

ENZYMES, HOST CELLS, AND METHODS FOR PRODUCING MOGROSIDESPRIORITY

This Application claims the benefit of, and priority to, US Application No. 63/596,330, filed November 6, 2023, which is hereby incorporated by reference in its entirety.

5

SEQUENCE LISTING

The instant application contains a sequence listing, which has been submitted in XML format via Patent Center. The contents of the XML file, named “MAN-039PC_107590-5038_Sequence_Listing,” which was created on November 5, 2024, is 123,589 bytes in size, are incorporated herein by reference in their entirety.

10

BACKGROUND

Mogrosides are triterpene-derived specialized secondary metabolites found in the fruit of the Cucurbitaceae family plant *Siraitia grosvenorii* (a/k/a monkfruit or Luo Han Guo). Their biosynthesis in fruit involves a number of consecutive glycosylations of the aglycone mogrol. The food industry is increasing its use of mogroside fruit extract as a natural non-sugar food sweetener. For example, mogroside V (Mog.V) has a sweetening capacity that is ~250 times that of sucrose (Kasai *et al.*, *Agric Biol Chem* (1989)). Moreover, additional health benefits of mogrosides have been revealed in recent studies (Li *et al.*, *Chin J Nat Med* (2014)).

15

20

A variety of factors are promoting a surge in interest in research and commercialization of the mogrosides and monkfruit in general, including, for example, the explosion in popularity of and demand for natural sweeteners; the difficulties in scalable sourcing of other promising natural sweeteners such as rebaudioside M (RebM) from the Stevia plant; the superior taste performance of Mog.V relative to other natural and artificial sweetener products on the market; and the medicinal potential of the plant and fruit.

25

Purified Mog.V has been approved as a high-intensity sweetening agent in Japan (Jakimovich *et al.*, *Journal of Natural Products* (1990)) and the extract has gained GRAS status in the USA as a non-nutritive sweetener and flavor enhancer (GRAS 522). Extraction of mogrosides from the fruit can yield a product of varying degrees of purity, often accompanied by undesirable aftertaste. In addition, yields of mogroside from cultivated fruit are limited due to low plant yields and particular cultivation requirements of the plant. Mogrosides are present at about 1% in the

fresh fruit and about 4% in the dried fruit (Li HB, *et al.*, 2006). Mog.V is the main component, with a content of 0.5% to 1.4% in the dried fruit. Moreover, purification difficulties limit purity for Mog.V, with commercial products from plant extracts being standardized to about 50% Mog.V. It is highly likely that a pure Mog.V product will achieve greater commercial success than the blend, since it is less likely to have off flavors, will be easier to formulate into products, and has good solubility potential. It is therefore advantageous to be able to produce sweet mogroside compounds via biotechnological processes.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1A shows biosynthetic path to Mog.V production *in vivo*. The enzymatic transformation required for each step is indicated, along with the type of enzyme required. Undesired reactions producing off pathway chemicals (*e.g.*, cucurbitadienol, hydrolyzed 2,3;22,23 dioxidosqualene, and glycosylated 24,25 dihydroxy cucurbitadienol) are shown by X. **FIG. 1B** illustrates glycosylation routes to Mog.V. Bubble structures represent different mogrosides. White tetra-cyclic core represents mogrol. The numbers below each structure indicate the particular glycosylated mogroside. Black circles represent C3 or C24 glycosylations. Dark grey vertical circles represent 1,6-glycosylations. Light grey horizontal circles represent 1,2-glycosylations. Abbreviations: Mog, mogrol; Sia, siamenoside.

FIG. 2A to **FIG. 2E** show the improvement in 2,3;22,23-dioxidosqualene production by a bacterial strain producing squalene and expressing derivatives of squalene epoxidases (SQE). **FIG. 2A** to **FIG. 2C** show the improvement in 2,3;22,23-dioxidosqualene production by a bacterial strain producing squalene and expressing derivatives of squalene epoxidase (SQE) made after successive rounds of engineering. **FIG. 2A** shows the fold improvement in the production of 2,3;22,23-dioxidosqualene by the bacterial strain expressing SQE_2 (SEQ ID NO: 3) compared to the bacterial strain expressing SQE_1 (SEQ ID NO: 2). **FIG. 2B** shows the fold improvement in the production of 2,3;22,23-dioxidosqualene by the bacterial strain expressing SQE_3 compared to the strain expressing SQE_2. **FIG. 2C** shows the fold improvement in the production of 2,3;22,23-dioxidosqualene by the bacterial strain expressing SQE_4 (SEQ ID NO: 4) compared to the strain expressing SQE_3. **FIG. 2D** and **FIG. 2E** show the improvement in 2,3;22,23-dioxidosqualene production by a bacterial strain producing squalene and expressing derivatives of a squalene epoxidase (McSQE_0; SEQ ID NO: 77) made after successive rounds of engineering.

FIG. 2D shows the fold improvement in the production of 2,3;22,23-dioxidosqualene by the bacterial strain expressing McSQE_1 (SEQ ID NO: 78) compared to the bacterial strain expressing McSQE_0. **FIG. 2E** shows the fold improvement in the production of 2,3;22,23-dioxidosqualene by the bacterial strain expressing McSQE_2 (SEQ ID NO: 79) compared to the bacterial strain expressing McSQE_1.

FIG. 3 shows the improvement in the production of 24,25-dihydroxy-cucurbitadienol by engineering of an epoxide hydrolase (EPH). The fold improvement in the production of 24,25-dihydroxy-cucurbitadienol by a bacterial strain that produces 24,25-epoxycucurbitadienol and expresses an engineered version EPH_1 (SEQ ID NO: 6) compared to the same strain expressing EPH_0 (SEQ ID NO: 5).

FIG. 4A to **FIG. 4E** show the improvement in 24,25-epoxy-cucurbitadienol production by a bacterial strain that produces 2,3;22,23-dioxidosqualene and expresses a derivative of cucurbitadienol synthase (CDS) made after successive rounds of engineering. **FIG. 4A** shows the fold improvement in the production of 24,25-epoxy-cucurbitadienol by the bacterial strain that expresses ECDS_1 (SEQ ID NO: 8) compared to the strain expressing ECDS_0 (SEQ ID NO: 7). **FIG. 4B** shows the fold improvement in the production of 24,25-epoxy-cucurbitadienol and the off-pathway compound cucurbitadienol by a bacterial strain expressing ECDS_2 (SEQ ID NO: 9) compared to the strain expressing ECDS_1. **FIG. 4C** shows the fold improvement in the production of 24,25-epoxy-cucurbitadienol by the strain expressing ECDS_3 (SEQ ID NO: 10) compared to the strain expressing ECDS_2. **FIG. 4D** shows the fold improvement in the production of 24,25-dihydroxy-cucurbitadienol by the strain expressing ECDS_4 (SEQ ID NO: 11) compared to the strain expressing ECDS_3. **FIG. 4E** shows the fold improvement in the production of 24,25-dihydroxy-cucurbitadienol by the strain expressing ECDS_5 (SEQ ID NO: 12) compared to the strain expressing ECDS_4.

FIG. 5 shows an alignment of CcCDS (SEQ ID NO: 7) with homologues SgCDS (SEQ ID NO: 68), CpCAS (SEQ ID NO: 69), PsCAS (SEQ ID NO: 70), and AsCAS (SEQ ID NO: 71). Some positions for mutagenesis are highlighted, as described in Table 11.

FIG. 6 shows the improvement in 24,25-dihydroxycucurbitadienol: cucurbitadienol ratio by a bacterial strain expressing AaSQS (SEQ ID NO: 13), SQE_3, ECDS_4 and EPH_1 compared to a bacterial strain expressing AaSQS, SQE_2, ECDS_3 and EPH_1.

FIG. 7A to **FIG. 7T** show the improvement in mogrol production by a bacterial strain producing 24,25-dihydroxycucurbitadienol and expressing derivatives of a cytochrome P450 enzyme created after successive rounds of engineering for hydroxylation of C11 of 24,25-dihydroxycucurbitadienol. **FIG. 7A** shows the fold improvement in the production of mogrol by the bacterial strain expressing sohB_CppCYP87A3 (SEQ ID NO: 15) compared to the bacterial strain expressing CppCYP87A3 (SEQ ID NO: 14). **FIG. 7B** shows the fold improvement in the production of mogrol by a strain expressing the bacterial strain expressing sohB_C11CYP_1 (SEQ ID NO: 16) compared to a strain expressing the parental enzyme sohB_CppCYP87A3. **FIG. 7C** shows the fold improvement in the production of mogrol by a strain expressing engineered enzyme sohB_C11CYP_2 (SEQ ID NO: 17) compared to a strain expressing the parental enzyme sohB_C11CYP_1. **FIG. 7D** shows the fold improvement in the production of mogrol by a strain expressing engineered enzyme sohB_C11CYP_3 (SEQ ID NO: 18) compared to a strain expressing the parental enzyme sohB_C11CYP_2. **FIG. 7E** shows the fold improvement in the production of mogrol by a strain expressing engineered enzyme sohB_C11CYP_4 (SEQ ID NO: 19) compared to a strain expressing the parental enzyme sohB_C11CYP_3. **FIG. 7F** shows the fold improvement in the production of mogrol by a strain expressing engineered enzyme sohB_C11CYP_5 (SEQ ID NO: 20) compared to a strain expressing the parental enzyme sohB_C11CYP_4. **FIG. 7G** shows the fold improvement in the production of mogrol, 11-hydroxyl-cucurbitadienol, and oxo-mogrol by a bacterial strain expressing engineered enzyme sohB_C11CYP_6 (SEQ ID NO: 21) compared to a strain expressing the parental enzyme sohB_C11CYP_5. **FIG. 7H** shows the fold improvement in the production of mogrol and 11-oxo mogrol by a strain expressing SohB_C11CYP_8 (SEQ ID NO: 45) compared to a strain expressing SohB_C11CYP_6. **FIG. 7I** shows the fold improvement in the production of mogrol and 11-oxo mogrol by a strain expressing SohB_C11CYP_9 (SEQ ID NO: 46) compared to a strain expressing SohB_C11CYP_8. **FIG. 7J** shows the fold improvement in the production of mogrol by a strain expressing SohB_C11CYP_10 (SEQ ID NO: 47) compared to a strain expressing SohB_C11CYP_9. **FIG. 7K** shows the fold improvement in the production of mogrol by a strain expressing SohB_C11CYP_11 (SEQ ID NO: 80) compared to a strain expressing SohB_C11CYP_10. **FIG. 7L** shows the fold improvement in the production of mogrol by a strain expressing SohB_C11CYP_12 (SEQ ID NO: 81) compared to a strain expressing SohB_C11CYP_11. **FIG. 7M** shows the fold improvement in the ratio mogrosides to

deoxymogrosides, and the production of 11-oxo mogroside by a strain expressing SohB_C11CYP_13 (SEQ ID NO: 82) compared to a strain expressing SohB_C11CYP_12. **FIG. 7N** shows the fold improvement in the production of mogrol by a strain expressing SohB_C11CYP_14 (SEQ ID NO: 83) compared to a strain expressing SohB_C11CYP_13. **FIG. 7O** shows the fold improvement in the production of mogrol by a strain expressing SohB_C11CYP_15 (SEQ ID NO: 84) compared to a strain expressing SohB_C11CYP_14. **FIG. 7P** shows the fold improvement in the production of mogrol by a strain expressing n20_C11CYP_15 (SEQ ID NO: 85) compared to a strain expressing SohB_C11CYP_15. **FIG. 7Q** shows the fold improvement in the production of mogrol by a strain expressing n20_C11CYP_16 (SEQ ID NO: 86) compared to a strain expressing n20_C11CYP_15. **FIG. 7R** shows the fold improvement in the production of mogrol by a strain expressing n20_C11CYP_17 (SEQ ID NO: 87) compared to a strain expressing n20_C11CYP_16. **FIG. 7S** shows the fold improvement in the production of mogrol by a strain expressing n20_C11CYP_18 (SEQ ID NO: 88) compared to a strain expressing n20_C11CYP_17. **FIG. 7T** shows the fold improvement in the production of mogrol by a strain expressing n20_C11CYP_19 (SEQ ID NO: 89) compared to a strain expressing n20_C11CYP_18.

FIG. 8A to **FIG. 8D** show the improvement in mogrol production by a bacterial strain expressing the biosynthetic pathway to mogrol and a derivative of cytochrome P450 reductase enzyme developed after successive rounds of engineering to support cytochrome P450-based hydroxylation at C11 of the 24,25-dihydroxycucurbitadienol with the engineered cytochrome P450 enzyme. **FIG. 8A** shows the fold improvement in the production of mogrol, including MogCPR1_1 (SEQ ID NO: 23) compared to a strain expressing MogCPR1_0 (SEQ ID NO: 22). **FIG. 8B** shows the fold improvement in the production of mogrol by a bacterial strain expressing MogCPR2_0 (SEQ ID NO: 24) compared to a strain expressing MogCPR1_0. **FIG. 8C** shows the fold improvement in the production of mogrol by a strain expressing MogCPR2_1 (SEQ ID NO: 25) compared to a strain expressing MogCPR2_0. **FIG. 8D** shows the fold improvement in the production of mogrol by a strain expressing MogCPR2_2/ SohB_C11CYP_6 (SEQ ID NO: 26) compared to a strain expressing MogCPR2_0.

FIG. 9A to **FIG. 9F** show the improvement in glycosylation of C3 position of mogrol after successive rounds of engineering of a uridine diphosphate dependent glycosyltransferase (UGT) for C3 glycosylation of mogrol. **FIG. 9A** shows the fold improvement in the production of MogIIIE

from MogIA substrate in an enzymatic assay with MogUGTC3_1 (SEQ ID NO: 28) compared to MogUGTc3_0 (SEQ ID NO: 27). **FIG. 9B** shows the fold improvement in the production of MogIE by a bacterial strain producing mogrol and expressing MogUGTC3_2 (SEQ ID NO: 29) compared to a strain expressing MogUGTC3_1. **FIG. 9C** shows the fold improvement in the production of MogIE compared to glycosylation of dihydroxycucurbitadiol off pathway products by a strain expressing MogUGTC3_3 (SEQ ID NO: 30) compared to a strain expressing MogUGTC3_2. **FIG. 9D** shows fold improvement in MogIE production from expression of CmUGTc3_1 as compared to CmUGTc3_0 in a mogrol production strain. **FIG. 9E** shows fold improvement in MogIE and deoxy-MogIE production from expression of CmUGTc3_1 variants in a mogrol production strain, and showing improved specificity towards mogrosides over deoxymogrosides. **FIG. 9F** shows fold improvement in MogIE and deoxy-MogIE production from expression of CmUGTc3_3 in a mogrol production strain, as compared to CmUGTc3_2.

FIG. 10A to FIG. 10E show the improvement in glycosylation of C24 position of mogrol after successive rounds of engineering of a uridine diphosphate dependent glycosyltransferase (UGT) for C24 glycosylation. **FIG. 10A** shows the fold improvement in the production of MogIIE from MogIE substrate in an enzymatic assay with MogUGTC24_1 (SEQ ID NO: 32) compared to MogUGTc24_0 (SEQ ID NO: 31). **FIG. 10B** shows the fold improvement in the production of MogIIE from MogIE substrate in an enzymatic assay with MogUGTC24_2 (SEQ ID NO: 33) compared to MogUGTC24_1. **FIG. 10C** shows the fold improvement in the production of MogIIE from MogIE substrate in an enzymatic assay with MogUGTC24_3 (SEQ ID NO: 34) compared to MogUGTC24_2. **FIG. 10D** shows fold improvement in MogIA production from expression of MogUGTc24_4 in a mogrol production strain, as compared to MogUGTc24_3. **FIG. 10E** shows fold improvement in MogIA and deoxy-MogIA production from expression of MogUGTc24_5 in a mogrol production strain.

FIG. 11A to FIG. 11D show the improvement in MogV production after successive rounds of engineering of a uridine diphosphate dependent glycosyltransferase (UGT). **FIG. 11A** shows the fold improvement in the production of MogV in an enzymatic assay for conversion of MogIIE to MogV by MogUGT1216_1 enzyme (SEQ ID NO: 37) compared to UGT94-289-1 enzyme (SEQ ID NO: 35) or MogUGT1216_0 (SEQ ID NO: 36). **FIG. 11B** shows the fold improvement in the production of MogIII, MogIV-A, Siamenoside, and MogV in an enzymatic assay for conversion of MogIIE by MogUGT1216_1, which is a double mutant (L14V-L266Q), compared to L14V

single substitution version (Mut1) or L266Q single substitution version (Mut16). **FIG. 11C** shows the fold improvement in the production of MogV in an enzymatic assay for conversion of MogIIE to MogV by MogUGT1216_2 (SEQ ID NO: 38) compared to MogUGT1216_1. **FIG. 11D** shows the fold improvement in the production of MogV in an enzymatic assay for conversion of MogIIE to MogV by MogUGT1216_3 (SEQ ID NO: 90) compared to MogUGT1216_2.

FIG. 12 shows the improvement in MogV production in an enzymatic assay for conversion of MogIIE to MogV by *Coffea arabica* uridine diphosphate dependent glycosyltransferase 1,6 (MogUGT16_0, SEQ ID NO: 39) and its engineered derivative MogUGT16_1 (SEQ ID NO: 40).

FIG. 13A to **FIG. 13C** show the improvement in 2,3;22,23-dioxidosqualene production after the expression of PntAB, a membrane-bound transhydrogenase that generates NADPH in *E. coli*, and successive rounds of engineering of PntAB. **FIG. 13A** shows the fold improvement in the production of 2,3;22,23-dioxidosqualene by a bacterial strain expressing SQS and SQE and overexpressing *pntAB* (SEQ ID NOs: 41 and 42), compared to a strain that does not overexpress *pntAB*. **FIG. 13B** shows the fold improvement in the production of 24,25-epoxy cucurbitadienol by a bacterial strain expressing SQS, SQE, ECDS, and EPH and overexpressing *pntAB* compared to a strain expressing SQS, SQE, ECDS, and EPH but not overexpressing *pntAB*. **FIG. 13C** shows the fold change in squalene, 2,3-oxidosqualene and 2,3:22,23-dioxidosqualene in bacterial strains overexpressing native *E. coli fldA* (SEQ ID NO: 43) or heterologous Cv.fdx gene (SEQ ID NO: 44) and expressing SQS and SQE, compared to a strain expressing SQS and SQE only.

FIG. 14A to **FIG. 14C** show the improvement in the production of mogrol and 24,24-dihydroxycucurbitadienol in a strain having reduced expression of an ABC type transporter encoded by the *ybhGFSR* operon. **FIG. 14A** shows the levels of 2,3;22,23-dioxidosqualene (OOSQ), and 24,24-dihydroxycucurbitadienol + mogrol in strain expressing the entire metabolic pathway for producing mogrol having the *ybhGFSR* operon deleted, in comparison with a *ybhGFSR*⁺ control strain that expresses the entire metabolic pathway for producing mogrol (CTL). **FIG. 14B** shows the levels of 2,3;22,23-dioxidosqualene (OOSQ), and 24,24-dihydroxycucurbitadienol + mogrol in strain expressing the entire metabolic pathway for producing mogrol having a hypomorphic *ybhG* mutation created by replacing the ribosome binding site in the chromosome with a weaker version than the native sequence, in comparison with a *ybhGFSR*⁺ control strain that expresses the entire metabolic pathway for producing mogrol

(CTL). **FIG. 14C** shows the levels of 24,24-dihydroxycucurbitadienol + mogrol (top panel), and 22,3;22,23-dioxidosqualene (OOSQ, bottom panel) in strain expressing the entire metabolic pathway for producing mogrol having a chromosomal *ybhF* gene deletion and having over-expression of the *ybhGFSR* operon from a plasmid, in comparison with a *ybhGFSR*⁺ control strain that expresses the entire metabolic pathway for producing mogrol (CTL).

DETAILED DESCRIPTION

The present disclosure is based on development of engineered enzymes and microbial cells that improve the microbial production of mogrol or mogrosides from squalene compared to known processes by increasing the production of various pathway intermediates. In various aspects, the engineered enzymes disclosed herein increase the production of pathway intermediates and reduce the formation of off pathway metabolites, *inter alia*, by increasing the substrate preference for preferred pathway intermediates, or decreasing substrate preference for undesired pathway intermediates or off pathway products, thereby channeling the metabolic resources of the microbial cells to the production of pathway intermediates while reducing the formation of undesired off pathway metabolites.

The pathway intermediates and certain off pathway metabolites are illustrated in **FIG. 1A** and **FIG. 1B**. **FIG. 1A** shows an *in vivo* biosynthetic path to mogrol and to MogV. This pathway involves condensation of farnesyl pyrophosphate to squalene, which is catalyzed by a squalene synthase (SQS); serial epoxidation of squalene to 2,3-oxidosqualene and 2,3;22,23-dioxidosqualene both of which are catalyzed by a squalene epoxidase (SQE); cyclization of 2,3;22,23-dioxidosqualene to 24,25-epoxy cucurbitadienol, which is catalyzed by an epoxy cucurbitadienol synthase (ECDS); hydration of 24,25-epoxy cucurbitadienol to 24,25-dihydroxy cucurbitadienol, which is catalyzed by an epoxide hydrolase (EPH); hydroxylation of 24,25-dihydroxy cucurbitadienol to mogrol, which are catalyzed by a cytochrome P450 and a cytochrome P450 reductase that are active on C11 position of 24,25-dihydroxy cucurbitadienol (**FIG. 1A**). **FIG. 1B** illustrates glycosylation routes to Mog.V. The conversion of mogrol to MogV involves C3 glucosylation, C24 glucosylation, 1,6-glucosylation, and 1,2-glucosylation reactions, that are catalyzed by uridine diphosphate dependent glycosyltransferases (UGTs, see **FIG. 1A** and **FIG. 1B**).

The pathway intermediates shown in **FIG. 1A**, *i.e.*, farnesyl pyrophosphate, squalene; 2,3-oxidosqualene, 2,3;22,23-dioxidosqualene, 24,25-epoxy cucurbitadienol, 24,25-dihydroxy cucurbitadienol, mogrol; and the mogrosides shown in **FIG. 1B** are referred to herein as “on pathway products.” As illustrated in **FIG. 1A**, the enzymes such as CDS, EPH, and UGTs catalyze the conversion of on pathway intermediates to undesired products, which include cucurbitadienol, hydrolyzed 2,3;22,23-dioxidosqualene, and glycosides of 24,25 dihydroxy cucurbitadienol. These undesired products are referred to herein as “off pathway” products. A bacterial cell expressing a squalene synthase (SQS), a squalene epoxidase (SQE); a cucurbitadienol synthase (CDS), an epoxide hydrolase (EPH) generally produces about 90% or more (by weight) off pathway products, and relatively low amounts of mogrol. Disclosed herein is a pathway comprising engineered enzymes that improve production of on pathway products. A bacterial cell comprising the engineered enzymes disclosed herein produces more than about 40% (by weight), or more than about 50%, or more than about 60%, or more than about 70%, or more than about 80% on pathway products, with respect to the total of the on pathway and off pathway products shown in **FIG. 1A** as 100%. A bacterial cell comprising the engineered enzymes disclosed herein produces less than about 50% (by weight), or less than about 40%, or less than about 30%, or less than about 20% off pathway products (and specifically the off-pathway products cucurbitadienol, hydrolyzed 2,3;22,23-dioxidosqualene, and glycosides of 24,25 dihydroxy cucurbitadienol).

Thus, the present disclosure, in various aspects and embodiments, provides enzymes, microbial strains, and methods for making mogrol and mogrol glycosides, using recombinant microbial processes. In other aspects, the invention provides methods for making products, including foods, beverages, and sweeteners (among others), by incorporating the mogrol glycosides produced according to the methods described herein. In still other aspects, the invention provides engineered UGT enzymes for glycosylating secondary metabolite substrates, such as mogrol or mogrosides.

As used herein, the terms “terpene” or “triterpene” are used interchangeably with the terms “terpenoid” or “triterpenoid,” respectively.

Accordingly, in one aspect, the present disclosure provides a method for producing mogrol or mogroside, comprising: providing a microbial cell producing squalene and expressing a biosynthetic pathway in the microbial cell that converts squalene to the mogrol or mogroside, and

culturing the microbial cell under conditions suitable for producing the mogrol or mogroside. The culturing produces squalene derivatives with less than about 40% by weight of off-pathway products cucurbitadienol, hydrolyzed 2,3;22,23-dioxidosqualene, and glycosylated 24,25-dihydroxycucurbitadienol. In some embodiments, the culturing produces at least 50% (by weight) or at least 60% of the on pathway products farnesyl pyrophosphate, squalene, 2,3-oxidosqualene, 2,3;22,23-dioxidosqualene, 24,25-epoxy cucurbitadienol, 24,25-dihydroxy cucurbitadienol, mogrol, and mogrosides; where the total of such on pathway products and the off pathway products cucurbitadienol, hydrolyzed 2,3;22,23-dioxidosqualene, and glycosides of 24,25 dihydroxy cucurbitadienol are considered as 100%.

In various aspects, the present disclosure provides a microbial cell producing squalene and expressing a biosynthetic pathway that converts squalene to mogrol or mogroside. When the microbial cell is cultured under conditions suitable for producing the mogrol or mogroside, the microbial cell produces squalene derivatives with less than 40% or less than 30% or less than 20% by weight of off-pathway products cucurbitadienol, hydrolyzed 2,3;22,23-dioxidosqualene, and glycosylated 24,25-dihydroxycucurbitadienol. In some embodiments, the microbial cell produces at least 50% or at least 60% or at least 70% by weight of the on pathway products farnesyl pyrophosphate, squalene, 2,3-oxidosqualene, 2,3;22,23-dioxidosqualene, 24,25-epoxy cucurbitadienol, 24,25-dihydroxy cucurbitadienol, mogrol, and mogrosides; where the total of such on pathway products and the off pathway products cucurbitadienol, hydrolyzed 2,3;22,23-dioxidosqualene, and glycosides of 24,25 dihydroxy cucurbitadienol are considered as 100%.

In some embodiments, the biosynthetic pathway comprises: a squalene synthase (SQS), a squalene epoxidase (SQE) enzyme that produces 2,3;22,23-dioxidosqualene from squalene; an epoxycucurbitadienol synthase (ECDS) enzyme that produces 24,25-epoxycucurbitadienol from 2,3;22,23-dioxidosqualene; an epoxide hydrolase (EPH) enzyme that produces 24,25-dihydroxy-cucurbitadienol from 24,25-epoxycucurbitadienol; and a cytochrome P450 enzyme (e.g., a P450 hydroxylase) that produces mogrol from 24,25-dihydroxy-cucurbitadienol. In some embodiments, the microbial cell expresses a P450 reductase partner that regenerates the P450 hydroxylase (See FIG. 1A).

In some embodiments, the microbial cell expresses one or more uridine diphosphate dependent glycosyltransferase (UGT) enzymes that glycosylate C-3 and C-24 positions of mogrol

to produce one or more mogrosides from mogrol, including but not limited to Mog V. In some embodiments, the microbial cell expresses a UGT enzyme that produces a C-3 mogrol glycoside from mogrol (UGTc3). In some embodiments, the microbial cell expresses a UGT enzyme that produces a C-24 mogrol glycoside from mogrol (UGTc24). In some embodiments, the microbial cell expresses a UGT enzyme that produces a 1,6 and/or 1,2 branched glycoside of a C-3 and/or C-24 glycoside of mogrol (UGT1216) (See **FIG. 1B**).

According to various aspects and embodiments, the disclosure provides engineered enzymes for use in the biosynthetic pathway, including SQE, ECDS, EPH, cytochrome P450 hydroxylase, P450 reductase, UGTc3, UGTc24, and UGT1216 enzymes. These engineered enzymes are described in detail below.

In various aspects and embodiments, the present disclosure provides a squalene epoxidase comprising an amino acid sequence that is at least 80%, or at least 85%, at least 90%, or at least 95%, or at least 97%, or at least 98%, or at least 99% identical to the amino acid sequence of SEQ ID NO: 2, and comprising one or more mutations or sets of mutations listed in Tables 1, 2, and 3. In some embodiments, squalene epoxidase comprises at least two, or at least three, or at least four, or at least five, or at least six, or more mutations or sets of mutations listed in Table 1, Table 2, and/or Table 3. In some embodiments, the squalene epoxidase comprises at least 2, at least 3, at least 4, or at least 5 mutations or sets of mutations from Tables 1, 2, and/or 3 that are shown to provide at least a 1.2 fold improvement in 2,3-dioxidosqualene titer and/or at least a 1.2 fold improvement in 2,3;22,23-dioxidosqualene titer compared to the parent enzyme. In some embodiments, the microbial cell expressing a squalene epoxidase disclosed herein provides for an improvement in 2,3;22,23-dioxidosqualene production compared to the parental enzyme.

In some embodiments, the squalene epoxidase comprises one or more substitutions at positions selected from S116, N134, I145, A164, A169, and A400 with respect to SEQ ID NO: 2. In some embodiments, S116 is substituted with a charged amino acid selected from Glu and Asp. In some embodiments, N134 is substituted with a non-polar amino acid selected from Gly, Ala, Ile, Leu, and Val. In some embodiments, I145 is substituted with a charged amino acid selected from His, Arg, and Lys. In some embodiments, A164 is substituted with a polar amino acid selected from Asn and Gln. In some embodiments, A169 is substituted with a charged amino acid selected from Arg, Lys, Glu, Asp, or His. In some embodiments, A400 is substituted with a non-

polar amino acid selected from Val, Leu, and Ile. In some embodiments, the squalene epoxidase comprises one or more mutations selected from S116E, N134G, I145H, A164N, A169R, and A400V with respect to SEQ ID NO: 2. In some embodiments, the squalene epoxidase comprises at least two, or at least 3, or at least 4, or at least 5, or all six mutations selected from S116E,
5 N134G, I145H, A164N, A169R, and A400V with respect to SEQ ID NO: 2.

In various embodiments, the SQE comprises the amino acid sequence of SEQ ID NO: 3 or SEQ ID NO: 4.

In various aspects and embodiments, the present disclosure provides a squalene epoxidase comprising an amino acid sequence that is at least 80%, or at least 85%, at least 90%, or at least
10 95%, or at least 97%, or at least 98%, or at least 99% identical to the amino acid sequence of SEQ ID NO: 77, and comprising one or more mutations or one or more sets of mutations listed in Tables 4 and/or 5.

In some embodiments, the squalene epoxidase comprises one or more substitutions at D41, M315, and/or S366 with respect to SEQ ID NO: 77. In some embodiments, S366 is substituted
15 with a charged amino acid selected from Arg, Lys and His. In some embodiments, D41 is substituted with a positively charged amino acid selected from His, Lys and Arg. In some embodiments, M315 is substituted with a non-polar amino acid selected from Leu, Ile, and Val. In some embodiments, the squalene epoxidase comprises one or more mutations selected from S366R, D41H, and M315L with respect to SEQ ID NO: 77.

20 In various embodiments, the SQE comprises the amino acid sequence of SEQ ID NO: 78 or SEQ ID NO: 79.

The pathway intermediate 2,3-dioxidosqualene can get diverted to an undesired cyclization leading to the production of the off-pathway product cucurbitadienol (See **FIG. 1A**). Production of the off pathway product cucurbitadienol can lead to increased impurities, increased cost of
25 production, and overall lower productivity of the strain and process. Without wishing to be bound by theory, it is believed that the squalene epoxidase disclosed herein catalyzes two consecutive epoxidation reactions at a rate that limits the availability of the 2,3-dioxidosqualene substrate for the undesired cyclization reaction. Accordingly, in some embodiments, the squalene epoxidase disclosed herein catalyzes two epoxidation reactions: the first catalyzing the epoxidation of
30 squalene to produce 2,3-dioxidosqualene, and the second catalyzing the epoxidation of 2,3-

dioxidosqualene to produce 2,3;22,23-dioxidosqualene. Accordingly, in some embodiments, the culturing of microbial cells expressing the squalene epoxidase disclosed herein and an ECDS produces at least 50%, or at least 60%, or at least 70% by weight of the on pathway products 2,3-oxidosqualene, 2,3;22,23-dioxidosqualene, and 24,25-epoxy cucurbitadienol; where the total of these on pathway products and the off pathway products cucurbitadienol is considered as 100%. In some embodiments, the culturing of microbial cells expressing the squalene epoxidase disclosed herein produces less than 40% (by weight), or less than 30%, or less than 20% cucurbitadienol.

Further amino acid modifications of SQE can be guided by available enzyme structures and homology models, including those described in Padyana AK, et al., Structure and inhibition mechanism of the catalytic domain of human squalene epoxidase, *Nat. Comm.* (2019) Vol. 10(97): 1-10; or Ruckenstein et al., Structure-Function Correlations of Two Highly Conserved Motifs in *Saccharomyces cerevisiae* Squalene Epoxidase, *Antimicrob. Agents and Chemo.* (2008) Vol. 52(4): 1496-1499.

Cucurbitadienol synthase enzymes are capable of catalyzing two cyclization reactions: cyclization of the on pathway intermediate 2,3-dioxidosqualene to an off-pathway product cucurbitadienol; and cyclization of the on pathway intermediate 2,3;22,23 dioxidosqualene to an on-pathway intermediate 24,25-epoxy-cucurbitadienol (See **FIG. 1A**). Disclosed herein are engineered cucurbitadienol synthase enzymes having an increased specificity for 2,3;22,23 dioxidosqualene substrate as compared to 2,3-dioxidosqualene substrate. The engineered cucurbitadienol synthase enzymes are referred to herein as 24, 25 epoxy cucurbitadienol synthase ("ECDS"). In some embodiments, the culturing of microbial cells expressing the ECDS disclosed herein produces less off pathway product cucurbitadienol compared to the culturing of microbial cells expressing a non-engineered cucurbitadienol synthase. In some embodiments, the culturing of microbial cells expressing the ECDS disclosed herein produces more of the on pathway product 24,25-epoxy-cucurbitadienol compared to the culturing of microbial cells expressing a non-engineered cucurbitadienol synthase.

In various aspects and embodiments, the present disclosure provides a 24, 25 epoxy cucurbitadienol synthase (ECDS) comprising an amino acid sequence that is at least 80%, or at least 85%, or at least 90%, or at least 95%, or at least 97%, or at least 98%, or at least 99% identical to the amino acid sequence of SEQ ID NO: 7, and comprising one or more mutations or sets of

mutations listed in Tables 7, 8, 9 and 10. The ECDS enzyme produces 24,25-epoxycucurbitadienol from 2,3;22,23-dioxidosqualene. In some embodiments, the ECDS enzyme has substrate preference for 2,3;22,23-dioxidosqualene over 2,3-oxidosqualene, thereby reducing production of the off-pathway product cucurbitadienol. In some embodiments, the ECDS enzyme comprises an amino acid sequence having a sequence identity of at least 80%, or at least 85%, or at least 90%, or at least 95%, or at least 98%, or at least 99% with respect to SEQ ID NO: 7, and having one or more mutations with respect to SEQ ID NO: 7 that increase enzyme productivity or substrate preference for 2,3;22,23-dioxidosqualene over 2,3-oxidosqualene. In some embodiments, the ECDS enzyme comprises one or more mutations or sets of mutations listed in Tables 7, 8, 9, and/or 10. In some embodiments, the ECDS enzyme comprises at least 1, at least 2, or at least 3, or at least 4, or at least 5 mutations listed in Tables 7, 8, 9, and/or 10 that increase 24,25-epoxycucurbitadienol titer by at least 1.2 fold and/or reduce cucurbitadienol production (as shown in the table), compared to the corresponding parent enzyme.

In some embodiments, the ECDS comprises one or more mutations at one or more positions selected from S24, C35, D50, N121, L245, W331, Q401, I490, I553, and A556, with respect to SEQ ID NO: 7. In some embodiments, S24 is substituted with a polar amino acid selected from Asn and Gln. In some embodiments, C35 is substituted with a charged amino acid selected from Asp and Glu. In some embodiments, D50 is substituted with Glu. In some embodiments, N121 is substituted with His, Lys or Arg. In some embodiments, L245 is substituted with Ile or Val. In some embodiments, W331 is substituted with Ser, Thr, or Tyr. In some embodiments, Q401 is substituted with Ala, Gly, or Val. In some embodiments, I490 is substituted with Val, Leu, or Ala. In some embodiments, I553 is substituted with Met. In some embodiments, A556 is substituted with a polar amino acid selected from Ser, Thr, and Tyr. In some embodiments, the ECDS comprises at least 1, or at least 2, or at least 3, or at least 4, or at least 5, or at least 6, or at least 7, or at least 8, or at least 9, or all 10 mutations selected from 24N, 35D, 50E, 121H, 245I, 331S, 401A, 490V, 553M, and 556S with respect to SEQ ID NO: 7. Exemplary ECDS enzymes in accordance with the various embodiments are disclosed herein as SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, and SEQ ID NO: 12.

In some embodiments, the ECDS enzyme comprises one or more mutations in a substrate access channel that improves substrate selectivity for 2,3;22,23-dioxidosqualene over 2,3-oxidosqualene. This access channel is shared with homologous enzymes (see FIG. 5). Amino acid

modification in the positions lining the substrate access channel can alter substrate selectivity. Exemplary positions include amino acids 245, 331, 553, and 556 with respect to SEQ ID NO: 7, which are close to the membrane harboring the ECDS. Indeed, the mutations 245I, 331S, and 553M (with respect to SEQ ID NO: 7) significantly improve production of 24,25-epoxy-cucurbitadienol and 24,25-dihydroxy cucurbitadienol, while reducing production of cucurbitadienol (see Table 10). Accordingly, in some embodiments, the ECDS enzyme is an enzyme having at least 65%, or at least 70%, or at least 75%, or at least 80%, or at least 85%, or at least 90%, or at least 95%, or at least 97%, or at least 98%, or at least 99% sequence identity to SEQ ID NO: 7, but having a mutation at a position corresponding to 245, 331, 553, and 556 of SEQ ID NO: 7. In some embodiments, the ECDS comprises the amino acid sequence of SEQ ID NO: 68, SEQ ID NO: 69, SEQ ID NO: 70, or SEQ ID NO: 71 (or an enzyme having at least 90%, or at least 95%, or at least 97%, or at least 98%, or at least 99% sequence identity thereto), and comprising one or more mutations (e.g., 1, 2, 3, or 4 mutations) at positions corresponding to positions 245, 331 553, and 556 of SEQ ID NO: 7. See Table 11. For example, the amino acid corresponding to position 245 of SEQ ID NO: 7 may be a hydrophobic amino acid, such as Ile, Val, Ala, Met or Phe; the amino acid corresponding to position 331 of SEQ ID NO: 7 may be a polar amino acid, such as but not limited to Ser or Thr; the amino acid corresponding to position 553 of SEQ ID NO: 7 may be a hydrophobic amino acid, such as Ile, Val, Ala, Met or Phe; and the amino acid corresponding to position 556 of SEQ ID NO: 7 may be a small polar amino acid, such as Ser or Thr. Alternatively or in addition, the ECDS comprises a mutation at a position corresponding to position 50, 490, 121, and/or 401, with respect to SEQ ID NO: 7. See Table 11. For example, the amino acid I490 of SEQ ID NO: 7 is believed to make up a portion of the active site, and thus mutations at this position may impact substrate preference and/or productivity. In some embodiments, the amino acid at the position corresponding to position 490 of SEQ ID NO: 7 is a hydrophobic amino acid, such as Val, Leu, or Ala. See Table 11. Other non-active site positions that may be important for controlling substrate preference include positions 50, 121, and 401 of SEQ ID NO: 7. In embodiments, the position corresponding to position 50 of SEQ ID NO: 7 is mutated to an amino acid selected from Asp, His, Gln, Asn, and Glu. See Table 11. In embodiments, the position corresponding to position 121 of SEQ ID NO: 7 is mutated to an amino acid selected from Asp, His, Gln, Asn, and Glu. See Table 11. In embodiments, the position

corresponding to position 401 of SEQ ID NO: 7 is mutated to an amino acid selected from Ala, Asp, His, Gln, Asn, Glu, and Pro. See Table 11.

Further amino acid modifications of ECDS enzyme can be guided by available enzyme structures and homology models, including those described in Swiss Model Q6BE24 and
5 AlphaFold model K7NBZ9.

An epoxide hydrolase (EPH) enzyme is capable of catalyzing two hydration reactions: hydration of the on pathway intermediate 2,3;22,23-dioxidosqualene to an off-pathway product 2,3;22,23 dioxidosqualene; and hydration the on pathway intermediate 24,25-epoxycucurbitadienol to an on-pathway intermediate 24,25-dihydroxycucurbitadienol (See **FIG. 1A**).
10 Disclosed herein are engineered EPH enzymes having an increased specificity for 24,25-dihydroxycucurbitadienol compared to 2,3;22,23 dioxidosqualene. Accordingly, in some embodiments, the culturing of microbial cells expressing the EPH disclosed herein (as well as SQE and ECDS enzymes) produces less than 30%, or less than 20% 2,3;22,23 dioxidosqualene, when the total of the substrates 2,3;22,23-dioxidosqualene and 24,25-epoxycucurbitadienol, and
15 products 2,3;22,23 dioxidosqualene and 24,25-dihydroxycucurbitadienol is considered as 100%. In some embodiments, the culturing of microbial cells expressing the EPH disclosed herein (as well as SQE and ECDS enzymes) produces less off pathway product 2,3;22,23 dioxidosqualene compared to the culturing of microbial cells expressing a non-engineered EPH. In some
20 embodiments, the culturing of microbial cells expressing the EPH disclosed herein produces more on pathway products 24,25-dihydroxycucurbitadienol compared to the culturing of microbial cells expressing a non-engineered EPH.

In various aspects and embodiments, the present disclosure provides an epoxide hydrolase (EPH) enzyme that produces 24,25-dihydroxycucurbitadienol from 24,25-epoxycucurbitadienol. In some embodiments, the EPH enzyme has substrate preference for 24,25-epoxycucurbitadienol
25 over 2,3;22,23-dioxidosqualene, and therefore produces less of the off pathway product hydrolyzed 2,3;22,23 dioxidosqualene. In some embodiments, the EPH enzyme comprises an amino acid sequence having a sequence identity of at least 80%, or at least 85%, or at least 90%, or at least 95%, or at least 98%, or at least 99% with respect to SEQ ID NO: 5. In some
30 embodiments, the EPH has one or more mutations with respect to SEQ ID NO: 5 that increases enzyme productivity or substrate preference for 24,25-epoxycucurbitadienol over 2,3;22,23-

dioxidosqualene, and wherein one or more of said mutations are selected from Table 6. In some embodiments, the EPH enzymes comprises at least 2, at least 3, at least 4, or at least 5 amino acid substitutions listed in Table 6, with respect to the amino acid sequence of SEQ ID NO: 5.

In some embodiments, the epoxide hydrolase comprises one or more mutations at one or more of positions selected from S68, Q128, G141, T144, A145, E163, E191, L262, C295, N299 with respect to SEQ ID NO: 5. In some embodiments, S68 is substituted with a non-polar amino acid selected from Ala, Gly, and Val. In some embodiments, Q128 is substituted with a non-polar amino acid selected from Ala, Pro, Gly and Val. In some embodiments, G141 is substituted with a polar amino acid selected from Ser, Thr, and Tyr. In some embodiments, T144 is substituted with an amino acid selected from Gln, Asn, Ala, Gly, and Val. In some embodiments, A145 is substituted with a non-polar amino acid selected from Val, Ile, and Leu. In some embodiments, E163 is substituted with a non-polar amino acid selected from Ala, Gly, and Val. In some embodiments, E191 is substituted with a non-polar amino acid selected from Gly, Ala, and Val. In some embodiments, L262 is substituted with Met. In some embodiments, C295 is substituted with a non-polar amino acid selected from Gly, Ala, and Val. In some embodiments, N299 is substituted with Gln. In some embodiments, the epoxide hydrolase has at least one, or at least two, or at least 3, or at least 4, or at least 5, or at least 6, or at least 7, or at least 8, or at least 9, or all 10 mutations selected from S68A, Q128A, G141S, T144Q, A145V, E163A, E191G, L262M, C295G, and N299Q with respect to SEQ ID NO: 5. An exemplary EPH enzyme according to this disclosure comprises the amino acid sequence of SEQ ID NO: 6.

Further amino acid modifications of the EPH enzyme can be guided by available enzyme structures and homology models, including those described in Mitusińska, K, et al., Structure-function relationship between soluble epoxide hydrolases structure and their tunnel network, *Comput Struct Biotechnol J*. 2022; 20: 193–205; Zou et al., Structure of Aspergillus niger epoxide hydrolase at 1.8 Å resolution: implications for the structure and function of the mammalian microsomal class of epoxide hydrolases, *Structure* 8(2): 111-122 (2000); Schulz et al., The crystal structure of mycobacterial epoxide hydrolase A, *Scientific Reports* 10: 16539 (2020); Gomez et al., Human soluble epoxide hydrolase: Structural basis of inhibition by 4-(3-cyclohexylureido)-carboxylic acids, *Protein Sci.* 15(1): 58-64 (2006); Biswal et al., The molecular structure of epoxide hydrolase B from Mycobacterium tuberculosis and its complex with a urea-based inhibitor, *JMol Biol* 381: 897-912 (2008); Nardini et al., The X-ray structure of epoxide hydrolase

from *Agrobacterium radiobacter* AD1. An enzyme to detoxify harmful epoxides, *J Biol Chem* 274(21):14579-86 (1999).

In various aspects and embodiments, the present disclosure provides a cytochrome P450 enzyme that produces mogrol from 24,25-dihydroxy-cucurbitadienol. In some embodiments, the cytochrome P450 enzyme selectively hydroxylates C11 of 24,25-dihydroxycucurbitadienol. In some embodiments, the cytochrome P450 enzyme comprises an amino acid sequence that is at least 70% identical, or at least 80% identical, or at least 90% identical, or at least 95% identical, or at least 97% identical, or at least 98% identical, or at least 99% identical to the amino acid sequence of SEQ ID NO: 14 or SEQ ID NO: 15.

Functional expression of plant cytochrome P450 in heterologous hosts has inherent limitations attributable to the heterologous hosts. For example, bacterial systems have translational incompatibility of the membrane signal modules of P450 enzymes due to the lack of an endoplasmic reticulum. In addition, the expression and function of the cytochrome P450 enzymes becomes limiting in the heterologous hosts. In various aspects and embodiments, the present disclosure provides engineered cytochrome P450 enzymes that efficiently express in microbial hosts (including bacterial hosts) and selectively hydroxylate C11 of 24,25-dihydroxycucurbitadienol.

In some embodiments, the cytochrome P450 enzyme has a replacement of all or a portion of the native transmembrane domain with the transmembrane domain from an *E. coli* inner membrane cytoplasmic C-terminus protein, as disclosed in U.S. Patent No. 10,774,314, which is hereby incorporated by reference in its entirety. In some embodiments, the *E. coli* inner membrane cytoplasmic C-terminus protein is selected from ycgG, yhcB, zipA, waaA, sohB, djlA, ycgG, lpxK, ypfN, ypfN, yhhM, and FliO. In some embodiments, the *E. coli* inner membrane cytoplasmic C-terminus protein is sohB or FliO.

In various embodiments, the P450 enzyme has a deletion or truncation of part or all of its native transmembrane domain. Generally, the deletion or truncation is about the first 15 to 75 amino acids, and the desired length can be determined using predictive tools known in the art. For example, the deletion/truncation of the native transmembrane region may be from about 15 to about 30 amino acids. The native P450 transmembrane region is replaced with a membrane anchor sequence derived from an *E. coli* protein that is an inner membrane protein with its C-terminus in

the cytoplasm, such as sohB or FliO. For example, the P450 enzyme provided herein as SEQ ID NO: 15 corresponds to CYP87A3 (*Cucurbita pepo* subsp. *pepo*) having a portion of its native transmembrane domain (18 amino acids) replaced with the transmembrane domain of *E. coli* sohB.

Cytochrome P450 enzymes can exhibit broad substrate specificity or low substrate selectivity. See, *e.g.*, Kim *et al.* Gene engineering, purification, crystallization and preliminary X-ray diffraction of cytochrome P450 p-coumarate-3-hydroxylase (C3H), the Arabidopsis membrane protein, *Protein Expression and Purification* 79(1): 149-155 (2011); Yamaguchi *et al.* Cytochrome P450 CYP71AT96 catalyses the final step of herbivore-induced phenylacetonitrile biosynthesis in the giant knotweed, *Fallopia sachalinensis*, *Plant Molecular Biology* 91: 229-239 (2016). The broad substrate specificity and low substrate selectivity of the cytochrome P450 enzymes leads to wasteful off pathway reactions and reduced levels of oxidation of the desired molecules. Disclosed herein are engineered cytochrome P450 enzymes having an increased specificity for the C-11 position of 24, 25 dihydroxy cucurbitadienol. In some embodiments, the engineered cytochrome P450 enzymes disclosed herein have an increased substrate selectivity to C-11 position of 24,25 dihydroxy cucurbitadienol, compared to C-11 position of cucurbitadienol, allowing the production of higher levels of mogrol, compared to 11-hydroxy cucurbitadienol (see, *e.g.*, **FIG. 7G** and **FIG. 7I**).

Accordingly, in some embodiments, the culturing of microbial cells expressing the engineered cytochrome P450 disclosed herein (as well as SQE, ECDS, and EPH enzymes) produces less than 40%, or less than 30%, or less than 20% 11-hydroxy cucurbitadienol and 11-oxo-mogrol, when the total of the substrates cucurbitadienol and 24,25-dihydroxy-cucurbitadienol, and products mogrol, 11-hydroxy cucurbitadienol, and 11-oxo-mogrol is considered as 100%. In some embodiments, the culturing of microbial cells expressing the engineered cytochrome P450 disclosed herein (as well as SQE, ECDS, and EPH enzymes) produces more than 60%, or more than 70%, or more than 80% mogrol, when the total of the products mogrol, 11-hydroxy cucurbitadienol, and 11-oxo-mogrol is as considered 100%. In some embodiments, the culturing of microbial cells expressing the engineered cytochrome P450 disclosed herein produces less off pathway products 11-hydroxy cucurbitadienol and 11-oxo-mogrol compared to the culturing of microbial cells expressing a non-engineered cytochrome P450 enzyme. In some embodiments, the culturing of microbial cells expressing the engineered cytochrome P450 disclosed herein produces

more mogrol compared to the culturing of microbial cells expressing a non-engineered cucurbitadienol synthase.

In various embodiments, the cytochrome P450 enzyme comprises an S193T substitution with respect to the enzyme of SEQ ID NO: 15, provided herein as SEQ ID NO: 16. In various
5 embodiments, the cytochrome P450 enzyme comprises one or more mutations or sets of mutations (e.g., at least 2, 3, 4, or 5 mutations or sets of mutations) listed in Tables 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, and/or 29 with respect to the corresponding parent enzyme. In various embodiments, the P450 enzyme comprises at least 2, at least 3, at least 4, or at least 5 mutations or sets of mutations listed in Tables 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26,
10 27, 28, and/or 29 that increase mogrol titer by at least 1.5 fold with respect to the corresponding parent enzyme.

In some embodiments, the cytochrome P450 enzyme comprises a substitution at one or more positions selected from A12, I27, S63, K75, V76, N102, G125, T128, W130, L131, K132, T192, S193, L218, G219, T223, V228, T231, T232, Y233, N234, K246, F354, K369, H433, F452,
15 V465, A468, and T482, with respect to SEQ ID NO: 15. In some embodiments, S63 is substituted with Asn, Phe, Gln, Tyr, or Trp. In some embodiments, A12 is substituted with Trp, Tyr or Phe. In some embodiments, I27 is substituted with Gly or Ala. In some embodiments, S63 is substituted with Asn, Phe, Gln, Tyr, or Trp. In some embodiments, K75 is substituted with Arg. In some embodiments, V76 is substituted with Met, Leu or Ile. In some embodiments, N102 is substituted
20 with His, Lys, or Arg. In some embodiments, G125 is substituted with an amino acid selected from Ala, Leu, Ile, and Val. In some embodiments, T128 is substituted with an amino acid selected from Gly, Ser, Pro, and Ala. In some embodiments, W130 is substituted with an amino acid selected from Asn, Gly, Ser, Leu, Gln, Ala, Val, and Thr. In some embodiments, L131 is substituted with an amino acid selected from Val, Ile, Pro, Thr, and Ser. In some embodiments, K132 is substituted
25 with an amino acid selected from Asn, Ser, Thr, and Gln. In some embodiments, T192 is substituted with Leu, Ile, or Val. In some embodiments, S193 is substituted with Thr. In some embodiments, L218 is substituted with Ile, Val, and Ala. In some embodiments, G219 is substituted with an amino acid selected from Ala, Gln, Arg, Val, Asn, His, and Lys. In some embodiments, T223 is substituted with Ser. In some embodiments, V228 is substituted with Ile or
30 Leu. In some embodiments, T231 is substituted with Phe, Trp, or Tyr. In some embodiments, T232 is substituted with Ala, Gly, or Val. In some embodiments, Y233 is substituted with Phe. In some

embodiments, N234 is substituted with His, Arg or Lys. In some embodiments, K246 is substituted with Val, Ile or Leu. In some embodiments, F354 is substituted with Tyr or Trp. In some embodiments, K369 is substituted with Arg, Lys, or His. In some embodiments, H433 is substituted with Phe, Trp, or Tyr. In some embodiments, F452 is substituted with Ile, Val, or Leu.

