENLSEVIKTGK
GLIVAWCKQLDVLAHESVGC FVTHCGFNSTLEAISLGVPVVAMPQFSDQTTNAKLLDEILGVGVRV
KADENGIVRRGNLASCIMIMEEERGVIIRKNAVKKWDLAKVAVHEGGSSDNDIVEFVSELIKA

>SrUGT76G1 [*Stevia rebaudiana*]

30 (SEQ ID NO:3)

MENKTETTVRRRRRIILFPVPFQGHINPILQLANVLYSKGFSITIFHTNFNKPKTSNYPHFTFRFI
LDNDPQDERISNLPTHGPLAGMRIPIINEHGADELRRLELLMLASEEDEEVSCITDALWYFAQS
35 VADSLNLRRLVLMTSSLFNFHAHVSLPQFDELGYLDPDDKTRLEEQASGFPMKVKDIKSAYSNNWQ
ILKEILGKMIKQTKASSGVIWNSFKELEESELETVIREIPAPSFLIPLPKHLTASSSSLLDHDRTV
FQWLDQQPPSSVLYVSFGSTSEVDEKDFLEIARGLVDSKQSFLWVVRPGFVKGSTWVEPLPDGFLG
ERGRIVKWVPQQEVLAHGAIGAFWTHSGWNSTLESVCEGVPMIFSDFGLDQPLNARYMSDVLKGV
YLENGWERGEIANAIRVMVDEEGEYIRQNARVLKQKADVSLMKGGSSYESLESLSVSYISSL
40

>SrUGT91D1 [*Stevia rebaudiana*]

45 (SEQ ID NO:4)

MYNVTYHQNSKAMATSDSIVDDRKLHVATFPWLAFGHILPFLQLSKLIAEKGHKVSFLSTTRNIQ
RLSSHISPLINVVQLTLPRVQELPEDAEATTDVHPEDIQYLKKAVDGLQPEVTRFLEQHSPDWIIY
DFTHYWLPSIAASLGISRAYFCVITPWTIAYLAPSSDAMINDSDGRTTVEDLTTPPKWFPFPTKVC
WRKHDLARMEPYEAPGISDGYRMGMVFKGSDCLLFKCYHEFGTQWLPLETLHQVPVVPVGLLPPE
IPGDEKDETWVSIKKWLDGKQKGSVVYVALGSEALVSQTEVVELALGLELSGLPFVWAYRKPKGPA
50 KSDSVELPDGFVERTRDRGLVWTSWAPQLRILSHESVCGFLTHCGSGSIVEGLMFGHPLIMLPIFC

DQPLNARLLEDKQVGIEIPRNEEDGCLTKESVARSLRSVVVENEGEIYKANARALSKIYNDTKVEK
EYVSQFVDYLEKNARAVAIDHES

>SrUGT91D2 [*Stevia rebaudiana*]

5

(SEQ ID NO:5)

MATSDSIVDDRKQLHVATFPWLAFGHILPYLQLSKLI AEKGHKVSFLSTTRNIQRLSSHISPLINV
VQLTLPRVQELPEDAEATTDVHPEDI PYLKKASDGLQPEVTRFLEQHSPDWI IYDYTHYWLPSIAA
10 SLGISRAHFSVTTTPWAIAYMGPSADAMINGSDGRTTVEDLTTPPKWFPFPTKVCWRKHDLARLV
KAPGISDGYRMGLVLKGSDCLLSKCYHEFGTQWLPLETLHQVPVVPVGLLPPEVPGDEKDETWVS
IKKWLDGKQKGSVVYVALGSEVLVSQTEVVELALGLELSGLPFVWAYRKPKGPAKSDSVELPDGFV
ERTRDRGLVWTSWAPQLRILSHESVCGFLTHCGSGSIVEGLMFGHPLIMLP IFGDQPLNARLLEDK
15 QVGIEIPRNEEDGCLTKESVARSLRSVVVEKEGEIYKANARELSKIYNDTKVEKEYVSQFVDYLEK
NTRAVIDHES

>SrUGT91D2e [*Stevia rebaudiana*]

(SEQ ID NO:6)

20

MATSDSIVDDRKQLHVATFPWLAFGHILPYLQLSKLI AEKGHKVSFLSTTRNIQRLSSHISPLINV
VQLTLPRVQELPEDAEATTDVHPEDI PYLKKASDGLQPEVTRFLEQHSPDWI IYDYTHYWLPSIAA
SLGISRAHFSVTTTPWAIAYMGPSADAMINGSDGRTTVEDLTTPPKWFPFPTKVCWRKHDLARLV
KAPGISDGYRMGLVLKGSDCLLSKCYHEFGTQWLPLETLHQVPVVPVGLLPPEIPGDEKDETWVS
25 IKKWLDGKQKGSVVYVALGSEVLVSQTEVVELALGLELSGLPFVWAYRKPKGPAKSDSVELPDGFV
ERTRDRGLVWTSWAPQLRILSHESVCGFLTHCGSGSIVEGLMFGHPLIMLP IFGDQPLNARLLEDK
QVGIEIPRNEEDGCLTKESVARSLRSVVVEKEGEIYKANARELSKIYNDTKVEKEYVSQFVDYLEK
NARAVAIDHES

30 > OsUGT1-2 [*Oryza sativa*]

(SEQ ID NO:7)

MDSGYSSSYAAAAGMHVVICPWLAFGHLLPCLDLAQLASRGHRVSFVSTPRNISRLPPVRPALAP
35 LVAFVALPLPRVEGLPDGAESTNDVPHDRPDMVELHRRAFDGLAAPFSEFLGTACADWVI DVVFHH
WAAAAALEHKVPCAMMLLGSAMIASIADRRLERAEETSPAAAGQGRPAAAPTFEVARMKLI RTKG
SSGMSLAERFSLTLSSSLVVGRSCVEFEPETVPLLSTLRGKPITFLGLMPPLHEGRREDGEDATV
RWLDAQPAKSVVYVALGSEVPLGVEKVHELALGLELAGTRFLWALRKPTGVSDADLLPAGFEERTR
GRGVVATRWVPQMSILAHAAVGAFLTHCGWNSTIEGLMFGHPLIMLP IFGDQGPNARLIEAKNAGL
40 QVARNDGDGSFDREGVAAAIRAVAVEEESSKVFQAKAKKLQEIVADMACHERYIDGFIQQLRSYKD

>MbUGTC19 [Circular permutant]

45 (SEQ ID NO:8)

MAECMNWLDDKPKE SVVYVAFGSLVKHGPEQVEEITRALIDSDVNFLWVIKHKEEGKLPENLSEVI
KTGKGLIVAWCKQLDVLAHESVGC FVTHCGFNSTLEAISLGVPVVAMPQFSDQTTNAKLLDEILGV
GVRVKADENGIVRRGNLASC IKMIMEEERGV IIRKNAVKKDLAKVAVHEGGSSDNDIVEFVSELI
50 KAGSGEQQKIKKSPHVLLIPFPLQGHINPFIQFGKRLISKGVKTTLVTTIHTLNSTLNHSNTTTTS
IEIQAISDGCDEGGFMSAGESYLETFKQVGSKSLADLIKKLQSEGTTIDAI IYDSMTEWVLDVAIE
FGIDGGSFFTQACVVNSLYYHVHKG L ISLPLGETVSVPGFPVLQRWETPLILQNHEQIQSPWSQML

FGQFANIDQARWVFTNSFYKLEEEVIEWTRKIWNLKVIGPTLPSMYLDKRLDDDDKDNGFNLYKANH
H

>MbUGT1,2 [Circular permutant]

5

(SEQ ID NO:9)

MAGSSGMSLAERFSLTSLRSSLVVGRCSEFEFETVPLLSTLRGKPITFLGLMPPLHEGRREDGED
ATVRWLDAQPAKSVVYVALGSEVPLGVEKVHELALGLELAGTRFLWALRKPTGVSDADLLPAGFEE
10 RTRGRGVVATRWVPQMSILAHAAVGAFETHCGWNSTIEGLMFHGHLIMLPIFGDQGNARLIEAKN
AGLQVARNDGDGSFDREGVAAAIRAVAVEEESSKVFQAKAKKLQEIVADMACHERYIDGFIQQLRS
YKDDSGYSSSYAAAAGMHVVICPWLAFGHLLPCLDLAQLASRGHRVSFVSTPRNISRLPPVRPAL
APLVAFAVALPLPRVEGLPDGAESTNDVPHDRPDMVELHRRAFDGLAAPFSEFLGTACADWVIVDVF
15 HHWAAAAALEHKVPCAMMLLGSAMHIAIADRRLEAETESPAAAGQGRPAAAPTFEVARMKLIRK

>MbUGT1-3 [Circular permutant]

(SEQ ID NO:10)

20

MANWQILKEILGKMIKQTKASSGVIWNSFKELEESELETVIREIPAPSFILPLPKHLTASSSSLLD
HDRTVFQWLDQQPPSSVLYVSFGSTSEVDEKDFLEIARGLVDSKQSFLWVVRPGFVKGSTWVEPLP
DGFLGERGRIVKWVPQQEVLAHGAIGAFWTHSGWNSTLESVCEGVPMIFSDFGLDQPLNARYMSDV
LKVGVYLENGWERGEIANAIRVMVDEEGEYIRQNARVLKQKADVSLMKGGSSYESLESLSVSYISS
25 LENKTETTTRRRRRRIILFPVPFQGHINPILQLANVLYSKGFSITIFHTNFNKPKTSNYPHFTFRFI
LDNDPQDERISNLPHGPLAGMRIPIINEHGADELRRLELLMLASEEDEEVSLITDALWYFAQS
VADSLNLRRLVLMTSSLFNFHAHVS LPQFDELGYLDPDDKTRLEEQASGFPMKVKDIKSAYS

