Use of octaketide synthases to produce kermesic acid and flavokermesic acid

Møller, Birger Lindberg ; Madsen,, Bjørn; Stærk, Dan; Okkels, Finn Thyge; Andersen-Ranberg, Johan; Kongstad, Kenneth Thermann; Binderup, Kim; Bennedsen, Mads; Nafisi, Majse; Khorsand-Jamal, Paiman; Kannangara, Rubini Maya; Mortensen, Uffe Hasbro; Thrane, Ulf; Frandsen, Rasmus John Normand

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 December 2016 (15.12.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/198564 A1

(51) International Patent Classification:
C12P 7/26 (2006.01) *C12P 17/06* (2006.01)
C12P 7/66 (2006.01)

(21) International Application Number:
PCT/EP2016/063242

(22) International Filing Date:
10 June 2016 (10.06.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
15171375.7 10 June 2015 (10.06.2015) EP

(71) Applicants: **DANMARKS TEKNISKE UNIVERSITET** [DK/DK]; Anker Engelundsvej 1, Bygning 101A, 2800 Lyngby (DK). **KØBENHAVNS UNIVERSITET** [DK/DK]; Nørregade 10, 1165 København K (DK).

(72) Inventors: **MØLLER, Birger Lindberg**; Kongstedvej 5, 2700 Brønshøj (DK). **MADSEN, Bjørn**; Tordenskjoldsvej 9, 3000 Helsingør (DK). **STÆRK, Dan**; Perlestensvej 7, 3540 Lynge (DK). **OKKELS, Finn Thyge**; Kongemarken 11, 4000 Roskilde (DK). **ANDERSEN-RANBERG, Johan**; 1535 Prince St., Berkeley, California 94703 (US). **KONGSTAD, Kenneth Thermann**; Ingerslevsgade 170, 3 sal, 1705 København V (DK). **BINDERUP, Kim**; Mosehøjvej 11, 2920 Charlottenlund (DK). **BENNEDSEN, Mads**; Hessemosevej 11, 3230 Græsted (DK). **NAFISI, Majse**; Byhøjen 28, 1tv, 2720 Vanløse (DK). **KHORSAND-JAMAL, Paiman**; Gedvad 3, 1. tv, 2800 Kgs. Lyngby (DK). **KANNANGARA, Rubini Maya**; Dronning Olgas Vej 34, 5tv, 2000 Frederiksberg (DK). **MORTENSEN, Uffe Hasbro**; Schleppegrellsgade 6, 4. th, 2200 København N (DK). **THRANE, Ulf**; Spurvevej 27, 3200 Helsinge (DK). **FRANDSEN, Rasmus John Normand**; Enevangen 55, 3450 Allerød (DK).

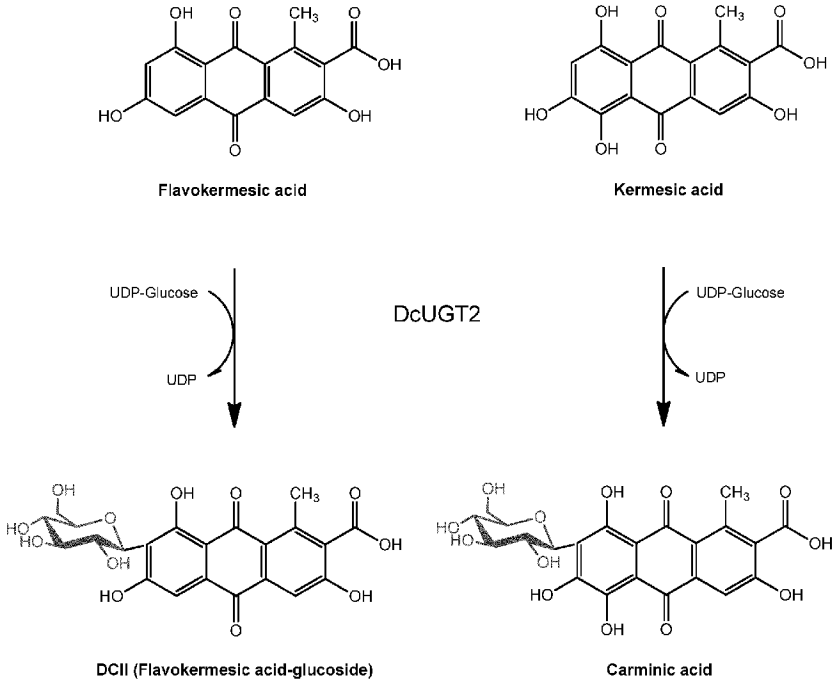
(74) Agent: **PLOUGMANN VINGTOFT A/S**; Rued Langgaards Vej 8, 2300 Copenhagen S (DK).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM,

[Continued on next page]

(54) Title: USE OF OCTAKETIDE SYNTHASES TO PRODUCE KERMESIC ACID AND FLAVOKERMESIC ACID

Figure 1



(57) Abstract: A method for producing an octaketide derived aromatic compound of interest (e.g. carminic acid), wherein the method comprises (I): heterologous expression of a recombinantly introduced Type III polyketide synthase (PKS) gene encoding an octaketide synthase (OKS) to obtain non-reduced octaketide in vivo within the recombinant host cell and (II): converting in vivo the non-reduced octaketide of step (I) into a C_{14}-C_{34} aromatic compound of interest (e.g. carminic acid).

WO 2016/198564 A1

DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

— *with information concerning incorporation by reference of missing parts and/or elements (Rule 20.6)*

USE OF OCTAKETIDE SYNTHASES TO PRODUCE KERMESIC ACID AND FLAVOKERMESIC ACID

FIELD OF THE INVENTION

The present invention relates to a method for producing an octaketide derived aromatic compound of interest (e.g. carminic acid), wherein the method comprises (I): heterologous expression of a recombinantly introduced Type III polyketide synthase (PKS) gene encoding an octaketide synthase (OKS) to obtain non-reduced octaketide *in vivo* within the recombinant host cell and (II): converting *in vivo* the non-reduced octaketide of step (I) into a C_{14}-C_{34} aromatic compound of interest (e.g. carminic acid).

BACKGROUND OF THE INVENTION

The natural pigment carminic acid is one of the most frequently used colorants of food, medicine, cosmetics and textiles.

Carminic acid is a colorant, which can be extracted from the female insect bodies of *Dactylopius coccus costa* (alternative name *Coccus cacti* L.). The insects live on *Nopalea coccinellifera, Opuntia fidus indica* and other plants of the family Cactaceae cultivated for instance in the desert areas of Mexico, Central and South America and Canary Islands. Depending on the pH the colorant may be a color in a spectrum from orange over red to purple and is generally known as cochineal or cochineal color. Carmine colorant is widely used in foods and beverages.

As known in the art *Porphyrophora polonica* is also producing carminic acid and was cultured for production of carminic acid in e.g. Poland.

In relation to current industrial relevant production, carminic acid is harvested by extraction from the insect's dried bodies with water or alcohol.

In order to try to resolve the problem of undesirable variations and price fluctuations - US5424421 (European Colour, published 1995) describes chemical synthesis of carminic acid by a route of synthesis involving different intermediates.

As discussed in e.g. WO2006/056585A1 (Chr. Hansen A/S), during the aqueous based extraction of carminic acid from the insect, an amount of insect protein is also released from the insect and will be contained in the color extract and it has been reported that the cochineal insect proteins could create some allergy related problems. In WO2006/056585A1 a special process to reduce the amount of insect protein from the insect extract solution is

described. However, the final produced color composition/product of WO2006/056585A1 will still comprise some amounts *Dactylopius coccus* costa insect proteins.

The structure of carminic acid is shown in Figure 1 herein. As can be seen from the Figure, it is a so-called *C*-glucoside (i.e. wherein the glucose is joined/conjugated to the aglucon by a carbon-carbon linkage).

As shown in Figure 1 herein, hydrolysis of the *C*-glucoside carminic acid can give glucose and the aglucon kermesic acid (KA).

The *in vivo* biosynthetic pathway of carminic acid in the insect (*Dactylopius coccus*) is currently not described in details. Accordingly, based on the prior art the skilled person does not know which compound is the aglucon during the *in vivo* biosynthetic production of carminic acid in *Dactylopius coccus*.

Analysis of *Dactylopius coccus* has shown that a broad range of compounds related to carminic acid are present in extracts from *Dactylopius coccus* and numerous of these compounds could in principle be glucosylated during the *in vivo* biosynthetic production of carminic acid.

For instance, in the article of *Stathopoulou* et al. (Analytica Chimica Acta 804 (2013) 264–272) six new anthraquinones were described in an extract from *D. coccus* and any of these six new anthraquinones (see e.g. figure 1 of the article) could in principle be the molecule which is glucosylated during the *in vivo* biosynthetic production of carminic acid in *Dactylopius coccus*.

Further, as known in the art the primary glucosylated compound formed during the *in vivo* biosynthetic production of the glucoside end product may be an unstable intermediate compound that will not be identified in an isolated extract from *Dactylopius coccus* as e.g. analyzed in the above discussed article of *Stathopoulou* et al.

As understood by the skilled person in the present context, based on the prior art, it could be speculated that a relevant primary glucosylated compound during the *in vivo* biosynthetic production of carminic acid in *Dactylopius coccus* could e.g. be an unstable intermediate polyketide compound with around the same number of carbon atoms as e.g. flavokermesic acid.

A herein relevant DNA or amino acid sequence of a glycosyltransferase involved in the *in vivo* insect (*Dactylopius coccus*) biosynthetic pathway of carminic acid is not explicitly described in the prior art.

As known in the art, for insects that accumulate low molecular weight chemicals the relevant biosynthetic pathway genes are sometimes not present in the insect genome.
For instance, some insects take up glycosides from the plants they feed on - see e.g. the article of Zagrobelny et al (Cyanogenic glucosides and plant-insect interactions; Phytochemistry. 2004 Feb;65(3):293-306) or the article of Geuder et al (Journal of Chemical Ecology, Vol. 23, No. 5, 1997).

Dactylopius coccus insects feed on cactus plants and it could be that *Dactylopius coccus* insects (like other insects) take up relevant glycosides from the cacti they feed on.

As known in the art, for insects that accumulate low molecular weight glycosides, the relevant biosynthetic pathway genes are sometimes found in the microorganisms living in the insects, see e.g. the article of *Genta* et al. (Potential role for gut microbiota in cell wall digestion and glucoside detoxification in Tenebrio molitor larvae), Journal of Insect Physiology 52 (2006) 593–601.

Accordingly, based on the prior art the skilled person could not know if the genome of *Dactylopius coccus* actually would comprise a gene encoding a glycosyltransferase involved in the *in vivo* biosynthetic pathway leading to carminic acid.

Polyketides are synthesized by a group of enzymes which commonly is referred to as polyketide synthases (PKS). All PKSs share the ability to catalyze Claisen condensation based fusion of acyl groups by the formation of carbon-carbon bonds with the release of carbon dioxide. This reaction is catalyzed by a beta-ketosynthase domain (KS). In addition to this domain/active site, synthesis can also depend on, but not exclusively, the action of Acyl-Carrier-Protein (ACP), Acyl-transferase (AT), Starter-Acyl-Transferase (SAT), Product Template (PT), ThioEsterase (TE), Chain Length Factor (CLF, also known as KSβ), Claisen CYClase (CYC), Ketoreductase (KR), dehydratase (DH), enoyl reductase (ER) and C-methyl transferase (Cmet). The substrates for polyketide synthesis are typically classified into starter and extender units, where the starter unit, including but not limited to acetyl-CoA is the first added unit of the growing polyketide chain; and extender units, e.g. but not exclusively malonyl-CoAs, are all subsequently added carbon-carbon units.
At the primary sequence level (amino acid sequence), secondary structure level (local fold), tertiary structure level (all over fold) and quaternary structure level (protein-protein interactions) the PKSs display a very large diversity, and are hence subdivided into different types.

Type I PKS systems are typically found in filamentous fungi and bacteria, where they are responsible for both the formation of aromatic, polyaromatic and reduced polyketides.

