

METHODS AND MATERIALS FOR BIOSYNTHESIS OF MOGROSIDE COMPOUNDS

Field of invention

5

The present invention relates to methods and materials for biosynthesis of mogroside compounds, and more particularly to methods involving use of cytochrome P450 enzymes to produce mogrol and/or using uridine-5'-diphospho (UDP) dependent glucosyltransferases (UGTs) to glycosylate mogrol and produce various mogrol glycosides (mogrosides). The methods may also involve use of enzymes involved in biosynthesis of substrates for mogrol production.

10

Background of invention

15

Mogrosides are a family of triterpene glycosides isolated from fruits of *Siraitia grosvenorii* (Swingle), also known as *Momordica grosvenori* (Swingle). Extracts of the fruits are commercially used as natural sweeteners. Four major compounds, Mogroside V, Mogroside IV, Siamenoside I, and 11-Oxomogroside V, have been identified from the fruits of *Siraitia grosvenorii* (Swingle) that are responsible for the sweetness of the fruits (see FIG. 1). Mogroside V is the most abundant of these four compounds at approximately 0.57% (w/w) of the dry fruit, followed by Mogroside IV and Siamenoside I, each of which contain four glucose moieties. 11-Oxomogroside V has a ketone group instead of a hydroxyl at C-11. See, e.g., Takemoto, *et al.*, *Yakugaku Zasshi*, 103, 1151-1154; 1155-1166; 1167-1173, (1983); Kasai, *et al.*, *Agric. Biol. Chem.* 53, 3347-3349 (1989); Matsumoto, *Chem. Pharm. Bull.* 38, 2030-2032 (1990); and Prakash, *et al.*, *J. Carbohydrate Chem.* 30, 16-26 (2011). However, the enzymes responsible for producing mogrosides have not been identified.

20

25

Tang et al. BMC Genomics 2011, 12:343 describes seven CYP450s and five UDPGs as potential candidates involved in mogroside biosynthesis. However, the document does not specifically identify any CYPs or UDPGs involved in mogroside biosynthesis.

30

Summary of invention

The present invention provides methods and materials for biosynthesis of mogroside compounds. Interestingly, the invention provides enzymes involved in mogroside biosynthesis.

5 Mogroside biosynthesis may involve several steps, and accordingly it is an aspect of the present invention to provide enzymes capable of catalysing each of these steps. It is however also foreseen that the methods may involve performing only some of the steps enzymatically, whereas others may be performed by other means.

10 In one aspect, this document features a method of producing a mogroside compound.

Thus, the invention provides a method of producing a mogroside, wherein the method comprises one or more of the following steps:

- 15 Step Ia. Enhancing levels of oxido-squalene
 Step Ib. Enhancing levels of dioxido-squalene
 Step IIa. Oxido-squalene -> cucurbitadienol
 Step IIb. Dioxido-squalene -> 24,25 epoxy cucurbitadienol
 Step IIIa. Cucurbitadienol -> 11-hydroxy-cucurbitadienol
20 Step IIIb. 24,25 epoxy cucurbitadienol -> 11-hydroxy-24,25 epoxy cucurbitadienol
 Step IVa. 11-hydroxy-cucurbitadienol -> mogrol
 Step IVb. 11-hydroxy-24,25 epoxy cucurbitadienol -> mogrol
 Step V mogrol -> mogroside

25 Methods for performing each of the above-mentioned steps are described herein below. In particular, enzymes or mixture of enzymes useful for each of above-mentioned steps are described in details herein below.

The invention also features a recombinant host comprising one or more of the following
30 heterologous nucleic acids:

- IIa. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes
 capable of catalysing Step IIa (Oxido-squalene -> cucurbitadienol)
 IIb. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes
 capable of catalysing Step IIb (Dioxido-squalene -> 24,25 epoxy
35 cucurbitadienol)

- 5
10
15
- IIIa. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IIIa (Cucurbitadienol -> 11-hydroxy-cucurbitadienol)
 - IIIb. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IIIb (24,25 epoxy cucurbitadienol -> 11-hydroxy-24,25 epoxy cucurbitadienol)
 - IVa. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IVa (11-hydroxy-cucurbitadienol -> mogrol)
 - IVb. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IVb (11-hydroxy-24,25 epoxy cucurbitadienol -> mogrol)
 - V Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step V (mogrol -> mogroside)

15 In addition to the heterologous nucleic acids, said recombinant host may have been modified to achieve Step Ia and/or Step Ib.

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

30

Description of Drawings

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

35 FIG. 2 is a schematic of the pathway for the production of mogrosides from glucose.

FIG. 3 is a schematic of the production of mogrol glycosides (mogrosides) from squalene.

FIG. 4 is a schematic of the pathway proposed herein by the inventors (top) and published (bottom) of a P450 pathway for formation of mogrol from cucurbitadienol.

5 FIG. 5 is a depiction of the biosynthesis of mogroside I E1, mogroside I A1, mogroside II E, mogroside III A2, mogroside III, mogroside IV, and mogroside V from mogrol using UGTs.

FIG. 6 is a schematic of the products obtained from mogroside V after incubation with a pectinase and/or a cellulase.

10 FIG. 7 shows the LC-MS mass peak 501 corresponding to the proton plus Na⁺ adduct of tetrahydroxysqualene in a sample from yeast strain EFSC3027 transformed with a plasmid expressing *S. grosvenorii* Epoxide hydrolase 2.

FIG.8 shows the LC-MS chromatogram peak of lanosterol in yeast strain (upper panel) and LC-MS chromatogram peaks of cucurbitadienol and lanosterol in yeast strain
15 EFSC3498, which expresses cucurbitadienol synthase (lower panel).

FIG.9 shows the LC-MS chromatogram with the three peaks made when CYP5491 and CPR4497 are expressed in yeast strain EFSC3498 (upper panel), while the three lower panels show the fragmentation spectrum of these three peaks. The masses of the 3 peaks (443.38, 441.37 and 457.36) correspond in weight to proton adducts of
20 hydroxylated cucurbitadienol, oxo cucurbitadienol and hydroxy plus oxo cucurbitadienol respectively.

FIG.10A shows a route from oxido-squalene to mogrol and 11-oxo-mogrol proposed by the present invention.

25 FIG.10B shows a route from dioxido-squalene to mogrol and 11-oxo-mogrol proposed by the present invention.

FIG.11A shows the LC-MS chromatogram of reference mogroside I A1, while FIG. 11B shows the LC-MS chromatogram of a sample of yeast strain EFSC1563 expressing UGT1576 in a culture fed 50uM mogrol.

30 FIG. 12A shows the LC-MS chromatograms of samples from yeast strain EFSC1563 co-expressing UGT SK98 with UGT1576 showing production of di-glycosylated mogrol (mogroside II A). FIG. 12B shows LC-MS chromatograms of samples from yeast strain EFSC1563 co-expressing UGT98 with UGT1576 showing production of di and tri-glycosylated mogrol (middle and lower frames).

35 FIG.13 shows a route from mogrol to Mogroside III A1 proposed by the present invention.

Detailed description of the invention

5 Method of producing a mogroside

This document is based on the invention that recombinant hosts such as microorganisms, plant cells, or plants can be developed that express polypeptides useful for the biosynthesis of mogrol (the triterpene core) and various mogrol
10 glycosides (mogrosides). The aglycone mogrol is glycosylated with different numbers of glucose moieties to form various mogroside compounds. Recombinant microorganisms are particularly useful hosts. The recombinant host may be any of the recombinant hosts described herein below in the section "Recombinant host".

Expression of these biosynthetic polypeptides in various microbial chassis allows
15 mogrol and its glycosides to be produced in a consistent, reproducible manner from energy and carbon sources such as sugars, glycerol, CO₂, H₂ and sunlight. FIG. 2 provides a schematic of the pathway for production of mogrol and various mogrosides from glucose.

20 It is one aspect of the invention to provide a method of producing a mogroside, wherein the method comprises one or more of the following steps:

- Step Ia. Enhancing levels of oxido-squalene
- Step Ib. Enhancing levels of dioxido-squalene
- Step IIa. Oxido-squalene -> cucurbitadienol
- 25 Step IIb. Dioxido-squalene -> 24,25 epoxy cucurbitadienol
- Step IIIa. Cucurbitadienol -> 11-hydroxy-cucurbitadienol
- Step IIIb. 24,25 epoxy cucurbitadienol -> 11-hydroxy-24,25 epoxy cucurbitadienol
- Step IVa. 11-hydroxy-cucurbitadienol -> mogrol
- Step IVb. 11-hydroxy-24,25 epoxy cucurbitadienol -> mogrol
- 30 Step V mogrol -> mogroside

Methods and materials for performing each of the steps are described in more detail herein below. Each of the steps of the method results in generation of a product. Said products may also be referred to as "intermediate products" herein. Each step uses a
35 substrate, which may also be referred to as "precursor molecules". It is clear from

above that the intermediate products also may serve as precursor molecules for a subsequent step.

Thus, the invention provides methods of producing mogrosides, wherein the method may comprise the steps of

5

Step Ia. Enhancing levels of oxido-squalene

Step IIa. Oxido-squalene -> cucurbitadienol

Step IIIa. Cucurbitadienol -> 11-hydroxy-cucurbitadienol

Step IVa. 11-hydroxy-cucurbitadienol -> mogrol

10

Step V mogrol -> mogroside

and optionally isolating said mogroside.

The invention also provides methods of producing mogrosides, wherein the method may comprise the steps of

15

a) Providing oxido-squalene

b) Performing Steps IIa, IIIa, IVa and V identified above

c) optionally isolating said mogroside

The invention also provides methods of producing mogrosides, wherein the method may comprise the steps of

20

Step Ib. Enhancing levels of dioxido-squalene

Step IIb. Dioxido-squalene -> 24,25 epoxy cucurbitadienol

Step IIIb. 24,25 epoxy cucurbitadienol -> 11-hydroxy-24,25 epoxy cucurbitadienol

25

Step IVb 11-hydroxy-24,25 epoxy cucurbitadienol -> mogrol

Step V mogrol -> mogroside

and optionally isolating said mogroside.

The invention also provides methods of producing mogrosides, wherein the method may comprise the steps of

30

a) providing dioxido-squalene

b) performing steps IIb, IIIb, IVb and V identified above

c) optionally isolating said mogroside

The invention also provides methods of producing mogrosides, wherein said mogroside may be a higher glycosylated mogroside, wherein the method may comprise the steps of

- a) providing cucurbitadienol
- 5 b) performing steps IIIa, IVa and V identified above
- c) optionally isolating said mogroside

The invention also provides methods of producing mogrosides, wherein said mogroside may be a higher glycosylated mogroside, wherein the method may comprise the steps of

- 10 a) providing 24,25 epoxy cucurbitadienol
- b) performing steps IIIb, IVb and V identified above
- c) optionally isolating said mogroside

15 The invention provides methods of producing mogrosides, wherein the method may comprise the steps of

- a) providing mogrol
- b) performing step V identified above
- 20 c) optionally isolating said mogroside

20

The invention provides methods of producing mogrol, wherein the method may comprise the steps of

- a) providing dioxido-squalene
- b) performing steps IIb, IIb and IVb identified above
- 25 c) optionally isolating said mogrol

25

In general, the method may be performed either in vitro or in vivo. It is also comprised within the invention that some steps are performed in vitro, whereas others may be performed in vivo. Thus, for example the first steps may be performed in vitro and
30 where after an intermediate product may be fed to recombinant host cells, capable of performing the remaining steps of the method. Alternatively, the first steps may be performed in vivo and where after an intermediate product may be used as substrate for the subsequent step(s) performed in vitro. Other combinations can also be envisaged.

35

When said methods are performed in vitro each of the steps of the methods may be performed separately. Alternatively, one or more of the steps may be performed within the same mixture. In embodiments wherein some or all of the steps of the methods are performed separately, then the intermediate product of each of the steps may be purified or partly purified before performing the next step.

When said methods are performed in vivo, the methods employ use of a recombinant host expressing one or more of said enzymes or the methods may employ use of several recombinant hosts expressing one or more of said enzymes. The methods may also employ a mixture of recombinant and non-recombinant host. If more than one host is used then the hosts may be co-cultivated, or they may be cultured separately. If the hosts are cultivated separately the intermediate products may be recovered and optionally purified and partially purified and fed to recombinant hosts using the intermediate products as substrates. Useful recombinant hosts to be used with the invention are described herein below.

Said oxido-squalene, dioxido-squalene, cucurbitadienol, 24,25 epoxy cucurbitadienol or mogrol may be provided in any suitable manner. For example said oxido-squalene, dioxido-squalene, cucurbitadienol, 24,25 epoxy cucurbitadienol or mogrol may be provided in isolated form or as part of a composition or an extract. In embodiments of the invention, wherein the methods are performed in vivo, said oxido-squalene, dioxido-squalene, cucurbitadienol, 24,25 epoxy cucurbitadienol or mogrol may be added to the cultivation medium. It is also comprised within the invention that a recombinant host is used, which endogenously expresses oxido-squalene, dioxido-squalene, cucurbitadienol, 24,25 epoxy cucurbitadienol or mogrol.

Recombinant hosts described herein below can be used in methods to produce mogroside compounds. For example, if the recombinant host is a microorganism, the method can include growing the recombinant microorganism in a culture medium under conditions in which one or more of the enzymes catalyzing step(s) of the methods of the invention, e.g. synthases, hydrolases, CYP450s and/or UGTs are expressed. The recombinant microorganism may be grown in a fed batch or continuous process. Typically, the recombinant microorganism is grown in a fermenter at a defined temperature(s) for a desired period of time.

A cell lysate can be prepared from the recombinant host expressing one or more enzymes and be used to contact a substrate, such that mogroside compounds can be produced. For example, a cell lysate can be prepared from the recombinant host expressing one or more UGTs and used to contact mogrol, such that mogroside compounds can be produced.

In some embodiments, mogroside compounds can be produced using whole cells that are fed raw materials that contain precursor molecules, e.g., mogrol. The raw materials may be fed during cell growth or after cell growth. The whole cells may be in suspension or immobilized. The whole cells may be in fermentation broth or in a reaction buffer. In some embodiments a permeabilizing agent may be required for efficient transfer of substrate into the cells.

Levels of products, substrates and intermediates can be determined by extracting samples from culture media for analysis according to published methods. Mogroside compounds can be recovered from the culture or culture medium using various techniques known in the art.

Recombinant host

This document also feature recombinant hosts. As used herein, the term recombinant host is intended to refer to a host, the genome of which has been augmented by at least one incorporated DNA sequence. Said incorporated DNA sequence may be a heterologous nucleic acid encoding one or more polypeptides. Such DNA sequences include but are not limited to genes that are not naturally present, DNA sequences that are not normally transcribed into RNA or translated into a protein ("expressed"), and other genes or DNA sequences which one desires to introduce into the non-recombinant host. It will be appreciated that typically the genome of a recombinant host described herein is augmented through the stable introduction of one or more recombinant genes. Said recombinant gene may also be a heterologous nucleic acid encoding one or more polypeptides. Generally, the introduced DNA or heterologous nucleic acid is not originally resident in the host that is the recipient of the DNA, but it is within the scope of the invention to isolate a DNA segment from a given host, and to subsequently introduce one or more additional copies of that DNA into the same host, e.g., to enhance production of the product of a gene or alter the expression pattern of a

gene. In some instances, the introduced DNA or heterologous nucleic acid will modify or even replace an endogenous gene or DNA sequence by, e.g., homologous recombination or site-directed mutagenesis.