30 >MbUGT1,2.2 [Circular permutant]

(SEQ ID NO:11)

MATKGSSGMSLAERFWLTLRSSLVVGRCSEFEFETVPLLSTLRGKPITFLGLMPPLHEGRREDG
35 EDATVRWLDAQPAKSVVYVALGSEVPLGVEKVHELALGLELAGTRFLWALRKPTGVSDADLLPAGF
EERTRGRGVVATRWVPQMSILAHAAVGAFETHCGWNSTIEGLMFHGHLIMLPIFGDQGNARLIEA
KNAGLQVARNDGDGSFDREGVAAAIRAVAVEEESSKVFQAKAKKLQEIVADMACHERYIDGFIQQL
RSYKDDSGYSSSYAAAAGMHVVICPWLAFGHLLPCLDLAQLASRGHRVSFVSTPRNISRLPPVRP
ALAPLVAFAVALPLPRVEGLPDGAESTNDVPHDRPDMVELHRRAFDGLAAPFSEFLGTACADWVID
40 VFHHWAAAAALEHKVPCAMMLLGSAMHIAIADRRLEHAETESPAAAGQGRPAAAPTFEVARMKLIR

>MbUGTC19-2 [Circular permutant]

45

(SEQ ID NO:12)

MANHHECMNWLDKPKESVVYVAFGSLVKHGPEQVEEITRALIDSDVNFLWVIKHKEEGKLPENLS
EVIKTGKGLIVAWCKQLDVLAHESVGC FVTHCGFNSTLEAISLGVPV VAMPQFSDQTTNAKL LDEI
50 LGVGVRVKADENGIVRRGNLASCIMIMEEERGVIIRKNAVWKDLAKVAVHEGGSSDNDIVEFVS
ELIKAGSGEQQKIKKSPHVLLIPFPLQGHINPFIQFGKRLISKGVKTTLVTTIHTLNSTLNHSNTT
TTSIEIQAISDGCDEGGFMSAGESYLETFKQVGSKSLADLIKKLQSEGTTIDAIYDSMTEWVLDV

AIEFGIDGGSFFTQACVVNSLYYHVHKGLISLPLGETVSVPGFPVLQRWETPLILQNHEQIQSPWS
QMLFGQFANIDQARWVFTNSFYKLEEEVIEWTRKIWNLKVIGPTLPSMYLDKRLDDDKDNGFNLYK
A

5

>MbUGTC13 (*Stevia rebaudiana* UGT85C2, P215T)

(SEQ ID NO:13)

10 MADAMATTEKKPHVIFIPFPAQSHIKAMLKLAQLLHHKGLQITFVNTDFIHNQFLESSGPHCLDGA
PGFRFETIPDGVSHSPEASIPRESLLRSIETNFLDRFIDLVTKLDPPTCIISDGFLSVFTIDAA
KKLGIPVMYWTAAACGFMGFYHIHSLIEKGFAPLKDASYLTNGYLDTVIDWVPGMEGIRLKDFPL
DWSTDLDNDKVLMTTEATQRSHKVSHHIFHTFDELEPSIIKTLSLRYNHIYTIGPLQLLLDQIPEE
KKQTGITSLSHGYSLVKEEPECQWLQSKEPNSVVYVNFSGSTVMSLEDMTFGWGLANSNHYFLWI
15 IRSNLVIGENAVLPPELEEHKKRGFIASWCSQEKVLKHPSVGGFLTHCGWGSTIESLSAGVPMIC
WPYSWDQLTNCRYICKEWVGLMGTKVKRDEVKRLVQELMGEGGHKMRNKAKDWEKARIAIAPN
GSSSLNIDKMVKEITVLARN

>76G1_L200A [*Stevia rebaudiana*, L200A]

20 (SEQ ID NO:14)

MAENKTETTVRRRRRIILFPVPFQGHINPILQLANVLYSKGFSITIFHTNFNPKPTSNNYPHFTFRF
ILDNDPQDERISNLPTHGPLAGMRIPIINEHGADELRRLELELLMLASEEDEEVSLITDALWYFAQ
SVADSLNLRRLVLMTSSLFNFHAHVSLPQFDELGYLDPDDKTRLEEQASGFPMMLKVVDIKSAYSNW
QIAKEILGKMIKQTKASSGVIWNSFKELEESELETVIREIPAPSFILPLPKHLTASSSSLLDHDRT
25 VFQWLDQQPPSSVLYVSFGSTSEVDEKDFLEIARGLVDSKQSFLLWVVRPGFVKGSTWVEPLPDGFL
GERGRIVKWVPQQEVLAHGAIGAFWTHSGWNSTLESVCEGVPMIFSDFGLDQPLNARYMSDVLKVG
VYLENGWERGEIANAIRVMVDEEGEYIRQNARVLKQKADVSLMKGSSYESLESLSVSYISSL

30 >MbUGT1-3_1 [Circular permutant]

(SEQ ID NO:15)

MAFLWVVRPGFVKGSTWVEPLPDGFLGERGRIVKWVPQQEVLAHGAIGAFWTHGGWNSTLESVCEG
VPMIFSDFGLDQPLNARYMSDVLKVG VYLENGWERGEIANAIRRLMVDEEGEYIRQNARVLKQKAD
35 VSLMKGSSYESLESLSVSYISSLGSGSGSGSRRRRRIILFPVPFQGHINPMLQLANVLYSKGFSIT
IFHTNFNPKPTSNNYPHFTFRFILDNDPQDERISNLPTHGPLAGMRIPIINEHGADELRRLELELLML
ASEEDEEVSLITDALWYFAQSVADSLNLRRLVLMTSSLFNFHAHVSLPQFDELGYLDPDDKTRLE
EQASGFPMMLKVVDIKSAYSNWQIAKEILGKMIKQTKASSGVIWNSFKELEESELETVIREIPAPSF
LIPLPKHLTASSSSLLDHDRTVFQWLDQQPPSSVLYVSFGSTSEVDEKDFLEIARGLVDSQS

40

>MbUGT1-3_1.5 [Circular permutant]

(SEQ ID NO:16)

45 MAFLWVVRPGFVKGSTWVEPLPDGFLGERGRIVKWVPQQEVLAHGAIGAFWTHGGWNSTLESVCEG
VPMIFSDFGLDQPLNARYMSDVLKVG VYLENGWERGEIANAIRRLMVDEEGEYIRQNARVLKQKAD

5 VSLMKGGSSYESLESLSVSYISSLGSGSGSGRRRIILFPVPFQGHINPMLQLANVLYSKGFSIT
IFHTNFNKPKTSNYPHFTFRFILDNDPQTTHGPLAGMRIPINEHGADELRRLELQMLASEEDEE
VSLITDALWYFAQSVADSLNLPRLVLTSSLFNFHHAHVSLPQFDELGYLDPDDKTRLEEQASGFP
MLKVKDIKSAYSNNWQIAKEILGMIKQTKASSGVIWNSFKELEESELETVIREIPAPSFLIPLPKH
LTASSSSLLDHDRTVTFQWLDQQPPSSVLYVSFGSTSEVDEKDFLEIARGLVDSQS

10 >MbUGT1-3_2 [Circular permutant]
(SEQ ID NO:17)
MAFLWVVRPGFVKGSTWVEPLPDGFLGERGRIVKWVPQQEVLAHGAIGAFWTHGGWNSTLESVCEG
VPMIFSDFGLDQPLNARYMSDVLKGVYLENGWERGEIANAIRRLMVDEEGEYIRQNARVLKQKAD
VSLMKGGSSYESLESLSVSYISSLGSGSGSGRRRIILFPVPFQGHINPMLQLANVLYSKGFSITIFH
15 TNFNKPKTSNYPHFTFRFILDNDPQDERISNLPTHGPLAGMRIPINEHGADELRRLELQMLASE
EDEEVSLITDALWYFAQSVADSLNLPRLVLTSSLFNFHHAHVSLPQFDELGYLDPDDKTRLEEQASGFP
MLKVKDIKSAYSNNWQIAKEILGMIKQTKASSGVIWNSFKELEESELETVIREIPAPSFLIPL
LPKHLTASSSSLLEHDRTVTFQWLDQQPPSSVLYVSFGSTSEVDEKDFLEIARGLVDSQS

20 >SgUGT720-269-1 (*Siraitia grosvenorii*)
(SEQ ID NO:18)
MEDRNAMDMSRIKYRPQPLRPASMVQPRVLLFPFPALGHVKPFLSLAELLSDAGIDVVFSTEYNH
RRISNTEALASRFPTLHFETIPDGLPPNESRALADGPLYFSMREGTKPRFRQLIQSLNDGRWPITC
IITDIMLSSPIEVAEEFGIPVIAFCPCSARYLSIHFFIPKLVEEGQIPYADDDPIGEIQGVPLFEG
LLRRNHLPGSWSDKSADISFSHGLINQTLAAGRASALILNTFDELEAPFLTHLSSI FNKIYTI GPL
25 HALSKSRLGDSSSSASALSGFWKEDRACMSWLDQPPRSVVFVSFGSTMKMADELREFWYGLVSS
GKPFLCVLRSDVVS GGEEAAELIEQMAEEEGAGGKLG MVVEWAAQEKVLSHPAVGGFLTHCGWNSTV
ESIAAGVPMMCWPILGDQPSNATWIDRVWKIGVERNREWDRLTVEKMVRALMEGQKRVEIQRSME
KLSKLANEKVVRGINLHPTISLKKDTPTTSEHPRHEFENMRGMNYEMLVGNAIKSPTLTCK