Members of the type I PKS possess several active sites on the same polypeptide chain and the individual enzyme is able to catalyze the repeated condensation of two-carbon units. The minimal set of domains in type I PKS includes KS, AT and ACP. The type I PKSs is further subdivided into modular PKSs and iterative PKSs, where iterative PKSs only possess a single copy of each active site type and reuse these repeatedly until the growing polyketide chain has reached its predetermined length. Type I iterative PKS that forms aromatic and polyaromatic compounds typically rely on the PT and CYC domain to direct folding of the formed non-reduced polyketide chain. Modular PKSs contain several copies of the same active sites, these are organized into repeated sequences of active sites which are called modules, each module is responsible for adding and modifying a single ketide unit. Each active site in the individual modules is only used once during synthesis of a single polyketide. Type I iterative PKS are typically found in fungi, while type I modular PKSs are typically found in bacteria.

Type II PKS systems are responsible for formation of aromatic and polyaromatic compounds in bacteria.

Type II PKSs are protein complexes where individual enzymes interact to form the functional PKS enzyme. The individual enzymes include activities for KS, CLF and ACP.

This type of PKS is characterized by being composed of multiple different enzymes that form a protein complex, which collectively is referred to as an active PKS. The type II PKSs form non-reduced polyketides that spontaneously folds into complex aromatic/cyclic compounds. However, in the bacterial systems folding of polyketide backbones is most often assisted/directed by different classes of enzymes, that act in trans (independent of the PKS enzyme) to promote a non-spontaneous fold. The involved enzyme classes are referred to as aromatases and cyclases. The biosynthesis of a single polyaromatic compound in these systems typically involves the successive action of multiple different aromatases/cyclases. The aromatases and cyclases can be divided into two groups based on which types of substrates they act on: where the first group only acts on linear polyketide chains and catalyzes formation of the first aromatic/cyclic group, the second group of enzymes only accepts substrates that include aromatic or cyclic groups (= products from the first group of aromatases/cyclases). It has proven impossible to functionally express type II PKS systems in a suitable production host (*E. coli*, *Bacillus*, yeast), likely due to the fact that these are multienzyme complexes which require a balanced expression level, and which may rely on unknown factors.

Type III PKSs generally only consist of a KS domain, which in the literature may e.g. be referred to as a KASIII or a Chalcone synthase domain that acts independently of the ACP domain. Type III PKS from bacteria, plant and fungi have been described.

Type III PKSs have long been known in plants, where they are responsible for formation of compounds such as flavonoids (pigments/anti-oxidants) and stilbenes, which are found in

many different plant species. The products of type III PKSs often spontaneously folds into complex aromatic/cyclic compounds.

The article of *Yu* et al. (2012) provides a review of Type III Polyketide synthases in natural product biosynthesis. The *Yu* et al. (2012) article reads: "Type III PKSs are selfcontained enzymes that form homodimers. Their single active site in each monomer catalyzes the priming, extension, and cyclization reactions iteratively to form polyketide products. Despite their structural simplicity, type III PKSs produce a wide array of compounds such as chalcones, pyrones, acridones, phloroglucinols, stilbenes, and resorcinolic lipids. In recent years, type III PKSs have drawn more attention due to their diverse products, wide distribution, relatively simple structures, and easy genetic manipulability. In this article, we will systematically discuss type III PKSs from plants, bacteria, and fungi as well as the recent progress in the type III PKS research."

In short, based on the prior art, the skilled person knows if a specific PKS of interest is a Type I, Type II or Type III PKS.

In addition to the protein structural and functional based classification of PKS systems, an alternative classification is based on the level of modifications found in the final polyketide product. Note that these modifications can either be introduced by the PKS itself or by post-acting enzymes. In this classification scheme the products are divided into two groups: (I) non-reduced and (II) reduced polyketides. The non-reduced type is characterized by the presence of ketone groups in the ketides (-CH2-CO-), originating from the starter or extender units, either as ketones or in the form of double bonds in aromatic groups. In reduced polyketides a single or all ketones have been reduced to alcohol (-CH2-CHOH-) groups by a KR domain/enzyme, or further to an alkene group (-C=C-) by a DH domain/enzyme, or even further to an alkane group (-CH2-CH2-) by an ER domain/enzyme. Based on these chemical features of the formed products the involved PKSs are categorized as either being a non-reducing PKS or a reducing PKS.

Folding of the formed polyketide chain into complex structures with cyclic motifs is typically a post-PKS enzyme guided and catalyzed process. The responsible enzymes belong to several different enzyme families, typically aromatases and/or cyclases. Fungal Type I iterative PKSs are special by posing a PT domain which is responsible for the formation of aromatic rings while CYC domains are responsible for product release coupled to formation of aromatic rings. The aromatases and cyclases acting on polyketides have been described from bacterial and plant systems. In addition, several examples exist where folding of the polyketide is a spontaneous process, e.g. flavonoids in plants.

PKSs have been isolated and functionally characterized from bacteria, fungi and plants. However, no PKS of animal origin has been described, and synthesis of polyketides in insects has in several instances been linked to the metabolic activity of endosymbiotic bacteria.

The article of *Tang*, Y. et al. (2004) describes that expression in the bacteria *Streptomyces coelicolor* CH999 strain, which contains chromosomal deletion affecting the entire Act gene cluster responsible for actinorhodin biosynthesis. The mini PKS (*Act* PKS = Act_KS, Act_CLF and Act_ACP), belonging to the type II PKSs, yields flavokermesic acid (FK) (called TMAC in bacterial articles) when combined with heterologous expression of the ZhuI aromatase/cyclase and ZhuJ cyclase from the *zhu* gene cluster in *Streptomyces* sp. R1128. Accordingly, this article describes recombinant introduction of a *Streptomyces* PKS gene into a *Streptomyces* host cell, so the PKS is not of a different genus than the host cell.

In Figure 2 herein is shown figure 5A of the *Tang*, Y. et al (2004) article. As can be seen in the figure and as further described in the article, the *Act* PKS (termed octaketide synthase (OKS)) creates a non-reduced octaketide and this octaketide is via the ZhuI aromatase/cyclase and ZhuJ cyclase converted into flavokermesic acid (FK) (called TMAC). The SEK4 and SEK4B compounds are also spontaneously produced (structures shown in Figure 2 may herein be termed shunt products).

In the plant *Aloe arborescens,* identified PKSs have been shown to produce polyketides of various lengths including octaketides, see e.g. Mizuuchi et al (2009) where it in figure 1 is illustrated that the octaketide synthases (OKSs) termed PKS4 and PKS5 may, by using malonyl-CoA as extender units, create a non-reduced octaketide. The SEK4 and SEK4B shunt compounds are also spontaneously formed.

The plant *Hypericum perforatum* (St. John's wort) also comprises octaketide synthases, see e.g. Karppinen et al (2008), where it is described that the PKS termed HpPKS2 was expressed in *E. coli*, followed by purification and *in vitro* biochemical characterization of the enzyme. In figure 1 of the article is illustrated that the PKS termed HpPKS2 creates a non-reduced octaketide (using acetyl-CoA as starter unit and malonyl-CoA as extender units) and the shunt products SEK4 and SEK4B are spontaneously formed.

The article of *Yu* et al. (2012) provides a review of Type III Polyketide synthases in natural product biosynthesis; the article reads on page 293: "Various type III PKSs have been engineered into *E. coli* to generate novel polyketides. The production of plant-specific curcuminoids has been reconstituted in *E. coli* by co-expressing CUS with phenylalanine ammonia-lyase from *Rhodotorula rubra* and 4-coumarate:CoA ligase (4CL) from

Lithospermum erythrorhizon". As explained in the article, the PKS termed "CUS" synthesizes a diketide-CoA and therefore CUS is not an octaketide synthase.

The article *Jadhav* et al (2014) describes that a type III hexaketide PKS from *Plumbago zeylanica* (PzPKS) was cloned and expressed in tobacco plants to study whether the transgenic tobacco plants expressing PzPKS synthesize the pharmacologically important polyketide, plumbagin.

In none of the above mentioned PKS related articles are discussed production of carminic acid.

Without being limited to theory, it is believed that the prior art does not describe that herein relevant type III PKS octaketide synthases (OKS) may be active *in vivo* in a heterologous production host cell of a different genus, e.g. a plant type III OKS may be able to create non-reduced octaketides *in vivo* in a heterologous production host cell, such as e.g. a recombinant *Aspergillus* production host cell.

The patent application PCT/EP2014/078540 was filed 18 December 2014. At the filing/priority date of the present patent application PCT/EP2014/078540 was not published. It describes a glycosyltransferase (GT) isolated from *Dactylopius coccus costa* insect which is capable of: (I): conjugating glucose to flavokermesic acid (FK); and/or (II): conjugating glucose to kermesic acid (KA) and use of this GT to e.g. make carminic acid.

PCT/EP2014/078540 does not directly and unambiguously describe herein discussed relevant non-reduced octaketides and/or polyketide synthases (PKS).

SUMMARY OF THE INVENTION

The problem to be solved by the present invention relates to the provision of a suitable biosynthetic pathway that may lead to carminic acid.

An advantage of the provision of such a suitable biosynthetic pathway as described herein is that it opens up the possibility for heterologous production (in e.g. *Aspergillus* or *yeast)* of carminic acid without the need to make an extraction from insects and thereby be able to make a carminic acid color composition/product that is free of unwanted *Dactylopius coccus costa* insect proteins.

One part of the solution relates to that the present inventors identified a *Dactylopius coccus* extract (including extracts of the endosymbionts present in *Dactylopius coccus*) with a herein relevant glycosyltransferase GT activity. As discussed herein, the present inventors

analyzed the GT and identified that it is capable of: (I): conjugating glucose to flavokermesic acid (FK); and/or (II): conjugating glucose to kermesic acid (KA). Accordingly, this GT can be used to e.g. make carminic acid.

The polynucleotide sequence encoding herein described isolated/cloned novel *Dactylopius coccus costa* glycosyltransferase is shown in SEQ ID NO: 1 herein and the polypeptide amino acid sequence is shown in SEQ ID NO: 2 herein.
The herein relevant glycosyltransferase enzyme of SEQ ID NO: 2 may herein be termed "DcUGT2" or simply "DcUGT".

Figure 1 shows a schematic presentation of the herein relevant glycosyltransferase activity of the herein described isolated/cloned DcUGT glycosyltransferase of SEQ ID NO:2 herein, as illustrated in the figure, it was found to be able to conjugate glucose to the aglycons flavokermesic acid (FK) and kermesic acid (KA).

Based on the prior art the skilled person does not know which compound is the primary glucosylated compound during the biosynthetic production of carminic acid *in vivo* in *Dactylopius coccus*.

The present inventors demonstrated that *Dactylopius coccus* contains a GT able to *C*-glycosylate flavokermesic acid (FK) and/or kermesic acid (KA). Accordingly, and without being limited to theory, it is plausible that flavokermesic acid (FK) and/or kermesic acid (KA) are suitable aglucons for e.g. *in vivo* heterologous biosynthesis of e.g. carminic acid.

Based on this knowledge, that it is plausible that flavokermesic acid (FK) and/or kermesic acid (KA) are suitable aglycons for e.g. *in vivo* heterologous biosynthesis of e.g. carminic acid, the present inventors identified that herein relevant glycosyltransferases may also be identified in *Aloe* plants, *Haworthia* plants and *Sorghum* or rice plants.

Having identified possible suitable aglucons, the present inventors could go back and try to identify suitable prior intermediate compounds that could be suitable for *in vivo* biosynthesis of the relevant aglucons (e.g. FK/KA).

As discussed in working Example 6 herein, the present inventors recombinantly expressed plant (*Aloe arborescens (AaOKS*) and/or *Hypericum perforatum (HpPKS2)*) type III polyketide synthases (PKS) octaketide synthases (OKS) in *Aspergillus nidulans* and identified that the OKS actually worked in the heterologous host, there were created non-reduced octaketides *in vivo* in the *Aspergillus nidulans* cells. The non-reduced octaketides are unstable and the identification that non-reduced octaketides were present *in vivo* was

verified by the accumulation of the shunt/degradation products SEK4 and SEK4B (see Figure 2 herein and above discussed prior art + Example 6 herein).