5 In some embodiments, V465 is substituted with an amino acid selected from Ile, Met, or Leu. In some embodiments, A468 is substituted with an amino acid selected from Thr, Ser, Asn, and Gln. In some embodiments, T482 is substituted with Ser. In some embodiments, the cytochrome P450 enzyme comprises at least one, or at least two, or at least 3, or at least 4, or at least 5, or at least 6, or at least 7, or at least 8, or at least 9, or at least 10, or at least 11, or at least 12, or at least 13, or

10 at least 14, or at least 15, or at least 16, or at least 17, or at least 18, or at least 19, or at least 20, or at least 21, or at least 22, or at least 23, or at least 24, or at least 25, or at least 26, or at least 27, or at least 28, or at least 29, or all 30 substitutions selected from A12W, I27G, S63F, S63N, K75R, V76M, N102H, G125A, T128G, W130N, L131I, K132N, T192L, S193T, L218I, G219A, T223S, V228I, T231F, T232A, Y233F, N234H, K246V, F354Y, K369R, H433F, F452I, V465I, A468T

15 and T482S, each with respect to SEQ ID NO: 15.

In some embodiments, the cytochrome P450 enzyme comprises an N-terminal deletion of at least about 20, or at least about 25, or at least about 30, or at least about 35, or at least about 40, or at least about 45, or at least about 50 amino acids with respect to SEQ ID NO: 15. In some embodiments, the cytochrome P450 enzyme comprises a deletion selected from a deletion of L3

20 to H29 with respect to SEQ ID NO: 15. In some embodiments, the cytochrome P450 enzyme comprises an amino acid sequence that is at least 70% identical, or at least 80% identical, or at least 90% identical, or at least 95% identical, or at least 97% identical, or at least 98% identical, or at least 99% identical to the amino acid sequence of SEQ ID NO: 85. In some embodiments, the cytochrome P450 enzyme comprises one or more mutations or one or more sets of mutations

25 listed in Tables 26, 27, 28, and/or 29.

In some embodiments, the cytochrome P450 enzyme comprises a substitution at one or more positions selected K8, D9, S10, N13, V15, I45, K46, K47, M49, K50, R51, I191, A192, F406, E411, Y412, and S413, with respect to SEQ ID NO: 85. In some embodiments, K8 is substituted with Arg. In some embodiments, D9 is substituted with Asn or Gln. In some

30 embodiments, S10 is substituted with Arg, Lys or His. In some embodiments, N13 is substituted with Lys, Arg, or His. In some embodiments, V15 is substituted with Lys, Arg, or His. In some

embodiments, I45 is substituted with Val, Ile, or Leu. In some embodiments, K46 is substituted with Arg, Lys or His. In some embodiments, K47 is substituted with Glu or Asp. In some embodiments, M49 is substituted with Val, Ile, or Leu. In some embodiments, K50 is substituted with Glu or Asp. In some embodiments, R51 is substituted with Lys. In some embodiments, I191
5 is substituted with Leu or Val. In some embodiments, A192 is substituted with Lys, Arg, or His. In some embodiments, F406 is substituted with Tyr or Trp. In some embodiments, E411 is substituted with Asp. In some embodiments, Y412 is substituted with Phe, Leu, Ile, or Trp. In some embodiments, S413 is substituted with Ala, Gly, or Val. In some embodiments, the cytochrome P450 enzyme comprises at least one, or at least two, or at least 3, or at least 4, or at
10 least 5, or at least 6, or at least 7, or at least 8, or at least 9, or at least 10, or at least 11, or at least 12, or at least 13, or at least 14, or at least 15, or at least 16, or all 17 substitutions selected from K8R, D9N, S10R, N13K, and V15K I45V, K46R, K47E, M49V, K50E, R51K, F406Y, E411D, Y412F, S413A, I191L and A192K with respect to SEQ ID NO: 85.

In accordance with embodiments of this disclosure, the cytochrome P450 enzyme
15 comprises an amino acid sequence of SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 47, SEQ ID NO: 80, SEQ ID NO: 81, SEQ ID NO: 82, SEQ ID NO: 83, or SEQ ID NO: 84.

In various embodiments, the cytochrome P450 enzyme comprises an N-terminal deletion
20 of at least about 20, or at least about 25, or at least about 30, or at least about 35, or at least about 40, or at least about 45, or at least about 50 amino acids with respect to SEQ ID NO: 15. In some embodiments, the cytochrome P450 enzyme comprises a deletion selected from a deletion of L3 to H29 with respect to SEQ ID NO: 15. In some embodiments, the cytochrome P450 enzyme comprises an amino acid sequence that is at least 70% identical, or at least 80% identical, or at
25 least 90% identical, or at least 95% identical, or at least 97% identical, or at least 98% identical, or at least 99% identical to the amino acid sequence of SEQ ID NO: 85, and comprising one or more mutations or one or more sets of mutations listed in Tables 26, 27, 28, and/or 29.

In some embodiments, the cytochrome P450 enzyme comprises a substitution at one or more positions selected K8, D9, S10, N13, V15, I45, K46, K47, M49, K50, R51, I191, A192,
30 F406, E411, Y412, and S413, with respect to SEQ ID NO: 85. In some embodiments, K8 is

substituted with Arg. In some embodiments, D9 is substituted with Asn or Gln. In some
embodiments, S10 is substituted with Arg, Lys or His. In some embodiments, N13 is substituted
with Lys, Arg, or His. In some embodiments, V15 is substituted with Lys, Arg, or His. In some
embodiments, I45 is substituted with Val, Ile, or Leu. In some embodiments, K46 is substituted
5 with Arg, Lys or His. In some embodiments, K47 is substituted with Glu or Asp. In some
embodiments, M49 is substituted with Val, Ile, or Leu. In some embodiments, K50 is substituted
with Glu or Asp. In some embodiments, R51 is substituted with Lys. In some embodiments, I191
is substituted with Leu or Val. In some embodiments, A192 is substituted with Lys, Arg, or His.
In some embodiments, F406 is substituted with Tyr or Trp. In some embodiments, E411 is
10 substituted with Asp. In some embodiments, Y412 is substituted with Phe, Leu, Ile, or Trp. In
some embodiments, S413 is substituted with Ala, Gly, or Val. In some embodiments, the
cytochrome P450 enzyme comprises at least one, or at least two, or at least 3, or at least 4, or at
least 5, or at least 6, or at least 7, or at least 8, or at least 9, or at least 10, or at least 11, or at least
12, or at least 13, or at least 14, or at least 15, or at least 16, or all 77 substitutions selected from
15 K8R, D9N, S10R, N13K, and V15K I45V, K46R, K47E, M49V, K50E, R51K, F406Y, E411D,
Y412F, S413A, I191L and A192K substitutions with respect to SEQ ID NO: 85. In accordance
with embodiments of this disclosure, the cytochrome P450 enzyme comprises an amino acid
sequence of SEQ ID NO: 86, or SEQ ID NO: 87, or SEQ ID NO: 88, or SEQ ID NO: 89.

In various aspects and embodiments, the present disclosure provides a cytochrome P450
20 reductase enzyme that is capable of reducing a P450 hydroxylase enzyme of any of the
embodiments disclosed herein. In some embodiments, the P450 reductase comprises an amino acid
sequence that is at least 80%, or at least 85%, or at least 90%, or at least 95%, or at least 97%, or
at least 98%, or at least 99% identical to SEQ ID NO: 22 or SEQ ID NO: 24.

In some embodiments, the cytochrome P450 reductase enzyme comprises a replacement
25 of all or a portion (e.g., 15 to 75 amino acids) of the native transmembrane domain with the
transmembrane domain from an *E. coli* inner membrane cytoplasmic C-terminus protein (see U.S.
Patent No. 10,774,314, which is hereby incorporated by reference in its entirety). In some
embodiments, the *E. coli* inner membrane cytoplasmic C-terminus protein is selected from YcgG,
YhcB, ZipA, WaaA, SohB, DjlA, LpxK, YpfN, and YhhM. Exemplary P450 reductase enzymes
30 (based on the CPR enzyme of SEQ ID NO: 22), and having 67 amino acids truncated from the
native transmembrane domain, and replaced with transmembrane sequences (21 to 30 amino acids)

from *E. coli* inner membrane cytoplasmic C-terminus proteins are shown in Table 30. In some embodiments, the transmembrane sequence is from *E. coli* protein YcgG. In some embodiments, from about 50 to about 80 amino acids, or from about 65 to about 78 amino acids (e.g., from 65 to 75 amino acids) of the N-terminus of SEQ ID NO: 22 are deleted and from 21 to 28 amino acids
5 from the transmembrane domain of an *E. coli* inner membrane cytoplasmic C-terminus protein (e.g., YcgG) are inserted. In some embodiments, the native N-terminus of ID NO: 22 (as described) is replaced with 20 to 30 amino acids from the N-terminus of YcgG. In some embodiments, the native N-terminus of ID NO: 22 is replaced with about 25 or 26 amino acids from the N-terminus of YcgG. In some embodiments, the engineered P450 reductase having the native N-terminus of
10 SEQ ID NO: 22 replaced with the N-terminus of YcgG further comprises one or more mutations or sets of mutations listed in Table 31, with respect to the corresponding parent enzyme. In some embodiments, the CPR enzyme comprises at least 2, at least 3, at least 4, or at least 5 mutations or sets of mutations shown in Table 31 that increase mogrol production by at least 1.2 fold with respect to the corresponding parent enzyme. Exemplary CPR enzymes according to the disclosure
15 include those comprising the amino acid sequence of SEQ ID NO: 23.

In other embodiments, the P450 reductase enzymes (based on the CPR enzyme of SEQ ID NO: 24), and having about 75 amino acids truncated from the native transmembrane domain, and replaced with transmembrane sequences (21 to 30 amino acids) from *E. coli* inner membrane cytoplasmic C-terminus proteins are shown in Table 32. In some embodiments, the transmembrane
20 sequence is from *E. coli* protein YcgG. In some embodiments, the native N-terminus of ID NO: 24 (about 75 amino acids) is replaced with about 25 amino acids from the N-terminus of YcgG. In some embodiments, the engineered P450 reductase having the native N-terminus of SEQ ID NO: 24 replaced with the N-terminus of YcgG further comprises one or more mutations or sets of mutations listed in Table 33, with respect to the corresponding parent enzyme. In some
25 embodiments, the CPR enzyme comprises at least 2, at least 3, at least 4, or at least 5 mutations or sets of mutations shown in Table 33 that increase mogrol production. Exemplary CPR enzymes according to the disclosure include those comprising the amino acid sequence of SEQ ID NO: 25 and SEQ ID NO: 26. In some embodiments, the cytochrome P450 reductase comprises a substitution of L90 residue with respect to SEQ ID NO: 24, optionally with Pro.

30 In various aspects and embodiments, the present disclosure provides a uridine diphosphate dependent glycosyltransferase (UGT) enzyme that produce mogrosides from mogrol. Exemplary

mogrosides that can be produced are shown in **FIG. 1B**, and include Mog. 1A, Mog. 1E, Mog. IIA1, Mog. IIE, Mog. IIA2, Mog. IIIA1, Mog. III, Mog. IIIA2, siamenoside, Mog. IVA, Mog. IV, and Mog. V. In certain embodiments, the mogroside is Mog. V. Other mogrosides can be produced, such as iso-Mog. V. In some embodiments, the UGT enzyme produces at least 2 times more of the desired mogroside (e.g., Mog. V) than glycosylated products of 24,25-dihydroxycucurbitadienol. In some embodiments, the UGT enzymes produce at least 3 times, or at least 5 times, or at least 10 times more of the desired mogroside (e.g., Mog. V) than glycosylated products of 24,25-dihydroxycucurbitadienol.

Uridine diphosphate dependent glycosyltransferase (UGT) enzymes can exhibit substrate promiscuity. See, e.g., Song *et al.*, Functional Characterization and Substrate Promiscuity of UGT71 Glycosyltransferases from Strawberry (*Fragaria x ananassa*). *Plant and Cell Physiology*, 56: 2478-2493 (2015); Bönisch *et al.*, Activity based profiling of a physiologic aglycone library reveals sugar acceptor promiscuity of family 1 UDP-glucosyltransferases from *Vitis vinifera*. *Plant Physiology*, 166: 23-39 (2014); Bönisch *et al.*, A UDP-glucose: Monoterpenol glucosyltransferase adds to the chemical diversity of the grapevine metabolome. *Plant Physiology*, 165: 561-581 (2014); Hefner *et al.*, Arbutin synthase, a novel member of the NRD1 β glycosyltransferase family, is a unique multifunctional enzyme converting various natural products and xenobiotics. *Bioorganic and Medicinal Chemistry* 10: 1731-1741(2002). The substrate promiscuity of the UGT enzymes leads to wasteful off pathway reactions and reduced levels of oxidation of the desired molecules. Disclosed herein are engineered cytochrome UGT enzymes having an increased specificity for the C-3 or C-24 positions of mogrol or for 1,6 and 1,2 branched glycosylations.

In some aspects and embodiments, the present disclosure provides engineered UGT enzymes that have an increased substrate selectivity to C-3 position of mogrol compared to, e.g., C-3 position of 24, 25 dihydroxy cucurbitadienol, or positions other than C-3 position of mogrol. Such engineered enzymes are referred to herein as UGTc3. Accordingly, in some embodiments, the UGTc3 enzymes disclosed herein provide for the production of higher levels of Mog1E compared to a glucosylated dihydroxycucurbitadienol and deoxy-Mog1E side products (see, e.g., **FIG. 9C**, **FIG. 9E** and **FIG. 9F**). In some embodiments, the UGTc3 enzymes disclosed herein, using an equal molar mixture comprising mogrol, deoxy-mogrol and dihydroxycucurbitadienol as a substrate, produce less than 60%, or less than 50%, or less than 40%, or less than 30%, or less than 20% glucosylated dihydroxycucurbitadienol and deoxy-Mog1E side products, when the total

glucosylated products is considered as 100%. In some embodiments, the UGTc3 enzymes disclosed herein, using a mixture comprising mogrol, deoxy-mogrol and dihydroxycucurbitadienol, produces less off pathway products glucosylated dihydroxycucurbitadienol and deoxy-MogIE compared to a non-engineered UGT enzyme. In
5 some embodiments, the UGTc3 enzyme disclosed herein, using a mixture comprising mogrol, deoxy-mogrol and dihydroxycucurbitadienol, produces more MogIE compared to the culturing of microbial cells expressing a non-engineered UGT enzyme.

In various aspects and embodiments, the present disclosure provides a uridine diphosphate dependent glycosyltransferase (UGT) enzyme that produces a C-3 mogrol glycoside from mogrol
10 (UGTc3). In some embodiments, the UGTc3 enzyme comprises an amino acid sequence that is at least 70% identical, or at least 80%, or at least 85%, or at least 90%, or at least 95%, or at least 97%, or at least 98%, or at least 99% identical to an amino acid sequence of SEQ ID NO: 27, SEQ ID NO: 52, SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59, SEQ ID NO: 60, SEQ ID NO: 61, SEQ ID NO: 62, SEQ ID
15 NO: 63, and SEQ ID NO: 64 (see Table 34). In some embodiments, the UGTc3 enzyme comprises an amino acid sequence that is at least 70% identical, or at least 80%, or at least 85%, or at least 90%, or at least 95%, or at least 97%, or at least 98%, or at least 99% identical to the amino acid sequence of SEQ ID NO: 27. In some embodiments, the UGT enzyme converts Mog. 1A to Mog. IIE, and converts mogrol to Mog. IE. In some embodiments, the UGT enzyme comprises one or
20 more mutations with respect to SEQ ID NO: 27 that result in greater specificity and/or activity for glycosylation of C-3 position (e.g., in the case of mogrol and/or Mog. IE substrate) compared to the UGT enzyme having an amino acid sequence of SEQ ID NO: 27. In some embodiments, one or more mutations (e.g., at least 2, at least 3, at least 4, or at least 5 mutations) are selected from Table 35 and/or Table 36. In some embodiments, at least 2, or at least 3, or at least 4, or at least 5
25 mutations are selected from Tables 35 and 36 which are shown to improve Mog. IIE titer or Mog. IE titer by at least 1.2 fold or at least 1.3 fold. In various embodiments, the UGT enzyme with activity at C3 has reduced specificity for 24,25-dihydroxy cucurbitadienol and its glycosylated derivatives.

In some embodiments, the UGTc3 enzyme comprises one or more mutations at one or more
30 positions selected from L41, D49, T74, C127, F303, and A307 with respect to SEQ ID NO: 27. In some embodiments, L41 is substituted with an aromatic amino acid selected from Phe, Tyr and

Trp. In some embodiments, D49 is substituted with Glu. In some embodiments, T74 is substituted with Met. In some embodiments, C127 is substituted with an aromatic amino acid selected from Phe, Tyr and Trp. In some embodiments, F303 is substituted with Tyr. In some embodiments, A307 is substituted with an amino acid selected from Ile, Leu, and Val. In some embodiments, the UGTc3 enzyme comprises at least one, or at least two, or at least 3, or at least 4, or at least 5, or all six of the following substitutions with respect to SEQ ID NO: 27: L41F, D49E, T74M, C127F, F303Y, and A307I. Exemplary engineered UGT enzymes having activity at C3 of mogrol according to the present disclosure comprise the amino acid sequence of SEQ ID NO: 28, SEQ ID NO: 29, or SEQ ID NO: 30.

In various aspects and embodiments, the present disclosure provides a UGT enzyme that produces a C-3 mogrol glycoside from mogrol (UGTc3). In some embodiments, the UGTc3 enzyme comprises an amino acid sequence that is at least 70% identical, or at least 80%, or at least 85%, or at least 90%, or at least 95%, or at least 97%, or at least 98%, or at least 99% identical to the amino acid sequence of SEQ ID NO: 60 or SEQ ID NO: 65. The enzyme defined by SEQ ID NO: 65 is a circular permutant of the enzyme defined by SEQ ID NO: 60. UGT circular permutants are described in U.S. Patent No. 10,463,062, which is hereby incorporated by reference in its entirety. In some embodiments, the UGT enzyme comprises one or more mutations with respect to SEQ ID NO: 60 or SEQ ID NO: 65 that result in greater specificity and/or activity for glycosylation of C-3 position compared to the UGT enzyme having an amino acid sequence of SEQ ID NO: 60 or SEQ ID NO: 65. In some embodiments, one or more mutations (e.g., at least 2, at least 3, at least 4, or at least 5 mutations) are selected from Table 37 with respect to SEQ ID NO: 60. In some embodiments, at least 1, or at least 2, or at least 3, or at least 4, or at least 5 mutations are selected from Table 37 and which are shown to increase Mog IE titer by at least 1.4 fold or at least 1.5 fold. In some embodiments, the mutation is a circular permutation with a cut site as shown in Table 38, and optionally a linking sequence. In some embodiments, one or more mutations or sets of mutations (e.g., at least 2, at least 3, at least 4, or at least 5 mutations or sets of mutations) are selected from Table 39 with respect to SEQ ID NO: 65. In some embodiments, at least 1, or at least 2, or at least 3, or at least 4, or at least 5 mutations or sets of mutations are selected from Table 39 and which are shown to increase Mog IE titer by at least 1.4 fold or at least 1.5 fold.

In some embodiments, the UGTc3 enzyme is a circular permutant with respect to SEQ ID NO: 60. In some embodiments, the circular permutant has a cut site corresponding to a position

between amino acid 150 to 180, or amino acid 160 to 180, or amino acid 165 to 180 of SEQ ID NO: 60 (e.g., cut site can be at position 169 or 177 in some embodiments). In some embodiments, the circular permutant further comprises a linker sequence between the amino acids that correspond to the N-terminal and C-terminal residues of SEQ ID NO: 60. For example, linkers in various embodiments may have a length of from 2 to 25 amino acids, and can be a rigid linker or a flexible linker. A flexible linker can be comprised predominately or entirely of Ser and Gly amino acids. More rigid linkers can be constructed of bulkier amino acids, and examples are shown in Table 38.

In some embodiments, the UGTc3 enzyme comprises one or more mutations at one or more positions selected from H40, I216, S241, and L352 with respect to SEQ ID NO: 65. In some embodiments, H40 is substituted with Asp, Gly, Gln, Ala or Val. In some embodiments, I216 is substituted with Ala, Gly, or Val. In some embodiments, S241 is substituted with Gly, Ala, or Val. In some embodiments, L352 is substituted with Met. In some embodiments, the UGTc3 enzyme comprises at least one, or at least two, or at least three, or all four of the following substitutions with respect to SEQ ID NO: 65: H40N, I216A, S241G, and L352M. Exemplary enzymes according to these embodiments are provided herein as SEQ ID NO: 66 and SEQ ID NO: 67.

In some aspects and embodiments, the present disclosure provides engineered UGT enzymes that have an increased substrate selectivity for C-24 position of mogrol compared to, e.g., C-24 position of 24,25 dihydroxy cucurbitadienol, or positions other than C-24 position of mogrol. Such engineered enzymes are referred to herein as UGTc24. Accordingly, in some embodiments, the UGTc24 enzymes disclosed herein provide for the production of higher levels of Mog1A compared to a deoxy-Mog1A side product (see, e.g., FIG. 10E). In some embodiments, the UGTc24 enzymes disclosed herein, using a mixture comprising equal amounts of mogrol and deoxy-mogrol (on a molar basis) as a substrate, produce less than 40%, or less than 30%, or less than 20% deoxy-Mog1A side product, when the total glucosylated product is considered 100%. In some embodiments, the UGTc24 enzymes disclosed herein, using a mixture comprising mogrol and deoxy-mogrol as a substrate, produces less off pathway product deoxy-Mog1A compared to a non-engineered UGT enzyme. In some embodiments, the UGTc24 enzyme disclosed herein, using a mixture comprising mogrol and deoxy-mogrol as a substrate, produces more Mog1A compared to the culturing of microbial cells expressing a non-engineered UGT enzyme.

In various aspects and embodiments, the present disclosure provides a UGT enzyme that produces a C-24 mogrol glycoside from mogrol (UGTc24). In some embodiments, the UGTc24 enzyme comprises an amino acid sequence that is at least 70% identical, or at least 80% identical, or at least 85% identical, or at least 90% identical, or at least 95% identical, or at least 97% identical, or at least 98% identical or at least 99% identical to an amino acid sequence selected from SEQ ID NO: 31, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 63, and SEQ ID NO: 64 (see Table 34). In some embodiments, the UGTc24 enzyme comprises an amino acid sequence that is at least 70% identical, or at least 80% identical, or at least 85% identical, or at least 90% identical, or at least 95% identical, or at least 97% identical, or at least 98% identical or at least 99% identical to the amino acid sequence of SEQ ID NO: 31. In some embodiments, the UGTc24 enzyme comprises one or more mutations resulting in greater specificity and/or activity for glycosylation of C-24 position (e.g., including but not limited to conversion of Mog. IE to Mog. IIE and/or conversion of mogrol to Mog. IA) compared to the UGT enzyme having an amino acid sequence of SEQ ID NO: 31. In some embodiments, the UGTc24 enzyme comprises one or more mutations or sets of mutations (e.g., at least 2, or at least 3, or at least 4, or at least 5 mutations or sets of mutations) selected from Table 40, Table 41, Table 42, Table 43, Table 44 and/or Table 45 (in each case with respect to the parent enzyme). In some embodiments, the UGTc24 enzyme comprises one or more mutations or sets of mutations (e.g., at least 2, or at least 3, or at least 4, or at least 5 mutations or sets of mutations) selected from Table 40, Table 41, Table 42, Table 43, Table 44 and/or Table 45 that are shown to increase Mog. IIE titer or Mog. IA titer by at least 1.5 fold or at least 2 fold. In some embodiments, at least one, at least two, at least 3, at least 4, or at least 5 mutations or sets of mutations improve the specificity for conversion to Mog. IA over deoxy-Mog. IA (e.g., see Tables 43 and 44).

In some embodiments, the UGTc24 enzyme comprises one or more mutations at positions selected from L18, A74, S88, I91, T95, H101, T184, F187, P190, T273, N306, M332, and Y381, with respect to SEQ ID NO: 31. In some embodiments, L18 is substituted with Met. In some embodiments, I63 is substituted with Val, Ala or Gly. In some embodiments, A74 is substituted with an acidic amino acid selected from Glu and Asp. In some embodiments, S88 is substituted with an amino acid selected from Ala, Gly, Leu, Val, and Ile. In some embodiments, I91 is substituted with an aromatic amino acid selected from Phe, Tyr and Trp. In some embodiments, T95 is substituted with an amino acid selected from Ala, Gly, and Val. In some embodiments,

H101 is substituted with Pro. In some embodiments, T184 is substituted with an amino acid selected from Phe, Tyr, and Trp. In some embodiments, F187Y is substituted with Tyr or Trp. In some embodiments, P190 is substituted with an amino acid selected from Glu and Asp. In some embodiments, T273 is substituted with Ser or is deleted. In some embodiments, N306 is substituted with Lys, His, or Arg. In some embodiments, M332 is substituted with an amino acid selected from Leu, Ile, Val, and Ala. In some embodiments, Y381 is substituted with an amino acid selected from Phe and Trp. In some embodiments, the UGTc24 enzyme comprises at least one, or at least two, or at least 3, or at least 4, or at least 5, or at least 6, or at least 7, or at least 8, or at least 9, or at least 10, or at least 11, or at least 12, or at least 13, or all 14 mutations selected from L18M, I63V, A74E, S88A, I91F, T95A, H101P, T184F, F187Y, P190E, T273S, N306K, M332L, and Y381F with respect to SEQ ID NO: 31. Exemplary enzymes according to embodiments of this disclosure comprise an amino acid sequence selected from SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 48, and SEQ ID NO: 49.

In various aspects and embodiments, the present disclosure provides a uridine diphosphate dependent glycosyltransferase (UGT) enzyme that produces a 1,6 and/or 1,2 branched glycoside of a C-3 and/or C-24 glycoside of mogrol (UGT1216). In some embodiments, the UGT1216 is a circular permutant of the enzyme represented by SEQ ID NO: 35, or a sequence having at least 70% sequence identity thereto, or at least 90% sequence identity thereto. In some embodiments, the circular permutant has a cut site at a position corresponding to amino acid 180 to 195 of SEQ ID NO: 35, or a sequence having at least 70% identity thereto. In some embodiments, the circular permutant has a cut site at a position corresponding to amino acid 185 to 190 of SEQ ID NO: 35, or a sequence having at least 70% identity thereto. In some embodiments, the circular permutant has a cut site at position 187 of SEQ ID NO: 35, or a sequence having at least 70% identity thereto.

In some embodiments, the UGT1216 enzyme comprises an amino acid sequence that is at least 70%, or at least 80%, or at least 90%, or at least 95%, or at least 97%, or at least 98%, or at least 99% identical to the amino acid sequence of SEQ ID NO: 36. In some embodiments, the UGT1216 enzyme has one or more mutations resulting in greater specificity and/or activity for to 1,6 and/or 1,2 glycosylation of a C-3 and/or C-24 glycoside of mogrol compared to the UGT enzyme having an amino acid sequence of SEQ ID NO: 36. In some embodiments, the UGT1216 enzyme has one or more mutations resulting in greater production of MogV from MogIIE compared to the UGT enzyme having an amino acid sequence of SEQ ID NO: 36. In some

embodiments, the UGT1216 enzyme has one or more mutations or sets of mutations (e.g., at least 2, or at least 3, or at least 4, or at least 5 mutations or sets of mutations) selected from Table 46, Table 47, Table 48, and/or Table 49, with respect to the respective parent enzyme. In some embodiments, the UGT1216 enzyme has one or more mutations or sets of mutations (e.g., at least 2, or at least 3, or at least 4, or at least 5 mutations or sets of mutations) selected from Table 46, Table 47, Table 48, and/or Table 49, with respect to the respective parent enzyme, and which are shown to increase Mog. V titer by at least 1.5 or by at least 2.0.

In some embodiments, the UGT1216 enzyme comprises mutations at one or more positions selected from L14, W160, L266, and H390 with respect to SEQ ID NO: 36. In some embodiments, L14 is substituted with a non-polar amino acid selected from Val, Ile, Ala, and Gly. In some embodiments, W160 is substituted with Pro, Ser, or Thr. In some embodiments, L266 is substituted with a polar amino acid selected from Gln and Asn. In some embodiments, H390 is substituted with an amino acid selected from Asp and Glu. In some embodiments, the UGT1216 enzyme comprises at least one, or at least two, or at least 3, or all 4 mutations selected from L14V, H19E, W160P, L266Q, T281S, S318K and H390D with respect to SEQ ID NO: 36. Exemplary UGT1216 enzymes according to the disclosure comprise an amino acid sequence selected from SEQ ID NO: 37, SEQ ID NO: 38 and SEQ ID NO: 90.

In various aspects and embodiments, the present disclosure provides a uridine diphosphate dependent glycosyltransferase (UGT) enzyme that produces a 1,6 and/or 1,2 branched glycoside of a C-3 and/or C-24 glycoside of mogrol (UGT1216). In some embodiments, the UGT1216 enzyme comprises an amino acid sequence that is at least 70%, or at least 80%, or at least 90%, or at least 95%, or at least 97%, or at least 98%, or at least 99% identical to the amino acid sequence of SEQ ID NO: 39. In some embodiments, the UGT1216 enzyme has one or more mutations resulting in greater specificity and/or activity for 1,6 and 1,2 branched glycosylation of a C-3 and/or C-24 glycoside of mogrol compared to the UGT enzyme having an amino acid sequence of SEQ ID NO: 39. In some embodiments, the UGT1216 enzyme has one or more mutations resulting in greater production of Mog. IVA from Mog. IIE and/or Mog. V from Mog. IIE, compared to the UGT enzyme having an amino acid sequence of SEQ ID NO: 39. In some embodiments, the UGT1216 enzyme has one or more mutations or sets of mutations (e.g., at least 2, at least 3, at least 4, or at least 5 mutations or sets of mutations) listed in Table 50, with respect to SEQ ID NO: 39. In some embodiments, the enzyme has one or more mutations or sets of mutations (e.g., at least

2, at least 3, at least 4, or at least 5 mutations or sets of mutations) listed in Table 50, with respect to SEQ ID NO: 39, and which are shown to provide at least 5 fold Mog. V titer.

In some embodiments, the UGT1216 enzyme comprises mutation(s) at positions selected from T147 and N207 with respect to SEQ ID NO: 39. In some embodiments, T147 is substituted
5 with a non-polar amino acid selected from Leu, Ile, and Val. In some embodiments, N207 is substituted with a basic amino acid selected from Lys, Arg, and His. In some embodiments, the UGT1216 enzyme comprises T147L and/or N207K substitutions with respect to SEQ ID NO: 39.

In various aspects, the present disclosure provides a uridine diphosphate dependent glycosyltransferase (UGT) enzyme that produces a 1,6 and/or 1,2 branched glycoside of a C-3
10 and/or C-24 glycoside of mogrol (UGT1216), wherein the UGT1216 enzyme comprises an amino acid sequence that is at least 80%, or at least 85%, at least 90%, or at least 95%, or at least 97%, or at least 98% identical to the amino acid sequence of SEQ ID NO: 38, and comprising T147L and/or N207K mutations with respect to SEQ ID NO: 39. In some embodiments, the enzyme comprises the amino acid sequence of SEQ ID NO: 40.

15 In various aspects, the present disclosure provides a microbial cell producing squalene and expressing a biosynthetic pathway that converts squalene to mogrol or mogroside, wherein when the microbial cell expresses one or more engineered enzymes of any of the aspects or embodiments disclosed herein.

In various embodiments, the microbial cell further expresses a heterologous farnesyl
20 diphosphate synthase (FPPS) and a squalene synthase enzyme (SQS), converting metabolic precursors of the MEP or MVA pathways to squalene. SQS enzymes and FPPS enzymes are known in the art. Exemplary SQS and FPPS enzymes (as well as other enzymes useful for the production of mogrol and mogrosides in microbial host cells) are described in U.S. 2021/0032669, which is hereby incorporated by reference in its entirety.

25 The microbial host cell in various embodiments may be prokaryotic or eukaryotic. In some embodiments, the microbial host cell is a bacterium, and which can be optionally selected from *Escherichia spp.*, *Bacillus spp.*, *Corynebacterium spp.*, *Rhodobacter spp.*, *Zymomonas spp.*, *Vibrio spp.*, and *Pseudomonas spp.* For example, in some embodiments, the bacterial host cell is a species selected from *Escherichia coli*, *Bacillus subtilis*, *Corynebacterium glutamicum*, *Rhodobacter capsulatus*, *Rhodobacter sphaeroides*, *Zymomonas mobilis*, *Vibrio natriegens*, or *Pseudomonas*
30

putida. In some embodiments, the bacterial host cell is *E. coli*. Alternatively, the microbial cell may be a yeast cell, such as but not limited to a species of *Saccharomyces*, *Pichia*, or *Yarrowia*, including *Saccharomyces cerevisiae*, *Pichia pastoris*, and *Yarrowia lipolytica*.

It was surprisingly found that overexpression of the *ybhGFSR* operon decreases the
5 production of mogrol and its precursor 24,25-dihydroxy cucurbitadienol. On the other hand, hypomorphic or null mutations in the *ybhGFSR* operon (e.g., reducing expression or activity of genes in the *ybhGFSR* operon) increase the production of mogrol and/or its immediate precursor 24,25-dihydroxy cucurbitadienol. It has been found that the overexpression of the *ybhGFSR* operon increases the production of 2,3;22,23-dioxidosqualene, which is the early intermediate of
10 mogrol synthesis. Reduced expression of the *ybhGFSR* transporter increases intracellular (int.) 2,3;22,23-dioxidosqualene, and improves conversion to downstream products 24,25-dihydroxy cucurbitadienol and mogrol.

Accordingly, in various aspects, the present disclosure relates to a microbial cell for producing mogrol or mogroside, wherein the microbial cell has reduced expression or activity of
15 one or more genes of the *ybhGFSR* operon, and wherein the microbial cell expresses a biosynthetic pathway that converts squalene to the mogrol or mogroside. In various embodiments, expression of one or more genes of the *ybhGFSR* operon (e.g., *ybhG* or *F*) is modified by replacing or modifying the native promoter, modifying the RBS, or modifying the gene to increase the degradation rate of the mRNA. In various embodiments, one or more *ybhGFSR* operon genes are
20 engineered to include a hypomorphic mutation to reduce activity. In some embodiments, one or more genes of the *ybhGFSR* operon are disrupted (e.g., deleted) (e.g., *ybhF*).

The YbhGFSR pump is a member of the ATP-binding cassette (ABC) transporter superfamily. The ABC transporter, which are also known as ABC efflux pumps, are very common in all organisms, as they actively transport various molecules, such as ions, xenobiotics, drugs,
25 sugars, amino acids, lipids and oligopeptides, *etc.*, across the lipid membrane using the energy derived from the hydrolysis of ATP to ADP. For example, it has been estimated that the *E. coli* has 79 ABC proteins, including 57 ABC pumps, making it the largest paralogous protein family. Linton, and Higgins, The *Escherichia coli* ATP-binding cassette (ABC) proteins, *Mol. Microbiol.*, 28: 5-13 (1998).

In various aspects and embodiments, the present disclosure provides a microbial cell for producing mogrol or mogroside, wherein the microbial cell harbors one or more genetic modifications that reduces the efflux of an intermediate of mogrol synthesis, the microbial cell thereby increases the efficiency of mogrol synthesis compared to a cell that does not have the genetic modifications. In some embodiments, the microbial cell comprises a mutation that reduces the amount or activity of a transporter. In some embodiments, the transporter is an ABC transporter. In some embodiments, the transporter is YbhGFSR or an ortholog thereof. In some embodiments, the microbial cell expresses a biosynthetic pathway that converts squalene to the mogrol or mogroside. In some embodiments, the microbial cell comprises a hypomorphic mutation or a null mutation in one or more of *ybhG*, *ybhF*, *ybhS*, and *ybhR*.

In some embodiments, the microbial cell comprises a null mutation in one or more of *ybhG*, *ybhF*, *ybhS*, and *ybhR*. In some embodiments, the null mutation is a complete or partial deletion of one or more of *ybhG*, *ybhF*, *ybhS*, and *ybhR*. In some embodiments, the null mutation is an *ybhF* ATPase null mutant created by mutating one or more of the Walker A, Q-loop, and Walker B, and signature motifs of D-loop and Switch motif.

In some embodiments, the microbial cell comprises a hypomorphic mutation in one or more of *ybhG*, *ybhF*, *ybhS*, and *ybhR*. In some embodiments, the hypomorphic mutation is a mutation that reduces expression of one or more genes, such as a promoter mutation, and/or a mutation selected from a mutant ribosome binding site, one or more mutations adding one or more rare codons, and adding a secondary RNA structure in one or more of *ybhG*, *ybhF*, *ybhS*, and *ybhR*. In some embodiments, the RBS for *ybhG* is modified to reduce expression, and/or expression of *ybhF* is disrupted (e.g., *ybhF* is deleted). In some embodiments, the expression and/or activity of one or more of *ybhG*, *ybhF*, *ybhS*, and *ybhR* is decreased using strategies including but not limited to mutating the *ybhGFSR* promoter to a weaker promoter compared to a native promoter, mutating the ribosome binding sites (RBSs) of to render one or more of *ybhG*, *ybhF*, *ybhS*, and *ybhR* mRNA to one or more weaker RBSs compared to the native RBS; mutating one or more of *ybhG*, *ybhF*, *ybhS*, and *ybhR* to include one or more protease recognition sites and thereby increase the degradation rate of the YbhGFSR transporter; mutating one or more of *ybhG*, *ybhF*, *ybhS*, and *ybhR* to include a secondary structure in their mRNA to decrease the rate of translation; mutating one or more of *ybhG*, *ybhF*, *ybhS*, and *ybhR* to include rare codons to decrease the rate of translation, or a combination thereof.

The microbial cell will produce MEP or MVA products, which act as substrates for the heterologous enzyme pathway. The MEP (2-C-methyl-D-erythritol 4-phosphate) pathway, also called the MEP/DOXP (2-C-methyl-D-erythritol 4-phosphate/l-deoxy-D-xylulose 5-phosphate) pathway or the non-mevalonate pathway or the mevalonic acid-independent pathway refers to the pathway that converts glyceraldehyde-3-phosphate and pyruvate to IPP and DMAPP. The pathway, which is present in bacteria, typically involves action of the following enzymes: 1-deoxy-D-xylulose-5-phosphate synthase (Dxs), 1-deoxy-D-xylulose-5-phosphate reductoisomerase (IspC), 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase (IspD), 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (IspE), 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (IspF), 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase (IspG), and isopentenyl diphosphate isomerase (IspH). The MEP pathway, and the genes and enzymes that make up the MEP pathway, are described in US 8,512,988, which is hereby incorporated by reference in its entirety. For example, genes that make up the MEP pathway include dxs, ispC, ispD, ispE, ispF, ispG, ispH, idi, and ispA. In some embodiments, the host cell expresses or overexpresses one or more of dxs, ispC, ispD, ispE, ispF, ispG, ispH, idi, ispA, or modified variants thereof, which results in the increased production of IPP and DMAPP. In some embodiments, the triterpenoid scaffold is produced at least in part by metabolic flux through an MEP pathway, and wherein the host cell has at least one additional gene copy of one or more of dxs, ispC, ispD, ispE, ispF, ispG, ispH, idi, ispA, or modified variants thereof.

The MVA pathway refers to the biosynthetic pathway that converts acetyl-CoA to IPP. The mevalonate pathway, which will be present in yeast, typically comprises enzymes that catalyze the following steps: (a) condensing two molecules of acetyl-CoA to acetoacetyl-CoA (e.g., by action of acetoacetyl-CoA thiolase); (b) condensing acetoacetyl-CoA with acetyl-CoA to form hydroxymethylglutaryl-CoenzymeA (HMG-CoA) (e.g., by action of HMG-CoA synthase (HMGS)); (c) converting HMG-CoA to mevalonate (e.g., by action of HMG-CoA reductase (HMGR)); (d) phosphorylating mevalonate to mevalonate 5-phosphate (e.g., by action of mevalonate kinase (MK)); (e) converting mevalonate 5-phosphate to mevalonate 5-pyrophosphate (e.g., by action of phosphomevalonate kinase (PMK)); and (f) converting mevalonate 5-pyrophosphate to isopentenyl pyrophosphate (e.g., by action of mevalonate pyrophosphate decarboxylase (MPD)). The MVA pathway, and the genes and enzymes that make up the MVA pathway, are described in US 7,667,017, which is hereby incorporated by reference in its entirety.

In some embodiments, the host cell expresses or overexpresses one or more of acetoacetyl-CoA thiolase, HMGS, HMGR, MK, PMK, and MPD or modified variants thereof, which results in the increased production of IPP and DMAPP. In some embodiments, the triterpenoid scaffold is produced at least in part by metabolic flux through an MVA pathway, and wherein the host cell
5 has at least one additional gene copy of one or more of acetoacetyl-CoA thiolase, HMGS, HMGR, MK, PMK, MPD, or modified variants thereof.

In some embodiments, the host cell is a bacterial host cell engineered to increase production of IPP and DMAPP from glucose as described in US 10,480,015 and US 10,662,442, the contents of which are hereby incorporated by reference in their entirety. For example, in some
10 embodiments the host cell overexpresses MEP pathway enzymes, with balanced expression to push/pull carbon flux to IPP and DMAPP. In some embodiments, the host cell is engineered to increase the availability or activity of Fe-S cluster proteins, so as to support higher activity of IspG and IspH, which are Fe-S enzymes. In some embodiments, the host cell is engineered to overexpress IspG and IspH, so as to provide increased carbon flux to 1-hydroxy-2-methyl-2-(E)-
15 butenyl 4-diphosphate (HMBPP) intermediate, but with balanced expression to prevent accumulation of HMBPP at an amount that reduces cell growth or viability, or at an amount that inhibits MEP pathway flux and/or terpenoid production. In some embodiments, the host cell exhibits higher activity of IspH relative to IspG. In some embodiments, the host cell is engineered to downregulate the ubiquinone biosynthesis pathway, e.g., by reducing the expression or activity
20 of IspB, which uses IPP and FPP substrate.

In still other embodiments, microbial cells expressing the mogrol or mogroside biosynthesis pathway co-express an isoprenol utilization pathway as described in US 2019/0367950, which is hereby incorporated by reference in its entirety. Such cells can produce IPP and DMAPP precursors from prenol and/or isoprenol substrate provided to the culture, thereby
25 bypassing the endogenous MEP pathway.

In some embodiments, the microbial cell contains one or more genetic modifications that enhance the supply and transfer of electrons for production of squalene and through the biosynthetic pathway to mogrol. In some embodiments, the microbial cell expresses at least one recombinant electron carrier. Such electron carriers support the function of various enzymes
30 disclosed herein, including enzymes of the MEP pathway. For example, functional expression of

plant cytochrome P450 in *E. coli* has inherent limitations attributable to the bacterial platforms (such as the absence of electron transfer machinery and cytochrome P450 reductases, and translational incompatibility of the membrane signal modules of P450 enzymes due to the lack of an endoplasmic reticulum). Accordingly, in some embodiments, the microbial cell overexpresses at least one endogenous electron carrier, such as one or more oxidoreductase enzymes, including oxidoreductases that oxidize pyruvate and/or lead to reduction of ferredoxin. In various embodiments, the microbial strain comprises an overexpression of, or complementation with, one or more of a flavodoxin (fldA), flavodoxin reductase, ferredoxin (fdx), and ferredoxin reductase. In some embodiments, the microbial cell expresses a heterologous *E. coli* fldA (SEQ ID NO: 43) or heterologous Cv.fdx (an fdx homolog from *Allochromatium vinosum*) (SEQ ID NO: 44), or enzyme having at least 80% or at least 90% or at least 95% or at least 97% sequence identity therewith. In some embodiments, the strain expresses one or more non-native fdx and/or fldA homologs. By way of example, the fdx homolog may be selected from Hm.fdx1 (*Heliobacterium modesticaldum*), Pa.fdx (*Pseudomonas aeruginosa*), Cv.fdx (*Allochromatium vinosum*), Ca.fdx (*Clostridium acetobutylicum*), Cp.fdx (*Clostridium pasteurianum*), Ev2.fdx (*Ectothiorhodospira shaposhnikovii*), Pp1.fdx (*Pseudomonas putida*) and Pp2.fdx (*Pseudomonas putida*). In some embodiments, the fldA homolog includes one or more selected from Ec.fldA (*E. coli*), Ac.fldA2 (*Azotobacter chroococcum*), Av.fldA2 (*Azotobacter vinelandii*), and Bs.fldA (*B. subtilis*). See WO 2018140778, which is hereby incorporated by reference in its entirety.

In some embodiments, the microbial cell is engineered to increase the metabolic supply of NADP by expressing or overexpressing an alternative or exogenous NADPH biosynthesis route. In some embodiments, the microbial cell expresses a bacterial *pntAB* or a homolog, or an ortholog thereof, or a variant thereof. In some embodiments, the microbial cell expresses a hyperactive variant of bacterial *pntAB*, or a homolog, or an ortholog thereof.

In various embodiments, the heterologous enzyme pathway produces mogrol, which may be an intermediate for downstream enzymes in the heterologous pathway, or in some embodiments is recovered from the culture. Mogrol may be recovered from host cells in some embodiments, and/or can be recovered from the culture media.

In some embodiments, the heterologous enzyme pathway further comprises one or more uridine diphosphate-dependent glycosyltransferase (UGT) enzymes, thereby producing one or

more mogrol glycosides (or “mogrosides”). The mogrol glycoside may be pentaglycosylated, hexaglycosylated, or more (e.g., 7, 8, or 9 glycosylations), in some embodiments. In other embodiments, the mogrol glycoside has two, three, or four glycosylations. The one or more mogrol glycosides may be selected from Mog.II-E, Mog.III, Mog.III-A1, Mog.III-A2, Mog.III, Mog.IV, Mog.IV-A, siamenoside, isomog.V, Mog.V, or Mog.VI. In some embodiments, the host cell produces Mog.V or siamenoside.

In some embodiments, the host cell expresses UGT enzymes that catalyze the primary glycosylation of mogrol at C24 and/or C3 hydroxyl groups, as well as one or more branching glycosylations, such as a beta 1,2 and/or beta 1,6 branching glycosylation at the primary C3 and C24 glucosyl groups. In some embodiments, at least one UGT enzyme comprises an amino acid sequence that is at least 70% identical to an amino acid sequence selected from SEQ ID NOs: 27-40 and 48-67. For example, in some embodiments, the UGT enzyme comprises an amino acid sequence that is at least 80%, or at least 85%, or at least 90%, or at least 95%, or at least 98%, or at least 99% identical to one of SEQ ID NOs: 27-40 and 48-67. Thus, at least one UGT enzyme comprises an amino acid sequence having from 1 to 20 amino acid modifications with respect to one of SEQ ID NOs: 27-40 and 48-67, the amino acid modifications being independently selected from amino acid substitutions, deletions, and insertions.

For example, in some embodiments, the microbial cell expresses at least four UGT enzymes, resulting in glucosylation of mogrol at the C3 hydroxyl group, the C24 hydroxyl group, as well as a further 1,6 glucosylation at the C3 glucosyl group, and a further 1,6 glucosylation and a further 1,2 glucosylation at the C24 glucosyl group. The product of such glucosylation reactions is Mog.V.

In some embodiments, the microbial host cell has one or more genetic modifications that increase the production of UDP-glucose, the co-factor employed by UGT enzymes. These genetic modifications may include one or more, or two or more (or all) of $\Delta galE$, $\Delta galT$, $\Delta galK$, $\Delta galM$, $\Delta ushA$, Δagp , Δpgm , duplication of *E. coli galU*, expression of *Bacillus subtilis* UGPA, and expression of *Bifidobacterium adolescentis* SPL.

In various embodiments, the reaction is performed in a microbial cell, and UGT enzymes are recombinantly expressed in the cell. In some embodiments, mogrol is produced in the cell by a heterologous mogrol synthesis pathway, as described herein. In other embodiments, mogrol or

mogrol glycosides (such as a monkfruit extract) are fed to the cells for glycosylation. In still other embodiments, the reaction is performed in vitro using purified UGT enzyme, partially purified UGT enzyme, or recombinant cell lysates.

5 The bacterial host cell is cultured to produce the triterpenoid product (e.g., mogroside). In some embodiments, carbon substrates such as C1, C2, C3, C4, C5, and/or C6 carbon substrates are employed for the production phase. In exemplary embodiments, the carbon source is glucose, sucrose, fructose, xylose, and/or glycerol. Culture conditions are generally selected from aerobic, microaerobic, and anaerobic.

10 In various embodiments, the bacterial host cell may be cultured at a temperature between 22° C and 37° C. While commercial biosynthesis in bacteria such as *E. coli* can be limited by the temperature at which overexpressed and/or foreign enzymes (e.g., enzymes derived from plants) are stable, recombinant enzymes may be engineered to allow for cultures to be maintained at higher temperatures, resulting in higher yields and higher overall productivity. In some embodiments, the culturing is conducted at about 22° C or greater, about 23° C or greater, about 24° C or greater,
15 about 25° C or greater, about 26° C or greater, about 27° C or greater, about 28° C or greater, about 29° C or greater, about 30° C or greater, about 31° C or greater, about 32° C or greater, about 33° C or greater, about 34° C or greater, about 35° C or greater, about 36° C or greater, or about 37° C.

20 In some embodiments, the bacterial host cells are further suitable for commercial production, at commercial scale. 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, or at least about 100,000 L, or at least about 500,000 L, or at least about 600,000 L. In an embodiment, the culturing may be conducted in batch culture, continuous culture, or semi-continuous culture.

25 In various embodiments, methods further include recovering the 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 of the terpenoid or terpenoid glycoside product.

30 In some embodiments, the production of indole (including prenylated indole) is used as a surrogate marker for terpenoid production, and/or the accumulation of indole in the culture is

controlled to increase production. For example, in various embodiments, accumulation of indole in the culture is controlled to below about 100 mg/L, or below about 75 mg/L, or below about 50 mg/L, or below about 25 mg/L, or below about 10 mg/L. The accumulation of indole can be controlled by balancing protein expression and activity using the multivariate modular approach as described in U.S. Pat. No. 8,927,241 (which is hereby incorporated by reference in its entirety), and/or is controlled by chemical means.

Other markers for efficient production of terpene and terpenoids, include accumulation of DOX or ME in the culture media. Generally, the bacterial strains may be engineered to accumulate less of these chemical species, which accumulate in the culture at less than about 5 g/L, or less than about 4 g/L, or less than about 3 g/L, or less than about 2 g/L, or less than about 1 g/L, or less than about 500 mg/L, or less than about 100 mg/L.

The optimization of terpene or terpenoid production by manipulation of MEP pathway genes, as well as manipulation of the upstream and downstream pathways, is not expected to be a simple linear or additive process. Rather, through combinatorial analysis, optimization is achieved through balancing components of the MEP pathway, as well as upstream and downstream pathways. Indole (including prenylated indole) accumulation and MEP metabolite accumulation (e.g., DOX, ME, MEcPP, and/or farnesol) in the culture can be used as surrogate markers to guide this process.

For example, in some embodiments, the bacterial strain has at least one additional copy of *dxs* and *idi* expressed as an operon/module; or *dxs*, *ispD*, *ispF*, and *idi* expressed as an operon or module (either on a plasmid or integrated into the genome), with additional MEP pathway complementation described herein to improve MEP carbon. For example, the bacterial strain may have a further copy of *dxr*, and *ispG* and/or *ispH*, optionally with a further copy of *ispE* and/or *idi*, with expressions of these genes tuned to increase MEP carbon and/or improve terpene or terpenoid titer. In various embodiments, the bacterial strain has a further copy of at least *dxr*, *ispE*, *ispG* and *ispH*, optionally with a further copy of *idi*, with expressions of these genes tuned to increase MEP carbon and/or improve terpene or terpenoid titer.

Manipulation of the expression of genes and/or proteins, including gene modules, can be achieved through various methods. 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, expression of genes or operons can be regulated through manipulating
5 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 protein expression can also be achieved through selection of appropriate promoters and ribosomal binding sites. In some embodiments, this may include the selection of
10 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.

Expression vectors containing all the necessary elements for expression are commercially available and known to those skilled in the art. See, e.g., Sambrook et al., *Molecular Cloning: A
15 Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, 1989. Cells are genetically engineered by the introduction into the cells of heterologous DNA. The heterologous DNA is placed under operable control of transcriptional elements to permit the expression of the heterologous DNA in the host cell.

In some embodiments, endogenous genes are edited, as opposed to gene complementation.
20 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.

25 In some embodiments, genes are overexpressed at least in part by controlling gene copy number. While gene copy number can be conveniently controlled using plasmids with varying copy number, gene duplication and chromosomal integration can also be employed. For example, a process for genetically stable tandem gene duplication is described in US 2011/0236927, which is hereby incorporated by reference in its entirety.

Mogrol glycosides can be recovered from the microbial culture. For example, mogrol glycosides may be recovered from microbial cells, or in some embodiments, are predominately available in the extracellular media, where they may be recovered or sequestered. In some embodiments, the mogrosides are recovered substantially as described in WO 2022/115527, which is hereby incorporated by reference in its entirety. For example, recovery can comprise adjusting the pH of the culture or media to below about pH 5 or above about 10, raising the temperature to at least about 50 °C, optionally adding one or more glycoside solubilizers; followed by biomass removal. By adjusting the properties of the culture or media, prior to biomass removal, it is possible to produce a product with desirable qualities, including: high purity of glycoside product, attractive color, easy solubilization, odorless, and/or high recovery yield. For example, initial pH and temperature adjustment of the culture can change fluid characteristics of the broth, and increase efficiency of a disc stack separator for biomass removal. Further, solubility and therefore yield of glycoside product can be substantially increased by the pH and/or temperature adjustment, and/or addition of a glycoside solubilizer.

