

METHODS AND MATERIALS FOR ENZYMATIC SYNTHESIS OF MOGROSIDE COMPOUNDS

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority from U.S. Serial No. 61/563,303, filed November 23, 2011, the disclosure of which is incorporated by reference in its entirety.

TECHNICAL FIELD

5 This invention relates to methods and materials for enzymatic synthesis of mogroside compounds, and more particularly to glycosylating mogrol using Uridine-5'-diphospho (UDP) dependent glucosyltransferases (UGTs) to produce various mogroside compounds.

BACKGROUND

10 Mogrosides are a family of triterpene glycosides isolated from fruits of *Siraitia grosvenorii* (Swingle), also known as *Momordica grosvenori* (Swingle). Extracts of the fruits are commercially used as natural sweeteners. Four major compounds, Mogroside V, Mogroside IV, Siamenoside I, and 11-Oxomogroside V, have been identified from the fruits of *Siraitia grosvenorii* (Swingle) that are responsible for the sweetness of the fruits.
15 See FIG. 1. Mogroside V is the most abundant of these four compounds at approximately 0.57% (w/w) of the dry fruit, followed by Mogroside IV and Siamenoside I, each of which contain four glucose moieties. 11-Oxomogroside V has a ketone group instead of a hydroxyl at C-11. See, e.g., Takemoto, *et al.*, *Yakugaku Zasshi*, 103, 1151-1154; 1155-1166; 1167-1173, (1983); Kasai, *et al.*, *Agric. Biol. Chem.* 53, 3347-3349 (1989);
20 Matsumoto, *Chem. Pharm. Bull.* 38, 2030-2032 (1990); and Prakash, *et al.*, *J. Carbohydrate Chem.* 30, 16-26 (2011).

All mogrosides share the same triterpene core, named mogrol. The aglycone mogrol is glycosylated with different numbers of glucose moieties to form various mogroside compounds. Mogroside is thought to be synthesized in the following manner:
25 synthesis of cucurbitadienol from the common triterpene precursor squalene; P450 oxidations of cucurbitadienol to produce the aglycone mogrol; and glycosylation of

mogrol to add the five glucoses one after another to produce mogroside V. See, Tang, *et al.*, *BMC Genomics*, 12, 343 (2011). Both intermediates cucurbitadienol and mogrol exist in the fruit as they have been isolated as minor products. See Ukiya, *et al.*, *J. Agric. Food Chem.* 50, 6710-6715 (2002). Glycoside intermediates exist in both 11-hydroxy and 11-oxo series, and gradually change from mogroside I to mogroside V as fruits ripen, which indicates that the triterpene core is fully oxidized by P450 enzymes before the subsequent glycosylations. However, the enzymes responsible for producing mogrosides have not been identified.

SUMMARY

In one aspect, this document features a method of producing a mogroside compound. The method includes incubating mogrol with a Uridine-5'-diphospho (UDP) dependent glucosyltransferase (UGT) to produce a mogroside compound (e.g., mogroside Ia, mogroside Ib, or a mogroside compound glycosylated at C25-OH). The UGT can be selected from the group consisting of 73C3, 73C6, 85C2, 73C5, and 73E1. The UGTs can be recombinantly produced or can be in a cell lysate of a recombinant host.

This document also features a method of producing a mogroside compound. The method includes contacting mogrol with a cell lysate prepared from a recombinant host expressing a UGT to produce a mogroside compound (e.g., mogroside Ia, mogroside Ib, or a mogroside compound glycosylated at C25-OH). The UGT can be selected from the group consisting of 73C3, 73C6, 85C2, 73C5, and 73E1.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the invention belongs. Although methods and materials similar or equivalent to those described herein can be used to practice the invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and are not intended to be limiting. Other features and advantages of the invention will be apparent from the following detailed

description. Applicants reserve the right to alternatively claim any disclosed invention using the transitional phrase “comprising,” “consisting essentially of,” or “consisting of,” according to standard practice in patent law.

DESCRIPTION OF DRAWINGS

5 FIG. 1 contains the chemical structure of Mogroside V, Mogroside IV, Siamenoside I, and 11-Oxomogroside V.

 FIG. 2 is a depiction of the biosynthesis of mogroside Ia and mogroside Ib from mogrol using UGTs.

 FIG. 3 contains the amino acid sequences of the followings UGTs: UGT73C3,
10 UGT73C5, UGT73C6, UGT73E1, and UGT85C2 (SEQ ID NOs: 1-5, respectively).