5 In particular, the recombinant host according to the present invention comprises one or more of the following heterologous nucleic acids:

- IIa. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IIa (Oxido-squalene -> cucurbitadienol)
- 10 IIb. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IIb (Dioxido-squalene -> 24,25 epoxy cucurbitadienol)
- IIIa. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IIIa (Cucurbitadienol -> 11-hydroxy-cucurbitadienol)
- 15 IIIb. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IIIb (24,25 epoxy cucurbitadienol -> 11-hydroxy-24,25 epoxy cucurbitadienol)
- IVa. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IVa (11-hydroxy-cucurbitadienol -> mogrol)
- 20 IVb. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IVb (11-hydroxy-24,25 epoxy cucurbitadienol -> mogrol)
- V Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step V (mogrol -> mogroside)

25 In addition to the heterologous nucleic acids, said recombinant host may have been modified to achieve Step Ia and/or Step Ib.

Enzymes capable of catalysing each of these steps are described herein below in more detail.

30 In one embodiment of the invention, the recombinant host according to the present invention may comprise the following heterologous nucleic acids:

- IIa. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IIa (Oxido-squalene -> cucurbitadienol)

IIIa. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IIIa (Cucurbitadienol -> 11-hydroxy-cucurbitadienol)

5 IVa. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IVa (11-hydroxy-cucurbitadienol -> mogrol)

V Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step V (mogrol -> mogroside)

and optionally said recombinant host may further have been modified to achieve Step Ia.

10

Said recombinant host cell is in particular useful in methods for producing mogrosides.

In one embodiment of the invention, the recombinant host according to the present invention may comprise the following heterologous nucleic acids:

15 IIIa. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IIIa (Cucurbitadienol -> 11-hydroxy-cucurbitadienol)

IVa. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IVa (11-hydroxy-cucurbitadienol -> mogrol)

20 V Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step V (mogrol -> mogroside)

Said recombinant host cell is in particular useful in methods for producing mogrosides comprising a step of providing curcubutadienol.

25

In one embodiment of the invention, the recombinant host according to the present invention may comprise the following heterologous nucleic acids:

V Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step V (mogrol -> mogroside)

30

Said recombinant host cell is in particular useful in methods for producing mogrosides comprising a step of providing mogrol.

In one embodiment of the invention, the recombinant host according to the present invention may comprise the following heterologous nucleic acids:

35

- 5 IIb. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IIb (Dioxido-squalene -> 24,25 epoxy cucurbitadienol)
- IIIb. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IIIb (24,25 epoxy cucurbitadienol -> 11-hydroxy-24,25 epoxy cucurbitadienol)
- IVb. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IVb (11-hydroxy-24,25 epoxy cucurbitadienol -> mogrol)
- 10 V Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step V (mogrol -> mogroside)

And optionally said recombinant host may have been modified to achieve Step Ib.

Said recombinant host cell is in particular useful in methods for producing mogrosides.

15

In one embodiment of the invention, the recombinant host according to the present invention may comprise the following heterologous nucleic acids:

- IIIb. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IIIb (24,25 epoxy cucurbitadienol -> 11-hydroxy-24,25 epoxy cucurbitadienol)
- 20 IVb. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IVb (11-hydroxy-24,25 epoxy cucurbitadienol -> mogrol)
- V Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step V (mogrol -> mogroside)
- 25

Said recombinant host cell is in particular useful in methods for producing mogrosides comprising a step of providing 24,25 epoxy cucurbitadienol.

30 In one embodiment of the invention, the recombinant host according to the present invention may comprise the following heterologous nucleic acids:

- IIb. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IIb (Dioxido-squalene -> 24,25 epoxy cucurbitadienol)

IIIb. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IIIb (24,25 epoxy cucurbitadienol -> 11-hydroxy-24,25 epoxy cucurbitadienol)

5 IVb. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IVb (11-hydroxy-24,25 epoxy cucurbitadienol -> mogrol)

and optionally said recombinant host may have been modified to achieve Step Ib.

Said recombinant host cell is in particular useful in methods for producing mogrol.

10

Suitable recombinant hosts include microorganisms, plant cells, and plants.

Thus, in one embodiment, a recombinant host that produces a mogroside compound can include a recombinant gene encoding at least a first UGT selected from the group consisting of 73C3, 73C6, 85C2, 73C5, and 73E1, and a recombinant gene encoding at least a second UGT selected from the group consisting of UGT98, UGT1495, UGT1817, UGT5914, UGT8468 and UGT10391. For example, a recombinant host can include a recombinant gene encoding at least one UGT selected from 73C3, 73C6, 85C2, and 73E1; a recombinant gene encoding 73C5; and a recombinant gene encoding at least one UGT selected from the group consisting of UGT98, UGT1495, UGT1817, UGT5914, UGT8468 and UGT10391. One or more of the following also can be included in a recombinant host: a recombinant gene encoding a cucurbitadienol synthase (e.g., from *Cucurbita pepo* or monk fruit); a recombinant gene encoding a cytochrome P450 polypeptide selected from the group CYP533, CYP937, CYP1798, CYP1994, CYP2048, CYP2740, CYP3404, CYP3968, CYP4112, CYP4149, CYP4491, CYP5491, CYP6479, CYP7604, CYP8224, CYP8728, CYP10020, and CYP10285 (SEQ ID NOs:3-20, respectively); and a recombinant gene encoding a squalene synthase (e.g., from *Gynostemma pentaphyllum* or *Arabidopsis thaliana*). CYP5491 has previously also been referred to as CYP87.

30

At least one of the genes in the recombinant host is a recombinant gene, the particular recombinant gene(s) depending on the species or strain selected for use. Additional genes or biosynthetic modules can be included in order to increase yield of

mogrol and mogrosides, improve efficiency with which energy and carbon sources are converted to mogrol and mogrosides, and/or to enhance productivity from the cell culture or plant.

5 The recombinant host further can include a recombinant gene encoding a cucurbitadienol synthase and/or a recombinant gene encoding a cytochrome P450 polypeptide (e.g., CYP533, CYP937, CYP1798, CYP1994, CYP2048, CYP2740, CYP3404, CYP3968, CYP4112, CYP4149, CYP4491, CYP5491, CYP6479, CYP7604, CYP8224, CYP8728, CYP10020, or CYP10285) and/or a recombinant gene encoding
10 a squalene synthase.

It is also comprised within the invention that the recombinant host may be modified in order to reduce glucanase activity, in particular glucanase activity, which may result in deglucosylation of mogrosides. Thus, the recombinant host may for example be
15 modified to reduce or even abolish exo-1,3-beta-Glucanase activity. In embodiments of the invention when the recombinant host is yeast, this may be accomplished by deletion of the EXG1 gene and/or of the EXG2 gene, both of which are encoding an exo-1,3-beta-Glucanase.

20 The term "recombinant gene" refers to a gene or DNA sequence that is introduced into a recipient host, regardless of whether the same or a similar gene or DNA sequence may already be present in such a host. The term "heterologous nucleic acid" refers to a nucleic acid that is introduced into a recipient host, wherein said host does not endogenously comprise said nucleic acid. "Introduced," or "augmented" in this context,
25 is known in the art to mean introduced or augmented by the hand of man. Thus, a recombinant gene may be a DNA sequence from another species, or may be a DNA sequence that originated from or is present in the same species, but has been incorporated into a host by recombinant methods to form a recombinant host. It will be appreciated that a recombinant gene that is introduced into a host can be identical to a
30 DNA sequence that is normally present in the host being transformed, and is introduced to provide one or more additional copies of the DNA to thereby permit overexpression or modified expression of the gene product of that DNA.

A recombinant gene encoding a polypeptide described herein comprises the coding
35 sequence for that polypeptide, operably linked in sense orientation to one or more

regulatory regions suitable for expressing the polypeptide. Because many microorganisms are capable of expressing multiple gene products from a polycistronic mRNA, multiple polypeptides can be expressed under the control of a single regulatory region for those microorganisms, if desired. A coding sequence and a regulatory region are considered to be operably linked when the regulatory region and coding sequence are positioned so that the regulatory region is effective for regulating transcription or translation of the sequence. Typically, the translation initiation site of the translational reading frame of the coding sequence is positioned between one and about fifty nucleotides downstream of the regulatory region for a monocistronic gene. In many cases, the coding sequence for a polypeptide described herein is identified in a species other than the recombinant host, *i.e.*, is a heterologous nucleic acid. Thus, if the recombinant host is a microorganism, the coding sequence can be from other prokaryotic or eukaryotic microorganisms, from plants or from animals. In some case, however, the coding sequence is a sequence that is native to the host and is being reintroduced into that organism. A native sequence can often be distinguished from the naturally occurring sequence by the presence of non-natural sequences linked to the exogenous nucleic acid, *e.g.*, non-native regulatory sequences flanking a native sequence in a recombinant nucleic acid construct. Such sequences may then also be considered heterologous nucleic acids. In addition, stably transformed exogenous nucleic acids typically are integrated at positions other than the position where the native sequence is found.

"Regulatory region" refers to a nucleic acid having nucleotide sequences that influence transcription or translation initiation and rate, and stability and/or mobility of a transcription or translation product. Regulatory regions include, without limitation, promoter sequences, enhancer sequences, response elements, protein recognition sites, inducible elements, protein binding sequences, 5' and 3' untranslated regions (UTRs), transcriptional start sites, termination sequences, polyadenylation sequences, introns, and combinations thereof. A regulatory region typically comprises at least a core (basal) promoter. A regulatory region also may include at least one control element, such as an enhancer sequence, an upstream element or an upstream activation region (UAR). A regulatory region is operably linked to a coding sequence by positioning the regulatory region and the coding sequence so that the regulatory region is effective for regulating transcription or translation of the sequence. For example, to operably link a coding sequence and a promoter sequence, the translation initiation site

of the translational reading frame of the coding sequence is typically positioned between one and about fifty nucleotides downstream of the promoter. A regulatory region can, however, be positioned at further distance, for example as much as about 5,000 nucleotides upstream of the translation initiation site, or about 2,000 nucleotides upstream of the transcription start site.

The choice of regulatory regions to be included depends upon several factors, including, but not limited to, efficiency, selectability, inducibility, desired expression level, and preferential expression during certain culture stages. It is a routine matter for one of skill in the art to modulate the expression of a coding sequence by appropriately selecting and positioning regulatory regions relative to the coding sequence. It will be understood that more than one regulatory region may be present, *e.g.*, introns, enhancers, upstream activation regions, transcription terminators, and inducible elements. One or more genes can be combined in a recombinant nucleic acid construct in "modules" useful for a discrete aspect of mogroside production. Combining a plurality of genes in a module, particularly a polycistronic module, facilitates the use of the module in a variety of species. In addition to genes useful for mogroside production, a recombinant construct typically also contains an origin of replication, and one or more selectable markers for maintenance of the construct in appropriate species.

It will be appreciated that because of the degeneracy of the genetic code, a number of nucleic acids can encode a particular polypeptide; *i.e.*, for many amino acids, there is more than one nucleotide triplet that serves as the codon for the amino acid. Thus, codons in the coding sequence for a given polypeptide can be modified such that optimal expression in a particular host is obtained, using appropriate codon bias tables for that host (*e.g.*, microorganism). Nucleic acids may also be optimized to a GC-content preferable to a particular host, and/or to reduce the number of repeat sequences. As isolated nucleic acids, these modified sequences can exist as purified molecules and can be incorporated into a vector or a virus for use in constructing modules for recombinant nucleic acid constructs.

A number of prokaryotes and eukaryotes are suitable for use as recombinant hosts with the present invention. Thus, the recombinant host may *e.g.* be selected from the group consisting of gram-negative bacteria, yeast and fungi. A species and strain selected for use as a mogroside production strain is first analyzed to determine which

production genes are endogenous to the strain and which genes are not present.

Genes for which an endogenous counterpart is not present in the strain are assembled in one or more recombinant constructs, which are then transformed into the strain in order to supply the missing function(s). Thus, it may be analysed which of steps IIa, IIIa, IVa and V are already performed by the host, and then said host may be modified by introduction of heterologous nucleic acids encoding enzymes catalyzing the remaining steps. Similarly, it may be analysed which of steps IIb, IIIb, IVb and V are already performed by the host, and then said host may be modified by introduction of heterologous nucleic acids encoding enzymes catalyzing the remaining steps. As mentioned before the recombinant host may also be modified to increase levels of oxido-squalene and/or dioxido-squalene.

Exemplary prokaryotic and eukaryotic species useful as recombinant with the present invention are described in more detail below. However, it will be appreciated that other species may be suitable. For example, the recombinant host may be in a genus selected from the group consisting of *Agaricus*, *Aspergillus*, *Bacillus*, *Candida*, *Corynebacterium*, *Escherichia*, *Fusarium*/*Gibberella*, *Kluyveromyces*, *Laetiporus*, *Lentinus*, *Phaffia*, *Phanerochaete*, *Pichia*, *Physcomitrella*, *Rhodoturula*, *Saccharomyces*, *Schizosaccharomyces*, *Sphaceloma*, *Xanthophyllomyces* and *Yarrowia*. Exemplary species from such genera useful as recombinant hosts include *Lentinus tigrinus*, *Laetiporus sulphureus*, *Phanerochaete chrysosporium*, *Pichia pastoris*, *Physcomitrella patens*, *Rhodoturula glutinis*, *Rhodoturula mucilaginosa*, *Phaffia rhodozyma*, *Xanthophyllomyces dendrorhous*, *Fusarium fujikuroi*/*Gibberella fujikuroi*, *Candida utilis* and *Yarrowia lipolytica*. In some embodiments, a recombinant host may be a microorganism, for example an Ascomycete such as *Gibberella fujikuroi*, *Kluyveromyces lactis*, *Schizosaccharomyces pombe*, *Aspergillus niger*, or *Saccharomyces cerevisiae*. In some embodiments, a recombinant host may be a microorganism for example a prokaryote such as *Escherichia coli*, *Rhodobacter sphaeroides*, or *Rhodobacter capsulatus*. It will be appreciated that certain microorganisms can be used to screen and test genes of interest in a high throughput manner, while other microorganisms with desired productivity or growth characteristics can be used for large-scale production of mogroside compounds. In particular, food grade microorganisms may be useful for large-scale production purposes.