In some embodiments, recovery comprises adding one or more glycoside solubility enhancers. Exemplary solubility enhancers include chemical reagents with alcohol functional groups (including organic acids and polymers) and/or polar reagents (including those with ether, ester, aldehyde, and ketone functional groups), and including but not limited to glycerol, 1,3-propanediol, polyvinyl alcohol, polyethylene glycol, among others. Other exemplary solubility enhancers include organic acids, saccharides, and polysaccharides. Other solubility enhancers are described in US 2020/0268026, which is hereby incorporated by reference in its entirety. Improvement in glycoside solubility facilitates biomass removal without large loss in product. Generally, solubility enhancers can be added to the harvested culture material in a range of from about 0.1 wt% to about 2 wt%, such as in a range of from about 0.1 wt% to about 1 wt% (e.g., about 0.5 wt%).

Subsequently, biomass is removed by centrifugation, thereby preparing a clarified broth. An exemplary process for biomass removal employs a disc stack centrifuge to separate liquid and solid phases. The clarified broth (liquid phase) is recovered for further processing to purify the glycoside product. The separated biomass (solid phase) can be reprocessed for further glycoside product recovery, or is alternatively processed as waste.

In some embodiments, glycosides are crystallized from the clarified broth. In some embodiments, the process includes 1, 2, or 3 crystallization steps. In some embodiments, glycoside products are purified from the clarified broth using one or more processes selected from filtration, ion exchange, activated charcoal, bentonite, affinity chromatography, and digestion, which can optionally be conducted prior to crystallization and/or prior to recrystallization. These processes can be selected to achieve a high product purity, attractive color, easy solubilization, odorless character, and high recovery yield. In some embodiments, the method employs affinity chromatography, such as with one or more of a styrene-divinylbenzene adsorbent resin, a strongly acidic cation exchange resin, a weakly acidic cation exchange resin, a strongly basic anion exchange resin, a weakly basic anion exchange resin, and a hydrophobic interaction resin. In still other embodiments, the process employs simulated moving bed chromatography, as described in US Patent 10,213,707, which is hereby incorporated by reference in its entirety. In still other embodiments, the recovery process is non-chromatographic (i.e., there are no chromatographic steps), providing substantial cost advantages. For example, the recovery process after biomass removal can consist essentially of, or consist of, filtration and crystallization steps. In some embodiments, the recovery process employs organic solvents (e.g., ethanol), but in other embodiments the process is entirely with aqueous solvents. In some embodiments, two crystallization steps are employed.

In some embodiments, the recovery process will include one or more steps of tangential flow filtration (TFF). For example, TFF with a filter having a pore size of about 5 kD can remove endotoxin, large proteins, and other cell debris, while also enhancing solubility of the final powdered product. In some embodiments, prior to initial crystallization, glycoside products are purified by tangential flow filtration, optionally having a membrane pore size of about 5 kD. TFF with a filter having a pore size of about 0.5 kD can also be employed downstream to remove small molecule impurities and salts, and/or to concentrate the mother liquor for recrystallization. In some embodiments, TFF with a pore size of about 0.5 kD is employed prior to recrystallization.

In each case, crystallization steps can include one or more phases of static crystallization, stirred crystallization, and evaporative crystallization. For example, crystallization steps may comprise a static phase followed by a stirred phase, to control crystal morphology. The static phase can grow large crystals with a high degree of crystalline domains. The crystallization process can

include seeding crystals, or in some embodiments, does not involve seeding crystals (i.e., crystals form spontaneously). In various embodiments, the crystallization solvent comprises water or water/ethanol. Exemplary crystallization solvents include water, optionally with from about 5% to about 50% ethanol by volume, or from about 25% to about 50% ethanol by volume (e.g., from about 30% to about 40% ethanol by volume). In some embodiments, after seeding crystals during a static phase, a stirred phase will rapidly grow the crystals, and increase the degree of amorphous domains. Using this process, resulting crystals may have better final solubility and a high purity of glycoside product, and may be easier to recover and wash.

In various embodiments, prior to recrystallization, glycoside products are resolubilized in a solvent (such as but not limited to water and/or ethanol), which may employ one or more of: lowering the pH of the solvent and glycoside product solution or suspension to below about pH 5 or raising the pH of the solution or suspension to above about pH 9, heating to at least about 50 °C, and adding one or more glycoside solubilizers. The targeted values for pH, temperature, glycoside solubilizer concentration can alternatively be as employed for biomass removal. For example, the glycoside product solution or suspension may be pH adjusted within the range of about 2 to about 5. In other embodiments, the pH is adjusted to the basic pH range, such as a pH within the range of about 9 to about 12, or within the range of about 9.5 or about 10 to about 12. pH adjustment may be by addition or titration of an organic or inorganic acid or hydroxide ions, according to known methods. In still other embodiments, recrystallization is performed at a pH of about 4 to about 12. Alternatively or in addition, the temperature of the solution or suspension is adjusted to a temperature between about 50°C and about 90°C, such as from about 50°C to about 80°C. Exemplary recrystallization solvents include water, optionally with from about 5% to about 50% ethanol by volume, or from about 25% to about 50% ethanol by volume (e.g., about 30% to about 40% ethanol by volume). Alternatively or in addition, solubility enhancers can be added to the solution/suspension in a range of from about 0.1wt% to about 2 wt%, such as in a range of from about 0.1 wt% to about 1 wt% (e.g., about 0.5 wt%), as described. Exemplary solubility enhancers include glycerol.

In some embodiments, after crystallization, resulting crystals are isolated, e.g., using basket centrifuges or belt filter, thereby isolating glycoside wet cake (e.g., mogrol glycoside wet cake). Washing at basket centrifuge steps can employ washes with water, or alternatively other rinses can

be employed (e.g., chilled water/ethanol). In some embodiments, the cake is dissolved and recrystallized. The wet cake from recrystallization may then be dried, optionally using a belt dryer, paddle dryer, or spray dryer. The dried cake can be milled and packaged.

Prior to recrystallization, the glycoside solution can be filtered to remove impurities. The
5 filter can be about a 0.2 micron filter in some embodiments. Alternatively, other pore sizes can be employed, such as about 0.45 micron filters and about 1.2 micron filters. In accordance with embodiments, the material of the filter can be selected to further remove impurities, such as by adsorption. For example, hydrophilic materials such as polyethersulfone (PES) have significant advantages over more hydrophobic materials such as polypropylene. Other exemplary hydrophilic
10 filter materials include nylon, cellulose acetate, cellulose nitrate, and normally hydrophobic materials that have been functionalized to result in a hydrophilic material (such PTFE or PVDF coated with fluoroalkyl terminated polyethylene glycol).

The similarity of nucleotide and amino acid sequences, i.e. the percentage of sequence identity, can be determined via sequence alignments. Such alignments can be carried out with
15 several art-known algorithms, such as with the mathematical algorithm of Karlin and Altschul (Karlin & Altschul (1993) Proc. Natl. Acad. Sci. USA 90: 5873-5877), with hmalign (HMMER package, <http://hmmer.wustl.edu/>) 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
20 algorithm is incorporated into the BLASTN and BLASTP programs of Altschul et al (1990) J. Mol. Biol. 215: 403-410. BLAST polynucleotide searches can be performed with the BLASTN program, score=100, word length=12.

BLAST protein searches may be performed with the BLASTP program, score=50, word length=3. To obtain gapped alignments for comparative purposes, Gapped BLAST is utilized as
25 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. Sequence matching analysis may be supplemented by established homology mapping techniques like Shuffle-LAGAN (Brudno M., Bioinformatics 2003b, 19 Suppl 1:154-162) or Markov random fields.

"Conservative substitutions" may be made, for instance, on the basis of similarity in polarity, charge, size, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the amino acid residues involved. The 20 naturally occurring amino acids can be grouped into the following six standard amino acid groups:

- 5 (1) hydrophobic: Met, Ala, Val, Leu, Ile;
- (2) neutral hydrophilic: Cys, Ser, Thr, Asn, Gln;
- (3) acidic: Asp, Glu;
- (4) basic: His, Lys, Arg;
- (5) residues that influence chain orientation: Gly, Pro; and
- 10 (6) aromatic: Trp, Tyr, Phe.

As used herein, "conservative substitutions" are defined as exchanges of an amino acid by another amino acid listed within the same group of the six standard amino acid groups shown above. For example, the exchange of Asp by Glu retains one negative charge in the so modified polypeptide. In addition, glycine and proline may be substituted for one another based on their

15 ability to disrupt α -helices. Some preferred conservative substitutions within the above six groups are exchanges within the following sub-groups: (i) Ala, Val, Leu and Ile; (ii) Ser and Thr; (ii) Asn and Gln; (iv) Lys and Arg; and (v) Tyr and Phe.

As used herein, "non-conservative substitutions" are defined as exchanges of an amino acid by another amino acid listed in a different group of the six standard amino acid groups (1) to

20 (6) shown above.

Modifications of enzymes as described herein can include conservative and/or non-conservative mutations. In some embodiments, an Alanine is substituted or inserted at position 2, to increase stability.

In some embodiments "rational design" is involved in constructing specific mutations in

25 enzymes. Rational design refers to incorporating knowledge of the enzyme, or related enzymes, such as its reaction thermodynamics and kinetics, its three dimensional structure, its active site(s), its substrate(s) and/or the interaction between the enzyme and substrate, into the design of the specific mutation. Based on a rational design approach, mutations can be created in an enzyme which can then be screened for increased production of a terpene or terpenoid relative to control

30 levels. In some embodiments, mutations can be rationally designed based on homology modeling.

As used herein, “homology modeling” refers to the process of constructing an atomic resolution model of one protein from its amino acid sequence and a three-dimensional structure of a related homologous protein.

In other aspects, the invention provides a method for making a product comprising a mogrol glycoside. The method comprises producing a mogrol glycoside in accordance with this disclosure, and incorporating the mogrol glycoside into a product. In some embodiments, the mogrol glycoside is siamenoside, Mog.V, Mog.VI, or Isomog.V. In some embodiments, the product is a sweetener composition, flavoring composition, food, beverage, chewing gum, texturant, pharmaceutical composition, tobacco product, nutraceutical composition, or oral hygiene composition.

The product may be a sweetener composition comprising a blend of artificial and/or natural sweeteners. For example, the composition may further comprise one or more of a steviol glycoside, aspartame, and neotame. Exemplary steviol glycosides comprises one or more of RebM, RebB, RebD, RebA, RebE, and RebI.

Non-limiting examples of flavors for which the products 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 starches, gums, texturizers, preservatives, antioxidants, emulsifiers, stabilizers, thickeners and gelling agents.

Mogrol glycosides obtained according to this invention may be incorporated as a high intensity natural sweetener in foodstuffs, beverages, pharmaceutical compositions, cosmetics, chewing gums, table top products, cereals, dairy products, toothpastes and other oral cavity compositions, etc.

Mogrol glycosides obtained according to this invention 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; vitamins; and minerals. Functional ingredients also may be

classified based on their health benefits, such as cardiovascular, cholesterol-reducing, and anti-inflammatory.

Mogrol glycosides 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 characteristics. It may also be used in drinks, foodstuffs, pharmaceuticals, and other products in which sugar cannot be used. In addition, highly purified target mogrol glycoside(s), particularly, Mog.V, Mog.VI, or Isomog.V, 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 mogrol glycoside(s) may be used as a sweetening compound 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 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; 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.

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.

As used in this specification and the appended claims, the singular forms “a”, “an” and “the” include plural referents unless the content clearly dictates otherwise. For example, reference to “a cell” includes a combination of two or more cells, and the like.

As used herein, the term “about” in reference to a number is generally taken to include numbers that fall within a range of 10% in either direction (greater than or less than) of the number.

Any aspect or embodiment disclosed herein may be combined with any aspect or embodiment disclosed herein.

EXAMPLES

FIG. 1A and **FIG. 1B** show an *in vivo* biosynthetic path to mogrol and to MogV that comprises condensation of farnesyl pyrophosphate to squalene, which is catalyzed by a squalene synthase (SQS); epoxidation of squalene to 2,3-oxidosqualene and 2,3;22,23-dioxidosqualene which is catalyzed by a squalene epoxidase (SQE); cyclization of 2,3;22,23-dioxidosqualene to 24,25-epoxy cucurbitadienol, which is catalyzed by an epoxy cucurbitadienol synthase (ECDS); hydration of 24,25-epoxy cucurbitadienol to 24,25-dihydroxy cucurbitadienol, which is catalyzed by an epoxide hydrolase (EPH); hydroxylation of 24,25-dihydroxy cucurbitadienol to mogrol, which are catalyzed by a cytochrome P450 and a cytochrome P450 reductase that are active on C11 position of 24,25-dihydroxy cucurbitadienol; and C3 glycosylation, C24 glycosylation, 1,6-glycosylation, and 1,2-glycosylation of mogrol, which are catalyzed by uridine diphosphate dependent glycosyltransferases (UGTs, see **FIG. 1A** and **FIG. 1B**).

As illustrated in **FIG. 1A**, the enzymes such as CDS, EPH, and UGTs, in addition to catalyzing the desired conversions, can divert the on pathway products to undesired off pathway products such as cucurbitadienol, hydrolyzed 2,3;22,23-dioxidosqualene, and glycosides of 24,25 dihydroxy cucurbitadienol. A bacterial cell expressing unengineered SQS, a SQE, a CDS, and an EPH generally produces about 90% or more of off pathway products, and relatively low amounts of mogrol. Disclosed herein are engineered enzymes showing greater substrate specificity and/or activity to shift production to on pathway products.

Example 1: Engineering of Squalene Epoxidase (SQE) to Improve 2,3;22,23-dioxidosqualene Production

Using an *E. coli* strain that produces high levels of the MEP pathway products IPP and DMAPP (see US 2018/0245103 and US 2018/0216137, which are hereby incorporated by reference), and with expression of ScFPPS and a squalene synthase (e.g., SEQ ID NO: 13), downstream enzymes were screened for their ability to convert squalene to on-pathway products. See **FIGS. 1A** and **1B**.

SQE_1 (SEQ ID NO: 2), which is a *Methylomonas lenta* SQE (SEQ ID NO: 1) derivative, has the following mutations H36R, F164A, M284L, V381L, F396Y with respect to SEQ ID NO: 1. This derivative is disclosed in WO 2021/126960, the entire contents of which are hereby incorporated by reference in their entirety. SQE_1 was further improved for the production of 2,3;22,23-dioxidosqualene.

E. coli strains producing squalene and expressing numerous SQE_1 mutant derivatives were screened in Round 1. For screening, fermentations were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by gas chromatography coupled with flame ionization detection (GC-FID) for 2,3-oxidosqualene, and 2,3;22,23-dioxidosqualene production. Effects of the amino acid substitutions were evaluated by comparing the titers of 2,3-oxidosqualene, and 2,3;22,23-dioxidosqualene produced in fermentations by bacterial strain expressing a given SQE derivative with bacterial strain expressing SQE_1. Table 1 shows fifty sets of amino acid substitutions that showed beneficial changes in the titers of 2,3-oxidosqualene and/or 2,3;22,23-dioxidosqualene compared to SQE_1:

Table 1: Round 1 of SQE Engineering

Mutations	2,3-dioxidosqualene titer	2,3;22,23-dioxidosqualene titer
A164N	1.37	0.74
A169R	1.71	0.94
N345S	1.16	0.67
I145H	1.55	0.94
A169E	1.44	0.93
Q175E	1.61	1.06
A348S	1.21	0.80
S116E	1.36	1.02
N134G	1.32	0.99

E114Q	1.37	1.03
H389P	1.17	0.95
L392V	1.18	0.97
Q410K	1.30	1.09
Q409R	1.17	0.98
H389P/K388N	1.24	1.06
K388N	1.20	1.02
L385I	1.15	0.98
C405G	1.17	1.02
F390Y	1.26	1.10
F398G	1.33	1.17
L408I	1.23	1.09
K388R	1.19	1.06
I393M	1.25	1.12
R36Q	1.18	1.09
I393L	1.10	1.02
I393V	1.22	1.14
A350V	1.13	1.06
A400L	1.19	1.17
N349G	1.25	1.22
I346T	1.12	1.11
A288S	1.38	1.39
I259L	1.19	1.20

K338Q	1.14	1.16
K321N	1.10	1.13
T323P	1.17	1.20
V278I	1.15	1.19
S253N	1.14	1.19
I393A	1.09	1.14
R339H	1.06	1.11
E261A	1.11	1.18
Y255F	1.10	1.17
C250E	1.09	1.19
F201I	1.06	1.16
L392A	1.19	1.31
I393T	1.12	1.26
Q395S	0.98	1.12
N233G	0.98	1.13
L202V	1.00	1.30
V200M	0.95	1.35
Y367A	0.52	1.19

Based on the above results, combination mutations were generated. SQE_2 (SEQ ID NO: 3), having the mutations S116E, N134G, I145H, A169R, A400V, was further studied in comparison with SQE_1. Fermentations using bacterial strains producing squalene and expressing SQE_1 or SQE_2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by GC-FID for 2,3;22,23-dioxidosqualene production and normalized to the level of 2,3;22,23-dioxidosqualene production by bacterial strains expressing SQE_1. As shown in **FIG. 2A**, the

bacterial strains expressing SQE_2 produced about 5-fold increased levels of 2,3;22,23-dioxidosqualene compared to the bacterial strains expressing SQE_1.

E. coli strains expressing further SQE mutant derivatives were screened in Round 2. Some of these mutant derivatives included clusters of mutations (see, *e.g.*, multiple mutations as shown in several rows of Table 2). For screening, fermentations were carried out in a 96-well plate at 37°C for 72 hrs using bacterial strains producing squalene and expressing various SQE_2 derivatives. Data were quantified by GC-FID for 2,3-oxidosqualene, and 2,3;22,23-dioxidosqualene production. Effects of the amino acid substitutions were evaluated by comparing the titers of 2,3-oxidosqualene, and 2,3;22,23-dioxidosqualene produced in fermentations by bacterial strain expressing a given SQE derivative with bacterial strain expressing SQE_2. Table 2 shows fourteen mutants that showed beneficial changes in the titers of 2,3-oxidosqualene and/or 2,3;22,23-dioxidosqualene compared to SQE_2:

Table 2: Round 2 of SQE Engineering

Mutations	2,3-dioxidosqualene titer	2,3;22,23-dioxidosqualene titer
E114Q/Q175E/F390Y/L408I	1.26	1.29
388N/F390Y	1.19	1.16
C405G	1.13	1.07
F390Y/I393M	1.12	0.97
I393L	1.11	0.87
R36Q/A164N/Q410K	1.10	1.12
F398G	1.07	1.22
Q175E/I393M/Q410K	1.06	1.10
A288S/F390Y/L408I	1.01	1.30
338Q/K388N/F390Y	1.00	1.22
H389P/K388N	0.91	1.10

H389P/K388N/F398G	0.87	1.22
F398G/Q410K	0.76	1.34
L392A	0.68	1.14

The SQE_2 sequence was recoded without making any further changes to amino acid sequence. The resultant sequence was called SQE_3. Fermentations using bacterial strains producing squalene and expressing SQE_2 or SQE_3 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by GC-FID for 2,3;22,23-dioxidosqualene production and normalized to the level of 2,3;22,23-dioxidosqualene production by bacterial strains expressing SQE_2. As shown in **FIG. 2B**, the bacterial strains expressing SQE_3 produced more than 2-fold increased levels of 2,3;22,23-dioxidosqualene compared to the bacterial strains expressing SQE_2.

E. coli strains expressing numerous SQE mutant derivatives were screened in Round 3. Some of these mutant derivatives included clusters of mutations (see, e.g., multiple mutations as shown in several rows of Table 3). For screening, fermentations were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by GC-FID for 2,3-oxidosqualene, and 2,3;22,23-dioxidosqualene production. Effects of the amino acid substitutions were evaluated by comparing the titers of 2,3-oxidosqualene, and 2,3;22,23-dioxidosqualene produced in fermentations by bacterial strains expressing a given SQE derivative with a bacterial strain expressing SQE_3. Table 3 shows eight mutants that showed beneficial changes in the titers of 2,3-oxidosqualene and/or 2,3;22,23-dioxidosqualene compared to SQE_3:

Table 3: Round 3 of SQE Engineering

Mutations	2,3-dioxidosqualene titer	2,3;22,23-dioxidosqualene titer
SQE3_A164N	1.47	1.06
SQE3_Q175E	1.30	1.09
SQE3_E114Q/A164N	1.46	1.11
SQE3_E114Q/Q175E	1.30	1.15

SQE3_A164N/Q175E	1.47	1.02
SQE3_E114Q/A164N/Q175E	1.55	1.09
SQE2_E114Q_Q175E	0.54	1.17
SQE2_E114Q/A164N/Q175E	0.75	1.25

Based on these results, SQE_4 (SEQ ID NO: 4), having the mutation A164N, in addition to the mutations present in SQE_3, was further studied in comparison with SQE_3. Fermentations using bacterial strains expressing SQE_3 or SQE_4 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by GC-FID for 2,3;22,23-dioxidosqualene production and normalized to the level of 2,3;22,23-dioxidosqualene production by bacterial strains expressing SQE_3. As shown in **FIG. 2C**, the bacterial strains expressing SQE_4 produced increased levels of 2,3;22,23-dioxidosqualene compared to the bacterial strains expressing SQE_3.

Example 2: Engineering of McSQE to Improve ,3;22,23-Dioxidosqualene Production

E. coli strains producing squalene and expressing mutant derivatives of *Methylococcus capsulatus* SQE (McSQE_0, SEQ ID NO: 77) were screened in Round 1. For screening, fermentations were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by liquid chromatography method with diode array detection (UPLC-DAD) for 2,3;22,23-dioxidosqualene production. Effects of the amino acid substitutions were evaluated by comparing the titer of 2,3;22,23-dioxidosqualene produced in fermentations by bacterial strain expressing a given SQE derivative with bacterial strain expressing McSQE_0. Table 4 shows 10 mutants that showed beneficial changes in the titers of 2,3;22,23-dioxidosqualene compared to McSQE_0:

Table 4: Round 1 of McSQE Engineering

Mutations	2,3;22,23-dioxidosqualene titer
S366R	3.80
I394P	3.20
G17A	2.74
T389G	2.06
D41H	1.84

I53L	1.21
M315L	1.20
M174N	1.15
G405A	1.15
G405V	1.11

McSQE_1 (SEQ ID NO: 78), having the mutation S366R with respect to SEQ ID NO: 77, was further studied in comparison with McSQE_0. Fermentations using bacterial strains producing squalene and expressing McSQE_0 or McSQE_1 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by UPLC-DAD for 2,3;22,23-dioxidosqualene production and normalized to the level of 2,3;22,23-dioxidosqualene production by bacterial strains expressing McSQE_0. As shown in **FIG. 2D**, the bacterial strains expressing McSQE_1 produced about 4-fold increased levels of 2,3;22,23-dioxidosqualene compared to the bacterial strains expressing McSQE_0.

E. coli strains producing squalene and expressing mutant derivatives of McSQE_1 were screened in Round 2. McSQE_2 (SEQ ID NO: 79), having the mutation D41H and M315L with respect to SEQ ID NO: 77, in addition to the mutation in McSQE_1 was further studied in comparison with McSQE_1. Fermentations using bacterial strains producing squalene and expressing McSQE_1 or McSQE_2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by UPLC-DAD for 2,3;22,23-dioxidosqualene production and normalized to the level of 2,3;22,23-dioxidosqualene production by bacterial strains expressing McSQE_0. As shown in **FIG. 2E**, the bacterial strains expressing McSQE_1 produced about 1.4-fold increased levels of 2,3;22,23-dioxidosqualene compared to the bacterial strains expressing McSQE_1.

E. coli strains producing squalene and expressing mutant derivatives of McSQE_2 were screened in Round 3. For screening, fermentations were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by UPLC-DAD for 2,3;22,23-dioxidosqualene production. Effects of the amino acid substitutions were evaluated by comparing the titer of 2,3;22,23-dioxidosqualene produced in fermentations by bacterial strain expressing a given SQE derivative with bacterial strain expressing McSQE_2. Table 5 shows mutants that showed beneficial changes in the titers of 2,3;22,23-dioxidosqualene compared to McSQE_2:

Table 5: Round 1 of McSQE Engineering

Mutations	2,3;22,23-dioxidosqualene titer
M232L	1.26
E112H	1.16
A273G	1.13
L48I	1.12
H68D	1.10
V38L	1.10

These results demonstrate that the above mutant derivatives provide a further improvement in the titers of 2,3;22,23-dioxidosqualene compared to McSQE_2.

Example 3: Engineering of Epoxide Hydrolase (EPH) to Improve 24,25-Dihydroxy-Cucurbitadienol Production

E. coli strains expressing amino acid substitution derivatives of *Siraitia grosvenorii* EPH3 (EPH_0; SEQ ID NO: 5) were screened. *Siraitia grosvenorii* EPH3 is disclosed in WO 2019169027, which is hereby incorporated by reference in its entirety. A bacterial strain that produces 24,25-epoxycucurbitadienol and expresses variants of EPH_0 were used for screening. Briefly, fermentations were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by gas chromatography coupled with flame ionization detection (GC-FID) for 24,25-dihydroxy-cucurbitadienol production. Effects of the amino acid substitutions were evaluated by comparing the titers of 24,25-dihydroxy-cucurbitadienol produced in fermentations by bacterial strain expressing a given EPH_0 derivative with bacterial strain expressing EPH_0. Table 6 shows twenty-two amino acid substitutions that showed beneficial changes in the titers of 24,25-dihydroxy-cucurbitadienol compared to EPH_0:

Table 6: EPH Engineering

Mutations	24,25-dihydroxy-cucurbitadienol titer
S68A	1.52
L78F	1.46
V127M	1.28

Q128A	2.19
Q128P	1.46
I130M	1.61
I130L	1.29
P137S	1.14
G141S	1.59
T144Q	1.89
T144A	1.41
A145V	1.43
E163A	1.59
E191G	1.70
F234M	1.42
S260L	1.26
L262M	1.40
F266M	1.13
V291L	1.34
D293G	1.28
C295G	2.25
N299Q	2.44

Based on these results, EPH_1 has the following mutations: S68A, Q128A, G141S, T144Q, A145V, E163A, E191G, L262M, C295G, N299Q with respect to EPH_0 was constructed. EPH_1 (SEQ ID NO: 6), was further studied. Fermentations using bacterial strains that produce 24,25-epoxycucurbitadienol and express EPH_0 or EPH_1 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by GC-FID for 24,25-dihydroxy-cucurbitadienol production. As shown in **FIG. 3**, the bacterial strains expressing EPH_1 produced about 3.5-fold increased levels of 24,25-dihydroxy-cucurbitadienol compared to the bacterial strains expressing EPH_0.

Example 4: Engineering of 24, 25 Epoxy Cucurbitadienol Synthase (ECDS) to Improve 24,25-Epoxy Cucurbitadienol Production

E. coli strains expressing mutant derivatives of *Citrullus colocynthis* cucurbitadienol synthase 2 (CcCDS_0 or ECDS_0; SEQ ID NO: 7) were screened. *Citrullus colocynthis* cucurbitadienol synthase 2 is disclosed in WO 2021/126960, which is hereby incorporated by reference in its entirety. Briefly, fermentations were carried out in a 96-well plate at 37°C for 72 hrs using bacterial strains that produce 2,3;22,23-dioxidosqualene and expresses ECDS_0 or derivatives thereof. Data were quantified by GC-FID for 24,25-epoxy-cucurbitadienol production. Effects of the amino acid substitutions were evaluated by comparing the titers of cucurbitadienol and 24,25-epoxy-cucurbitadienol produced in fermentations by bacterial strain expressing a given ECDS_0 derivative with bacterial strain expressing ECDS_0. Table 7 shows fourteen amino acid substitutions that showed beneficial changes in the titers of 24,25-epoxy-cucurbitadienol compared to ECDS_0:

Table 7: Round 1 of ECDS Engineering

Mutations	Cucurbitadienol titer	24,25-epoxy-cucurbitadienol titer
S24N	1.40	1.17
C35D	1.42	1.16
V105L	1.16	1.04
C267T	1.10	1.01
I334L	1.29	1.16
R415K	1.18	1.08
M503L	1.13	1.06
C522Y	1.31	1.08
A556S	2.05	1.44
T558V	0.93	1.14

G560A	1.00	1.14
T575A	1.41	1.11
L605I	1.17	1.09
C618S	1.32	1.18

Based on these results, ECDS_1 (SEQ ID NO: 8), which harbors S24N, C35D and A556S substitutions, was further generated and studied in comparison with ECDS_0. Fermentations using bacterial strains that produce 2,3;22,23-dioxidosqualene and express ECDS_0 or ECDS_1 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by GC-FID for 24,25-epoxy cucurbitadienol production. As shown in **FIG. 4A**, the bacterial strains expressing ECDS_1 produced about 6-fold increased levels of 24,25-epoxy cucurbitadienol compared to the bacterial strains expressing ECDS_0.

E. coli strains expressing further mutant derivatives of ECDS_1 were constructed and screened as discussed above, using bacterial strains that produce 2,3;22,23-dioxidosqualene and express mutant derivatives of ECDS_1. Based on these results, ECDS_2 (SEQ ID NO: 9), which harbors I490V and I553M substitutions in addition to the substitutions present in ECDS_1, was further studied in comparison with ECDS_1. Fermentations were carried out in a 96-well plate at 37°C for 72 hrs using bacterial strains that produce 2,3;22,23-dioxidosqualene and express ECDS_1 or ECDS_2. Data were quantified by GC-FID for cucurbitadienol and 24,25-epoxy-cucurbitadienol production. As shown in **FIG. 4B**, the bacterial strains expressing ECDS_2 produced about 3.5-fold increased levels of 24,25-epoxy-cucurbitadienol, while producing less than about 2-fold increased levels of cucurbitadienol, compared to the bacterial strains expressing ECDS_1. These results demonstrate, *inter alia*, that ECDS_2 has increased activity or specificity towards 2,3;22,23-dioxidosqualene substrate compared to 2,3-oxidosqualene substrate.

E. coli strains expressing further mutants of ECDS_2 were screened using bacterial strains that produce 2,3;22,23-dioxidosqualene and express mutant derivatives of ECDS_2. Based on these results, ECDS_3 (SEQ ID NO: 10), which harbors D50E, N121H, and Q401A substitutions in addition to the substitutions present in ECDS_2, was constructed and further studied in comparison with ECDS_2. Fermentations were carried out in a 96-well plate at 37°C for 72 hrs bacterial strains that produce 2,3;22,23-dioxidosqualene and express ECDS_2 or ECDS_3. Data

were quantified by GC-FID for 24,25-epoxy-cucurbitadienol production. As shown in **FIG. 4C**, the bacterial strains expressing ECDS_3 produced over 2-fold increased levels of 24,25-epoxy-cucurbitadienol compared to the bacterial strains expressing ECDS_2.

E. coli strains that produce 2,3;22,23-dioxidosqualene and express mutant derivatives of ECDS_3 as well as a downstream EPH were screened. Briefly, fermentations were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified using ultra-performance liquid chromatography method with diode array detection (UPLC-DAD) for cucurbitadienol, 24,25-epoxy-cucurbitadienol, and 24,25-dihydroxy cucurbitadienol production. Effects of the amino acid substitutions were evaluated by comparing the titers of cucurbitadienol, 24,25-epoxy-cucurbitadienol, and 24,25-dihydroxy cucurbitadienol produced in fermentations by bacterial strain expressing a given ECDS_3 derivative with bacterial strain expressing ECDS_3. Table 8 shows sixteen amino acid substitutions that showed beneficial changes in the titers of 24,25-epoxy cucurbitadienol and 24,25-dihydroxy-cucurbitadienol compared to ECDS_3:

Table 8: Round 4 of ECDS Engineering

Mutations	Cucurbitadienol titer	24,25-epoxy-cucurbitadienol titer	24,25-dihydroxy-cucurbitadienol titer
H475T	1.30	1.81	1.34
W470Y	1.48	1.59	1.06
V469E	1.26	1.46	1.32
N468E	1.28	1.43	1.60
V469K	1.18	1.37	1.24
L245I	0.90	1.31	1.39
V469A	1.31	1.29	1.29
F471H	0.95	1.28	1.07
F258V	1.25	1.28	1.26
T224P	1.02	1.23	1.23
S556A	1.06	1.22	0.87
G261S	1.03	1.20	1.16
T483S	0.88	1.18	0.68
C373N	1.28	1.14	1.45

F258L	1.14	1.13	1.15
H473Q	0.95	1.11	0.83

Based on these results, ECDS_4 (SEQ ID NO: 11), which harbors L245I substitution in addition to the substitutions present in ECDS_3, was constructed and further studied in comparison with ECDS_3. Fermentations using bacterial strains that produce 2,3;22,23-dioxidosqualene and express ECDS_3 or ECDS_4 (in addition to EPH) were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by UPLC-DAD for 24,25-dihydroxy-cucurbitadienol production. As shown in **FIG. 4D**, the bacterial strains expressing ECDS_4 produced about 1.4-fold increased levels of 24,25-dihydroxy-cucurbitadienol compared to the bacterial strains expressing ECDS_3.

E. coli strains producing 2,3;22,23-dioxidosqualene and expressing mutant derivatives of ECDS_4 (as well as EPH) were screened. Some of these mutant derivatives included double mutants (e.g., double mutants as shown in several rows of Table 9). Briefly, fermentations were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by UPLC-DAD for cucurbitadienol, 24,25-epoxy-cucurbitadienol, and 24,25-dihydroxy cucurbitadienol production. Effects of the amino acid substitutions were evaluated by comparing the titers of cucurbitadienol, 24,25-epoxy-cucurbitadienol, and 24,25-dihydroxy cucurbitadienol produced in fermentations by bacterial strain expressing a given ECDS_4 mutant with bacterial strain expressing ECDS_4. Table 9 shows twenty-three mutants that showed beneficial changes in the titers of 24,25-dihydroxy-cucurbitadienol compared to ECDS_4:

Table 9: Round 5 of ECDS Engineering

Mutations	Cucurbitadienol titer	24,25-epoxy-cucurbitadienol titer	24,25-dihydroxy-cucurbitadienol titer
E729N	1.27	1.18	1.30
S670A	0.82	0.71	1.29
A745N	1.81	1.58	1.23
W331S	0.85	1.22	1.22
E729Q	1.01	1.29	1.22

A745T	1.89	1.67	1.21
S670G	1.73	1.50	1.21
M553T	1.30	1.23	1.20
E729K	2.00	1.48	1.20
G733I	3.10	1.31	1.19
F258T	1.01	1.25	1.17
K506Q/L585I	1.16	1.36	1.17
S670T	1.75	1.58	1.17
V469A	0.54	0.44	1.15
T742D	2.98	1.28	1.14
C522F	0.42	0.99	1.14
A574S	0.53	0.86	1.13
V645E/K649R	1.81	1.48	1.13
A745L	1.56	1.55	1.13
I564T	0.71	1.42	1.12
E50Q	1.23	1.00	1.12
Y254E	1.07	1.19	1.12
F249Y	0.81	1.12	1.12

Based on these results, ECDS_5 (SEQ ID NO: 12), which harbors W331S substitution in addition to the substitutions present in ECDS_4, was further studied in comparison with ECDS_4. Fermentations using bacterial strains expressing ECDS_4 or ECDS_5 (in addition to EPH) were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by UPLC-DAD for 24,25-dihydroxy-cucurbitadienol production. As shown in **FIG. 4E**, the bacterial strains expressing ECDS_5 produced about 1.25-fold increased levels of 24,25-dihydroxy-cucurbitadienol compared to the bacterial strains expressing ECDS_4.

A molecular model of *Citrullus colocynthis* cucurbitadienol synthase 2 was developed. Without wishing to be bound by theory, the enzyme may comprise an access channel at the bottom of the cucurbitadienol synthase structure near its surface that interacts with the cell membrane (on the cytosolic side), where squalene substrate accesses the active site. Interestingly, the mutations I553M, L245I, and W331S disclosed above, which were identified in rounds 2, 4 and 5 of ECDS engineering, respectively, alter the topology of the access channel. To directly compare the effect of those mutations on substrate specificity (and thus, product specificity), the production of 24,25-dihydroxy-cucurbitadienol, and 24,25-epoxy-cucurbitadienol, both of which are derived through cyclization of 2,3;22,23-dioxidosqualene by ECDS, was compared to production of cucurbitadienol, which is derived from cyclization of 2,3-oxidosqualene by ECDS. ECDS variants were compared over different generations of 24, 25 epoxy cucurbitadienol synthases (ECDS) disclosed herein, compared to ECDS_0. Briefly, fermentations were carried out in a 96-well plate at 37°C for 72 hrs using bacterial strains that produce 2,3;22,23-dioxidosqualene and express ECDS_0 or its derivatives having I553M, L245I, and W331S substitutions. Data were quantified using UPLC-DAD for cucurbitadienol, 24,25-epoxy-cucurbitadienol, and 24,25-dihydroxy cucurbitadienol production. Effects of the amino acid substitutions were evaluated by comparing the titers of cucurbitadienol or a total of 24,25-epoxy-cucurbitadienol + 24,25-dihydroxy cucurbitadienol produced in fermentations by bacterial strain expressing a given ECDS derivative with bacterial strain expressing ECDS_0. Table 10 shows the results:

Table 10: Effect of mutations in substrate access channel of cucurbitadienol synthase.

Mutation	Cucurbitadienol titer	24,25-epoxy-cucurbitadienol + 24,25-dihydroxy cucurbitadienol titer
I553M	0.97	2.08
L245I	0.81	1.28
W331S	0.90	1.30

As shown above, each of the mutations show an increase in the production of 24,25-dihydroxy-cucurbitadienol + 24,25-epoxy-cucurbitadienol, both of which are derived from 2,3;22,23-dioxidosqualene, and/or a decrease in the production of cucurbitadienol, which is derived from 2,3-oxidosqualene. These results illustrate, *inter alia*, that mutations of a squalene-

access channel in cucurbitadienol synthase (CDS) provide improved substrate specificity that increase the production of 24,25-dihydroxy-cucurbitadienol and 24,25-epoxy-cucurbitadienol (which are intermediates of mogrol production) and decrease the production of undesirable off-pathway side product cucurbitadienol.

5 **FIG. 5** shows an alignment CcCDS (SEQ ID NO: 7) with homologues SgCDS (SEQ ID NO: 68), CpCAS (SEQ ID NO: 69), PsCAS (SEQ ID NO: 70), and AsCAS (SEQ ID NO: 71). Positions corresponding to D50, N121, L245, W331, Q401, I490, I553, and A556 of SEQ ID NO: 7 are highlighted, which are useful positions for mutagenesis as impacting the putative substrate access channel (e.g., due to proximity to the membrane), impacting the active site, or other
10 positions shown to impact activity or substrate specificity.

Table 11: ECDS, positions for mutagenesis to impact productivity or substrate specificity.

	Mutation close to membrane				Active site	Non-active site		
ECDS Key Mutations	L245I	I553M	A556S	W331S	I490V	D50E	N121H	Q401A
SgCDS	L240	I548	A551	W326	I485	H48	N119	Q396
CpCAS	L245	M553	A556	W331	I490	Q51	H122	P401
PsCAS	M236	I544	A547	W322	I481	E42	H113	P392
AaCAS	L236	I544	A547	W322	I481	E42	H113	P392

Table 12: Percent Identity of ECDS enzymes with each other.

Identity(%)	ECDS(CcCDS)	SgCDS	PsCAS_mut	CpCAS_mut	AaCAS_mut
ECDS(CcCDS)	100	88	70.5	71.2	69
SgCDS	44.3	100	38.6	38	39.4
PsCAS_mut	71.3	77.5	100	77.2	81.3
CpCAS_mut	71.1	75.4	76.1	100	76.5
AaCAS_mut	69.7	79	81.1	77.4	100

Example 5: Cooperation of Engineered Squalene Epoxidase (SQE) and 24,25 Epoxy Cucurbitadienol Synthase (ECDS)

Two *E. coli* strains expressing different engineered versions of squalene epoxidase (SQE) and 24,25 epoxy cucurbitadienol synthase (ECDS) were constructed. These two strains expressed

5 (1) *Artemisia annua* squalene synthase (AaSQS; SEQ ID NO: 13), SQE_3, ECDS_4 and EPH_1; and (2) AaSQS, SQE_2, ECDS_3 and EPH_1. The production of cucurbitadienol and 24,25-dihydroxy-cucurbitadienol by these strains was analyzed. Briefly, fermentations were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified using UPLC-DAD for cucurbitadienol and 24,25-dihydroxy cucurbitadienol production. The ratios of cucurbitadienol to 24,25-dihydroxy

10 cucurbitadienol produced by these strains were determined and plotted. As shown in **FIG. 6**, the strain expressing AaSQS, SQE_2, ECDS_3 and EPH_1 produced less than half as much 24,25-dihydroxy-cucurbitadienol as cucurbitadienol. On the other hand the strain expressing AaSQS, SQE_3, ECDS_4 and EPH_1 produced more than two times as much 24,25-dihydroxy-cucurbitadienol as cucurbitadienol (**FIG. 6**). Thus, the latter strain was capable of producing over

15 4-times more 24,25-dihydroxy-cucurbitadienol with respect to cucurbitadienol, compared to the former strain (**FIG. 6**). Given that a strain expressing SQE_3 produces about 2-fold more 2,3;22,23-dioxidosqualene compared to a strain expressing SQE_2 (See **FIG. 2B**); and a strain expressing ECDS_4 alone produces about 1.4-fold more 24,25-dihydroxy-cucurbitadienol compared to a strain expressing ECDS_3 (See **FIG. 4D**), these results show a synergistic effect of

20 the engineering of SQE and ECDS.

Example 6: Engineering of Cytochrome P450 Enzyme for C11-Specificity

A *Cucurbita pepo* subsp. *pepo* cytochrome P450 enzyme CYP87A3 (CppCYP87A3, SEQ ID NO: 14) was engineered. As a first step, the native transmembrane domain of CppCYP87A3 was replaced with the transmembrane domain from *E. coli* sohB to generate sohB_CppCYP87A3

25 (SEQ ID NO: 15). See U.S. Patent 10,774,314, which is hereby incorporated by reference in its entirety. Effect of these cytochrome P450 enzymes on the production of 11 hydroxy-cucurbitadienol was compared in *E. coli* strains producing cucurbitadienol. Briefly, fermentations were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified UPLC-DAD for 11 hydroxy-cucurbitadienol production. As shown in **FIG. 7A**, the strain expressing

sohB_CppCYP87A3 produced about 1.75-fold more 11 hydroxy-cucurbitadienol compared to the strain expressing CppCYP87A3.

E. coli strains expressing mutant derivatives of sohB_CppCYP87A3 were screened in a bacterial strain producing 24,25-dihydroxycucurbitadienol. Based on these results, 5
sohB_C11CYP_1 (SEQ ID NO: 16), which harbors a S193T substitution, was further studied in comparison with sohB_CppCYP87A3. Briefly, fermentations using bacterial strains producing 24,25-dihydroxy-cucurbitadienol and expressing sohB_CppCYP87A3 or sohB_C11CYP_1 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by GC-FID for mogrol production. As shown in **FIG. 7B**, the bacterial strains expressing sohB_C11CYP_1 produced 10
about 5-fold increased levels of mogrol compared to the bacterial strains expressing sohB_CppCYP87A3. These results demonstrate, *inter alia*, that sohB_C11CYP_1 has improved specificity or activity towards the hydroxylation of C-11 position.

E. coli strains expressing mutant derivatives of SohB_C11CYP_1 were screened. Briefly, fermentations using bacterial strains producing 24,25-dihydroxy-cucurbitadienol and expressing 15
sohB_C11CYP_1 derivatives were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by GC-FID for mogrol production. Effects of the amino acid substitutions in SohB_C11CYP_1 were evaluated by comparing the titers of mogrol produced by bacterial strain expressing a given sohB_C11CYP_1 derivative with bacterial strain expressing sohB_C11CYP_1. Table 13 shows sixty-four amino acid substitutions that showed beneficial changes in the titers of 20
mogrol compared to sohB_C11CYP_1:

Table 13: Round 2 of C11CYP Engineering

Mutation	Mogrol titer
W130N	4.16
I141L	3.02
W130G	2.87
A468S	2.74
A468T	2.54

K132S	2.42
W130S	2.37
N234H	2.27
A269V	2.23
I470G	2.06
L120F	1.99
S184A	1.97
G219Q	1.94
A468N	1.92
L120A	1.90
K246M	1.89
V228I	1.82
L131P	1.80
Y309F	1.78
L106Q	1.76
A330N	1.74
A466V	1.69
G253E	1.67
L255R	1.66
L206S	1.63
I291L	1.63
T207S	1.63
H469S	1.63

I243A	1.62
F452V	1.61
H433F	1.58
F123I	1.57
L120M	1.57
S280T	1.56
F158L	1.54
T371L	1.49
R155K	1.43
S393A	1.43
A174S	1.38
L460I	1.37
A151P	1.34
K35R	1.34
T145V	1.34
N329K	1.30
A111L	1.29
Q421S	1.28
I418N	1.25
L131I	1.23
Q277T	1.22
A332E	1.22
R394V	1.21

G209N	1.20
R467L	1.19
R191D	1.19
L374I	1.18
H469P	1.17
W130L	1.17
D36N	1.16
I279L	1.15
Q286D	1.15
I450V	1.14
G216A	1.14
L117M	1.12
S292L	1.11

Based on these results, SohB_C11CYP_2 (SEQ ID NO: 17), which harbors W130N, N234H, and A468T substitutions in addition to the substitution present in SohB_C11CYP_1, was further studied in comparison with sohB_CppCYP87A3. Fermentations using bacterial strains producing 24,25-dihydroxy-cucurbitadienol and expressing sohB_CppCYP87A3 or
5 SohB_C11CYP_2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by GC-FID for mogrol production. As shown in **FIG. 7C**, the bacterial strains expressing SohB_C11CYP_2 produced about 2-fold increased levels of mogrol compared to the bacterial strains expressing sohB_CppCYP87A3.

E. coli strains expressing mutant derivatives of SohB_C11CYP_2 were screened. Briefly,
10 fermentations using bacterial strains producing 24,25-dihydroxy-cucurbitadienol and expressing sohB_C11CYP_2 or sohB_C11CYP_2 derivatives were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by GC-FID for mogrol production. Effects of the amino acid substitutions in SohB_C11CYP_2 were evaluated by comparing the titers of mogrol produced in

fermentations by bacterial strain expressing a given sohB_C11CYP_2 derivative with bacterial strain expressing sohB_C11CYP_2. Table 14 shows fifty-three amino acid substitutions that showed beneficial changes in the titers of mogrol compared to sohB_C11CYP_2:

Table 14: Round 3 of C11CYP Engineering

Mutation	Mogrol titer
H433F	2.20
T128G	2.15
L106Q	2.03
K132N	1.79
N329K	1.78
L131T	1.73
I291L	1.70
L310V	1.69
F452V	1.68
T468N	1.67
F214Y	1.62
A151P	1.56
G219Q	1.54
L477Y	1.54
D312N	1.53
V228I	1.50
V465M	1.49
A311S	1.48
E325M	1.47

E474G	1.43
A466I	1.36
L477F	1.36
G125A	1.34
A330N	1.34
A269V	1.33
T215D	1.33
L206S	1.32
Q58E	1.31
L50F	1.28
D282E	1.28
T302V	1.28
F283M	1.28
T302A	1.27
V42K	1.27
G253E	1.25
S184A	1.25
A98P	1.24
G216N	1.24
I279M	1.23
G219R	1.22
L255R	1.22
L52F	1.22

A311T	1.21
T223A	1.21
T304S	1.19
K38R	1.19
H469S	1.19
K33R	1.18
T119A	1.17
N130G	1.15
G476P	1.14
T301S	1.13
K132T	1.13

Based on these results, SohB_C11CYP_3 (SEQ ID NO: 18), which harbors H433F substitution in addition to the substitution present in SohB_C11CYP_2, was further studied in comparison with SohB_C11CYP_2. Fermentations using bacterial strains producing 24,25-dihydroxy-cucurbitadienol and expressing SohB_C11CYP_3 or SohB_C11CYP_2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by GC-FID for mogrol production. As shown in **FIG. 7D**, the bacterial strains expressing SohB_C11CYP_3 produced over about 2-fold increased levels of mogrol compared to the bacterial strains expressing SohB_C11CYP_2.

E. coli strains expressing mutant derivatives of SohB_C11CYP_3 were screened. Some of the mutant derivatives included mutants having clusters of mutations (see, *e.g.*, clusters of mutations as shown in some of the rows of Table 15). Briefly, fermentations using bacterial strains producing 24,25-dihydroxy-cucurbitadienol and expressing SohB_C11CYP_3 or SohB_C11CYP_3 derivatives were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by GC-FID for mogrol production. Effects of the amino acid substitutions in SohB_C11CYP_3 were evaluated by comparing the titers of mogrol produced by bacterial strain expressing a given SohB_C11CYP_3 derivative with bacterial strain expressing

SohB_C11CYP_3. Table 15 shows thirty-two mutants that showed beneficial changes in the titers of mogrol compared to SohB_C11CYP_3:

Table 15: Round 4 of C11CYP Engineering

Mutation	Mogrol titer
T304S	1.76
G125A_T128G	1.61
F452V_T468N	1.53
T128S	1.51
L106Q_K132N_F452V	1.49
V228F	1.48
F452I	1.46
A311T	1.42
A311S	1.40
L310V	1.39
G125A_K132T	1.36
K132S	1.33
T128P	1.33
V228A	1.31
A466S	1.30
A330N	1.29
F473I	1.24
K246M	1.23
T128G_V228I	1.22
T128A	1.20

V228M	1.18
L134V	1.17
L477V	1.17
D475N	1.16
G253E_L255R	1.14
F452V	1.14
F452L	1.12
F452V_T468N_L477Y	1.11
K464N	1.11
E325M	1.11
A133S	1.11
T128V	1.10

Based on these results, SohB_C11CYP_4 (SEQ ID NO: 19), which harbors G125A and T128G substitutions in addition to the substitutions present in SohB_C11CYP_3, was further studied in comparison with SohB_C11CYP_3. Fermentations using bacterial strains producing 24,25-dihydroxy-cucurbitadienol and expressing SohB_C11CYP_4 or SohB_C11CYP_3 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by GC-FID for mogrol production. As shown in **FIG. 7E**, the bacterial strains expressing SohB_C11CYP_4 produced over about 1.6-fold increased levels of mogrol compared to the bacterial strains expressing SohB_C11CYP_3.

E. coli strains expressing mutant derivatives, recoded versions, or derivatives having alternative N-terminal anchors of SohB_C11CYP_4 were screened. Some of the mutant derivatives included mutants having clusters of mutations (see, *e.g.*, clusters of mutations as shown in some of the rows of Table 16). Briefly, fermentations using bacterial strains producing 24,25-dihydroxy-cucurbitadienol and expressing SohB_C11CYP_4 or SohB_C11CYP_4 derivatives were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by GC-FID for mogrol

production. Effects of the mutants of SohB_C11CYP_4 were evaluated by comparing the titers of mogrol produced in fermentations by bacterial strain expressing a given SohB_C11CYP_4 derivative with bacterial strain expressing SohB_C11CYP_4. Table 16 shows fourteen derivatives that showed beneficial changes in the titers of mogrol compared to SohB_C11CYP_4:

5

Table 16: Round 5 of C11CYP Engineering

Mutation	Mogrol titer
L131V	1.44
I299T	1.34
V465I	1.24
K132N	1.20
K464R	1.20
L134T	1.17
T304S_A330N	1.15
L218I	1.14
D127S	1.13
E325M_A330N	1.13
I299V	1.11
G253E	1.11
D127Q	1.11
anchor_swap_n39_fliO	1.10

Based on these results, SohB_C11CYP_5 (SEQ ID NO: 20), which harbors an L131V substitution, in addition to the substitutions present in SohB_C11CYP_4, was further studied in comparison with SohB_C11CYP_4. Fermentations using bacterial strains producing 24,25-dihydroxy-cucurbitadienol and expressing SohB_C11CYP_5 or SohB_C11CYP_4 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by GC-FID for mogrol production.

10

As shown in **FIG. 7F**, the bacterial strains expressing SohB_C11CYP_5 produced over about 1.3-fold increased levels of mogrol compared to the bacterial strains expressing SohB_C11CYP_4.

E. coli strains expressing mutant derivatives of SohB_C11CYP_5 were screened for improved substrate specificity. Based on these data, SohB_C11CYP_6 (SEQ ID NO: 21), which harbors V131L, K132N, and V465I substitutions, in addition to the substitution present in SohB_C11CYP_5, was further studied in comparison with SohB_C11CYP_5. Fermentations using bacterial strains producing 24,25-dihydroxy-cucurbitadienol and expressing SohB_C11CYP_6 or SohB_C11CYP_5 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by GC-FID for mogrol, 11-hydroxyl-cucurbitadienol, and oxo-mogrol production. As shown in **FIG. 7G**, the bacterial strains expressing SohB_C11CYP_6 produced decreased levels of 11-hydroxyl-cucurbitadienol and oxo-mogrol, without affecting the levels of mogrol, compared to the bacterial strains expressing SohB_C11CYP_5. These results indicated, *inter alia*, that SohB_C11CYP_6 had an improved substrate specificity (*e.g.*, discrimination of cucurbitadienol over 24,25-dihydroxy-cucurbitadienol) compared to SohB_C11CYP_5.