 FIG. 4 is a schematic of the products obtained from mogroside V after incubation with a pectinase and/or a cellulase.

DETAILED DESCRIPTION

 This document provides methods and materials for glycosylating mogrol using
15 one or more Uridine-5'-diphospho (UDP) dependent glucosyltransferases (UGTs). As indicated below, at least five UGTs have been identified that glycosylate the aglycone mogrol. See FIG. 2. Each of the UGTs identified herein are in glycosyltransferase family I. UGTs 73C3, 73C6, 85C2 and 73E1 glycosylate at the C24-OH position (UGT#2 in FIG. 2), while UGT73C5 glycosylates at both the C3-OH (UGT#1 in FIG. 2) and C24-
20 OH position (UGT#2). UGTs 73C3, 73C5, and 73C6 are from *Arabidopsis thaliana*. UGT 73E1 and 85C2 are from *Stevia rebaudiana*. The amino acid sequences of UGTs 73C3, 73C5, 73C6, 73E1, and 85C2 (SEQ ID NOs:1-5) are set forth in FIG. 3.

 UGT polypeptides described herein can be produced using any suitable method. For example, UGT polypeptides can be produced by chemical synthesis. Alternatively, a
25 UGT polypeptide described herein can be produced by standard recombinant technology using heterologous expression vectors encoding that UGT polypeptide. Expression vectors can be introduced into host cells (e.g., by transformation or transfection) for expression of the encoded polypeptide, which then can be purified. Expression systems

that can be used for small or large scale production of UGT polypeptides include, without limitation, microorganisms such as bacteria (e.g., *E. coli* and *B. subtilis*) transformed with recombinant bacteriophage DNA, plasmid DNA, or cosmid DNA expression vectors containing the nucleic acid molecules described herein. Useful expression systems also include insect cell systems infected with recombinant virus expression vectors (e.g., baculovirus) containing the nucleic acid molecules described herein, and plant cell systems infected with recombinant virus expression vectors (e.g., tobacco mosaic virus) or transformed with recombinant plasmid expression vectors (e.g., Ti plasmid) containing the nucleic acid molecules described herein. UGT polypeptides also can be produced using mammalian expression system harboring recombinant expression constructs containing promoters derived from the genome of mammalian cells (e.g., the metallothionein promoter) or from mammalian viruses (e.g., the adenovirus late promoter and the cytomegalovirus promoter), along with the nucleic acids described herein. UGT polypeptides can have an N-terminal or C-terminal tag as discussed below.

This document also provides isolated nucleic acids encoding the UGT polypeptides. An “isolated nucleic acid” refers to a nucleic acid that is separated from other nucleic acid molecules that are present in a genome, including nucleic acids that normally flank one or both sides of the nucleic acid in a genome. The term “isolated” as used herein with respect to nucleic acids also includes any non-naturally-occurring nucleic acid sequence, since such non-naturally-occurring sequences are not found in nature and do not have immediately contiguous sequences in a naturally-occurring genome.

An isolated nucleic acid can be, for example, a DNA molecule, provided one of the nucleic acid sequences normally found immediately flanking that DNA molecule in a naturally-occurring genome is removed or absent. Thus, an isolated nucleic acid includes, without limitation, a DNA molecule that exists as a separate molecule (e.g., a chemically synthesized nucleic acid, or a cDNA or genomic DNA fragment produced by PCR or restriction endonuclease treatment) independent of other sequences as well as DNA that is incorporated into a vector, an autonomously replicating plasmid, a virus (e.g., any paramyxovirus, retrovirus, lentivirus, adenovirus, or herpes virus), or into the genomic

DNA of a prokaryote or eukaryote. In addition, an isolated nucleic acid can include an engineered nucleic acid such as a DNA molecule that is part of a hybrid or fusion nucleic acid. A nucleic acid existing among hundreds to millions of other nucleic acids within, for example, cDNA libraries or genomic libraries, or gel slices containing a genomic DNA restriction digest, is not considered an isolated nucleic acid.

In some embodiments, a nucleic acid sequence encoding a UGT polypeptide can include a tag sequence that encodes a “tag” designed to facilitate subsequent manipulation (e.g., to facilitate purification or detection), secretion, or localization of the encoded polypeptide. Tag sequences can be inserted in the nucleic acid sequence encoding the UGT polypeptide such that the encoded tag is located at either the carboxyl or amino terminus of the UGT polypeptide. Non-limiting examples of encoded tags include green fluorescent protein (GFP), glutathione S transferase (GST), HIS tag, and Flag™ tag (Kodak, New Haven, CT). Other examples of tags include a chloroplast transit peptide, a mitochondrial transit peptide, an amyloplast peptide, signal peptide, or a secretion tag.