30 >SgUGT94-289-3 (*Siraitia grosvenorii*)
(SEQ ID NO:19)
MTIFFSVEILVLGIAEFAAIAMDAAQQGDTTITILMLPWLGYGHLSAFLELAKSLSRNRFHIYFCST
SVNLDAIKPKLPSSFSDSIQFVELHLPSSPEFPPHLHTTNGLPPTLMPALHQAFSMAAQHFESILQ
TLAPHLLIYDSLQPWAPRVASSLKIPAINFNTTG VFVISQGLHPIHYPHSKFPFSEFVLHNHWKAM
35 YSTADGASTERTRKRGEAFLYCLHASCSVILINSFRELEGKYMDYLSVLLNKKVVPVGPLVYEPNQ
DGEDEGYSSIKNWLDKKEPSSTVFVSFGSEYFPSKEEMEEIAHGLEASEVNFIVVVRFPQGDNTSG
IEDALPKGFLERAGERGMVVKGWAPQAKILKHWS TGGFVSHCGWNSVMESMMFGVPIIGVPMHVDQ
PFNAGLVEEAGVGVEAKRDPDGKIQRDEVAKLIKEVVVEKTREDVRKKAREMSEILRSKGEEKFDE
MVAEISL L L L K I

40 >SgUGT74-345-2 (*Siraitia grosvenorii*)
(SEQ ID NO:20)

MDETTVNGGRRASDVVVFAPRHHGMSMPLQFSKRLVSKGLRVTFLITTSATESLRLNLPPSSSLD
 LQVISDVPESENDIATLEGYLRFSKATVSKTLADFIDGIGNPPKFIVYDSVMPWVQEVARGRLDAA
 PFFTQSSAVNHILNHVYGGSL SIPAPENTAVSLPSMPVLQAEDLPAFPDDPEVVMNFMTSQFSNFQ
 DAKWIFFNTFDQLECKKQSQVWNWMAWRPPIKTVGPTIPSAYLDDGRLEDDRAFGNLNLLKPEDGKN
 5 TRQWQWLDKSDTASVLYISFGSLAILQEEQVKELAYFLKDTNLSFLWVLRDSELQKLPHNFVQETS
 HRGLVVNWCSQLQVLSHRAVSCFVTHCGWNSTLEALS LGVPMVAIPQWVDQTTNAKFVADVVRVGV
 RVKKKDERIVTKEELEASIRQVVQGEGRNEFKHNAIKWKKLAKEAVDEGGSSDKNIEEFVKTI A

>SgUGT75-281-2 (*Siraitia grosvenorii*)

10 (SEQ ID NO:21)

MGDNQDGGGKELKENVKKGKELGRQAIGEGYINPSLQLARRLISLGVNVTFATTVLAGRMMKNKT
 HQTATTPGLSFATFSDGFDDETLKPNGDLTHYFSELRRCGSESLTHLITSAANEGRPITFVIYSL
 LSWAADIASTYDIPSAFFAQPATVLALYFYYFHGYGDTICSKLQDPSSYIELPGLPLLT SQDMPS
 FFSPSGPHAFILPPMREQAEFLGRQSQPKVLNVTDALEADALRAIDKLKMLAIGPLIPSALLGGN
 15 DSSDASFCDLQVSSDYIEWLNSKPDSSVYISVGSICVLSDEQEDELVHALLNSGHTFLWVKR
 SKENNEGKQETDEEKLKKLEEQGMVSWCRQVEVLKHPALGCFLTHCGWNSTIESLVSGLPVVA
 PQQIDQATNAKLIEDVWKTGVRVKANTEGIVEREEIRCLDLVMGSRDQKEEIERNAKKWKELAR
 QAIGEGGSSDSNLKTFLEWIDLEI

20 >SgUGT720-269-4 (*Siraitia grosvenorii*)

(SEQ ID NO:22)

MAEQAHDLHLVLLFPFPAEGHIKPFCLAEELCNAGFHVTFNLTDYNHRRHLNLHLLAARFPSLHF
 ESISDGLPPDQPRDILDPKFFISICQVTKPLFRELLLSYKRISSVQTGRPPITCVITDVI FRFPID
 VAEELDIPVFSFCTFSARFMFLYFWIPKLIEDGQLPYPNGNINQKLYGVAPEAEGLLRCKDLPGHW
 25 AFADKDDQLNFVDQTTASSRSSLILNTFDDLEAPFLGRLSTIFKKIYAVGPIHSLNLSHHCGL
 WKEDHSLAWLDSRAAKSVVFSFGSLVKITSRQLMEFWHGLNSGKSFLFVLRSDVVEGDDEKQV
 VKEIYETKAEGKWL VVGWAPQEKVLAHEAVGGFLTHSGWNSILESIAAGVPMISCPKIGDQSSNCT
 WISKVWKIGLEMEDRYDRVSVETMVR SIMEQEGEKMQKTIAELAKQAKYKVSKDGT SYQNLECLIQ
 DIKKLNQIEGFINNPNFSDLLRV

30

>SgUGT94-289-2 (*Siraitia grosvenorii*)

(SEQ ID NO:23)

MDAQQGHGTTTILMLPWVG YGHLLPFLELAKSLSRKLFHIYFCSTSVSLDAIKPKLPSSISSDDSI
 QLVELRLPSSPELPPHLHTTNGLP SHLMPALHQA FVMAAQHFQVILQTLAPHL LIYDILQPWAPQV
 35 ASSLNIPAINFSTTGASMLSR TLHPTHYPSSKFP ISEFVLHNHWRAMYTTADGALTEEGHKIEETL
 ANCLHTSCGVVLVNSFRELETKYIDYLSVLLNKKVVPVGPLVYEPNQEGEDEGYSSIKNWL DKKEP
 SSTVFVFSFGTEYFPSKEEMEEIAYGLELSEVNFIWVLRFPQGDSTSTIEDALPKGFLERAGERAMV
 VKGWAPQAKILKHWSTGGLVSHCGWNSMMEGMMFGVPII AVPMHLDQPFNAGLVEEAGVGVEAKRD
 SDGKIQREEVAKSIKEVVIEKTREDVRKKAREMDTKHGPTYFSRSKVSSFGRLYKINRPTTLTVGR
 40 FWSKQIKMKRE

>SgUGT94-289-1 (*Siraitia grosvenorii*)

(SEQ ID NO:24)

MDAQRGHTTTILMFPWLGYGHLSAFLELAKSLSRNFHIYFCSTSVNLDAIKPKLPSSSSSDSIQL
5 VELCLPSSPDQLPPHLHTTNALPPHLMPTLHQAFSMAAQHFAAILHTLAPHLLIYDSFQPWAPQLA
SSLNIPAINFNTTGASVLTRMLHATHYPSSKFPISEFVLHDYWKAMYSAAAGGAVTKKDHKIGETLA
NCLHASCSVILINSFRELEEKYMDYLSVLLNKKVVPVGPLVYEPNQDGEDEGYSSIKNWLDKKEPS
STVFVSFGSEYFPSKEEMEEIAHGLEASEVHFIWVVRFPQGDNTSAIEDALPKGFLERVGERGMVV
10 KGWAPQAKILKHWSTGGFVSHCGWNSVMESMMFGVPIIGVPMHLDQPFNAGLAEEAGVGVEAKRDP
DGKIQRDEVAKLIKEVVVEKTREDVRKKAREMSEILRSKGEEKMDEMVA AISLFLKI

>XP_022151474.1 7-deoxyloganetic acid glucosyltransferase-like
[*Momordica charantia*]

(SEQ ID NO:25)

MAQPQTQARVLVFPYPTVGHKIPFLSLAELLADGGLDVVFLSTEYNHRRIPNLEALASRFPTLHFD
15 TIPDGLPIDKPRV IIGGELYTSMRDGVKQRLRQVLQSYNDGSSPITCVICDVMLSGPIEAAEELGI
PVVTFCPYSARYLCAHFVMPKLIIEGQIPFTDGNLAGEIQGVPLFGGLLRDHLPGFWFVKSLSDE
VWSHAFLNQTLAVGRTSALIINTLDELEAPFLAHLSSSTFDKIYPIGPLDALSKSRLGDSSTSVL
TAFWKEDQACMSWLD SQPPKSVIFVSFGSTMRTADKLVEFWHGLVNSGTRFLCVLRSDIVEGGGA
20 ADLIKQVGETGNGIVVEWAAQEKVLAHRAVGGLTHCGWNSTMESIAAGVPMMCWQIYGDQMINAT
WIGKVWKIGIERDDKWDRSTVEKMIKELMEGEKGAEIQRSMKEFSKLANDKVVKGGTSFENLELIV
EYLKKLKPSN

>XP_022151546.1 7-deoxyloganetic acid glucosyltransferase-like
25 [*Momordica charantia*]

(SEQ ID NO:26)

MAQPRVLLFFFPAMGHVVKPFLSLAELLSDAGVEVVFLSTEYNHRRIPDIGALAARFPTLHFETIPD
GLPPDQPRVLADGHLYFSMLDGTKPRFRQLIQSLNGNPRPITCIINDVMLSSPIEVAEEFGIPVIA
FCPCSARFLSVHFFMPNFIEEAQIPYTDENPMGKIEEATVFEGLLRKDLPLGLWCAKSSNISFSHR
30 FINQTIAAGRASALILNTFDELESPFLNHLSSIFPKIYCIGPLNALSRSLGKSSSSSSSALAGFWK
EDQAYMSWLESQPPRSVIFVSFGSTMKMEAWKLAEFWYGLVNSGSPFLFVFRPDCVINSGDAAEVM
EGRGRGMVVEWASQEKVLAHPAVGGFLTHCGWNSTVESIVAGVPMCCPIVADQLSNATWIHKVWK
IGIEGDEKWDRSTVEMMIKELMESQKGTEIRTSIEMLSKLANEKVVKGGTSLNNFELLVEDIKTLR
RPYT