E. coli strains expressing mutant derivatives of SohB_C11CYP_6 were screened. Briefly, fermentations using bacterial strains producing mogrol through expression of SohB_C11CYP_6 derivatives and expressing MogCPR2_0 or MogCPR2_2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by GC-FID for mogrol production. Effects of the amino acid substitutions in SohB_C11CYP_6 were evaluated by comparing the titers of mogrol produced. Table 17 shows nine sets of amino acid substitutions that showed beneficial changes in the titers of mogrol compared to SohB_C11CYP_6:

Table 17: Round 6 of C11CYP Engineering

Mutation	Mogrol titer
A25E_A26R	1.123
S63F	1.122
N102H	1.136
D183E	1.462
L218I_G219Q	1.733

K238Q	1.548
G253E	1.778
S472Y	1.352
P483E	1.358

E. coli strains expressing additional mutant derivatives of SohB_C11CYP_6 were screened. Briefly, fermentations using bacterial strains producing mogrol through expression of SohB_C11CYP_6 derivatives and expressing MogCPR2_0 or MogCPR2_2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by GC-FID for mogrol and 11-oxo mogrol production. Effects of the amino acid substitutions in SohB_C11CYP_6 were evaluated by comparing the titers of mogrol and 11-oxo mogrol produced. Table 18 shows eight sets of amino acid substitutions that showed beneficial changes in the titers of mogrol compared to SohB_C11CYP_6:

Table 18: Round 7 of C11CYP Engineering

Mutation	Mogrol titer	11-oxo mogrol titer
R61I	1.319	0.000
A88V	1.095	0.000
A125D	1.053	0.000
C236A	1.233	0.000
S290A	1.294	0.000
A125N	1.043	0.370
L288I_S292L	1.379	0.471
K464E	1.108	0.661

Based on these and other results results, SohB_C11CYP_8 (SEQ ID NO: 45), which has L218I and G219A substitutions, was further studied in comparison with SohB_C11CYP_6. Fermentations using bacterial strains producing mogrol through expression of SohB_C11CYP_6 or SohB_C11CYP_8 and expressing MogCPR2_0 or MogCPR2_2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by GC-FID for mogrol and 11-oxo mogrol production. As shown in **FIG. 7H**, the bacterial strains expressing SohB_C11CYP_8 produced

about 1.2-fold increased levels of mogrol and about 2.3-fold increased levels of 11-oxo mogrol compared to the bacterial strains expressing SohB_C11CYP_6.

E. coli strains expressing mutant derivatives of SohB_C11CYP_8 were screened. Briefly, fermentations using bacterial strains producing mogrol through expression of SohB_C11CYP_8 derivatives and expressing MogCPR2_1 or MogCPR2_1 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by GC-FID for mogrol production. Effects of the amino acid substitutions in SohB_C11CYP_8 were evaluated by comparing the titers of mogrol produced. Table 19 shows seventeen sets of amino acid substitutions that showed beneficial changes in the titers of mogrol compared to SohB_C11CYP_8:

10

Table 19: Round 8 of C11CYP Engineering

Mutation	Mogrol titer
N102H_S63F	1.265
T446V	1.251
R61I_N102H_A25E_A26R	1.222
K464D	1.202
A25E_A26R_S63F	1.169
K464R	1.169
N102H_A25E_A26R	1.161
K238Q_D183E	1.155
F354Y	1.133
F433L	1.130
R61I_N102H	1.127
E438D	1.125
K273M	1.109
S292L	1.106
S63Y	1.103
K238Q	1.102
G209N	1.102

Based on these results, SohB_C11CYP_9 (SEQ ID NO: 46), which has N102H and S63F substitutions, was further studied in comparison with SohB_C11CYP_8. Fermentations using bacterial strains producing mogrol through expression of SohB_C11CYP_8 or SohB_C11CYP_9 and expressing MogCPR2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by UPLC-DAD for mogrol and 11-oxo mogrol production. As shown in **FIG. 7I**, the bacterial strains expressing SohB_C11CYP_9 produced about 1.25-fold increased levels of mogrol and a 0.25-fold decrease in 11-oxo mogrol compared to the bacterial strains expressing SohB_C11CYP_8.

E. coli strains expressing mutant derivatives of SohB_C11CYP_9 were screened. Briefly, fermentations using bacterial strains producing mogrol through expression of SohB_C11CYP_9 derivatives and expressing MogCPR2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by UPLC-DAD for mogrol production. Effects of the amino acid substitutions in SohB_C11CYP_9 were evaluated by comparing the titers of mogrol produced. Table 20 shows twenty-three sets of amino acid substitutions that showed beneficial changes in the titers of mogrol compared to SohB_C11CYP_9:

Table 20: Round 9 of C11CYP Engineering

Mutation	Mogrol titer
F63N_K75R_F354Y	1.41
K35R_F63N_K75R_F158M_G208A_F354Y	1.37
F63N_K75R_K238Q_G253E	1.34
K35R_F63N_K75R_K238Q_G253E	1.34
K75R_K238Q_G253E_Q277K_F354Y	1.32
F63N_K238Q_G253E	1.32
F63N_K75R_F158M_Q277K	1.29
F63N_K75R	1.28
K75R_F158M_F354Y	1.26
K75R_F158M	1.24
K75R_F354Y	1.24
K35R_F63N_K75R_F158M_K238Q_G253E_F354Y	1.23
F63N_F158M_K238Q_G253E_F354Y	1.22

K238Q_G253E_F354Y	1.22
K35R_F63N_K75R_F158M_G208A_K238Q_G253E_Q277K_F354Y	1.21
F63N_F158M	1.20
F354Y	1.18
K75R_K238Q_G253E	1.15
F158M	1.14
K75R	1.13
E485D	1.12
H102Y	1.11
V53I	1.10

Based on these results, SohB_C11CYP_10 (SEQ ID NO: 47), which has F63N, K75R, and F354Y substitutions, was further studied in comparison with SohB_C11CYP_9. Fermentations using bacterial strains producing mogrol through expression of SohB_C11CYP_9 or SohB_C11CYP_10 and expressing MogCPR2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by UPLC-DAD for mogrol production. As shown in **FIG. 7J**, the bacterial strains expressing SohB_C11CYP_10 produced about 1.4-fold increased levels of mogrol compared to the bacterial strains expressing SohB_C11CYP_9.

E. coli strains expressing mutant derivatives of SohB_C11CYP_10 were screened for the increase in mogrol and mogrol derived mogrosides relative to 24,25 dihydroxycucurbitadienol and deoxymogroside derived compounds. Briefly, fermentations using bacterial strains producing mogrol through expression of SohB_C11CYP_10 derivatives, expressing MogCPR2 and two UGTs (MogUGTc24_3 and MogUGTc3_2) were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by liquid chromatography–mass spectrometry (LC-MS) for the production of mogrol and mogrol derived mogrosides. Effects of the amino acid substitutions in SohB_C11CYP_10 were evaluated by comparing the titers of mogrol and mogrol derived mogrosides. Table 21 shows twelve sets of amino acid substitutions that showed beneficial changes in the titers of mogrol compared to SohB_C11CYP_10:

Table 21: Round 10 of C11CYP Engineering

Mutation	Mogrol titer
----------	--------------

K369R	1.55
I299V	1.38
T207S_G208A_G209N	1.34
D201Q_S202E	1.33
V187L	1.27
K35R_D36N_S37W_V42K	1.27
R191D	1.22
M105L	1.19
S165Q	1.17
K77E	1.15
K74E	1.13
E438D	1.10

Based on these results, SohB_C11CYP_11 (SEQ ID NO: 80), which has K369R substitution, in addition to the mutations in SohB_C11CYP_10, was further studied in comparison with SohB_C11CYP_10. Fermentations using bacterial strains producing mogrol through expression of SohB_C11CYP_11 or SohB_C11CYP_11 and expressing MogCPR2 and the UGTs MogUGTc24_3 and MogUGTc3_2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by LC-MS for mogroside and deoxymogroside production. As shown in **FIG. 7K**, the bacterial strains expressing SohB_C11CYP_11 produced about 1.6-fold increased levels of mogrol compared to the bacterial strains expressing SohB_C11CYP_10.

E. coli strains expressing mutant derivatives of SohB_C11CYP_11 were screened for the increase in mogrol and mogrol derived mogrosides relative to 24,25 dihydroxycucurbitadienol and deoxymogroside derived compounds. Briefly, fermentations using bacterial strains producing mogrol through expression of SohB_C11CYP_11 derivatives, expressing MogCPR2 and the UGTs MogUGTc24_3 and MogUGTc3_2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by liquid chromatography–mass spectrometry (LC-MS) for the production of mogroside and deoxymogroside. Effects of the amino acid substitutions in SohB_C11CYP_11 were evaluated by comparing the titers of mogroside and deoxymogroside. Table 22 shows eleven sets of amino acid substitutions that showed beneficial changes in the titers of mogrol compared to SohB_C11CYP_10:

Table 22: Round 11 of C11CYP Engineering

Mutation	mogroside + deoxymogroside titer
A60L	1.30
T223S_V228I	1.27
T482S	1.22
V306A	1.10
L52F	1.10
I14A_N32R_S37P_T47S_K167R	1.45
A12W	1.36
I14A_N32R_S37P_T47S	1.33
I14F_A256K	1.31
T192L	1.23
K246V_T347K	1.14

Based on these results, SohB_C11CYP_12 (SEQ ID NO: 81), which has T223S and V228I substitutions, in addition to the mutations in SohB_C11CYP_11, was further studied in comparison with SohB_C11CYP_11. Fermentations using bacterial strains producing mogrol through expression of SohB_C11CYP_11 or SohB_C11CYP_12 and expressing MogCPR2 and the UGTs MogUGTc24_3 and MogUGTc3_2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by LC-MS for mogroside and deoxymogroside production. As shown in **FIG. 7L**, the bacterial strains expressing SohB_C11CYP_12 produced about 1.3-fold increased levels of mogrol compared to the bacterial strains expressing SohB_C11CYP_11.

E. coli strains expressing mutant derivatives of SohB_C11CYP_12 were screened for the increase in mogrol and mogrol derived mogrosides relative to 24,25 dihydroxycucurbitadienol and deoxymogroside derived compounds. Briefly, fermentations using bacterial strains producing mogrol through expression of SohB_C11CYP_12 derivatives, expressing MogCPR2 and the UGTs MogUGTc24_3 and MogUGTc3_2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by liquid chromatography–mass spectrometry (LC-MS) for the production of mogroside and deoxymogroside. Effects of the amino acid substitutions in SohB_C11CYP_12 were evaluated by comparing the titers of mogrosides to deoxymogrosides. Table 23 shows six

sets of amino acid substitutions that showed beneficial changes in the titers of mogrol compared to SohB_C11CYP_10:

Table 23: Round 12 of C11CYP Engineering

Mutation	mogroside titer
R90Q	1.90
A12W	1.28
T482S	1.14
A12M	1.13
T145A	1.11
T231F_T232A_Y233F	1.03

Based on these results, SohB_C11CYP_13 (SEQ ID NO: 82), which has T231F, T232A, and Y233F substitutions, in addition to the mutations in SohB_C11CYP_12, was further studied in comparison with SohB_C11CYP_12. Fermentations using bacterial strains producing mogrol through expression of SohB_C11CYP_12 or SohB_C11CYP_13 and expressing MogCPR2 and the UGTs MogUGTc24_3 and MogUGTc3_2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by LC-MS for mogroside, deoxymogroside, and 11-oxo mogroside production. As shown in **FIG. 7M**, the bacterial strains expressing SohB_C11CYP_13 produced about 60% reduced levels of 11-oxomogrosides compared to the bacterial strains expressing SohB_C11CYP_12.

E. coli strains expressing mutant derivatives of SohB_C11CYP_13 were screened for the increase in mogrol and mogrol derived mogrosides relative to 24,25 dihydroxycucurbitadienol and deoxymogroside derived compounds. Briefly, fermentations using bacterial strains producing mogrol through expression of SohB_C11CYP_13 derivatives, expressing MogCPR2 and the UGTs MogUGTc24_3 and MogUGTc3_2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by liquid chromatography–mass spectrometry (LC-MS) for the production of mogroside and deoxymogroside. Effects of the amino acid substitutions in SohB_C11CYP_13 were evaluated by comparing the titers of mogrosides to deoxymogrosides. Table 24 shows twenty-three sets of amino acid substitutions that showed beneficial changes in the titers of mogrol compared to SohB_C11CYP_10:

Table 24: Round 13 of C11CYP Engineering

Mutation	Mogroside titer
A12W_F452I_T192L	1.64
A12W_F452V_T192L	1.64
A174S_T175S_R191E_T192L_K246V_A330N_T406A_L444M_F452V	1.53
F452I_T192L_T482S	1.45
I228F_S202E_T482F	1.41
A174S_T175S_R191E_T192L_K246V_T406A_L444M_F452V	1.37
F452V_R191D_T446V	1.36
A26R_K77E_L460I	1.34
G209N_T47S_Y455F	1.27
A26R_K77E_Q375H	1.26
F452V_K246M_T482S	1.24
K246I	1.20
I228F_K77E_S184A	1.19
V53L	1.18
A26R_L460I_Q375H	1.16
G209N_R456N_T47S	1.16
K77D_K246V	1.14
R191E	1.14
A12W_R191D_T446V	1.13
L266M_L374I_N32R	1.13
A322E_I14F_I228M	1.13
G209N_T482F_V187L	1.12
F452I_K246M_T482S	1.12

Based on these results, SohB_C11CYP_14 (SEQ ID NO: 83), which has A12W, F452I, and T192L substitutions, in addition to the mutations in SohB_C11CYP_13, was further studied in comparison with SohB_C11CYP_13. Fermentations using bacterial strains producing mogrol

through expression of SohB_C11CYP_13 or SohB_C11CYP_14 and expressing MogCPR2 and the UGTs MogUGTc24_3 and MogUGTc3_2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by LC-MS for mogrol production. As shown in **FIG. 7N**, the bacterial strains expressing SohB_C11CYP_14 produced about 1.7-fold increased levels of mogrol compared to the bacterial strains expressing SohB_C11CYP_13.

E. coli strains expressing mutant derivatives of SohB_C11CYP_14 were screened for the increase in mogrol and mogrol derived mogrosides relative to 24,25 dihydroxycucurbitadienol and deoxymogroside derived compounds. Briefly, fermentations using bacterial strains producing mogrol through expression of SohB_C11CYP_14 derivatives, expressing MogCPR2 and the UGTs MogUGTc24_3 and MogUGTc3_2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by liquid chromatography–mass spectrometry (LC-MS) for the production of mogroside and deoxymogroside. Effects of the amino acid substitutions in SohB_C11CYP_14 were evaluated by comparing the titers of mogrosides to deoxymogrosides. Table 25 shows twenty-four sets of amino acid substitutions that showed beneficial changes in the titers of mogrol compared to SohB_C11CYP_10:

Table 25: Round 14 of C11CYP Engineering

Mutation	Mogroside titer
I27G_K246V_T482S_V76M	1.37
L57F_G209N_I27G_T482S	1.32
L57F_I27G_T482S_V76M	1.29
A26G_E342D_I14F_T482F	1.29
Q375H_A26G_E342D_H29V	1.28
N32R_E383A_H29V_I14F	1.27
G209N_A26R_K246V_T482S	1.25
A22L_Q375H_A26G_E342D	1.24
I27G_K246V_T446V_S202E	1.23
A22L_Q375H_A26G_T482F	1.23
L57F_G209N_A26R_T482S	1.22
R191E_A26G_E342D_S37P	1.21

K246V_T446V_S202E_V76M	1.17
A22L_R191E_N32R_T482F	1.17
K246V_T446V_S202E_T47K	1.15
101_MB4773_P1G11_v87	1.15
I27G_T446V_T482S_V76M	1.14
E383K_W30N_T169S_T47K	1.13
L57F_G209N_A26R_T446V	1.13
T47S_R155K_A185T_I228F	1.13
I27G_T482S_S202E_V76M	1.12
I27G_T446V_S202E_V76M	1.11
E383K_W30N_T169S_K246I	1.11
T47S_L460V_R155K_I228F	1.10

Based on these results, SohB_C11CYP_15 (SEQ ID NO: 84), which has I27G, K246V, T482S, and V76M substitutions, in addition to the mutations in SohB_C11CYP_14, was further studied in comparison with SohB_C11CYP_14. Fermentations using bacterial strains producing mogrol through expression of SohB_C11CYP_14 or SohB_C11CYP_15 and expressing MogCPR2 and the UGTs MogUGTc24_3 and MogUGTc3_2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by LC-MS for mogrol production. As shown in **FIG. 7O**, the bacterial strains expressing SohB_C11CYP_15 produced about 1.4-fold increased levels of mogrol compared to the bacterial strains expressing SohB_C11CYP_14.

A version of SohB_C11CYP_15 having a deletion of amino acids L3 to H29 n20_C11CYP_15 (SEQ ID NO: 85) was constructed. This mutant beneficial for solubility compared to SohB_C11CYP_15. Bacterial strains producing SohB_C11CYP_15 mutant were further studied in comparison with SohB_C11CYP_15. Fermentations using bacterial strains producing mogrol through expression of SohB_C11CYP_15 or n20_C11CYP_15 and expressing MogCPR2 and the UGTs MogUGTc24_3 and MogUGTc3_2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by LC-MS for mogrol production. As shown in **FIG. 7P**, the bacterial strains expressing n20_C11CYP_15 produced similar levels of mogrol compared to the bacterial strains expressing SohB_C11CYP_15.

E. coli strains expressing mutant derivatives of n20_C11CYP_15 were screened for the increase in mogrol and mogrol derived mogrosides relative to 24,25 dihydroxycucurbitadienol and deoxymogroside derived compounds. Briefly, fermentations using bacterial strains producing mogrol through expression of n20_C11CYP_15 derivatives, expressing MogCPR2 and the UGTs MogUGTc24_3 and MogUGTc3_2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by liquid chromatography–mass spectrometry (LC-MS) for the production of mogroside and deoxymogroside. Effects of the amino acid substitutions in n20_C11CYP_15 were evaluated by comparing the titers of mogrosides to deoxymogrosides. Table 26 shows four amino acid substitutions with respect to SEQ ID NO: 85 that showed beneficial changes in the titers of mogrol compared to n20_C11CYP_15:

Table 26: Round 15 of C11CYP Engineering

Mutation	Mogroside titer
K8R/D9N/S10R/N13K/V15K	2.26
Q31E/A33L/R34K	2.23
S390D	1.20
D235G/Q241T/A242L/G247E/S248K/Q249E/Q250E/S253T	1.10

Based on these results, n20_C11CYP_16 (SEQ ID NO: 86), which has K8R, D9N, S10R, N13K, and V15K substitutions with respect to SEQ ID NO: 85, in addition to the mutations in n20_C11CYP_15, was further studied in comparison with n20_C11CYP_15. Fermentations using bacterial strains producing mogrol through expression of n20_C11CYP_15 or n20_C11CYP_16 and expressing MogCPR2 and the UGTs MogUGTc24_3 and MogUGTc3_2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by LC-MS for mogrol production. As shown in **FIG. 7Q**, the bacterial strains expressing n20_C11CYP_16 produced about 2.25-fold increased levels of mogrol compared to the bacterial strains expressing n20_C11CYP_15.

E. coli strains expressing mutant derivatives of n20_C11CYP_16 were screened for the increase in mogrol and mogrol derived mogrosides relative to 24,25 dihydroxycucurbitadienol and deoxymogroside derived compounds. Briefly, fermentations using bacterial strains producing mogrol through expression of n20_C11CYP_16 derivatives, expressing MogCPR2 and the UGTs MogUGTc24_3 and MogUGTc3_2 were carried out in a 96-well plate at 37°C for 72 hrs. Data

were quantified by liquid chromatography–mass spectrometry (LC-MS) for the production of mogroside and deoxymogroside. Effects of the amino acid substitutions in n20_C11CYP_16 were evaluated by comparing the titers of mogrosides to deoxymogrosides. Table 27 shows five sets of amino acid substitutions that showed beneficial changes in the titers of mogrol compared to n20_C11CYP_16.

Table 27: Round 16 of C11CYP Engineering

Mutation	Mogroside titer
I45V_K46R_K47E_M49V_K50E_R51K	1.37
K50E_R51K_I55V_C59K	1.33
A33L_R34K	1.22
R8K_R10K_R34K_R51K	1.21
L79Q	1.16

Based on these results, n20_C11CYP_17 (SEQ ID NO: 87), which has I45V, K46R, K47E, M49V, K50E, and R51K substitutions with respect to SEQ ID NO: 85, in addition to the mutations in n20_C11CYP_16, was further studied in comparison with n20_C11CYP_16. Fermentations using bacterial strains producing mogrol through expression of n20_C11CYP_16 or n20_C11CYP_17 and expressing MogCPR2 and the UGTs MogUGTc24_3 and MogUGTc3_2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by LC-MS for mogrol production. As shown in **FIG. 7R**, the bacterial strains expressing n20_C11CYP_17 produced about 1.4-fold increased levels of mogrol compared to the bacterial strains expressing n20_C11CYP_16.

E. coli strains expressing mutant derivatives of n20_C11CYP_17 were screened for the increase in mogrol and mogrol derived mogrosides relative to 24,25 dihydroxycucurbitadienol and deoxymogroside derived compounds. Briefly, fermentations using bacterial strains producing mogrol through expression of n20_C11CYP_17 derivatives, expressing MogCPR2 and the UGTs MogUGTc24_3 and MogUGTc3_2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by liquid chromatography–mass spectrometry (LC-MS) for the production of mogroside and deoxymogroside. Effects of the amino acid substitutions in n20_C11CYP_17 were evaluated by comparing the titers of mogrosides to deoxymogrosides. Table 28 shows five sets of

amino acid substitutions that showed beneficial changes in the titers of mogrol compared to n20_C11CYP_16.

Table 28: Round 17 of C11CYP Engineering

Mutation	Mogroside titer
R164D_V168R_K169A	1.37
I191L_A192K	1.32
F163L_L165T_V168E_K169Q_E173R_D174E_ S175N_S176A_L179I_G181P_G182E	1.19
A33L_R34K	1.17
A33L	1.17

Based on these results, n20_C11CYP_18 (SEQ ID NO: 88), which has I191L and A192K substitutions with respect to SEQ ID NO: 85, in addition to the mutations in n20_C11CYP_17, was further studied in comparison with n20_C11CYP_17. Fermentations using bacterial strains producing mogrol through expression of n20_C11CYP_17 or n20_C11CYP_18 and expressing MogCPR2 and the UGTs MogUGTc24_3 and MogUGTc3_2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by LC-MS for mogrol production. As shown in **FIG. 7S**, the bacterial strains expressing n20_C11CYP_18 produced about 1.35-fold increased levels of mogrol compared to the bacterial strains expressing n20_C11CYP_17.

E. coli strains expressing mutant derivatives of n20_C11CYP_18 were screened for the increase in mogrol and mogrol derived mogrosides relative to 24,25 dihydroxycucurbitadienol and deoxymogroside derived compounds. Briefly, fermentations using bacterial strains producing mogrol through expression of n20_C11CYP_18 derivatives, expressing MogCPR2 and the UGTs MogUGTc24_3 and MogUGTc3_2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by liquid chromatography–mass spectrometry (LC-MS) for the production of mogroside and deoxymogroside. Effects of the amino acid substitutions in n20_C11CYP_18 were evaluated by comparing the titers of mogrosides to deoxymogrosides. Table 29 shows twenty-three sets of amino acid substitutions that showed beneficial changes in the titers of mogrol compared to n20_C11CYP_16.

Table 29: Round 18 of C11CYP Engineering

Mutation	Mogroside titer
F406Y_E411D_Y412F_S413A	1.47
V26I_Q31E_A33L_R34K	1.37
G247E_S248D_Q249N_Q250D	1.32
G247E_S248E_Q250D	1.31
L179M_G181A_G182E	1.30
A33R_R34K	1.26
E411D_Y412L_S413A	1.23
F109H	1.19
R144D_Y145S_A147S	1.19
E225D_G226E_L228R_S230N	1.17
A303S_A305K_P310A_I311L	1.17
S175E	1.16
L79A	1.14
F238M_A242M	1.13
L25F_V26I_A33L_R34K	1.13
F109N	1.13
S157A_A158S_V160H	1.13
K218Q_G226A_L228R	1.13
E236D_A242M_K244E	1.12
G247D	1.12
R164D_V160L	1.12
R144N_Y145H_A147S_P150D	1.11

Based on these results, n20_C11CYP_19 (SEQ ID NO: 89), which has F406Y, E411D, Y412F, and S413A substitutions with respect to SEQ ID NO: 85, in addition to the mutations in n20_C11CYP_17, was further studied in comparison with n20_C11CYP_17. Fermentations using bacterial strains producing mogrol through expression of n20_C11CYP_17 or n20_C11CYP_18 and expressing MogCPR2 and the UGTs MogUGTc24_3 and MogUGTc3_2 were carried out in

a 96-well plate at 37°C for 72 hrs. Data were quantified by LC-MS for mogrol production. As shown in **FIG. 7T**, the bacterial strains expressing n20_C11CYP_18 produced about 1.5-fold

Example 7: Engineering of Cytochrome P450 Reductase Enzyme

A *Cucurbita pepo* subsp. *pepo* cytochrome P450 reductase enzyme (CppCPR4, also referred to herein as MogCPR1_0, SEQ ID NO: 22) was engineered. As a first step, the native transmembrane domain of CppCPR4 was replaced with the transmembrane domains from various *E. coli* membrane proteins. In the process, CppCPR4 was truncated removing the N-terminal 67 amino acids. Effect of these transmembrane domains on the production of mogrol was compared in *E. coli* strains producing mogrol through expression of SohB_C11CYP_4. Briefly, fermentations were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified UPLC-DAD for mogrol production. As shown in Table 30, the strains expressing CppCPR4 having the transmembrane domains from various *E. coli* membrane proteins produced more mogrol compared to the strain expressing CppCPR4. *E. coli* membrane proteins and the use of their N-terminal transmembrane domains or portions thereof as membrane anchors for cytochrome P450 enzymes and reductases is described in U.S. Patent No. 10,774,314, which is hereby incorporated by reference in its entirety.

Table 30: Round 1 of C11CPR Engineering

Transmembrane domain	Mogrol titer
ycgG	6.87
yhcB	4.34
zipA	3.51
waaA	3.11
sohB	3.03
djlA	2.35
lpxK	2.35
ypfN	1.93

yhhM	1.33
------	------

E. coli strains expressing mutant derivatives of CppCPR4 were screened. Briefly, fermentations using bacterial strains producing mogrol expressing CppCPR4 or CppCPR4 derivatives were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by UPLC-DAD for mogrol production. Effects of the amino acid substitutions in CppCPR4 were evaluated by comparing the titers of mogrol produced in fermentations by bacterial strain expressing a given CppCPR4 derivative with bacterial strain expressing CppCPR4. Table 31 shows eighteen mutants that showed beneficial changes in the titers of mogrol compared to CppCPR4:

Table 31: Round 2 of C11CPR Engineering

Mutation	Mogrol titer
N601T	3.66
Q218I	2.02
F48L	1.23
D171P	1.23
Q74R	1.20
L66F	1.17
S634T	1.16
T701D	1.15
I86V	1.14
L441S	1.14
R706K	1.12
S370P	1.12
L653I_A657G	1.16
A397P	1.11

V504I	1.11
S440A	1.10
K69R	1.10

Based these results from Tables 30 and 31, MogCPR1_1 (SEQ ID NO: 23), which has a N-terminal deletion of 67 amino acids and a transmembrane domain based on the 25 N-terminal amino acids of *E. coli ycgG* (i.e., n25_ycgG), was further studied in comparison with MogCPR1_0 (CpCPR4). Fermentations using bacterial strains producing mogrol and expressing MogCPR1_0 or MogCPR1_1 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by UPLC-DAD for mogrol production. As shown in **FIG. 8A**, the bacterial strains expressing MogCPR1_1 produced over about 1.25-fold increased levels of mogrol compared to the bacterial strains expressing MogCPR1_0.

A *Phaseolus vulgaris* cytochrome P450 reductase enzyme (PvCPR, also referred to herein as MogCPR2_0, SEQ ID NO: 24) was compared with MogCPR1_0. Fermentations using bacterial strains producing mogrol and expressing MogCPR1_0 or MogCPR2_0 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by UPLC-DAD for mogrol production. As shown in **FIG. 8B**, the bacterial strains expressing MogCPR2_0 produced over about 1.4-fold increased levels of mogrol compared to the bacterial strains expressing MogCPR1_0.

The native transmembrane domain of MogCPR2_0 was replaced with the transmembrane domains from *E. coli ycgG*. In the process, MogCPR2_0 was truncated removing the N-terminal 75 amino acids. Effect of these transmembrane domains with various point mutations on the production of mogrol was compared in *E. coli* strains producing mogrol through expression of SohB_C11CYP_6. Briefly, fermentations were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified UPLC-DAD for mogrol production. As shown in Table 32, the strains expressing MogCPR2_0 having the transmembrane domains from *E. coli ycgG* and various point mutations produced more mogrol compared to the strain expressing MogCPR2_0.

Table 32: Round 3 of C11CPR Engineering

Transmembrane domain	Mogrol titer
----------------------	--------------

ycgG_V219I	1.35
ycgG_L166F	1.33
ycgG_Q107K	1.26
ycgG_S163T	1.25
ycgG_K202Q	1.23
ycgG_D225E	1.20
ycgG_M257L	1.19
ycgG_D129E	1.19
ycgG_T226I	1.18
ycgG_Q82R	1.17
ycgG_ΔT273	1.15
ycgG_ΔK139	1.15
ycgG_S272A	1.14
ycgG_I228A	1.10

Based on these results, MogCPR2_1 (SEQ ID NO: 25), which has a N-terminal deletion of 75 amino acids and a transmembrane domain based on the N-terminal 25 amino acids of *E. coli* *ycgG* (i.e., n25_ycgG), was further studied in comparison with MogCPR2_0. Fermentations using bacterial strains producing mogrol through expression of SohB_C11CYP_6 and expressing MogCPR2_0 or MogCPR2_1 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by UPLC-DAD for mogrol production. As shown in **FIG. 8C**, the bacterial strains expressing MogCPR2_1 produced increased levels of mogrol compared to the bacterial strains expressing MogCPR2_0.

E. coli strains expressing mutant derivatives of MogCPR2_1 were screened. Briefly, fermentations using bacterial strains producing mogrol through expression of SohB_C11CYP_6 and expressing MogCPR2_1 or MogCPR2_1 derivatives were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by UPLC-DAD for mogrol production. Effects of the amino

acid substitutions in MogCPR2_1 were evaluated by comparing the titers of mogrol produced. Table 33 shows nine amino acid substitutions that showed beneficial changes in the titers of mogrol compared to MogCPR2_1:

Table 33: Round 4 of C11CPR Engineering

Mutation	Mogrol titer
D129E	1.23
T94V	1.18
I228A	1.17
L90P	1.17
Q82R	1.15
S163T	1.14
T226I	1.13
Q107K	1.12
M257L	1.11

5 Based on these results, MogCPR2_2 (SEQ ID NO: 26), which has a L90P substitution, was further studied in comparison with MogCPR2_0. Fermentations using bacterial strains producing mogrol through expression of SohB_C11CYP_6 and expressing MogCPR2_0 or MogCPR2_2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by UPLC-DAD for mogrol production. As shown in **FIG. 8D**, the bacterial strains expressing MogCPR2_2 produced
 10 about 1.25-fold increased levels of mogrol compared to the bacterial strains expressing MogCPR2_0.

Example 8: Engineering of Uridine Diphosphate Dependent Glycosyltransferase Enzymes (UGT) for C3-Specificity

Various UGT homologs were screened for their o-glycosylation activity on c3 or c24
 15 positions of mogrol. Briefly, enzyme assays for conversion of mogrol to MogIE or MogIA were carried out in a 96-well plate at 37°C for 48 hrs. Data were quantified by ultra-high performance

liquid chromatography coupled with triple quadrupole mass spectrometry (UPLC-QQQ) for MogIE or MogIE production. Table 34 shows the results:

Table 34: Screening of various UGT homologs for their o-glycosylation activity on c3 or c24 of mogrol

Name	C3 Activity	C24 Activity
AtUGT73C3 (SEQ ID NO: 50)	No	Yes
SgUGT720_269_1_E (SEQ ID NO: 51)	No	Yes
CmOUGT (SEQ ID NO: 52)	Yes	No
SgUGT720_269_4_Manus (SEQ ID NO: 53)	Yes	No
SgUGT720_269_4_E (SEQ ID NO: 54)	Yes	No
SgUGT74_345_2_I (SEQ ID NO: 55)	Yes	No
XP_023543158.1 (SEQ ID NO: 56)	Yes	No
XP_022151514.1 (SEQ ID NO: 57)	Yes	No
XP_023764969.1 (SEQ ID NO: 58)	Yes	No
XP_043637218.1 (SEQ ID NO: 59)	Yes	No
XP_008445481.1 (SEQ ID NO: 60)	Yes	No
XP_022132196.1 (SEQ ID NO: 61)	Yes	No
XP_023538293.1 (SEQ ID NO: 62)	Yes	No
UGT74AC1 (SEQ ID NO: 63)	Yes	Yes
UGT74AC1_M7 (SEQ ID NO: 64)	Yes	Yes

- 5 *Stevia rebaudiana* UGT85C1 (MogUGTc3_0, SEQ ID NO: 27) was engineered. Mutant derivatives of MogUGTc3_0 were screened in an enzyme assay for conversion of MogIA to MogIIE. Assays were carried out in a 96-well plate at 37°C for 48 hrs. Data were quantified by ultra-high performance liquid chromatography coupled with triple quadrupole mass spectrometry (UPLC-QQQ) for MogIIE production. Effects of the amino acid substitutions were evaluated by
- 10 comparing the production of MogIIE by a given MogUGTc3_0 amino acid substitution with that produced by MogUGTc3_0. Table 35 shows three amino acid substitutions that showed beneficial changes in conversion of MogIA to MogIIE compared to MogUGTc3_0:

Table 35: Round 1 of MogUGTc3 Engineering

Mutation	MogIIE titer
L41F	1.35
D49E	1.1
C127F	1.27

Based on these results, MogUGTc3_1 (SEQ ID NO: 28), which has L41F, D49E, and C127F substitutions, was further studied in comparison with MogUGTc3_0. Enzyme assays with MogUGTc3_0 or MogUGTc3_1 and MogIA substrate were carried out in a 96-well plate at 37°C for 48 hrs. Data were quantified by UPLC-QQQ for MogIIE production. As shown in **FIG. 9A**, MogUGTc3_1 enzyme produced about 2-fold increased levels of MogIIE compared to the MogUGTc3_0 enzyme.

E. coli strains producing mogrol and expressing mutant derivatives of MogUGTc3_1 were screened. Briefly, fermentations were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by ultra-high performance liquid chromatography coupled with triple quadrupole mass spectrometry (UPLC-QQQ) for MogIE production. Effects of the amino acid substitutions were evaluated by comparing the titers of MogIE produced in fermentations by bacterial strain expressing a given MogUGTc3_1 derivative with bacterial strain producing mogrol and expressing MogUGTc3_1. Table 36 shows thirty-three amino acid substitutions that showed beneficial changes in the titers of MogIE compared to MogUGTc3_1:

Table 36: Round 2 of MogUGTc3 Engineering

Mutation	MogIE titer
A307I	1.65
F219I	1.40
N302S	1.36
F215L	1.34
Y212N	1.33
F303Y	1.32

A222S	1.32
Y212D	1.31
F215V	1.27
E220D	1.26
E220S	1.23
F219Q	1.21
S400T	1.20
A214I	1.18
Q223H	1.18
R94V	1.17
L96I	1.17
A214V	1.17
Y212D	1.16
T221S	1.16
N211K	1.16
A222I	1.16
E220Q	1.16
L218V	1.15
I307V	1.12
Q223K	1.12
I205M	1.11
T102K	1.11
S400V	1.10

Y212F	1.10
E216D	1.10
E220G	1.10
R94A	1.10

Based on these results, MogUGTc3_2 (SEQ ID NO: 29), which has an A307I substitution, was further studied in comparison with MogUGTc3_1. Fermentations using bacterial strains expressing MogUGTc3_1 or MogUGTc3_2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by UPLC-QQQ for MogIE production. As shown in **FIG. 9B**, the bacterial strains expressing MogUGTc3_2 produced about 1.6-fold increased levels of MogIE compared to the bacterial strains expressing MogUGTc3_1.

E. coli strains expressing mutant derivatives of MogUGTc3_2 were screened for improved substrate selectivity. These efforts generated MogUGTc3_3 (SEQ ID NO: 30), which has T74M and F303Y substitutions. MogUGTc3_3 was further studied in comparison with MogUGTc3_2. Fermentations using bacterial strains producing mogrol and expressing MogUGTc3_2 or MogUGTc3_3 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by UPLC-QQQ for the production of MogIE and two glucosylated 24,25-dihydroxy cucurbitadienol derivatives, which are undesired side products. As shown in **FIG. 9C**, the bacterial strains expressing MogUGTc3_3 produced decreased levels of glucosylated 24,25-dihydroxy cucurbitadienol derivatives and about 1.25-fold increased levels of MogIE compared to the bacterial strains expressing MogUGTc3_2. These results suggest, *inter alia*, that MogUGTc3_3 exhibits improved substrate specificity (*e.g.*, discrimination of mogrol over 24,25-dihydroxy-cucurbitadienol) compared to MogUGTc3_2.

Cucumis melo UGTc3 (XP_008445481.1, SEQ ID NO: 60, “CmUGTc3_0”) was engineered. Mutant derivatives of CmUGTc3_0 were screened in an enzyme assay for conversion of Mogrol to MogIE. Assays were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by ultra-high performance liquid chromatography coupled with triple quadrupole mass spectrometry (UPLC-QQQ) for MogIE production. Effects of the amino acid substitutions were evaluated by comparing the production of MogIE by a given CmUGTc3_0 variant with that

produced by CmUGTc3_0. Table 37 shows thirty-four amino acid substitutions and a circular permutant (CP16) that showed beneficial changes in conversion of Mogrol to MogIE compared to CmUGTc3_0:

Table 37: Round 1 of CmUGTc3_0 Engineering

Mutation	MogIE titer
S408G	1.93
H207N	1.86
S17A	1.74
L47V	1.69
K225E	1.66
G161K	1.57
L46K	1.56
CP8	1.54
CP9	1.53
F71V	1.52
R450G	1.50
I422V	1.50
V173P	1.47
S178G	1.45
S224E	1.44
Δ V423	1.43
Q16K	1.42
CP10	1.37
CP7	1.36
Q347E	1.33
V211A	1.32
L59M	1.32
R397K	1.32
K254D	1.29
I41A	1.27

A257D	1.25
T303A	1.23
ΔT110	1.22
CP16	1.22
R138F	1.21
S350A	1.21
L155I	1.18
S418C	1.16
N228D	1.15
Q183E	1.10

Table 38: Circular Permutant Constructs

CP	Cut Site (Lower)	Linker	SEQ ID NO
7	P169	VEMTAANGGGERIK	SEQ ID NO: 73
8	P169	VAANGGGERGK	SEQ ID NO: 65
9	P177	VEMTAANGGGERIK	SEQ ID NO: 74
10	P177	VAANGGGERGK (	SEQ ID NO: 75
16	G250	VAANGGGERGK	SEQ ID NO: 76

CmUGTc3_1 (SEQ ID NO: 65), which is a circular permutation of CmUGTc3_0 having a cut-site at position 169, was further studied in comparison with CmUGTc3_0. A circular permutant retains the same basic fold of the parent enzyme, but has a different position of the N-terminus (e.g., “cut-site”), with the original N- and C-termini connected, optionally with an additional linker sequence, with respect to the wild type enzyme, optionally by a linking sequence. For example, in the circular permutants, the N-terminal Methionine is positioned at a site in the protein other than the natural N-terminus. Enzyme assays with CmUGTc3_0 or CmUGTc3_1 and mogrol substrate were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by UPLC-QQQ for MogIE production. As shown in **FIG. 9D**, CmUGTc3_1 enzyme produced over 1.5-fold increased levels of MogIE compared to the CmUGTc3_0 enzyme.

Mutant derivatives of CmUGTc3_1 were screened in an enzyme assay for conversion of Mogrol to MogIE. Assays were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by ultra-high performance liquid chromatography coupled with triple quadrupole mass spectrometry (UPLC-QQQ) for MogIE production. Effects of the amino acid substitutions were evaluated by comparing the production of MogIE by a given CmUGTc3_1 variant with that produced by CmUGTc3_1. Table 39 shows thirty-seven variants that showed beneficial changes in conversion of Mogrol to MogIE compared to CmUGTc3_1:

Table 39: Round 2 of CmUGTc3_0 Engineering

Mutation	MogIE titer
S11G	1.71
H40N_S310A_S241G_K58E	1.70
S241G_L352M_S11G_S57E	1.65
L352M_S11G	1.58
H40N_K58E	1.55
L448I	1.54
E441Q	1.54
H40N_G454K	1.53
H40N_S310A_K58E_S241G_L340V	1.52
S241G_S11G_K58E	1.51
H40N_S241G	1.50
H40N_S310A	1.49
H40N_S310A_S241G	1.47
H40N_L352M_G454K_S241G	1.46
H40N_L352M_G454K	1.45
K58E_L352M_S11G	1.44
H40N	1.42
L448V	1.32
H40N_L340V_S11G_S57E	1.29
G454K	1.27
L340V	1.22

L340I	1.22
I255V	1.21
H40G	1.19
S310P	1.17
H450Y	1.15
L352M	1.15
S57E	1.15
S310V	1.15
F364V	1.14
L339Q	1.12
S310A_S241G	1.11
A377E	1.11
S241G	1.10
R283G	1.10
L339K	1.10
A377D	1.10

Based on these results, CmUGTc3_2 (SEQ ID NO: 66), which has an L352M substitution with respect to CmUGTc3_1, was further studied in comparison with CmUGTc3_1 and another CmUGTc3_1 derivative having a L448V substitution. Enzyme assays with CmUGTc3_2, the L448V derivative of CmUGTc3_1, or CmUGTc3_1 and mogrol substrate were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by UPLC-QQQ for MogIE and deoxy-MogIE production. As shown in **FIG. 9E**, CmUGTc3_2 enzyme produced about 1.2-fold increased levels of MogIE and reduced levels of deoxy-MogIE to about 0.9 fold compared to the CmUGTc3_1 enzyme. The L448V derivative of CmUGTc3_1 produced about 1.3-fold increased levels of MogIE and about 1.2-fold increased levels of levels of deoxy-MogIE compared to the CmUGTc3_1 enzyme.

Based on the above results, CmUGTc3_3 (SEQ ID NO: 67), which has H40N, I216A, S241G, and M352L substitutions with respect to CmUGTc3_1 was further studied in comparison with CmUGTc3_2. Enzyme assays with CmUGTc3_3 or CmUGTc3_2 and mogrol substrate were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by UPLC-QQQ for MogIE

and deoxy-MogIE production. As shown in **FIG. 9F**, CmUGTc3_3 enzyme produced about 1.6-fold increased levels of MogIE and about 1.1-fold increased levels of levels of deoxy-MogIE compared to the CmUGTc3_2 enzyme.

Example 9: Engineering of Uridine Diphosphate Dependent Glycosyltransferase Enzyme (UGT) for C24-Specificity

Mutant derivatives of *Glycyrrhiza uralensis* (UGT73F24) (MogUGTc24_0, SEQ ID NO: 31) were screened. Briefly, enzyme assays for conversion of MogIE to MogIIE were carried out in a 96-well plate at 37°C for 48 hrs. Data were quantified by ultra-high performance liquid chromatography coupled with triple quadrupole mass spectrometry (UPLC-QQQ) for MogIIE production. Effects of the amino acid substitutions were evaluated by comparing the production of MogIIE produced by a given MogUGTc24_0 derivative with that produced by MogUGTc24_0. Table 40 shows five amino acid substitutions that showed beneficial changes in the titers of MogIIE compared to MogUGTc24_0:

Table 40: Round 1 of MogUGTc24 Engineering

Mutation	MogIIE titer
H101P	2.21
A74E	2.14
I91F	1.87
I436L	1.67
Q241E	1.54

Based on these results, MogUGTc24_1 (SEQ ID NO: 32), which has A74E, I91F, and H101P substitutions, was further studied in comparison with MogUGTc24_0. Enzyme assays for conversion of MogIE to MogIIE by MogUGTc24_1 or MogUGTc24_0 were carried out in a 96-well plate at 37°C for 48 hrs. Data were quantified by UPLC-QQQ for MogIIE production. As shown in **FIG. 10A**, MogUGTc24_1 enzyme produced about 3.5-fold increased levels of MogIIE compared to MogUGTc24_0 enzyme.

Mutant derivatives of MogUGTc24_1 were screened. Briefly, enzyme assays for conversion of MogIE to MogIIE were carried out in a 96-well plate at 37°C for 48 hrs. Data were quantified by ultra-high performance liquid chromatography coupled with triple quadrupole mass spectrometry (UPLC-QQQ) for MogIIE production. Effects of the amino acid substitutions were evaluated by comparing the conversion of MogIE to to MogIIE by a given MogUGTc24_1 derivative with that converted by MogUGTc24_1. Table 41 shows ninety-seven amino acid substitutions that showed beneficial changes in the titers of MogIIE compared to MogUGTc24_0:

Table 41: Round 2 of MogUGTc24 Engineering

Mutation	MogIIE titer
S88A	2.42
T184F	2.16
Y381F	2.09
S164P	2.09
H57N	2.05
H151V	1.97
D86E	1.96
I403V	1.94
H237N	1.90
T95A	1.86
P190E	1.86
F382Y	1.82
Q61H	1.80
H93Q	1.80
D379E	1.79
M96T	1.79
S256K	1.78
D109H	1.78
F196L	1.74
L18M	1.72
N306K	1.72

F105L	1.72
V174I	1.71
A161D	1.70
E219R	1.69
A240T	1.67
S287C	1.66
H93K	1.65
L230I	1.63
T273S	1.60
N78S	1.60
T80S	1.58
A160S	1.58
R108E	1.58
Q449R	1.58
M332L	1.58
H38V	1.56
L267V	1.53
Q452E	1.53
K100Q	1.53
S250D	1.53
V60I	1.52
H458Y	1.52
R441K	1.51
R421K	1.50
C214Y	1.49
R418A	1.47
E107Q	1.45
N327-	1.45
L97M	1.45
R220K	1.44

K100E	1.43
T459S	1.43
K329G	1.43
E74D	1.42
E400Q	1.42
E295Q	1.42
G163E	1.41
T83S	1.41
T179K	1.40
D417E	1.39
K182T	1.39
V76I	1.35
W471L	1.34
P157H	1.34
V188L	1.34
C27V	1.34
D104Q	1.33
K280S	1.32
G415K	1.32
N466E	1.32
I176L	1.31
D431E	1.30
Q241E	1.29
V106L	1.28
D467E	1.28
V434E	1.25
I221V	1.25
N326M	1.24
R99Q	1.24
Q435E	1.24

R438S	1.23
V426L	1.23
V236C	1.20
I114L	1.20
S47L	1.18
T51L	1.18
I387V	1.17
G345D	1.16
E212P	1.15
K185T	1.14
D169N	1.14
E216D	1.14
H56F	1.13
D104K	1.12
L202I	1.10
H173E	1.10

Based on these results, MogUGTc24_2 (SEQ ID NO: 33), which has T184F, Y381F, T95A, P190E, and L18M substitutions, in addition to the substitutions present in MogUGTc24_1, was further studied in comparison with MogUGTc24_1. Enzyme assays for conversion of MogIIE to MogIIE by MogUGTc24_2 or MogUGTc24_1 were carried out in a 96-well plate at 37°C for 48 hrs. Data were quantified by UPLC-QQQ for MogIIE production. As shown in **FIG. 10B**, MogUGTc24_2 enzyme produced about 4.5-fold increased levels of MogIIE compared to MogUGTc24_1 enzyme.

E. coli strains expressing mutant derivatives of MogUGTc24_2 were screened. Briefly, fermentations of mogrol producing strains were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by ultra-high performance liquid chromatography coupled with triple quadrupole mass spectrometry (UPLC-QQQ) for MogIA production. Effects of the amino acid substitutions were evaluated by comparing the titers of MogIA produced in fermentations by bacterial strain expressing a given MogUGTc24_2 derivative with bacterial strain expressing

MogUGTc24_2. Table 42 shows twenty-nine mutants that showed beneficial changes in the titers of MogIA compared to MogUGTc24_2:

Table 42: Round 3 of MogUGTc24 Engineering

Mutation	MogIA titer
D379E	2.75
T83S/S88A	2.70
S88A	2.61
H57N/S88A	2.52
I403V/R421K/Q449R	2.37
I403V	2.34
T273S/N306K/M332L	2.34
T273S/S287C/N306K	2.24
T80S_S88A_H93Q	2.13
D379E/F382Y	2.10
D86E/D87D	1.98
L230I/H237N/A240T	1.87
H38V/N78S/D109H	1.83
N306K	1.77
H38V/S88A	1.75
S88A/M96T/K100Q	1.72
S250D/T273S/N306K	1.69
Q449R	1.62
L230I	1.59
H38V/H57N/Q61H	1.59

Q452E	1.54
L267V	1.49
H237N	1.48
F196L/E219R	1.48
E219R	1.47
S250D/S256K	1.43
H57N	1.41
I145L	1.38
H38V	1.37

Based on these results, MogUGTc24_3 (SEQ ID NO: 34), which has S88A, T273S, N306K, and M332L substitutions, in addition to the substitutions present in MogUGTc24_2, was further studied in comparison with MogUGTc24_2. Fermentations using bacterial strains producing mogrol and expressing MogUGTc24_3 or MogUGTc24_2 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by UPLC-QQQ for MogIA production. As shown in **FIG. 10C**, the bacterial strains expressing MogUGTc24_3 produced about 4-fold increased levels of MogIA compared to the bacterial strains expressing MogUGTc24_2.

E. coli strains expressing mutant derivatives of MogUGTc24_3 were screened. Briefly, fermentations of mogrol producing strains were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by ultra-high performance liquid chromatography coupled with triple quadrupole mass spectrometry (UPLC-QQQ) for MogIA production. Effects of the amino acid substitutions were evaluated by comparing the titers of MogIA produced in fermentations by bacterial strain expressing a given MogUGTc24_3 derivative with bacterial strain expressing MogUGTc24_3. Table 43 shows ten mutants that showed beneficial changes in the titers of MogIA compared to MogUGTc24_3:

Table 43: Round 4 of MogUGTc24 Engineering

Mutation	MogIA titer
----------	-------------

C270S	1.251
S250D	1.207
F58I	1.180
H57N	1.155
H237N	1.150
I403V_R421K_Q449R	1.145
F196L_E219R	1.137
T405S	1.129
Y92L	1.115
Y92S	1.114

Based on these and other results, MogUGTc24_4 (SEQ ID NO: 48), which has F187Y substitution, in addition to the substitutions present in MogUGTc24_3, was further studied in comparison with MogUGTc24_3. Fermentations using bacterial strains producing mogrol and expressing MogUGTc24_3 or MogUGTc24_4 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by UPLC-QQQ for MogIA production. As shown in **FIG. 10D**, the bacterial strains expressing MogUGTc24_4 produced about 1.2-fold increased levels of MogIA compared to the bacterial strains expressing MogUGTc24_3.

E. coli strains expressing mutant derivatives of MogUGTc24_4 were screened. Briefly, fermentations of mogrol producing strains were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by ultra-high performance liquid chromatography coupled with triple quadrupole mass spectrometry (UPLC-QQQ) for MogIA and deoxy-MogIA production. Effects of the amino acid substitutions were evaluated by comparing the titers of MogIA and deoxy-MogIA produced in fermentations, and by calculating the improvement in the specificity to MogIA conversion by bacterial strain expressing a given MogUGTc24_4 derivative with bacterial strain expressing MogUGTc24_4. Table 44 shows twenty-three mutants that showed beneficial changes in the titers of MogIA compared to MogUGTc24_4, with several also showing improved specificity:

Table 44: Round 5 of MogUGTc24 Engineering

Mutation	MogIA titer	Deoxy MogIA titer	Specificity Improvement
----------	-------------	-------------------	-------------------------

S67A	1.40	1.33	Yes
I63V	1.39	1.08	Yes
T62Q	1.38	1.20	Yes
Q68K	1.24	1.16	Yes
Q50A	1.21	1.36	No
S247A	1.21	1.29	No
S162T	1.20	1.17	Yes
I132V	1.19	1.01	Yes
L230I	1.19	0.90	Yes
R108K	1.18	1.00	Yes
S35Q	1.16	1.35	No
S256K	1.14	1.22	No
T239N	1.14	1.08	No
Y283I	1.13	1.24	No
E253Q	1.12	1.30	No
V414I	1.12	1.33	No
L52I	1.12	1.23	No
V188M	1.12	1.40	No
M23L	1.11	1.06	No
S198A	1.11	1.08	No
H237N	1.10	1.10	No
D124S	1.10	1.02	No
D379E	0.95	0.59	Yes

Based on these results, MogUGTc24_5 (SEQ ID NO: 49), which has I63V substitution, in addition to the substitutions present in MogUGTc24_4, was further studied in comparison with MogUGTc24_4. Fermentations using bacterial strains producing mogrol and expressing MogUGTc24_5 or MogUGTc24_4 were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by UPLC-QQQ for MogIA and Deoxy MogA production. As shown in **FIG. 10E**, the bacterial strains expressing MogUGTc24_5 produced about 1.4-fold increased levels of

MogIA and only about 1.1 increased levels of Deoxy MogA compared to the bacterial strains expressing MogUGTc24_4.