Saccharomyces cerevisiae

As described above the recombinant host may for example be *Saccharomyces cerevisiae*. *Saccharomyces cerevisiae* is a widely used chassis organism in synthetic biology, and can be used as the recombinant microorganism platform. There are libraries of mutants, plasmids, detailed computer models of metabolism and other information available for *S. cerevisiae*, allowing for rational design of various modules to enhance product yield. Methods are known for making recombinant microorganisms. The VG4 strain of *S. cerevisiae* from Brochado et al. 2010 (Microb Cell Fact. 9:84) may be particularly useful. VG4 has the genotype of *pdh1Δgdh1Δ↑GDH2*. Another very useful strain of *S. cerevisiae* is BY4742 described herein below in Example 9, or the yeast strain described in Kirby, J et al in FEBS Journal **275** (2008) 1852-1859.

Aspergillus spp.

The recombinant host may also be a *Aspergillus* species such as *A. oryzae*, *A. niger* and *A. sojae*. *Aspergillus* spp, such as the aforementioned are widely used microorganisms in food production, and can also be used as the recombinant microorganism platform. Nucleotide sequences are available for genomes of *A. nidulans*, *A. fumigatus*, *A. oryzae*, *A. clavatus*, *A. flavus*, *A. niger*, and *A. terreus*, allowing rational design and modification of endogenous pathways to enhance flux and increase product yield. Any of these may be used recombinant hosts. Metabolic models have been developed for *Aspergillus*, as well as transcriptomic studies and proteomics studies. *A. niger* is cultured for the industrial production of a number of food ingredients such as citric acid and gluconic acid, and thus species such as *A. niger* are generally suitable for the production of food ingredients.

Escherichia coli

The recombinant host may also be *Escherichia coli*, which is another widely used platform organism in synthetic biology. Similar to *Saccharomyces*, there are libraries of mutants, plasmids, detailed computer models of metabolism and other information available for *E. coli*, allowing for rational design of various modules to enhance product yield. Methods similar to those described above for *Saccharomyces* can be used to make recombinant *E. coli* microorganisms.

Rhodobacter spp.

The recombinant host may also be *Rhodobacter*. Similar to *E. coli*, there are libraries of mutants available as well as suitable plasmid vectors, allowing for rational design of

various modules to enhance product yield. Methods similar to those described above for *E. coli* can be used to make recombinant *Rhodobacter* microorganisms.

Physcomitrella spp.

- 5 The recombinant host may also be *Physcomitrella* mosses. *Physcomitrella* mosses, when grown in suspension culture, have characteristics similar to yeast or other fungal cultures. This genera is becoming an important type of cell for production of plant secondary metabolites, which can be difficult to produce in other types of cells.

10 **Step Ia - Enhancing levels of oxido-squalene**

- As described herein above the methods of the invention may comprise a step of enhancing the levels of oxido-squalene. This is in particular relevant in methods comprising step IIa, wherein step IIa is performed in vivo. Step Ia may in particular be performed by modifying the recombinant host to be used with the methods in a manner enhancing the levels of oxido-squalene in said recombinant host. The invention also relates to recombinant hosts modified to enhance the levels of oxido-squalene.
- 15

- Thus, the methods may also comprise one or more steps leading to formation of oxido-squalene, in particular to the formation of 2,3-oxido-squalene. Said steps are preferably performed prior to step IIa described below, or simultaneously herewith. FIG. 3 provides a schematic of the pathway from squalene to mogrosides.
- 20

- One step in the production of oxido-squalene may be production of squalene from farnesyl pyrophosphate. One enzyme that catalyzes the production of squalene from farnesyl pyrophosphate is squalene synthase (also referred to as squalene synthase). Said squalene synthase may be any enzyme classified under EC 2.5.1.21. The reaction is typically thought to proceed using NADPH as a cosubstrate. Accordingly, the method may comprise a step of production of squalene from farnesyl pyrophosphate catalyzed by a squalene synthase in the presence of NADPH. In embodiments of the invention wherein the methods are performed in vivo, the recombinant host may thus comprise a heterologous nucleic acid encoding a squalene synthase. Some recombinant hosts may comprise an endogenous squalene synthase in which case the endogenous enzyme may suffice. Endogenous squalene production
- 25
- 30

pathways exist in yeast metabolism, and accordingly, if the recombinant host is yeast, then said step may be endogenous to the recombinant host.

5 The squalene synthase may be any useful squalene synthase. For example the squalene synthase may be squalene synthase from *Gynostemma pentaphyllum* (protein accession number C4P9M2), another cucurbitaceae family plant. The squalene synthase may also be selected from the the groups consisting of squalene synthase of *Arabidopsis thaliana* (protein accession number C4P9M3), *Brassica napus*, *Citrus macrophylla*, *Euphorbia tirucalli* (protein accession number B9WZW7),
10 *Glycine max*, *Glycyrrhiza glabra* (protein accession number Q42760, Q42761), *Glycrrhiza uralensis* (protein accession number D6QX40, D6QX41, D6QX42, D6QX43, D6QX44, D6QX45, D6QX47, D6QX39, D6QX55, D6QX38, D6QX53, D6QX37, D6QX35, B5AID5, B5AID4, B5AID3, C7EDD0, C6KE07, C6KE08 , C7EDC9), *Lotusjaponicas* (protein accession number Q84LE3), *Medicago truncatula* (protein
15 accessionnumber Q8GSL6), *Pisum sativum*, *Ricinus communis* (protein accession number B9RHC3), and *Prunus mume* and functional homologues of any of the aforementioned sharing at least at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith. Increased copy numbers, heterologous nucleic acids encoding squalene synthase, or
20 increased expression of the native squalene synthase may improve levels of mogrosides produced in a recombinant host.

Another step in the production of oxido-squalene may be production of oxido-squalene from squalene. One enzyme that catalyzes the production of oxido-squalene from
25 squalene is squalene epoxidase (also referred to as squalene monooxygenase). Said squalene epoxidase may be any enzyme classified under EC 1.4.99.7. The reaction is typically thought to proceed using NADPH as a cosubstrate. Accordingly, the method may comprise a step of production of oxido-squalene from squalene catalyzed by a squalene epoxidase in the presence of NADPH. In embodiments of the invention
30 wherein the methods are performed in vivo, the recombinant host may thus comprise a heterologous nucleic acid encoding a squalene epoxidase. Some recombinant hosts may comprise an endogenous squalene epoxidase, in which case the endogenous enzyme may suffice. Endogenous oxido-squalene production pathways exist in yeast metabolism, and accordingly, if the recombinant host is yeast, then said step may be
35 endogenous to the recombinant host. However, in order to enhance the level of oxido-

squalene it may never-the-less be advantageous to express addition squalene epoxidase.

- 5 The squalene epoxidase may be any useful squalene epoxidase. The squalene epoxidase may for example be squalene epoxidase from *Gynostemma pentaphyllum* (protein accession number C4P9M2), a cucurbitaceae family plant. The squalene epoxidase may also be selected from the group consisting of squalene epoxidase of *Arabidopsis thaliana* (protein accession number Q9SM02, O65403, O65402, O65404, O81000, or Q9T064), *Brassica napus* (protein accession number 10 O65727, O65726), *Euphorbia tirucalli* (protein accession number A7VJN1), *Medicago truncatula* (protein accession number Q8GSM8, Q8GSM9), *Pisum sativum*, and *Ricinus communis* (protein accession number B9R6V0, B9S7W5, B9S6Y2, B9T0Y3, B9S7T0, B9SX91) and functional homologues of any of the aforementioned sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for 15 example at least 98% sequence identity therewith. Increased copy numbers, heterologous nucleic acids encoding squalene epoxidase, or increased expression of the native squalene epoxidase may improve levels of mogrosides produced in a recombinant host.
- 20 The squalene epoxidase may also be the product of the ERG1 gene of *S. cerevisiae*. Thus, the squalene epoxidase may be a polypeptide of SEQ ID NO:54 or a functional homologues thereof sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.
- 25 In one embodiment the recombinant host comprises a heterologous nucleic acid encoding a squalene epoxidase operably linked to sequence directing high expression of said squalene epoxidase in said host cell. Thus, the squalene epoxidase may be endogenous to the recombinant host, but the expression level may be increased by additional copies of nucleic acids encoding the squalene epoxidase and/or by use of 30 stronger promoters.

- Oxido-squalene serves as a substrate for production of lanosterol. Thus, in one embodiment the level of oxido-squalene may be increased by reducing the activity of lanosterol synthase. In recombinant hosts expressing an endogenous lanosterol 35 synthase, this may be achieved by substituting the endogenous promoter directed

expression of lanosterol synthase with a weaker promoter directing expression of a lower level of lanosterol synthase. In yeast the ERG7 gene encodes lanosterol synthase. Thus, when the recombinant host is yeast, then the promoter of the ERG7 gene may be substituted for another promoter, which directs a level of expression, which is lower than the endogenous expression level of ERG7. The lanosterol synthase may thus be the product of the ERG7 gene of *S. cerevisiae*, the sequence of which is provided herein as SEQ ID NO:55 or a functional homologues thereof sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.

Examples of useful weak promoters include the methionine-repressible promoter of the MET3 gene or the CUP1 copper inducible promoter. Non-limiting examples of how to reduce the activity of lanosterol synthase are described in Example 9 herein below or in Kirby et al., 2008 (vide supra) both of which are incorporated by reference herein. The sequence of *S. cerevisiae* lanosterol synthase is provided as SEQ ID NO:55. Thus, when the recombinant host is *S. cerevisiae*, then it is preferred that the polypeptide of SEQ ID NO:55 is expressed at a lower level than the level of said polypeptide in wild type *S. cerevisiae*. Similarly, when the recombinant host expresses a polypeptide similar to the polypeptide of SEQ ID NO:55 (e.g. at least 70% identical to SEQ ID NO:55), then it is preferred that said polypeptide at least 70% identical to SEQ ID NO:55 is expressed at a lower level than the level of said polypeptide in the wild type host.

In addition, expression of a truncated form of the enzyme 3-hydroxy-3-methylglutaryl-CoA reductase (tHMG1) may also lead enhanced levels of oxido-squalene. A useful truncated form of yeast HMG reductase (tHMG1) is described in Donald et al., 1997, Appl.Environ. Microbiol. 63, 3341-3344.

Step Ib - Enhancing levels of dioxido-squalene

As described herein above the methods of the invention may comprise a step of enhancing the levels of dioxido-squalene. This is in particular relevant in methods comprising step IIb, wherein step IIb is performed in vivo. Step Ib may in particular be performed by modifying the recombinant host to be used with the methods in a manner

enhancing the levels of dioxido-squalene in said recombinant host. The invention also relates to recombinant hosts modified to enhance the levels of dioxido-squalene.

5 Thus, the methods may also comprise one or more steps leading to enhanced levels of dioxido-squalene. Said steps are preferably performed prior to step IIb described below, or simultaneously herewith.

10 The present invention describes that the levels of dioxido-squalene in particular may be enhanced by high expression of a squalene epoxidase. Said squalene epoxidase may be any of the squalene epoxidase described herein above in the section "Step Ia - Enhancing levels of oxido-squalene". In particular, the squalene epoxidase may be the product of the ERG1 gene of *S. cerevisiae*. Thus, the squalene epoxidase may be a polypeptide of SEQ ID NO:54 or a functional homologues thereof sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at
15 least 98% sequence identity therewith. High expression level may be achieved by introducing a heterologous nucleic acid encoding a squalene epoxidase into the host cell operably linked to sequence directing high expression of said squalene epoxidase in said host cell. Thus, the squalene epoxidase may be endogenous to the recombinant host, but the expression level may be increased by additional copies of nucleic acids
20 encoding the squalene epoxidase and/or by use of stronger promoters.

The levels of dioxido-squalene may also be enhanced by reducing the activity of lanosterol synthase. The activity of lanosterol synthase may be reduced by any of the methods described herein above in the section "Step Ia - Enhancing levels of oxido-
25 squalene".

The levels of dioxido-squalene may also be enhanced by expression of a truncated form of the enzyme 3-hydroxy-3-methylglutaryl-CoA reductase (tHMG1) may also lead enhanced levels of oxido-squalene. A useful truncated form of yeast HMG reductase
30 (tHMG1) is described in Donald et al., 1997, Appl. Environ. Microbiol. 63, 3341-3344.

Step IIa - Oxido-squalene -> cucurbitadienol

35 As described herein above the methods of the invention may comprise a step of producing cucurbitadienol from oxido-squalene, and in particular from 2,3-oxido-

squalene using an enzyme or mixture of enzymes capable of catalysing conversion of oxido-squalene to form cucurbitadienol. The invention also relates to recombinant hosts comprising a heterologous nucleic acid encoding an enzyme capable of catalysing conversion of oxido-squalene to cucurbitadienol.

5

The step may be performed in vitro by incubating a composition comprising oxido-squalene with said enzyme or a mixture of enzymes capable of catalyzing conversion of oxido-squalene to form cucurbitadienol.

10 The step may also be performed in vivo in a recombinant host comprising heterologous nucleic acid(s) encoding an enzyme or a mixture of enzymes capable of catalyzing conversion of oxido-squalene to form cucurbitadienol. Said recombinant host may be capable of producing oxido-squalene, for example because the recombinant host expresses one or more enzymes of the oxido-squalene biosynthesis pathway.

15 Alternatively, oxido-squalene may be provided to said recombinant host for example in the cultivation medium.

Said enzyme or mixture of enzyme capable of catalyzing conversion of oxido-squalene to form cucurbitadienol preferably comprises or consists of a cucurbitadienol synthase.

20

Said cucurbitadienol synthase may be any useful cucurbitadienol synthase, for example a cucurbitadienol synthase, which has been classified as an oxidosqualene cyclase, such as the oxidosqualene cyclase described by Shibuya, *Tetrahedron*, Vol 60: pp. 6995-7003 (2004).

25

The amino acid sequence of a cucurbitadienol synthase from *Cucurbita pepo* is provided herein as SEQ ID NO:1 and also is provided in GenBank under Protein Accession No. BAD34645.1. In one embodiment of the invention the cucurbitadienol synthase is a polypeptide of SEQ ID NO:1 or a functional homologue thereof sharing at
30 least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.

As described in Example 5, the cucurbitadienol synthase from monk fruit was identified and the sequence of the C-terminal portion of the polypeptide determined.

35 The amino acid sequence of the C-terminal portion of the monk fruit polypeptide is

provided herein as SEQ ID NO:2. SEQ ID NO:2 is 97.5% identical to residues 515 to 764 of the *C. pepo* polypeptide set forth in SEQ ID NO:1. Thus, in one embodiment of the invention the cucurbitadienol synthase is a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:2.

5

In a preferred embodiment the cucurbitadienol synthase is the polypeptide of SEQ ID NO:43 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.

10

Other homologous proteins can be found of similar length and having approximately 70% homology or higher to SEQ ID NO:1. Such homologs include the polypeptides from *Lotus japonicas* (BAE53431), *Populus trichocarpa* (XP_002310905), *Actaea racemosa* (ADC84219), *Betula platyphylla* (BAB83085), *Glycyrrhiza glabra* (BAA76902), *Vitis vinifera* (XP_002264289), *Centella asiatica* (AAS01524), *Panax ginseng* (BAA33460), and *Betula platyphylla* (BAB83086). The cucurbitadienol synthase may be any of the aforementioned or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.