As discussed in the conclusion paragraph of working Example 6 herein, expression of plant type III PKS (*HpPKS2* or *AaOKS*) resulted in the production of different compounds including the compounds SEK4, SEK4B and flavokermesic acid (FK) *in vivo* in *Aspergillus nidulans*. Since there in this Example 6, were not inserted heterologous cyclases and/or aromatases into the *Aspergillus* strains and FK compound was identified, it indicates that the *Aspergillus* strains may comprise homologous cyclases and/or aromatases capable of converting non-reduced octaketide into FK compound *in vivo*.

As discussed in the conclusion paragraph of working Example 8 herein, the present inventors made a co-expression of the heterologous plant PKS (*AaOKS)* and glycosyltransferase *(DcUGT2)* in *Aspergillus nidulans* and it resulted in the *in vivo* production of carminic acid (CA) and DcII.

With respect to experimental work of the present inventors in relation to heterologous expression/production in *Nicotiana benthamiana* plant (a close relative of tobacco plant), as discussed in the conclusion paragraph of working Example 11 herein, the results of this Example 11 demonstrated that:

(i): The plant PKS *AaOKS* gene of *Aloe arborescens* was transiently expressed in *N. benthamiana* and *in vivo* this resulted in formation of SEK4 and SEK4B, which demonstrated that AaOKS can function as an active octaketide synthase *in vivo* in *N. benthamiana*;

(ii): Since no flavokermesic acid (FK) anthrone or FK could be observed when *AaOKS* was agroinfiltrated alone, *N. benthamiana* may lack endogenous enzymes to further metabolize the non-reduced octaketide into these compounds.

(iii): The *Streptomyces* sp. R1128 cyclase genes, *ZhuI* and *ZhuJ,* were co-agroinfiltrated (i.e. *in vivo* co-expressed) with *AaOKS* and *in vivo* production/accumulation of different compounds including FK was observed. Accordingly, heterologous expression of *Streptomyces* R1128 cyclase genes resulted in the *in vivo* production of different compounds including FK;

(iv): *In vivo* production of DcII and carminic acid (CA) was detected when *DcUGT2*_was co-expressed with *AaOKS*, *ZhuI* and *ZhuJ* in *N. benthamiana*.

Accordingly, the present inventors demonstrated:

(I): A recombinantly introduced Type III-like polyketide synthase (PKS) gene encoding an octaketide synthase (OKS) (such as e.g. *Aloe arborescens (AaOKS*) *and/or Hypericum perforatum (HpPKS2)*) and wherein the OKS is of a different genus than the host cell (such as e.g. a fungal *Aspergillus nidulans* or *Nicotiana benthamiana* plant host cell) is capable of

converting suitable starter and extender units into a non-reduced octaketide under suitable growth conditions and there is then *in vivo* produced the non-reduced octaketide; and

(II): It is possible to convert *in vivo* within the growing recombinant host cell the non-reduced octaketide of step (I) into a C_{14}-C_{34} aromatic aglycon compound of interest (e.g. flavokermesic acid (FK) or kermesic acid (KA)), wherein the aromatic aglycon compound of interest is not SEK4 and/or SEK4B (i.e. SEK4 and/or SEK4B are derived spontaneously from the non-reduced octaketide and may herein be termed shunt products);

(III): If the recombinant host cell also comprises a glycosyltransferase gene encoding a glycosyltransferase then the in step (II) produced aromatic aglycon compound of interest may be *in vivo* glycosylated by the glycosyltransferase to produce a C_{14}-C_{34} aromatic glycoside compound of interest (e.g. carminic acid (CA) or DcII)).

It is believed that above step (II) may by itself be seen as a novel significant contribution to the art by the present invention, since based on the prior art the skilled person could not know for sure that the in step (I) created non-reduced octaketide would in fact be "freely" available *in vivo* within the recombinant host cell of the different genus to actually be converted into a different aromatic aglycon compound of interest (e.g. flavokermesic acid (FK)) and not only the spontaneously produced SEK4 and/or SEK4B compounds.

As discussed herein, it may be preferred that *in vivo* conversion of the non-reduced octaketide is done via involvement of *in trans* acting (independent of the PKS enzyme) aromatases/cyclases.

As discussed above, the experimental work of the present inventors as discussed in working examples herein demonstrated that it is possible to make heterologous *in vivo* production of e.g. insect *Dactylopius costa* carminic acid (CA) in different recombinant production host cells such as fungal *Aspergillus nidulans* cells or *Nicotiana benthamiana* plant (a close relative of tobacco plant) cells.

As discussed above, today one may only get carminic acid (CA) by direct isolation from *Dactylopius* insect bodies or via chemical synthesis of carminic acid by a route involving different intermediates (US5,424,421).

Accordingly, it may be seen as a major contribution to the art that it is herein demonstrated that it is possible to make *in vivo* heterologous production of carminic acid (CA) in e.g. fungal *Aspergillus* cells or *Nicotiana* plant cells.

As understood by the skilled person in the present context, based on the novel technical information herein there is no reason to believe that it should not be possible to make *in vivo* heterologous recombinant cell production of different octaketide derived compounds of interest – such as e.g. a C_{14}-C_{34} aromatic compounds of interest, where examples of such

C_{14}-C_{34} aromatic compound could e.g. be Mutactin (see e.g. Figure 2 in above discussed Tang, Y. et al (2004) article); Emodin, Hypericin or Pseudohypericin (see e.g. Figure 1 in above discussed Karppinen et al (2008) article); or Barbaloin (see e.g. Figure 1 and 4 in above discussed Mizuuchi et al (2009) article); or carminic acid (CA).

As understood by the skilled person in the present context, in step (II) of the method of the first aspect as described herein, the specific types of *in vivo* created C_{14}-C_{34} aromatic aglycon specific compounds of interest (such e.g. flavokermesic acid (FK) or kermesic acid (KA)) will generally depend on the post-PKS enzymes (e.g. homologous or heterologous aromatases and/or cyclases) present within the used recombinant host cell.

In the prior art is described numerous different e.g. aromatases and/or cyclases that a skilled person may use to create specific C_{14}-C_{34} aromatic aglycon of interest in step (II) of the method of the first aspect as described herein.

Based on the results discussed herein, there is in Figure 3 herein shown an example of a model for a suitable biosynthetic pathway that may lead to carminic acid.

In short, the example of a biosynthetic pathway model comprises following steps:
(i): using a type III PKS octaketide synthase to create a non-reduced octaketide;
(ii): conversion of the non-reduced octaketide via *in trans* acting (independent of the PKS enzyme) aromatases/cyclases into flavokermesic acid (FK) and/or kermesic acid (KA); and
(iii): (1): the kermesic acid (KA) aglucon is glucosylated to yield carminic acid CA; or
(2): the FK is glucosylated to yield DcII, which is further hydroxylated to yield carminic acid (CA).

As discussed above and without being limited to theory, it is believed that the prior art does not describe that herein relevant type III PKS octaketide synthases (OKS) may be active *in vivo* in a heterologous production host cell of a different genus, e.g. a plant type III OKS may be able to create non-reduced octaketides *in vivo* in a heterologous production host cell such as e.g. a recombinant *Aspergillus* production host cell.

As discussed above, in the prior art it is known that non-reduced octaketides may *in vivo* be converted into different octaketide derived aromatic compounds of interest, the specific type of aromatic compound of interest will generally depend on the specific post-PKS enzymes (e.g. aromatases and/or cyclases) present *in vivo* in the host cell.
For instance, in Figure 1 of above discussed Karppinen et al (2008) article is shown that non-reduced octaketides may *in vivo* in *H. perforatum* be converted into the aromatic compounds Emodin anthrone and/or Emodin (an Anthraquinone).

In Figure 1 of above discussed *Mizuuchi* et al. (2009) article is shown that non-reduced octaketide may *in vi*vo in *A. arborescens* be converted into the aromatic octaketide anthrone barbaloin compound.

Accordingly, a first aspect of the present invention relates to a method for producing an octaketide derived aromatic compound of interest, wherein the method comprises following steps:

(I): contacting *in vivo* in a recombinant host cell comprising a recombinantly introduced Type III polyketide synthase (PKS) gene encoding an octaketide synthase (OKS) and wherein the OKS is of a different genus than the host cell:

(i): suitable starter unit and extender units with the recombinantly introduced OKS capable of converting the starter and extender units into a non-reduced octaketide under suitable conditions wherein there *in vivo* is produced the non-reduced octaketide; and

(II): converting *in vivo* within the recombinant host cell the non-reduced octaketide of step (I) into a C_{14}-C_{34} aromatic aglycon compound of interest, wherein the aromatic aglycon compound of interest is not SEK4 and/or SEK4B; and

(III): optionally, the recombinant host cell also comprises a glycosyltransferase gene encoding a glycosyltransferase and the in step (II) produced aromatic aglycon compound of interest is *in vivo* glycosylated by the glycosyltransferase to produce a C_{14}-C_{34} aromatic glycoside compound of interest; and

(IV): isolating the aromatic aglycon compound of interest of step (II) and/or isolating the aromatic glycoside compound of interest of step (III) to get a composition, wherein the composition comprises less than 1% w/w dry matter of recombinant host cell material.

As discussed above, the present inventors recombinantly expressed plant (*Aloe arborescens* and/or *Hypericum perforatum*) Type III polyketide synthases (PKS) octaketide synthases (OKS) in *Aspergillus nidulans* and identified that the OKS actually worked in the heterologous host. Accordingly, an example of a host cell in step (I) could e.g. be *Aspergillus nidulans* and an example of a Type III PKS/OKS of step (I) could e.g. be an OKS from *Aloe arborescens* and/or *Hypericum perforatum*.

As discussed above and as known in the art, non-reduced octaketides may *in vivo* be converted spontaneously into the SEK4 and SEK4B compounds (structures shown in Figure 2 may be termed shunt products).

As understood by the skilled person in the present context, the identification of SEK4 and/or SEK4B *in vivo* within the recombinant host cell demonstrates that non-reduced octaketide is present *in vivo*.

Accordingly and as understood by the skilled person, a Type III octaketide synthase (OKS) of step (I) may be defined as an OKS, which *in vitro* is capable of producing the octaketide-derived shunt products SEK4 and/or SEK4B.
An assay for such an *in vitro* analysis is described e.g. in working Example 9 herein.

As discussed above, the term Type III polyketide synthase (PKS) is well-known to the skilled person and the skilled person can therefore routinely identify if a PKS of interest is understood to be a Type III PKS and therefore not a Type I or a Type II PKS.

As discussed above, in the prior art it is known that non-reduced octaketides may *in vivo* be converted into different octaketide-derived aromatic compounds of interest and this will generally depend on the specific post-PKS enzymes (e.g. aromatases and/or cyclases) present *in vivo* in the host cell.
Accordingly and as understood by the skilled person, in the present context, the conversion in step (II) of the non-reduced octaketide into a C_{14}-C_{34} aromatic aglycon compound of interest is generally done by involvement of suitable post-PKS enzymes (e.g. aromatases and/or cyclases) present *in vivo* in the host cell.

The aromatic aglycon compound of interest of step (II) may in principle be any herein relevant C_{14}-C_{34} aromatic aglycon compound of interest, such as e.g. an aromatic anthraquinone compound or an aromatic anthrone compound (for further details see below).

As discussed above - the present inventors identified that flavokermesic acid (FK) was present in the *Aspergillus nidulans* cells comprising recombinantly introduced Type III OKS from *Aloe arborescens* and/or *Hypericum perforatum*.

Flavokermesic acid (FK) and kermesic acid (KA) both have 16 carbon (C) atoms, - they are understood to be examples of C_{16} aromatic aglycon compounds and as such examples of C_{14}-C_{34} aromatic aglycon compounds of interest in step (II).

As discussed herein, if the in step (IV) isolated aromatic aglycon compound of interest is FK this may e.g. *in vitro* via glucosylation be converted into DcII and thereafter into carminic acid (CA) (see e.g. Figure 3 herein).