Functional homologs

Functional homologs of the polypeptides described above are also suitable for use in the methods and recombinant hosts described herein. A functional homolog is a polypeptide that has sequence similarity to a reference polypeptide, and that carries out one or more of the biochemical or physiological function(s) of the reference polypeptide. A functional homolog and the reference polypeptide may be natural occurring polypeptides, and the sequence similarity may be due to convergent or divergent evolutionary events. As such, functional homologs are sometimes designated in the literature as homologs, or orthologs, or paralogs. Variants of a naturally occurring functional homolog, such as polypeptides encoded by mutants of a wild type coding sequence, may themselves be functional homologs. Functional homologs can also be created via site-directed mutagenesis of the coding sequence for a polypeptide, or by combining domains from the coding sequences for different naturally-occurring polypeptides (“domain swapping”). Techniques for modifying genes encoding functional

UGT polypeptides described herein are known and include, *inter alia*, directed evolution techniques, site-directed mutagenesis techniques and random mutagenesis techniques, and can be useful to increase specific activity of a polypeptide, alter substrate specificity, alter expression levels, alter subcellular location, or modify polypeptide:polypeptide interactions in a desired manner. Such modified polypeptides are considered functional homologs. The term “functional homolog” is sometimes applied to the nucleic acid that encodes a functionally homologous polypeptide.

Functional homologs can be identified by analysis of nucleotide and polypeptide sequence alignments. For example, performing a query on a database of nucleotide or polypeptide sequences can identify homologs of UGT polypeptides. Sequence analysis can involve BLAST, Reciprocal BLAST, or PSI-BLAST analysis of nonredundant databases using an UGT amino acid sequence as the reference sequence. Amino acid sequence is, in some instances, deduced from the nucleotide sequence. Those polypeptides in the database that have greater than 40% sequence identity are candidates for further evaluation for suitability as a UGT polypeptide. Amino acid sequence similarity allows for conservative amino acid substitutions, such as substitution of one hydrophobic residue for another or substitution of one polar residue for another. If desired, manual inspection of such candidates can be carried out in order to narrow the number of candidates to be further evaluated. Manual inspection can be performed by selecting those candidates that appear to have domains present in UGT polypeptides, *e.g.*, conserved functional domains.

Conserved regions can be identified by locating a region within the primary amino acid sequence of a polypeptide that is a repeated sequence, forms some secondary structure (*e.g.*, helices and beta sheets), establishes positively or negatively charged domains, or represents a protein motif or domain. See, *e.g.*, the Pfam web site describing consensus sequences for a variety of protein motifs and domains on the World Wide Web at sanger.ac.uk/Software/Pfam/ and pfam.janelia.org/. The information included at the Pfam database is described in Sonnhammer *et al.*, *Nucl. Acids Res.*, 26:320-322 (1998); Sonnhammer *et al.*, *Proteins*, 28:405-420 (1997); and Bateman *et al.*, *Nucl. Acids Res.*, 27:260-262 (1999). Conserved regions also can be determined by aligning sequences of

the same or related polypeptides from closely related species. Closely related species preferably are from the same family. In some embodiments, alignment of sequences from two different species is adequate.

Typically, polypeptides that exhibit at least about 40% amino acid sequence identity are useful to identify conserved regions. Conserved regions of related polypeptides exhibit at least 45% amino acid sequence identity (*e.g.*, at least 50%, at least 60%, at least 70%, at least 80%, or at least 90% amino acid sequence identity). In some embodiments, a conserved region exhibits at least 92%, 94%, 96%, 98%, or 99% amino acid sequence identity. Sequence identity can be determined as set forth above.

Methods of Producing Mogroside Compounds

Mogroside compounds can be produced by incubating a mogrol substrate with one or more of the UGT polypeptides described herein, resulting in the production of a mogroside product. In some embodiments, the reaction mixture contains a plurality of UGT polypeptides such that a plurality of glycosylations occur in the reaction vessel. In other embodiments, the reaction mixture contains a single UGT polypeptide and one or more glycosylations catalyzed by that polypeptide take place. For example, a first reaction vessel can comprise a substrate and one or more UGT polypeptides for producing an intermediate, which can be introduced into a second reaction vessel containing one or more other UGT polypeptides to produce a subsequent intermediate or a mogroside product. The product produced in the second reaction vessel then can be recovered.