20

Step IIb - Dioxido-squalene -> 24,25 epoxy cucurbitadienol

As described herein above the methods of the invention may comprise a step of producing 24,25 epoxy cucurbitadienol from dioxido-squalene using an enzyme or mixture of enzymes capable of catalysing conversion of oxido-squalene to form cucurbitadienol. The invention also relates to recombinant hosts comprising a heterologous nucleic acid encoding an enzyme capable of catalysing conversion of dioxido-squalene to 24,25 epoxy cucurbitadienol.

25

30

The step may be performed in vitro by incubating a composition comprising dioxido-squalene with said enzyme or a mixture of enzymes capable of catalyzing conversion of dioxido-squalene to form 24,25 epoxy cucurbitadienol.

35

The step may also be performed in vivo in a recombinant host comprising heterologous nucleic acid(s) encoding an enzyme or a mixture of enzymes capable of catalyzing

conversion of dioxido-squalene to 24,25 epoxy cucurbitadienol. Said recombinant host may be capable of producing dioxido-squalene, for example because the recombinant host expresses one or more enzymes of the dioxido-squalene biosynthesis pathway. However, it is preferred that said recombinant host has been modified to enhance
5 levels of dioxido-squalene in any of the manners described herein above in the section "Step 1b Enhancing levels of dioxido-squalene". Alternatively, dioxido-squalene may be provided to said recombinant host for example in the cultivation medium.

Said enzyme or mixture of enzyme capable of catalyzing conversion of dioxido-squalene to 24,25 epoxy cucurbitadienol preferably comprises or consists of a
10 cucurbitadienol synthase.

Said cucurbitadienol synthase may be any useful cucurbitadienol synthase, for example a cucurbitadienol synthase, which has been classified as an oxidosqualene cyclase, such as the oxidosqualene cyclase described by Shibuya, *Tetrahedron*, Vol
15 60: pp. 6995-7003 (2004). In one embodiment of the invention the cucurbitadienol synthase is a polypeptide of SEQ ID NO:1 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.

In a preferred embodiment the cucurbitadienol synthase is the polypeptide of SEQ ID NO:43 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence
20 identity therewith.

Other homologous proteins can be found of similar length and having approximately 70% homology or higher to SEQ ID NO:1. Such homologs include the polypeptides from *Lotus japonicas* (BAE53431), *Populus trichocarpa* (XP_002310905), *Actaea racemosa* (ADC84219), *Betula platyphylla* (BAB83085), *Glycyrrhiza glabra*
25 (BAA76902), *Vitis vinifera* (XP_002264289), *Centella asiatica* (AAS01524), *Panax ginseng* (BAA33460), and *Betula platyphylla* (BAB83086). The cucurbitadienol synthase may be any of the aforementioned or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.
30
35

Step IIIa - Cucurbitadienol -> 11-hydroxy-cucurbitadienol

As described herein above the methods of the invention may comprise a step of producing 11-hydroxy-cucurbitadienol from cucurbitadienol using an enzyme or a mixture of enzymes capable of catalysing hydroxylation of cucurbitadienol to form 11-hydroxy-cucurbitadienol.

The step may be performed *in vitro* by incubating a composition comprising cucurbitadienol with said enzyme capable of catalyzing hydroxylation of cucurbitadienol to form 11-hydroxy-cucurbitadienol.

The step may also be performed *in vivo* in a recombinant host comprising heterologous nucleic acid(s) encoding an enzyme capable of catalyzing hydroxylation of cucurbitadienol to form 11-hydroxy-cucurbitadienol. Said recombinant host may be capable of producing cucurbitadienol, for example because the recombinant host expresses one or more enzymes of the cucurbitadienol biosynthesis pathway. Alternatively, cucurbitadienol may be provided to said recombinant host for example in the cultivation medium.

Said enzyme capable of catalyzing hydroxylation of cucurbitadienol to form 11-hydroxy-cucurbitadienol preferably is selected from the group of cytochrome P450 enzymes.

As indicated in Example 7, one or more of CYP533, CYP937, CYP1798, CYP1994, CYP2048, CYP2740, CYP3404, CYP3968, CYP4112, CYP4149, CYP4491, CYP5491, CYP6479, CYP7604, CYP8224, CYP8728, CYP10020, and CYP10285 (encoded by SEQ ID NOs: 3-20, respectively) can be used to produce mogrol. eYAC technology can be used to assess activity of the cytochrome P450 enzymes as set forth in Example 8. Alternatively, an *in vitro* reaction can be used to assess the activity. Thus, in one embodiment of the invention at least one cytochrome P450 enzyme is selected from the group consisting of polypeptides encoding by the nucleic acid sequence SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20 or a or a functional homologue thereof sharing at least 70%,

such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith at the amino acid level.

In a preferred embodiment of the invention the enzyme capable of catalyzing hydroxylation of cucurbitadienol to form 11-hydroxy-cucurbitadienol is CYP5491. Thus, the enzyme catalyzing hydroxylation of cucurbitadienol to form 11-hydroxy-cucurbitadienol may be a polypeptide of SEQ ID NO:44 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.

In one embodiment of the invention this step may be aided by at least one CYP activator. This step of the methods of the invention may thus comprise use of a cytochrome P450 enzyme as described above in combination with at least one CYP activator. Thus, the recombinant host may in addition to containing heterologous nucleic acids encoding the cytochrome P450 enzymes described herein above also contain a heterologous nucleic acid encoding a CYP activator. Said CYP activator may be any useful CYP activator, for example it may be a polypeptide be a polypeptide of SEQ ID NO:46 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.

Step IIIb 24,25 epoxy cucurbitadienol -> 11-hydroxy-24,25 epoxy cucurbitadienol

As described herein above the methods of the invention may comprise a step of producing 11-hydroxy-24,25 epoxy cucurbitadienol from 24,25 epoxy cucurbitadienol using an enzyme capable of catalysing hydroxylation of 24,25 epoxy cucurbitadienol to form 11-hydroxy-24,25 epoxy cucurbitadienol.

The step may be performed in vitro by incubating a composition comprising 24,25 epoxy cucurbitadienol with said enzyme capable of catalyzing hydroxylation of 24,25 epoxy cucurbitadienol to form 11-hydroxy-24,25 epoxy cucurbitadienol.

The step may also be performed in vivo in a recombinant host comprising heterologous nucleic acid(s) encoding an enzyme capable of catalyzing hydroxylation of 24,25 epoxy cucurbitadienol to form 11-hydroxy-24,25 epoxy cucurbitadienol. Said recombinant host

may be capable of producing 24,25 epoxy cucurbitadienol, for example because the recombinant host expresses one or more enzymes of the 24,25 epoxy cucurbitadienol biosynthesis pathway, e.g. cucurbitadienol synthase. Alternatively, 24,25 epoxy cucurbitadienol may be provided to said recombinant host for example in the cultivation medium.

Said enzyme capable of catalyzing hydroxylation of cucurbitadienol to form 11-hydroxy-cucurbitadienol preferably is selected from the group of cytochrome P450 enzymes.

In a preferred embodiment of the invention the enzyme capable of catalyzing hydroxylation of 24,25 epoxy cucurbitadienol to form 11-hydroxy-24,25 epoxy cucurbitadienol is CYP5491. Thus, the enzyme catalyzing hydroxylation of 24,25 epoxy cucurbitadienol to form 11-hydroxy-24,25 epoxy cucurbitadienol may be a polypeptide of SEQ ID NO:44 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.

In one embodiment of the invention this step may be aided by at least one CYP activator. This step of the methods of the invention may thus comprise use of a cytochrome P450 enzyme as described above in combination with at least one CYP activator. Thus, the recombinant host may in addition to containing heterologous nucleic acids encoding the cytochrome P450 enzymes described herein above also contain a heterologous nucleic acid encoding a CYP activator. Said CYP activator may be any useful CYP activator, for example it may be a polypeptide be a polypeptide of SEQ ID NO:46 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.

Step IVa - 11-hydroxy-cucurbitadienol -> mogrol

As described herein above the methods of the invention may comprise a step of producing mogrol from 11-hydroxy-cucurbitadienol using an enzyme or a mixture of enzymes capable of catalysing conversion of 11-hydroxy-cucurbitadienol to form mogrol.

The step may be performed in vitro by incubating a composition comprising 11-hydroxy-cucurbitadienol with said enzyme or mixture of enzymes capable of catalyzing conversion of 11-hydroxy-cucurbitadienol to form mogrol.

- 5 The step may also be performed in vivo in a recombinant host comprising heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalyzing conversion of 11-hydroxy-cucurbitadienol to form mogrol. Said recombinant host may be capable of producing 11-hydroxy-cucurbitadienol, for example because the recombinant host expresses one or more enzymes of the 11-hydroxy-cucurbitadienol biosynthesis pathway. Alternatively, 11-hydroxy-cucurbitadienol may be provided to
10 said recombinant host for example in the cultivation medium.

- Said enzyme or mixture of enzymes capable of catalyzing conversion of 11-hydroxy-cucurbitadienol to form mogrol preferably comprises one or more enzymes with
15 together has CYP450 activity and epoxide hydrolase activity.

- Enzymes with CYP450 include for example the polypeptides encoding by the nucleic acid sequence SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20 or a or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith at the amino acid level.

- 25 Another enzyme with CYP450 activity is CYP5491. Thus, the enzyme with CYP450 activity may be a polypeptide of SEQ ID NO:44 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.

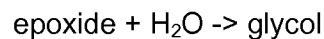
- 30 In one embodiment of the invention this step may be aided by at least one CYP activator. This step of the methods of the invention may thus comprise use of a cytochrome P450 enzyme as described above in combination with at least one CYP activator. Thus, the recombinant host may in addition to containing heterologous nucleic acids encoding the cytochrome P450 enzymes described herein above also
35 contain a heterologous nucleic acid encoding a CYP activator. Said CYP activator may

be any useful CYP activator, for example it may be a polypeptide be a polypeptide of SEQ ID NO:46 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.

5

The enzyme having epoxide hydrolase activity may for example be an enzyme classified under EC 3.3._._. Said epoxide hydrolase preferably catalyses the following reaction:

10



Examples of enzymes with epoxide hydrolase activity includes *S. grosvenorii* Epoxide hydrolase 1 and *S. grosvenorii* Epoxide hydrolase 2. Thus, the enzyme with epoxide hydrolase activity may be selected from the group consisting of polypeptides of SEQ ID NO:38, SEQ ID NO:40 and functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.

15

Step IVa - 11-hydroxy-24,25 epoxy cucurbitadienol -> mogrol

20

As described herein above the methods of the invention may comprise a step of producing mogrol from 11-hydroxy-24,25 epoxy cucurbitadienol using an enzyme or a mixture of enzymes capable of catalysing conversion of 11-hydroxy-24,25 epoxy cucurbitadienol to form mogrol.

25

The step may be performed in vitro by incubating a composition comprising 11-hydroxy-24,25 epoxy cucurbitadienol with said enzyme or mixture of enzymes capable of catalyzing conversion of 11-hydroxy-24,25 epoxy cucurbitadienol to form mogrol.

30

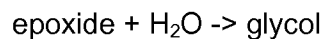
The step may also be performed in vivo in a recombinant host comprising heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalyzing conversion of 11-hydroxy-24,25 epoxy cucurbitadienol to form mogrol. Said recombinant host may be capable of producing 11-hydroxy-24,25 epoxy cucurbitadienol, for example because the recombinant host expresses one or more enzymes of the 11-hydroxy-24,25 epoxy cucurbitadienol biosynthesis pathway.

35

Alternatively, 11-hydroxy-24,25 epoxy cucurbitadienol may be provided to said recombinant host for example in the cultivation medium.

5 Said enzyme or mixture of enzymes capable of catalyzing conversion of 11-hydroxy-24,25 epoxy cucurbitadienol to form mogrol preferably comprises an enzyme with epoxide hydrolase activity.

10 The enzyme having epoxide hydrolase activity may for example be an enzyme classified under EC 3.3._._. Said epoxide hydrolase preferably catalyses the following reaction:



15 Examples of enzymes with epoxide hydrolase activity includes *S. grosvenorii* Epoxide hydrolase 1 and *S. grosvenorii* Epoxide hydrolase 2. Thus, the enzyme with epoxide hydrolase activity may be selected from the group consisting of polypeptides of SEQ ID NO:38, SEQ ID NO:40 and functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.

20

Step V - Mogrol -> mogroside

25 The methods of invention may involve a step of glycosylating mogrol to form mogroside. This step is in general accomplished with the aid of an enzyme or a mixture of enzymes capable of catalyzing glycosylation of mogrol and/or of glycosylated mogrol.

30 The mogroside may be any of the mogrosides described herein below in the section "Mogrosides".

Step V may be performed in vitro by incubating a composition comprising mogrol with said enzyme or a mixture of enzymes capable of catalyzing glycosylation of mogrol. The step may also be divided into separate steps, wherein each step involves glycosylation of mogrol or glycosylated mogrol.

35

The step may also be performed in vivo in a recombinant host comprising heterologous nucleic acid(s) encoding an enzyme or a mixture of enzymes capable of catalyzing glycosylation of mogrol and optionally also of glycosylated mogrol. Said recombinant host may be capable of producing mogrol, for example because the recombinant host expresses one or more enzymes of the mogrol biosynthesis pathway. Alternatively, mogrol may be provided to said recombinant host for example in the cultivation medium.

Said enzyme or mixture of enzyme capable of catalyzing glycosylation of mogrol preferably comprises a Uridine-5'-diphospho (UDP) dependent glucosyltransferase (UGT). In particular, it is preferred that step V comprises use of a UGT.

Thus, step V may include incubating mogrol with at least one Uridine-5'-diphospho (UDP) dependent glucosyltransferase (UGT) to produce a mogroside compound (e.g., mogroside I E1, mogroside I A1, mogroside II E, mogroside III A2, mogroside III, mogroside IV, mogroside V, or a mogroside compound glycosylated at C24-OH).

The UGT may for example be selected from the group consisting of 73C3, 73C6, 85C2, 73C5, and 73E1. The UGT may also be UGT73C3 of SEQ ID NO:21, UGT73C6 of SEQ ID NO:23, UGT85C2 of SEQ ID NO:25, UGT73C5 of SEQ ID NO: 22, UGT73E1 of SEQ ID NO:24 or a functional homologue of any of the aforementioned sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.

The UGT may also be selected from the group consisting of UGT98, UGT1495, UGT1817, UGT5914, UGT8468 and UGT10391. The UGT may also be UGT98 of SEQ ID NO:53, UGT1495 encoded by SEQ ID NO:27, UGT1817 encoded by SEQ ID NO:28, UGT5914 encoded by SEQ ID NO:30, UGT8468 encoded by SEQ ID NO:31 and UGT10391 encoded by SEQ ID NO:32 or a functional homologue of any of the aforementioned sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith at the amino acid level.