E. coli strains expressing mutant derivatives of MogUGTc24_5 were screened. Briefly, fermentations of mogrol producing strains were carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by ultra-high performance liquid chromatography coupled with triple quadrupole mass spectrometry (UPLC-QQQ) for MogIA and deoxy-MogIA production. Effects of mutations were evaluated by comparing the titers of MogIA and deoxy-MogIA produced in fermentations, and by calculating the improvement in the specificity to MogIA conversion by bacterial strain expressing a given MogUGTc24_5 derivative with bacterial strain expressing MogUGTc24_5. Table 45 shows mutants that showed beneficial changes in the titers of MogIA compared to MogUGTc24_5:

Table 45: Round 6 of MogUGTc24 Engineering

Mutation	MogIA titer	Deoxy MogIA titer	Specificity Improvement
E190A	1.34	1.83	No
V404I	1.28	2.33	No
E190H	1.23	2.38	No
M371V	1.20	2.30	No
M23L	1.19	1.50	No
K306N	1.19	1.96	No
S247G	1.19	2.49	No
T239N	1.16	1.94	No
S35Q	1.16	1.38	No
V76C	1.15	1.95	No
H237A	1.15	1.98	No
F58V	1.13	1.20	No

F58I	1.10	1.17	No
T62Q/S67A/Q68K/R108K/I132V/S162T/L230I/D379E	1.10	0.79	Yes
S67A/L230I/D379E	1.04	0.68	Yes
R108A	0.95	0.70	Yes
M371L	0.95	0.75	Yes
V188L	0.88	0.54	Yes
I132V/S162T/L230I	0.85	0.43	Yes
Q68K/I132V/S162T/L230I/D379E	0.77	0.34	Yes

Example 10: Engineering of Uridine Diphosphate Dependent Glycosyltransferase Enzyme (UGT) for 1,6 and 1,2 Branched Glycosylations

Derivatives of *Siraitia grosvenorii* UGT94-289 (SEQ ID NO: 35) including a circular permutant thereof with cut-site at V187 (MogUGT1216_0, SEQ ID NO: 36) were screened. A circular permutant strategy for derivatizing UGT enzymes is disclosed in U.S. Patent No. 10,463,062, which is hereby incorporated by reference in its entirety. Enzymatic assays for conversion of MogIII_E to MogV were carried out using MogUGT1216_0 and its variants, in a 96-well plate at 37°C for 48 hrs. Data were quantified by UPLC-QQQ for MogV. Table 46 shows numerous derivatives that showed beneficial changes in the titers of MogV compared to MogUGT1216_0:

Table 46: Round 1 of Uridine Diphosphate Dependent Glycosyltransferase Enzyme (UGT) Engineering for 1,6 and 1,2 Branched Glycosylations

Mutation	MogV titer
L14V	3.71
H390D	3.06
I270K	2.89
ΔS332	2.86

M257I	2.73
A458S	2.68
L295I	2.63
E254D	2.57
M179K	2.53
A429L	2.35
Q121K	2.35
G279Q	2.27
G457E	2.20
N414E	2.18
D322S	2.18
L266Q	2.17
Δ G271-S274	2.17
L285V	2.12
A455L	2.12
Δ A127	2.02
N28K	1.98
V49I	1.97
D335S	1.91
Δ D349	1.90
M451K	1.88
S318P	1.86
S330E	1.84

A377S	1.83
F397L	1.81
T283R	1.80
S421A	1.80
Ins275_GS	1.80
S29T	1.71
$\Delta V3$	1.66
K72I	1.65
T4A	1.64
W160S	1.63
A108L	1.63
A450K	1.62
H112N	1.61
T281S	1.60
K6G	1.59
L355Y	1.57
$\Delta G457/\Delta A458$	1.56
S453R	1.55
Q61D	1.55
$\Delta E64$	1.54
$\Delta D7$	1.53
S23D	1.53
S251K	1.53

S374D	1.52
A388K	1.51
P326K	1.50
T430F	1.50
Ins3_GA	1.49
G63D	1.47
Δ G271	1.46
Ins459_V	1.46
S440P	1.44
S435E	1.42
Ins459_VT	1.37
Δ V3-K5	1.37
L423T	1.36
Q403E	1.33
S265V	1.33
P94L	1.30
T418S	1.29
G186A	1.28
P213E	1.27
Δ D275	1.26
F267L	1.24
H431K	1.23
S126K	1.23

V422M	1.21
N73E	1.18
V443Y	1.18
L427F	1.18
V24I	1.16
A297P	1.16
V3S	1.15
K148E	1.14
D214N	1.12
I248K	1.11

Based on these results, MogUGT1216_1 (SEQ ID NO: 37), which has L14V and L266Q substitutions, in addition to the circular permutation at V187 of SgUGT94_289_1, was further studied in comparison with SgUGT94_289_1 and MogUGT1216_0. Enzymatic assays for conversion of MogIIE to MogV were carried out using MogUGT1216_0, SgUGT94_289_1, and MogUGT1216_1 in a 96-well plate at 37°C for 48 hrs. Data were quantified by UPLC-QQQ for MogV. As shown in **FIG. 11A**, SgUGT94_289_1 produced about 10-fold increased levels of MogV compared to the circular permutant MogUGT1216_0. Interestingly, MogUGT1216_1 produced about 1.4-fold increased levels of MogV compared to SgUGT94_289_1 (**FIG. 11A**).

The effects of L14V and L266Q mutations present in MogUGT1216_1 were separately explored. Towards that, MogUGT1216_0 derivatives having L14V and L266Q single mutants were created and studied in comparison with MogUGT1216_0 (having no mutation). Enzymatic assays for conversion of MogIIE to various mogrosides were carried out using in a 96-well plate at 37°C for 48 hrs. Data were quantified by UPLC-QQQ for different mogrosides and plotted, normalizing the amounts of MogIII, MogIV-A, Siamenoside, and MogV produced to the productivity of MogUGT1216_0. L14V (Mut1) and L266Q (Mut16) single mutants produced an about 2 fold and about 5 fold increased amount of MogV, respectively (**FIG. 11B**). In contrast, as

shown in **FIG. 11B**, the L14V/L266Q double mutant produced about 18-fold increased levels of MogV compared to MogUGT1216_0. L14V single mutant and L266Q single mutant also produced substantially higher amounts of siamenoside (**FIG. 11B**).

Derivatives of MogUGT1216_1 were screened using enzymatic assays for conversion of MogIVA to MogV. The enzymatic assays were carried out using MogUGT1216_1 and its variants, in a 96-well plate at 37°C for 48 hrs. Data were quantified by UPLC-QQQ for MogV. Table 47 shows derivatives that showed beneficial changes in the titers of MogV compared to MogUGT1216_1:

Table 47: Round 2 of Uridine Diphosphate Dependent Glycosyltransferase Enzyme (UGT)

Engineering for 1,6 and 1,2 Branched Glycosylation

Mutation	MogV titer
W160P	2.93
ΔD349	2.28
ΔG271-G273	2.22
G419S	2.16
G457E	2.06
N16E	1.99
A15L	1.97
H390D	1.97
P329S	1.81
L249M	1.71
I25V	1.69
Q338E	1.68
ΔS332	1.68

P326H	1.65
M260A	1.63
I10L	1.49
P329N	1.49
L295V	1.48
H370L	1.48
K269G	1.47
C22S	1.47
P329G	1.47
E254D	1.47
C17S	1.44
T368A	1.44
P329A	1.44
K256D	1.41
S333Y	1.41
K227R	1.36
D335S	1.36
M257I	1.34
Q350E	1.31
A458S	1.31
A263E	1.31
L392V	1.28
S330E	1.28

K9S	1.25
K327R	1.22
Q121K	1.22
M145L	1.20
L295I	1.16
E58D	1.15
M38I	1.12
H19K	1.12
V222I	1.12
S374Q	1.10
P326S	1.10

Based on these results, MogUGT1216_2 (SEQ ID NO: 38), which has W160P and H390D substitutions, in addition to the mutations present in MogUGT1216_1, was further studied in comparison with MogUGT1216_1. Enzymatic assays for conversion of MogIIE to MogV were carried out using MogUGT1216_0, and MogUGT1216_1 in a 96-well plate at 37°C for 48 hrs. Data were quantified by UPLC-QQQ for MogV. As shown in **FIG. 11C**, MogUGT1216_2 (SEQ ID NO: 38) produced about 19-fold increased levels of MogV compared to MogUGT1216_1.

Derivatives of MogUGT1216_2 were screened in a Mog IIE production strain. Production of downstream mogrosides including MogV were quantified and the total UDP-glucose consumption was determined by the quantity of product produced and the number of total glycosylations catalyzed by the enzyme. The assay was carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by UPLC-QQQ for conversion of MogIIE to MogV. Table 48 shows derivatives that showed beneficial changes in the titers of MogV compared to MogUGT1216_2:

Table 48: Round 3 of Uridine Diphosphate Dependent Glycosyltransferase Enzyme (UGT)

Engineering for 1,6 and 1,2 Branched Glycosylation

Mutation	MogV titer
1216_2	1.03
C17S	1.42
H19E	1.66
H19K	1.14
C22S	1.29
T281S	1.49
S318K	1.97
Q338E	1.43
C343P	1.31
S421A	1.40
A450K	1.30
G457E	1.42

Based on these results, MogUGT1216_3 (SEQ ID NO: 90), which has H19E, T281S, and S318K substitutions, in addition to the mutations present in MogUGT1216_2, was further studied in comparison with MogUGT1216_2. Enzymatic assays for conversion of MogIIE to MogV were carried out using MogUGT1216_0, and MogUGT1216_1 in a 96-well plate at 37°C for 48 hrs.

5 Data were quantified by UPLC-QQQ for MogV. As shown in **FIG. 11D**, MogUGT1216_3 (SEQ ID NO: 90) produced about 1.6-fold increased levels of MogV compared to MogUGT1216_2.

Derivatives of MogUGT1216_3 were screened in a Mog IIE production strain. Production of downstream mogrosides including MogV were quantified and the total UDP-glucose consumption was determined by the quantity of product produced and the number of total

10 glycosylations catalyzed by the enzyme. The assay was carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by UPLC-QQQ for conversion of MogIIE to MogV. Table 49 shows derivatives that showed beneficial changes in the titers of MogV compared to MogUGT1216_2:

Table 49: Round 4 of Uridine Diphosphate Dependent Glycosyltransferase Enzyme (UGT)
Engineering for 1,6 and 1,2 Branched Glycosylation

Mutation	MogV titer
A203V_A209V	1.21
P94M	1.20
Y57R	1.20
Y57L	1.17
Q218S	1.16
Y432D	1.16
F267E	1.13
A108L	1.12
S333Y_D335Q	1.12
S435V_K436E_S440P	1.11
C17S	1.11
V43S	1.10

These results demonstrate that the above mutant derivatives provide a further improvement in the titers of MogV compared to MogUGT1216_2.

5 *Example 11: Engineering of Coffea arabica Uridine Diphosphate Dependent Glycosyltransferase*

Derivatives of *Coffea arabica* uridine diphosphate dependent glycosyltransferase 1,6 (MogUGT16_0, SEQ ID NO: 39) were constructed and screened using enzymatic assays for conversion of MogIIE to MogV. The enzymatic assays were carried out using MogUGT16_0 and its variants, in a 96-well plate at 37°C for 48 hrs. Data were quantified by UPLC-QQQ for MogV.

10 Table 50 shows derivatives that showed beneficial changes in the titers of MogV compared to MogUGT16_0:

Table 50: Round 1 of *Coffea arabica* Uridine Diphosphate Dependent Glycosyltransferase
Engineering

Mutation	MogV titer
----------	------------

G150F	64.02
T147L	54.17
N207K	37.82
K270E	30.96
L354V	24.03
V281L	22.90
L13F	15.37
M394V	9.76
T32A	9.63
V281Q	7.81
K101A	7.80
T256V	7.30
C219E	6.60
V175K	6.38
K270P	6.01
S43T	5.97
D285P	5.65
A377V	5.27
E74G	5.25
T303A	4.87
N283G	4.84
D95K	3.75
Y452N	1.75

F217L	1.06
-------	------

Based on these results, MogUGT16_1 (SEQ ID NO: 40), which has T147L and N207K substitutions was further studied in comparison with MogUGT16_0. Enzymatic assays for conversion of MogIIE to MogV were carried out using MogUGT16_0 and MogUGT16_1 in a 96-well plate at 37°C for 48 hrs. Data were quantified by UPLC-QQQ for MogV. As shown in **FIG. 12**, MogUGT16_1 produced more than about 40-fold increased levels of MogV compared to MogUGT16_0.

Example 12: Strategies to Increase the Production of 2,3;22,23-Dioxidosqualene Precursor

PntAB is a membrane-bound transhydrogenase that generates NADPH in *E. coli*. Since NADPH is required as a cofactor for each of the two oxygenation reactions catalysed by the squalene epoxidase (**FIG. 1**), to provide additional NADPH for the two oxygenation reactions catalysed by the squalene epoxidase, a strain that overexpresses native *E. coli pntA-pntB* genes (by gene complementation) was constructed and the levels of 2,3;22,23-dioxidosqualene produced by this strain were analysed. Briefly, fermentation was carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by GC-FID for 2,3;22,23-dioxidosqualene production. As shown in **FIG. 13A**, the strain overexpressing *pntAB* (SEQ ID NOs: 41 and 42) produced about 1.4-fold more 2,3;22,23-dioxidosqualene, compared to a strain that does not overexpress *pntAB*. These results indicate, *inter alia*, that extra NADPH produced by PntAB enabled more squalene to be converted into 2,3-oxidosqualene and then subsequently to 2,3;22,23-dioxidosqualene.

The next question addressed was whether the overexpression of native *E.coli pntA-pntB* genes would increase 24,25-epoxy cucurbitadienol production if enzymes that catalyze epoxidation of squalene to 2,3;22,23-dioxidosqualene and cyclization of 2,3;22,23-dioxidosqualene to 24,25-epoxy cucurbitadienol were coexpressed. Therefore, *E. coli* strains expressing SQS, SQE, ECDS, and EPH with or without the overexpression of native *E.coli pntA-pntB* were constructed and studied. Briefly, fermentation was carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by GC-FID for 24,25-epoxy cucurbitadienol production (24,25 dihydroxy cucurbitadienol measured as a proxy). As shown in **FIG. 13B**, shows the strain expressing *pntAB* produced about 1.9-fold more 24,25-epoxy cucurbitadienol, compared to a strain that does not overexpress *pntAB*. These results indicate, *inter alia*, that extra NADPH produced by

PntAB enabled more squalene to be converted into 2,3-oxidosqualene and then subsequently to 2,3;22,23-dioxidosqualene.

To further explore the role of supplementation of reduction power, strain having genomically integrated cassettes overexpressing native *E.coli fldA* (SEQ ID NO: 43) or
5 heterologous Cv.fdx (an fdx homolog from *Allochromatium vinosum*, SEQ ID NO: 44) were constructed in strains that express SQS and SQE. These strains were studied in comparison with a wild type strain that overexpresses SQS and SQE without overexpression of *E.coli fldA* or *Cy.fdx*. Briefly, fermentation was carried out in a 96-well plate at 37°C for 72 hrs. Data were quantified by GC-FID for squalene, 2,3-oxidosqualene, and 2,3;22,23-dioxidosqualene production. As shown
10 in **FIG. 13C**, the strain overexpressing native *E.coli fldA* produced about 4-fold increased levels of 2,3;22,23-dioxidosqualene compared to the strain that only overexpresses SQS and SQE. Similarly, the strain overexpressing heterologous Cv.fdx also produced more 2,3-oxidosqualene, and 2,3;22,23-dioxidosqualene compared to the strain that only that overexpresses SQS and SQE. There results indicate, *inter alia*, that provision of extra electron carrier such as ferredoxin or
15 flavodoxin, may results in availability of extra NADPH, which enabled more squalene to be converted into 2,3-oxidosqualene and then subsequently to 2,3;22,23-dioxidosqualene.

Example 13: Increased Production of Mogrol and 24,24-Dihydroxycucurbitadienol in a Strain having Reduced Expression of the YbhGFSR ABC Type Transporter

Overexpression of the ABC pumps, such as that encoded by the *ybhGFSR* operon, can be
20 exploited for increasing the production of compounds in microbial cells. Accordingly, derivatives of a strain expressing the entire metabolic pathway for producing mogrol were constructed and tested.

For these experiments, a parental strain that expresses the entire metabolic pathway for producing mogrol was modified and evaluated. To evaluate the role of the *ybhGFSR* operon, a
25 strain having the *ybhGFSR* operon deleted from the chromosome was constructed and evaluated in comparison with the parental *ybhGFSR*⁺ strain. The strains were grown in a 96-well plate at 37°C for 72 hrs. Samples of the entire cultures, including cell and culture supernatant, were extracted and analyzed. Data were quantified by LC-DAD for 2,3;22,23-dioxidosqualene, 24,25-dihydroxy cucurbitadienol, and mogrol production and normalized with the production by the
30 parental strain. Surprisingly, as shown in **FIG. 14A**, the *ybhGFSR* operon deletant strain exhibited

increased 24,25-dihydroxy cucurbitadienol + mogrol production compared to the parental strain. The *ybhGFSR* operon deletant strain exhibited decreased production of 2,3;22,23-dioxidosqualene compared to the parental strain (**FIG. 14A**).

To further evaluate this phenomenon, the parental strain that expresses the entire metabolic pathway for producing mogrol was modified to create an hypomorphic *ybhG* allele. Briefly, the *ybhG* ribosome binding site in the chromosome was replaced with a weaker version than the native sequence to create the *ybhG* hypomorphic allele. The strain having the hypomorphic *ybhG* allele was evaluated in comparison with the parental *ybhGFSR*⁺ strain. The strains were grown in a 96-well plate at 37°C for 72 hrs. Samples of the entire cultures, including cell and culture supernatant, were extracted and analyzed. Data were quantified by LC-DAD for 2,3;22,23-dioxidosqualene, 24,25-dihydroxy cucurbitadienol, and mogrol production and normalized with the production by the parental strain. As shown in **FIG. 14B**, the strain harboring the *ybhG* hypomorphic allele exhibited increased 24,25-dihydroxy cucurbitadienol + mogrol production compared to the parental strain. The strain harboring the *ybhG* hypomorphic allele also exhibited decreased production of 2,3;22,23-dioxidosqualene compared to the parental strain (**FIG. 14B**).

To understand the effect of overexpression of the *ybhGFSR* operon, the parental strain that expresses the entire metabolic pathway for producing mogrol was modified to overexpress the *ybhGFSR* operon. Briefly, the *ybhF* gene deleted from the chromosome and the *ybhGFSR* operon was overexpressed from a plasmid. The strain overexpressing the *ybhGFSR* operon was evaluated in comparison with the parental *ybhGFSR*⁺ strain. The strains were grown in a 96-well plate at 37°C for 72 hrs. Samples of the entire cultures, including cell and culture supernatant, were extracted and analyzed. Data were quantified by LC-DAD for 2,3;22,23-dioxidosqualene, 24,25-dihydroxy cucurbitadienol, and mogrol production and normalized with the production by the parental strain. As shown in **FIG. 14C** (top panel), the strain that overexpresses of the *ybhGFSR* operon exhibited decreased 24,25-dihydroxy cucurbitadienol + mogrol production compared to the parental strain, with an increase in 2,3;22,23-dioxidosqualene, 24,25-dihydroxy cucurbitadienol, and mogrol collectively. The strain that overexpresses of the *ybhGFSR* operon is confirmed to exhibit a substantial increased production of 2,3;22,23-dioxidosqualene compared to the parental strain (**FIG. 14C** (bottom panel)).

These results demonstrate that hypomorphic or null mutations in the *ybhGFSR* operon increase the production of mogrol and/or its immediate precursor 24,25-dihydroxy cucurbitadienol, while decreasing the production of 2,3;22,23-dioxidosqualene. These results also demonstrate that the overexpression of the *ybhGFSR* operon decreases the production of mogrol
5 and/or its immediate precursor 24,25-dihydroxy cucurbitadienol, while increasing the production of 2,3;22,23-dioxidosqualene. Without wishing to be bound by theory, it is believed that the YbhGFSR ABC type transporter increases efflux of 2,3;22,23-dioxidosqualene, preventing its conversion to downstream products. Reduced expression of the *ybhGFSR* transporter increases intracellular (int.) 2,3;22,23-dioxidosqualene, and improves conversion to downstream products
10 24,25-dihydroxy cucurbitadienol and mogrol.

SEQUENCES

Squalene Epoxidase

Methylomonas lenta SQE (SEQ ID NO: 1)

5 MAKEEFDICIIIGAGMAGATISAYLAPKGIKIALIDHCYKEKKRIVGELLQPGAVLSLEQMGLSH
LLDGFEAQTVKGYALLQGNEKTTIPYPSQHEGIGLHNGRFLQQIRASALENSSVTQIHGKALQL
LENERNEIIGVSYRESITSQIKSIYAPLTITSDGFFSNFRAHLSNNQKTVTSYFIGLILKDCM
PFPKHGHVFLSGPTPFICYPISDNEVRLLIDFPGEQLPRKNLLQEHLDTNVTPIPECMRSSYA
QAIQEGGFKVMPNHYMAAKPIVRKGAVMLGDALNMRHPLTGGGLTAVFSDIQILSAHLLAMPDF
10 KNTDLIHEKIEAYYRDRKRANANLNILANALYAVMSNDLLKTAVFKYLQCGGANAQESI AVL
LNRKHFSLIKQFCFLAVFGACNLLQQSISNIPKALKLLKDAFVIIKPLIKNELS

SQE_1 (SEQ ID NO: 2)

MAKEEFDICIIIGAGMAGATISAYLAPKGIKIALIDRCYKEKKRIVGELLQPGAVLSLEQMGLSH
LLDGFEAQTVKGYALLQGNEKTTIPYPSQHEGIGLHNGRFLQQIRASALENSSVTQIHGKALQL
LENERNEIIGVSYRESITSQIKSIYAPLTITSDGFASNFRAHLSNNQKTVTSYFIGLILKDCM
15 PFPKHGHVFLSGPTPFICYPISDNEVRLLIDFPGEQLPRKNLLQEHLDTNVTPIPECMRSSYA
QAIQEGGFKVMPNHYMAAKPIVRKGAVLLGDALNMRHPLTGGGLTAVFSDIQILSAHLLAMPDF
KNTDLIHEKIEAYYRDRKRANANLNILANALYAVMSNDLLKTAVFKYLQCGGANAQESI ALLAG
LNRKHFSLIKQYCFVAVFGACNLLQQSISNIPKALKLLKDAFVIIKPLIKNELS

SQE_2 (SEQ ID NO: 3)

20 MAKEEFDICIIIGAGMAGATISAYLAPKGIKIALIDRCYKEKKRIVGELLQPGAVLSLEQMGLSH
LLDGFEAQTVKGYALLQGNEKTTIPYPSQHEGIGLHNGRFLQQIRASALENESVTQIHGKALQL
LENERGEIIGVSYRESHTSQIKSIYAPLTITSDGFASNFRRHLSNNQKTVTSYFIGLILKDCM
PFPKHGHVFLSGPTPFICYPISDNEVRLLIDFPGEQLPRKNLLQEHLDTNVTPIPECMRSSYA
QAIQEGGFKVMPNHYMAAKPIVRKGAVLLGDALNMRHPLTGGGLTAVFSDIQILSAHLLAMPDF
25 KNTDLIHEKIEAYYRDRKRANANLNILANALYAVMSNDLLKTAVFKYLQCGGANAQESI ALLAG
LNRKHFSLIKQYCFVAVFGACNLLQQSISNIPKALKLLKDAFVIIKPLIKNELS

SQE_4 (SEQ ID NO: 4)

MAKEEFDICIIIGAGMAGATISAYLAPKGIKIALIDRCYKEKKRIVGELLQPGAVLSLEQMGLSH
LLDGFEAQTVKGYALLQGNEKTTIPYPSQHEGIGLHNGRFLQQIRASALENESVTQIHGKALQL
30 LENERGEIIGVSYRESHTSQIKSIYAPLTITSDGFNSNFRRHLSNNQKTVTSYFIGLILKDCM
PFPKHGHVFLSGPTPFICYPISDNEVRLLIDFPGEQLPRKNLLQEHLDTNVTPIPECMRSSYA
QAIQEGGFKVMPNHYMAAKPIVRKGAVLLGDALNMRHPLTGGGLTAVFSDIQILSAHLLAMPDF
KNTDLIHEKIEAYYRDRKRANANLNILANALYAVMSNDLLKTAVFKYLQCGGANAQESI ALLAG
LNRKHFSLIKQYCFVAVFGACNLLQQSISNIPKALKLLKDAFVIIKPLIKNELS

35

Methylococcus capsulatus SQE (McSQE_0) (SEQ ID NO: 77)

MASSIELGEWDVLIAGGSVAGSAAAAAALSGLGRLVLIVEPDPDPGRRLAGELIHPPGIDGLLEL
GLIHDDVPQGSVVNGFAIFPNDGEGAPATLLPYGEIHGRQRCGRVIEHTLLKSHLLETVRGFE
RVSVWLGARVTGMEHEDGKGYVATVTHEGTETRMEVRLIIGADGPMSQLRKMVGISHETQRYSG
MIGLEVEDTHLPNPGYGNIFLNPAGVSYAYGIGGGRRARVMFEVLKGADSKESIRDHLRLFPAPF
5 RGDIEAVLAQ GKPLAAANYCIVPEASVKANVALVGDARGCCHPLTASGITAAVKDAFVMRDALQ
ATGLNFEAALKRYSVQCGRLQLTRRTLAEELREAFLAQTPEAELLSQCIFSYWRNSPKGRQASM
ALLSTLDSSIFSLASQYTLVGLQAFRLLPQWLGA KMGGDWFRGVAQLVSKSLKFQQDALNQALR
AK

10 *McSQE_1* (SEQ ID NO: 78)

MASSIELGEWDVLIAGGSVAGSAAAAAALSGLGRLVLIVEPDPDPGRRLAGELIHPPGIDGLLEL
GLIHDDVPQGSVVNGFAIFPNDGEGAPATLLPYGEIHGRQRCGRVIEHTLLKSHLLETVRGFE
RVSVWLGARVTGMEHEDGKGYVATVTHEGTETRMEVRLIIGADGPMSQLRKMVGISHETQRYSG
MIGLEVEDTHLPNPGYGNIFLNPAGVSYAYGIGGGRRARVMFEVLKGADSKESIRDHLRLFPAPF
15 RGDIEAVLAQ GKPLAAANYCIVPEASVKANVALVGDARGCCHPLTASGITAAVKDAFVMRDALQ
ATGLNFEAALKRYSVQCGRLQLTRRTLAEELREAFLAQTPEAELLRQCIFSYWRNSPKGRQASM
ALLSTLDSSIFSLASQYTLVGLQAFRLLPQWLGA KMGGDWFRGVAQLVSKSLKFQQDALNQALR
AK

20 *McSQE_2* (SEQ ID NO: 79)

MASSIELGEWDVLIAGGSVAGSAAAAAALSGLGRLVLIVEPHDPDPGRRLAGELIHPPGIDGLLEL
GLIHDDVPQGSVVNGFAIFPNDGEGAPATLLPYGEIHGRQRCGRVIEHTLLKSHLLETVRGFE
RVSVWLGARVTGMEHEDGKGYVATVTHEGTETRMEVRLIIGADGPMSQLRKMVGISHETQRYSG
MIGLEVEDTHLPNPGYGNIFLNPAGVSYAYGIGGGRRARVMFEVLKGADSKESIRDHLRLFPAPF
25 RGDIEAVLAQ GKPLAAANYCIVPEASVKANVALVGDARGCCHPLTASGITAAVKDAFVLRDALQ
ATGLNFEAALKRYSVQCGRLQLTRRTLAEELREAFLAQTPEAELLRQCIFSYWRNSPKGRQASM
ALLSTLDSSIFSLASQYTLVGLQAFRLLPQWLGA KMGGDWFRGVAQLVSKSLKFQQDALNQALR
AK

30

Epoxide Hydrolases

Siraitia grosvenorii EPH3 (EPH_0) (SEQ ID NO: 5)

MADQIEHITINTNGIKMHIA SVGTGPVVLHGHFPELWYSWRHQLLYLSSVGYRAIAPDLRGY
DTDSPASPTSYTALHIVGDLV GALDELGIEKVFVGHWDGAI IAWYFCLFRPDRIKALVNLSVQ
35 FIPRNPAIPFIEGFRTAFGDDFYMC RFQVPGEAEEDFASIDTAQLFKTSLCNRSSAPPCLPKEI
GFRAIPPPENLPSWLTEEDINYAAKFKQTGFTGALNYYRAFDLTWELTAPWTGAQIQVPVKFI
VGDSDLTYHFPGAKEYIHNGGFKKDVPLLEEVVVVKDACHFINQERPQEINAHIHDFINKF

EPH_1 (SEQ ID NO: 6)

MADQIEHITINTNGIKMHIA SVGTGPV VLL LHGFPELWYSWRHQ LLYLSSVGYRAIAPDLRGY
 DTDAPASPTSYTALHIVGDLV GALDELGIEKVFLVGH DWGAI IAWYFCLFRPDRIKALVNLSVA
 FIPRNP AIPFIESFRQVFGDDFYMCRFQVPGEAEADFASIDTAQLFKTSLCNRSSAPPCLPKGI
 5 GFRAIPPPENLP SWLTEEDINYYAAKFQGTGTGALNYYRAFDLTWELTAPWTGAQIQVPVKFI
 VGDSDMTYHFPGAKEYIHNGGFKKDVP LLEE VVVKDAGHFIQ QERPQEINAHIHDFINKF

Cucurbitadienol Synthases (CDS)

Citrullus colocynthis (CcCDS2; ECDS_0) (SEQ ID NO: 7)

10 MAWRLKVGAE SVGEKEEKWLKSI SNHLGRQVWEFCAHQPTASPNHLQQIDNARNHFRNNRFHRK
 QSSDLFLAIQNEKEIANVTKGGGIKVKEEEDVRKETVKNTVERALSFYSAIQ TNDGNWASDLGG
 PMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYLYNHQNE DGGWGLHIEGTSTMFGSALNYVAL
 RLLGEDADGGEGGAMTKARSWILDRGGATAITSWGKLWLSVLGVYEWSGNNPLPPEFWLLPYCL
 PFHPGRMWCHCRMVYLPMSYLYGKRFGVGPITPIVLSLRKELYTIPYHEIDWNRSRNTCAKEDLY
 15 YPHPKMQDILWGSIIYHLYEPLFTRWPGKRLREKALQMAMKHIHYEDENSRYICLGPVNKVLNML
 CCWVEDPYSDAFKFHLQRVPDYLWVAEDGMRMQGYNGSQLWDTAFSVQAIISTKLIDSFGTTLK
 KAHDFVKDSQIQQDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLMLS KLPSKIV
 GEPLEKSRLCDAVNVLLSLQNEGGFASYELTRSYPWLELINPAETFGDIVIDYPYVECTSATM
 EALTLFKKLHPGHRTKEIDIAVARAANFLENMQRTDGSWYGCWGVCFITYAGWFGIKGLVAAGRT
 20 YNSCVAIRKACDFLLSKELPGGGWGESYLSQCQNKVYTNLEGNRPHLVNTAWVLMALIEAGQAER
 DPAPLHRAARLLINSQLENGDFPQEEIMGVFNKNCMITAAYRNI FPIWALGEYFHRVLTE

ECDS_1 (SEQ ID NO: 8)

MAWRLKVGAE SVGEKEEKWLKSI NNHLGRQVWEFDAHQPTASPNHLQQIDNARNHFRNNRFHRK
 QSSDLFLAIQNEKEIANVTKGGGIKVKEEEDVRKETVKNTVERALSFYSAIQ TNDGNWASDLGG
 25 PMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYLYNHQNE DGGWGLHIEGTSTMFGSALNYVAL
 RLLGEDADGGEGGAMTKARSWILDRGGATAITSWGKLWLSVLGVYEWSGNNPLPPEFWLLPYCL
 PFHPGRMWCHCRMVYLPMSYLYGKRFGVGPITPIVLSLRKELYTIPYHEIDWNRSRNTCAKEDLY
 YPHPKMQDILWGSIIYHLYEPLFTRWPGKRLREKALQMAMKHIHYEDENSRYICLGPVNKVLNML
 CCWVEDPYSDAFKFHLQRVPDYLWVAEDGMRMQGYNGSQLWDTAFSVQAIISTKLIDSFGTTLK
 30 KAHDFVKDSQIQQDCPGDPNVWFRHIHKGAWPFSTRDHGWLISDCTAEGLKASLMLS KLPSKIV
 GEPLEKSRLCDAVNVLLSLQNEGGFASYELTRSYPWLELINPSETFGDIVIDYPYVECTSATM
 EALTLFKKLHPGHRTKEIDIAVARAANFLENMQRTDGSWYGCWGVCFITYAGWFGIKGLVAAGRT
 YNSCVAIRKACDFLLSKELPGGGWGESYLSQCQNKVYTNLEGNRPHLVNTAWVLMALIEAGQAER
 DPAPLHRAARLLINSQLENGDFPQEEIMGVFNKNCMITAAYRNI FPIWALGEYFHRVLTE

35 ECDS_2 (SEQ ID NO: 9)

MAWRLKVGAE SVGEKEEKWLKSI NNHLGRQVWEFDAHQPTASPNHLQQIDNARNHFRNNRFHRK
 QSSDLFLAIQNEKEIANVTKGGGIKVKEEEDVRKETVKNTVERALSFYSAIQ TNDGNWASDLGG
 PMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYLYNHQNE DGGWGLHIEGTSTMFGSALNYVAL
 RLLGEDADGGEGGAMTKARSWILDRGGATAITSWGKLWLSVLGVYEWSGNNPLPPEFWLLPYCL

PFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIVLSLRKELYTIPYHEIDWNRSRNTCAKEDLY
YPHPKMQDILWGSIIYHLYEPLFTRWPGKRLREKALQMAMKHIHYEDENSRYICLGPVNKVLNML
CCWVEDPYSDAFKFHLLQRPDYLWVAEDGMRMQGYNGSQLWDTAFSVQAIISTKLIDSFGTTLK
KAHDFVKDSQIQQDCPGDPNVWFRHIHKGAWPFSTRDHGWLVS DCTAEG LKASLMLS KLPSKIV
5 GEPLKSR L CDAVNVL LSLQ NENG GFAS YEL TRSY PWLE LMNP SETFG DIVIDY PYVECT SATM
EALTLFKKLHPGHRTKEIDIAVARAANFLENMQRTDGSWYGCWGVCFITYAGWFGIKGLVAAGRT
YNSCVAIRKACDFLLSKELPGGGWGESYLSCQNKVYTNLEGNRPHLVNTAWVLMALIEAGQAER
DPAPLHRAARLLINSQLENGDFPQEEIMGVFNKNCMITYAAYRNIFPIWALGEYFHRVLTE

ECDS_3 (SEQ ID NO: 10)

10 MAWRLKVGAE SVGEKEEKWLKS INNHLGRQVWEFDAHQPTASPNHLQQIENARNHFRNNRFHRK
QSSDLFLAIQNEKEIANVTKGGGIKVKEEEDVRKETVKNTVERALSFYSAIQTNDGHWASDLGG
PMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYLYNHQNE DGGWGLHIEGTSTMFGSALNYVAL
RLLGEDADG GEGGAMTKARSWILDRGGATAITSWGKLWLSVLGVYEWSGNNPLPPEFWLLPYCL
PFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIVLSLRKELYTIPYHEIDWNRSRNTCAKEDLY
15 YPHPKMQDILWGSIIYHLYEPLFTRWPGKRLREKALQMAMKHIHYEDENSRYICLGPVNKVLNML
CCWVEDPYSDAFKFHLLARVPDYLWVAEDGMRMQGYNGSQLWDTAFSVQAIISTKLIDSFGTTLK
KAHDFVKDSQIQQDCPGDPNVWFRHIHKGAWPFSTRDHGWLVS DCTAEG LKASLMLS KLPSKIV
GEPLKSR L CDAVNVL LSLQ NENG GFAS YEL TRSY PWLE LMNP SETFG DIVIDY PYVECT SATM
EALTLFKKLHPGHRTKEIDIAVARAANFLENMQRTDGSWYGCWGVCFITYAGWFGIKGLVAAGRT
20 YNSCVAIRKACDFLLSKELPGGGWGESYLSCQNKVYTNLEGNRPHLVNTAWVLMALIEAGQAER
DPAPLHRAARLLINSQLENGDFPQEEIMGVFNKNCMITYAAYRNIFPIWALGEYFHRVLTE

ECDS_4 (SEQ ID NO: 11)

MAWRLKVGAE SVGEKEEKWLKS INNHLGRQVWEFDAHQPTASPNHLQQIENARNHFRNNRFHRK
QSSDLFLAIQNEKEIANVTKGGGIKVKEEEDVRKETVKNTVERALSFYSAIQTNDGHWASDLGG
25 PMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYLYNHQNE DGGWGLHIEGTSTMFGSALNYVAL
RLLGEDADG GEGGAMTKARSWILDRGGATAITSWGKLWLSVLGVYEWSGNNPIPPEFWLLPYCL
PFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIVLSLRKELYTIPYHEIDWNRSRNTCAKEDLY
YPHPKMQDILWGSIIYHLYEPLFTRWPGKRLREKALQMAMKHIHYEDENSRYICLGPVNKVLNML
CCWVEDPYSDAFKFHLLARVPDYLWVAEDGMRMQGYNGSQLWDTAFSVQAIISTKLIDSFGTTLK
30 KAHDFVKDSQIQQDCPGDPNVWFRHIHKGAWPFSTRDHGWLVS DCTAEG LKASLMLS KLPSKIV
GEPLKSR L CDAVNVL LSLQ NENG GFAS YEL TRSY PWLE LMNP SETFG DIVIDY PYVECT SATM
EALTLFKKLHPGHRTKEIDIAVARAANFLENMQRTDGSWYGCWGVCFITYAGWFGIKGLVAAGRT
YNSCVAIRKACDFLLSKELPGGGWGESYLSCQNKVYTNLEGNRPHLVNTAWVLMALIEAGQAER
DPAPLHRAARLLINSQLENGDFPQEEIMGVFNKNCMITYAAYRNIFPIWALGEYFHRVLTE

35 ECDS_5 (SEQ ID NO: 12)

MAWRLKVGAE SVGEKEEKWLKS INNHLGRQVWEFDAHQPTASPNHLQQIENARNHFRNNRFHRK
QSSDLFLAIQNEKEIANVTKGGGIKVKEEEDVRKETVKNTVERALSFYSAIQTNDGHWASDLGG
PMFLLPGLVIALYVTGVLNSVLSKHHRQEMCRYLYNHQNE DGGWGLHIEGTSTMFGSALNYVAL
RLLGEDADG GEGGAMTKARSWILDRGGATAITSWGKLWLSVLGVYEWSGNNPIPPEFWLLPYCL
40 PFHPGRMWCHCRMVYLPMSYLYGKRFVGPITPIVLSLRKELYTIPYHEIDWNRSRNTCAKEDLY
YPHPKMQDILSGSIIYHLYEPLFTRWPGKRLREKALQMAMKHIHYEDENSRYICLGPVNKVLNML
CCWVEDPYSDAFKFHLLARVPDYLWVAEDGMRMQGYNGSQLWDTAFSVQAIISTKLIDSFGTTLK

KAHDFVKDSQIQQDCPGDPNVWFRHIHKGAWPFSTRDHGWLVS DCTA EGLKASLM LSKLP SKIV
 GEPLK SRLCD AVNVL LSLQ NENG GFASYEL TRSY PWLE LMNP SETFGD IVIDY PYVECT SATM
 EAL TLFK LHPG HRTKE IDIA VARA ANFLEN MQRTD GSWYGC WGVCF TYAGW FGIKGLVAAGR T
 YNSCVAIRKACDFLLSKELPGGGWGESYLS CQNKVY TNLEGNR PHLVNTAWVLMALIEAGQAER
 5 DPAPLHRAARLLINSQLENGDFPQEEIMGVFNKNCMITYAAYRNI FPIWALGEYFHRVLTE

SgCDS (SEQ ID NO: 68)

MAWRLKVGAE SVGENDEKWLKSI SNHLGRQVWEFCPDAGTQQQLLQVHKARKAFHDDRFRHKQS
 SDFITIQYKGEVENGKGTAGVKLKEGEEVRKEAVESSLERALS FYSSIQTSDGNWASDLGGPM
 FLLPGLVIALYVTGVLNSVLSKHHRQEMCRYVYNHQ NEDGGWGLHIEGPSTMFGSALNYVALRL
 10 LGEDANAGAMPKARAWILDHGGATGITSWGLWLSVLGVYEWSGNNPLPPEFWLFPYFLPFHPG
 RMWCHCRMVYLPMSYLYGKRFGVGPITPIVLSLRKELYAVPYHEIDWNKSRNTCAKEDLYPHPK
 MQDILWGLSLHHVYEPLFTRWPAKRLREKALQTAMQHIHYEDENTRYICLGPVNKVLNLLCCWVE
 DPYSDAFKLHLQRVHDYLWVAEDGMKMQGYNGSQLWDTAFS IQAIVSTKLVDNYGPTLRKAHDF
 VKSSQIQQDCPGDPNVWYRHIHKGAWPFSTRDHGWLIS DCTA EGLKAALM LSKLP SETVGESE
 15 RNRLCD AVNVL LSLQ NDNG GFASYEL TRSY PWLE LINPAETFGD IVIDY PYVECT SATMEALTL
 FKKLHPG HRTKE IDTAIVRAANFLENMQRTD GSWYGC WGVCF TYAGW FGIKGLVAAGR TYNNCL
 AIRKACDFLLSKELPGGGWGESYLS CQNKVY TNLEGNR PHLVNTAWVLMALIEAGQAERDPTPL
 HRAARLLINSQLENGDFPQEEIMGVFNKNCMITYAAYRNI FPIWALGEYCHRVLTE

PsCAS_mut (SEQ ID NO: 69)

MWKLKVAEGGTPWLRTLNNHVGRQVWEFDPHSGSPQDLDDIETARRNFHDNRFTHKHSDDLMLR
 LQFAKENPMNEVL PKVKVDVEDVTEEAVATTLRRGLNFYSTIQSHDGHWPGLGGPMFLMPGL
 VITLSVTGALNAVLTD EHRKEMRRYLYNHQNKDGGWGLHIEGPSTMFGSVLCYVTLRLLGEPN
 DGECDMERGRDWILEHGGATYITSWGKMWLSVLGVFEWSGNNPMPPEIWLLPYALPVHPGRMWC
 HCRMVYLPMSYLYGKRFGVGPITPTVLSLRKELFTVPYHDIDWNQARNLCAKEDLYPHPLVQDI
 25 LWATLHKFVEPVFMNWP GKKLREKAIKTAIEHIHYEDENTRYICIGPVNKVLNMLCCWVEDPNS
 EAFKLHLPRIYDYLWVAEDGMKMQGYNGSQLWDTAFAAQAIISTNLIDEFGPTLKAHAFIKNS
 QVSEDCPGDL SKYRHI SKGAWPFSTADHGWPIS DCTA EGLKAVLLLSKIAPEIVGEPLDSKRL
 YDAVNVL LSLQ NENGGLATYELTRSYTWLEIINPAETFGD IVIDCPYVECTSAAIQALATFGKL
 YPGHRREEIQCCIEKAVAFIEKIQASDGSWYGSWGVCF TYGTWFGIKGLIAAGKNFSNCL SIRK
 30 ACEFLLSKQLPSGGWAESYLS CQNKVYSNLEGNRSHVVNTGWAMLALIEAEQAKRDPTPLHRAA
 VCLINSQLENGDFPQEEIMGVFNKNCMITYAAYRCIFPIWALGEYRRVLQAC

CpCAS_mut (SEQ ID NO: 70)

MAWQLKIGADTVPSDPSNAGGWLSTLNNHVGRQVWHFHPGLGSPEDLQQIQQARQHFSDHRFEK
 KHSADLLMRMQFAKENSSFVNLPQVKVKDKEDVTEEAVTRTLRRAINFYSTIQADDGHWPGLG
 35 GPMFLIPGLVITLSITGALNAVLSTEHQREICRYLYNHQNKDGGWGLHIEGPSTMFGSVLNVT
 LRLLGEEAEDGQGAVDKARKWILDHGGAAAITSWGKMWLSVLGVYEWAGNNPLPPELWLLPYLL
 PCHPGRMWCHCRMVYLPMSYLYGKRFGVGPITPIIRSLRKELYLVYPYHEVDWNKARNQCAKEDLY
 YPHPLVQDILWATLHHVYEPLFMHWP AKRLREKALQSV MQHIHYEDENTRYICIGPVNKVLNML
 CCWAEDPHSEAFKLHIPRIYDYLWIAEDGMKMQGYNGSQLWDTAFVQAIISTELAEYETTLR
 40 KAHKYIKDSQVLEDCPGDLQSWYRHI SKGAWPFSTADHGWPIS DCTA EGLKAVLLLSKLPSEIV
 GKSID EQQLYNAVNVILSLQNTDGGFATYELTRSYRWLELMNPAETFGD IVIDY PYVECSAAI
 QALAAFFKKLYPGHRRDEIDNCIAEAAADFIESIQATDGSWYGSWGVCF TYGGWFGIRGLVAAGRR

YNNCSSLRKACDFLLSKELAAGGWGESYLSQNKVYTNIKDDRPHIVNTGWAMLSLIDAGQSER
DPTPLHRAARVLINSQMEDGDFPQEEIMGVFNKNCMISYSAYRNIFPIWALGEYRSRVLKPLK

AaCAS_mut (SEQ ID NO: 71)

5 MWKLKIAEGGDPWLRRTTNDHIGRQIWEFDPTLGSVEELAEIEKLRKTRDNRFEKKHSADLLMR
SQFAKENSVSVPKVNKDVEDITEDKVTNVLRRRAIGFSTLQADDGHWPGDLGGPMFLLPGL
VITLSITGALNAVLSKEHKREMCRYLYNHQNIIDGGWGLHIEGHSTMFGSALNYVTLRLLGEGAN
DGEGAMEKGRKWILDHGGATAITSWGKFWLSVLGVFEWPGNNPLPPEMWLLPYFLPVHPGRMWC
HCRMVYLPMSYLYGKRFGPITSTVLALRKEFTVPYHDIDWNEARNLCAKEDLYYPHPLIQDV
10 LWATLDFVEPVLMSWPGKKLREKALRTAMEHIHYEDENTRYICIGPVNKVLNMLCCWVEDPNS
EAFKLHLPRIQDYLWIAEDGMKMQGYNGSQLWDAAFTVQAIMSTNLIEFGPTLKKGHIFIKKS
QVLNDCYGDLDYWRHISKGAWPFSTADHGWPISDCTAEGKKAALLSKLPSEIVDEPLDAKRF
YDAVNVLISLMNADGSFATYELTRSYSWLELINPAETFGDIVIDYPYVECTSAAIQALVAFKRL
YPGHRDEVOGCIDKAAAFLEKIQEADGSWYGSWAVCFTYGTWFGVKGLVAAGKNYSNCSSIRK
ACNFLLSKQLASGGWGESYLSQVCKVYTNLEGNRSHVNTGWAMLALIDAEQAKRDPTPLHRAA
15 RVLINSQMENGEFPQEEIMGVFNRCMITYAAYRNIFPIWALGEYRCRVLKVET

Squalene Synthase (SQS)

Artemisia annua SQS (SEQ ID NO: 13)

20 MASSLKAVLKHPDDFYPLLKLKMAAKKAEKQIPSQPHWAFSYSMLHKVSRSFALVIQQLNQPQLR
DAVCIFYLVLRALDTEDDTSIAADIKVPIILIAFHKHIYNRDWHFACGTKEYKVLMDDQFHHVST
AFLELKRGYQEAIEDITMRMGAGMAKFICKEVETVDDYDEYCHYVAGLVGIGLSKLFHSSGTEI
LFSDSISNSMGLFLQKTNIIRDYLEINEIPKSRMFWPRIWISKYVNKLEDLKYEENSEKAVQC
LNDMVTNALIHIEDCLKYMSQLKDPAIFRFAIPQIMAIGTLALCYNNIEVFRGVVKLRRGLTA
KVIDRKTMDADVQAQFSDFDMLKSKVDMHPNAQTTITRLEAAQKICKDSGTLSNRKSYIVKR
ESSYSAALLALLFTILAILYAYLSANRPNKIKFTL

25 Cytochrome P450

Cucurbita pepo subsp. *pepo* cytochrome P450 enzyme CYP87A3 (SEQ ID NO: 14)

30 MAWAIIVGLATLAVAYYIHWINKWKDSKFNGLVPPGTMGLPLVGETLQLARPSDSLVDVHPFIKK
KVKRYGPIFKTCLAGRPVVVSTDAEFNNYIMLQEGRAVEMWYLDTLKFFGLDTEWLKALGFH
KYIRSITLNHFGAESLRERFLPRIEESAKETLRYWATQPSVEVKDSAAVMVFRTSMVKMVEDS
SKLLTGGLTKKFTGLGGFLTLPINVPGTTYNKCMKDMKEIQKKLREILEGRLASGGGSDDEDFL
GQAIKDKGSQQFISDDFIQLLFSISFASFESISTTLTLVLNYLADHPDVVKELEAEHEAIRNA
RADPDGPITWEEYKSMTFTLHVIFETLRLGSVTPALLRKTTKELQINGYTIPEGWTVMLVTASR
HRDPAVYKDPHTFNPWRWKELDSITIQKNFMPFGGLRHCAGAEYSKVYLCTFLHILFTKYRWT
35 KLKGGKVARAHILSFEDGLHVKFTPKE

sohB_CppCYP87A3 (SEQ ID NO: 15)

MALLSEYGLFLAKIVTVVLAIAAIAAIIHWINKWKDSKFNGLVPPGTMGLPLVGETLQLARPSD
SLDVHPFIKKKVKRYGPIFKTCLAGRPVVVSTDAEFNNYIMLQEGRAVEMWYLDTLKFFGLDT

EWLKALGFIHKYIRSITLNFHGAESLRERFLPRIEESAKETLRYWATQPSVEVKDSAAVMVFRT
SMVKMVSEDSSKLLTGGLTKKFTGLLGGFLLPINVPGTTYNKCMKDMKEIQKKLREILEGRLA
SGGGSDEDFLGQAIKDKGSQQFISDDFIQLLFSISFASFESISTTLTLVLNYLADHPDVVKEL
EAEHEAIRNARADPDGPITWEEYKSMTFTLHVIFETLRLGSVTPALLRKTTELQINGYTIPEG
5 WTVMLVTASRRDPVAVYKDPHTFNPWRWKELDSITIQKNFMPFGGGLRHCAGAEYSKVYLCTFL
HILFTKYRWTCLKGGKVARAHILSFEDGLHVKFTPKE

sohB_C11CYP_1 (SEQ ID NO: 16)

MALLSEYGLFLAKIVTVVLAIAAIAAIIHWINKWKDSKFNGVLPPGTMGMLPLVGETLQLARPSD
SLDVHVPFIKKKVKRYGPIFKTCLAGRPVVVSTDAEFNNYIMLQEGRAVEMWYLDTLKFFGLDT
10 EWLKALGFIHKYIRSITLNFHGAESLRERFLPRIEESAKETLRYWATQPSVEVKDSAAVMVFRT
TMVKMVSEDSSKLLTGGLTKKFTGLLGGFLLPINVPGTTYNKCMKDMKEIQKKLREILEGRLA
SGGGSDEDFLGQAIKDKGSQQFISDDFIQLLFSISFASFESISTTLTLVLNYLADHPDVVKEL
EAEHEAIRNARADPDGPITWEEYKSMTFTLHVIFETLRLGSVTPALLRKTTELQINGYTIPEG
WTVMLVTASRRDPVAVYKDPHTFNPWRWKELDSITIQKNFMPFGGGLRHCAGAEYSKVYLCTFL
15 HILFTKYRWTCLKGGKVARAHILSFEDGLHVKFTPKE

SohB_C11CYP_2 (SEQ ID NO: 17)

MALLSEYGLFLAKIVTVVLAIAAIAAIIHWINKWKDSKFNGVLPPGTMGMLPLVGETLQLARPSD
SLDVHVPFIKKKVKRYGPIFKTCLAGRPVVVSTDAEFNNYIMLQEGRAVEMWYLDTLKFFGLDT
ENLKALGFIHKYIRSITLNFHGAESLRERFLPRIEESAKETLRYWATQPSVEVKDSAAVMVFRT
20 TMVKMVSEDSSKLLTGGLTKKFTGLLGGFLLPINVPGTTYHKCMKDMKEIQKKLREILEGRLA
SGGGSDEDFLGQAIKDKGSQQFISDDFIQLLFSISFASFESISTTLTLVLNYLADHPDVVKEL
EAEHEAIRNARADPDGPITWEEYKSMTFTLHVIFETLRLGSVTPALLRKTTELQINGYTIPEG
WTVMLVTASRRDPVAVYKDPHTFNPWRWKELDSITIQKNFMPFGGGLRHCAGAEYSKVYLCTFL
HILFTKYRWTCLKGGKVARTHILSFEDGLHVKFTPKE

25 sohB_C11CYP_3 (SEQ ID NO: 18)

MALLSEYGLFLAKIVTVVLAIAAIAAIIHWINKWKDSKFNGVLPPGTMGMLPLVGETLQLARPSD
SLDVHVPFIKKKVKRYGPIFKTCLAGRPVVVSTDAEFNNYIMLQEGRAVEMWYLDTLKFFGLDT
ENLKALGFIHKYIRSITLNFHGAESLRERFLPRIEESAKETLRYWATQPSVEVKDSAAVMVFRT
TMVKMVSEDSSKLLTGGLTKKFTGLLGGFLLPINVPGTTYHKCMKDMKEIQKKLREILEGRLA
30 SGGGSDEDFLGQAIKDKGSQQFISDDFIQLLFSISFASFESISTTLTLVLNYLADHPDVVKEL
EAEHEAIRNARADPDGPITWEEYKSMTFTLHVIFETLRLGSVTPALLRKTTELQINGYTIPEG
WTVMLVTASRRDPVAVYKDPHTFNPWRWKELDSITIQKNFMPFGGGLRHCAGAEYSKVYLCTFL
HILFTKYRWTCLKGGKVARTHILSFEDGLHVKFTPKE-

sohB_C11CYP_4 (SEQ ID NO: 19)