Each of the UGT polypeptides can be a purified polypeptide, *e.g.*, can be added to a reaction mixture as a solution containing 80%, 90%, 95%, or greater than 99% by weight of the desired UGT. Alternatively, the UGT polypeptide(s) can be present in a cell lysate prepared from a recombinant host expressing the UGT(s), and can be added to a reaction mixture as a cell lysate for incubation with the mogrol substrate.

Levels of products, substrates and intermediates can be determined by extracting samples from the reaction vessel for analysis according to published methods.

Mogroside compounds can be recovered from the reaction vessel using various techniques known in the art.

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

5

EXAMPLES

Example 1 -Purification of mogroside V

Mogroside V was purified from commercially available monk fruit extracts (PureLo®, Swanson) as follows. Three bottles of PureLo® (240 grams) were dissolved in
10 water (900 mL), then loaded on a column of HP-20 resin (400 gram resin). The column was washed with water (2.5 liters); then further washed with 20% methanol-water. The product was eluted with methanol. After evaporation of solvents and drying under high vacuum, mogroside V (2.5 grams, ~80% purity, 11-oxomogroside V was the major impurity) was obtained.

15

Example 2 - Enzymatic synthesis of mogrol from mogroside V

Mogroside V (300 mg) was dissolved in 0.1M sodium acetate buffer (pH 4.5, 100mL), and crude pectinase from *Aspergillus niger* (25mL, Sigma P2736) was added. The mixture was stirred at 50°C for 48 hours. The reaction mixture was extracted with
20 ethyl acetate (2x100ml). The organic extract was dried under vacuum then purified with preparative HPLC. Pure mogrol (40 mg) was obtained and its structure confirmed by NMR and mass spectroscopy. See FIG. 4.

25

Example 3 - Enzymatic synthesis of mogrol 3-O-glucoside (mogroside Ia) and mogrol 24-O-glucoside (mogroside Ib) from mogroside V

Mogroside V (300 mg) was dissolved in 0.1M sodium acetate buffer (pH 4.5, 100ml), and crude pectinase from *Aspergillus niger* (25ml, Sigma P2736) was added. The mixture was stirred at 50°C for 6.5 hours. The reaction mixture was extracted with ethyl acetate (2x100ml). The organic extract was dried under vacuum then purified with

preparative HPLC. Pure mogroside Ia (11.0 mg) and mogroside Ib (8.0 mg) were obtained. Their structures were confirmed by NMR and mass spectroscopy. See FIG. 4.

Example 4 - In vitro UGT screening and reactions

5 *In vitro* reactions of mogrol with a panel of 230 UGT enzymes were performed
and the products were analyzed with LC-MS. The *in vitro* UGT reaction mixtures included
4X Tris buffer, mogrol (250 μ M), UDP-glucose (750 μ M) and 1% alkaline phosphatase.
Five μ L of each partially purified UGT enzyme or crude enzyme extract was added to the
reaction, and the reaction volume brought to 50 μ L with water. The reactions were
10 incubated overnight at 30°C and performed in sterilized 96 well plates. After the
incubation, 25 μ L of DMSO were added into each reaction and the reaction plates were
centrifuged for 5 min. Forty μ L samples were taken from each well and filtered, and were
used for LC-MS analysis.

15 UGTs 73C3, 73C6 and 85C2 were found to convert all the mogrol substrate to mogroside Ib. UGT 73C5 makes both mogroside Ia and Ib. In the reaction with UGT 73E1, although the reaction was not complete, mogroside Ib was found as the major product, together with a new glycosylated mogrol (neither mogroside Ia nor Ib; exact mass shown as a mogroside I, presumably caused by a glycosylation event on C25-OH).