When the methods are performed in vitro the UGTs can for example be recombinantly

produced or can be in a cell lysate of a recombinant host. This document also features a method of producing a mogroside compound, wherein the method includes contacting mogrol with a cell lysate prepared from a recombinant host expressing a UGT to produce a mogroside compound (e.g., mogroside I E1, mogroside I A1, mogroside II E, mogroside III A2, mogroside III, mogroside IV, mogroside V, or a mogroside compound glycosylated at C24-OH). The UGT can be any of the above mentioned UGTs.

This document provides methods and materials for glycosylating mogrol using one or more Uridine-5'-diphospho (UDP) dependent glucosyltransferases (UGTs). As indicated below, at least five UGTs have been identified that glycosylate the aglycone mogrol. Each of the UGTs identified herein are in glycosyltransferase family I. Thus, in one preferred embodiment the UGT is a UGT in glycosyltransferase family I.

UGTs 73C3, 73C6, 85C2 and 73E1 are capable of catalyzing glycosylation at the C24-OH position of mogrol or mogroside (UGT#2 in FIG. 4). Accordingly, in methods of the invention wherein the mogroside to be produced comprises a glycosylation at the C24-OH position then at least one UGT may be UGT73C3 of SEQ ID NO:21, UGT73C6 of SEQ ID NO:23, UGT85C2 of SEQ ID NO:25 or UGT73E1 of SEQ ID NO:24 or a functional homologue of any of the aforementioned sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.

UGT73C5 is capable of catalyzing glycosylation at both the C3-OH of mogrol and mogroside (UGT#1 in FIG. 4) and C24-OH position (UGT#2). Accordingly, in methods of the invention wherein the mogroside to be produced comprises a glycosylation at the C24-OH position and/or a glycosylation at the C3-OH position, then at least one UGT may be UGT73C5 of SEQ ID NO:22 or a functional homologue of any of the aforementioned sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.

UGTs 73C3, 73C5, and 73C6 are from *Arabidopsis thaliana*. UGT 73E1 and 85C2 are from *Stevia rebaudiana*. The amino acid sequences of UGTs 73C3, 73C5, 73C6, 73E1, and 85C2 are provided herein as SEQ ID NOs: 21-25, respectively). Thus, UGTs 73C3, 73C6, 85C2, or 73E1 can be used to produce mogroside I E1 from mogrol, and

UGT73C5 can be used to produce mogroside I A1 from mogrol. Mogroside I E1 can be converted to mogroside II E using UGT73C5. Mogroside I A1 can be converted to mogroside II E using UGTs 73C3, 73C6, 85C2, or 73E1.

- 5 In one preferred embodiment of the invention at least one UGT is UGT1576 of SEQ ID NO:48 or a functional homologue of any of the aforementioned sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith. This is in particular the case in embodiments of the invention, wherein the mogroside to be produced comprises a glycosylation at the
- 10 C24-OH position, because UGT1576 is a glycosyltransferase with mogrol 24-OH UDP-glycosyltransferase activity.

- In one preferred embodiment of the invention at least one UGT is UGT98 of SEQ ID NO:53 or a functional homologue thereof sharing at least 70%, such as at least 80%,
- 15 for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith. This is in particular the case in embodiments of the invention, wherein the mogroside to be produced comprises a 1,2 glucosylation and a 1,6 glycosylation of the glucose at position C-24 to form mogroside III A1.

- 20 In one preferred embodiment of the invention at least one UGT is UGT SK98 of SEQ ID NO:50 or a functional homologue of any of the aforementioned sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith. This is in particular the case in embodiments of the invention, wherein the mogroside to be produced comprise a 1,2 glycosylation of
- 25 the glucose at position C-24 to form mogroside II A.

- As shown in FIG. 4, three enzymatic glycosylations convert mogroside II E into mogroside V or 11-Oxo-mogroside V. First, two glucoses are attached with 1,6-bonds to the two glucose molecules already present in mogroside II E. Second, another
- 30 glucose is added to the C24-bound glucose, with a 1,2 bond. Mogroside IV is an intermediate in which the 1,2-bound glucose is missing at the C24-bound glucose. In siamenoside this glucose is present, but the 1,6-bound glucose at the C3-bound glucose is missing. 11-Oxo-mogroside V is identical to mogroside V, only the 11-OH is oxidized. One or more of the following UGTs can be used to convert mogroside II E to
- 35 mogroside IV, mogroside V, 11-oxo-mogroside V, and siamenoside I: UGT98, UGT1495, UGT1817, UGT3494, UGT5914, UGT8468, UGT10391, UGT11789,

UGT11999, UGT13679 and UGT15423 (SEQ ID NOs: 26-36, respectively) or functional . For example, one or more of UGT98, UGT1495, UGT1817, UGT5914, UGT8468 and UGT10391 can be used to produce mogroside IV, mogroside V, 11-oxo-mogroside V, or siamenoside I.

5

In one embodiment of the invention step V comprises one or more of the following steps:

- a) Glucosylation of mogrol at C24 to form mogroside I A1
- b) 1,6 glucosylation of the C24 bound glucose of mogroside I A1 to form mogroside II A
- c) 1,2 glucosylation of the C24 bound glucose of mogroside IIa to form mogroside III A1
- d) Glucosylation of mogroside III A1 at the C3 to form siamenoside 1
- e) 1,6 glucosylation of the C3 bound glucose of siamenoside 1 to form mogroside V

15

These steps may each be catalyzed by a UGT capable of catalyzing said step. Thus, for example step a) may for example be catalyzed by UGT1576 of SEQ ID NO:48 or a functional homologue of any of the aforementioned sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith. Step b) may for example be catalyzed by UGT98 of SEQ ID NO:53 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith. Step c) may for example be catalyzed by UGT98 of SEQ ID NO:53, UGT SK98 of SEQ ID NO:50 or a functional homologue of any of the aforementioned sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith. Step d) may for example be catalyzed by UGT73C5 of SEQ ID NO:22 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith. Step e) may for example be catalyzed by UGT of the UGT91 family. For example step e) may be catalyzed by UGT98 of SEQ ID NO:53 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.

20

25

30

Activity of the UGTs can be assessed *in vitro*. For example, an *in vitro* UGT reaction mixture can include UGT enzyme, 4X Tris buffer, substrate (250 μ M), UDPglucose (750 μ M) and 1% alkaline phosphatase, in a total reaction volume of about 50 μ l. The reactions can be performed in sterilized 96 well plates, and incubated overnight at 30°C. After the incubation, 25 μ L of DMSO can be added to each reaction and the reaction plates centrifuged for 5 min. Samples can be taken from each well, filtered, and then analyzed via LC-MS.

10 Production of Polypeptides

As described herein above, the methods of the invention may be performed in *in vitro* or *in vivo*. In embodiments of the invention where the methods are performed *in vitro* one or more of the enzymes to be used in the methods may be prepared using any conventional method for producing polypeptides.

Thus, enzymes, such as synthases, hydrolyases, UGTs and CYP450 polypeptides described herein can be produced using any method. For example, enzymes, such as synthases, hydrolyases, UGT or CYP450 polypeptides can be produced by chemical synthesis. Alternatively, enzymes, such as synthases, hydrolyases, UGT or CYP450 polypeptides described herein can be produced by standard recombinant technology using heterologous expression vectors encoding enzymes, such as synthases, hydrolyases, UGT or CYP450 polypeptides. Expression vectors can be introduced into host cells (e.g., by transformation or transfection) for expression of the encoded polypeptide, which then optionally can be purified or partly purified. Crude extracts comprising the enzymes may also be used with the methods of the invention.

Expression systems that can be used for small or large scale production of enzymes, such as synthases, hydrolyases, UGT and CYP450 polypeptides include, without limitation, microorganisms such as bacteria (e.g., *E. coli* and *B. subtilis*) transformed with recombinant DNA, such as bacteriophage DNA, plasmid DNA, or cosmid DNA expression vectors containing the nucleic acid molecules described herein, or yeast (e.g., *S. cerevisiae* or *S. pombe*) transformed with recombinant yeast expression vectors containing the nucleic acid molecules described herein. Useful expression systems also include insect cell systems infected with recombinant virus expression vectors (e.g., baculovirus) containing the nucleic acid molecules described herein, or plant cell systems infected with recombinant virus expression vectors (e.g., tobacco

mosaic virus) or transformed with recombinant plasmid expression vectors (e.g., Ti plasmid) containing the nucleic acid molecules described herein. Enzymes, such as synthases, hydrolyases, UGT or CYP450 polypeptides also can be produced using mammalian expression system harboring recombinant expression constructs
5 containing promoters derived from the genome of mammalian cells (e.g., the metallothionein promoter) or from mammalian viruses (e.g., the adenovirus late promoter and the cytomegalovirus promoter), along with the nucleic acids described herein. Enzymes, such as synthases, hydrolyases, UGT or CYP450 polypeptides to be used with the methods of the invention may have an N-terminal or C-terminal tag as
10 discussed below.

This document also provides isolated nucleic acids encoding the enzymes to be used in each of steps Ia, Ib, IIa, IIb, IIIa, IIIb, IVa, IVb and V described herein above, such as synthases, hydrolyases, UGT or CYP450 polypeptides. An "isolated nucleic acid" refers
15 to a nucleic acid that is separated from other nucleic acid molecules that are present in a genome, including nucleic acids that normally flank one or both sides of the nucleic acid in a genome. The term "isolated" as used herein with respect to nucleic acids also includes any non-naturally-occurring nucleic acid sequence, since such non-naturally-occurring sequences are not found in nature and do not have immediately contiguous
20 sequences in a naturally-occurring genome. Thus, the isolated nucleic acid may be cDNA encoding any of the enzymes to be used with the methods of the invention.

An isolated nucleic acid can be, for example, a DNA molecule, provided one of the nucleic acid sequences normally found immediately flanking that DNA molecule in a
25 naturally-occurring genome is removed or absent. Thus, an isolated nucleic acid includes, without limitation, a DNA molecule that exists as a separate molecule (e.g., a chemically synthesized nucleic acid, or a cDNA or genomic DNA fragment produced by PCR or restriction endonuclease treatment) independent of other sequences as well as DNA that is incorporated into a vector, an autonomously replicating plasmid, a virus
30 (e.g., any paramyxovirus, retrovirus, lentivirus, adenovirus, or herpes virus), or into the genomic DNA of a prokaryote or eukaryote. In addition, an isolated nucleic acid can include an engineered nucleic acid such as a DNA molecule that is part of a hybrid or fusion nucleic acid. A nucleic acid existing among hundreds to millions of other nucleic acids within, for example, cDNA libraries or genomic libraries, or gel slices containing a
35 genomic DNA restriction digest, is not considered an isolated nucleic acid.

In some embodiments, a nucleic acid sequence encoding an enzyme to be used with the methods of the invention, such as synthases, hydrolyases, UGT or CYP450 polypeptides can include a tag sequence that encodes a "tag" designed to facilitate subsequent manipulation (e.g., to facilitate purification or detection), secretion, or localization of the encoded polypeptide. Tag sequences can be inserted in the nucleic acid sequence encoding the enzyme, such that the encoded tag is located at either the carboxyl or amino terminus of the enzyme. Non-limiting examples of encoded tags include green fluorescent protein (GFP), glutathione S transferase (GST), HIS tag, and Flag™ tag (Kodak, New Haven, CT). Other examples of tags include a chloroplast transit peptide, a mitochondrial transit peptide, an amyloplast peptide, signal peptide, or a secretion tag.

Functional homologs

Functional homologs of the polypeptides described above are also suitable for use in the methods and recombinant hosts described herein. A functional homolog is a polypeptide that has sequence similarity to a reference polypeptide, and that carries out one or more of the biochemical or physiological function(s) of the reference polypeptide. Thus, functional homologues of the enzymes described herein are polypeptides that have sequence similarity to the reference enzyme, and which are capable of catalyzing the same step or part of a step of the methods of the invention as the reference enzyme.

In general it is preferred that functional homologues share at least some degree of sequence identity with the reference polypeptide. Thus, it is preferred that a functional homologues of any of the polypeptides described herein shares at least 70%, such as at least 75%, such as at least 80%, for example at least 85%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.

Amino acid sequence identity requires identical amino acid sequences between two aligned sequences. Thus, a candidate sequence sharing 80% amino acid identity with a reference sequence, requires that, following alignment, 80% of the amino acids in the candidate sequence are identical to the corresponding amino acids in the reference sequence. Identity according to the present invention is determined by aid of computer analysis, such as, without limitations, the ClustalW computer alignment program

(Higgins D., Thompson J., Gibson T., Thompson J.D., Higgins D.G., Gibson T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673-4680), and the default parameters suggested therein. The ClustalW software is available from as a ClustalW WWW Service at the European Bioinformatics Institute <http://www.ebi.ac.uk/clustalw>. Using this program with its default settings, the mature (bioactive) part of a query and a reference polypeptide are aligned. The number of fully conserved residues are counted and divided by the length of the reference polypeptide. The sequence identity is determined over the entire length of the reference polypeptide

A functional homolog and the reference polypeptide may be natural occurring polypeptides, and the sequence similarity may be due to convergent or divergent evolutionary events. As such, functional homologs are sometimes designated in the literature as homologs, or orthologs, or paralogs. Variants of a naturally occurring functional homolog, such as polypeptides encoded by mutants of a wild type coding sequence, may themselves be functional homologs. Functional homologs can also be created via site-directed mutagenesis of the coding sequence for a polypeptide, or by combining domains from the coding sequences for different naturally-occurring polypeptides ("domain swapping"). Techniques for modifying genes encoding functional homologues of an enzyme to be used with the methods of the invention, such as synthases, hydrolyases, UGT or CYP450 polypeptides described herein are known and include, *inter alia*, directed evolution techniques, site-directed mutagenesis techniques and random mutagenesis techniques, and can be useful to increase specific activity of a polypeptide, alter substrate specificity, alter expression levels, alter subcellular location, or modify polypeptide:polypeptide interactions in a desired manner. Such modified polypeptides are considered functional homologs. The term "functional homolog" is sometimes applied to the nucleic acid that encodes a functionally homologous polypeptide.