As discussed herein, another optional alternative example could be that the host cell also *in vivo* expresses a herein relevant glycosyltransferase and there in step (III) *in vivo* in the host cell is produced e.g. DcII and/or carminic acid (CA). This may be seen as an example of optional step (III) of the first aspect.
Working Examples herein show preferred examples of this optional alternative.

DcII and/or carminic acid (CA) are both glucosides and the aglycon part (i.e. not including the glucose) of these compounds have 16 carbon (C) atoms.
Accordingly, both of these DcII and/or carminic acid (CA) compounds are herein understood to be examples of C_{16} aromatic glycoside compounds and as such examples of C_{14}-C_{34} aromatic glycoside compounds of interest in optional step (III).

As understood by the skilled person, in the present context, the isolating step (IV) essentially relates to a step to isolate/purify the aromatic compound of interest from recombinant host cell material (or cultivation/growth media if the host cell is e.g. a fungal/microorganism cell).

The present inventors identified that flavokermesic acid (FK) was present in the *Aspergillus nidulans* cells comprising recombinantly introduced Type III OKS from *Aloe arborescens* and/or *Hypericum perforatum*.

One may say that it is surprising that the prior art does not describe that herein relevant type III PKS octaketide synthases (OKS) may be active *in vivo* in a heterologous production host cell of a different genus, e.g. a plant type III OKS may be able to create non-reduced octaketides *in vivo* in a heterologous production host cell such as e.g. a recombinant *Aspergillus* production host cell.

As discussed above, the prior art describes numerous examples of *in vitro* tests on different OKS enzymes. However, no prior art describes that the OKS works *in vivo* in a host cell of a different genus.

Without being limited to theory, type III OKS may work in a heterologous host when the OKS is being protected *in vivo* against degradation (e.g. via formation of metabolons/complexes with other proteins). This degradation mechanism is not present *in vitro*.

The fact, that an OKS works in *Aspergillus nidulans* and *Nicotiana* plant cells demonstrates that the plant cells are capable of protecting a heterologous recombinant introduced OKS.

It is believed that this may be considered a general matter, i.e. that such protective systems will also be present in many other host cell types (such as e.g. yeast, tobacco plants etc).

In addition the octaketide product of the OKS is spontaneously degraded into SEK4 and/or SEK4B, which may inhibit the enzyme at the *in vivo* concentrations. When enzymes metabolizing the octaketide are present the PKS will not be inhibited by SEK4 and/or SEK4B and remain active.

In conclusion, a suitable method to produce an active OKS is to express it together with enzymes metabolizing the octaketide and/or forming complexes with the enzymes. These enzymes may be present natively in the cells expressing the introduced OKS or may be recombinantly introduced.

The fact as shown herein, that flavokermesic acid (FK) was present in the *Aspergillus nidulans* cells comprising recombinantly introduced Type III OKS from *Aloe arborescens* and/or *Hypericum perforatum* shows that *Aspergillus nidulans* cells comprise enzymes (e.g. aromatases/cyclase) metabolizing the produced non-reduced octaketide.

Without being limited to theory, it is believed that this may be considered a general matter, i.e. that such cyclases and/or aromatases will also be present in many other host cell types (such as e.g. yeast, tobacco plants etc).

DEFINITIONS

All definitions of herein relevant terms are in accordance of what would be understood by the skilled person in relation to the herein relevant technical context.

The term "aglycon" denotes non-carbohydrate part of the corresponding glycosylated form of the aglycon. When the sugar is glucose the aglycon may be termed aglucon. Further, when the sugar is glucose the term glucosylated may be used instead of glycosylated.

When the aglycon is glycosylated at a hydroxy group there is generally created a so-called *O*-glycosidic bond, i.e. a so-called *O*-Glycoside (or *O*-Glucoside if the sugar is glucose).

When the aglycon is glycosylated by a carbon-carbon linkage it may herein be termed a *C*-glycosidic bond, i.e. a so-called *C*-Glycoside (or *C*-Glucoside if the sugar is glucose).

The term "glycoside" denotes a compound, which on hydrolysis can give a sugar and a non-sugar (aglycon) residue, e. g. glucosides can give glucose, galactosides can give galactose.

The term "glycosyltransferase" denotes an enzyme capable of conjugating a nucleotide activated sugar to a compound (e.g. an aglycon compound). The sugar may e.g. be D and L isomers of galactose, glucosamine, *N*-acetylglusamine, xylose, glucuronic acid, rhamnose, arabinose, mannose or glucose. Alternatively the sugar may be a carbohydrate derivative

such as e.g. inositol, olivose, rhodinose and etc. available as nucleotide diphosphates. Further the sugar may for instance be a monosaccharide, a disaccharide or a trisaccharide. In the case of oligo- and polysaccharides the sugars are linked one by one by e.g. involving the use of one or several glycosyltransferases. If the sugar is glucose the glycosyltransferase may be termed a glucosyltransferase.

When the glycosyltransferase conjugates a nucleotide-activated sugar to a compound via a *C*-glycosidic bond it may herein be termed a *C*-glycosyltransferase.

When the glycosyltransferase conjugates a sugar to an aglycon via an *O*-glycosidic bond it may herein be termed an *O*-glycosyltransferase.

The term "hybridizes" in relation to a polynucleotide which hybridizes under at least medium stringency conditions with (i) nucleotides 1 to 1548 of SEQ ID NO:1 or (ii) a complementary strand of (i) relates to the nucleotide sequence hybridizes to a labeled nucleic acid probe corresponding to the nucleotide sequence shown in SEQ ID NO:1 or its complementary strand under medium to very high stringency conditions. Molecules to which the nucleic acid probe hybridizes under these conditions can be detected using e.g. X-ray film.

Herein relevant hybridization stringency conditions are stringency conditions that the skilled person normally would understand are relevant, i.e. for medium stringency conditions the conditions that skilled person would understand are medium stringency conditions. The skilled person knows herein relevant hybridization stringency conditions, see e.g. (J. Sambrook, E.F. Fritsch, and T. Maniatus, 1989, Molecular Cloning, A Laboratory Manual, 2d edition, Cold Spring Harbor, New York).

According to the art, for long probes of at least 100 nucleotides in length, very low to very high stringency conditions are defined as prehybridization and hybridization at 42°C in 5X SSPE, 0.3% SDS, 200 µg/ml sheared and denatured salmon sperm DNA, and either 25% formamide for very low and low stringencies, 35% formamide for medium and medium-high stringencies, or 50% formamide for high and very high stringencies, following standard Southern blotting procedures for 12 to 24 hours optimally.

For long probes of at least 100 nucleotides in length, the carrier material is finally washed three times each for 15 minutes using 2X SSC, 0.2% SDS preferably at least at 45°C (very low stringency), more preferably at least at 50°C (low stringency), more preferably at least at 55°C (medium stringency), more preferably at least at 60°C (medium-high stringency), even more preferably at least at 65°C (high stringency), and most preferably at least at 70°C (very high stringency).

The term "*in vitro*" (Latin: in glass) relates to studies that are conducted using components of an organism that have been isolated from their usual biological surroundings in order to permit a more detailed or more convenient analysis than can be done with whole organisms. Colloquially, these experiments are commonly called "test tube experiments". In

contrast, *in vivo* studies are those that are conducted with living organisms in their normal intact state.

The term "*in vivo*" (Latin for "within the living") relates to experimentation using a whole, living organism as opposed to a partial or dead organism, or an *in vitro* ("within the glass", e.g., in a test tube or petri dish) controlled environment.

The term "isolated polynucleotide" essentially relates herein to that the polynucleotide is isolated from its natural environment. Said in other words that the polynucleotide preparation is essentially free of other polynucleotide material with which it is natively associated. The polynucleotide sequence encoding the herein described isolated/cloned novel glycosyltransferase is shown in SEQ ID NO:1 and it was isolated from the insect *Dactylopius coccus*. Accordingly, as understood by the skilled person, the term "isolated polynucleotide" as used herein does not cover the polynucleotide of SEQ ID NO:1 as it is naturally present in the genome of *Dactylopius coccus*. The term "isolated polynucleotide" essentially relates to that the isolated polynucleotide is in a form suitable for use within genetically engineered protein production systems. Thus, an isolated polynucleotide contains at most 10%, preferably at most 8%, more preferably at most 6%, more preferably at most 5%, more preferably at most 4%, more preferably at most 3%, even more preferably at most 2%, most preferably at most 1%, and even most preferably at most 0.5% by weight of other polynucleotide material with which it is natively associated. The term "isolated polynucleotide" may herein alternatively be termed "cloned polynucleotide".

The term "isolated polypeptide" essentially relates herein to that the polypeptide is isolated from its natural environment. The herein described novel glycosyltransferase polypeptide as shown in SEQ ID NO: 2 herein was isolated from the insect *Dactylopius coccus*. Accordingly, as understood by the skilled person in the present context, the term "isolated polypeptide" as used herein does not cover the glycosyltransferase polypeptide of SEQ ID NO:2 as it is naturally present in the genome of *Dactylopius coccus*. The term "isolated polypeptide" denotes herein a polypeptide preparation which contains at most 10%, preferably at most 8%, more preferably at most 6%, more preferably at most 5%, more preferably at most 4%, at most 3%, even more preferably at most 2%, most preferably at most 1%, and even most preferably at most 0.5% by weight of other polypeptide material with which it is natively associated, as understood by the skilled person in the present context, the term "other polypeptide material with which it is natively associated" may in relation to the herein described novel glycosyltransferase polypeptide as shown in SEQ ID NO: 2 be seen in relation to other polypeptide material with which it is natively associated in *Dactylopius coccus*. In some case, it may be preferred that the "isolated polypeptide" refers to a polypeptide which is at least 20% pure, preferably at least 40% pure, more preferably at

least 60% pure, even more preferably at least 80% pure, most preferably at least 90% pure, and even most preferably at least 95% pure, as determined by SDS-PAGE.

The term "non-reduced octaketide" denotes a non-reduced octaketide, wherein the non-reduced type is characterized by the presence of the originally ketone groups in the ketides (-CH2-CO-), originating from the starter or extender units, either as ketones or in the form of double bonds in aromatic groups. In reduced polyketides a single or all ketones have been reduced to alcohol (-CH2-CHOH-) groups by e.g. a KR domain/enzyme, or further to an alkene group (-C=C-) by e.g. a DH domain/enzyme, or even further to an alkane group (-CH2-CH2-) by e.g. an ER domain/enzyme. Based on these chemical features of the formed products the involved PKSs are categorized as either being a non-reducing PKS or a reducing PKS.

The term "non-reducing PKS" or "non-reducing polyketide synthase" denotes a PKS which does not reduce the ketone groups in the formed polyketide chain. The lack of reductions can for instance be due to (I) a lack of the necessary keto-reductase (KR) active sites in the enzyme; and/or (II) lack of tailoring enzymes capable of catalyzing the keto-reduction reaction.

The term "nucleic acid construct" as used herein refers to a nucleic acid molecule, either single- or double-stranded, which is isolated from a naturally occurring gene or which is modified to contain segments of nucleic acids in a manner that would not otherwise exist in nature. The term nucleic acid construct is synonymous with the term "expression cassette" when the nucleic acid construct contains the control sequences required for expression of a coding sequence of the present invention. As known in the art control sequences include all components, which are necessary or advantageous for the expression of a polynucleotide encoding a polypeptide of the present invention. Each control sequence may be native or foreign to the nucleotide sequence encoding the polypeptide. Such control sequences include, but are not limited to, a leader, polyadenylation sequence, propeptide sequence, promoter, signal peptide sequence, and transcription terminator. At a minimum, the control sequences include a promoter, and transcriptional and translational stop signals. The control sequences may be provided with linkers for the purpose of introducing specific restriction sites facilitating ligation of the control sequences with the coding region of the nucleotide sequence encoding a polypeptide.

The term "octaketide" (greek for "eight") denotes a polyketide chain consisting of eight ketide units, meaning that the polyketide backbone consists of 16 carbon atoms. The term "ketide" refers to a –CH2-CO- unit or modification of this group.