20 **OTHER EMBODIMENTS**

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

Claims

1. A method of producing a mogroside compound, comprising contacting and glycosylating a mogrol with a Uridine-5'-diphospho (UDP) dependent glucosyltransferase (UGT) polypeptide that is:
 - (a) UGT73C3 polypeptide having at least 90% sequence identity to the amino acid sequence set forth in SEQ ID NO:1;
 - (b) UGT73C5 polypeptide having at least 90% sequence identity to the amino acid sequence set forth in SEQ ID NO:2;
 - (c) UGT73C6 polypeptide having at least 90% sequence identity to the amino acid sequence set forth in SEQ ID NO:3;
 - (d) UGT73E1 polypeptide having at least 90% sequence identity to the amino acid sequence set forth in SEQ ID NO:4; or
 - (e) UGT85C2 polypeptide having at least 90% sequence identity to the amino acid sequence set forth in SEQ ID NO:5.
2. A method of producing a mogroside compound, comprising contacting and glycosylating a mogrol with a cell lysate prepared from a recombinant host expressing a UGT polypeptide that is:
 - (a) UGT73C3 polypeptide having at least 90% sequence identity to the amino acid sequence set forth in SEQ ID NO:1;
 - (b) UGT73C5 polypeptide having at least 90% sequence identity to the amino acid sequence set forth in SEQ ID NO:2;
 - (c) UGT73C6 polypeptide having at least 90% sequence identity to the amino acid sequence set forth in SEQ ID NO:3;
 - (d) UGT73E1 polypeptide having at least 90% sequence identity to the amino acid sequence set forth in SEQ ID NO:4; or
 - (e) UGT85C2 polypeptide having at least 90% sequence identity to the amino acid sequence set forth in SEQ ID NO:5.
3. A method of producing a mogroside compound, comprising:

- (a) contacting and glycosylating a mogrol with a one or more Uridine-5'-diphospho (UDP) dependent glucosyltransferase (UGT) polypeptide;
- (b) contacting and glycosylating a mogrol with a cell lysate prepared from a recombinant host expressing a one or more UGT polypeptide capable of glycosylating the mogrol to produce a one or more mogroside compound;
- (c)
 - (i) contacting and glycosylating a mogrol in a first reaction vessel with a one or more Uridine-5'-diphospho (UDP) dependent glucosyltransferase (UGT) polypeptide capable of glycosylating the mogrol to produce an intermediate compound;
 - (ii) contacting and glycosylating the intermediate compound in a second reaction vessel with the one or more UGT polypeptides capable of glycosylating the intermediate compound to produce a one or more mogroside compound; and
 - (iii) recovering the one or more mogroside compound produced in step (ii);or
- (d)
 - (i) contacting and glycosylating a mogrol in a first reaction vessel with a cell lysate prepared from a recombinant host expressing a one or more UGT polypeptide capable of glycosylating the mogrol to produce an intermediate compound;
 - (ii) contacting and glycosylating the intermediate compound in a second reaction vessel with the one or more UGT polypeptides capable of glycosylating the intermediate compound to produce a one or more mogroside compound; and
 - (iii) recovering the one or more mogroside compound produced in step (ii),wherein the UGT polypeptide is selected from:
 - (a) UGT73C3 polypeptide having at least 90% sequence identity to the amino acid sequence set forth in SEQ ID NO:1;
 - (b) UGT73C5 polypeptide having at least 90% sequence identity to the amino acid sequence set forth in SEQ ID NO:2;
 - (c) UGT73C6 polypeptide having at least 90% sequence identity to the amino acid sequence set forth in SEQ ID NO:3;

- (d) UGT73E1 polypeptide having at least 90% sequence identity to the amino acid sequence set forth in SEQ ID NO:4; and
 - (e) UGT85C2 polypeptide having at least 90% sequence identity to the amino acid sequence set forth in SEQ ID NO:5.
4. The method of any one of claims 1 to 3, wherein the mogroside compound is mogroside 1a; mogroside 1b or mogroside compound glycosylated at C25-OH.
 5. The method of any one of claims 1 to 3, wherein the UGT polypeptide is a recombinant UGT polypeptide.
 6. The method of claims 2 or 3, wherein the UGT polypeptide is produced by chemical synthesis.
 7. The method of claim 5, wherein the recombinant UGT polypeptide is produced in a cell lysate of a recombinant host.
 8. The method of any one of claims 1 to 3, wherein the UGT polypeptide contains a tag sequence located at either a carboxyl or an amino terminus of the UGT polypeptide.
 9. The method of claim 8, wherein the tag sequence comprises a green fluorescent protein (GFP), a glutathione S transferase (GST), a HIS tag, a Flag tag, a chloroplast transit peptide, a mitochondrial transit peptide, an amyloplast peptide, a signal peptide, or a secretion tag.
 10. The method of any one of claims 2, 3 or 7, wherein the recombinant host is a microorganism, transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA expression vectors.

11. The method of any one of claims 2, 3 or 7, wherein the recombinant host is an insect cell system infected with recombinant virus expression vectors.
12. The method of any one of claims 2, 3 or 7, wherein the recombinant host is a plant cell system infected with recombinant virus expression vectors or transformed with recombinant virus expression vectors.
13. The method of any one of claims 2, 3 or 7, wherein the recombinant host is a mammalian expression system harboring recombinant expression constructs containing promoters derived from the genome of mammalian cells or from mammalian viruses.