Functional homologs can be identified by analysis of nucleotide and polypeptide sequence alignments. For example, performing a query on a database of nucleotide or polypeptide sequences can identify homologs of enzymes to be used with the methods of the invention, such as synthases, hydrolyases, UGT or CYP450 polypeptides. Sequence analysis can involve BLAST, Reciprocal BLAST, or PSI-BLAST analysis of

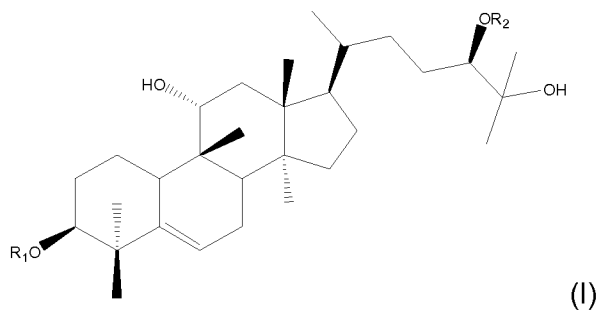
nonredundant databases using one of the sequences identified herein encoding an enzyme to be used with the methods of the invention, such as synthases, hydrolyases, UGT amino acid sequence as the reference sequence. Amino acid sequence is, in some instances, deduced from the nucleotide sequence. Those polypeptides in the database that have greater than 40% sequence identity are candidates for further evaluation for suitability as synthases, hydrolyases, UGT or CYP450 polypeptide. Amino acid sequence similarity allows for conservative amino acid substitutions, such as substitution of one hydrophobic residue for another or substitution of one polar residue for another. If desired, manual inspection of such candidates can be carried out in order to narrow the number of candidates to be further evaluated. Manual inspection can be performed by selecting those candidates that appear to have domains present in enzymes to be used with the methods of the invention, such as synthases, hydrolyases, UGT or CYP450 polypeptides, *e.g.*, conserved functional domains. Conserved regions can be identified by locating a region within the primary amino acid sequence of a polypeptide that is a repeated sequence, forms some secondary structure (*e.g.*, helices and beta sheets), establishes positively or negatively charged domains, or represents a protein motif or domain. See, *e.g.*, the Pfam web site describing consensus sequences for a variety of protein motifs and domains on the World Wide Web at sanger.ac.uk/Software/Pfam/ and pfam.janelia.org/. The information included at the Pfam database is described in Sonnhammer *et al.*, Nucl. Acids Res., 26:320-322 (1998); Sonnhammer *et al.*, Proteins, 28:405-420 (1997); and Bateman *et al.*, Nucl. Acids Res., 27:260-262 (1999). Conserved regions also can be determined by aligning sequences of the same or related polypeptides from closely related species. Closely related species preferably are from the same family. In some embodiments, alignment of sequences from two different species is adequate. Typically, polypeptides that exhibit at least about 40% amino acid sequence identity are useful to identify conserved regions. Conserved regions of related polypeptides exhibit at least 45% amino acid sequence identity (*e.g.*, at least 50%, at least 60%, at least 70%, at least 80%, or at least 90% amino acid sequence identity). In some embodiments, a conserved region exhibits at least 92%, 94%, 96%, 98%, or 99% amino acid sequence identity. Sequence identity can be determined as set forth above.

Mogrosides

The present invention relates to methods for producing mogrosides and materials for use in such methods. The term "mogroside" as used herein refers to mogrol

glycosylated at one or more positions. In particular, mogrosides according to the present invention may be mogrol glycosylated with one or more glucose residues at the positions 3 and/or 24. It is less preferred that mogrosides are glycosylated at the 11 and 25 positions. Mogrol is a compound of formula I provided below, wherein both R_1 and R_2 are $-H$.

It is preferred that the mogroside is a compound of the following formula I:



wherein R_1 and R_2 independently are $-H$, mono-glucoside, di-glucoside, tri-glucoside, and at least one of R_1 and R_2 is not $-H$.

In particular the mogroside may be one the mogrosides described in Table 1 herein below.

Table 1 Mogrosides of formula I

Name	R_1	R_2
mogroside V	Glc6-Glc-	Glc6-Glc2-Glc
siamenoside I	Glc-	Glc6-Glc2-Glc-
mogroside IV	Glc6-Glc-	Glc2-Glc-
mogroside IV A	Glc6-Glc-	Glc6-Glc-
mogroside III	Glc-	Glc6-Glc-
mogroside III A1	H	Glc6-Glc2-Glc-
mogroside III A2	Glc6-Glc-	Glc-
mogroside III E	Glc-	Glc2-Glc-
mogroside II A	H	Glc2-Glc-
mogroside II A1	H	Glc6-Glc-
mogroside II A2	Glc6-Glc-	H
mogroside II E	Glc-	Glc-

mogroside I A1	H	Glc-
mogroside I E1	Glc-	H

Glc = glucose

- 5 Mogroside I A1 may sometimes be referred to as mogroside Ib. Mogroside I E1 may sometimes be referred to as mogroside Ia. Mogroside II E may sometimes be referred to as mogroside II. Mogroside III A2 may sometimes be referred to as mogroside IIIa. Mogroside III may sometimes be referred to as mogroside IIIb. This alternative nomenclature is for example used in USSN 61/733,220.

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

Sequence listing

SEQ ID NO:1	Amino acid sequence of <i>C. pepo</i> cucurbitadienol synthase
SEQ ID NO:2	Amino acid sequence of C-terminal portion of <i>Siraitia grosvenorii</i> cucurbitadienol synthase
SEQ ID NO:3	DNA sequence encoding CYP533 (nucleotide sequence of CYP533 gene)
SEQ ID NO:4	DNA sequence encoding CYP937 (nucleotide sequence of CYP937 gene)
SEQ ID NO:5	DNA sequence encoding CYP1798 (nucleotide sequence of CYP1798 gene)
SEQ ID NO:6	DNA sequence encoding CYP1994 (nucleotide sequence of CYP1994 gene)
SEQ ID NO:7	DNA sequence encoding CYP2048 (nucleotide sequence of CYP2048 gene)
SEQ ID NO:8	DNA sequence encoding CYP2740 (nucleotide sequence of CYP2740 gene)
SEQ ID NO:9	DNA sequence encoding CYP3404 (nucleotide sequence of CYP3404 gene)
SEQ ID NO:10	DNA sequence encoding CYP3968 (nucleotide sequence of CYP3968 gene)

SEQ ID NO:11	DNA sequence encoding CYP4112 (nucleotide sequence of CYP4112 gene)
SEQ ID NO:12	DNA sequence encoding CYP4149 (nucleotide sequence of CYP4149 gene)
SEQ ID NO:13	DNA sequence encoding CYP4491 (nucleotide sequence of CYP4491 gene)
SEQ ID NO:14	DNA sequence encoding CYP5491 (nucleotide sequence of CYP5491 gene)
SEQ ID NO:15	DNA sequence encoding CYP6479 (nucleotide sequence of CYP6479 gene)
SEQ ID NO:16	DNA sequence encoding CYP7604 (nucleotide sequence of CYP7604 gene)
SEQ ID NO:17	DNA sequence encoding CYP8224 (nucleotide sequence of CYP8224 gene)
SEQ ID NO:18	DNA sequence encoding CYP8728 (nucleotide sequence of CYP8728 gene)
SEQ ID NO:19	DNA sequence encoding CYP10020 (nucleotide sequence of CYP10020 gene)
SEQ ID NO:20	DNA sequence encoding CYP10285 (nucleotide sequence of CYP10285 gene)
SEQ ID NO:21	Amino acid sequence of UGT73C3
SEQ ID NO:22	Amino acid sequence of UGT73C5
SEQ ID NO:23	Amino acid sequence of UGT73C6
SEQ ID NO:24	Amino acid sequence of UGT73E1
SEQ ID NO:25	Amino acid sequence of UGT85C2
SEQ ID NO:26	Nucleotide sequence encoding <i>Siraitia grosvenorii</i> UGT98
SEQ ID NO: 27	Nucleotide sequence encoding <i>Siraitia grosvenorii</i> UGT1495
SEQ ID NO:28	Nucleotide sequence encoding <i>Siraitia grosvenorii</i> UGT1817
SEQ ID NO:29	Partial gene sequence - nucleotide sequence encoding fragment of <i>Siraitia grosvenorii</i> UGT3494
SEQ ID NO:30	Nucleotide sequence encoding <i>Siraitia grosvenorii</i> UGT5914
SEQ ID NO:31	Nucleotide sequence encoding <i>Siraitia grosvenorii</i> UGT8468
SEQ ID NO:32	Nucleotide sequence encoding <i>Siraitia grosvenorii</i> UGT10391
SEQ ID NO:33	Partial gene sequence - nucleotide sequence encoding

	fragment of <i>Siraitia grosvenorii</i> UGT11789
SEQ ID NO:34	Partial gene sequence - nucleotide sequence encoding fragment of <i>Siraitia grosvenorii</i> UGT11999
SEQ ID NO:35	Partial gene sequence - Nucleotide sequence encoding fragment of <i>Siraitia grosvenorii</i> UGT13679
SEQ ID NO:36	Partial gene sequence - Nucleotide sequence encoding fragment of <i>Siraitia grosvenorii</i> UGT15423
SEQ ID NO:37	DNA sequence encoding <i>S. grosvenorii</i> Epoxide hydrolase 1 codon optimised for expression in <i>S. cerevisiae</i>
SEQ ID NO:38	Amino acid sequence of <i>S. grosvenorii</i> Epoxide hydrolase 1
SEQ ID NO:39	DNA sequence encoding <i>S. grosvenorii</i> Epoxide hydrolase 2 codon optimised for expression in <i>S. cerevisiae</i>
SEQ ID NO:40	Amino acid sequence of <i>S. grosvenorii</i> Epoxide hydrolase 2
SEQ ID NO:41	DNA sequence encoding CYP10969 (nucleotide sequence of CYP10969 gene)
SEQ ID NO:42	DNA sequence encoding <i>Siraitia grosvenorii</i> cucurbitadienol synthase codon optimized for expression in <i>S. cerevisiae</i>
SEQ ID NO:43	Amino acid sequence of <i>Siraitia grosvenorii</i> cucurbitadienol synthase
SEQ ID NO:44	Amino acid sequence of <i>S. grosvenorii</i> CYP5491
SEQ ID NO:45	DNA sequence encoding <i>S. grosvenorii</i> CPR4497
SEQ ID NO:46	Amino acid sequence of <i>S. grosvenorii</i> CPR4497
SEQ ID NO:47	DNA sequence encoding <i>S. grosvenorii</i> UGT1576
SEQ ID NO:48	Amino acid sequence of <i>S. grosvenorii</i> UGT1576
SEQ ID NO:49	DNA sequence encoding <i>S. grosvenorii</i> UGT SK98
SEQ ID NO:50	Amino acid sequence of <i>S. grosvenorii</i> UGT SK98
SEQ ID NO:51	DNA sequence encoding <i>S. grosvenorii</i> UGT98
SEQ ID NO:52	DNA sequence encoding <i>S. grosvenorii</i> UGT98 codon optimised for expression in <i>S. cerevisiae</i>
SEQ ID NO:53	Amino acid sequence of <i>S. grosvenorii</i> UGT98
SEQ ID NO:54	Amino acid sequence of <i>S. cerevisiae</i> squalene epoxidase encoded by the ERG1 gene
SEQ ID NO:55	Amino acid sequence of <i>S. cerevisiae</i> lanosterol synthase encoded by the ERG7 gene

Examples

Example 1 - Purification of mogroside V

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

Example 2 - Enzymatic synthesis of mogrol from mogroside V

Mogroside V (300 mg) was dissolved in 0.1M sodium acetate buffer (pH 4.5, 100mL),
15 and crude pectinase from *Aspergillus niger* (25mL, Sigma P2736) was added. The mixture was stirred at 50°C for 48 hours. The reaction mixture was extracted with ethyl acetate (2x100ml). The organic extract was dried under vacuum then purified with preparative HPLC. Pure mogrol (40 mg) was obtained and its structure confirmed by NMR and mass spectroscopy. See FIG. 6.

20

Example 3 - Enzymatic synthesis of mogrol 3-O-glucoside (mogroside I E1) and mogrol 24-O-glucoside (mogroside I A1) from mogroside V

Mogroside V (300 mg) was dissolved in 0.1M sodium acetate buffer (pH 4.5, 100ml),
and crude pectinase from *Aspergillus niger* (25ml, Sigma P2736) was added. The
25 mixture was stirred at 50°C for 6.5 hours. The reaction mixture was extracted with ethyl acetate (2x100ml). The organic extract was dried under vacuum then purified with preparative HPLC. Pure mogroside I E1 (11.0 mg) and mogroside I A1 (8.0 mg) were obtained. Their structures were confirmed by NMR and mass spectroscopy. See FIG. 6.

30

Example 4 - In vitro UGT screening and reactions

In vitro reactions of mogrol with a panel of 230 UGT enzymes were performed and the products were analyzed with LC-MS. The *in vitro* UGT reaction mixtures included 4X Tris buffer, mogrol (250 µM), UDP-glucose (750 µM) and 1% alkaline phosphatase.
35 Five µl of each partially purified UGT enzyme or crude enzyme extract was added to

the reaction, and the reaction volume brought to 50 µl with water. The reactions were incubated overnight at 30°C and performed in sterilized 96 well plates. After the incubation, 25 µL of DMSO were added into each reaction and the reaction plates were centrifuged for 5 min. Forty µL samples were taken from each well and filtered, and
5 were used for LC-MS analysis. UGTs 73C3, 73C6 and 85C2 were found to convert all the mogrol substrate to mogroside I A1. UGT 73C5 makes both mogroside I E1 and I A1. In the reaction with UGT 20 73E1, although the reaction was not complete, mogroside I A1 was found as the major product, together with a new glycosylated mogrol (neither mogroside I E1 nor I A1; exact mass shown as a mogroside I,
10 presumably caused by a glycosylation event on C11-OH).

Example 5 – Identifying the monk fruit cucurbitadienol synthase

The gene in monk fruit that codes for cucurbitadienol synthase is CirCS, and the partial gene sequence covering 338 of the supposedly 764 amino acids was identified by
15 doing a tBLASTn analysis of the assembled data with a query cucurbitadienol synthase from *Cucurbita pepo* (accession number BAD34645.1, SEQ ID NO:1). The partial CirCS is 97.5% identical to the *C. pepo* gene at the protein level (SEQ ID NO:2; from residues 515 to 764 of SEQ ID NO:1).

Example 6 – Identifying monk fruit candidate genes for P450 enzymes catalyzing formation of mogrol from cucurbitadienol

A pathway from cucurbitadienol to mogrol has been proposed by Tang *et al.*, *BMC Genomics*, 12, 343 (2011). The intermediates cucurbitadienol and mogrol exist in the fruit as they have been isolated as minor products. See Ukiya, *et al.*, *J. Agric. Food*
25 *Chem.* 50, 6710-6715 (2002). Glycoside intermediates exist in both 11-hydroxy and 11-oxo series, and gradually change from mogroside I to mogroside V as fruits ripen, which indicates that the triterpene core is fully oxidized by P450 enzymes before the subsequent glycosylations. According to the scheme proposed by Tang *et al.*, three independent cytochrome P450 enzyme-catalyzed oxidations results in mogrol
30 formation from cucurbitadienol (lower route in FIG. 4). The present inventors have found that the proposed primary reaction is highly unlikely. It is therefore submitted that the route may involve epoxidation by one cytochrome P450 enzyme, followed by a spontaneous or enzyme catalyzed hydration, and another P450 enzyme-catalyzed oxidation (visualized in the upper route in FIG. 4), or the route may comprise similar

steps in another order as shown in fig. 10A. The present inventors also propose another route starting from dioxido-squalene, which is shown in fig. 10B.