35

>XP_022151514.1 7-deoxyloganetic acid glucosyltransferase-like
[*Momordica charantia*]

(SEQ ID NO:27)

- MEQSDSNSDDHQHHVLLFFPPAKGHIKPFCLQAQLLCGAGLQVTF LN TDH NHRRIDDRHRRL LATQ
 FPM LHF KSI SDGLPPDHPRDLLDGKLIASMRV TESLFRQLLLSYNGYNGGTNNVSN SGRPPISC
 VITDVIFSF PVEVAEELGIPVFSFATFSARFLFLYFWIPKLIQEGQLPFPDGKTNQELYGVPGAEG
 IIRCKDLPGSWSVEAVAKNDPMNFVKQTLASSRSSGLILNTFEDLEAPFVTHLSNTFDKIYTIGPI
 5 HSL LGTSHCGLWKEDYACLA WLDARPRKSVV FVSFGSLVKTTSRELMELWHGLVSSGKS FLLV LRS
 DVVEGEDEEQVVKEILESNGEGKWL VVGWAPQEEVLAHEAIGGFLTHSGWNSTMESIAAGVPMVCW
 PKIGDQPSNCTWVSRVWKVGLEMEERYDRSTVARMARSMMEQEGKEMERRIAELAKRVKYRVGKDG
 ESYRNLES LIRDIKITKSSN
- 10 >XP_004147933.2 PREDICTED: 7-deoxyloganetic acid
 glucosyltransferase-like [*Cucumis sativus*]
 (SEQ ID NO:28)
- 15 MGLSPTD HVLLFFPPAKGHIKPFCLAHLLCNAGLRVTF LSTEHHHQKLHNLTHLAAQIPSLHFQS
 ISDGLSLDHP RNLLDGQLFKSMPQVTKPLFRQLLLSYKDGTS PITCVITDLILRFPMDVAQELDIP
 VFCFSTFSARFLFLYFSIPK LLEDGQIPYPEGNSNQVLHGIPGAEGLLRCKDLPGYWSVEAVANYN
 PMNFVNQTIATSKSHGLILNTFDELEVPPFITNLSKIYKKVYTIGPIHSL LKKSQVTQYEFWKEDHS
 CLAWLDSQPPRSVMFVSFGSIVKLKSSQLKEFWNGLVDSGKAFLLVLRSDALVEETGEDEKQKEL
 VIKEIMETKEEGRWVIVNWAPQEKVLEHKAIGGFLTHSGWNSTLESVAVGVPMVSWPQIGDQPSNA
 TWLSKVWKIGVEMEDSYDRSTVESKVR SIMEHEDKKMENAIVELAKRVDDRVSKEGTSYQNLQRLI
 20 EDIEGFKLN
- >XP_022978164.1 7-deoxyloganetic acid glucosyltransferase-like
 [*Cucurbita maxima*]
 (SEQ ID NO:29)
- 25 MELSHTHHVLLFFPPAKGHIKPFSLAQLLCNAGLRVTF LN TDH HHHRIHDLNRLAAQLPTLHFDS
 VSDGLPPDEPRNVFDGKLYESIRQVTSSLFRELLVSYNNGTSSGRPPITCVITDVMFRFPIDIAEE
 LGIPVFTFSTFSARFLFLIFWIPK LLEDGQLRYPEQELHGVPGAEG LIRWKDLPGFWSVEDVADWD
 PMNFVNQTLATSRSSGLILNTFDELEAPFLTSLSKIYKKIYSLGPINSLLKNFQSQPQYNLWKEDH
 SCMAWLDSQPRKSVV FVSFGSVVKLT SRQLMEFWNGLVNSGMPFLLVLRSDVIEAGEEVVREIMER
 30 KAEGRWVIVSWAPQEEVLAHDAVGGFLTHSGWNSTLES LAAGVPMISWPQIGDQTSNSTWISKVWR
 IGLQLEDGFDSS TIETMVR S IMDQTMEKTVAELAERAKNRASKNGTSYRN FQTLIQDITNIIETHI
- >XP_022950128.1 7-deoxyloganetic acid glucosyltransferase-like
 [*Cucurbita moschata*]
 35 (SEQ ID NO:30)
- MELSPTHHLLLFFPPAKGHIKPFSLAQLLCNAGARVTF LN TDH HHHRIHDLRLAAQLPTLHFDS
 VSDGLPPDESRNVFDGKLYESIRQVTSSLFRELLVSYNNGTSSGRPPITCVITDCMFRFPIDIAEE
 LGIPVFTFSTFSARFLFLFFWIPK LLEDGQLRYPEQELHGVPGAEG LIRCKDLPGFLSDEDVAHWK
 PINFVNQILATSRSSGLILNTFDELEAPFLTSLSKIYKKIYSLGPINSLLKNFQSQPQYNLWKEDH
 40 SCMAWLDSQPPKSVV FVSFGSVVKLTNRQLVEFWNGLVNSGKPFLLVLRSDVIEAGEEVVRENMER

KAEGRWMIWSWAPQEEVLAHDAVGGFLTHSGWNSTLES LAAGVPMISWTQIGDQTSNSTWVSKVWR
IGLQLEDGFDSTIETMVRVMDQTMKTVAE LAERAKNRASKNGTSYRNFQTLIQDITNIIETHI

5 >XP_020422423.1 7-deoxyloganetic acid glucosyltransferase [*Prunus persica*]

(SEQ ID NO:31)

MAMKQPHVIIFFPFLQGHMKPLLCLAE LLCHAGLHVTVNTHHHQRLANRQALSTHFPTLHFESI
SDGLPEDDPRTLNSQLLIALKTSIRPHFRELLKTISLKAESNDTLVPPPPSCIMTDGLVTFADFVAE
ELGLPILSFNVPCPRYLWTCLCLPKLIENGQLPFQDDDMNVEITGVPGMEGLLHRQDLPGFCRVKQ
10 ADHPSLQFAINETQTLKRASALILDTVYELDAPCISHMALMFPKIYTLGPLHALLNSQIGDMSRGL
ASHGSLWKSDLNMTWLD SQPSKSIIYVSFGTLVHLTRAQVIEFWYGLVNSGHPFLWVMRSDITSG
DHQIPAELENGTKERGCIVDWVSQEEVLAHKS VGGFLTHSGWNSTLESIVAGLPMICWPKLGDHYI
ISSTVCRQWKIGLQLNENCDRSNIESMVQTLMGSKREEIQSSMDAISKL SRDSVAEGGSSHNNLEQ
LIEYIRNLQHQN

15

>EOY07351.1 UDP-glucosyl transferase 85A3, putative [*Theobroma cacao*]

(SEQ ID NO:32)

MRQPHVLVLPFPAQGH IKPMLCLAE LLCCAGLRVTF LNTHHSHRRLNNLQDLSTRFPTLHFESVSD
20 GLPEDHPRNLVHFMHLVHSIKNVTKPLLRDLLTSLSLKTDIPPVSCIIADGILSFAIDVAEELQIK
VIIIFRTISSCCLWSYLCVPKLIQQGELQFSDSDMGQKVSSVPEMKGSLRLHDRPYSFGLKQLEDPN
FQFFVSETQAMTRASAVIFNTFDSLEAPVLSQMIPLLPKVYTIGPLHALRKARLGDL SQHSSFN GN
LREADHNCITWLD SQPLRSVVYVSFGSHVLTSEELLEFWHGLVNSGKRFLWVLRPDI IAGEKDHN
QIIAREPDLGTKEKGLLDVWAPQEEVLAHPSVGGFLTHCGWNSTLES MVAGVPMLCWPKL PDQLVN
25 SSCVSEVWKIGLDL KDMCDRSTVEKMVRALMEDRREEVMRSVDGISKLA RESVSHGGSSSSNLEML
IQELET

>XP_022155979.1 beta-D-glucosyl crocetin beta-1,6-
glucosyltransferase-like [*Momordica charantia*]

30 (SEQ ID NO:33)

MDAHQQAEHTTTILMLPWVG YGHLTAYLELAKALSRRNFHIYYCSTPVNIESIKPKLTIPCSSIQF
VELHLPSSDDLPPNLHTTNGLP SHLMPTLHQAFSAAAPLFEEILQTLCPHLLIYDSLQPWAPKIAS
SLKIPALNFNTSGVSVIAQALHAIHHPDSKFPLSDFILHNYWKSTYTTADGGASEKTRRAREAFLY
CLNSSGNAILINTFRELEGEYIDYLSLLL NKKVIPIGPLVYEPNQDEDQDEEYRSIKNWL DKKEPC
35 STVFVSFGSEYFPSNEEMEEIAPGLEESGANFIWVVRFPKLENRNGIIEEGLLERAGERGMV KEW
APQARILRHGSIGGFVSHCGWNSVMESIICGVPVIGVPMRVDQPYNAGLVEEAGVGVEAKRDPDGK
IQRHEVSKLIKQVVVEKTRDDVRKKVAQMSEILRRKGDEKIDEMVALISLLPKG

>XP_022986080.1 beta-D-glucosyl crocetin beta-1,6-
40 glucosyltransferase-like [*Cucurbita maxima*]