The term "recombinant expression vector" relates to recombinant expression vectors comprising a polynucleotide of the present invention, a promoter, and transcriptional and translational stop signals. The various nucleic acids and control sequences described above may be joined together to produce a recombinant expression vector which may include one or more convenient restriction sites to allow for insertion or substitution of the nucleotide sequence encoding the polypeptide at such sites.

The term "recombinant host cell" should herein be understood according to the art. As known in the art, recombinant polynucleotide (e.g. DNA) molecules are polynucleotide (e.g. DNA) molecules formed by laboratory methods of genetic recombination (such as molecular cloning) to bring together genetic material from multiple sources, creating sequences that would not otherwise be found in biological organisms. As understood by the skilled person, a recombinant host cell comprises recombinant polynucleotide (e.g. DNA) molecules and a recombinant host cell will therefore not be understood as covering a natural wildtype cell as such, like e.g. a natural wildtype *Dactylopius coccus* cell.

The term "Sequence Identity" relates to the relatedness between two amino acid sequences or between two nucleotide sequences.

For purposes of the present invention, the degree of sequence identity between two amino acid sequences is determined using the Needleman-Wunsch algorithm (Needleman and Wunsch, 1970, *J. Mol. Biol.* 48: 443-453) as implemented in the Needle program of the EMBOSS package (EMBOSS: The European Molecular Biology Open Software Suite, Rice *et al.*, 2000, *Trends Genet.* 16: 276-277), preferably version 3.0.0 or later. The optional parameters used are gap open penalty of 10, gap extension penalty of 0.5, and the EBLOSUM62 (EMBOSS version of BLOSUM62) substitution matrix. The output of Needle labeled "longest identity" (obtained using the –nobrief option) is used as the percent identity and is calculated as follows:

(Identical Residues x 100)/(Length of Alignment – Total Number of Gaps in Alignment).

For purposes of the present invention, the degree of sequence identity between two nucleotide sequences is determined using the Needleman-Wunsch algorithm (Needleman and Wunsch, 1970, *supra*) as implemented in the Needle program of the EMBOSS package (EMBOSS: The European Molecular Biology Open Software Suite, Rice *et al.*, 2000, *supra*), preferably version 3.0.0 or later. The optional parameters used are gap open penalty of 10, gap extension penalty of 0.5, and the EDNAFULL (EMBOSS version of NCBI NUC4.4) substitution matrix. The output of Needle labeled "longest identity" (obtained using the –nobrief option) is used as the percent identity and is calculated as follows:

(Identical Deoxyribonucleotides x 100)/(Length of Alignment – Total Number of Gaps in Alignment).

As understood by the skilled person in the present context, for both "sequence identity between two nucleotide sequences" and "sequence identity between two amino acid

sequences", the term "Length of Alignment" should be understood as the actual length of alignment between the two sequences to be compared for sequence identity.
For instance, if a reference sequence is a specific SEQ ID of e.g. 100 amino acids and the other sequence is an identical sequence with 25 amino acids less at one end (i.e. the other sequence is of a length of 75 amino acids) then the "Length of Alignment" will be 75 amino acids and the percent identity will be 100%.
Another example is for instance, if a reference sequence is a specific SEQ ID of e.g. 100 amino acids and the other sequence is an identical sequence with 25 amino acids extra at one end (i.e. the other sequence is of a length of 125 amino acids) then will the "Length of Alignment" be 100 amino acids and the percent identity will be 100%.

The term "Type III polyketide synthase (PKS)" is, as discussed herein, a well-known term to the skilled person and the skilled person will know if a specific PKS of interest is a Type III PKS. As discussed in the review article of *Yu* et al (2012), Type III PKSs are self-contained enzymes that form homodimers. Their single active site in each monomer catalyzes the priming and extension to form polyketide products.

Embodiments of the present invention is described below, by way of examples only.

DRAWINGS

Figure 1: Schematic presentation of the herein relevant glycosyltransferase activity of the herein described isolated/cloned novel glycosyltransferase of SEQ ID NO:2 herein. As illustrated in the figure, it was found to be able to conjugate glucose to the aglycons flavokermesic acid (FK) and kermesic acid (KA).

Figure 2: Shows figure 5A of the *Tang*, Y. et al (2004) article. As can be seen in the figure and as further described in the article, the *Act* PKS (termed octaketide synthase (OKS)) creates a non-reduced octaketide and this octaketide is via the ZhuI aromatase/cyclase and ZhuJ cyclase converted into flavokermesic acid (FK) (called TMAC). The SEK4 and SEK4B compounds are also formed spontaneously (structures shown in Figure 2 may be termed shunt products).

Figure 3: Shows an example of a model for a suitable biosynthetic pathway that may lead to carminic acid.

Figure 4: Shows 9,10-anthraquinone skeleton; anthrone skeleton and FK antrone compound structure.

DETAILED DESCRIPTION OF THE INVENTION

A recombinant host cell – e.g. step (I)

Based on e.g. the sequence information disclosed herein, it is routine work for the skilled person to make a herein relevant recombinant host cell. As an example, based on the prior art the skilled person knows numerous different suitable recombinant host cells that for years have been used as recombinant host cells for e.g. expression of different polypeptides of interest.

The recombinant host cell in the method of the first aspect may be a growing recombinant host cell or e.g. in a so-called stationary phase.

Preferably, the recombinant host cell in the method of the first aspect is a growing recombinant host cell and step (I) and step (II) of the first aspect are:

(I): contacting *in vivo* in a growing recombinant host cell comprising a recombinantly introduced Type III polyketide synthase (PKS) gene encoding an octaketide synthase (OKS) and wherein the OKS is of a different genus than the host cell:

(i): suitable starter unit and extender units with the recombinantly introduced OKS capable of converting the starter and extender units into a non-reduced octaketide under suitable growth conditions wherein there *in vivo* is produced the non-reduced octaketide; and

(II): converting *in vivo* within the growing recombinant host cell the non-reduced octaketide of step (I) into a C_{14}-C_{34} aromatic aglycon compound of interest, wherein the aromatic aglycon compound of interest is not SEK4 and/or SEK4B.

The recombinant host cell may be any suitable cell such as any eukaryotic cell [e.g. mammalian cells (such as e.g. Chinese hamster ovary (CHO) cells) or a plant cell] or any prokaryotic cell.

It may be preferred that the recombinant host cell is a plant cell, preferably wherein the plant cell is a plant cell selected from the group consisting of: *Nicotiana sp.* (e.g. *Nicotiana benthamiana* cells); rhubarb, buckweed, *Hypericum* and *Aloe sp*.

Preferably, the plant cell is a *Nicotiana sp.*, more preferably the plant cell is *Nicotiana benthamiana*.

The recombinant host cell may be a cell selected from the group consisting of a filamentous fungal cell and a microorganism cell.

Filamentous fungi include all filamentous forms of the subdivision Eumycota and Oomycota (as defined by Hawksworth et al. , 1995, supra). The filamentous fungi are characterized by a vegetative mycelium composed of chitin, cellulose, glucan, chitosan, mannan, and other complex polysaccharides. Vegetative growth is by hyphal elongation and carbon catabolism is obligately aerobic. In contrast, vegetative growth by yeasts such as *Saccharomyces cerevisiae* is by budding of a unicellular thallus and carbon catabolism may be fermentative.

It may be preferred that the filamentous fungal cell is a cell of a species selected form the group consisting of, but not limited to, *Acremonium*, *Aspergillus*, *Fusarium*, *Humicola*, *Mucor*, *Myceliophthora*, *Neurospora*, *Penicillium*, *Thielavia*, *Tolypocladium*, and *Trichoderma* or a teleomorph or synonym thereof.

A preferred *Aspergillus* cell is *Aspergillus niger*, *Aspergillus oryzae or Aspergillus nidulans*. The most preferred *Aspergillus* cell is *Aspergillus nidulans*.

A preferred microorganism cell herein is a microorganism cell selected from the group consisting of a yeast cell and a prokaryotic cell.

A preferred yeast cell is a yeast cell of a phylum selected from the group consisting of Ascomycetes, Basidiomycetes and fungi imperfecti. Preferably the yeast cell is of the phylum Ascomycetes.

A preferred Ascomycetes yeast cell is of a genus selected from the group consisting of *Ashbya*, *Botryoascus*, *Debaryomyces*, *Hansenula*, *Kluveromyces*, *Lipomyces*, *Saccharomyces*, sush as *Saccharomyces cerevisiae*, *Pichia* and *Schizosaccharomyces*.

A preferred yeast cell is a yeast cell of a genus selected from the group consisting of *Saccharomyces,* such as *Saccharomyces cerevisiae* and *Pichia*.

A preferred prokaryotic cell is selected from the group consisting of: *Bacillus*, *Streptomyces*, *Corynebacterium*, *Pseudomonas*, lactic acid bacteria and an *E. coli* cell.

A preferred *Bacillus* cell is *B. subtilis*, *Bacillus amyloliquefaciens* or *Bacillus licheniformis*.
A preferred *Streptomyces* cell is *Streptomyces setonii* or *Streptomyces coelicolor*.
A preferred *Corynebacterium* cell is *Corynebacterium glutamicum*.
A preferred *Pseudomonas* cell is *Pseudomonas putida* or *Pseudomonas fluorescens*.

<u>Polyketide synthase (PKS) gene encoding an octaketide synthase (OKS) – e.g. step (I)</u>
Step (I) of the first aspect relates to "Type III polyketide synthase (PKS) gene encoding an octaketide synthase (OKS) and wherein the OKS is of a different genus than the host cell".

As discussed above and as known in the art - non-reduced octaketides may spontaneously be converted *in vivo* to the SEK4 and SEK4B compounds (structures shown in Figure 2 herein – may herein be termed shunt products).

As understood by the skilled person in the present context – identification of SEK4 and/or SEK4B *in vivo* within the recombinant host cell demonstrates that non-reduced octaketide is present *in vivo*.

Accordingly, the Type III octaketide synthase (OKS) of step (I) may be defined as an OKS, which *in vitro* is capable of producing the octaketide-derived shunt products SEK4 and/or SEK4B.

Assay for such an *in vitro* analysis is described e.g. in working Example 9 herein.

As discussed above, the skilled person knows if a specific PKS of interests is a Type III PKS.

As discussed above, in the prior art is known several different type III PKSs from plants, bacteria and/or fungi.

Accordingly, it may be preferred that the Type III polyketide synthase (PKS) gene of step (I) is a PKS gene from a plant, a bacterium or a fungi,

Preferably, the Type III polyketide synthase (PKS) gene of step (I) is a PKS gene from a plant.

If the Type III polyketide synthase (PKS) gene of step (I) is a PKS gene from a plant it is preferred that the plant is a plant selected from the group consisting of: *Aloe spp.* (preferably *Aloe arborescens*), *Hypericum spp.* (preferably *Hypericum perforatum*), rhubarb, buckweed and *Hawortia spp*.

Preferably the plant is a plant selected from the group consisting of: *Aloe spp*. and *Hypericum spp*.

More preferably, the plant is a plant selected from the group consisting of: *Aloe arborescens* and *Hypericum perforatum*.

If the Type III polyketide synthase (PKS) gene of step (I) is a PKS gene from a bacterium it is preferred that the bacterium is a bacterium selected from the group consisting of: *Streptomyces spp. (*preferably *Streptomyces coelicolor*).

If the Type III polyketide synthase (PKS) gene of step (I) is a PKS gene from a fungi it is preferred that the fungi is a fungi selected from the group consisting of: *Acremonium*,

Aspergillus, *Fusarium*, *Humicola*, *Mucor*, *Myceliophthora*, *Neurospora*, *Penicillium*, *Thielavia*, *Tolypocladium*, and *Trichoderma*.
A preferred *Aspergillus* is *Aspergillus niger*, *Aspergillus oryzae or Aspergillus nidulans*.
The most preferred *Aspergillus* is *Aspergillus nidulans*.

As known in the art, a Type I PKS and/or a Type II PKS may be mutated to remove relevant elements (e.g. active sites) to convert it into what the skilled person in the present context would functionally consider being a Type III PKS. A PKS, which by the skilled person is functionally considered being a Type III PKS is herein understood to be a Type III PKS.