To identify the most likely candidate P450 genes from monk fruit, a BLAST database was made consisting of the polypeptide sequences of the 239 public domain *Arabidopsis thaliana* cytochrome P450 enzymes, representing most known enzyme subfamilies and variations. The sequences were used in a tBLASTn (translated nucleotide database) analysis of the assembled monk fruit transcriptome data to identify all sequences with a homology to any of the database query sequences with an E value of 10E-10 or lower. Seventy-two sequences were identified. Typically, the ability to assemble full or long gene lengths of expressed sequence tags in a transcriptome study means that many sequence tags of the gene in question were present. In the current experiment, this indicates that the gene was highly expressed in the monk fruit tissue and thus has a high probability of being a candidate for one of the two P450 enzymes of interest. Of the 72 sequences, 18 were full length or almost full length. The assembled genes were designated CYP533, CYP937, CYP1798, CYP1994, CYP2048, CYP2740, CYP3404, CYP3968, CYP4112, CYP4149, CYP4491, CYP5491, CYP6479, CYP7604, CYP8224, CYP8728, CYP10020, and CYP10285.

These are candidate genes for two P450 enzymes involved in catalyzing conversion of cucurbitadienol into mogrol. Full length gene sequences were amplified by PCR for the gene contigs CYP533, CYP937, CYP1798, CYP1994, CYP2740, CYP4112, CYP4149, CYP4491, CYP5491, CYP7604, CYP8224, and CYP10285, using monk fruit leaf genomic DNA or root cDNA and sequence overlap extension technology to remove resident intron sequences. The nucleotide sequences of CYP533, CYP937, CYP1798, CYP1994, CYP2048, CYP2740, CYP3404, CYP3968, CYP4112, CYP4149, CYP4491, CYP5491, CYP6479, CYP7604, CYP8224, CYP8728, CYP10020, and CYP10285 are provided as SEQ ID NOs: 3-20, respectively.

Example 7 – Identifying monk fruit candidate genes for glycosyltransferase enzymes catalyzing formation of mogroside V, 11-Oxo-mogroside V, mogroside IV, mogrosides III A2 and b and siamenoside from mogroside II E

Three enzymatic glycosylations are needed to convert mogroside II E into mogroside V or 11-Oxo-mogroside V. Two glucoses are attached with 1,6-bonds to the two glucose

molecules already present in mogroside II E. This may be done by one UGT enzyme. Another glucose is added to the C24-bound glucose, with a 1,2 bond. Mogroside IV is an intermediate in which the 1,6-bound glucose is missing at the C24-bound glucose. In siamenoside this glucose is present, but the 1,6-bound glucose at the C3-bound glucose is missing. 11-Oxo-mogroside V is identical to mogroside V, only the 11-OH is oxidized. See, FIG. 1 for the structures of mogroside IV, mogroside V, 11-Oxo-mogroside V, and siamenoside.

To identify all possible UGT genes in the assembled monk fruit transcriptome data, a database was assembled consisting of the polypeptide sequences of glycosyltransferases (UGTs) of all known sub-families, a total of 160 sequences. A tBLASTn analysis was performed between this database and the assembled monk fruit data. UGTs performing di-glycosylation (i.e., attaching a sugar to another sugar which in turn resides on an aglycon) invariably come from Family 1 UGT sub-families 76, 79, 91 or 94 (with the latter three forming the "orthology group 8"). While sub-family 76 enzymes usually make 1,3 bonds, orthology group 8 UGTs always make 1,2 or 1,6 bonds.

Sequences were identified that showed more homology to orthology group 8 enzymes than to any other UGT enzymes or any non-UGT genes. Thus 11 contigs were identified as likely candidates to encode the two glycosyltransferase genes needed to turn mogroside II E into mogroside V: UGT98, UGT1495, UGT1817, UGT3494, UGT5914, UGT8468, UGT10391, UGT11789, UGT11999, UGT13679 and UGT15423 (SEQ ID NOs: 26-36, respectively). Of these we were able to amplify by PCR UGT98, UGT1495, UGT1817, UGT5914, UGT8468 and UGT10391, using monk fruit leaf genomic DNA or root cDNA. The amplified genes were inserted into *E. coli* expression plasmid vectors.

The enzymes are expressed and purified on nickel columns. *In vitro* reactions of mogroside I A1, I E1 and II E with the panel of 6 purified UGT enzymes are performed and the products analyzed with LC-MS. The *in vitro* UGT reaction mixtures include 4X Tris buffer, substrate (250 μ M), UDP-glucose (750 μ M) and 1% alkaline phosphatase. Five μ l of each partially purified UGT enzyme are added to the reaction, and the reaction volume brought to 50 μ l with water. The reactions are incubated overnight at 30°C and performed in sterilized 96 well plates. After the incubation, 25 μ L of DMSO

are added into each reaction and the reaction plates are centrifuged for 5 min. Forty μ L samples are taken from each well and filtered, and then analyzed via LC-MS. The UGT catalyzing the 1,6-bond formation as well as the enzyme catalyzing the 1,2-bond formation are identified based on the LC-MS analysis.

5

Example 8 – Using eYAC technology to identify the cytochrome P450 enzymes responsible for turning cucurbitadienol into mogrol

eYAC gene expression technology was used to identify the active cytochrome P450 enzymes within a collection of candidate genes. The following genes were inserted into “Entry vectors” (a collection of plasmid vectors containing gene promoter and terminator sequences which have different nucleotide sequence but which are all 30 repressible by the addition of the amino acid methionine): the *Cucurbita pepo* cucurbitadienol synthase gene, CYP533 (SEQ ID NO:3), CYP937 (SEQ ID NO:4), CYP1798 (SEQ ID NO:5), CYP1994 (SEQ ID NO:6), CYP2740 (SEQ ID NO:8), CYP4112 (SEQ ID NO:11), CYP4149 (SEQ ID NO:12), CYP4491 (SEQ ID NO:13), CYP5491 (SEQ ID NO:14), CYP7604 (SEQ ID NO:16), CYP8224 (SEQ ID NO:17), and CYP10285 (SEQ ID NO:20), the two cytochrome P450 oxidoreductase (CPR) genes from *Arabidopsis thaliana* (ATR1 and ATR2), a CPR from *Stevia rebaudiana* (CPR8), a CPR isolated from monk fruit, and the glycosyltransferases UGT73C5 (SEQ ID NO: 22) and UGT73C6 (SEQ ID NO:23) from *A. thaliana* and UGT85C2 (SEQ ID NO:25) from *S. rebaudiana*.

The expression cassettes from these 17 plasmids are excised after an *Ascl* + *SrfI* digestion, purified and then randomly concatenated in ligation reactions to create artificial yeast chromosomes (“eYACs”). From 30 to 200 μ g of DNA are prepared from 10 each of the cassette-containing entry vectors and the cassettes are randomly concatenated into eYACs by ligation with T4 ligase in a 3 hour reaction. The success of the concatenation reaction is assessed by the viscosity of the reaction mixture, since concatenated DNA is highly viscous. DNA fragments (“arms”) containing a centromere, two telomeres and the LEU2 and TRP1 selection markers are added to the end of the 15 concatenated expression cassettes, thereby creating functional eYACs. The eYACs are transformed into transformation-competent spheroplasts of yeast strain *erg7* by zymolyase digestion of the yeast cell wall, followed by treatment with a CaCl_2 /PEG buffer, making the spheroplasts permeable to large molecules such as eYACs. After

transformation, the yeast spheroplasts are embedded in a “noble agar” based solid growth medium, in which regeneration of the cell wall can take place. Colonies appear from 4-8 days after inoculation. The regeneration medium lacks the amino acids leucine and tryptophan, thus selecting for the presence of double-armed eYACs in the yeast cells. One hundred transformants are selected and analyzed for production of mogrosides I E1, I A1 and II E, LC-MS (Liquid Chromatography-coupled Mass Spectrometry (Triple Quadropole)).

Each transformant is re-streaked and tested for yeast strain markers and the genetic presence of both arms of the eYAC, i.e., the LEU2 and TRP1 markers. More than 95% of the transformants has the correct genotype. Each transformant is given a CEY designation number. Initially, 48 CEYs are grown in 50 ml of Synthetic Complete medium (SC) in 100 ml Ehrlenmeyer flasks, without methionine, so as to induce gene expression from the eYACs, and without tryptophan, leucine and histidine, so as to counter-select for loss of eYACs. The cultures have a start density corresponding to an OD600 of 0.25, and they are inoculated for 48 h at 30 C, with slow shaking (150 rpm). After 24 hours, 1 ml supernatant from each culture is collected and subjected to LC-MS analysis. Positive CEYs (i.e., those producing any of the mogrosides assayed for) are subjected to PCR analysis in order to assess which CYP genes are present on the harbored eYAC and thus identifying the mogrol pathway P450 enzymes.

Example 9

Boosting mogrol pathway precursor availability

The background strain used in this study is the BY4742 strain deleted for the *TRP1* gene. This strain is called EFSC301. To increase the availability of oxidosqualene and dioxidosqualene in this laboratory yeast strain, the promoter of the endogenous *ERG7* gene was displaced by a PCR fragment consisting of the Nurseothricin marker (NatMX) and the CUP1 copper inducible promoter. This displacement gives low transcription and thereby low expression of *ERG7* when the yeast strain is grown in normal growth medium like Synthetic Complete medium (SC medium). *ERG7* encode the lanosterol synthase and lowered expression is known to result in accumulation of oxidosqualene and dioxidosqualene in baker's yeast. Oxidosqualene is generally the precursor of triterpenoids and possibly a precursor of the mogrol pathway. To further increase oxidosqualene and dioxidosqualene availability the squalene epoxidase encoded by *ERG1* was overexpressed by a GPD1 promoter from a gene copy integrated into the

genome. The sequence of the squalene epoxidase encoded by ERG1 is provided herein as SEQ ID NO:54. Furthermore a truncated copy of the yeast HMG reductase (tHMG1) was expressed from a genomically integrated gene copy, with expression from a GPD1 promoter. The resulting strain is called EFSC3027.

5

The successful boosting of oxidosqualene and dioxidosqualene production in the strain EFSC3027 was demonstrated by production of tetrahydroxysqualene when either one of two soluble *S. grosvenorii* epoxide hydrolases was expressed in this strain. One epoxide hydrolase was *S. grosvenorii* Epoxide hydrolase 1 of SEQ ID NO:38. In order to prepare yeast expressing this a *S. cerevisiae* codon optimized *S. grosvenorii* Epoxide hydrolase 1 gene sequence of SEQ ID NO:37 was introduced in the yeast strain EFSC3027. The other epoxide hydrolase was *S. grosvenorii* Epoxide hydrolase 2 of SEQ ID NO:40. In order to prepare yeast expressing this a *S. cerevisiae* codon optimized *S. grosvenorii* Epoxide hydrolase 1 gene sequence of SEQ ID NO:39 was introduced in the yeast strain EFSC3027. Figure 7 shows the LC-MS mass peak 501 corresponding to the proton plus Na⁺ adduct of tetrahydroxysqualene in a sample from yeast strain EFSC3027 transformed with a plasmid expressing *S. grosvenorii* Epoxide hydrolase 2. Tetrahydroxysqualene is made by the hydrolysis of 2,3 and 22,23 epoxide bonds of dioxidosqualene. No accumulation of tetrahydroxy squalene was detected in the EFSC301 background strain. Samples were made by boiling culture aliquots in 50% DMSO and then pelleting of cell material by centrifugation. Supernatants were then measured by ESI LC-MS.

A similar system for boosting oxidosqualene availability for β -amyrin production was described by Kirby, J *et al* in FEBS Journal **275** (2008) 1852-1859

Example 10

Production of cucurbitadienol in yeast strain EFSC3027

When a *S. cerevisiae* codon optimized gene copy of the *Siraitia grosvenorii* cucurbitadienol synthase of Accession No HQ128567 (sequence provided herein as SEQ ID NO:42) is integrated into the genome of yeast strain EFSC3027 and transcription of this gene is driven by the GPD1 promoter, the expression of the cucurbitadienol synthase results in production of cucurbitadienol in the yeast strain in amounts that are easily detectable by ESI LC-MS (see fig. 8). The amino acid sequence of *Siraitia grosvenorii* cucurbitadienol synthase is provided herein as SEQ

35

ID NO:43. The strain comprising SEQ ID NO:42 producing cucurbitadienol is called EFSC3498. Yeast strains were grown at 30°C for 5 days in synthetic complete medium containing 2% glucose, and cucurbitadienol was extracted by boiling a culture sample in 50% ethanol/20% KOH for 5 minutes followed by extraction with an equal volume of hexane and then evaporation of hexane and resuspension of dried extract in methanol. Fig. 8 shows the LC-MS chromatogram peak of lanosterol in EFSC3027 (upper frame) and the LC-MS chromatogram peaks of cucurbitadienol and lanosterol in EFSC3498 (lower frame). The peak corresponding to lanosterol shows a retention time of ~8.05 whereas the peak corresponding to cucurbitadienol has a retention time of 7.85. Both lanosterol and cucurbitadienol shows a mass in the LC-MS chromatogram of 409.4 (proton adduct minus one H₂O)

Example 11

Production of oxo and hydroxy cucurbitadienol in *S. cerevisiae*

When the cucurbitadienol producing yeast strain EFSC3498 (prepared as described in Example 10) is transformed with two plasmids, one expressing the *S. grosvenorii* CYP5491 from a TEF1 promoter, the other expressing the *S. grosvenorii* CPR4497 also from a TEF1 promoter (DNA sequence encoding CPR4497 provided as SEQ ID NO:14) three conspicuous peaks emerge (see Fig.9 for LC-MS chromatogram peaks). The amino acid sequence of *S. grosvenorii* CYP5491 is provided herein as SEQ ID NO:44 and the DNA sequence encoding *S. grosvenorii* CYP5491 is provided as SEQ ID NO:14. The amino acid sequence of *S. grosvenorii* CPR4497 is provided herein as SEQ ID NO:46 and the DNA sequence encoding *S. grosvenorii* CPR4497 is provided as SEQ ID NO:45. The upper frame in Fig.9 shows the LC-MS chromatogram with the three peaks made when CYP5491 and CPR4497 are expressed in EFSC3498, while the three lower frames show the fragmentation spectrum of these three peaks. CYP5491 is 99% identical to acc. no. HQ128570 and HQ128571 at both the amino acid and nucleotide sequence level. The masses of the 3 peaks (443.38, 441.37 and 457.36) correspond in weight to proton adducts of hydroxylated cucurbitadienol, oxo cucurbitadienol and hydroxy plus oxo cucurbitadienol respectively. Without being bound by theory it is believed that the hydroxylated cucurbitadienol (protonated mass 443.38) and oxidated cucurbitadienol (protonated mass 441.37) is 11-hydroxy-cucurbitadienol and 11-oxo-cucurbitadienol, respectively. The peak that corresponds to both oxo plus hydroxy cucurbitadienol (protonated mass 457.36) could be 11-oxo-24,25 epoxy cucurbitadienol, formed, either from cyclization of dioxidosqualene by the

cucurbitadienol synthase and 11 hydroxylation by CYP5491 (Fig. 10B) or by CYP5491 being multifunctional, making both the 11-oxidation and the 24,25-epoxidation (Fig. 10A).