(SEQ ID NO:34)

MDAQKAVDTPPTTVLMLPWIGYGHL SAYLELAKALSRRNFHVYFCSTPVNLDSIKPNLI PPPPSIQ
FVDLHLPSSPELPPHLHTTNGLP SHLKPTLHQAFSAAQHF EAILQTLSPHLLIYDSLQPWAPRIA
SSLNIPAINFNTTAVSIIAHALHSVHY PDSKFPFSD FVLHDYWKAKYTTADGATSEKIRRGAE AFL
5 YCLNASCDVVLVNSFRELEGEYMDYLSVLLKKKVSVG PLVYEPSEGEDEEYWR IKKWLDEKEAL
STVLVSFGSEYFPSKEEMEEIAHGLEE SEANFIWVVRFPKGEE SCRGIEEALPKG FVERAGERAMV
VKKWAPQGKILKHGSIGGFVSHCGWNSVLESIRFGVPVIGVPMHLDQPYNAGLLEEAGIGVEAKRD
ADGKIQRDQVASLIKRVVVEKTREDIWKTVREMREVLRRRDDDMIDEMVAEISVVLKI

10 >XP_022156002.1 beta-D-glucosyl crocetin beta-1,6-
glucosyltransferase-like [*Momordica charantia*]

(SEQ ID NO:35)

MDARQQA EHTTTILMLPWVG YGHL SAYLELAKALSRRNFHIYYCSTPVNIESIKPKLTIPCSSIQF
VELHLPFSDDLPPNLHTTNGLP SHLMPALHQAFSAAAPLFEAILQTLCPHLLIYDSLQPWAPQIAS
15 SLKIPALNFNTTGVSVIARALHTIHH PDSKFP LSEIVLHNYWKATHATADGANPEKFRDLEALLC
CLHSSCNAILINTFRELEGEYIDYLSLLLNKKVTPIGPLVYEPNQDEEQDEEYRSIKNWL DKKEPY
STIFVSFGSEYFPSNEEMEEIARGLEESGANFIWVVRFHKL ENGNGITEEGLLERAGERGMVIQGW
APQARILRHGSIGGFVSHCGWNSVMESIICGVPVIGVPMGLDQPYNAGLVEEAGVGVEAKRDPDGK
IQRHEVSKLIKQVVVEKTRDDVRKKV AQMSEILRRKGDEKIDEMVALISLLLKG

20

>XP_022943327.1 beta-D-glucosyl crocetin beta-1,6-
glucosyltransferase-like [*Cucurbita moschata*]

(SEQ ID NO:36)

MDAQKAVDTPPTTVLMLPWIGYGHL SAYLELAKALSRRNFHVYFCSTPVNLDSIKPNLI PPPPSIQ
25 FVDLHLPSSPELPPHLHTTNGLP SHLKPTLHQAFSAAQHF EAILQTLSPHLLIYDSLQPWAPRIA
SSLNIPAINFNTTAVSIIAHALHSVHY PDSKFPFSD FVLHDYWKAKYTTADGATSEKTRRGVE AFL
YCLNASCDVVLVNSFRELEGEYMDYLSVLLKKKVSVG PLVYEPSEGEDEEYWR IKKWLDEKEAL
STVLVSFGSEYFPPKEEMEEIAHGLEE SEANFIWVVRFPKGEESSRGIEEALPKG FVERAGERAM
VVKKWAPQGKILKHGSIGGFVSHCGWNSVLESIRFGVPVIGAPMHL DQPYNAGLLEEAGIGVEAKR
30 DADGKIQRDQVASLIKQVVVEKTREDIWKKVREMREVLRRRDDDDMMIDEMVAVISVVLKI

>XP_022996307.1 beta-D-glucosyl crocetin beta-1,6-
glucosyltransferase-like [*Cucurbita maxima*]

(SEQ ID NO:37)

MSSNLFLKISIPFGRRLRDSALNCSVFHCKLHLAIAIAMDAQQAANKSPTATTIFMLPWAGYGHLSA
YLELAKALSTRNFHIYFCSTPVSLASIKPRLIPSCSSIQFVELHLPSSDEFPPHLHTTNGLP SRLV
PTFHQAFSEAAQTFEAFLQTLRPHLLIYDSLQPWAPRIASSLNIPAINFFTAGAFVSHVLR AFHY
PDSQFPSSDFVLHSRWKIKNTTAE SPTQAKLPKIGE AIGYCLNASRGVILTNSFRELEGKYIDYLS
VILKKRVFPIGPLVYQPNQDEEDEDYSRIKNWLD RKEASSTVLVSFGSEFFLSKEETE AIAHGLEQ
40 SEANFIWGIRFPKGAKKNAIEEALPEGFLERAGGRAMVVEEWV PQGKILKHGSIGGFVSHCGWNSA

MESIVCGVPIIGIPMQVDQPFNAGILEEAGVGVEAKRDSGKIQRDEVAKLIKEVVVERTREDIRN
KLEKINEILRSRREEKLDELATEISLLSRN

5 >XP_022957664.1 beta-D-glucosyl crocetin beta-1,6-
glucosyltransferase-like [*Cucurbita moschata*]
(SEQ ID NO:38)

MDAQQAAANKSPTASTIFMLPWVGYGHL SAYLELAKALSTRNFHVYFCSTPVSLASIKPRLIPSCSS
IQFVELHLPSSDEFPPHLHTTNGLPAHLVPTIHQAFAAAAQTFEAFLQTLRPHLLIYDSLQPWAPR
IASSLNIPAINFFTAGAFVSHVLR AFHY PDSQFPSSDFVLHSRWKIKNTTAESPTQVKIPKIGEA
10 IGYCLNASRGVILTNSFRELEGKYIDYLSVILKKRVLP IGPPLVYQPNQDEEDEDYSRIKNWLD RKE
ASSTVLVSFGSEFFLSKEETEIAHGLEQSEANFIWGIRFPKGAKKNAIEEALPEGFLERVGGRAM
VVEEWVPQGKILKHGNI GG FVSHCGWNSAMESIMCGVPVIGIPMQVDQPFNAGILEEAGVGVEAKR
DSDGKIQRDEVAKLIKEVVVERTREDIRNKLEEINEILRTRREEKLDELATEISLLCKN

15 >OMO57892.1 UDP-glucuronosyl/UDP-glucosyltransferase [*Corchorus*
capsularis]
(SEQ ID NO:39)

MDSKQKMSVLMFPWLA YGHISPFLELAKKLSKRNFHTFFFSTPINLSIKSKLSPKYAQSIQFVE
LHLP SLPDLP PHYHTTNGLP PHLMTLKKAFDMSSLQFSKILKTLNPDLLVYDFIQPWAPLLALSN
20 KIPAVHFLCTSAAMSSFSVHAFKKPCEDFPFPNIYVHGFMNAKFN NMENCSSDDSI SDQDRVLQC
FERSTKIILVKT FEELEGKFMDYLSVLLNKKIVPTGPLTQDPNEDEGDDDDERTKLLLEWLNKKS
STVFVSFGSEYFLSKEEREEIAYGLELSKVNFIWVIRFPLGENKTNLEEALPQGFLQRVSEGLVV
ENWAPQAKILQHSSIGGFVSHCGWSSVMESLKFVPIIAIPMHLDDQPLNARLVVDVGVGLEVIRNH
GSLEREEIAKLIKEVVLGNGNDGEIVRRKAREMSNHIKKKGKMDDELVEELMLICKMKPNSCHLS
25

>XP_015886141.1 PREDICTED: beta-D-glucosyl crocetin beta-1,6-
glucosyltransferase-like isoform X1 [*Ziziphus jujuba*]
(SEQ ID NO:40)

MMERQRSIKVLMFPWLAHGHISPFLELAKRLTDRNFQIYFCSTPVNLT SVKPKLSQKYSSSIKLVE
30 LHLP SLPDLP PHYHTTNGLALNLIPTLKKAFDMSSSFSTILSTIKPDLLIYDFLQPWAPQLASCM
NIPAVNFLSAGASMVSFVLHSIKYNGDDHDEFLTTELHLSDSMEAKFAEMTESSPDEHIDRAVTC
LERSNSLILIKSFRELEGKYLDYLSLSFAKKVVP IGPPLVAQDTNPEDDSMDIINWLDKKEKSSTVF
VSFGSEYYLTNEEMEEIAYGLELSKVNFIWVVRFP LGQKMAVEEALPKGFLERVGEKGMVVEDWAP
QMILGHSSIGGFVSHCGWSSLMESLKLGVPIIAMPMLDQPINAKLVERSGVGLVVKRDKNRIE
35 REYLAKVIREIVVEKARQDIEKKAREMSNIITEKGEEIDNVVEELAKLCGM

>XP_002271587.3 PREDICTED: beta-D-glucosyl crocetin beta-1,6-
glucosyltransferase-like [*Vitis vinifera*]
(SEQ ID NO:41)