As discussed above and in working examples herein, the present inventors recombinantly expressed plant (*Aloe arborescens* and/or *Hypericum perforatum*) type III polyketide synthases (PKS) octaketide synthases (OKS) in *Aspergillus nidulans* and identified that the OKS actually worked in the heterologous host, i.e. there were created non-reduced octaketides *in vivo* in the *Aspergillus nidulans* cells.
As discussed herein, flavokermesic acid (FK) was also identified in the *Aspergillus nidulans* cells.

Public available *Aloe arborescens* OKS sequences are herein shown with following SED ID numbers:
Herein termed *AaOKS:* SEQ ID NO:6 [Genbank ID number AY567707 (nucleotide seq.) and SEQ ID NO:7: [Genbank ID number AAT48709 (polypeptide seq)];
Herein termed *AaOKS2*(PKS4)*:* SEQ ID NO:8 [Genbank ID number FJ536166] (nucleotide seq.) and SEQ ID NO:9 [Genbank ID number ARC19997] (polypeptide seq);
Herein termed *AaOKS3*(PKS5)*:* SEQ ID NO:10 [Genbank ID number FJ536167] (nucleotide seq.) and SEQ ID NO:11 [Genbank ID number ARC19998] (polypeptide seq);
Herein termed *AaPKS2_A207G*: SEQ ID NO:12 (nucleotide seq.) and SEQ ID NO:13: (polypeptide seq).

Public available *Hypericum perforatum* OKS amino acid sequence is herein shown with following SED ID Number:
Herein termed *HpPKS2*: SEQ ID NO:14 [Genbank ID number HQ529467] (nucleotide seq.) and SEQ ID NO:15: [Genbank ID number AEE69029] (polypeptide seq);

The level of identities (%) between Type III PKSs at amino acid level:

	AaOKS	**AaPKS2**	**AaOKS3**	**HpPKS2**	**AaPKS2_A207G**
AaOKS	-	96.28%	89.88%	43.07%	91.32%
AaOKS2		-	90.12%	43.81%	91.07%
AaOKS3			-	45.07%	85.93%
HpPKS2				-	43.56%
AaPKS2_A207G					-

It may be preferred that the octaketide synthase (OKS) in step (I) of the first aspect is a OKS comprising an amino acid sequence which has at least 70% (preferably at least 80%, more preferably at least 90% and even more preferably at least 98%) identity with amino acids 1 to 403 of SEQ ID NO:7.

It may be preferred that the octaketide synthase (OKS) in step (I) of the first aspect is a OKS comprising an amino acid sequence which has at least 70% (preferably at least 80%, more preferably at least 90% and even more preferably at least 98%) identity with amino acids 1 to 403 of SEQ ID NO:9.

It may be preferred that the octaketide synthase (OKS) in step (I) of the first aspect is a OKS comprising an amino acid sequence which has at least 70% (preferably at least 80%, more preferably at least 90% and even more preferably at least 98%) identity with amino acids 1 to 405 of SEQ ID NO:11.

It may be preferred that the octaketide synthase (OKS) in step (I) of the first aspect is a OKS comprising an amino acid sequence which has at least 70% (preferably at least 80%, more preferably at least 90% and even more preferably at least 98%) identity with amino acids 1 to 403 of SEQ ID NO:13.

It may be preferred that the octaketide synthase (OKS) in step (I) of the first aspect is a OKS comprising an amino acid sequence which has at least 70% (preferably at least 80%, more preferably at least 90% and even more preferably at least 98%) identity with amino acids 1 to 393 of SEQ ID NO:15.

Step (I) – other matter

As discussed above, step (I)(i) of the first aspect reads: "suitable starter units and extender units with the recombinantly introduced OKS capable of converting the starter and extender units into a non-reduced octaketide under suitable conditions wherein there *in vivo* is produced the non-reduced octaketide".

As discussed above, suitable starter units and extender units are known in the art.

According to the art, suitable starter units may e.g. be acetyl-CoA or malonyl-CoA and suitable extender units may e.g. malonyl-CoA.

It is routine work for the skilled person to identify suitable growth conditions wherein there *in vivo* is produced the non-reduced octaketide as known to the skilled person, such suitable growth conditions will generally depend on the specific used recombinant host cell.

A recombinant host cell and OKS of different genus

As discussed above, step (I) of the first aspect reads: "contacting *in vivo* in a recombinant host cell comprising a recombinantly introduced Type III polyketide synthase (PKS) gene encoding an octaketide synthase (OKS) and wherein the OKS is of a different genus than the host cell"

It is herein most preferred to use a preferred recombinant host cell as discussed herein and a preferred OKS of a different genus as discussed herein.

For instance, a preferred embodiment herein relates to wherein the:
- recombinant host cell is a host cell selected from the group consisting of: *Aspergillus* (preferably *Aspergillus nidulans*) and *Nicotiana sp.* (preferably *Nicotiana benthamiana*); and
- the OKS of a different genus than the host cell is an OKS selected from the group consisting of: OKS from *Aloe spp.* (preferably *Aloe arborescens*) and *Hypericum spp.* (preferably *Hypericum perforatum*).

Preferably, the octaketide synthase (OKS) from *Aloe spp* is:
- an OKS comprising an amino acid sequence which has at least 70% (preferably at least 80%, more preferably at least 90% and even more preferably at least 98%) identity with amino acids 1 to 403 of SEQ ID NO:7; or
- an OKS comprising an amino acid sequence which has at least 70% (preferably at least 80%, more preferably at least 90% and even more preferably at least 98%) identity with amino acids 1 to 403 of SEQ ID NO:9; or
- an OKS comprising an amino acid sequence which has at least 70% (preferably at least 80%, more preferably at least 90% and even more preferably at least 98%) identity with amino acids 1 to 405 of SEQ ID NO:11; or
an OKS comprising an amino acid sequence which has at least 70% (preferably at least 80%, more preferably at least 90% and even more preferably at least 98%) identity with amino acids 1 to 403 of SEQ ID NO:13.

Preferably, the octaketide synthase (OKS) from *Hypericum spp.* is an OKS comprising an amino acid sequence which has at least 70% (preferably at least 80%, more preferably at least 90% and even more preferably at least 98%) identity with amino acids 1 to 393 of SEQ ID NO:15.

Aromatic aglycon compound of interest – step (II) of first aspect

As discussed above, step (II) of the first aspect reads: "converting *in vivo* within the recombinant host cell the non-reduced octaketide of step (I) into a C_{14}-C_{34} aromatic aglycon compound of interest, wherein the aromatic aglycon compound of interest is not SEK4 and/or SEK4B".

In a preferred embodiment, the aromatic aglycon compound of interest is a C_{14}-C_{18} aromatic aglycon compound of interest.

Preferably the aromatic aglycon compound of interest is at least one compound selected from the group consisting of: anthraquinone compound and anthrone compound.

Within the group above, it is preferred that the aromatic aglycon compound of interest of step (II) is an anthraquinone compound.

According to the art, anthraquinones (also known as anthraquinonoids) are a class of phenolic compounds based on the 9,10-anthraquinone skeleton (see Figure 4 herein).
For example and as understood by the skilled person, flavokermesic acid (FK) and kermesic acid (KA) are examples of anthraquinones.

Anthrones are a class of phenolic compounds based on the anthrone skeleton (see figure 4 herein).
In Figure 4 is shown FK antrone compound structure – i.e an example of a herein relevant anthrone compound.

Preferably, the aromatic aglycon compound of interest is a C_{16} aromatic aglycon compound of interest.
In a preferred embodiment, the C_{16} aromatic aglycon compound of interest is flavokermesic acid (FK) or kermesic acid (KA).

As discussed above, the conversion in step (II) of the non-reduced octaketide into a C_{14}-C_{34} (preferably C_{14}-C_{18}) aromatic aglycon compound of interest is generally done by involvement of suitable post-PKS enzymes (e.g. aromatases and/or cyclases) present *in vivo* in the host cell.

Accordingly, it may be preferred that the conversion *in vivo* in step (II) of the method of the first aspect of the non-reduced octaketide into a C_{14}-C_{34} aromatic aglycon compound of interest is done via involvement of at least one aromatase/cyclase.
Preferably, it is done via involvement of at least one *in trans* acting (independent of the PKS enzyme) aromatase/cyclase.

It may be preferred that step (II) of the method of the first aspect is:
(II): converting *in vivo* within the recombinant host cell the non-reduced octaketide of step (I) into an aromatic aglycon compound of interest, wherein the conversion *in vivo* of the non-reduced octaketide into an aromatic aglycon compound of interest is done via involvement of at least one *in trans* acting (independent of the PKS enzyme) aromatase/cyclase and wherein the aromatase/cyclase is from a fungi or a bacteria and heterologous to the recombinant host cell and from a different genus than the PKS.

SEQ ID NO: 17 herein is the public available amino sequence of *Streptomyces* ZhuI aromatase/cyclase (Genbank accession AAG30197) and SEQ ID NO: 19 herein is the public available amino sequence of *Streptomyces* ZhuJ aromatase/cyclase (Genbank accession AAG30196) (see e.g. herein discussed article of Tang, Y. et al (2004)).
As discussed below, these aromatases/cyclases were successfully used in working examples herein.

It may be preferred that the aromatase/cyclase is a aromatase/cyclase comprising an amino acid sequence which has at least 70% (preferably at least 80%, more preferably at least 90% and even more preferably at least 98%) identity with amino acids 1 to 169 of SEQ ID NO:17.

It may be preferred that the aromatase/cyclase is a aromatase/cyclase comprising an amino acid sequence which has at least 70% (preferably at least 80%, more preferably at least 90% and even more preferably at least 98%) identity with amino acids 1 to 256 of SEQ ID NO:19.

<u>Aromatic glycoside compound of interest – step (III) of first aspect</u>
As discussed above, optional step (III) of the first aspect reads: "the recombinant host cell also comprises a glycosyltransferase gene encoding a glycosyltransferase and the in step (II) produced aromatic aglycon compound of interest is *in vivo* glycosylated by the glycosyltransferase to produce a C_{14}-C_{34} aromatic glycoside compound of interest".

It may be preferred that this step (III) is fulfilled (i.e. not optional).

In a preferred embodiment, the aromatic glycoside compound of interest is a C_{14}-C_{18} aromatic glycoside compound of interest.

As understood by the skilled person in the present context – preferred aromatic aglycon compound of interest of step (II) may be converted into corresponding preferred aromatic glycoside compound of interest in step (III).

Preferably the aromatic glycoside compound of interest is at least one compound selected from the group consisting of: anthraquinone compound, anthrone compound.

Within the group above it is preferred that the aromatic glycoside compound of interest of step (III) is an anthraquinone compound.

For example and as understood by the skilled person - DcII and carminic acid (CA) are examples of anthraquinones.

Preferably, the aromatic glycoside compound of interest is a C_{16} aromatic glycoside compound of interest.
In a preferred embodiment, the C_{16} aromatic glycoside compound of interest is a flavokermesic acid glycoside or a kermesic acid glycoside.
In a preferred embodiment, the flavokermesic acid glycoside is DcII.
In a preferred embodiment, the kermesic acid glycoside is carminic acid (CA).

The glycosyltransferase (GT) of step (III) may e.g. be herein described GT from *Dactylopius coccus*.
Alternatively, it may e.g. be herein discussed glycosyltransferases from *Aloe* plants, *Haworthia* plants, *Sorghum* and/or rice plants.

Preferably, the glycosyltransferase is a glycosyltransferase polypeptide capable of:
(I): conjugating nucleotide activated glucose to flavokermesic acid (FK); and/or
(II): conjugating nucleotide activated glucose to kermesic acid (KA).

The polynucleotide sequence encoding herein described isolated/cloned novel *Dactylopius coccus costa* glycosyltransferase is shown in SEQ ID NO: 1 herein and the polypeptide amino acid sequence is shown in SEQ ID NO: 2 herein.
The herein relevant glycosyltransferase enzyme of SEQ ID NO: 2 may herein be termed "DcUGT2".