MALLSEYGLFLAKIVTVVLAIAAIAAIIHWINKWKDSKFNGVLPPGTMGMLPLVGETLQLARPSD
SLDVHVPFIKKKVKRYGPIFKTCLAGRPVVVSTDAEFNNYIMLQEGRAVEMWYLDTLKFFALDG
ENLKALGFIHKYIRSITLNFHGAESLRERFLPRIEESAKETLRYWATQPSVEVKDSAAVMVFRT
TMVKMVSEDSSKLLTGGLTKKFTGLLGGFLLPINVPGTTYHKCMKDMKEIQKKLREILEGRLA
SGGGSDEDFLGQAIKDKGSQQFISDDFIQLLFSISFASFESISTTLTLVLNYLADHPDVVKEL
40 EAEHEAIRNARADPDGPITWEEYKSMTFTLHVIFETLRLGSVTPALLRKTTELQINGYTIPEG

WTVMLVTASRHRDPAVYKDPHTFNPWRWKELDSITIQKNFMPFGGGLRFCAGAEYSKVYLCTFL
HILFTKYRWTKLKGKGVARTHILSFEDGLHVKFTPKE

sohB_C11CYP_5 (SEQ ID NO: 20)

5 MALLSEYGLFLAKIVTVVLAIAAAIAAIHWINWKWDSKFNGVLPPGTMGLPLVGETLQLARPSD
SLDVHPFIKKKVKRYGPIFKTCLAGRPVVVSTDAEFNNYIMLQEGRAVEMWYLDTLSKFFALDG
ENVKALGFIHKYIRSITLNHFGAESLRERFLPRIEESAKETLRYWATQPSVEVKDSAAVMVVRT
TMVKMVSEDSSKLLTGGLTKKFTGLLGGFLTLPINVPGTTHKCMKDMKEIQKKLREILEGRLA
SGGGSDEDFLGQAIDKGSQQFISDDFIQLLFSISFASFESISTTLTLVLNLYADHPDVVKEL
EAEHEAIRNARADPDGPITWEEYKSMTFTLHVIFETLRLGSVTPALLRKTTELQINGYTIPEG
10 WTVMLVTASRHRDPAVYKDPHTFNPWRWKELDSITIQKNFMPFGGGLRFCAGAEYSKVYLCTFL
HILFTKYRWTKLKGKGVARTHILSFEDGLHVKFTPKE

sohB_C11CYP_6 (SEQ ID NO: 21)

15 MALLSEYGLFLAKIVTVVLAIAAAIAAIHWINWKWDSKFNGVLPPGTMGLPLVGETLQLARPSD
SLDVHPFIKKKVKRYGPIFKTCLAGRPVVVSTDAEFNNYIMLQEGRAVEMWYLDTLSKFFALDG
ENLNALGFIHKYIRSITLNHFGAESLRERFLPRIEESAKETLRYWATQPSVEVKDSAAVMVVRT
TMVKMVSEDSSKLLTGGLTKKFTGLLGGFLTLPINVPGTTHKCMKDMKEIQKKLREILEGRLA
SGGGSDEDFLGQAIDKGSQQFISDDFIQLLFSISFASFESISTTLTLVLNLYADHPDVVKEL
EAEHEAIRNARADPDGPITWEEYKSMTFTLHVIFETLRLGSVTPALLRKTTELQINGYTIPEG
WTVMLVTASRHRDPAVYKDPHTFNPWRWKELDSITIQKNFMPFGGGLRFCAGAEYSKVYLCTFL
20 HILFTKYRWTKLKGKGIARTHILSFEDGLHVKFTPKE

sohB_C11CYP_8 (SEQ ID NO: 45)

25 MALLSEYGLFLAKIVTVVLAIAAAIAAIHWINWKWDSKFNGVLPPGTMGLPLVGETLQLARPSD
SLDVHPFIKKKVKRYGPIFKTCLAGRPVVVSTDAEFNNYIMLQEGRAVEMWYLDTLSKFFALDG
ENLNALGFIHKYIRSITLNHFGAESLRERFLPRIEESAKETLRYWATQPSVEVKDSAAVMVVRT
TMVKMVSEDSSKLLTGGLTKKFTGLIAGFLTLPINVPGTTHKCMKDMKEIQKKLREILEGRLA
SGGGSDEDFLGQAIDKGSQQFISDDFIQLLFSISFASFESISTTLTLVLNLYADHPDVVKEL
EAEHEAIRNARADPDGPITWEEYKSMTFTLHVIFETLRLGSVTPALLRKTTELQINGYTIPEG
WTVMLVTASRHRDPAVYKDPHTFNPWRWKELDSITIQKNFMPFGGGLRFCAGAEYSKVYLCTFL
HILFTKYRWTKLKGKGIARTHILSFEDGLHVKFTPKE

30 sohB_C11CYP_9 (SEQ ID NO: 46)

35 MALLSEYGLFLAKIVTVVLAIAAAIAAIHWINWKWDSKFNGVLPPGTMGLPLVGETLQLARPSD
SLDVHPFIKKKVKRYGPIFKTCLAGRPVVVSTDAEFNNYIMLQEGRAVEMWYLDTLSKFFALDG
ENLNALGFIHKYIRSITLNHFGAESLRERFLPRIEESAKETLRYWATQPSVEVKDSAAVMVVRT
TMVKMVSEDSSKLLTGGLTKKFTGLIAGFLTLPINVPGTTHKCMKDMKEIQKKLREILEGRLA
SGGGSDEDFLGQAIDKGSQQFISDDFIQLLFSISFASFESISTTLTLVLNLYADHPDVVKEL
EAEHEAIRNARADPDGPITWEEYKSMTFTLHVIFETLRLGSVTPALLRKTTELQINGYTIPEG
WTVMLVTASRHRDPAVYKDPHTFNPWRWKELDSITIQKNFMPFGGGLRFCAGAEYSKVYLCTFL
HILFTKYRWTKLKGKGIARTHILSFEDGLHVKFTPKE

sohB_C11CYP_10 (SEQ ID NO: 47)

MALLSEYGLFLAKIVTVVLAIAAAIAAI IHWINKWKDSKFNGVLPPGTMGLPLVGETLQLARPND
SLDVHVPFIKKRVKRYGPIFKTCLAGRPVVVSTDAEFNHYIMLQEGRAVEMWYLDTLSKFFALDG
ENLNALGFIHKYIRSITLNFHGAESLRERFLPRIEESAKETLRYWATQPSVEVKDSAAVMVVRT
TMVKMVSEDSSKLLTGGLTKKFTGLIAGFLTLPINVPGTTYHKCMKDMKEIQKKLREILEGRLA
5 SGGGSDEDFLGQAIKDKGSQQFISDDFI IQLLFSISFASFESISTTLTLVLNLYLADHPDVVKEL
EAEHEAIRNARADPDGPITWEEYKSMTFTLHVIYETLRLGSVTPALLRKTTELQINGYTIPEG
WTVMMLVTASRRDPAVYKDPHTFNPWRWKELDSITI QKNFMPFGGGLRFCAGAEYSKVYLCTFL
HILFTKYRWTCLKGGKIARTHILSFEDGLHVKFTPKE

10 sohB_C11CYP_11 (SEQ ID NO: 80)

MALLSEYGLFLAKIVTVVLAIAAAIAAI IHWINKWKDSKFNGVLPPGTMGLPLVGETLQLARPND
SLDVHVPFIKKRVKRYGPIFKTCLAGRPVVVSTDAEFNHYIMLQEGRAVEMWYLDTLSKFFALDG
ENLNALGFIHKYIRSITLNFHGAESLRERFLPRIEESAKETLRYWATQPSVEVKDSAAVMVVRT
TMVKMVSEDSSKLLTGGLTKKFTGLIAGFLTLPINVPGTTYHKCMKDMKEIQKKLREILEGRLA
15 SGGGSDEDFLGQAIKDKGSQQFISDDFI IQLLFSISFASFESISTTLTLVLNLYLADHPDVVKEL
EAEHEAIRNARADPDGPITWEEYKSMTFTLHVIYETLRLGSVTPALLRRTTELQINGYTIPEG
WTVMMLVTASRRDPAVYKDPHTFNPWRWKELDSITI QKNFMPFGGGLRFCAGAEYSKVYLCTFL
HILFTKYRWTCLKGGKIARTHILSFEDGLHVKFTPKE

sohB_C11CYP_12 (SEQ ID NO: 81)

20 MALLSEYGLFLAKIVTVVLAIAAAIAAI IHWINKWKDSKFNGVLPPGTMGLPLVGETLQLARPND
SLDVHVPFIKKRVKRYGPIFKTCLAGRPVVVSTDAEFNHYIMLQEGRAVEMWYLDTLSKFFALDG
ENLNALGFIHKYIRSITLNFHGAESLRERFLPRIEESAKETLRYWATQPSVEVKDSAAVMVVRT
TMVKMVSEDSSKLLTGGLTKKFTGLIAGFLSLPINIPGTTYHKCMKDMKEIQKKLREILEGRLA
SGGGSDEDFLGQAIKDKGSQQFISDDFI IQLLFSISFASFESISTTLTLVLNLYLADHPDVVKEL
25 EAEHEAIRNARADPDGPITWEEYKSMTFTLHVIYETLRLGSVTPALLRRTTELQINGYTIPEG
WTVMMLVTASRRDPAVYKDPHTFNPWRWKELDSITI QKNFMPFGGGLRFCAGAEYSKVYLCTFL
HILFTKYRWTCLKGGKIARTHILSFEDGLHVKFTPKE

sohB_C11CYP_13 (SEQ ID NO: 82)

30 MALLSEYGLFLAKIVTVVLAIAAAIAAI IHWINKWKDSKFNGVLPPGTMGLPLVGETLQLARPND
SLDVHVPFIKKRVKRYGPIFKTCLAGRPVVVSTDAEFNHYIMLQEGRAVEMWYLDTLSKFFALDG
ENLNALGFIHKYIRSITLNFHGAESLRERFLPRIEESAKETLRYWATQPSVEVKDSAAVMVVRT
TMVKMVSEDSSKLLTGGLTKKFTGLIAGFLSLPINIPGF AFHKCMKDMKEIQKKLREILEGRLA
SGGGSDEDFLGQAIKDKGSQQFISDDFI IQLLFSISFASFESISTTLTLVLNLYLADHPDVVKEL
EAEHEAIRNARADPDGPITWEEYKSMTFTLHVIYETLRLGSVTPALLRRTTELQINGYTIPEG
35 WTVMMLVTASRRDPAVYKDPHTFNPWRWKELDSITI QKNFMPFGGGLRFCAGAEYSKVYLCTFL
HILFTKYRWTCLKGGKIARTHILSFEDGLHVKFTPKE

sohB_C11CYP_14 (SEQ ID NO: 83)

MALLSEYGLFLWKIVTVVLAIAAAIAAI IHWINKWKDSKFNGVLPPGTMGLPLVGETLQLARPND
SLDVHVPFIKKRVKRYGPIFKTCLAGRPVVVSTDAEFNHYIMLQEGRAVEMWYLDTLSKFFALDG

ENLNALGFIHKYIRSITLNFHGAESLRERFLPRIEESAKETLRYWATQPSVEVKDSAAVMVFRL
TMVKMVSEDSSKLLTGGLTKKFTGLIAGFLSLPINIPGFAFHKCMKDMKEIQKKLREILEGRLA
SGGGSDEDFLGQAIKDKGSQQFISDDFI IQLLFSISFASFESISTTTLVLVNLADHPDVVKEL
EAEHEAIRNARADPDGPITWEEYKSMTFTLHVIYETLRLGSVTPALLRRTTKELQINGYTIPEG
5 WTVMLVTASRRDPVYKDPHTFNPWRWKELDSITI QKNFMPFGGGLRFCAGAEYSKVYLCTFL
HILITKYRWTCLKGGKIARTHILSFEDGLHVKFTPKE

sohB_C11CYP_15 (SEQ ID NO: 84)

MALLSEYGLFLWKIVTVVLAIAAIAAGIHWINKWKDSKFNGVLPPGTMGLPLVGETLQLARPND
SLDVHPPFIKKRMKRYGPIFKTCLAGRPVVVSTDAEFNHYIMLQEGRAVEMWYLDTLSKFFALD
10 ENLNALGFIHKYIRSITLNFHGAESLRERFLPRIEESAKETLRYWATQPSVEVKDSAAVMVFRL
TMVKMVSEDSSKLLTGGLTKKFTGLIAGFLSLPINIPGFAFHKCMKDMKEIQKVLREILEGRLA
SGGGSDEDFLGQAIKDKGSQQFISDDFI IQLLFSISFASFESISTTTLVLVNLADHPDVVKEL
EAEHEAIRNARADPDGPITWEEYKSMTFTLHVIYETLRLGSVTPALLRRTTKELQINGYTIPEG
WTVMLVTASRRDPVYKDPHTFNPWRWKELDSITI QKNFMPFGGGLRFCAGAEYSKVYLCTFL
15 HILITKYRWTCLKGGKIARTHILSFEDGLHVKFSPKE

n20_C11CYP_15 (SEQ ID NO: 85)

MAWINKWKDSKFNGVLPPGTMGLPLVGETLQLARPNDSLDVHPPFIKKRMKRYGPIFKTCLAGRP
VVVSTDAEFNHYIMLQEGRAVEMWYLDTLSKFFALDGENLNALGFIHKYIRSITLNFHGAESLR
ERFLPRIEESAKETLRYWATQPSVEVKDSAAVMVFRLTMVKMVSEDSSKLLTGGLTKKFTGLIA
20 GFLSLPINIPGFAFHKCMKDMKEIQKVLREILEGRLASGGGSDEDFLGQAIKDKGSQQFISDDF
IIQLLFSISFASFESISTTTLVLVNLADHPDVVKELEAEHEAIRNARADPDGPITWEEYKSMT
FTLHVIYETLRLGSVTPALLRRTTKELQINGYTIPEGWTVMLVTASRRDPVYKDPHTFNPWR
WKELDSITI QKNFMPFGGGLRFCAGAEYSKVYLCTFLHILITKYRWTCLKGGKIARTHILSFED
GLHVKFSPKE

25 n20_C11CYP_16 (SEQ ID NO: 86)

MAWINKWRNRKFKGKLPPGTMGLPLVGETLQLARPNDSLDVHPPFVRERVEKYGPIFKTCLAGRP
VVVSTDAEFNHYIMLQEGRAVEMWYLDTLSKFFALDGENLNALGFIHKYIRSITLNFHGAESLR
ERFLPRIEESAKETLRYWATQPSVEVKDSAAVMVFRLTMVKMVSEDSSKLLTGGLTKKFTGLIA
GFLSLPINIPGFAFHKCMKDMKEIQKVLREILEGRLASGGGSDEDFLGQAIKDKGSQQFISDDF
30 IIQLLFSISFASFESISTTTLVLVNLADHPDVVKELEAEHEAIRNARADPDGPITWEEYKSMT
FTLHVIYETLRLGSVTPALLRRTTKELQINGYTIPEGWTVMLVTASRRDPVYKDPHTFNPWR
WKELDSITI QKNFMPFGGGLRFCAGAEYSKVYLCTFLHILITKYRWTCLKGGKIARTHILSFED
GLHVKFSPKE

n20_C11CYP_17 (SEQ ID NO: 87)

MAWINKWRNRKFKGKLPPGTMGLPLVGETLQLARPNDSLDVHPPFVRERVEKYGPIFKTCLAGRP
VVVSTDAEFNHYIMLQEGRAVEMWYLDTLSKFFALDGENLNALGFIHKYIRSITLNFHGAESLR
ERFLPRIEESAKETLRYWATQPSVEVKDSAAVMVFRLTMVKMVSEDSSKLLTGGLTKKFTGLIA
GFLSLPINIPGFAFHKCMKDMKEIQKVLREILEGRLASGGGSDEDFLGQAIKDKGSQQFISDDF
35 IIQLLFSISFASFESISTTTLVLVNLADHPDVVKELEAEHEAIRNARADPDGPITWEEYKSMT
FTLHVIYETLRLGSVTPALLRRTTKELQINGYTIPEGWTVMLVTASRRDPVYKDPHTFNPWR
40

WKELDSITIQKNFMPFGGLRFCAGAEYSKVYLCTFLHILITKYRWTKLKGGKIARTHILSFED
GLHVKFSPKE

n20_C11CYP_18 (SEQ ID NO: 88)

MAWINKWRNRKFKGKLPPGTMGLPLVGETLQLARPNDSLDVHPFVRERVEKYGPIFTCTLAGRP
5 VVVS TDAEFNHYIMLQEGRAVEMWYLD TLSKFFALDGENLNALGFIHKYIRSITLNHFGAESLR
ERFLPRIEESAKETLRYWATQPSVEVKDSAAVMVFR LTMVKMVSEDSSKLLTGGLTKKFTGLLK
GFLSLPINIPGF AFHKCMKDMKEIQKVLREILEGR LASGGGSD EDFLGQA IKDKGSQQFISDDF
IIQLLFSISFASFESISTTTLTLVLN LADHPDVVKELEAEHEAIRNARADPDGPITWEEYKSM T
FTLHV IYETLRLGSVTPALLRRTTKELQINGYTIPEGWTVMLVTASRHRDPAVYKDPHTFNPWR
10 WKELDSITIQKNFMPFGGLRFCAGAEYSKVYLCTFLHILITKYRWTKLKGGKIARTHILSFED
GLHVKFSPKE

n20_C11CYP_19 (SEQ ID NO: 89)

MAWINKWRNRKFKGKLPPGTMGLPLVGETLQLARPNDSLDVHPFVRERVEKYGPIFTCTLAGRP
VVVS TDAEFNHYIMLQEGRAVEMWYLD TLSKFFALDGENLNALGFIHKYIRSITLNHFGAESLR
15 ERFLPRIEESAKETLRYWATQPSVEVKDSAAVMVFR LTMVKMVSEDSSKLLTGGLTKKFTGLLK
GFLSLPINIPGF AFHKCMKDMKEIQKVLREILEGR LASGGGSD EDFLGQA IKDKGSQQFISDDF
IIQLLFSISFASFESISTTTLTLVLN LADHPDVVKELEAEHEAIRNARADPDGPITWEEYKSM T
FTLHV IYETLRLGSVTPALLRRTTKELQINGYTIPEGWTVMLVTASRHRDPAVYKDPHTFNPWR
WKELDSITIQKNFMPFGGLRYCAGADFAKVYLCTFLHILITKYRWTKLKGGKIARTHILSFED
20 GLHVKFSPKE

Cytochrome P450 Reductases

Cucurbita pepo subsp. *pepo* cytochrome P450 reductase enzyme
25 CppCPR4 or MogCPR1_0 (SEQ ID NO: 22)

MAQSESRS MKVSPLEMSAIIRKAMDPSQDSSSVREVATLILENREFVMILTTSIAVLIGCVV
VLVWKRSSDQAKSFEPPKQLIVKKPEPEVDGKKKVTVFVGTQTGTAE GF AKALAEFEAKARYE
KATFRVVDLDDYAADDDEYEEKLKKETLAIFFLATYGDGEPTDNAARFYKWFSEGEKKEGEWISN
LQYAVFGLGNRQYEHFNKIAKVVD EQLAEQGGKRLVPVGLGDDDQCIEDDFS AWREALWPELDK
30 LLREED EFTTVSTPYTA AVLEYRVV FYDAADVSGDKK WAFANGHAVYDIQHPCRANVA VRKEL
HTSASDRSCTHLEFDISGTGLTYETGDHVGVCENLDEVVEEAIRLIGLSPETYFSIHTDKEDG
TPLSGSSLPFPFAPCTLR TALTQYADLLSSPKKSALVALAAHASDPAEADRLRHLSLPAGKDEY
SQWIVASQRSLLLEVMAEFPSAR PPLGVFFAAVAPRLQPRYYSISSSPRMAPSRIHVTCALVYDK
TPTGRIHKGLCSTWMKNAIPLEESQACSWAPIYVRQSNFKLPTDSKLPIIMIGPGTGLAPFRGF
35 LQERLALKESGVELGHSILFFGCRNRKMDYIYEDELNNFVETGALSELIVAFSREGPSKEYVQH
KMVEKASDIWNLLSQGAYIYVCGDAKG MARDVHRTLHNIVQE QGSLDSSKAESMVKNLQMTGRY
LRD VW

MogCPR1_1 (SEQ ID NO: 23)

MANTLIPILVAICLFITGVAILNIQWKRSSDQAKSFEPPKQLIVKKPEPEVDDGKKKVTVFFG
TQTGTAEGFAKALAEAKARYEKATFRVVDLDDYAADDDEYEEKLKKETLAIFFLATYGDGEPT
DNAARFYKWFSEGEKEGEWISNLQYAVFGLGNRQYEHFNKIAKVVDQLAEQGGKRLVPVGLGD
DDQCIEDDFSAREALWPELDKLLREDEFTTVSTPYTAAVLEYRVVFYDAADVSGGDKKWAF
5 NGHAVYDIQHPCANAVRRELHTSASDRSCTHLEFDISGTGLTYETGDHVGVCENLDEVVEE
AIRLIGLSPETYFSIHTDKEDGTPLSGSSLPPPFAPCTLRALTQYADLLSSPKKSALVALAAH
ASDPAEADRLRHLSPAGKDEYSQWIVASQRSLEVMAEFSPARPLGVFFAAVAPRLQPRYYS
ISSSPRMAPSRIHVTCALVYDKTPTGRIHKGLCSTWMKNAIPLEESQACSWAPIYVRQSNFKLP
TDSKLPIIMIGPGTGLAPFRGFLQERLALKESGVELGHSILFFGCRNRKMDYIYEDELNNFVET
10 GALSELIVAFSREGPSKEYVQHKMVEKASDIWNLLSQGAYIYVCGDAKGMDVHRTLHNIVQE
QGSLDSSKAESMVKNLQMTGRYL RDVW

Phaseolus vulgaris cytochrome P450 reductase enzyme (PvCPR or MogCPR2_0) (SEQ ID NO: 24)

MAQSSSSSSSSSSSSMSPFDLMAAIIKGKKVLDPSNVSSDSSVSEVANIIFENREFVMILTSI
15 AVLIGCVVLIWRRSSGQKVKPVEPLKPLTVKEPEVEVDDGKQKVTIFFGTQTGTAEGFAKALA
DEAKARYEKAKFRVVDLDDYAADDDEYEEKLKKESLALFFLATYGDGEPTDNAARFYKWFTEGK
ERGEWLQNLKYGVFGLGNRQYEHFNKIAKVVDLTLIEQGAKRLVPVGLGDDDDQCIEDDFTAWRE
MLWPELDQLLRDEDDSTTVSTPYTAASEYRVVFYDPADAPLEDKSWGNANGHAVHDAQHPCRS
NVAVRRELHTPQSDRSCTHLEFDIAGTGLSYETGDHVGVCENLIETVEEALKLLGLSPDTYFS
20 IHSDEKEDGTPLGGSSLPPTFPCTLRALTQYADLLSSPKKSALLALAAHASDPTEADRLRYLA
SPAGKDEYAQWIVASQRSLEVMAEFSPAKPPLGVFFAAVVPRLQPRYYSISSSPRMAPSRIHV
TCALVYEKTPAGRIHKGLCSTWMKNCVPLEKSSDCSWAPIFVRQSNFKLPDPEVPVIMIGPGT
GLAPFRGFLQERFAMKEDGVELGPSILFFGCRNRQMDYIYEDELNNFVQSGALSELVVAFSREG
PTKEYVQHKMMEKASDIWNMISQGGYLYVCGDAKGMDVHRTLHTIVQE QGSLDSSKAESMVK
25 NLQMTGRYL RDVW

MogCPR2_1 (SEQ ID NO: 25)

MANTLIPILVAICLFITGVAILNIQWRRSSGQKVKPVEPLKPLTVKEPEVEVDDGKQKVTIFFG
TQTGTAEGFAKALADEAKARYEKAKFRVVDLDDYAADDDEYEEKLKKESLALFFLATYGDGEPT
DNAARFYKWFTEGKERGEWLQNLKYGVFGLGNRQYEHFNKIAKVVDLTLIEQGAKRLVPVGLGD
30 DDQCIEDDFTAWREMLWPELDQLLRDEDDSTTVSTPYTAASEYRVVFYDPADAPLEDKSWGNA
NGHAVHDAQHPCRSNVAVRRELHTPQSDRSCTHLEFDIAGTGLSYETGDHVGVCENLIETVEE
ALKLLGLSPDTYFSIHSDEKEDGTPLGGSSLPPTFPCTLRALTQYADLLSSPKKSALLALAAH
ASDPTEADRLRYLASPAGKDEYAQWIVASQRSLEVMAEFSPAKPPLGVFFAAVVPRLQPRYYS
ISSSPRMAPSRIHVTCALVYEKTPAGRIHKGLCSTWMKNCVPLEKSSDCSWAPIFVRQSNFKLP
35 TDPEVPVIMIGPGTGLAPFRGFLQERFAMKEDGVELGPSILFFGCRNRQMDYIYEDELNNFVQS
GALSELVVAFSREGPTKEYVQHKMMEKASDIWNMISQGGYLYVCGDAKGMDVHRTLHTIVQE
QGSLDSSKAESMVKNLQMTGRYL RDVW

MogCPR2_2 (SEQ ID NO: 26)

MAQSSSSSSSSSSSSMSPFDLMAAIIKGKKVLDPSNVSSDSSVSEVANIIFENREFVMILTSI
40 AVLIGCVVLIWRRSSGQKVKPVEPPKPLTVKEPEVEVDDGKQKVTIFFGTQTGTAEGFAKALA
DEAKARYEKAKFRVVDLDDYAADDDEYEEKLKKESLALFFLATYGDGEPTDNAARFYKWFTEGK
ERGEWLQNLKYGVFGLGNRQYEHFNKIAKVVDLTLIEQGAKRLVPVGLGDDDDQCIEDDFTAWRE

MLWPELDQLLRDEDDSTTVSTPYTAAISEYRVVFYDPADAPLEDKSWGNANGHAVHDAQHPCRS
 NVAVRKELHTPQSDRSCTHLEFDIAGTGLSYETGDHVGVCENLIETVEEALKLLGLSPDTYFS
 IHSDKEDGTPLGGSSLPPTFPCTLRALTALTKYADLLSSPKKSALLALAAHASDPTEADRLRYLA
 SPAGKDEYAQWIVASQRSLLLEVMAEFPSAKPPLGVFFAAVVPRLQPRYYSISSSPRMAPSRIHV
 5 TCALVYEKTPAGRIHKGLCSTWMKNCVPLEKSSDCSWAPIFVRQSNFKLPTDPEVPVIMIGPGT
 GLAPFRGFLQERFAMKEDGVELGPSILFFGCRNRQMDYIYEDELNNFVQSGALSELVVAFSREG
 PTKEYVQHKMMEKASDIWNMISQGGYLYVCGDAKGMARDVHRTLHTIVQEQQSLDSSKAESMVK
 NLQMTGRYL RDVW

10 Uridine diphosphate dependent glycosyltransferases (UGT)

AtUGT73C3 (SEQ ID NO: 50)

MATEKTHQFHPSLHFVLPFMAQGHMIPMIDIARLLAQRGVTITIVTTPHNAARFKNVNLRAIE
 SGLAINILHVKFPYQEFGLPEGKENIDSLDSTELMVPFFKAVNLLLEDPMVKLMEEMKPRPSCLI
 SDWCLPYTSIIAKNFNIPKIVFHGMGCFNLLCMHVLRRNLEILENVKSDEEYFLVPSFPDRVEF
 15 TKLQLPVKANASGDWKEIMDEMVKAEYTSYGIVVNTFQELEPPYVKDYKEAMDGKVWSIGPVSL
 CNKAGADKAERGSKAAIDQDECLQWLDSKEEGSVLYVCLGSICNLPLSQLKELGLGLEESRRSF
 IWVIRGSEKYKELFEWMLESGFEERIKERGLLIKGWAPQVLILSHPSVGGFLTHCGWNSTLEGI
 TSGIPLITWPLFGDQFCNQKLVVQVLKAGVSAGVEEVMKWGEEDKIGVLVDKEGVKKAVEELMG
 DSDDAKERRRRVKELGELAHKAVEKGGSSHSNITLLLQDIMQLAQFKN

20 SgUGT720_269_1_E (SEQ ID NO: 51)

MAVQPRVLLFPFPALGHVKPFLSLAELLSDAGIDVVFLSTEYNHRRISNTEALASRFPTLHFET
 IPDGLPPNESRALADGPLYFSMREGTKPRFRQLIQSLNDGRWPITCIITDIMLSSPIEVAEEFG
 IPVIAFCPCSARYLSIHFFIPKLVEEGQIPYADDDPIGEIQGVPLFEGLLRNHLPGSWSDKSA
 DISFSHGLINQTLAAGRASALILNTFDELEAPFLTHLSSIFNKIYTIGPLHALSKSRLGDSSSS
 25 ASALSGFWKEDRACMSWLDQPPRSVVFVSFGSTMKMADELREFWYGLVSSGKPFCLVLRSDV
 VSGGEAAELIEQMAEEEGAGGKLG MVVEWAAQEKVLSHPAVGGFLTHCGWNSTVESIAAGVPM
 CWPILGDQPSNATWIDRVWKIGVERNREWDRLTVEKMVRALMEGQKRVEIQRSMEKLSKLANE
 KVVRRGLSFDNLEVLVEDIKKLKPYKF

CmOUGT (SEQ ID NO: 52)

MAELSHTHHVLLFPFPAKGHIKPFSLAQLLCNAGLRVTFLNTDHHHRIHDLNRLAAQLPTLH
 FDSVSDGLPPDEPRNVFDGKLYESIRQVTSSLFRELLVSYNNGTSSGRPPITCVITDVMFRFPI
 DIAEELGIPVFTFSTFSARFLFLIFWIPKLLLEDGQLRYPEQELHGVPGAEGIRWKDLPGFWSV
 EDVADWDPMNFVNQTLATSRSSGLILNTFDELEAPFLTSLSKIYKKIYSLGPIINSLLKNFQSQP
 QYNLWKEDHSCMAWLDSQPRKSVVFVSFGSVVKLTSLRQLMEFWNGLVNSGMPFLLVLRSDVIEA
 35 GEEVVREIMERKAEGRWVIVSWAPQEEVLAHDAVGGFLTHSGWNSTLES LAAGVPMISWPQIGD
 QTSNSTWISKVVRIGLQLEDGFDSSSTIETMVRSIMDQTMKTVAE LAERAKNRASKNGTSYRNF
 QTLIQDITNIIETHI

SgUGT720_269_4_Manus (SEQ ID NO: 53)

MAEQAHDLLHVLLFPFPAEGHIKPFLCLAELLNAGFHVTFNLTDYNHRRRLHNLHLLAARFPSL
 HFESISDGLPPDQPRDILDPKFFISICQVTKPLFRELLLSYKRISVQVQTRPPITCVITDVI FR
 FPIDVAEELDIPVFSFCTFSARFMFLYFWIPKLIEDGQLPYPNGNINQKLYGVAPEAEGLLRCK
 DLPGHWAFADELKDDQLNFVDQTTASSRSSGLILNTFDDLEAPFLGRLSTIFKKIYAVGPIHSL
 5 LNSHHCGWLKEDHSCLAWLDSRAAKSVVFSFGSLVKITSRQLMEFWHGLLNSGKSFLFVLRSD
 VVEGDDEKQVVKEIYETKAEGKWL VVGWAPQEKVLAHEAVGGFLTHSGWNSILESIAAGVPMIS
 CPKIGDQSSNCTWISKVWKIGLEMEDRYDRVSVETMVRSIMQEKEKMQKTIAELAKQAKYKVS
 KDGTSYQNLECLIQDIKKLNQIEGFINNPNFSDLLRV

SgUGT720_269_4_E (SEQ ID NO: 54)

10 MAEQAHDLLHVLLFPYPAGHIKPFLCLAELLNAGLNVTFNLTDYNHRRRLHNLHLLAACFPSL
 HFESISDGLQPDQPRDILDPKFYISICQVTKPLFRELLLSYKRTSSVQVQTRPPITCVITDVI FR
 FPIDVAEELDIPVFSFCTFSARFMFLYFWIPKLIEDGQLPYPNGNINQKLYGVAPEAEGLLRCK
 DLPGHWAFADELKDDQLNFVDQTTASLRSSGLILNTFDDLEAPFLGRLSTIFKKIYAVGPIHAL
 LNSHHCGWLKEDHSCLAWLDSRAARSVVFSFGSLVKITSRQLMEFWHGLLNSGTSFLFVLRSD
 15 VVEGDGEKQVVKEIYETKAEGKWL VVGWAPQEKVLAHEAVGGFLTHSGWNSILESIAAGVPMIS
 CPKIGDQSSNCTWISKVWKIGLEMEDQYDRATVEAMVRSIMKHEGEKIQKTIAELAKRAKYKVS
 KDGTSYRNLEILIEDIKKIKPN

SgUGT74_345_2_I (SEQ ID NO: 55)

MADETTVNGGRRASDVVVFAPPRHGHMS PMLQFSKRLVSKGLRVTFLLITTSATESLRNLPPSS
 20 SLDLQVISDVPESENDIATLEGYLRSEKATVSKTLADFIDGIGNPPKFIVYDSVMPWVQEVARGR
 GLDAAPFFTQSSAVNHILNHVYGGSL SIPAPENTAVSLPSMPVLQAE DLPAFPDDPEVVMNFMF
 SQFSNFQDAKWI FFNTFDQLECKVVNW MADRWPIKTVGPTIP SAYLDDGRLEDDRAFGNL LKPK
 EDGKNTRQWQWLDSKDTASVLYISFGSLAILQEEQVKELAYFLKDTNLSFLWVLRDSELQKLP
 NFVQETSHRQLVWNWCSQLQVLSHRAVSCFVTHCGWNSTLEALSLGVPMVAIPQWVDQTTNAKF
 25 VADVWRVGRVKKKDERIVTKEELEASIRQVVQGEGRNEFKHNAIKWKKLAKEAVDEGGSSDKN
 IEEFVKTIA

XP_023543158.1 (SEQ ID NO: 56)

MAELSHTHHVLLFPFPAKGHIKPPFSLAQLLCNAGLRVTFNLTDHHRRIHDLRLAAQLPTLH
 FDSVSDGLPPDEPRNVFDGKLYESIRQVTSSLFRDLLVSYNNGTSSGRPPITCVITDCMFRFP
 30 DIAEELGIPVFTFSTFSARFLFLFFWIPKLIEDGQLRYPEQELHGVPGAEG LIRCKDLPGFLSD
 EDVAHWKPMNFVNQILATSRSSGLILNTFDELEAPFLTSLSKIYKKIYSLGPINSLLKNFQSQP
 QYNLWKEDHSCMAWLDSQPRKSVVFSFGSVVKLT SRQLMEFWNGLVNSGKPFLLVLRSDVTEA
 GEEVVREIMERKAEGRWVIVNWAPQEEVLAHDAVGGFLTHSGWNSTLESIAAGVPMISWPQIGD
 QTSNSTWVSKVWRIGLQLEDGFDSS TIETMVRSIMDQKMEKTVAELAERAKNRASKNGTSYRNF
 35 QTLIQDITNIIETHI

XP_022151514.1 (SEQ ID NO: 57)

MANSDDHQHHVLLFPFPAKGHIKPFLCLAQLLCGAGLQVTFNLTDHNRRIIDRRHRRLLATQFP
 MLHFKSISDGLPPDHPRLDLDGKLIASMRRVTESLFRQLLLSYNGYGNGTNNVSNSGRPPISC
 VITDVI FSPVEVAEELGIPVFSFATFSARFLFLYFWIPKLIQEGQLPFPDGKTNQELYGVPGA
 40 EGIIRCKDLPGSWSVEAVAKNDPMNFVKQTLASSRSSGLILNTFEDLEAPFVTHLSNTFDKIYT

IGPIHSLLGTS HCGLWKEDYACLAWL DARPRKSVV FVSFGSLVKTT SRELMELWHGLVSSGKSF
LLVLRSDVVEGEDEEQVVK EILESNGEGKWL VVGWAPQEEVLAHEAIGGFLTHSGWNSTMESIA
AGVPMVCWPKIGDQPSNCTWVSRVWKVGLEMEERYDRSTVARMARSMMEQEGKEMERRIAELAK
RVKYRVGKDGESYRNLESLIRDIKITKSSN

5 XP_023764969.1 (SEQ ID NO: 58)

MAGDAVPIIGQKQPHVVFVFPFPAQSHIKCMLKLARILHHNGLHVTFINTHSNQKRLVKSNGLLG
LDKAPGFQFKTVPDGLSSATDDGVEHTQTMAELWTYLGANFLGSFLDVVSGLEIPVTCIICDGF
MTYTNTIHAAEKLNIPIILFWTMAASGFMGFYQVKVLAEKGILPLKDEIYLTNGYLDMEIDWIP
GMEGIRLKELEPEIKLFTKHDNPAFKFLL ETAQLAHKVSHMI IHTFEELEASLIKELKSIFPNVY
10 SVGPLELLLNQITEKETNKS LCNYS LWKEEPECVQWLQSKEPNSVVYVNFSGIAVMSLQDLLE
FGWGLVNSKHEFLWII RTDLVDGKPVVLPQELEDAMKGKGFVASWCSQEEVLNHSSVGGFLTHG
GWGSIIESLSAGVPMICWPVSGDQQTNCRQMCKEWGVGMEISRNVKRDEVEKLVKELMEGMEGK
RMRKKALEWKKVAEKATGSNGSSWIDAEKLANQIVKLSTKFPTV

XP_043637218.1 (SEQ ID NO: 59)

MANTMTRVDEKQPHVVFIPFPAQSHIKCMLKLAE LLHHKGIHITFINTRSNHKKLVESGGTHDF
KDIPSFQFRTVSDCPSGAEDDGVEPVTPTLVEVWMYLADNFFDSFLDIVSGLESTPATCIICDG
FMTYTNTMITAAEKLKIPIILFWTMAACGFMGFYQAKVLTDKGLLPLQDES YLTNGYLEMEIDWI
PGMEGCRLKDLPE DMLVTKVDDPGYRYLLETAQA AHNVSYIIMHTFEELEARLVNEIKSIFPNV
YTVGPLQLLLN QIKEKENKKTAFNCYSLWNEEPECIQWLESKEPSSVVYVNFGLAVMSLEDLI
20 EFGWGLVNSDYEF LWIIRADLVKGNPVVLPKELEEEIGKKGFLAKWCSQEEVLNHPSVGGFLTH
GGWGSTIESLSAGLPMICWPSIGDQRANCRQMCKQWQVGMEIGKNVKRDEVEKLVKLLMDGLEG
KRMKEKARHWKKMAEKATSSNGSSSIDVEKLANQIINF

XP_008445481.1 (CmUGTc3_0) (SEQ ID NO: 60)

MAEMTAANGGERIKQSHVIVFPFPRHGHMSPMLQFSKRLISKGLLLTFLITSSASQSLTINIP
25 PPSFHFKIIISDLPESDDVATL DAYLRSFRAAVTKSLSNFIDEVLTSSSNEEVPTLIVYDSVM
PWVQSVAAERGLDSAPFFTESAAVNHLHLVYGGSL SIPPPDNVVVSLPSEIVLQPEDLPSFPD
DPEVVLD FMTSQFSHLENVKWIFINTFDRLESKVWNMAKTLPIKTVGPTIP SAYLDGRLEKDK
AYGLNVSKSNNGKCPIKWLD SKETASVIYISFGSLVILSEEQVKELTNLLRDTDFSFLWVLRES
EMVKLPKNFVQDTS DRGLIVNWCCQLQVL SHKAVSCFVTHCGWNSTLEALS LGVPMVAIPQWID
30 QTTNAKFVADVVRVGVRVKKNEKSVAIKEE LEASIRKIVVQNGTNEFKQNAIKWKNLAKEAVD
ERGSSDKNIEEFVQALV

XP_022132196.1 (SEQ ID NO: 72)

MADNVVSLPSEIVLQPEDLPSFPDDPEVVLD FMTSQFSHLENVKWIFINTFDRLESKVWNMA
KTLPIKTVGPTIP SAYLDGRLEKDKAYGLNVSKSNNGKCPIKWLD SKETASVIYISFGSLVILS
35 EEQVKELTNLLRDTDFSFLWVLRESEMVKLPKNFVQDTS DRGLIVNWCCQLQVL SHKAVSCFVT
HCGWNSTLEALS LGVPMVAIPQWIDQTTNAKFVADVVRVGVRVKKNEKSVAIKEE LEASIRKIV
VQNGTNEFKQNAIKWKNLAKEAVDERGSSDKNIEEFVQALVAANGGGERGKQSHVIVFPFPRH
GHMSPMLQFSKRLISKGLLLTFLITSSASQSLTINIP PPSFHFKIIISDLPESDDVATL DAYLR
SFRAAVTKSLSNFIDEVLTSSSNEEVPTLIVYDSVM PWVQSVAAERGLDSAPFFTESAAVNHL
40 LHLVYGGSL SIPPP

XP_022132196.1 (SEQ ID NO: 61)

MAEKATANGRRSSHVLLFAYPMHGHMSFPMQLQFAKRLASKGLLVTFLLTSSVTESLQIDLPPSY
PIHLRFISDFHTEVIETLKQRHEAFAAAVSRSLGEFLDGALINGDHPPRLMVFDSPMPWAMEVA
RSRGLEAAPFFTESAAVNHILNQVYEGSLSIPAPENAAVSIPSLPNLEAEDLPYFSPVIREVTL
5 EFMTQRQFSSFKDAKWIFINTFDQLEPQIVNWMGERWPIKTVGPTVPSAYLDGRLEKDKTYGLKR
QKPEDGRAVEWLDSKETASVVYISFGSLVMLAEKQVKELTNFLTESGLPFLWVLRSEMEKLPE
NFIQETSGKGLVVNWCSQLEVLSHKAVGCFVTHGGWNSTLEALSSGVPMVAVPQWIDQTTNAKF
IADVWEIGVRVKLNEKHEIATKDELEASIRQVIEGREKKNSIKWRKLAKEAVDEGGSSDKNIED
FAKTIM

10 XP_023538293.1 (SEQ ID NO: 62)

MASNTTLNGRRSSHVLLFAYPKQGHLSFPMQLQFAKRLASKGLRITFLTTTSTTKSLEIDLPA
QIDLRFISDVRTEPILSLKDEHESFEAVVSKSFGDFIDGTLRSSGFDPPRFVIFDSVMPWAMDV
ARVRGIDSAPFFTESCVVNHILNQVYEGSFSIPPVENVAAGISIPPLPVLQTEDLPYFSYEPEL
VLKFMTDQFSSFKNAKWIFVNTFDQLEMKVWNWMTQKWPIKTIGPNIPSAYLDGRLKDDKTYGL
15 NHQNLNNCKIFQWLDSKEIASVIYLSFGSLVILPEEQVNELARFFKDTNFSFLWVLRSEMEKLE
PNNFVQQTSHKGLVVKWCCQLQVLVSHKAVSCFVTHCGWNSTIEALSLGVPMVAVPQWIDQTTNA
KFVADVWKVGARVKMNDKGIATKLELEASIRHVSQGYRQNEIKQNSIKLRNLAKEAMDEGGSSD
KNIEQFVKELD

UGT74AC1 (SEQ ID NO: 63)

20 MAEKGDTLHILVFPFSPQGHINPLLQLSKRLIAKGIVSLVTTLHVS NHLQLQGAYSNSVKIEVI
SDGSEDRLETDTMRQTLDRFRQKMTKNLEDFLQKAMVSSNPPKFIIYDSTMPWVLEVAKEFGD
RAPFYTQSCALNSINYHVLHGQLKLPPETPTISLPSMPLLRPSDLPAYDFDPASTDTIIDLLTS
QYSNIQDANLLFCNTFDKLEGEIIQWMETLGRPVKTVGPTVPSAYLDKRVENDKHYGLSLFKPN
EDVCLKWLDSKPSGSVLVSYGSLVEMGEEQLKELALGIKETGKFFLWVVRDTEAEKLPNFVE
25 SVAEKGLVSWCSQLEVLAHPSVGCFFTHCGWNSTLEALCLGVPVVAFPQWADQVTNAKFLEDV
WKVGKRVKRNEQRLASKEEVRSCIWEVMEGERASEFKSNSMEWKKWAKEAVDEGGSSDKNIEEF
VAMLKQT

UGT74AC1_M7 (SEQ ID NO: 64)

30 MAEKGDTLHILVFPFPAQGHINPLLQLSKHLIAKGIVSLVTTLHVS NRMQLQGAYSNSVKIEVI
SDGSEDRLETDTLRQYLDRFRQKMTKNLEDFLQKAMVSSNPPKFIIYDSTMPWVLEVAKEFGD
RAPFYTQSCALNSINYHVLHGQLKLPPETPTISLPSMPLLRPSDLPAYDFDPASTDTIIDLLTS
QYSNIQDANLLFCNTFDKLEGEIIQWMETLGRPVKTVGPTVPSAYLDKRVENDKHYGLSLFKPN
EDVCLKWLDSKPSGSVLVSYGSLVEMGEEQLKELALGIKETGKFFLWVVRDTEAEKLPNFVE
SVAEKGLVSWCSQLEVLAHPSVGCFFTHCGWNSTLEALCLGVPVVAFPQWADQVTNAKFLEDV
35 WKVGKRVKRNEQRLASKEEVRSCIWEVMEGERASEFKSNSMEWKKWAKEAVDEGGSSDKNIEEF
VAMLKQT

Stevia rebaudiana UGT85C1 MogUGTc3_0 (SEQ ID NO: 27)

MADQMAKIDEKKPHVVFIIPFPAQSHIKCMLKLARILHQKGLYITFINTDTNHERLVASGGTQWL
ENAPGFWFKTVPDGFGSAKDDGVKPTDALREIMDYLKTNFFDLFLDLVLKLEVPATCIICDGCM

TFANTIRAAEKLNI PVILFWTMAACGFMAFYQAKVLKEKEIVPVKDETYLTNGYLDMEIDWIPG
MKRIRLRDLPEFILATKQNYFAFEFLFETAQLADKVSHMI IHTFEELEASLVSEIKSIFPNVYT
IGPLQLLL NKITQKETNND SYSLWKEEPECVEWLNSKEPNSVVYVNFGLAVMSLQDLVEFGWG
LVNSNH YFLWII IRANLIDGKPAVMPQELKEAMNEKGFVGSWCSQEEVLNHPAVGGFLTHCGWGS
5 IIESLSAGVPMLGWPSIGDQRANCRQMCKEWEVGM EIGKNVKRDEVEKLVRLMEGLEGERMRK
KALEWKKSATLATCCNGSSSLDVEKLANEIKKLSRN

MogUGTC3_1 (SEQ ID NO: 28)

MADQMAKIDEKKPHVVFIPFPAQSHIKCMLKLARILHQKGFYITFINTETNHERLVASGGTQWL
ENAPGFWFKTVPDGFSGAKDDGVKPTDALREIMDY LKTNFFDLFLDLVLKLEV PATCIICDGF M
10 TFANTIRAAEKLNI PVILFWTMAACGFMAFYQAKVLKEKEIVPVKDETYLTNGYLDMEIDWIPG
MKRIRLRDLPEFILATKQNYFAFEFLFETAQLADKVSHMI IHTFEELEASLVSEIKSIFPNVYT
IGPLQLLL NKITQKETNND SYSLWKEEPECVEWLNSKEPNSVVYVNFGLAVMSLQDLVEFGWG
LVNSNH YFLWII IRANLIDGKPAVMPQELKEAMNEKGFVGSWCSQEEVLNHPAVGGFLTHCGWGS
IIESLSAGVPMLGWPSIGDQRANCRQMCKEWEVGM EIGKNVKRDEVEKLVRLMEGLEGERMRK
15 KALEWKKSATLATCCNGSSSLDVEKLANEIKKLSRN

MogUGTc3_2 (SEQ ID NO: 29)

MADQMAKIDEKKPHVVFIPFPAQSHIKCMLKLARILHQKGFYITFINTETNHERLVASGGTQWL
ENAPGFWFKTVPDGFSGAKDDGVKPTDALREIMDY LKTNFFDLFLDLVLKLEV PATCIICDGF M
20 TFANTIRAAEKLNI PVILFWTMAACGFMAFYQAKVLKEKEIVPVKDETYLTNGYLDMEIDWIPG
MKRIRLRDLPEFILATKQNYFAFEFLFETAQLADKVSHMI IHTFEELEASLVSEIKSIFPNVYT
IGPLQLLL NKITQKETNND SYSLWKEEPECVEWLNSKEPNSVVYVNFGLIVMSLQDLVEFGWG
LVNSNH YFLWII IRANLIDGKPAVMPQELKEAMNEKGFVGSWCSQEEVLNHPAVGGFLTHCGWGS
IIESLSAGVPMLGWPSIGDQRANCRQMCKEWEVGM EIGKNVKRDEVEKLVRLMEGLEGERMRK
KALEWKKSATLATCCNGSSSLDVEKLANEIKKLSRN

25 MogUGTc3_3 (SEQ ID NO: 30)

MADQMAKIDEKKPHVVFIPFPAQSHIKCMLKLARILHQKGFYITFINTETNHERLVASGGTQWL
ENAPGFWFKMVPDGFSGAKDDGVKPTDALREIMDY LKTNFFDLFLDLVLKLEV PATCIICDGF M
TFANTIRAAEKLNI PVILFWTMAACGFMAFYQAKVLKEKEIVPVKDETYLTNGYLDMEIDWIPG
MKRIRLRDLPEFILATKQNYFAFEFLFETAQLADKVSHMI IHTFEELEASLVSEIKSIFPNVYT
30 IGPLQLLL NKITQKETNND SYSLWKEEPECVEWLNSKEPNSVVYVNYGSLIVMSLQDLVEFGWG
LVNSNH YFLWII IRANLIDGKPAVMPQELKEAMNEKGFVGSWCSQEEVLNHPAVGGFLTHCGWGS
IIESLSAGVPMLGWPSIGDQRANCRQMCKEWEVGM EIGKNVKRDEVEKLVRLMEGLEGERMRK
KALEWKKSATLATCCNGSSSLDVEKLANEIKKLSRN

MogUGTC24_0 (SEQ ID NO: 31)

35 MADVAEEQPLKIYFIPYLAAGHMIPLCDIATLFASRGHHVTIITTPSNAQTLRESHHFRVQTIQ
FPSQEVGLPAGVQNLTAVTNLDDSYKIIYHATMLLRKHIEDFVERDPPDCIVADFLFPWDDVAT
KLHIPRLVFNGFTLFTICAMESHKAHPLPVDAASGSFVIPDFPHHVTINSTPPKRTKEFVDPLL
TEAFKSHGFLINSFVELDGEECVEHYERITGGHKAHWLGP AFLVHRTAQDRGEKSVVSTQECLS
WLD SKRDNSVLYICFGTICYFPDKQLYEIASAIEASGHEFIWVVPEKRGNADESEEEKEKWLPK
40 GFEERNNGKKGMIIRGWAPQVAILGHPAVGGFLTHCGWNSTVEAVSAGVPMITWPVHSDQYFNE

KLITQVRGIGVEVGAEWIVTAFRETEKLVGRDRIERAVRRVMDGGDEAVQIRRRARELGEMAR
QAVQEGGSSHTNLTLALINDLKRWRDSKQLN

MogUGTC24_1 (SEQ ID NO: 32)

5 MADVAEEQPLKIYFIPYLAAGHMIPLCDIATLFASRGHHVTIITTPSNAQTLRESHHFRVQTIQ
FPSQEVGLPEGVQNLTAVTNLDDSYKFYHATMLLRKPIEDFVERDPPDCIVADFLFPWVDDVAT
KLHIPRLVFNGFTLFTICAMESHKAHPLPVDAASGSFVIPDFPHHVTINSTPPKRTKEFVDPLL
TEAFKSHGFLINSFVELDGEECVEHYERITGGHKAWHLGPAFLVHRTAQDRGEKSVVSTQECLS
WLSKRDNNSVLYICFGTICYFPDKQLYEIASAIEASGHEFIWVVPEKRGNADESEEEKEKWLPK
10 GFEERNNGKKGMIIRGWAPQVAILGHPAVGGFLTHCGWNSTVEAVSAGVPMITWVPVHSDQYFNE
KLITQVRGIGVEVGAEWIVTAFRETEKLVGRDRIERAVRRVMDGGDEAVQIRRRARELGEMAR
QAVQEGGSSHTNLTLALINDLKRWRDSKQLN

MogUGTC24_2 (SEQ ID NO: 33)

15 MADVAEEQPLKIYFIPYMAAGHMIPLCDIATLFASRGHHVTIITTPSNAQTLRESHHFRVQTIQ
FPSQEVGLPEGVQNLTAVTNLDDSYKFYHAAMLLRKPIEDFVERDPPDCIVADFLFPWVDDVAT
KLHIPRLVFNGFTLFTICAMESHKAHPLPVDAASGSFVIPDFPHHVTINSTPPKRFKEFVDELL
TEAFKSHGFLINSFVELDGEECVEHYERITGGHKAWHLGPAFLVHRTAQDRGEKSVVSTQECLS
WLSKRDNNSVLYICFGTICYFPDKQLYEIASAIEASGHEFIWVVPEKRGNADESEEEKEKWLPK
GFEERNNGKKGMIIRGWAPQVAILGHPAVGGFLTHCGWNSTVEAVSAGVPMITWVPVHSDQFFNE
20 KLITQVRGIGVEVGAEWIVTAFRETEKLVGRDRIERAVRRVMDGGDEAVQIRRRARELGEMAR
QAVQEGGSSHTNLTLALINDLKRWRDSKQLN