- MDARQSDGISVLMFPWLAHGHI SPFLQLAKKLSKRNF SIYFCSTPVNLDPIKGKLSSESYLSIQLV
KLHLPSLPPELPPQYHTTNGLPPLMPTLKM AFDMASPNFSN ILKTLHPDLLIYDFLQWPAPAAASS
LNIPAVQFLSTGATLQSFLAHRHRKPGIEFFQEIHL PDYEIGRLNRFLEPSAGRISDRDRANQCL
ERSSRFSLIKTFREIEAKYLDYVSDLTKKKMVTVG PLLQDPEDEDEATDIVEWLNKKCEASAVFVS
5 FGSEYFVSKEEMEEIAHGLELSNVDFIWVVRFP MGEKIRLEDALPPGFLHRLGDRGMVVEGWAPQR
KILGHSSIGGFVSHCGWSSVMEGMKFGVPIIAMP MHLDPINAKLVEAVGVGREVKRDENRKLERE
EIAKVIKEVVGEKNGENVRRKARELSETLRKKGDEE IDVVVEELKQLCSY
- >XP_018840205.1 PREDICTED: beta-D-glucosyl crocetin beta-1,6-
10 glucosyltransferase-like [*Juglans regia*]
(SEQ ID NO:42)
- MDTARKRIRVVMLPWLAHGHI SPFLELSKKLAKRNF HIYFCSTPVNLSSIKPKLSGKYSRSIQLVE
LHLPSLPPELPPQYHTTKGLPPLNATLKRAFD MAGPHFSN ILKTLSPDLLIYDFLQWPAPAI AASQ
NIPAINFLSTGAAMTSFVLHAMKKPGDEF PFPEIHLDECMKTRFVDLPEDHSPSDDHNHISDKDRA
15 LKCFERSSGFVMMKTFEELEGKYINFLSHLMQ KKI VPGPLVQNPVRGDHEKAKTLEWLDKRKQSS
AVFVSFGTEYFLSKEEMEEIAYGLELSNVNFIW VVRFPGEKVKLEEALPEGFLQRVGEKGMVVEG
WAPQAKILMHPSIGGFVSHCGWSSVMESIDFGV PIVAI PMQLDQPVNAKVVEQAGVGVEVKRDRDG
KL EREEVATVIREVVMGNIGESVRKKEREMRDN IRKKGEEKMDGVAQELVQLYGNGIKNV
- >XP_021652171.1 beta-D-glucosyl crocetin beta-1,6-
20 glucosyltransferase-like [*Hevea brasiliensis*]
(SEQ ID NO:43)
- METLQRRKISVLMFPWLAHGHL SPFLELSKKLNKRNF HVYFCSTPVNLDSIKPKLSAEYSFSIQLV
ELHLPSPELPLHYHTTNGLPPLMKNLKNAFDMA SSSFFN ILKTLKPDLLIYDFIQWPAPALASS
25 LNIPAVNFLCTSMAMSCFGLHLNNQEAKFPF PG IYPRDYMRMKVFGALESSSNDIKDGERAGRCMD
QSFHLILAKTFRELEGKYIDYLSVKLMKKI VPGPLVQDP I FEDDEKIMDHHQVIKWLEKKERLST
VFVSFGTEYFLSTEEMEEIAYGLELSKAHFIW VVRFP TGEKINLEESLPKRYLERVQERGKIVEGW
APQQKILRHSSIGGFVSHCGWSSIMESMKFGVPI IAMP MNLDQPVNSRIVEDAGVGIEVRNKSGE
LEREEIAKTIRKVVVEKDGNVSRKAREMSD TIRKKGEEEIDGVVDELLQLCDVKTNYLQ
- >XP_021619073.1 beta-D-glucosyl crocetin beta-1,6-
30 glucosyltransferase-like [*Manihot esculenta*]
(SEQ ID NO:44)
- MATAQTRKISVLMFPWLAHGHL SPFLELSKKLANRNF HVYFCSTPVNLDSIKPKLSPEYHFSIQFV
35 ELHLPSPELPSHYHTTNGLPPLMKTLKKA FDMAS SSSFFN ILKTLNPDLLIYDFLQWPAPALASS
LNIPAVNFLCSSMAMSCFGLNLNKNKEIKFLF PEIYPRDYMEMKLF RVFESSNQIKDGERAGRCI
DQSFHVILAKTFRELEGKYIDYVSVKCNKKI VPGPLVEDTIHEDDEKTMDDHHHHHHDEVIKWLEK
KERSTTVFVSFGSEYFLSKEEMEEIAHGLELSKV NFIWVVRFPKG EKINLEESLPEGYLERIQERG
KIVEGWAPQRKILGHSSIGGFVSHCGWSSIMES MKLGVPIIAMP MNLDQPINSRIVEAAGVGIEVS
40 RNQSGELEREEMA KTIRKVVVEREGVYVRRKAREMSD VLRKKGEEEIDGVVDELVLQCDMKTNYL

>GAV83746.1 UDPGT domain-containing protein [*Cephalotus follicularis*]

(SEQ ID NO:45)

5 MDLKRRSIRVLMPLWLAHGHISPFLELAKKLTNRNFLIYFCSTPINLNSIKPKLSSKYSFSIQLVE
LHLPSPLELPPIHYHTTNGLPPLHLMNTLKTAFDMASPSFLNLIKTLKPDLLICDHLQPWAPSLASSL
NIPAIIFPTNSAIMMAFSLHHAKNPGEEFPFPSININDDMVKSINFLHSASNGLTDMDRVLCCLER
SSNTMLLKTFRQLEAKYVDYSSALLKKKIVLAGPLVQVPDNEDEKIEI IKWLDSRGQSSTVFVSFG
SEYFLSKEEREDIAHGLELSKVNFIWVVRFPVGEKVKLEEALPNGFAERIGERGLVVEGWAPQAMI
10 LSHSSIGGFVSHCGWSSMMESMKFGVPPIIAMPMHIDQPLNARLVEDVGVGLEIKRNKDGREFEREEL
ARVIKEVLVYKNGDAVRSKAREMSEHIKKNQDQIDGVADALVKLCCKMTNSLNQD

>*Coffea Arabica* UGT_1,6

(SEQ ID NO:46)

15 MENHATFNVLMPLWLAHGHVSPYLELAKKLTARNFNVYLCSSPATLSSVRSKLTEKFSQSIHLVEL
HLPKLPPELPAEYHTTNGLPPLHLMPTLKDAFDMAKPNFCNVLSLKPDLIIYDLLQPWAPSAASAFN
IPAVVFISSSATMTSFGHLHFFKNPGTKYPYGNAI FYRDYESVFVENLTRDRDRTYRVINCMERSSK
IILIKGFNEIEGKYFDYFSCLTGKKVVPVGPLVQDPVLDDDCRIMQWLNKKEKGSTVFVSFGSEY
FLSKKDMEEIAHGLEVSNNVDFIWVVRFPKGENIVIEETLPKGFFERVGERGLVVNGWAPQAKILTH
20 PNVGGFVSHCGWNSVMESMKFGLPIIAMPMHLDQPINARLIEEVGAGVEVLRDSKGLHRERMAET
INKVMKEASGESVRKKARELQEKLELKGDEEIDDVVKELVQLCATKNKRNLHYY

>Circular permutant linker sequence

(SEQ ID NO:47)

25 GSGGSG

>Circular permutant linker sequence

(SEQ ID NO:48)

GSGGSGGSG

30

CLAIMS

1. A method for producing a glycosylated product from glycoside intermediates, comprising:
providing a microbial cell expressing one or more UDP-dependent glycosyl
5 transferase enzymes intracellularly,
incubating the microbial strain with the glycoside intermediates, and
recovering the glycosylated product.
2. The method of claim 1, wherein the glycoside intermediates are provided as a plant
10 extract or fraction thereof.
3. The method of claim 1, wherein the glycoside intermediate is a glycosylated
secondary metabolite, optionally selected from a glycosylated terpenoid, a glycosylated
flavonoid, a glycosylated cannabinoid, a glycosylated polyketide, a glycosylated stilbenoid,
15 and a glycosylated polyphenol.
4. The method of claim 3, wherein the glycoside intermediate is a glycosylated
terpenoid.
- 20 5. The method of claim 4, wherein the glycoside intermediate is steviol glycoside or
mogroside.
6. The method of any one of claims 1, wherein the glycoside intermediate has one, two,
three, four, or five, glycosyl groups.
25
7. The method of claim 6, wherein the glycosyl groups are independently selected from
glucosyl, galactosyl, mannosyl, xylosyl, and rhamnosyl groups.
8. The method of claim 6, wherein the glycosylated product has at least five glycosyl
30 groups, or at least six glycosyl groups.

9. The method of claim 8, wherein the glycosylated product has seven, eight, or nine glycosyl groups.
10. The method of claim 8, wherein biosynthesis of the product involves at least two glycosylation reactions of the intermediate in the microbial cell.
11. The method of any one of claims 1, wherein the glycoside intermediates are steviol glycosides from stevia leaf extract.
- 10 12. The method of claim 11, wherein the steviol glycoside intermediates comprise one or more of stevioside, steviolbioside, rebaudioside A, dulcoside A, dulcoside B, rebaudioside C, and rebaudioside F.
13. The method of claim 12, wherein the extract comprises stevioside, steviolbioside, and Rebaudioside A as prominent components.
- 15 14. The method of claim 13, wherein the glycosylated product comprises RebM.
15. The method of claim 12, wherein the glycosylated product comprises RebK, RebC+1, and/or RebC+2.
- 20 16. The method of any one of claims 1 to 15, wherein the microbial cell is a bacterial cell.
17. The method of claim 16, wherein the bacterial cell is *Escherichia* spp., *Bacillus* spp., *Rhodobacter* spp., *Zymomonas* spp., or *Pseudomonas* spp.
- 25 18. The method of claim 17, wherein the bacterial cell is *Escherichia coli*, *Bacillus subtilis*, *Rhodobacter capsulatus*, *Rhodobacter sphaeroides*, *Zymomonas mobilis*, or *Pseudomonas putida*.
19. The method of claim 18, wherein the bacterial cell is *E. coli*.