Accordingly, it may be preferred that the glycosyltransferase is a glycosyltransferase polypeptide capable of:

(I): conjugating nucleotide activated glucose to flavokermesic acid (FK); and/or

(II): conjugating nucleotide activated glucose to kermesic acid (KA);

and wherein the glycosyltransferase polypeptide is at least one polypeptide selected from the group consisting of:

(a) a polypeptide comprising an amino acid sequence which has at least 70% identity with amino acids 1 to 515 of SEQ ID NO:2;

(b) a polypeptide comprising an amino acid sequence which has at least 70% identity with amino acids 1 to 468 of SEQ ID NO:2;

(c) a polypeptide which is encoded by a polynucleotide that hybridizes under at least medium stringency conditions with (i) nucleotides 1 to 1548 of SEQ ID NO:1 or (ii) a complementary strand of (i); and

(d) a fragment of amino acids 1 to 515 of SEQ ID NO:2, which has the glycosyltransferase activity as specified in (I) and/or (II).

A preferred embodiment relates to wherein the glycosyltransferase polypeptide of is:

(a) a polypeptide comprising an amino acid sequence which has at least 80% identity with amino acids 1 to 515 of SEQ ID NO:2; more preferably

(a) a polypeptide comprising an amino acid sequence which has at least 90% identity with amino acids 1 to 515 of SEQ ID NO:2; even more preferably

(a) a polypeptide comprising an amino acid sequence which has at least 95% identity with amino acids 1 to 515 of SEQ ID NO:2; and most preferably

(a) a polypeptide comprising an amino acid sequence which has at least 98% identity with amino acids 1 to 515 of SEQ ID NO:2.

It may be preferred that the glycosyltransferase polypeptide of the first aspect is a polypeptide comprising an amino acid sequence with amino acids 1 to 515 of SEQ ID NO:2.

As discussed herein, the identified data/results of working Examples 4 show that herein relevant GT enzymes can be identified in e.g. *Sorghum* and rice plants.

The *Sorghum* polypeptide sequence (Genbank ID number: AAF17077.1) is shown as SEQ ID NO: 4 herein.

The rice polypeptide sequence (Genbank ID number: CAQ77160.1) is shown as SEQ ID NO: 5 herein.

It may be relevant that the glycosyltransferase is a glycosyltransferase comprising an amino acid sequence which has at least 70% (preferably at least 80%, more preferably at least 90% and even more preferably at least 98%) identity with amino acids 1 to 492 of SEQ ID NO:4.

It may be relevant that the glycosyltransferase is a glycosyltransferase comprising an amino acid sequence which has at least 70% (preferably at least 80%, more preferably at

least 90% and even more preferably at least 98%) identity with amino acids 1 to 471 of SEQ ID NO:5.

Alternatively, glycosyltransferases may also be a glycosyltransferase from a *Haworthia* plant.

Isolating the aromatic compound of interest – step (IV)

As discussed above, step (IV) of the first aspect reads: "isolating the aromatic aglycon compound of interest of step (II) and/or isolating the aromatic glycoside compound of interest of step (III) to get a composition, wherein the composition comprises less than 1% w/w dry matter of recombinant host cell material."

In the present context, the skilled person routinely know how to isolate a specific herein relevant aromatic compound of interest in relation to step (IV), i.e. it is not believed herein necessary to discuss this in great details.

In a preferred embodiment of the method as described herein is, wherein the isolated composition in step (IV) comprises aromatic glycoside compound of interest and wherein the aromatic glycoside compound of interest is flavokermesic acid glycoside (preferably DcII) and/or a kermesic acid glycoside (preferably carminic acid (CA)).

It may be preferred to have a relative pure composition, accordingly it may be preferred that composition comprising at least 10% w/w dry matter of the aromatic compound of interest and wherein the composition comprises less than 0.5% w/w dry matter of recombinant host cell material; or that composition comprising at least 50% w/w dry matter of the aromatic compound of interest and wherein the composition comprises less than 0.1% w/w dry matter of recombinant host cell material.

If there in step (IV) is obtained a composition comprising isolated aromatic aglycon compound of interest (e.g. flavokermesic acid (FK) or kermesic acid (KA)) the method of the first aspect may comprise an extra step of:

(IVa): glycosylating the aromatic aglycon compound of interest to produce an aromatic glycoside compound of interest.

This step may be done by chemical synthesis according to the art.

Alternatively, it may be done by use of a glycosyltransferase capable of glycosylating the aglycon under suitable conditions wherein there is produced the aglycon glycoside.

An embodiment of the invention relates to wherein the aromatic aglycon compound of interest of the isolated composition of step (IV) is flavokermesic acid (FK) and/or kermesic acid (KA) and the method of the first aspect comprises an extra step of:

(IVa) contacting:

(a1): flavokermesic acid (FK) with a glycosyltransferase capable of glycosylating the flavokermesic acid under suitable conditions wherein there is produced the flavokermesic acid glycoside; and/or

(a2): kermesic acid (KA) with a glycosyltransferase capable of glycosylating the kermesic acid under suitable conditions wherein there is produced the kermesic acid glycoside.

In a preferred embodiment, the flavokermesic acid glycoside is DcII.

In a preferred embodiment, the kermesic acid glycoside is carminic acid (CA).

As discussed herein, the glycosyltransferase (GT) of this step may e.g. be the herein described GT from *Dactylopius coccus*.

Alternatively, it may e.g. be the herein discussed glycosyltransferases from *Aloe* plants, *Haworthia* plants, *Sorghum* and/or rice plants.

Aspects and preferred embodiments – in so-called claim format/language

Herein described aspect(s) and preferred embodiments thereof may be presented in so-called claim format/language. This is done below for some of the herein described aspect(s) and preferred embodiments thereof.

1. A method for producing an octaketide derived aromatic compound of interest, wherein the method comprises the following steps:

(I): contacting *in vivo* in a recombinant host cell comprising a recombinantly introduced Type III polyketide synthase (PKS) gene encoding an octaketide synthase (OKS) and wherein the OKS is of a different genus than the host cell:

(i): suitable starter unit and extender units with the recombinantly introduced OKS capable of converting the starter and extender units into a non-reduced octaketide under suitable conditions wherein there *in vivo* is produced the non-reduced octaketide; and

(II): converting *in vivo* within the recombinant host cell the non-reduced octaketide of step (I) into a C_{14}-C_{34} aromatic aglycon compound of interest, wherein the aromatic aglycon compound of interest is not SEK4 and/or SEK4B; and

(III): optionally, the recombinant host cell also comprises a glycosyltransferase gene encoding a glycosyltransferase and the in step (II) produced aromatic aglycon compound of interest is *in vivo* glycosylated by the glycosyltransferase to produce a C_{14}-C_{34} aromatic glycoside compound of interest; and

(IV): isolating the aromatic aglycon compound of interest of step (II) and/or isolating the aromatic glycoside compound of interest of step (III) to get a composition, wherein the composition comprises less than 1% w/w dry matter of recombinant host cell material.

2. The method according to claim 1, wherein the recombinant host cell is a plant cell, a filamentous fungal cell, a yeast cell or a prokaryotic cell.

3. The method according to claim 2, wherein the recombinant host cell is a plant cell.

4. The method according to claim 3, wherein the plant cell is a plant cell selected from the group consisting of: *Nicotiana sp.* (e.g. *Nicotiana benthamiana* cells); rhubarb, buckweed, *Hypericum* and *Aloe sp*.

5. The method according to claim 4, wherein the plant cell is *Nicotiana sp.*

6. The method according to claim 5, wherein the plant cell is *Nicotiana benthamiana*.

7. The method according to claim 2, wherein the recombinant host cell is a filamentous fungal cell.

8. The method according to claim 7, wherein the filamentous fungal cell is *Acremonium*, *Aspergillus*, *Fusarium*, *Humicola*, *Mucor*, *Myceliophthora*, *Neurospora*, *Penicillium*, *Thielavia*, *Tolypocladium*, or *Trichoderma*.

9. The method according to claim 8, wherein the filamentous fungal cell is an *Aspergillus* cell.

10. The method according to claim 9, wherein the *Aspergillus* cell is *Aspergillus niger*, *Aspergillus oryzae or Aspergillus nidulans*.

11. The method according to claim 10, wherein the *Aspergillus* cell is *Aspergillus nidulans*.

12. The method according to any of the preceding claims, wherein the Type III polyketide synthase (PKS) gene of step (I) is a PKS gene from a plant, a bacterium or a fungi.

13. The method of claim 12, wherein the Type III polyketide synthase (PKS) gene of step (I) is a PKS gene from a plant.

14. The method of claim 13, wherein the plant is a plant selected from the group consisting of: *Aloe spp.* (preferably *Aloe arborescens*), *Hypericum spp.* (preferably *Hypericum perforatum*), rhubarb, buckweed and *Hawortia spp*.

15. The method of claim 14, wherein the plant is a plant selected from the group consisting of: *Aloe spp.* and *Hypericum spp.*

16. The method of claim 15, wherein the plant is a plant selected from the group consisting of: *Aloe arborescens* and *Hypericum perforatum.*

17. The method of claim 13, wherein the octaketide synthase (OKS) in step (I) of claim 1 is a OKS comprising an amino acid sequence which has at least 70% (preferably at least 80%, more preferably at least 90% and even more preferably at least 98%) identity with amino acids 1 to 403 of SEQ ID NO:7.

18. The method of claim 13, wherein the octaketide synthase (OKS) in step (I) of claim 1 is a OKS comprising an amino acid sequence which has at least 70% (preferably at least 80%, more preferably at least 90% and even more preferably at least 98%) identity with amino acids 1 to 403 of SEQ ID NO:9.

19. The method of claim 13, wherein the octaketide synthase (OKS) in step (I) of claim 1 is a OKS comprising an amino acid sequence which has at least 70% (preferably at least 80%, more preferably at least 90% and even more preferably at least 98%) identity with amino acids 1 to 405 of SEQ ID NO:11.

20. The method of claim 13, wherein the octaketide synthase (OKS) in step (I) of claim 1 is a OKS comprising an amino acid sequence which has at least 70% (preferably at least 80%, more preferably at least 90% and even more preferably at least 98%) identity with amino acids 1 to 403 of SEQ ID NO:13.

21. The method of claim 13, wherein the octaketide synthase (OKS) in step (I) of claim 1 is a OKS comprising an amino acid sequence which has at least 70% (preferably at least 80%, more preferably at least 90% and even more preferably at least 98%) identity with amino acids 1 to 393 of SEQ ID NO:15.

22. The method according to any of the preceding claims, wherein the suitable starter unit is acetyl-CoA and/or malonyl-CoA.

23. The method according to any of the preceding claims, wherein the suitable extender units is malonyl-CoA.

24. The method according to any of the preceding claims, wherein the:
- recombinant host cell is a host cell selected from the group consisting of: *Aspergillus* (preferably *Aspergillus nidulans*) and *Nicotiana sp.* (preferably *Nicotiana benthamiana*); and
- the OKS of a different genus than the host cell is an OKS selected from the group consisting of: OKS from *Aloe spp.* (preferably *Aloe arborescens*) and *Hypericum spp.* (preferably *Hypericum perforatum*).

25. The method according to any of the preceding claims, wherein the aromatic aglycon compound of interest is a C_{14}-C_{18} aromatic aglycon compound of interest.

26. The method according to claim 25, wherein the aromatic aglycon compound of interest is at least one compound selected from the group consisting of: anthraquinone compound and anthrone compound.

27. The method of claim 26, wherein the aromatic aglycon compound of interest is an anthraquinone and the anthraquinone is flavokermesic acid (FK) or kermesic acid (KA).

28. The method according to claim 25, wherein the aromatic aglycon compound of interest is a C_{16} aromatic aglycon compound of interest.

29. The method according to claim 28, wherein the C_{16} aromatic aglycon compound of interest is flavokermesic acid (FK) or kermesic acid (KA).

30. The method according to any of the preceding claims, wherein the conversion *in vivo* in step (II) of claim 1 of the non-reduced octaketide into a C_{14}-C_{34} aromatic aglycon compound of interest is done via involvement of at least one aromatase/cyclase.