5 Example 12

Glycosylation of mogrol in *S. cerevisiae* by expression of *S. grosvenorii* UGTs

UGTs 98, SK98 and 1576 were cloned from *S. grosvenorii* leaf and root cDNA by primers designed from fruit gene contigs assembled from illumina sequencing data. *S. grosvenorii* was purchased from Horizon Herbs, LLC, United States. The DNA sequence and protein sequence of UGT98 are provided herein as SEQ ID NO:51 and 10 53, respectively, whereas a SEQ ID NO:52 provides a DNA sequence encoding UGT98 codon optimised for expression in *S. cerevisiae*. The DNA sequence and protein sequence of UGTSK98 are provided herein as SEQ ID NO:49 and 50, respectively, The DNA sequence and protein sequence of UGT1576 are provided 15 herein as SEQ ID NO:47 and 48, respectively. Yeast strain EFSC1563 has a deletion of the EXG1 gene and of the EXG2 gene both encoding and exo-1,3-beta-Glucanase. When yeast strain **EFSC1563 (EFSC301 *exg1 exg2*)** is transformed with a plasmid expressing UGT1576 driven by a GPD1 promoter and fed mogrol to a concentration in the growth medium of 10-100uM, a clear formation of mogroside I A1 is detected by 20 LC-MS (Fig11B). The produced mogroside I A1 shows the same retention time as the reference mogroside I A1 in the LC-MS analysis. Figure 11A shows the LC-MS chromatogram of reference mogroside I A1, while 11B shows the peak from a sample of EFSC1563 expressing UGT1576 in a culture fed 50uM mogrol. These data show that the UGT1576 gene encodes a glycosyltransferase with mogrol 24-OH UDP- 25 glycosyltransferase activity. Samples were made by mixing culture aliquot 1:1 with DMSO followed by boiling (80°C) for 5 minutes and pelleting by centrifugation. Supernatants were then subjected to ESI LC-MS.

When UGTs 98 and SK98 cloned into yeast expression plasmids with expression from GPD1 promoters are transformed into EFSC1563 without co-transformation of a 30 UGT1576 expression plasmid, no conversion of fed mogrol is detected. In contrast, co-expression of UGT98 or UGT SK98 with UGT1576 in EFSC1563 fed with mogrol results in further glycosylation of mogroside I A1. UGT SK98 co-expressed with UGT1576 results in production of di-glycosylated mogrol (mogroside II A, Fig.12A)^[LAD1], while co-expression with UGT98 results in di and tri-glycosylated mogrol 35 (middle and lower frames, Fig.12B^[LAD2]). The di-glycosylated mogrol that is formed by

both UGT98 and UGT SK98 has a different retention time than mogroside II E and mogroside II A1 during LC-MS, making it likely that it is mogroside II A. This means that both UGT98 and UGT SK98 can catalyse a 1,2 glycosylation of the glucose of mogroside I A1. UGT98 appears to be multifunctional, catalysing 1,2 glycosylation of mogroside I A1 resulting in production of mogroside II A, followed by what may be a 1,6 glycosylation of mogroside II A to form mogroside III A1 (Figure 12B). We believe that UGT98 catalyses 1,6 glycosylation of mogroside II because mogrol itself is not glycosylated by the UGT98. It is therefore likely that the UGT98 is multifunctional, being both a 1,2 and 1,6 UDP-glucose glycosyl transferase of the 24-glucose moiety of mogrosides. UGTs 98 and SK98 belong to the UGT91 family of UDP-glucose glycosyltransferases and members of this family are known to be 1,2 and 1,6 glycosyltransferases.

Figure 13 schematically summarizes the glycosylation reactions from mogrol to mogroside III A1

15

OTHER EMBODIMENTS

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

20

WHAT IS CLAIMED IS:

1. A method of producing a mogroside compound, said method comprising one or more of the following steps:

- 5 Step IIb. Contacting dioxido-squalene with an enzyme capable of catalyzing conversion of Dioxido-squalene -> 24,25 epoxy cucurbitadienol;
- Step IVa. Contacting 11-hydroxy-cucurbitadienol with an enzyme or mixture of enzymes capable of catalysing conversion of 11-hydroxy-cucurbitadienol -> mogrol;
- 10 Step IVb. Contacting 11-hydroxy-24,25 epoxy cucurbitadienol with an enzyme capable of catalysing conversion of 11-hydroxy-24,25 epoxy cucurbitadienol -> mogrol;
- Step V. Contacting mogrol and/or glucosylated mogrol with an enzyme or mixture of enzymes capable of catalysing glucosylation of mogrol and/or
- 15 glucosylated mogrol to form mogroside, wherein said enzyme or mixture of enzyme comprises a UGT selected from the group consisting of UGT1576, UGT98 and UGT SK98.

2. The method according to claim 1, wherein the method comprises the steps of

- 20 Step IIa. Contacting oxido-squalene with an enzyme capable of catalysing conversion of Oxido-squalene -> cucurbitadienol
- Step IIIa. Contacting said cucurbitadienol with an enzyme capable of catalysing hydroxylation of Cucurbitadienol -> 11-hydroxy-cucurbitadienol
- Step IVa. Contacting said 11-hydroxy-cucurbitadienol with an enzyme capable of
- 25 catalysing conversion of 11-hydroxy-cucurbitadienol -> mogrol
- Step V Contacting said mogrol with an enzyme or mixture of enzymes capable of catalysing glucosylation of mogrol -> mogroside
- and optionally isolating said mogroside.

30 3. The method according to claim 1, wherein the method comprises the steps of

- Step IIb. Contacting dioxido-squalene with an enzyme capable of catalysing conversion of dioxido-squalene -> 24,25 epoxy cucurbitadienol
- Step IIIb. Contacting said 24,25 epoxy cucurbitadienol with an enzyme capable of catalysing hydroxylation of 24,25 epoxy cucurbitadienol -> 11-hydroxy-
- 35 24,25 epoxy cucurbitadienol

- Step IVb. Contacting said 11-hydroxy-24,25 epoxy cucurbitadienol with an enzyme capable of catalysing conversion of 11-hydroxy-24,25 epoxy cucurbitadienol -> mogrol
- 5 Step V Contacting said mogrol with an enzyme or mixture of enzymes capable of catalysing glucosylation of mogrol -> mogroside and optionally isolating said mogroside.
4. The method according to claim 1, wherein the method comprises the step of
- 10 Step IIIa. Contacting cucurbitadienol with an enzyme capable of catalysing hydroxylation of Cucurbitadienol -> 11-hydroxy-cucurbitadienol
- Step IVa. Contacting said 11-hydroxy-cucurbitadienol with an enzyme capable of catalysing conversion of 11-hydroxy-cucurbitadienol -> mogrol
- Step V Contacting said mogrol with an enzyme or mixture of enzymes capable of catalysing glucosylation of mogrol -> mogroside
- 15 and optionally isolating said mogroside.
5. The method according to claim 1, wherein the method comprises the steps of
- Step IIIb. Contacting 24,25 epoxy cucurbitadienol with an enzyme capable of catalysing hydroxylation of 24,25 epoxy cucurbitadienol -> 11-hydroxy-24,25 epoxy cucurbitadienol
- 20 Step IVb. Contacting said 11-hydroxy-24,25 epoxy cucurbitadienol with an enzyme capable of catalysing conversion of 11-hydroxy-24,25 epoxy cucurbitadienol -> mogrol
- Step V Contacting said mogrol with an enzyme or mixture of enzymes capable of catalysing glucosylation of mogrol -> mogroside
- 25 and optionally isolating said mogroside.
6. The method according to claim 1, wherein the methods comprises the step of
- Step V Contacting mogrol and/or glucosylated mogrol with an enzyme or mixture of enzymes capable of catalysing glucosylation or mogrol and/or glucosylated mogrol to form mogroside, wherein said enzyme or mixture of enzyme comprises a UGT selected from the group consisting of UGT1576, UGT98 and UGT SK98.
- 30

7. The method according to any one of claims 1 to 6, wherein said mogroside is glycosylated at C24.

5 8. The method according to any one of claims 1 to 7, wherein said mogroside is glycosylated at C3.

9. The method according to any one of claims 1 to 8, wherein the mogroside is selected from the group consisting of mogroside I E1, mogroside I A1, mogroside II E, mogroside III A2, mogroside III, mogroside IV and mogroside V.

10

10. The method according to any one of claims 1 to 8, wherein the mogroside is selected from the group consisting of mogroside I A1, mogroside II A, mogroside III A1, siamenoside and mogroside V.

15 11. A recombinant host cell comprising one or more of the following heterologous nucleic acids:

IIb. Heterologous nucleic acid encoding an enzyme capable of catalyzing conversion of Dioxido-squalene -> 24,25 epoxy cucurbitadienol;

20 IVa Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing conversion of 11-hydroxy-cucurbitadienol -> mogrol;

IVb Heterologous nucleic acid encoding an enzyme capable of catalysing conversion of 11-hydroxy-24,25 epoxy cucurbitadienol -> mogrol;

25 V. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing glucosylation of mogrol and/or glucosylated mogrol to form mogroside, wherein said enzyme or mixture of enzyme comprises a UGT selected from the group consisting of UGT1576, UGT98 and UGT SK98.

30 12. The recombinant host according to claim 11, wherein the recombinant host cell comprises the following heterologous nucleic acids:

IIa. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IIa (Oxido-squalene -> cucurbitadienol)

35 IIIa. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IIIa (Cucurbitadienol -> 11-hydroxy-cucurbitadienol)

- IVa. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IVa (11-hydroxy-cucurbitadienol -> mogrol)
- V Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step V (mogrol -> mogroside).

5

13. The recombinant host according to claim 11, wherein the recombinant host comprises the following heterologous nucleic acids:

- IIIa. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IIIa (Cucurbitadienol -> 11-hydroxy-cucurbitadienol)
- IVa. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IVa (11-hydroxy-cucurbitadienol -> mogrol)
- V Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step V (mogrol -> mogroside)

10

15

14. The recombinant host according to claim 11, wherein the recombinant host comprises the following heterologous nucleic acids:

- IIb. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IIb (Dioxido-squalene -> 24,25 epoxy cucurbitadienol)
- IIIb. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IIIb (24,25 epoxy cucurbitadienol -> 11-hydroxy-24,25 epoxy cucurbitadienol)
- IVb. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IVb (11-hydroxy-24,25 epoxy cucurbitadienol -> mogrol)
- V Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step V (mogrol -> mogroside)

20

25

30

15. The recombinant host according to claim 11, wherein the recombinant host comprises the following heterologous nucleic acids:

- IIIb. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IIIb (24,25 epoxy cucurbitadienol -> 11-hydroxy-24,25 epoxy cucurbitadienol)

- IVb. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step IVb (11-hydroxy-24,25 epoxy cucurbitadienol -> mogrol)
- 5 V Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing Step V (mogrol -> mogroside)

16. The recombinant host according to claim 11, wherein the recombinant host comprises the following heterologous nucleic acids:

- 10 V. Heterologous nucleic acid(s) encoding an enzyme or mixture of enzymes capable of catalysing glucosylation of mogrol and/or glucosylated mogrol to form mogroside, wherein said enzyme or mixture of enzyme comprises a UGT selected from the group consisting of UGT1576, UGT98 and UGT SK98.

- 15 17. The recombinant host according to any one of claims 11 to 16, wherein the recombinant host furthermore comprises a heterologous nucleic acid encoding a squalene epoxidase.

- 20 18. The recombinant host according to any one of claims 11 to 7, wherein the recombinant host has been modified to reduced expression of a lanosterol synthase.

19. The method according to any one of claims 1 to 10, wherein the method comprises
- i) providing a recombinant host according to any one of claims 11 to 18
 - ii) cultivating said recombinant host in a cultivation medium
 - 25 iii) optionally isolating said mogroside from the cultivation medium and/or the recombinant host.

20. The method according to any one of claims 1 to 10 or the recombinant host according to any one of claims 11 to 18, wherein the enzyme capable of catalyzing
- 30 conversion of Oxido-squalene -> cucurbitadienol is cucurbitadienol synthase of SEQ ID NO:43 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.

21. The method according to any one of claims 1 to 10 or the recombinant host according to any one of claims 11 to 18, wherein the enzyme capable of catalysing conversion of Dioxido-squalene -> 24,25 epoxy cucurbitadienol is cucurbitadienol synthase of SEQ ID NO:43 or a functional homologue thereof sharing at least 70%,
5 such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.

22. The method according to any one of claims 1 to 10 or the recombinant host according to any one of claims 11 to 18, wherein the enzyme capable of catalysing
10 conversion of Cucurbitadienol -> 11-hydroxy-cucurbitadienol is CYP5491 of SEQ ID NO:44 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.

23. The method according to any one of claims 1 to 10 or the recombinant host according to any one of claims 11 to 18, wherein the enzyme capable of catalysing conversion of 24,25 epoxy cucurbitadienol -> 11-hydroxy-24,25 epoxy cucurbitadienol is CYP5491 of SEQ ID NO:44 or a functional homologue thereof sharing at least 70%,
15 such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.
20

24. The method according to any one of claims 1 to 10 or the recombinant host according to any one of claims 11 to 18, wherein the enzyme or mixture of enzymes capable of catalysing conversion of 11-hydroxy-cucurbitadienol -> mogrol comprises a
25 polypeptide selected from the group consisting of Epoxide hydrolase 1 of SEQ ID NO:38, Epoxide hydrolase 2 of SEQ ID NO:40 and functional homologues of the aforementioned sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.

25. The method according to any one of claims 1 to 10 or the recombinant host according to any one of claims 11 to 18, wherein the enzyme capable of catalysing conversion of 11-hydroxy-24,25 epoxy cucurbitadienol -> mogrol is a polypeptide selected from the group consisting of Epoxide hydrolase 1 of SEQ ID NO:38, Epoxide hydrolase 2 of SEQ ID NO:40 and functional homologues of the aforementioned
30

sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.

26. The method according to any one of claims 1 to 10 or the recombinant host
- 5 according to any one of claims 11 to 18, wherein the enzyme or mixture of enzymes capable of catalysing conversion of mogrol -> mogroside comprises one or more UGTs selected from the group consisting of UGT1576 of SEQ ID NO:48, UGT98 of SEQ ID NO:53, UGT SK98 of SEQ ID NO:50 and functional homologues of the aforementioned
- 10 sharing at least 70%, such as at least 80%, for example at least 90%, such as at least 95%, for example at least 98% sequence identity therewith.
- .