20. The method of claim 19, wherein the bacterial cell comprises the genetic modifications: *ushA* and *galETKM* are deleted, inactivated, or reduced in expression; *pgi* is deleted, inactivated, or reduced in expression; and *pgm* and *galU* are overexpressed.
21. The method of any one of claims 1 to 15, wherein the microbial cell is a yeast cell.
- 5 22. The method of claim 21, wherein the yeast is *Saccharomyces spp.*, *Schizosaccharomyces spp.*, *Pichia spp.*, *Phaffia spp.*, *Kluyveromyces spp.*, *Candida spp.*, *Talaromyces spp.*, *Brettanomyces spp.*, *Pachysolen spp.*, *Debaryomyces spp.*, or *Yarrowia spp.*
- 10 23. The method of claim 22, wherein the yeast is *Saccharomyces cerevisiae*, *Pichia pastoris*, or *Yarrowia lipolytica*.
24. The method of any one of claims 1 to 23, wherein the microbial cell expresses one or more UGT enzymes comprising the amino acid sequence of any one of SEQ ID NOS:1 to 45, or a derivative thereof.
- 15 25. The method of claim 24, wherein the microbial cell expresses a 1-3' glycosylating UGT enzyme.
26. The method of claim 25, wherein the 1-3' glycosylating UGT enzyme is selected from SrUGT76G1, MbUGT1-3, MbUGT1-3_1, MbUGT1-3_1.5, MbUGT1-3_2, and derivatives thereof.
- 20 27. The method of claim 26, wherein the 1-3' glycosylating UGT enzyme is SrUGT76G1 comprising a L200A substitution.
28. The method of claim 26, wherein the 1-3' glycosylating UGT enzyme comprises an amino acid sequence that is at least 70% identical to the amino acid sequence of SEQ ID NO:15, SEQ ID NO:16, or SEQ ID NO:17.
- 25 29. The method of any one of claims 1 to 28, wherein the microbial cell expresses a 1-2' glycosylating UGT enzyme.

30. The method of claim 29, wherein the 1-2' glycosylating UGT enzyme is selected from SrUGT91D2, SrUGT91D1, SrUGT91D2e, OsUGT1-2, MbUGT1,2, MbUGT1,2,2, and derivatives thereof.
31. The method of any one of claims 1 to 30, wherein the microbial cell expresses a C13 O-glycosylating UGT enzyme.
32. The method of claim 31, wherein the C13 O-glycosylating UGT enzyme is SrUGT85C2 or a derivative thereof, optionally comprising a P215T substitution.
33. The method of any one of claims 1 to 32, wherein the microbial cell expresses a C19 O-glycosylating UGT enzyme.
34. The method of claim 33, wherein the C19 O-glycosylating UGT enzyme is SrUGT74G1, MbUGTc19, or a derivative thereof.
35. The method of claim 24, wherein the microbial cell expresses a 1-3' glycosylating UGT enzyme and a 1-2' glycosylating UGT enzyme, and optionally a C13 O-glycosylating UGT enzyme and a C19 O-glycosylating enzyme.
36. The method of claim 35, wherein the microbial cell expresses a 1-3' glycosylating UGT enzyme, a 1-2' glycosylating UGT enzyme, a C19 O-glycosylating UGT enzyme, and a C13 O-glycosylating UGT enzyme.
37. The method of claim 36, wherein the microbial cell expresses: a SrUGT85C2 or derivative thereof; MbUGT1,2 or derivative thereof; SrUGT74G1 or derivative thereof; and an enzyme selected from SrUGT76G1 or derivative thereof, MbUGT1-3_1 or derivative thereof, MbUGT1-3_1.5 or derivative thereof, and MbUGT1-3_2 or derivative thereof.
38. The method of any one of claims 1 to 37, wherein the UGT enzymes are integrated into the chromosome of the microbial cell.
39. The method of any one of the claims 1 to 37, wherein the UGT enzymes are expressed extrachromosomally.

40. The method of any one of claims 1 to 39, wherein the microbial cell has one or more genetic modification that increase UDP-glucose availability.
41. The method of claim 40, wherein the microbial cell has a deletion, inactivation, or reduced activity or expression of a gene encoding an enzyme that consumes UDP-glucose.
- 5 42. The method of claim 41, wherein the microbial cell has a deletion, inactivation, or reduced activity of *ushA* or ortholog thereof and/or one or more of *galE*, *galT*, *galK*, and *galM*, or ortholog.
43. The method of claim 42, wherein *galETKM* genes are inactivated, deleted, or reduced in expression.
- 10 44. The method of claim 40, wherein the microbial cell has a deletion, inactivation, or reduced activity or expression of *otsA* (trehalose-6-phosphate synthase) or ortholog thereof.
45. The method of claim 40, wherein the microbial cell has a deletion, inactivation, or reduced activity or expression of *ugd* (UDP-glucose 6-dehydrogenase) or ortholog thereof.
46. The method of claim 40, wherein the microbial cell has a deletion, inactivation, or
15 reduced expression of *waaG*, *waaO*, *waaJ*, *yfdG*, *yfdI*, and/or *wcaJ*, or ortholog(s) thereof.
47. The method of claim 40, wherein the microbial cell has a deletion, inactivation, or reduced activity or expression of *pgi* (glucose-6 phosphate isomerase) or ortholog thereof.
48. The method of claim 40, wherein the microbial cell has an overexpression or increased activity of one or more genes encoding an enzyme involved in converting glucose-
20 6-phosphate to UDP-glucose, which is optionally *pgm* (phosphoglucomutase) and/or *galU* (UTP-glucose-1-phosphate uridylyltransferase), or ortholog or derivative thereof.
49. The method of claim 40, wherein the microbial cell has an overexpression or increased activity or expression of *ycjU* (β -phosphoglucomutase) and *Bifidobacterium bifidum* *ugpA*, or ortholog or derivative thereof.

50. The method of claim 40, wherein the microbial cell has one or more genetic modifications that increase flux to the pentose phosphate pathway (PPP), and which is optionally an overexpression or activity of *E. coli* zwf or homologue or engineered derivative thereof.
- 5 51. The method of claim 40, wherein the microbial cell has one or more genetic modifications that increase glucose transport, and which optionally includes increased expression or activity of *E. coli* galP and *E. coli* glk (glucokinase), or alternatively *Zymomonas mobilis* glf and *E. coli* glk, or homologs, orthologs, or engineered derivatives of thereof.
- 10 52. The method of claim 40, wherein the microbial cell has one or more genetic modifications that increase UTP production and recycling, and which optionally include increased expression or activity of *E. coli* pyrH (UMP kinase), *E. coli* cmk (cytidylate kinase), *E. coli* adk (adenylate kinase), or *E. coli* ndk (nucleoside diphosphate kinase), or homologs, orthologs, or engineered derivatives of these enzymes.
- 15 53. The method of claim 40, wherein the microbial cell has one or more genetic modifications that increase UDP production, and which optionally include: overexpression or increased activity of upp (uracil phosphoribosyltransferase), pyrH (UMP kinase), and cmk (cytidylate kinase), or homolog or engineered derivatives thereof; or overexpression or increased activity of dctA (C4 dicarboxylate/orotate:H⁺ symporter), pyrE (orotate phosphoribosyltransferase), pyrH (UMP kinase) and cmk (cytidylate kinase), or homolog or
20 engineered derivatives thereof.
54. The method of claim 40, wherein the microbial cell has one or more genetic modifications to remove or reduce regulation of glucose uptake, which optionally include deletion, inactivation, or reduced expression of sgrS small regulatory RNA.
- 25 55. The method of claim 40, wherein the microbial cell has one or more genetic modifications that reduce dephosphorylation of glucose-1-phosphate, which optionally include deletion, inactivation, or reduced expression or activity of one or more of *E. coli* agp (glucose-1-phosphatase), *E. coli* yihX (α -D-glucose-1-phosphate phosphatase), *E. coli* ybiV

(sugar phosphatase), *E. coli* yidA (sugar phosphatase), *E. coli* yigL (phosphosugar phosphatase), and *E. coli* phoA (alkaline phosphatase), or an ortholog thereof.

56. The method of claim 40, wherein the microbial cell has one or more genetic modifications that reduce conversion of glucose-1-phosphate to TDP-glucose, which
5 optionally include a deletion, inactivation, or reduced expression or activity of one or more of *E. coli* rffH (dTDP-glucose pyrophosphorylase) and *E. coli* rfbA (dTDP glucose pyrophosphorylase), or an ortholog thereof.

57. The method of claim 40, wherein the microbial cell has one or more genetic modifications that reduce conversion of glucose-1-phosphate to ADP-glucose, and which
10 optionally include deletion, inactivation, or reduced expression or activity of *E. coli* gigC (glucose-1-phosphate adenylyltransferase), or an ortholog thereof.

58. The method of claim 40, wherein the microbial cell is a bacterial cell comprising the genetic modifications: *ushA* and *galETKM* are deleted, inactivated, or reduced in expression; *pgi* is deleted, inactivated, or reduced in expression; and *pgm* and *galU* are
15 overexpressed.

59. The method of any one of claims 1 to 58, wherein the microbial cell expresses an endogenous transport protein that transports the glycoside intermediate(s) into the cell.

60. The method of any one of claims 1 to 58, wherein the microbial cell overexpresses a recombinant transport protein that transports the glycoside intermediate(s) into the cell.

20 61. The method of any one of claims 59 or 60, wherein the microbial strain expresses a transport protein selected from *E. coli* *acrAD*, *xylE*, *ascF*, *bglF*, *chbA*, *ptsG/crr*, *wzxE*, *rfbX*, or an ortholog or derivative thereof.