MogUGTC24_3 (SEQ ID NO: 34)

25 MADVAEEQPLKIYFIPYMAAGHMIPLCDIATLFASRGHHVTIITTPSNAQTLRESHHFRVQTIQ
FPSQEVGLPEGVQNLTAVTNLDDAYKFYHAAMLLRKPIEDFVERDPPDCIVADFLFPWVDDVAT
KLHIPRLVFNGFTLFTICAMESHKAHPLPVDAASGSFVIPDFPHHVTINSTPPKRFKEFVDELL
TEAFKSHGFLINSFVELDGEECVEHYERITGGHKAWHLGPAFLVHRTAQDRGEKSVVSTQECLS
WLSKRDNNSVLYICFGSICYFPDKQLYEIASAIEASGHEFIWVVPEKRGKADESEEEKEKWLPK
GFEERNNGKKGLIIRGWAPQVAILGHPAVGGFLTHCGWNSTVEAVSAGVPMITWVPVHSDQFFNE
30 KLITQVRGIGVEVGAEWIVTAFRETEKLVGRDRIERAVRRVMDGGDEAVQIRRRARELGEMAR
QAVQEGGSSHTNLTLALINDLKRWRDSKQLN

30 MogUGTC24_4 (SEQ ID NO: 48)

35 MADVAEEQPLKIYFIPYMAAGHMIPLCDIATLFASRGHHVTIITTPSNAQTLRESHHFRVQTIQ
FPSQEVGLPEGVQNLTAVTNLDDAYKFYHAAMLLRKPIEDFVERDPPDCIVADFLFPWVDDVAT
KLHIPRLVFNGFTLFTICAMESHKAHPLPVDAASGSFVIPDFPHHVTINSTPPKRFKEYVDELL
TEAFKSHGFLINSFVELDGEECVEHYERITGGHKAWHLGPAFLVHRTAQDRGEKSVVSTQECLS
WLSKRDNNSVLYICFGSICYFPDKQLYEIASAIEASGHEFIWVVPEKRGKADESEEEKEKWLPK
GFEERNNGKKGLIIRGWAPQVAILGHPAVGGFLTHCGWNSTVEAVSAGVPMITWVPVHSDQFFNE
KLITQVRGIGVEVGAEWIVTAFRETEKLVGRDRIERAVRRVMDGGDEAVQIRRRARELGEMAR
QAVQEGGSSHTNLTLALINDLKRWRDSKQLN

MogUGTC24_5 (SEQ ID NO: 49)

MADVAAEQPLKIYFIPYMAAGHMIPLCDIATLFASRGHHVTIITTPSNAQTLRESHHFRVQTVQ
 FPSQEVGLPEGVQNLTAVTNLDDAYKFYHAAMLLRKPIEDFVERDPPDCIVADFLFPWVDDVAT
 KLHIPRLVFNGFTLFTICAMESHKAHPLPVDAASGSFVIPDFPHHVTINSTPPKRFKEYVDELL
 TEAFKSHGFLINSFVELDGEECVEHYERITGGHKAWHLGPAFLVHRTAQDRGEKSUVSTQECLS
 5 WLDKSRDNSVLYICFGSICYFPDKQLYEIASAIEASGHEFIWVVPKRGKADESEEEKEKWLPK
 GFEERNNGKKGLIIRGWAPQVAILGHPAVGGFLTHCGWNSTVEAVSAGVPMITWVPVHSDQFFNE
 KLITQVRGIGVEVGAEWIVTAFRETEKLVGRDRIERAVRRVMDGGDEAVQIRRRARELGEMAR
 QAVQEGGSSHTNLTLALINDLKRWRDSKQLN

Siraitia grosvenorii UGT94-289, SEQ ID NO: 35)

10 MADAQRGHTTTILMFPWLGYGHLSAFLELAKSLSRNRFHIYFCSTSVNLDAIKPKLPSSSSSDS
 IQLVCLPSSPDQLPPLHTTNALPPLMPTLHQAFSMAAQHFAAILHTLAPHLIIYDSFQPW
 APQLASSLNIPAINFNTTGASVLTRMLHATHYPSSKFPISFVLHDYWKAMYSAGGAVTKKDH
 KIGETLANCLHASCSVILINSFRELEEKYMDYLSVLLNKKVVPVGPLVYEPNQDGEDEGYSSIK
 NWLDKKEPSSTVFVSFGSEYFPSKEEMEEIAHGLEASEVHFIWVVRFPQGDNTSAIEDALPKGF
 15 LERVGERGMVVKGWAPQAKILKHWSGGFVSHCGWNSVMESMMFGVPIIGVPMHLDQFPNAGLA
 EEAGVGVEAKRDPDGKIQRDEVAKLIKEVVVEKTREDVRKKAREMSEILRSKGEEKMDEMVAAI
 SLFLKI

MogUGT1216_0, SEQ ID NO: 36)

MAVTKKDHKIGETLANCLHASCSVILINSFRELEEKYMDYLSVLLNKKVVPVGPLVYEPNQDGE
 DEGYSSIKNWLDKKEPSSTVFVSFGSEYFPSKEEMEEIAHGLEASEVHFIWVVRFPQGDNTSAI
 EDALPKGFLEERVGERGMVVKGWAPQAKILKHWSGGFVSHCGWNSVMESMMFGVPIIGVPMHLD
 QFPNAGLAEEAGVGVEAKRDPDGKIQRDEVAKLIKEVVVEKTREDVRKKAREMSEILRSKGEEK
 MDEMVAAISLFLKIGGSDAQRGHTTTILMFPWLGYGHLSAFLELAKSLSRNRFHIYFCSTSVN
 LDAIKPKLPSSSSSDSIQLVELCLPSSPDQLPPLHTTNALPPLMPTLHQAFSMAAQHFAAIL
 25 HTLAPHLIIYDSFQPWAPQLASSLNIPAINFNTTGASVLTRMLHATHYPSSKFPISFVLHDYW
 KAMYSAGGA

MogUGT1216_1 (SEQ ID NO: 37)

MAVTKKDHKIGETVLANCLHASCSVILINSFRELEEKYMDYLSVLLNKKVVPVGPLVYEPNQDGE
 DEGYSSIKNWLDKKEPSSTVFVSFGSEYFPSKEEMEEIAHGLEASEVHFIWVVRFPQGDNTSAI
 30 EDALPKGFLEERVGERGMVVKGWAPQAKILKHWSGGFVSHCGWNSVMESMMFGVPIIGVPMHLD
 QFPNAGLAEEAGVGVEAKRDPDGKIQRDEVAKLIKEVVVEKTREDVRKKAREMSEILRSKGEEK
 MDEMVAAISQFLKIGGSDAQRGHTTTILMFPWLGYGHLSAFLELAKSLSRNRFHIYFCSTSVN
 LDAIKPKLPSSSSSDSIQLVELCLPSSPDQLPPLHTTNALPPLMPTLHQAFSMAAQHFAAIL
 HTLAPHLIIYDSFQPWAPQLASSLNIPAINFNTTGASVLTRMLHATHYPSSKFPISFVLHDYW
 35 KAMYSAGGA

MogUGT1216_2 (SEQ ID NO: 38)

MAVTKKDHKIGETVLANCLHASCSVILINSFRELEEKYMDYLSVLLNKKVVPVGPLVYEPNQDGE
 DEGYSSIKNWLDKKEPSSTVFVSFGSEYFPSKEEMEEIAHGLEASEVHFIWVVRFPQGDNTSAI
 EDALPKGFLEERVGERGMVVKGWAPQAKILKHPSTGGFVSHCGWNSVMESMMFGVPIIGVPMHLD
 40 QFPNAGLAEEAGVGVEAKRDPDGKIQRDEVAKLIKEVVVEKTREDVRKKAREMSEILRSKGEEK

MDEMVA AISQFLKIGGGSDAQRGHTTTILMFPWLGYGHLSAFLELAKSLSRNFHIYFCSTSVN
LDAIKPKLPSSSSSDSIQLVELCLPSSPDQLPPLHTTNALPPLMPTLHQAFSMAAQHFAAIL
HTLAPDLLIYDSFQWPAPQLASSLNIPAINFNTTGASVLTRMLHATHYPSSKFP ISEFVLHDYW
KAMYS AAGGA

5 MogUGT1216_3 (SEQ ID NO: 90)

MAVTKKDHKIGETVANCLEASCSVILINSFRELEEKYMDYLSVLLNKKVVPVGPLVYEPNQDGE
DEGYSSIKNWLDKKEPSSTVFVSFGSEYFPSKEEMEEIAHGLEASEVHFIWVVRFPQGDNTSAI
EDALPKGFLERVGERGMVVKGWAPQAKILKHPSTGGFVSHCGWNSVMESMMFGVPIIGVPMHLD
QPFNAGLAEAGVGVEAKRDPDGKIQRDEVAKLIKEVVVEKTREDVRKKAREMSEILRSKGEEK
10 MDEMVA AISQFLKIGGGSDAQRGHSTTILMFPWLGYGHLSAFLELAKSLSRNFHIYFCSTKVN
LDAIKPKLPSSSSSDSIQLVELCLPSSPDQLPPLHTTNALPPLMPTLHQAFSMAAQHFAAIL
HTLAPDLLIYDSFQWPAPQLASSLNIPAINFNTTGASVLTRMLHATHYPSSKFP ISEFVLHDYW
KAMYS AAGGA

Coffea arabica uridine diphosphate dependent glycosyltransferase
1,6 MogUGT16_0 (SEQ ID NO: 39)

MAENHATFNVLMPLPWL AHGHVSPYLELAKKLTARNFNVYLCSSPATLSSVRSKLTEKFSQSIHL
VELHLPKLPELPAEYHTTNGLPPLMPTLKDADFMAKPNFCNVLKS LKPDLLIYDLLQPWAP EA
ASAFNIPAVFIFSSATMTS FGLHFFKNPGTKYPYGN AIFYRDYESVFVENLTRDRD TYRVIN
CMERSSKIILIKGFNEIEGKYFDYFSCLTGKKVVPVGPLVQDPVLDD EDCRIMQWL NKKEKGST
20 VVVSFGSEYFLSKKDME EIAHGLEVSNVDFIWVVRFPKGENIVIEETLPKGFFERVGERGLV N
GWAPQAKILTHPNVGGFVSHCGWNSVMESMKFGLPIIAMPMLDQPINARLIEEVGAGVEVLRD
SKGKLHRERMAETINKVMKEASGESVRKKARELQEKLELKGDEEIDDVVKELVQLCATKNKRNG
LHYY

MogUGT16_1 (SEQ ID NO: 40)

25 MAENHATFNVLMPLPWL AHGHVSPYLELAKKLTARNFNVYLCSSPATLSSVRSKLTEKFSQSIHL
VELHLPKLPELPAEYHTTNGLPPLMPTLKDADFMAKPNFCNVLKS LKPDLLIYDLLQPWAP EA
ASAFNIPAVFIFSSATMTS FGLHFFKNPGTKYPYGN AIFYRDYESVFVENLTRDRD TYRVIN
CMERSSKIILIKGFKEIEGKYFDYFSCLTGKKVVPVGPLVQDPVLDD EDCRIMQWL NKKEKGST
VVVSFGSEYFLSKKDME EIAHGLEVSNVDFIWVVRFPKGENIVIEETLPKGFFERVGERGLV N
30 GWAPQAKILTHPNVGGFVSHCGWNSVMESMKFGLPIIAMPMLDQPINARLIEEVGAGVEVLRD
SKGKLHRERMAETINKVMKEASGESVRKKARELQEKLELKGDEEIDDVVKELVQLCATKNKRNG
LHYY

CmUGTc3_1 (SEQ ID NO: 65)

MADNVVVS LPS EIVLQPEDLPSFPDDPEVVLD FMTSQFSHLENVKWIFINTFDRLESKV VNWMA
35 KTLPIKTVGPTIP SAYLDGRLEKDKAYGLNVSKSNNGKCP IKWLDSKETASVIYISFGSLVILS
EEQVKELTNLLRDTDFS FLWVLRESEMVKLPKNFVQDTS DRGLIVNWCCQLQVL SHKAVSCFVT
HCGWNSTLEALS LGVPMVAIPQWIDQTTNAKFVADVVRVGV RVKKNEKSV AIKEELEASIRKIV
VQNGTNEFKQNAIKWKNLAKEAVDERGSSDKNIEEFVQALVAANGGGERGKQSHVIVFPFPRH
GHMSPMLQFSKRLISKGLLLTFLITSSASQSLTINIPSPSFHFKIISDLPESDDVATL DAYLR

SFRAAVTKSLSNFIDEVLTSSSNEEVPPTLIVYDSVMPWVQSVAERGLDSAPFFTESAAVNHL
LHLVYGGSLSIPPP

CmUGTc3_2 (SEQ ID NO: 66)

5 MADNVVSLPSEIVLQPEDLPSFPDDPEVVLD FMTSQFSHLENVKWIFINTFDRLESKVWNWMA
KTLPIKTVGPTIPSAAYLDGRLEKDKAYGLNVSKSNNGKCPIKWLD SKETASVIYISFGSLVILS
EEQVKELTNLLRDTDFSLWVLRESEMVKLPKNFVQDTS DRGLIVNWCCQLQVL SHKAVSCFVT
HCGWNSTLEALSLGVPMVAIPQWIDQTTNAKFVADVWRVGV RVKKNEKSVAIKEELEASIRKIV
VQNGTNEFKQNAIKWKNLAKEAVDERGSSDKNIEEFVQALVAANGGGERGKQSHVIVFPFPRH
10 GHMSPMLQFSKRLISKGLLLTFLITSSASQSMTINIPSPSFHFKIISDL PESDDVATL DAYLR
SFRAAVTKSLSNFIDEVLTSSSNEEVPPTLIVYDSVMPWVQSVAERGLDSAPFFTESAAVNHL
LHLVYGGSLSIPPP

CmUGTc3_3 (SEQ ID NO: 67)

15 MADNVVSLPSEIVLQPEDLPSFPDDPEVVLD FMTSQFSNLENVKWIFINTFDRLESKVWNWMA
KTLPIKTVGPTIPSAAYLDGRLEKDKAYGLNVSKSNNGKCPIKWLD SKETASVIYISFGSLVILS
EEQVKELTNLLRDTDFSLWVLRESEMVKLPKNFVQDTS DRGLIVNWCCQLQVL SHKAVSCFVT
HCGWNSTLEALSLGVPMVAIPQWADQTTNAKFVADVWRVGV RVKKNEKGVAIKEELEASIRKIV
VQNGTNEFKQNAIKWKNLAKEAVDERGSSDKNIEEFVQALVAANGGGERGKQSHVIVFPFPRH
GHMSPMLQFSKRLISKGLLLTFLITSSASQSLTINIPSPSFHFKIISDL PESDDVATL DAYLR
20 SFRAAVTKSLSNFIDEVLTSSSNEEVPPTLIVYDSVMPWVQSVAERGLDSAPFFTESAAVNHL
LHLVYGGSLSIPPP

CmUGTc3_CP7 (SEQ ID NO: 73)

25 MADNVVSLPSEIVLQPEDLPSFPDDPEVVLD FMTSQFSHLENVKWIFINTFDRLESKVWNWMA
KTLPIKTVGPTIPSAAYLDGRLEKDKAYGLNVSKSNNGKCPIKWLD SKETASVIYISFGSLVILS
EEQVKELTNLLRDTDFSLWVLRESEMVKLPKNFVQDTS DRGLIVNWCCQLQVL SHKAVSCFVT
HCGWNSTLEALSLGVPMVAIPQWIDQTTNAKFVADVWRVGV RVKKNEKSVAIKEELEASIRKIV
VQNGTNEFKQNAIKWKNLAKEAVDERGSSDKNIEEFVQALVEMTAANGGGERIKQSHVIVFPF
PRHGHMSPMLQFSKRLISKGLLLTFLITSSASQSLTINIPSPSFHFKIISDL PESDDVATL DA
YLR SFRAAVTKSLSNFIDEVLTSSSNEEVPPTLIVYDSVMPWVQSVAERGLDSAPFFTESAAV
NHL LHLVYGGSLSIPPP

30 CmUGTc3_CP9 (SEQ ID NO: 74)

35 MASEIVLQPEDLPSFPDDPEVVLD FMTSQFSHLENVKWIFINTFDRLESKVWNWMAKTLPIKTV
GPTIPSAAYLDGRLEKDKAYGLNVSKSNNGKCPIKWLD SKETASVIYISFGSLVILSEEQVKELT
NLLRDTDFSLWVLRESEMVKLPKNFVQDTS DRGLIVNWCCQLQVL SHKAVSCFVTHCGWNSTL
EALSLGVPMVAIPQWIDQTTNAKFVADVWRVGV RVKKNEKSVAIKEELEASIRKIVVQNGTNE
FKQNAIKWKNLAKEAVDERGSSDKNIEEFVQALVEMTAANGGGERIKQSHVIVFPFPRHGHMSP
MLQFSKRLISKGLLLTFLITSSASQSLTINIPSPSFHFKIISDL PESDDVATL DAYLR SFRAA
VTKSLSNFIDEVLTSSSNEEVPPTLIVYDSVMPWVQSVAERGLDSAPFFTESAAVNHL LHLVY
GGSLSIPPPDNVVVSLP

CmUGTc3_CP10 (SEQ ID NO: 75)

MASEIVLQPEDLPSFPDDPEVVLD FMTSQFSHLENVKWIFINTFDRLESKVVNWMAKTLPIKTV
 GPTIPSAYLDGRLEKDKAYGLNVSKSNNGKCPIKWLD SKETASVIYISFGSLVILSEEQVKELT
 NLLRDTDFS FLWVLRESEMVKLPKNFVQDTS DRGLIVNWCCQLQVL SHKAVSCFVTHCGWNSTL
 EALSLGVPMVAIPQWIDQTTNAKFVADVVRVGV RVKKNEKSVAIKEELEASIRKIVVQNGTNE
 5 FKQNAIKWK NLAKEAVDERGSSDKNIEEFVQALVAANGGGERGKQSHVIVFPFPRHGHMSPMLQ
 FSKRLISKGLLLTFLITSSASQSLTINIPSPSFHFKIISDLPE SDDVATL DAYLRSFRAAVTK
 SLSNFIDEVLTSSSNEEVPTLIVYDSVMPWVQSVAAERGLDSAPFFTESAAVNHLHLVYGG
 LSIPPPDNVVVSLP

CmUGTc3_CP16 (SEQ ID NO: 76)

10 MARLEKDKAYGLNVSKSNNGKCPIKWLD SKETASVIYISFGSLVILSEEQVKELTNLLRDTDFS
 FLWVLRESEMVKLPKNFVQDTS DRGLIVNWCCQLQVL SHKAVSCFVTHCGWNSTLEALSLGVPM
 VAIPQWIDQTTNAKFVADVVRVGV RVKKNEKSVAIKEELEASIRKIVVQNGTNEFKQNAIKWK
 NLAKEAVDERGSSDKNIEEFVQALVAANGGGERGKQSHVIVFPFPRHGHMSPMLQFSKRLISK
 LLLTFLITSSASQSLTINIPSPSFHFKIISDLPE SDDVATL DAYLRSFRAAVTKSLSNFIDEV
 15 LTSSSNEEVPTLIVYDSVMPWVQSVAAERGLDSAPFFTESAAVNHLHLVYGGSL SIPPPDNV
 VVSLPSEIVLQPEDLPSFPDDPEVVLD FMTSQFSHLENVKWIFINTFDRLESKVVNWMAKTLPI
 KTVGPTIPSAYLDG

Other sequences

pntA (SEQ ID NO: 41)

20 MRIGIPRERLTNETRVAATPKTVEQLLKLGF TVAVESGAGQLASFDDKAFVQAGAEIVEGNSVW
 QSEIILKVNAPLDDEIALLNPGTTLV SFIWPAQNPELMQKLAERNVTVMAMDSVPRI SRAQSLD
 ALSSMANIAGYRAIVEAAHEFGRFFT GQITAGKVPPAKVMVIGAGVAGLAAIGAANSLGAI VR
 AFDTRPEVKEQVQSMGAEFLELDFKEEAGSGDGYAKVMSDAFIKAEMELFAAQAKEVDIIVTTA
 LIPGKPAPKLITREMVD SMKAGSVIVDLAAQNGGNCEYTPGEI FTENGVKVIGYTDLPGRLP
 25 TQSSQLYGTNLVNLLKLLCKEKDGNITVD FDDVIRGVTVIRAGEITWPAPPIQVSAQPQAAQK
 AAPEVKTEEKCTCSPWRKYALMALAIILFGWMASVAPKEFLGHFTVFALACVVGYYVWNVSHA
 LHTPLMSVTNAISGIIIVGALLQIGQGGWVSFLSFI AVLIASINIFGGFTVTQRMLKMFRKN

pntB (SEQ ID NO: 42)

30 MSGGLVTAAYIVAAILFIFSLAGLSKHETS RQGNNGFIAGMAIALIATIFGPDTGNVGWILLAM
 VIGGAIGIRLAKKVENTEMPELVAILHSFVGLAAVLVGFNSYLHHDAGMAPILVNIHLTEVFLG
 IFIGAVTFTGSVVAFGKLCGKISSKPLMLPNRHKMNLAA LVVSFLLLIVFVRTDSVGLQVLALL
 IMTAIALVFGWHLVASIGGADMPVVV SMLNSYSGWAAAAAGFMLSNDLLIVTGALVGSSGAILS
 YIMCKAMNRSFISVIAGGFGTDGSSTGDDQEVGEHREITAEETAELLKNSHSV IITPGYGMAVA
 35 QAQYPVAEITEKL RARGINVRFGIHPVAGRLPGH MNVLLAEAKVPYDIVLEMDEINDDFADTDT
 VLVIGANDTVNPAAQDDPKSPIAGMPVLEVWKAQNVIVFKRSMNTGYAGVQNPLFFKENTHMLF
 GDAKASVDAILKAL

E. coli fldA (SEQ ID NO: 43)

MAITGIFFGSDTGNTENIAKMIQKQLGKDVADVHDIAKSSKEDLEAYDILLLLGIPTWYYGEAQC
DWDDFFPTLEEIDFNGKLVALFGCGDQEDYAEYFCDALGTIRDIIIEPRGATIVGHWPTAGYHFE
ASKGLADDDHVFVGLAIDEDRQPELTAERVEKWVKQISEELHLDEILNA

5 *Allochromatium vinosum* Cv.fdx (SEQ ID NO: 44)

MALMITDECINCDVCEPECPNGAISQGDETYVIEPSLCTECVGHYETSQCVEVCPVDCIIKDPS
HEETEDELRAKYERITGEG

CLAIMS

1. A method for producing mogrol or mogroside, comprising:
providing a microbial cell producing squalene and expressing a biosynthetic pathway that converts squalene to the mogrol or mogroside,
culturing the microbial cell under conditions suitable for producing the mogrol or mogroside, wherein the culturing produces squalene derivatives with less than about 40% by weight of off-pathway products cucurbitadienol, hydrolyzed 2,3;22,23-dioxidosqualene, and glycosylated 24,25-dihydroxycucurbitadienol.
2. The method of claim 1, wherein the biosynthetic pathway comprises a squalene epoxidase (SQE) enzyme that produces 2,3;22,23-dioxidosqualene from squalene.
3. The method of claim 2, wherein the squalene epoxidase comprises an amino acid sequence that is at least 80%, or at least 85%, at least 90%, or at least 95% identical to the amino acid sequence of SEQ ID NO: 2, and comprising one or more mutations or one or more sets of mutations listed in Tables 1, 2, and/or 3.
4. The method of claim 3, wherein the squalene epoxidase comprises one or more substitutions at positions selected from with respect to SEQ ID NO: 2, optionally wherein:
 - (i) S116 is substituted with a charged amino acid selected from Glu and Asp;
 - (ii) N134 is substituted with a non-polar amino acid selected from Gly, Ala, Ile, Leu, and Val;
 - (iii) I145 is substituted with a charged amino acid selected from His, Arg, and Lys;
 - (iv) A164 is substituted with a polar amino acid selected from Asn and Gln.
 - (v) A169 is substituted with a charged amino acid selected from Arg, Lys, Glu, Asp, or His; and/or
 - (vi) A400 is substituted with a non-polar amino acid selected from Val, Leu, and Ile.
5. The method of claim 4, wherein the squalene epoxidase comprises one or more mutations selected from S116E, N134G, I145H, A164N, A169R, and A400V with respect to SEQ ID NO: 2.

6. The method of claim 2, wherein the squalene epoxidase comprises an amino acid sequence that is at least 80%, or at least 85%, at least 90%, or at least 95% identical to the amino acid sequence of SEQ ID NO: 77, and comprising one or more mutations or one or more sets of mutations listed in Tables 4 and/or 5.
7. The method of claim 6, wherein the squalene epoxidase comprises one or more substitutions at positions selected from with respect to SEQ ID NO: 77, optionally wherein:
- (i) S366 is substituted with a charged amino acid selected from Arg, Lys and His;
 - (ii) D41 is substituted with a positively charged amino acid selected from His, Lys and Arg;
 - (iii) M315 is substituted with a non-polar amino acid selected from Leu, Ile, and Val;
8. The method of claim 7, wherein the squalene epoxidase comprises one or more mutations selected from S366R, D41H, and M315L with respect to SEQ ID NO: 77.
9. The method of any one of claims 1 to 8, wherein the biosynthetic pathway comprises a 24,25-epoxycucurbitadienol synthase (ECDS) enzyme that produces 24,25-epoxycucurbitadienol from 2,3;22,23-dioxidosqualene.
10. The method of claim 9, wherein the ECDS enzyme has substrate preference for 2,3;22,23-dioxidosqualene over 2,3-oxidosqualene.
11. The method of claim 10, wherein the ECDS enzyme comprises an amino acid sequence having a sequence identity of at least 80%, or at least 85%, or at least 90%, or at least 95%, or at least 98%, or at least 99% with respect to SEQ ID NO: 7, and having one or more mutations with respect to SEQ ID NO: 7 that increase enzyme productivity or substrate preference for 2,3;22,23-dioxidosqualene over 2,3-oxidosqualene.
12. The method of claim 11, wherein the ECDS enzyme comprises one or more mutations or sets of mutations listed in Tables 7, 8, 9, and/or 10.

13. The method of claim 11 or claim 12, wherein the ECDS comprises one or more mutations at one or more of positions selected from S24, C35, D50, N121, L245, W331, Q401, I490, I553, and A556, with respect to SEQ ID NO: 7, optionally wherein:

- (i) S24 is substituted with a polar amino acid selected from Asn and Gln;
- (ii) C35 is substituted with a charged amino acid selected from Asp and Glu;
- (iii) D50 is substituted with Glu;
- (iv) N121 is substituted with His, Lys and Arg,
- (v) L245 is substituted with Ile or Val;
- (vi) W331 is substituted with Ser, Thr, or Tyr;
- (vii) Q401 is substituted with Ala, Gly, or Val;
- (viii) I490 is substituted with Val, Leu, or Ala;
- (ix) I553 is substituted with Met;

and/or

- (x) A556 is substituted with a polar amino acid selected from Ser, Thr, and Tyr.

14. The method of claim 13, wherein the ECDS comprises one or more mutations selected from S24N, C35D, D50E, N121H, L245I, W331S, Q401A, I490V, I553M, and A556S with respect to SEQ ID NO: 7.

15. The method of any one of claims 9 to 14, wherein the ECDS enzyme comprises one or more mutations in a substrate access channel that improves substrate selectivity for 2,3;22,23-dioxidosqualene over 2,3-oxidosqualene.

16. The method of any one of claims 1 to 15, wherein the biosynthetic pathway comprises an epoxide hydrolase (EPH) enzyme that produces 24,25-dihydroxy-cucurbitadienol from 24,25-epoxycucurbitadienol.

17. The method of claim 16, wherein the EPH enzyme has substrate preference for 24,25-epoxycucurbitadienol over 2,3;22,23-dioxidosqualene.

18. The method of claim 17, wherein the EPH enzyme comprises an amino acid sequence having a sequence identity of at least 80%, or at least 85%, or at least 90%, or at least 95%, or at least 98%, or at least 99% with respect to SEQ ID NO: 5.
19. The method of claim 18, wherein the EPH has one or more mutations with respect to SEQ ID NO: 5 that increases enzyme productivity or substrate preference for 24,25-epoxycucurbitadienol over 2,3;22,23-dioxidosqualene, and wherein one or more of said mutations are selected from Table 6.
20. The method of claim 19, wherein the epoxide hydrolase comprises one or more mutations at one or more of positions selected from S68, Q128, G141, T144, A145, E163, E191, L262, C295, N299 with respect to SEQ ID NO: 5, optionally wherein:
- (i) S68 is substituted with a non-polar amino acid selected from Ala, Gly, and Val;
 - (ii) Q128 is substituted with a non-polar amino acid selected from Ala, Pro, Gly and Val;
 - (iii) G141 is substituted with a polar amino acid selected from Ser, Thr, and Tyr;
 - (iv) T144 is substituted with an amino acid selected from Gln, Asn, Ala, Gly, and Val;
 - (v) A145 is substituted with a non-polar amino acid selected from Val, Ile, and Leu;
 - (vi) E163 is substituted with a non-polar amino acid selected from Ala, Gly, and Val;
 - (vii) E191 is substituted with a non-polar amino acid selected from Gly, Ala, and Val;
 - (viii) L262 is substituted with Met;
 - (ix) C295 is substituted with a non-polar amino acid selected from Gly, Ala, and Val; and/or
 - (x) N299 is substituted with Gln.
21. The method of claim 20, wherein the epoxide hydrolase comprises one or more mutations selected from S68A, Q128A, G141S, T144Q, A145V, E163A, E191G, L262M, C295G, and N299Q with respect to SEQ ID NO: 5.
22. The method of any one of claims 1 to 21, wherein the biosynthetic pathway comprises a cytochrome P450 enzyme that produces mogrol from 24,25-dihydroxy-cucurbitadienol.

23. The method of claim 22, wherein the cytochrome P450 enzyme selectively hydroxylates C11 of 24,25-dihydroxycucurbitadienol.
24. The method of claim 23, wherein the cytochrome P450 enzyme comprises an amino acid sequence that is at least 70% identical, or at least 80% identical, or at least 90% identical, or at least 95% identical, or at least 97% identical, or at least 98% identical, or at least 99% identical to the amino acid sequence of SEQ ID NO: 14 or SEQ ID NO: 15.
25. The method of claim 24, wherein the cytochrome P450 enzyme has a replacement of all or a portion of the native transmembrane domain with the transmembrane domain from an *E. coli* inner membrane cytoplasmic C-terminus protein, wherein the *E. coli* protein is optionally SohB or FliO.
26. The method of claim 24 or 25, wherein the cytochrome P450 enzyme comprises a mutation or set of mutations listed in Tables 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, and/or 29.
27. The method of any one of claims 24 to 26, wherein the cytochrome P450 enzyme comprises a substitution at one or more positions selected from A12, I27, S63, K75, V76, N102, G125, T128, W130, L131, K132, T192, S193, L218, G219, T223, V228, T231, T232, Y233, N234, K246, F354, K369, H433, F452, V465, A468, and T482, with respect to SEQ ID NO: 15, wherein the one or more substitutions are optionally selected from:
- (i) A12 is substituted with Trp, Tyr or Phe;
 - (ii) I27 is substituted with Gly or Ala;
 - (iii) S63 is substituted with Asn, Phe, Gln, Tyr, or Trp;
 - (iv) K75 is substituted with Arg;
 - (v) V76 is substituted with Met, Leu or Ile;
 - (vi) N102 is substituted with His, Lys, or Arg;
 - (vii) G125 is substituted with an amino acid selected from Ala, Leu, Ile, and Val;
 - (viii) T128 is substituted with an amino acid selected from Gly, Ser, Pro, and Ala;

- (ix) W130 is substituted with an amino acid selected from Asn, Gly, Ser, Leu, Gln, Ala, Val, and Thr;
- (x) L131 is substituted with an amino acid selected from Val, Ile, Pro, Thr, and Ser;
- (xi) K132 is substituted with an amino acid selected from Asn, Ser, Thr, and Gln;
- (xii) T192 is substituted with Leu, Ile, or Val;
- (xiii) S193 is substituted with Thr;
- (xiv) L218 is substituted with Ile, Val, and Ala;
- (xv) G219 is substituted with an amino acid selected from Ala, Gln, Arg, Val, Asn, His, and Lys;
- (xvi) T223 is substituted with Ser;
- (xvii) V228 is substituted with Ile or Leu;
- (xviii) T231 is substituted with Phe, Trp, or Tyr;
- (xix) T232 is substituted with Ala, Gly, or Val;
- (xx) Y233 is substituted with Phe;
- (xxi) N234 is substituted with His, Arg or Lys;
- (xxii) K246 is substituted with Val, Ile or Leu;
- (xxiii) F354 is substituted with Tyr or Trp;
- (xxiv) K369 is substituted with Arg, Lys, or His;
- (xxv) H433 is substituted with Phe, Trp, or Tyr;
- (xxvi) F452 is substituted with Ile, Val, or Leu;
- (xxvii) V465 is substituted with an amino acid selected from Ile, Met, or Leu;
- (xxviii) A468 is substituted with an amino acid selected from Thr, Ser, Asn, and Gln; and
- (xxix) T482 is substituted with Ser.

28. The method of claim 27, wherein the cytochrome P450 enzyme comprises one or more substitutions selected from A12W, I27G, S63F, S63N, K75R, V76M, N102H, G125A, T128G, W130N, L131I, K132N, T192L, S193T, L218I, G219A, T223S, V228I, T231F, T232A, Y233F, N234H, K246V, F354Y, K369R, H433F, F452I, V465I, A468T and T482S with respect to SEQ ID NO: 15.

29. The method of any one of claims 24 to 28, wherein the cytochrome P450 enzyme comprises an N-terminal deletion of at least about 20, or at least about 25, or at least about 30, or at least about 35, or at least about 40, or at least about 45, or at least about 50 amino acids with respect to SEQ ID NO: 15.
30. The method of claim 29, wherein the cytochrome P450 enzyme comprises a deletion selected from a deletion of L3 to H29 with respect to SEQ ID NO: 15.
31. The method of claim 30, wherein the cytochrome P450 enzyme comprises an amino acid sequence that is at least 70% identical, or at least 80% identical, or at least 90% identical, or at least 95% identical, or at least 97% identical, or at least 98% identical, or at least 99% identical to the amino acid sequence of SEQ ID NO: 85, and comprising one or more mutations or one or more sets of mutations listed in Tables 26, 27, 28, and/or 29.
32. The method of claim 31, wherein the cytochrome P450 enzyme comprises a substitution at one or more positions selected K8, D9, S10, N13, V15, I45, K46, K47, M49, K50, R51, I191, A192, F406, E411, Y412, and S413, with respect to SEQ ID NO: 85, wherein the one or more substitutions are optionally selected from:
- (i) K8 is substituted with Arg;
 - (ii) D9 is substituted with Asn or Gln;
 - (iii) S10 is substituted with Arg, Lys or His;
 - (iv) N13 is substituted with Lys, Arg, or His;
 - (v) V15 is substituted with Lys, Arg, or His;
 - (vi) I45 is substituted with Val, Ile, or Leu;
 - (vii) K46 is substituted with Arg, Lys or His;
 - (viii) K47 is substituted with Glu or Asp;
 - (ix) M49 is substituted with Val, Ile, or Leu;
 - (x) K50 is substituted with Glu or Asp;
 - (xi) R51 is substituted with Lys;
 - (xii) I191 is substituted with Leu or Val;
 - (xiii) A192 is substituted with Lys, Arg, or His;

- (xiv) F406 is substituted with Tyr or Trp;
- (xv) E411 is substituted with Asp;
- (xvi) Y412 is substituted with Phe, Leu, Ile, or Trp; and
- (xvii) S413 is substituted with Ala, Gly, or Val.

33. The method of claim 32, wherein the cytochrome P450 enzyme comprises one or more substitutions selected from K8R, D9N, S10R, N13K, and V15K I45V, K46R, K47E, M49V, K50E, R51K, F406Y, E411D, Y412F, S413A, I191L and A192K substitutions with respect to SEQ ID NO: 85.

34. The method of any one of claims 22 to 33, wherein the microbial cell expresses a P450 reductase.

35. The method of claim 34, wherein the P450 reductase comprises an amino acid sequence that is at least 80%, or at least 85%, or at least 90%, or at least 95%, or at least 97%, or at least 98%, or at least 99% identical to SEQ ID NO: 22 or SEQ ID NO: 24.

36. The method of claim 35, wherein the cytochrome P450 reductase enzyme comprises a replacement all or a portion of the native transmembrane domain with the transmembrane domain from an *E. coli* inner membrane cytoplasmic C-terminus protein, wherein the *E. coli* protein is selected from YcgG, yhcB, zipA, waaA, sohB, djlA, lpxK, ypfN, and yhhM.

37. The method of claim 36, wherein the *E. coli* protein is YcgG.

38. The method of claim 36 or 37, wherein from about 60 to about 80 amino acids of the N-terminus of SEQ ID NO: 22 or SEQ ID NO: 24 are deleted, and optionally from about 65 to about 78 amino acids of the N-terminus of SEQ ID NO: 22 or SEQ ID NO: 24.

39. The method of claim 37 or 38, wherein: the native N-terminus of SEQ ID NO: 22 or SEQ ID NO: 24 is replaced with 20 to 30 amino acids from the N-terminus of YcgG, and optionally replaced with 25 amino acids from the N-terminus of YcgG.

40. The method of any one of claims 37 to 39, wherein the cytochrome P450 reductase comprises one or more mutations listed in Table 31 with respect to SEQ ID NO: 22, or comprises one or more mutation listed in Table 32 or 33 with respect to SEQ ID NO: 24, and optionally a substitution of L90 residue with respect to SEQ ID NO: 24.

41. The method of any one of claims 1 to 40, wherein the biosynthetic pathway comprises one or more uridine diphosphate dependent glycosyltransferase (UGT) enzymes that produce one or more mogrosides from mogrol, where the one or more mogrosides are selected from Mog 1A, Mog. 1E, Mog. IIA1, Mog. IIE, Mog. IIA2, Mog. IIIA1, Mog. III, Mog. IIIA2, siamenoside, Mog. IVA, Mog. IV, Mog. V, and iso-Mog. V, and optionally wherein the mogroside is substantially Mog. V.

42. The method of claim 41, wherein the UGT enzymes produce at least 2 times more MogV than glycosylated products of 24,25-dihydroxycucurbitadienol.

43. The method of claim 42, wherein the UGT enzymes produce at least 3 times, or at least 5 times, or at least 10 times more MogV than glycosylated products of 24,25-dihydroxycucurbitadienol.

44. The method of any one of claims 1 to 43, wherein the biosynthetic pathway comprises a UGT enzyme that produces a C-3 mogrol glycoside from mogrol (UGTc3).

45. The method of claim 44, wherein the UGTc3 enzyme comprises an amino acid sequence that is at least 70% identical, or at least 80%, or at least 85%, or at least 90%, or at least 95%, or at least 97%, or at least 98%, or at least 99% identical to the amino acid sequence of SEQ ID NO: 27, wherein the UGT enzyme optionally comprises one or more mutations with respect to SEQ ID NO: 27 that result in greater specificity and/or activity for glycosylation of C-3 position compared to the UGT enzyme having an amino acid sequence of SEQ ID NO: 27.

46. The method of claim 45, wherein one or more mutations are selected from Table 35 and/or Table 36.

47. The method of claim 46, wherein the UGTc3 enzyme comprises one or more mutations at one or more positions selected from L41, D49, T74, C127, F303, and A307 with respect to SEQ ID NO: 27, optionally wherein the mutations are selected from:

- (i) L41 is substituted with an aromatic amino acid selected from Phe, Tyr and Trp;
- (ii) D49 is substituted with Glu;
- (iii) T74 is substituted with Met;
- (iv) C127 is substituted with an aromatic amino acid selected from Phe, Tyr and Trp;
- (v) F303 is substituted with Tyr; and
- (vi) A307 is substituted with an amino acid selected from Ile, Leu, and Val.

48. The method of claim 47, wherein the UGTc3 enzyme comprises one or more of the following substitutions with respect to SEQ ID NO: 27: L41F, D49E, T74M, C127F, F303Y, and A307I.

49. The method of claim 44, wherein the UGTc3 enzyme comprises an amino acid sequence that is at least 70% identical, or at least 80%, or at least 85%, or at least 90%, or at least 95%, or at least 97%, or at least 98%, or at least 99% identical to the amino acid sequence of SEQ ID NO: 60 or SEQ ID NO: 65, wherein the UGTc3 enzyme optionally comprises one or more mutations with respect to SEQ ID NO: 60 or SEQ ID NO: 65 that result in greater specificity and/or activity for glycosylation of C-3 position compared to the UGT enzyme having an amino acid sequence of SEQ ID NO: 60 or SEQ ID NO: 65.

50. The method of claim 49, wherein one or more mutations are selected from Table 37 with respect to SEQ ID NO: 60, and/or Table 39 with respect to SEQ ID NO: 65.

51. The method of claim 49 or claim 50, wherein the UGTc3 enzyme is a circular permutant with respect to SEQ ID NO: 60.

52. The method of claim 51, wherein the circular permutant has a cut site corresponding to a position between amino acid 150 to 180, or amino acid 155 to 180, or amino acid 160 to 180, or amino acid 165 to 180, or amino acid 169 to 180 of SEQ ID NO: 60, and optionally having a linker sequence between the amino acids that correspond to the N-terminal and C-terminal residues of SEQ ID NO: 60, and wherein the linker optionally has a length of from 2 to 25 amino acids.

53. The method of claim 52, wherein the UGTc3 enzyme comprises one or more mutations at one or more positions selected from H40, I216, S241, and L352 with respect to SEQ ID NO: 65, optionally wherein the mutations are selected from:

- (i) H40 is substituted with Asp, Gly, Gln, Ala or Val;
- (ii) I216 is substituted with Ala, Gly, or Val;
- (iii) S241 is substituted with Gly, Ala, or Val; and
- (iv) L352 is substituted with Met.

54. The method of claim 53, wherein the UGTc3 enzyme comprises one or more of the following substitutions with respect to SEQ ID NO: 65: H40N, I216A, S241G, and L352M.

55. The method of any one of claims 1 to 54, wherein the biosynthetic pathway comprises a UGT enzyme that produces a C-24 mogrol glycoside from mogrol (UGTc24).

56. The method of claim 55, wherein the UGTc24 enzyme comprises an amino acid sequence that is at least 70% identical, or at least 80% identical, or at least 85% identical, or at least 90% identical, or at least 95% identical, or at least 97% identical, or at least 98% identical or at least 99% identical to the amino acid sequence of SEQ ID NO: 31, wherein the UGTc24 enzyme optionally comprises one or more mutations resulting in greater specificity and/or activity for glycosylation of C-24 position compared to the UGT enzyme having an amino acid sequence of SEQ ID NO: 31.

57. The method of claim 56, wherein one or more mutations are selected from Table 40, Table 41, Table 42, Table 43, Table 44 and/or Table 45.

58. The method of claim 56 or 57, wherein the UGTc24 enzyme comprises one or more mutations at positions selected from L18, A74, S88, I91, T95, H101, T184, F187, P190, T273, N306, M332, and Y381, with respect to SEQ ID NO: 31, wherein one or more mutations are optionally selected from:

- (i) L18 is substituted with Met;
- (ii) I63 is substituted with Val, Ala or Gly;
- (iii) A74 is substituted with an acidic amino acid selected from Glu and Asp;
- (iv) S88 is substituted with an amino acid selected from Ala, Gly, Leu, Val, and Ile;
- (v) I91 is substituted with an aromatic amino acid selected from Phe, Tyr and Trp;
- (vi) T95 is substituted with an amino acid selected from Ala, Gly, and Val;
- (vii) H101 is substituted with Pro;
- (viii) T184 is substituted with an amino acid selected from Phe, Tyr, and Trp;
- (ix) F187 is substituted with Tyr or Trp
- (x) P190 is substituted with an amino acid selected from Glu and Asp;
- (xi) T273 is substituted with Ser or is deleted;
- (xii) N306 is substituted with Lys, His, or Arg;
- (xiii) M332 is substituted with an amino acid selected from Leu, Ile, Val, and Ala; and
- (xiv) Y381 is substituted with an amino acid selected from Phe and Trp.

59. The method of claim 58, wherein the UGTc24 enzyme comprises one or more mutations selected from L18M, I63V, A74E, S88A, I91F, T95A, H101P, T184F, F187Y, P190E, T273S, N306K, M332L, and Y381F with respect to SEQ ID NO: 31.

60. The method of any one of claims 1 to 59, wherein the biosynthetic pathway comprises a uridine diphosphate dependent glycosyltransferase (UGT) enzyme that produces a 1,6 and/or 1,2 branched glycoside of a C-3 and/or C-24 glycoside of mogrol (UGT1216).

61. The method of claim 60, wherein the UGT1216 is a circular permutant of the enzyme represented by SEQ ID NO: 35, or a sequence having at least 70% identity thereto.

62. The method of claim 61, wherein the circular permutant has a cut site at a position corresponding to amino acid 180 to 195 of SEQ ID NO: 35, or a sequence having at least 70% identity thereto, and

optionally at a position corresponding to 185 to 190 of SEQ ID NO: 35, or a sequence having at least 70% identity thereto, and

optionally at position 187 of SEQ ID NO: 35, or a sequence having at least 70% identity thereto.

63. The method of claim 62, wherein the UGT1216 enzyme comprises an amino acid sequence that is at least 70%, or at least 80%, or at least 90%, or at least 95%, or at least 97%, or at least 98%, or at least 99% identical to the amino acid sequence of SEQ ID NO: 36, wherein the UGT1216 enzyme has one or more mutations resulting in greater specificity and/or activity for to 1,6 and/or 1,2 glycosylation of a C-3 and/or C-24 glycoside of mogrol compared to the UGT enzyme having an amino acid sequence of SEQ ID NO: 36.

64. The method of claim 63, wherein the UGT1216 enzyme has one or more mutations resulting in greater production of MogV from MogIIE compared to the UGT enzyme having an amino acid sequence of SEQ ID NO: 36.

65. The method of claim 64, wherein one or more mutations or sets of mutations of the UGT1216 are selected from Table 46, Table 47, Table 48, and/or Table 49.

66. The method of claim 64 or 65, wherein the UGT1216 enzyme comprises mutations at one or more positions selected from L14, H19, W160, L266, T281, S318, and H390 with respect to SEQ ID NO: 36, optionally wherein:

- (i) L14 is substituted with a non-polar amino acid selected from Val, Ile, Ala, and Gly;
- (ii) H19 is substituted with a charged amino acid selected from Glu, Lys, Asp and Glu;
- (iii) W160 is substituted with Pro, Ser, or Thr;
- (iv) L266 is substituted with a polar amino acid selected from Gln and Asn;
- (v) T281 is substituted with Ser;
- (vi) S318 is substituted with an amino acid selected from Lys, Pro, and Arg; and

(iv) H390 is substituted with an amino acid selected from Asp and Glu.

67. The method of claim 66, wherein the UGT1216 enzyme comprises one or more substitutions selected from L14V, H19E, W160P, L266Q, T281S, S318K and H390D with respect to SEQ ID NO: 36.

68. The method of claim 67, wherein the UGT1216 enzyme comprises the mutations L14V, W160P, L266Q, and H390D substitutions with respect to SEQ ID NO: 36.

69. The method of claim 62, wherein the UGT1216 enzyme comprises an amino acid sequence that is at least 70%, or at least 80%, or at least 90%, or at least 95%, or at least 97%, or at least 98%, or at least 99% identical to the amino acid sequence of SEQ ID NO: 39, wherein the UGT1216 enzyme as one or more mutations resulting in greater specificity and/or activity for 1,6 and 1,2 branched glycosylation of a C-3 and/or C-24 glycoside of mogrol compared to the UGT enzyme having an amino acid sequence of SEQ ID NO: 39.

70. The method of claim 69, wherein the UGT1216 enzyme has one or more mutations resulting in greater production of MogIVA from MogIIE or MogV from MogIIE, compared to the UGT enzyme having an amino acid sequence of SEQ ID NO: 39.

71. The method of claim 70, wherein the UGT1216 enzyme comprises one or more mutations listed in Table 50.

72. The method of claim 70 or 71, wherein the UGT enzyme comprises mutation(s) at positions selected from T147 and N207 with respect to SEQ ID NO: 39, optionally wherein:

- (i) T147 is substituted with a non-polar amino acid selected from Leu, Ile, and Val; and/or
- (ii) N207 is substituted with a basic amino acid selected from Lys, Arg, and His.

73. The method of claim 72, wherein the UGT enzyme comprises T147L and/or N207K substitutions with respect to SEQ ID NO: 39.

74. The method of any one of claims 1 to 73, wherein the microbial cell expresses a squalene synthase (SQS) enzyme.
75. The method of claim 74, wherein the microbial cell expresses a farnesyl pyrophosphate (FPP) synthase.
76. The method of claim 74 or 75, wherein the microbial cell overexpresses one or more enzymes of an MEP or MVA pathway.
77. The method of claim 76, wherein the microbial cell is a bacterium, and is optionally selected from *Escherichia spp.*, *Bacillus spp.*, *Corynebacterium spp.*, *Rhodobacter spp.*, *Zymomonas spp.*, *Vibrio spp.*, and *Pseudomonas spp.*, and is optionally a species selected from *Escherichia coli*, *Bacillus subtilis*, *Corynebacterium glutamicum*, *Rhodobacter capsulatus*, *Rhodobacter sphaeroides*, *Zymomonas mobilis*, *Vibrio natriegens*, and *Pseudomonas putida*.
78. The method of claim 77, wherein the microbial cell is a yeast cell, optionally selected from a species of *Saccharomyces*, *Pichia*, or *Yarrowia*, and is optionally *Saccharomyces cerevisiae*, *Pichia pastoris*, and *Yarrowia lipolytica*.
79. The method of any one of claims 1 to 78, wherein the microbial cell expresses or a recombinant *pntAB* or a variant thereof.
80. The method of any one of claims 1 to 79, wherein the microbial cell expresses at least one recombinant electron carrier.
81. The method of claim 80, wherein the at least one recombinant electron carrier is a recombinant ferredoxin (fdx) or recombinant flavodoxin (Fld).
82. The method of any one of claims 1 to 81, wherein the microbial cell produces more than about 40%, or more than about 50%, or more than about 60%, or more than about 70%, or more than about 80% on pathway products 2,3-oxidosqualene, 2,3;22,23-dioxidosqualene, 24,25-epoxy cucurbitadienol, 24,25-dihydroxy cucurbitadienol, mogrol, and mogrosides.

83. The method of any one of claims 1 to 82, wherein the microbial cell produces less than about 35%, or less than about 30%, or less than about 25%, or less than about 20% off pathway products cucurbitadienol, hydrolyzed 2,3;22,23-dioxidosqualene, and glycosylated 24,25-dihydroxycucurbitadienol.
84. The method of any one of claims 1 to 83, wherein the microbial cell comprises a mutation that reduces the amount or activity of a transporter.
85. The method of claim 84, wherein the transporter is an ABC transporter.
86. The method of claim 84 or claim 85, wherein the transporter is YbhGFSR or an ortholog thereof.
87. The method of claim 86, wherein the microbial cell comprises a hypomorphic mutation or a null mutation in one or more of *ybhG*, *ybhF*, *ybhS*, and *ybhR*.
88. The method of claim 87, wherein the null mutation is a complete or partial deletion of one or more of *ybhG*, *ybhF*, *ybhS*, and *ybhR*.
89. The method of claim 87, wherein the hypomorphic mutation is a promoter mutation, and/or a mutation selected from a mutant ribosome binding site, one or more mutations adding one or more rare codons, and adding a secondary RNA structure in one or more of *ybhG*, *ybhF*, *ybhS*, and *ybhR*.
90. A method for producing mogrol or mogroside, comprising:
providing a microbial cell producing squalene and expressing a biosynthetic pathway that converts squalene to the mogrol or mogroside,
wherein the microbial cell comprises one or more genetic modifications that reduces the efflux of an intermediate of mogrol synthesis, optionally wherein the intermediate is 2,3;22,23-dioxidosqualene,

wherein the microbial cell exhibits an increase in mogrol synthesis compared to a cell that does not have the genetic modifications.

91. The method of claim 90, wherein the genetic modification comprises a mutation that reduces the amount or activity of a transporter.

92. The method of claim 90 or claim 91, wherein the transporter is YbhGFSR or an ortholog thereof.

93. The method of claim 92, wherein the microbial cell comprises a hypomorphic mutation or a null mutation in one or more of *ybhG*, *ybhF*, *ybhS*, and *ybhR*.

94. The method of claim 93, wherein the null mutation is a complete or partial deletion of one or more of *ybhG*, *ybhF*, *ybhS*, and *ybhR*.

95. The method of claim 93, wherein the hypomorphic mutation is a promoter mutation, and/or a mutation selected from a mutant ribosome binding site, one or more mutations adding one or more rare codons, and adding a secondary RNA structure in one or more of *ybhG*, *ybhF*, *ybhS*, and *ybhR*.

96. The method of any one of claims 90 to 95, wherein the microbial cell is *E. coli*.

97. The methods of any one of claims 90 to 96, wherein the biosynthetic pathway comprises one or more enzymes disclosed herein.

98. A microbial cell producing squalene and expressing a biosynthetic pathway that converts squalene to mogrol or mogroside, wherein when the microbial cell is cultured under conditions suitable for producing the mogrol or mogroside, the microbial cell produces squalene derivatives with less than about 40% by weight of off-pathway products cucurbitadienol, hydrolyzed 2,3;22,23-dioxidosqualene, and glycosylated 24,25-dihydroxycucurbitadienol.

99. The microbial cell of claim 98, wherein the biosynthetic pathway comprises a squalene epoxidase (SQE) enzyme that produces 2,3;22,23-dioxidosqualene from squalene.

100. The microbial cell of claim 99, wherein the squalene epoxidase comprises an amino acid sequence that is at least 80%, or at least 85%, at least 90%, or at least 95% identical to the amino acid sequence of SEQ ID NO: 2, and comprising one or more mutations or one or more sets of mutations listed in Tables 1, 2, and 3.

101. The microbial cell of claim 100, wherein the squalene epoxidase comprises one or more substitutions at positions selected from with respect to SEQ ID NO: 2, optionally wherein:

- (i) S116 is substituted with a charged amino acid selected from Glu and Asp;
- (ii) N134 is substituted with a non-polar amino acid selected from Gly, Ala, Ile, Leu, and Val;
- (iii) I145 is substituted with a charged amino acid selected from His, Arg, and Lys;
- (iv) A164 is substituted with a polar amino acid selected from Asn and Gln.
- (v) A169 is substituted with a charged amino acid selected from Arg, Lys, Glu, Asp, or His; and/or
- (vi) A400 is substituted with a non-polar amino acid selected from Val, Leu, and Ile.

102. The microbial cell of claim 101, wherein the squalene epoxidase comprises one or more mutations selected from S116E, N134G, I145H, A164N, A169R, and A400V with respect to SEQ ID NO: 2.

103. The microbial cell of claim 99, wherein the squalene epoxidase comprises an amino acid sequence that is at least 80%, or at least 85%, at least 90%, or at least 95% identical to the amino acid sequence of SEQ ID NO: 77, and comprising one or more mutations or one or more sets of mutations listed in Tables 4 and/or 5.

104. The microbial cell of claim 103, wherein the squalene epoxidase comprises one or more substitutions at positions selected from with respect to SEQ ID NO: 77, optionally wherein:

- (i) S366 is substituted with a charged amino acid selected from Arg, Lys and His;

(ii) D41 is substituted with a positively charged amino acid selected from His, Lys and Arg;

(iii) M315 is substituted with a non-polar amino acid selected from Leu, Ile, and Val;

105. The microbial cell of claim 104, wherein the squalene epoxidase comprises one or more mutations selected from S366R, D41H, and M315L with respect to SEQ ID NO: 77.

106. The microbial cell of any one of claims 98 to 105, wherein the biosynthetic pathway comprises a 24,25-epoxycucurbitadienol synthase (ECDS) enzyme that produces 24,25-epoxycucurbitadienol from 2,3,22,23-dioxidosqualene.

107. The microbial cell of claim 106, wherein the ECDS enzyme has substrate preference for 2,3,22,23-dioxidosqualene over 2,3-oxidosqualene.

108. The microbial cell of claim 107, wherein the ECDS enzyme comprises an amino acid sequence having a sequence identity of at least 80%, or at least 85%, or at least 90%, or at least 95%, or at least 98%, or at least 99% with respect to SEQ ID NO: 7, and having one or more mutations with respect to SEQ ID NO: 7 that increase enzyme productivity or substrate preference for 2,3,22,23-dioxidosqualene over 2,3-oxidosqualene.

109. The microbial cell of claim 108, wherein the ECDS enzyme comprises one or more mutations or sets of mutations listed in Tables 7, 8, 9, and/or 10.

110. The microbial cell of claim 108 or claim 109, wherein the ECDS comprises one or more mutations at one or more of positions selected from S24, C35, D50, N121, L245, W331, Q401, I490, I553, and A556, with respect to SEQ ID NO: 7, optionally wherein:

(i) S24 is substituted with a polar amino acid selected from Asn and Gln;

(ii) C35 is substituted with a charged amino acid selected from Asp and Glu;

(iii) D50 is substituted with Glu;

(iv) N121 is substituted with His, Lys and Arg,

(v) L245 is substituted with Ile or Val;

- (vi) W331 is substituted with Ser, Thr, or Tyr;
- (vii) Q401 is substituted with Ala, Gly, or Val;
- (viii) I490 is substituted with Val, Leu, or Ala;
- (ix) I553 is substituted with Met;

and/or

- (x) A556 is substituted with a polar amino acid selected from Ser, Thr, and Tyr.

111. The microbial cell of claim 110, wherein the ECDS comprises one or more mutations selected from S24N, C35D, D50E, N121H, L245I, W331S, Q401A, I490V, I553M, and A556S with respect to SEQ ID NO: 7.

112. The microbial cell of any one of claims 106 to 111, wherein the ECDS enzyme comprises one or more mutations in a substrate access channel that improves substrate selectivity for 2,3;22,23-dioxidosqualene over 2,3-oxidosqualene.

113. The microbial cell of any one of claims 98 to 112, wherein the biosynthetic pathway comprises an epoxide hydrolase (EPH) enzyme that produces 24,25-dihydroxy-cucurbitadienol from 24,25-epoxycucurbitadienol.

114. The microbial cell of claim 113, wherein the EPH enzyme has substrate preference for 24,25-epoxycucurbitadienol over 2,3;22,23-dioxidosqualene.

115. The microbial cell of claim 114, wherein the EPH enzyme comprises an amino acid sequence having a sequence identity of at least 80%, or at least 85%, or at least 90%, or at least 95%, or at least 98%, or at least 99% with respect to SEQ ID NO: 5, optionally wherein the EPH has one or more mutations with respect to SEQ ID NO: 5 that increases enzyme productivity or substrate preference for 24,25-epoxycucurbitadienol over 2,3;22,23-dioxidosqualene, and wherein one or more of said mutations are selected from Table 6.

116. The microbial cell of claim 115, wherein the epoxide hydrolase comprises one or more mutations at one or more of positions selected from S68, Q128, G141, T144, A145, E163, E191, L262, C295, N299 with respect to SEQ ID NO: 5, optionally wherein:

- (i) S68 is substituted with a non-polar amino acid selected from Ala, Gly, and Val;
- (ii) Q128 is substituted with a non-polar amino acid selected from Ala, Pro, Gly and Val;
- (iii) G141 is substituted with a polar amino acid selected from Ser, Thr, and Tyr;
- (iv) T144 is substituted with an amino acid selected from Gln, Asn, Ala, Gly, and Val;
- (v) A145 is substituted with a non-polar amino acid selected from Val, Ile, and Leu;
- (vi) E163 is substituted with a non-polar amino acid selected from Ala, Gly, and Val;
- (vii) E191 is substituted with a non-polar amino acid selected from Gly, Ala, and Val;
- (viii) L262 is substituted with Met;
- (ix) C295 is substituted with a non-polar amino acid selected from Gly, Ala, and Val;
- and/or
- (x) N299 is substituted with Gln.

117. The microbial cell of claim 116, wherein the epoxide hydrolase comprises one or more mutations selected from S68A, Q128A, G141S, T144Q, A145V, E163A, E191G, L262M, C295G, and N299Q with respect to SEQ ID NO: 5.

118. The microbial cell of any one of claims 98 to 117, wherein the biosynthetic pathway comprises a cytochrome P450 enzyme that produces mogrol from 24,25-dihydroxycucurbitadienol.

119. The microbial cell of claim 118, wherein the cytochrome P450 enzyme selectively hydroxylates C11 of 24,25-dihydroxycucurbitadienol, optionally wherein the cytochrome P450 enzyme comprises an amino acid sequence that is at least 70% identical, or at least 80% identical, or at least 90% identical, or at least 95% identical, or at least 97% identical, or at least 98% identical, or at least 99% identical to the amino acid sequence of SEQ ID NO: 14 or SEQ ID NO: 15.

120. The microbial cell of claim 119, wherein the cytochrome P450 enzyme has a replacement of all or a portion of the native transmembrane domain with the transmembrane domain from an *E. coli* inner membrane cytoplasmic C-terminus protein, wherein the *E. coli* protein is optionally SohB or FliO.

121. The microbial cell of claim 119 or 120, wherein the cytochrome P450 enzyme comprises a mutation or set of mutations listed in Tables 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, and/or 29.

122. The microbial cell of any one of claims 118 to 121, wherein the cytochrome P450 enzyme comprises a substitution at one or more positions selected from A12, I27, S63, K75, V76, N102, G125, T128, W130, L131, K132, T192, S193, L218, G219, T223, V228, T231, T232, Y233, N234, K246, F354, K369, H433, F452, V465, A468, and T482, with respect to SEQ ID NO: 15, wherein the one or more substitutions are optionally selected from:

- (i) A12 is substituted with Trp, Tyr or Phe;
- (ii) I27 is substituted with Gly or Ala;
- (iii) S63 is substituted with Asn, Phe, Gln, Tyr, or Trp;
- (iv) K75 is substituted with Arg;
- (v) V76 is substituted with Met, Leu or Ile;
- (vi) N102 is substituted with His, Lys, or Arg;
- (vii) G125 is substituted with an amino acid selected from Ala, Leu, Ile, and Val;
- (viii) T128 is substituted with an amino acid selected from Gly, Ser, Pro, and Ala;
- (ix) W130 is substituted with an amino acid selected from Asn, Gly, Ser, Leu, Gln, Ala, Val, and Thr;
- (x) L131 is substituted with an amino acid selected from Val, Ile, Pro, Thr, and Ser;
- (xi) K132 is substituted with an amino acid selected from Asn, Ser, Thr, and Gln;
- (xii) T192 is substituted with Leu, Ile, or Val;
- (xiii) S193 is substituted with Thr;
- (xiv) L218 is substituted with Ile, Val, and Ala;
- (xv) G219 is substituted with an amino acid selected from Ala, Gln, Arg, Val, Asn, His, and Lys;

- (xvi) T223 is substituted with Ser;
- (xvii) V228 is substituted with Ile or Leu;
- (xviii) T231 is substituted with Phe, Trp, or Tyr;
- (xix) T232 is substituted with Ala, Gly, or Val;
- (xx) Y233 is substituted with Phe;
- (xxi) N234 is substituted with His, Arg or Lys;
- (xxii) K246 is substituted with Val, Ile or Leu;
- (xxiii) F354 is substituted with Tyr or Trp;
- (xxiv) K369 is substituted with Arg, Lys, or His;
- (xxv) H433 is substituted with Phe, Trp, or Tyr;
- (xxvi) F452 is substituted with Ile, Val, or Leu;
- (xxvii) V465 is substituted with an amino acid selected from Ile, Met, or Leu;
- (xxviii) A468 is substituted with an amino acid selected from Thr, Ser, Asn, and Gln; and
- (xxix) T482 is substituted with Ser.

123. The microbial cell of claim 122, wherein the cytochrome P450 enzyme comprises one or more substitutions selected from A12W, I27G, S63F, S63N, K75R, V76M, N102H, G125A, T128G, W130N, L131I, K132N, T192L, S193T, L218I, G219A, T223S, V228I, T231F, T232A, Y233F, N234H, K246V, F354Y, K369R, H433F, F452I, V465I, A468T and T482S with respect to SEQ ID NO: 15.

124. The microbial cell of any one of claims 118 to 123, wherein the cytochrome P450 enzyme comprises an N-terminal deletion of at least about 20, or at least about 25, or at least about 30, or at least about 35, or at least about 40, or at least about 45, or at least about 50 amino acids with respect to SEQ ID NO: 15.

124. The microbial cell of claim 123, wherein the cytochrome P450 enzyme comprises a deletion selected from a deletion of L3 to H29 with respect to SEQ ID NO: 15, optionally wherein the cytochrome P450 enzyme comprises an amino acid sequence that is at least 70% identical, or at least 80% identical, or at least 90% identical, or at least 95% identical, or at least 97% identical, or at least 98% identical, or at least 99% identical to the amino acid sequence of SEQ ID NO: 85,

and comprising one or more mutations or one or more sets of mutations listed in Tables 26, 27, 28, and/or 29.

126. The microbial cell of claim 125, wherein the cytochrome P450 enzyme comprises a substitution at one or more positions selected K8, D9, S10, N13, V15, I45, K46, K47, M49, K50, R51, I191, A192, F406, E411, Y412, and S413, with respect to SEQ ID NO: 85, wherein the one or more substitutions are optionally selected from:

- (i) K8 is substituted with Arg;
- (ii) D9 is substituted with Asn or Gln;
- (iii) S10 is substituted with Arg, Lys or His;
- (iv) N13 is substituted with Lys, Arg, or His;
- (v) V15 is substituted with Lys, Arg, or His;
- (vi) I45 is substituted with Val, Ile, or Leu;
- (vii) K46 is substituted with Arg, Lys or His;
- (viii) K47 is substituted with Glu or Asp;
- (ix) M49 is substituted with Val, Ile, or Leu;
- (x) K50 is substituted with Glu or Asp;
- (xi) R51 is substituted with Lys;
- (xii) I191 is substituted with Leu or Val;
- (xiii) A192 is substituted with Lys, Arg, or His;
- (xiv) F406 is substituted with Tyr or Trp;
- (xv) E411 is substituted with Asp;
- (xvi) Y412 is substituted with Phe, Leu, Ile, or Trp; and
- (xvii) S413 is substituted with Ala, Gly, or Val.

127. The microbial cell of claim 126, wherein the cytochrome P450 enzyme comprises one or more substitutions selected from K8R, D9N, S10R, N13K, and V15K I45V, K46R, K47E, M49V, K50E, R51K, F406Y, E411D, Y412F, S413A, I191L and A192K substitutions with respect to SEQ ID NO: 85.

128. The microbial cell of any one of claims 118 to 127, wherein the microbial cell expresses a P450 reductase.

129. The microbial cell of claim 128, wherein the P450 reductase comprises an amino acid sequence that is at least 80%, or at least 85%, or at least 90%, or at least 95%, or at least 97%, or at least 98%, or at least 99% identical to SEQ ID NO: 22 or SEQ ID NO: 24.

130. The microbial cell of claim 129, wherein the cytochrome P450 reductase enzyme comprises a replacement all or a portion of the native transmembrane domain with the transmembrane domain from an *E. coli* inner membrane cytoplasmic C-terminus protein, wherein the *E. coli* protein is selected from YcgG, yhcB, zipA, waaA, sohB, djlA, lpxK, ypfN, and yhhM.

131. The microbial cell of claim 130, wherein the *E. coli* protein is YcgG.

132. The microbial cell of claim 130 or 131, wherein from about 60 to about 80 amino acids of the N-terminus of SEQ ID NO: 22 or SEQ ID NO: 24 are deleted, and optionally from about 65 to about 78 amino acids of the N-terminus of SEQ ID NO: 22 or SEQ ID NO: 24.

133. The microbial cell of claim 132, wherein: the native N-terminus of SEQ ID NO: 22 or SEQ ID NO: 24 is replaced with 20 to 30 amino acids from the N-terminus of YcgG, and optionally replaced with 25 amino acids from the N-terminus of YcgG.

134. The microbial cell of any one of claims 128 to 133, wherein the cytochrome P450 reductase comprises one or more mutations listed in Table 31 with respect to SEQ ID NO: 22, or comprises one or more mutation listed in Table 32 or 33 with respect to SEQ ID NO: 24, and optionally a substitution of L90 residue with respect to SEQ ID NO: 24.

135. The microbial cell of any one of claims 98 to 134, wherein the biosynthetic pathway comprises one or more uridine diphosphate dependent glycosyltransferase (UGT) enzymes that produce MogV from mogrol.

136. The microbial cell of claim 135, wherein the UGT enzymes produce at least 2 times more MogV than glycosylated products of 24,25-dihydroxycucurbitadienol.

137. The microbial cell of claim 136, wherein the UGT enzymes produce at least 3 times, or at least 5 times, or at least 10 times more MogV than glycosylated products of 24,25-dihydroxycucurbitadienol.

138. The microbial cell of any one of claims 135 to 137, wherein the biosynthetic pathway comprises a UGT enzyme that produces a C-3 mogrol glycoside from mogrol (UGTc3).

139. The microbial cell of claim 138, wherein the UGTc3 enzyme comprises an amino acid sequence that is at least 70% identical, or at least 80%, or at least 85%, or at least 90%, or at least 95%, or at least 97%, or at least 98%, or at least 99% identical to the amino acid sequence of SEQ ID NO: 27, wherein the UGT enzyme optionally comprises one or more mutations with respect to SEQ ID NO: 27 that result in greater specificity and/or activity for glycosylation of C-3 position compared to the UGT enzyme having an amino acid sequence of SEQ ID NO: 27.

140. The microbial cell of claim 139, wherein one or more mutations are selected from Table 35 and/or Table 36.

141. The microbial cell of claim 140, wherein the UGTc3 enzyme comprises one or more mutations at one or more positions selected from L41, D49, T74, C127, F303, and A307 with respect to SEQ ID NO: 27, optionally wherein the mutations are selected from:

- (i) L41 is substituted with an aromatic amino acid selected from Phe, Tyr and Trp;
- (ii) D49 is substituted with Glu;
- (iii) T74 is substituted with Met;
- (iv) C127 is substituted with an aromatic amino acid selected from Phe, Tyr and Trp;
- (v) F303 is substituted with Tyr; and
- (vi) A307 is substituted with an amino acid selected from Ile, Leu, and Val.

142. The microbial cell of claim 141, wherein the UGTc3 enzyme comprises one or more of the following substitutions with respect to SEQ ID NO: 27: L41F, D49E, T74M, C127F, F303Y, and A307I.

143. The microbial cell of claim 142, wherein the UGTc3 enzyme comprises an amino acid sequence that is at least 70% identical, or at least 80%, or at least 85%, or at least 90%, or least 95%, or at least 97%, or at least 98%, or at least 99% identical to the amino acid sequence of SEQ ID NO: 60 or SEQ ID NO: 65, wherein the UGT enzyme optionally comprises one or more mutations with respect to SEQ ID NO: 60 or SEQ ID NO: 65 that result in greater specificity and/or activity for glycosylation of C-3 position compared to the UGT enzyme having an amino acid sequence of SEQ ID NO: 60 or SEQ ID NO: 65, optionally wherein one or more mutations are selected from Table 24 with respect to SEQ ID NO: 60, and/or Table 25 with respect to SEQ ID NO: 65.

144. The microbial cell of claim 143, wherein the UGTc3 enzyme is a circular permutant with respect to SEQ ID NO: 60.

145. The microbial cell of claim 144, wherein the circular permutant has a cut site corresponding to a position between amino acid 150 to 170, or amino acid 155 to 175, or amino acid 160 to 185, or amino acid 165 to 190, or amino acid 165 to 175 of SEQ ID NO: 60, and optionally having a linker sequence between the amino acids that correspond to the N-terminal and C-terminal residues of SEQ ID NO: 60, and wherein the linker optionally has a length of from 2 to 25 amino acids.

146. The microbial cell of claim 145, wherein the UGTc3 enzyme comprises one or more mutations at one or more positions selected from H40, I216, S241, and L352 with respect to SEQ ID NO: 65, optionally wherein the mutations are selected from:

- (i) H40 is substituted with Asp, Gly, Gln, Ala or Val;
- (ii) I216 is substituted with Ala, Gly, or Val;
- (iii) S241 is substituted with Gly, Ala, or Val; and
- (iv) L352 is substituted with Met.

147. The microbial cell of claim 146, wherein the UGTc3 enzyme comprises one or more of the following substitutions with respect to SEQ ID NO: 65: H40N, I216A, S241G, and L352M.

148. The microbial cell of any one of claims 135 to 147, wherein the biosynthetic pathway comprises a UGT enzyme that produces a C-24 mogrol glycoside from mogrol (UGTc24).

149. The microbial cell of claim 148, wherein the UGTc24 enzyme comprises an amino acid sequence that is at least 70% identical, or at least 80% identical, or at least 85% identical, or at least 90% identical, or at least 95% identical, or at least 97% identical, or at least 98% identical or at least 99% identical to the amino acid sequence of SEQ ID NO: 31, wherein the UGTc24 enzyme optionally comprises one or more mutations resulting in greater specificity and/or activity for glycosylation of C-24 position compared to the UGT enzyme having an amino acid sequence of SEQ ID NO: 31, optionally wherein one or more mutations are selected from Table 40, Table 41, Table 42, Table 43, Table 44 and/or Table 45.

150. The microbial cell of claim 149, wherein the UGTc24 enzyme comprises one or more mutations at positions selected from L18, A74, S88, I91, T95, H101, T184, P190, T273, N306, M332, and Y381, with respect to SEQ ID NO: 31, wherein one or more mutations are optionally selected from:

- (i) L18 is substituted with Met;
- (ii) I63 is substituted with Val, Ala or Gly;
- (iii) A74 is substituted with an acidic amino acid selected from Glu and Asp;
- (iv) S88 is substituted with an amino acid selected from Ala, Gly, Leu, Val, and Ile;
- (v) I91 is substituted with an aromatic amino acid selected from Phe, Tyr and Trp;
- (vi) T95 is substituted with an amino acid selected from Ala, Gly, and Val;
- (vii) H101 is substituted with Pro;
- (viii) T184 is substituted with an amino acid selected from Phe, Tyr, and Trp;
- (ix) F187 is substituted with Tyr or Trp
- (x) P190 is substituted with an amino acid selected from Glu and Asp;
- (xi) T273 is substituted with Ser or is deleted;
- (xii) N306 is substituted with Lys, His, or Arg;

- (xiii) M332 is substituted with an amino acid selected from Leu, Ile, Val, and Ala; and
- (xiv) Y381 is substituted with an amino acid selected from Phe and Trp.

151. The microbial cell of claim 150, wherein the UGTc24 enzyme comprises one or more mutations selected from L18M, I63V, A74E, S88A, I91F, T95A, H101P, T184F, F187Y, P190E, T273S, N306K, M332L, and Y381F with respect to SEQ ID NO: 31.

152. The microbial cell of any one of claims 135 to 151, wherein the biosynthetic pathway comprises a uridine diphosphate dependent glycosyltransferase (UGT) enzyme that produces a 1,6 and/or 1,2 branched glycoside of a C-3 and/or C-24 glycoside of mogrol (UGT1216).

153. The microbial cell of claim 152, wherein the UGT1216 is a circular permutant of the enzyme represented by SEQ ID NO: 35, or a sequence having at least 70% identity thereto.

154. The microbial cell of claim 153, wherein the circular permutant has a cut site at a position corresponding to amino acid 180 to 195 of SEQ ID NO: 35, or a sequence having at least 70% identity thereto, and

optionally at a position corresponding to 185 to 190 of SEQ ID NO: 35, or a sequence having at least 70% identity thereto, and

optionally at position 187 of SEQ ID NO: 35, or a sequence having at least 70% identity thereto.

155. The microbial cell of claim 154, wherein the UGT1216 enzyme comprises an amino acid sequence that is at least 70%, or at least 80%, or at least 90%, or at least 95%, or at least 97%, or at least 98%, or at least 99% identical to the amino acid sequence of SEQ ID NO: 36, wherein the UGT1216 enzyme has one or more mutations resulting in greater specificity and/or activity for to 1,6 and/or 1,2 glycosylation of a C-3 and/or C-24 glycoside of mogrol compared to the UGT enzyme having an amino acid sequence of SEQ ID NO: 36.

156. The microbial cell of claim 155, wherein the UGT1216 enzyme has one or more mutations resulting in greater production of MogV from MogIIE compared to the UGT enzyme having an

amino acid sequence of SEQ ID NO: 36, optionally wherein one or more mutations or sets of mutations of the UGT1216 are selected from Table 46, Table 47, Table 48, and/or Table 49.

157. The microbial cell of claim 155 or 156, wherein the UGT1216 enzyme comprises mutations at one or more selected from L14, H19, W160, L266, T281, S318, and H390 with respect to SEQ ID NO: 36, optionally wherein:

- (i) L14 is substituted with a non-polar amino acid selected from Val, Ile, Ala, and Gly;
- (ii) H19 is substituted with a charged amino acid selected from Glu, Lys, Asp and Glu;
- (iii) W160 is substituted with Pro, Ser, or Thr;
- (iv) L266 is substituted with a polar amino acid selected from Gln and Asn;
- (v) T281 is substituted with Ser;
- (vi) S318 is substituted with an amino acid selected from Lys, Pro, and Arg; and
- (iv) H390 is substituted with an amino acid selected from Asp and Glu.

158. The microbial cell of claim 157, wherein the UGT1216 enzyme comprises one or more substitutions selected from L14V, H19E, W160P, L266Q, T281S, S318K and H390D with respect to SEQ ID NO: 36.

159. The microbial cell of claim 158, wherein the UGT1216 enzyme comprises the mutations L14V, W160P, L266Q, and H390D substitutions with respect to SEQ ID NO: 36.

160. The microbial cell of claim 154, wherein the UGT1216 enzyme comprises an amino acid sequence that is at least 70%, or at least 80%, or at least 90%, or at least 95%, or at least 97%, or at least 98%, or at least 99% identical to the amino acid sequence of SEQ ID NO: 30, wherein the UGT1216 enzyme as one or more mutations resulting in greater specificity and/or activity for 1,6 and 1,2 branched glycosylation of a C-3 and/or C-24 glycoside of mogrol compared to the UGT enzyme having an amino acid sequence of SEQ ID NO: 39.

161. The microbial cell of claim 160, wherein the UGT1216 enzyme has one or more mutations resulting in greater production of MogIVA from MogIIE or MogV from MogIIE, compared to the UGT enzyme having an amino acid sequence of SEQ ID NO: 39.

162. The microbial cell of claim 161, wherein the UGT1216 enzyme comprises one or more mutations listed in Table 50.

163. The microbial cell of claim 161 or 162, wherein the UGT enzyme comprises mutation(s) at positions selected from T147 and N207 with respect to SEQ ID NO: 39, optionally wherein:

- (i) T147 is substituted with a non-polar amino acid selected from Leu, Ile, and Val; and/or
- (ii) N207 is substituted with a basic amino acid selected from Lys, Arg, and His.

164. The microbial cell of claim 163, wherein the UGT enzyme comprises T147L and/or N207K substitutions with respect to SEQ ID NO: 39.

165. The microbial cell of any one of claims 98 to 164, wherein the microbial cell expresses a squalene synthase (SQS) enzyme.

166. The microbial cell of claim 165, wherein the microbial cell expresses a farnesyl pyrophosphate (FPP) synthase.

167. The microbial cell of claim 165 or 166, wherein the microbial cell overexpresses one or more enzymes of an MEP or MVA pathway.

168. The microbial cell of claim 167, wherein the microbial cell is a bacterium, and is optionally selected from *Escherichia* spp., *Bacillus* spp., *Corynebacterium* spp., *Rhodobacter* spp., *Zymomonas* spp., *Vibrio* spp., and *Pseudomonas* spp., and is optionally a species selected from *Escherichia coli*, *Bacillus subtilis*, *Corynebacterium glutamicum*, *Rhodobacter capsulatus*, *Rhodobacter sphaeroides*, *Zymomonas mobilis*, *Vibrio natriegens*, and *Pseudomonas putida*.

169. The microbial cell of claim 168, wherein the microbial cell is a yeast cell, optionally selected from a species of *Saccharomyces*, *Pichia*, or *Yarrowia*, and is optionally *Saccharomyces cerevisiae*, *Pichia pastoris*, and *Yarrowia lipolytica*.

170. The microbial cell of any one of claims 98 to 169, wherein the microbial cell expresses or a recombinant *pntAB* or a variant thereof.

171. The microbial cell of any one of claims 98 to 170, wherein the microbial cell expresses at least one recombinant electron carrier.

172. The microbial cell of claim 171, wherein the at least one recombinant electron carrier is a recombinant ferredoxin (fdx) or recombinant flavodoxin (Fld).

173. The microbial cell of any one of claims 98 to 172, wherein when the microbial cell is cultured under conditions suitable for producing the mogrol or mogroside, the microbial cell produces more than about 40%, or more than about 50%, or more than about 60%, or more than about 70%, or more than about 80% on pathway products 2,3-oxidosqualene, 2,3;22,23-dioxidosqualene, 24,25-epoxy cucurbitadienol, 24,25-dihydroxy cucurbitadienol, mogrol, and mogrosides.

174. The microbial cell of any one of claims 98 to 173, wherein when the microbial cell is cultured under conditions suitable for producing the mogrol or mogroside, the microbial cell produces less than about 35%, or less than about 30%, or less than about 25%, or less than about 20% of pathway products cucurbitadienol, hydrolyzed 2,3;22,23-dioxidosqualene, and glycosylated 24,25-dihydroxycucurbitadienol.

175. The microbial cell of any one of claims 98 to 174, wherein the microbial cell comprises a mutation that reduces the amount or activity of a transporter.

176. The microbial cell of claim 175, wherein the transporter is an ABC transporter.

177. The microbial cell of claim 175 or claim 176, wherein the transporter is YbhGFSR or an ortholog thereof.

178. The microbial cell of claim 177, wherein the microbial cell comprises a hypomorphic mutation or a null mutation in one or more of *ybhG*, *ybhF*, *ybhS*, and *ybhR*.

179. The microbial cell of claim 178, wherein the null mutation is a complete or partial deletion of one or more of *ybhG*, *ybhF*, *ybhS*, and *ybhR*.

180. The microbial cell of claim 178, wherein the hypomorphic mutation is a promoter mutation, and/or a mutation selected from a mutant ribosome binding site, one or more mutations adding one or more rare codons, and adding a secondary RNA structure in one or more of *ybhG*, *ybhF*, *ybhS*, and *ybhR*.

181. A microbial cell for producing mogrol or mogroside expressing a biosynthetic pathway that converts squalene to the mogrol or mogroside,

wherein the microbial cell comprises one or more genetic modifications that reduces the efflux of an intermediate of mogrol synthesis, optionally wherein the intermediate is 2,3;22,23-dioxidosqualene,

wherein the microbial cell exhibits an increase in mogrol synthesis compared to a cell that does not have the genetic modifications.

182. The microbial cell of claim 181, wherein the genetic modification comprises a mutation that reduces the amount or activity of a transporter.

183. The microbial cell of claim 181 or claim 182, wherein the transporter is YbhGFSR or an ortholog thereof.

184. The microbial cell of claim 183, wherein the microbial cell comprises a hypomorphic mutation or a null mutation in one or more of *ybhG*, *ybhF*, *ybhS*, and *ybhR*.

185. The microbial cell of claim 184, wherein the null mutation is a complete or partial deletion of one or more of *ybhG*, *ybhF*, *ybhS*, and *ybhR*.

186. The microbial cell of claim 184, wherein the hypomorphic mutation is a promoter mutation, and/or a mutation selected from a mutant ribosome binding site, one or more mutations adding one or more rare codons, and adding a secondary RNA structure in one or more of *ybhG*, *ybhF*, *ybhS*, and *ybhR*.

187. The microbial cell of any one of claims 181 to 186, wherein the microbial cell is *E. coli*.

188. The microbial cell of any one of claims 181 to 187, wherein the biosynthetic pathway comprises one or more enzymes disclosed herein.

189. A squalene epoxidase comprising an amino acid sequence that is at least 80%, or at least 85%, at least 90%, or at least 95% identical to the amino acid sequence of SEQ ID NO: 2, and comprising one or more mutations or sets of mutations listed in Tables 1, 2, and 3.

190. The squalene epoxidase of claim 189, comprising at least two, or at least three, or at least four, or at least five, or at least six, or more mutations or sets of mutations listed in Table 1, Table 2, and/or Table 3 with respect to the amino acid sequence of SEQ ID NO: 2.

191. The squalene epoxidase of claim 189, wherein the squalene epoxidase comprises one or more substitutions with respect to the amino acid sequence of SEQ ID NO: 2 selected from:

- (i) S116 is substituted with a charged amino acid selected from Glu and Asp;
- (ii) N134 is substituted with a non-polar amino acid selected from Gly, Ala, Ile, Leu, and Val;
- (iii) I145 is substituted with a charged amino acid selected from His, Arg, and Lys;
- (iv) A164 is substituted with a polar amino acid selected from Asn and Gln.
- (v) A169 is substituted with a charged amino acid selected from Arg, Lys, Glu, Asp, or His; and/or
- (vi) A400 is substituted with a non-polar amino acid selected from Val, Leu, and Ile.

192. The squalene epoxidase of claim 191, wherein the squalene epoxidase comprises the substitutions S116E, N134G, I145H, A164N, A169R, and A400V with respect to SEQ ID NO: 2.

193. A squalene epoxidase comprising an amino acid sequence that is at least 80%, or at least 85%, at least 90%, or at least 95% identical to the amino acid sequence of SEQ ID NO: 77, and comprising one or more mutations or one or more sets of mutations listed in Tables 4 and/or 5.

194. The squalene epoxidase of claim 193, wherein the squalene epoxidase comprises one or more substitutions at positions selected from with respect to SEQ ID NO: 77, optionally wherein:

- (i) S366 is substituted with a charged amino acid selected from Arg, Lys and His;
- (ii) D41 is substituted with a positively charged amino acid selected from His, Lys and Arg;
- (iii) M315 is substituted with a non-polar amino acid selected from Leu, Ile, and Val;

195. The squalene epoxidase of claim 194, wherein the squalene epoxidase comprises one or more mutations selected from S366R, D41H, and M315L with respect to SEQ ID NO: 77.

196. A 24, 25 epoxy cucurbitadienol synthase (ECDS) comprising an amino acid sequence that is at least 80%, or at least 85%, at least 90%, or at least 95% identical to the amino acid sequence of SEQ ID NO: 7, and comprising one or more mutations or sets of mutations listed in Tables 7, 8, 9 and 10.

197. The ECDS of claim 196, comprising at least two, or at least three, or at least four, or at least five, or at least six, or at least seven, or at least eight, or at least nine, or at least ten, or more mutations or sets of mutations listed in Table 7, Table 8, Table 9 and/or Table 10 with respect to SEQ ID NO: 7.

198. The ECDS of claim 197, wherein the ECDS comprises one or more substitutions selected from:

- (i) S24 is substituted with a polar amino acid selected from Asn and Gln;
- (ii) C35 is substituted with a charged amino acid selected from Asp and Glu;
- (iii) D50 is substituted with Glu;
- (iv) N121 is substituted with His, Lys and Arg,

- (v) L245 is substituted with Ile or Val;
- (vi) W331 is substituted with Ser, Thr, or Tyr;
- (vii) Q401 is substituted with Ala, Gly, or Val;
- (viii) I490 is substituted with Val, Leu, or Ala;
- (ix) I553 is substituted with Met;

and/or

- (x) A556 is substituted with a polar amino acid selected from Ser, Thr, and Tyr.

199. The ECDS of claim 198, wherein the ECDS comprises the substitutions S24N, C35D, D50E, N121H, L245I, W331S, Q401A, I490V, I553M, and A556S with respect to SEQ ID NO: 7.

200. A epoxide hydrolase comprising an amino acid sequence that is at least 80%, or at least 85%, at least 90%, or at least 95% identical to the amino acid sequence of SEQ ID NO: 5, and comprising one or more mutations or sets of mutations listed in Table 6.

201. The epoxide hydrolase of claim 200, comprising at least two, or at least three, or at least four, or at least five, or at least six, or at least seven, or at least eight, or at least nine, or at least ten, or more mutations or sets of mutations listed in Table 6 respect to SEQ ID NO: 5.

202. The epoxide hydrolase of claim 201, wherein the epoxide hydrolase comprises one or more substitutions selected from:

- (i) S68 is substituted with a non-polar amino acid selected from Ala, Gly, and Val;
- (ii) Q128 is substituted with a non-polar amino acid selected from Ala, Pro, Gly and Val;
- (iii) G141 is substituted with a polar amino acid selected from Ser, Thr, and Tyr;
- (iv) T144 is substituted with an amino acid selected from Gln, Asn, Ala, Gly, and Val;
- (v) A145 is substituted with a non-polar amino acid selected from Val, Ile, and Leu;
- (vi) E163 is substituted with a non-polar amino acid selected from Ala, Gly, and Val;
- (vii) E191 is substituted with a non-polar amino acid selected from Gly, Ala, and Val;
- (viii) L262 is substituted with Met;

(ix) C295 is substituted with a non-polar amino acid selected from Gly, Ala, and Val; and/or

(x) N299 is substituted with Gln.

203. The epoxide hydrolase of claim 202, wherein the epoxide hydrolase comprises the substitutions S68A, Q128A, G141S, T144Q, A145V, E163A, E191G, L262M, C295G, and N299Q with respect to SEQ ID NO: 5.

204. A cytochrome P450 comprising an amino acid sequence that is at least 80%, or at least 85%, at least 90%, or at least 95% identical to the amino acid sequence of SEQ ID NO: 14 or 15, and comprising one or more mutations or sets of mutations listed in Tables 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, and/or 29.

205. The cytochrome P450 of claim 204, wherein the cytochrome P450 enzyme has a replacement of all or a portion of the native transmembrane domain with the transmembrane domain from an *E. coli* inner membrane cytoplasmic C-terminus protein, wherein the *E. coli* protein is optionally SohB or FliO, optionally wherein about 15 to about 20 amino acids of the cytochrome P450 enzyme are replaced with about 25 to about 30 amino acids of SohB or about 36 to about 42 amino acids of FliO.

206. The cytochrome P450 of claim 205, comprising at least two, or at least three, or at least four, or at least five, or at least six, or at least seven, or at least eight, or at least nine, or at least ten, or more mutations or sets of mutations listed in Table 10, Table 11, Table 12, Table 13, Table 14, Table 15, Table 16, and/or Table 17.

207. The cytochrome P450 of claim 206, wherein the cytochrome P450 enzyme comprises a substitution at one or more positions selected from A12, I27, S63, K75, V76, N102, G125, T128, W130, L131, K132, T192, S193, L218, G219, T223, V228, T231, T232, Y233, N234, K246, F354, K369, H433, F452, V465, A468, and T482, with respect to SEQ ID NO: 15, wherein the one or more substitutions are optionally selected from:

(i) A12 is substituted with Trp, Tyr or Phe;

- (ii) I27 is substituted with Gly or Ala;
- (iii) S63 is substituted with Asn, Phe, Gln, Tyr, or Trp;
- (iv) K75 is substituted with Arg;
- (v) V76 is substituted with Met, Leu or Ile;
- (vi) N102 is substituted with His, Lys, or Arg;
- (vii) G125 is substituted with an amino acid selected from Ala, Leu, Ile, and Val;
- (viii) T128 is substituted with an amino acid selected from Gly, Ser, Pro, and Ala;
- (ix) W130 is substituted with an amino acid selected from Asn, Gly, Ser, Leu, Gln, Ala, Val, and Thr;
- (x) L131 is substituted with an amino acid selected from Val, Ile, Pro, Thr, and Ser;
- (xi) K132 is substituted with an amino acid selected from Asn, Ser, Thr, and Gln;
- (xii) T192 is substituted with Leu, Ile, or Val;
- (xiii) S193 is substituted with Thr;
- (xiv) L218 is substituted with Ile, Val, and Ala;
- (xv) G219 is substituted with an amino acid selected from Ala, Gln, Arg, Val, Asn, His, and Lys;
- (xvi) T223 is substituted with Ser;
- (xvii) V228 is substituted with Ile or Leu;
- (xviii) T231 is substituted with Phe, Trp, or Tyr;
- (xix) T232 is substituted with Ala, Gly, or Val;
- (xx) Y233 is substituted with Phe;
- (xxi) N234 is substituted with His, Arg or Lys;
- (xxii) K246 is substituted with Val, Ile or Leu;
- (xxiii) F354 is substituted with Tyr or Trp;
- (xxiv) K369 is substituted with Arg, Lys, or His;
- (xxv) H433 is substituted with Phe, Trp, or Tyr;
- (xxvi) F452 is substituted with Ile, Val, or Leu;
- (xxvii) V465 is substituted with an amino acid selected from Ile, Met, or Leu;
- (xxviii) A468 is substituted with an amino acid selected from Thr, Ser, Asn, and Gln; and
- (xxix) T482 is substituted with Ser.

208. The cytochrome P450 of claim 207, wherein the cytochrome P450 comprises the substitutions A12W, I27G, S63F, S63N, K75R, V76M, N102H, G125A, T128G, W130N, L131I, K132N, T192L, S193T, L218I, G219A, T223S, V228I, T231F, T232A, Y233F, N234H, K246V, F354Y, K369R, H433F, F452I, V465I, A468T and T482S with respect to SEQ ID NO: 14 or 15.

209. A cytochrome P450 enzyme comprising an N-terminal deletion of at least about 20, or at least about 25, or at least about 30, or at least about 35, or at least about 40, or at least about 45, or at least about 50 amino acids with respect to SEQ ID NO: 15.

210. The cytochrome P450 enzyme of claim 209, wherein the cytochrome P450 enzyme comprises a deletion selected from a deletion of L3 to H29 with respect to SEQ ID NO: 15.

211. The cytochrome P450 enzyme of claim 210, wherein the cytochrome P450 enzyme comprises an amino acid sequence that is at least 70% identical, or at least 80% identical, or at least 90% identical, or at least 95% identical, or at least 97% identical, or at least 98% identical, or at least 99% identical to the amino acid sequence of SEQ ID NO: 85, and comprising one or more mutations or one or more sets of mutations listed in Tables 26, 27, 28, and/or 29.

212. The cytochrome P450 enzyme of claim 211, wherein the cytochrome P450 enzyme comprises a substitution at one or more positions selected K8, D9, S10, N13, V15, I45, K46, K47, M49, K50, R51, I191, A192, F406, E411, Y412, and S413, with respect to SEQ ID NO: 85, wherein the one or more substitutions are optionally selected from:

- (i) K8 is substituted with Arg;
- (ii) D9 is substituted with Asn or Gln;
- (iii) S10 is substituted with Arg, Lys or His;
- (iv) N13 is substituted with Lys, Arg, or His;
- (v) V15 is substituted with Lys, Arg, or His;
- (vi) I45 is substituted with Val, Ile, or Leu;
- (vii) K46 is substituted with Arg, Lys or His;
- (viii) K47 is substituted with Glu or Asp;
- (ix) M49 is substituted with Val, Ile, or Leu;

- (x) K50 is substituted with Glu or Asp;
- (xi) R51 is substituted with Lys;
- (xii) I191 is substituted with Leu or Val;
- (xiii) A192 is substituted with Lys, Arg, or His;
- (xiv) F406 is substituted with Tyr or Trp;
- (xv) E411 is substituted with Asp;
- (xvi) Y412 is substituted with Phe, Leu, Ile, or Trp; and
- (xvii) S413 is substituted with Ala, Gly, or Val.

213. The cytochrome P450 enzyme of claim 212, wherein the cytochrome P450 enzyme comprises one or more substitutions selected from K8R, D9N, S10R, N13K, and V15K I45V, K46R, K47E, M49V, K50E, R51K, F406Y, E411D, Y412F, S413A, I191L and A192K substitutions with respect to SEQ ID NO: 85.

214. A cytochrome P450 reductase comprising an amino acid sequence that is at least 80%, or at least 85%, at least 90%, or at least 95% identical to the amino acid sequence of SEQ ID NO: 22 or SEQ ID NO: 24, and comprising one or more mutations or sets of mutations listed in Tables 19 with respect to SEQ ID NO: 22, or Tables 20 and/or 21 with respect to SEQ ID NO: 24.

215. The cytochrome P450 reductase of claim 214, wherein the cytochrome P450 reductase enzyme comprises a replacement all or a portion of the native transmembrane domain with the transmembrane domain from an *E. coli* inner membrane cytoplasmic C-terminus protein, wherein the *E. coli* protein is selected from YcgG, yhcB, zipA, waaA, sohB, djIA, lpxK, ypfN, and yhhM.

216. The cytochrome P450 reductase of claim 215, comprising at least two, or at least three, or at least four, or at least five, or at least six, or at least seven, or at least eight, or at least nine, or at least ten, or more mutations or sets of mutations listed in Table 18, Table 19, Table 20, and Table 21.

217. The cytochrome P450 reductase of claim 216, wherein the cytochrome P450 reductase comprises a substitution of L90 residue with respect to SEQ ID NO: 24.

218. The cytochrome P450 reductase of claim 217, wherein the cytochrome P450 reductase comprises the substitution L90P with respect to SEQ ID NO: 24.

219. A uridine diphosphate dependent glycosyltransferase enzyme capable of producing a C-3 mogrol glycoside from mogrol (UGTc3) comprising an amino acid sequence that is at least 80%, or at least 85%, at least 90%, or at least 95% identical to the amino acid sequence of SEQ ID NO: 27, and comprising one or more mutations or sets of mutations listed in Tables 23, and/or 24.

220. The UGTc3 enzyme of claim 219, comprising at least two, or at least three, or at least four, or at least five, or at least six, or at least seven, or at least eight, or at least nine, or at least ten, or more mutations or sets of mutations listed in Table 35 and/or Table 36.

221. The UGTc3 enzyme of claim 220, wherein the UGTc3 enzyme comprises one or more substitutions with respect to SEQ ID NO: 27 selected from:

- (i) L41 is substituted with an aromatic amino acid selected from Phe, Tyr and Trp;
- (ii) D49 is substituted with Glu;
- (iii) T74 is substituted with Met;
- (iv) C127 is substituted with an aromatic amino acid selected from Phe, Tyr and Trp;
- (v) F303 is substituted with Tyr; and
- (vi) A307 is substituted with an amino acid selected from Ile, Leu, and Val.

222. The UGTc3 enzyme of claim 221, wherein the UGTc3 enzyme comprises the substitutions L41F, D49E, T74M, C127F, F303Y, and A307I with respect to SEQ ID NO: 27.

223. A uridine diphosphate dependent glycosyltransferase enzyme capable of producing a C-3 mogrol glycoside from mogrol (UGTc3) comprising an amino acid sequence that is at least 80%, or at least 85%, at least 90%, or at least 95% identical to the amino acid sequence of SEQ ID NO: 65, SEQ ID NO: 66, or SEQ ID NO: 67.

224. A uridine diphosphate dependent glycosyltransferase enzyme capable of producing a C-3 mogrol glycoside from mogrol (UGTc3) comprising an amino acid sequence that is at least 80%, or at least 85%, at least 90%, or at least 95% identical to the amino acid sequence of SEQ ID NO: 60, and comprising one or more mutations or sets of mutations listed in Table 37 with respect to SEQ ID NO: 60, and/or Table 39 with respect to SEQ ID NO: 65.

225. A uridine diphosphate dependent glycosyltransferase enzyme capable of producing a C-3 mogrol glycoside from mogrol (UGTc3), wherein the UGTc3 enzyme is a circular permutant with respect to SEQ ID NO: 60.

226. The uridine diphosphate dependent glycosyltransferase enzyme of claim 225, wherein the circular permutant has a cut site corresponding to a position between amino acid 150 to 170, or amino acid 155 to 175, or amino acid 160 to 185, or amino acid 165 to 190, or amino acid 165 to 175 of SEQ ID NO: 60, and optionally having a linker sequence between the amino acids that correspond to the N-terminal and C-terminal residues of SEQ ID NO: 60, and wherein the linker optionally has a length of from 2 to 25 amino acids.

227. The uridine diphosphate dependent glycosyltransferase enzyme of claim 225, wherein the UGTc3 enzyme comprises one or more mutations at one or more positions selected from H40, I216, S241, and L352 with respect to SEQ ID NO: 65, optionally wherein the mutations are selected from:

- (i) H40 is substituted with Asp, Gly, Gln, Ala or Val;
- (ii) I216 is substituted with Ala, Gly, or Val;
- (iii) S241 is substituted with Gly, Ala, or Val; and
- (iv) L352 is substituted with Met.

228. The uridine diphosphate dependent glycosyltransferase enzyme of claim 227, wherein the UGTc3 enzyme comprises one or more of the following substitutions with respect to SEQ ID NO: 65: H40N, I216A, S241G, and L352M.

229. A uridine diphosphate dependent glycosyltransferase enzyme capable of producing a C-24 mogrol glycoside from mogrol (UGTc24) comprising an amino acid sequence that is at least 80%, or at least 85%, at least 90%, or at least 95% identical to the amino acid sequence of SEQ ID NO: 31, and comprising one or more mutations or sets of mutations listed in Tables 27, 28, 29, 30, 31, and 32.

230. The UGTc24 enzyme of claim 229, comprising at least two, or at least three, or at least four, or at least five, or at least six, or at least seven, or at least eight, or at least nine, or at least ten, or at least eleven, or at least twelve, or at least thirteen or more mutations or sets of mutations listed in Table 40, Table 41, Table 42, Table 43, Table 44 and/or Table 45.

231. The UGTc24 enzyme of claim 230, wherein the UGTc24 enzyme comprises one or more substitutions selected from:

- (i) L18 is substituted with Met;
- (ii) I63 is substituted with Val, Ala or Gly;
- (iii) A74 is substituted with an acidic amino acid selected from Glu and Asp;
- (iv) S88 is substituted with an amino acid selected from Ala, Gly, Leu, Val, and Ile;
- (v) I91 is substituted with an aromatic amino acid selected from Phe, Tyr and Trp;
- (vi) T95 is substituted with an amino acid selected from Ala, Gly, and Val;
- (vii) H101 is substituted with Pro;
- (viii) T184 is substituted with an amino acid selected from Phe, Tyr, and Trp;
- (ix) F187 is substituted with Tyr or Trp
- (x) P190 is substituted with an amino acid selected from Glu and Asp;
- (xi) T273 is substituted with Ser or is deleted;
- (xii) N306 is substituted with Lys, His, or Arg;
- (xiii) M332 is substituted with an amino acid selected from Leu, Ile, Val, and Ala; and
- (xiv) Y381 is substituted with an amino acid selected from Phe and Trp.

232. The UGTc24 enzyme of claim 231, wherein the UGTc24 enzyme comprises the substitutions L18M, A74E, S88A, I91F, T95A, H101P, T184F, P190E, T273S, N306K, M332L, and Y381F with respect to SEQ ID NO: 31.

233. A uridine diphosphate dependent glycosyltransferase enzyme capable of producing a 1,6 and/or 1,2 branched glycoside of a C-3 and/or C-24 glycoside of mogrol (UGT1216) comprising an amino acid sequence that is at least 80%, or at least 85%, at least 90%, or at least 95% identical to the amino acid sequence of SEQ ID NO: 35 or SEQ ID NO: 36, and comprising one or more mutations or sets of mutations listed in Tables 33 and/or 34.

234. The UGT1216 of claim 233, wherein the UGT1216 is a circular permutant of the enzyme represented by SEQ ID NO: 35, or a sequence having at least 70% identity thereto.

235. The UGT1216 of claim 234, wherein the circular permutant has a cut site at a position corresponding to amino acid 180 to 195 of SEQ ID NO: 35, or a sequence having at least 70% identity thereto, and

optionally at a position corresponding to 185 to 190 of SEQ ID NO: 35, or a sequence having at least 70% identity thereto, and

optionally at position 187 of SEQ ID NO: 35, or a sequence having at least 70% identity thereto.

236. The UGT1216 enzyme of claim 235, comprising at least two, or at least three, or at least four, or more mutations or sets of mutations listed in Table 33, and/or Table 34.

237. The UGT1216 enzyme of claim 236, wherein the UGT1216 enzyme comprises one or more substitutions selected from:

- (i) L14 is substituted with a non-polar amino acid selected from Val, Ile, Ala, and Gly;
- (ii) W160 is substituted with Pro, Ser, or Thr;
- (iii) L266 is substituted with a polar amino acid selected from Gln and Asn; and
- (iv) H390 is substituted with an amino acid selected from Asp and Glu.

238. The UGT1216 enzyme of claim 237, wherein the UGT1216 enzyme comprises the substitutions L14V, W160P, L266Q, and H390D with respect to SEQ ID NO: 36.

239. A uridine diphosphate dependent glycosyltransferase enzyme capable of producing a 1,6 and/or 1,2 branched glycoside of a C-3 and/or C-24 glycoside of mogrol (UGT1216) comprising an amino acid sequence that is at least 80%, or at least 85%, at least 90%, or at least 95% identical to the amino acid sequence of SEQ ID NO: 39, and comprising one or more mutations or sets of mutations listed in Table 50.

240. The UGT1216 enzyme of claim 239, comprising at least two, more mutations or sets of mutations listed in Table 50.

241. The UGT1216 enzyme of claim 240, wherein the UGT1216 enzyme comprises one or more substitutions selected from:

- (i) T147 is substituted with a non-polar amino acid selected from Leu, Ile, and Val; and/or
- (ii) N207 is substituted with a basic amino acid selected from Lys, Arg, and His.

242. The UGT1216 enzyme of claim 241, wherein the UGT1216 enzyme comprises the substitutions T147L and/or N207K substitutions with respect to SEQ ID NO: 39.

243. A method of producing a mogroside comprising C3 glucosylated mogrol, the method comprising contacting mogrol with a uridine diphosphate dependent glycosyltransferase enzyme (UGT) enzyme of any one of claims 233 to 236.

244. A method of producing a mogroside comprising one or more of C24 glucosylated mogrol, the method comprising contacting mogrol with a uridine diphosphate dependent glycosyltransferase enzyme (UGT) enzyme of any one of claims 206 to 208.

227. A method of producing a mogroside comprising a 1,6 and 1,2 glucosylation of a C3- or C24-glucosylated mogrol, the method comprising contacting mogrol, C3- glucosylated mogrol or C24-glucosylated mogrol with a uridine diphosphate dependent glycosyltransferase enzyme (UGT) enzyme any one of claims 209 to 218.