62. The method of any one of claims 1 to 61, wherein the microbial cell expresses an endogenous or recombinant transporter that transports glycosylated product outside the cell.

25 63. The method of any one of claims 1 to 62, wherein the method results in at least 40% conversion of glycoside intermediate to glycosylated product by weight.

64. The method of claim 63, wherein the method results in a least 50% conversion of glycoside intermediate to glycosylated product by weight.
65. The method of claim 64, wherein the method results in at least 75% conversion of glycoside intermediate to glycosylated product by weight.
- 5 66. The method of any one of claims 1 to 65, wherein a ratio of RebM to RebD of at least 5:1 is recovered as the glycosylated product.
67. The method of any one of claim 1 to 66, wherein the method is performed by batch fermentation, continuous fermentation, or semi-continuous fermentation.
68. The method of claim 67, wherein the method is performed by fed batch fermentation.
- 10 69. The method of claim 68, wherein the glycoside intermediates are incubated with the cell for about 72 hours or less.
70. A microbial cell comprising:
one or more recombinant genes encoding UDP-dependent glycosyl transferase (UGT) enzyme(s) for intracellular expression; and
15 one or more genetic modifications that increase UDP-glucose availability.
71. The microbial cell of claim 70, wherein the microbial cell converts fed glycoside intermediates into a glycosylated product.
72. The microbial cell of claim 71, wherein the microbial cell is a bacterial cell.
- 20 73. The microbial cell of claim 72, wherein the bacterial cell is *Escherichia* spp., *Bacillus* spp., *Rhodobacter* spp., *Zymomonas* spp., or *Pseudomonas* spp.
74. The microbial cell of claim 73, wherein the bacterial cell is selected from *Escherichia coli*, *Bacillus subtilis*, *Rhodobacter capsulatus*, *Rhodobacter sphaeroides*, *Zymomonas mobilis*, or *Pseudomonas putida*.
- 25 75. The microbial cell of claim 74, wherein the bacterial cell is *E. coli*.

76. The microbial cell of claim 75, wherein the bacterial cell comprises the genetic modifications: *ushA* and *galETKM* are deleted, inactivated, or reduced in expression; *pgi* is deleted, inactivated, or reduced in expression; and *pgm* and *galU* are overexpressed.
77. The microbial cell of claim 70, wherein the microbial cell is a yeast cell.
- 5 78. The microbial cell of claim 77, wherein the yeast is *Saccharomyces spp.*, *Schizosaccharomyces spp.*, *Pichia spp.*, *Phaffia spp.*, *Kluyveromyces spp.*, *Candida spp.*, *Talaromyces spp.*, *Brettanomyces spp.*, *Pachysolen spp.*, *Debaryomyces spp.*, or *Yarrowia spp.*
- 10 79. The microbial cell of claim 78, wherein the yeast cell is selected from *Saccharomyces cerevisiae*, *Pichia pastoris*, and *Yarrowia lipolytica*.
80. The microbial cell of any one of claims 70 to 79, wherein the microbial cell does not express a plant biosynthetic pathway producing a glycoside intermediate.
81. The microbial cell of any one of claims 70 to 80, wherein the microbial cell expresses an endogenous or recombinant transport protein that transports glycoside intermediate into
15 the cell.
82. The microbial cell of any one of claims 70 to 81, wherein the microbial cell expresses a 1-3' glycosylating UGT enzyme, which is optionally selected from SrUGT76G1, MbUGT1-3, MbUGT1-3_1, MbUGT1-3_1.5, MbUGT1-3_2, or a derivative thereof.
- 20 83. The microbial cell of claim 82, wherein the microbial cell expresses a 1-2' glycosylating UGT enzyme, which is optionally selected from SrUGT91D2, SrUGT91D1, SrUGT91D2e, OsUGT1-2, MbUGT1,2, MbUGT1,2.2, and derivatives thereof.
84. The microbial cell of any one of claims 83, wherein the microbial cell expresses a C13 O-glycosylating UGT enzyme, which is optionally selected from SrUGT85C2 and derivatives thereof.

85. The microbial cell of claim 84, wherein the microbial cell expresses a C19 O-glycosylating UGT enzyme, which is optionally selected from SrUGT74G1, MbUGTc19, and derivatives thereof.
86. The microbial cell of claim 82, wherein the microbial cell expresses a 1-3' glycosylating UGT enzyme, a 1-2' glycosylating UGT enzyme, a C19 O-glycosylating UGT enzyme, and a C13 O-glycosylating UGT enzyme.
87. The microbial cell of claim 86, wherein the microbial cell expresses a SrUGT85C2 or derivative thereof, MbUGT1,2 or derivative thereof, SrUGT74G1 or derivative thereof, and SrUGT76G1 or derivative thereof.
88. The microbial cell of claim 86 or 87, wherein the UGT enzymes are integrated into the chromosome of the microbial cell.
89. The microbial cell of claim 86 or 87, wherein the UGT enzymes are expressed extrachromosomally.
90. The microbial cell of claim 70, wherein the microbial cell has a deletion, inactivation, or reduced activity or expression of a gene encoding an enzyme that consumes UDP-glucose.
91. The method of claim 90, wherein the microbial cell has a deletion, inactivation, or reduced activity of *ushA* or ortholog thereof and/or one or more of *galE*, *galT*, *galK*, and *galM*, or ortholog.
92. The method of claim 91, wherein *galETKM* genes are inactivated, deleted, or reduced in expression.
93. The method of claim 70, wherein the microbial cell has a deletion, inactivation, or reduced activity or expression of *otsA* (trehalose-6-phosphate synthase) or ortholog thereof.
94. The method of claim 70, wherein the microbial cell has a deletion, inactivation, or reduced activity or expression of *ugd* (UDP-glucose 6-dehydrogenase) or ortholog thereof.

95. The method of claim 70, wherein the microbial cell has a deletion, inactivation, or reduced expression of waaG, waaO, waaJ, yfdG, yfdI, and/or wcaJ, or ortholog(s) thereof.
96. The method of claim 70, wherein the microbial cell has a deletion, inactivation, or reduced activity or expression of pgi (glucose-6 phosphate isomerase) or ortholog thereof.
- 5 97. The method of claim 70, wherein the microbial cell has an overexpression or increased activity of one or more genes encoding an enzyme involved in converting glucose-6-phosphate to UDP-glucose, which is optionally pgm (phosphoglucomutase) and/or galU (UTP-glucose-1-phosphate uridylyltransferase), or ortholog or derivative thereof.
98. The method of claim 70, wherein the microbial cell has an overexpression or
10 increased activity or expression of ycjU (β -phosphoglucomutase) and *Bifidobacterium bifidum* ugpA, or ortholog or derivative thereof.
99. The method of claim 70, wherein the microbial cell has one or more genetic modifications that increase flux to the pentose phosphate pathway (PPP), and which is optionally an overexpression or increased activity of *E. coli* zwf or homologue or engineered
15 derivative thereof.
100. The method of claim 70, wherein the microbial cell has one or more genetic modifications that increase glucose transport, and which optionally includes increased expression or activity of *E. coli* galP and *E. coli* glk (glucokinase), or alternatively *Zymomonas mobilis* glf and *E. coli* glk, or homologs, orthologs, or engineered derivatives of
20 these genes.
101. The method of claim 70, wherein the microbial cell has one or more genetic modifications that increase UTP production and recycling, and which optionally include increased expression or activity of *E. coli* pyrH (UMP kinase), *E. coli* cmk (cytidylate kinase), *E. coli* adk (adenylate kinase), or *E. coli* ndk (nucleoside diphosphate kinase), or
25 homologs, orthologs, or engineered derivatives of these enzymes.
102. The method of claim 70, wherein the microbial cell has one or more genetic modifications that increase UDP production, and which optionally include: overexpression

or increased activity of upp (uracil phosphoribosyltransferase), pyrH (UMP kinase), and cmk (cytidylate kinase), or homolog or engineered derivatives thereof; or overexpression or increased activity of dctA (C4 dicarboxylate/orotate:H⁺ symporter), pyrE (orotate phosphoribosyltransferase), pyrH (UMP kinase) and cmk (cytidylate kinase), or homolog or
5 engineered derivatives thereof.

103. The method of claim 70, wherein the microbial cell has one or more genetic modifications to remove or reduce regulation of glucose uptake, which optionally include deletion, inactivation, or reduced expression of sgrS small regulatory RNA.

104. The method of claim 70, wherein the microbial cell has one or more genetic
10 modifications that reduce dephosphorylation of glucose-1-phosphate, which optionally include deletion, inactivation, or reduced expression or activity of one or more of *E. coli* agp (glucose-1-phosphatase), *E. coli* yihX (α -D-glucose-1-phosphate phosphatase), *E. coli* ybiV (sugar phosphatase), *E. coli* yidA (sugar phosphatase), *E. coli* yigL (phosphosugar phosphatase), and *E. coli* phoA (alkaline phosphatase), or an ortholog thereof in the
15 microbial cell.

105. The method of claim 70, wherein the microbial cell has one or more genetic modifications that reduce conversion of glucose-1-phosphate to TDP-glucose, which optionally include a deletion, inactivation, or reduced expression or activity of one or more of *E. coli* rffH (dTDP-glucose pyrophosphorylase) and *E. coli* rfbA (dTDP glucose
20 pyrophosphorylase), or an ortholog thereof.

106. The method of claim 70, wherein the microbial cell has one or more genetic modifications that reduce conversion of glucose-1-phosphate to ADP-glucose, and which optionally include deletion, inactivation, or reduced expression or activity of *E. coli* gigC (glucose-1-phosphate adenylyltransferase), or an ortholog thereof.