31. The method according to claim 30, wherein the conversion *in vivo* in step (II) of claim 1 of the non-reduced octaketide into a C_{14}-C_{34} aromatic aglycon compound of interest is done via involvement of at least one *in trans* acting (independent of the PKS enzyme) aromatase/cyclase.

32. The method of claim 31,
wherein the aromatase/cyclase is a aromatase/cyclase comprising an amino acid sequence which has at least 70% (preferably at least 80%, more preferably at least 90% and even more preferably at least 98%) identity with amino acids 1 to 169 of SEQ ID NO:17 (*Streptomyces* ZhuI); and/or
wherein the aromatase/cyclase is a aromatase/cyclase comprising an amino acid sequence which has at least 70% (preferably at least 80%, more preferably at least 90% and even

more preferably at least 98%) identity with amino acids 1 to 256 of SEQ ID NO:19 (*Streptomyces* ZhuJ).

33. The method according to any of the preceding claims, wherein step (III) is fulfilled (i.e. not optional) – i.e. wherein step (III) is: the recombinant host cell also comprises a glycosyltransferase gene encoding a glycosyltransferase and the in step (II) produced aromatic aglycon compound of interest is *in vivo* glycosylated by the glycosyltransferase to produce a C_{14}-C_{34} aromatic glycoside compound of interest.

34. The method of claim 33, wherein the aromatic glycoside compound of interest is a C_{14}-C_{18} aromatic glycoside compound of interest.

35. The method of claim 34, wherein the aromatic glycoside compound of interest is a C_{16} aromatic glycoside compound of interest.

36. The method of claim 35, wherein the C_{16} aromatic glycoside compound of interest is a flavokermesic acid glycoside or a kermesic acid glycoside.

37. The method of claim 36, wherein the C_{16} aromatic glycoside compound of interest is a flavokermesic acid glycoside and the flavokermesic acid glycoside is DcII.

38. The method of claim 36, wherein the C_{16} aromatic glycoside compound of interest is a kermesic acid glycoside and the kermesic acid glycoside is carminic acid (CA).

39. The method according to any of the claims 33 to 38, wherein the glycosyltransferase (GT) of step (III) is a GT from *Dactylopius coccus*, a GT from *Aloe* plants, a GT from *Haworthia* plants, a GT from Sorghum or a GT from rice plant.

40. The method of claim 36, wherein the glycosyltransferase (GT) is a glycosyltransferase polypeptide capable of:

(I): conjugating nucleotide activated glucose to flavokermesic acid (FK); and/or

(II): conjugating nucleotide activated glucose to kermesic acid (KA).

41. The method according to any of the claims 33 to 40, wherein the glycosyltransferase is a glycosyltransferase polypeptide capable of:

(I): conjugating nucleotide activated glucose to flavokermesic acid (FK); and/or

(II): conjugating nucleotide activated glucose to kermesic acid (KA);

and wherein the glycosyltransferase polypeptide is at least one polypeptide selected from the group consisting of:

(a) a polypeptide comprising an amino acid sequence which has at least 70% identity with amino acids 1 to 515 of SEQ ID NO:2;

(b) a polypeptide comprising an amino acid sequence which has at least 70% identity with amino acids 1 to 468 of SEQ ID NO:2;

(c) a polypeptide which is encoded by a polynucleotide that hybridizes under at least medium stringency conditions with (i) nucleotides 1 to 1548 of SEQ ID NO:1 or (ii) a complementary strand of (i); and

(d) a fragment of amino acids 1 to 515 of SEQ ID NO:2, which has the glycosyltransferase activity as specified in (I) and/or (II).

42. The method according to any of the claims 33 to 40, wherein the glycosyltransferase is a glycosyltransferase comprising an amino acid sequence which has at least 70% (preferably at least 80%, more preferably at least 90% and even more preferably at least 98%) identity with amino acids 1 to 492 of SEQ ID NO:4.

43. The method according to any of the claims 33 to 40, wherein the glycosyltransferase is a glycosyltransferase comprising an amino acid sequence which has at least 70% (preferably at least 80%, more preferably at least 90% and even more preferably at least 98%) identity with amino acids 1 to 471 of SEQ ID NO:5.

44. The method according to any of the claims 33 to 43, wherein the isolated composition in step (IV) comprises aromatic glycoside compound of interest and wherein the aromatic glycoside compound of interest is flavokermesic acid glycoside (preferably DcII) and/or a kermesic acid glycoside (preferably carminic acid (CA)).

45. The method according to claim 44, wherein the aromatic glycoside compound of interest is flavokermesic acid glycoside and the flavokermesic acid glycoside is DcII.

46. The method according to claim 44, wherein the aromatic glycoside compound of interest is kermesic acid glycoside and the kermesic acid glycoside is carminic acid (CA).

47. The method according to any of the preceding claims, wherein the isolated composition in step (IV) of claim 1 is comprising at least 10% w/w dry matter of the aromatic compound of interest and wherein the composition comprises less than 0.5% w/w dry matter of recombinant host cell material.

48. The method according to any of the preceding claims, wherein there in step (IV) is obtained a composition comprising isolated aromatic aglycon compound of interest (e.g. flavokermesic acid (FK) or kermesic acid (KA)) and the method then comprise an extra step of:

(IVa): glycosylating the aromatic aglycon compound of interest to produce an aromatic glycoside compound of interest.

49. The method of claim 48, wherein the glycosylating the aromatic aglycon compound of interest is done by use of a glycosyltransferase capable of glycosylating the aglycon under suitable conditions wherein there is produced the aglycon glycoside.

50. The method of claim 49, wherein the aromatic aglycon compound of interest of the isolated composition of step (IV) is flavokermesic acid (FK) and/or kermesic acid (KA) and the method of the first aspect comprises an extra step of:

(IVa): contacting:

(a1): flavokermesic acid (FK) with a glycosyltransferase capable of glycosylating the flavokermesic acid under suitable conditions wherein there is produced the flavokermesic acid glycoside; and/or

(a2): kermesic acid (KA) with a glycosyltransferase capable of glycosylating the kermesic acid under suitable conditions wherein there is produced the kermesic acid glycoside.

51. The method of any of claims 48 to 50, wherein the flavokermesic acid glycoside is DcII.

52. The method of any of claims 48 to 50, wherein the kermesic acid glycoside is carminic acid (CA).

53. The method of any of the preceding claims, wherein the recombinant host cell in claim 1 is a growing recombinant host cell and step (I) and step (II) of claims 1 are:

(I): contacting *in vivo* in a growing recombinant host cell comprising a recombinantly introduced Type III polyketide synthase (PKS) gene encoding an octaketide synthase (OKS) and wherein the OKS is of a different genus than the host cell:

(i): suitable starter unit and extender units with the recombinantly introduced OKS capable of converting the starter and extender units into a non-reduced octaketide under suitable growth conditions wherein there *in vivo* is produced the non-reduced octaketide; and

(II): converting *in vivo* within the growing recombinant host cell the non-reduced octaketide of step (I) into a C_{14}-C_{34} aromatic aglycon compound of interest, wherein the aromatic aglycon compound of interest is not SEK4 and/or SEK4B.

EXAMPLES

Relevant sequence information can be found in herein provided Sequence Listing document

Example 1 – *Cloning of D. coccus GT and test of its FK and KA activity*

Materials and methods

Purification of DNA and mRNA

Fresh frozen *Dactylopius coccus (*were obtained from Lanzarote). Fresh frozen *Porphyrophora polonica* were obtained from Poland. The frozen insects were ground into powder under liquid nitrogen and DNA/RNA was purified: DNA was purified using DNeasy Blood & Tissue kit (Qiagen), according to the protocol of the supplier. RNA was purified using RNeasy mini kit (Qiagen) according to the protocol of the supplier.

Eucaryote mRNA was made into cDNA using RT² Easy First Strand Kit (Qiagen) according to the protocol of the supplier using poly-dT priming of the revers transcriptase reaction.

Sequencing of DNA and RNA:

DNA and cDNA were sent for sequencing at BGI (Shenzen, China) for sequencing using 100 bp paired-end Illumina technology according to the protocol of Illumina at a coverage of approximately 60-100X and the output in fastq data format.

Analysis of DNA and RNA/cDNA sequences:

The obtained fastq-sequences of DNA and RNA/cDNA were imported into Genomic Workbench version 5.4 (CLC-bio, Århus, Denmark) and assembled using the *de novo* assembling algorithm into contigs. Output files were exported as FASTA format.

RNA/cDNA FASTA files were then imported into IOGMA v. 10 (Genostar, Grenoble, France) and putitative genes were identified using the "hidden Markov-Matrix-based prokaryote gene-finder.

Putative genes were annotated using BLAST (basic local alignment sequence tool) against genbank (NCBI) using as well the nucleotide sequence as the translated protein sequence. The putative genes were also annotated by similarity comparison to PFAM databases of protein families.

Preparation of protein fractions from D. coccus

Three grams of fresh *D. coccus insects* were homogenized in 120 mL of isolation buffer [350 mM sucrose, 20 mM Tricine (pH 7.9), 10 mM NaCl, 5 mM DTT, 1 mM PMSF) containing 0.3 g polyvinylpolypyrrolidone. The homogenate was filtered through a nylon cloth (22 µm mesh) and centrifuged for (10 min, 10,000 x g at 4 °C). The supernatant was centrifuged (1h, 105,000xg, at 4 °C), yielding a soluble and a membrane bound protein fraction. The soluble

protein fraction was concentrated to 1 mL and buffer-exchanged with 20 mM Tricine (pH 7.9), 5 mM DTT by using Amicon ultrafugation-3K devices (Millipore). The membrane bound protein pellet was washed 3 times by resuspending the pellet in 60 mL of 20 mM Tricine (pH 7.9), 5 mM DTT using a marten paintbrush followed by re-centrifugation. The membrane bound protein pellet was finally resuspended in 1 mL 20 mM Tricine (pH 7.9), 5 mM DTT. The soluble protein fraction and the membrane bound protein fraction were analyzed for glycosylation activity.

Purification of a flavokermesic acid/kermesic acid-specific GT activity from *D. coccus* membrane proteins

A membrane bound protein fraction isolated from 3 g fresh *D. coccus insects* was solubilized by adding 1 % (v/v) Triton x-100 (reduced form) and gently stirring for 1.5 h in the cold. The Triton x-100 treated solution was centrifuged (1h, 105,000xg, at 4 °C) and the supernatant was isolated and applied to a column packed with 2 mL Q-sepharose Fast flow (Pharmacia). The column was washed in 4 mL of buffer A [20 mM Tricine (pH 7.9), 0.1 % (v/v) Triton x-100 (reduced form), 50 mM NaCl] and proteins were eluted with 20 mM Tricine (pH 7.9), 0.1 % (v/v) Triton x-100 (reduced form) using a discontinuous NaCl gradient from 100 mM-500 mM,(with 50 mM increments). 0.5-ml-fractions were collected, desalted, analyzed by SDS-PAGE and monitored for glucosylation activity using the described radiolabeled glucosylation enzyme assay. A fraction containing enriched flavokermesic acid/kermesic acid-specific GT activity was subjected to peptide mass fingerprinting analysis.

Enzyme assays and glucoside product detection

Glucosylation of flavokermesic acid and kermesic acid was performed in assay mixtures of 60 µL containing 20 mM Tricine(pH 7.9), 3.3 µm UDP[14C]glucose and 20 uL protein extract (membrane bound or soluble protein). Reactions were incubated for 0.5 h at 30°C and terminated by adding 180 µL of methanol. Samples were centrifuged at 16,000xg for 5 min at 4°C and supernatant was spotted on TLC plates (silica gel 60 F254 plates; Merck). Assay products were resolved in dichloromethane:methanol:formic acid (7:2:2, by volume). Radiolabeled products were visualized using a STORM 840 PhosphorImager (Molecular Dynamics, http://www.moleculardynamics.com).

Expression of codon optimized DcUGT2, DcUGT4 and DcUGT5 in S. cerevisiae

A synthetic codon optimized version of DcUGT2 and two other putative GT sequences from the *Dactylopius coccus* transcriptome termed DcUGT4 and DcUGT5 for yeast expression was purchased from GenScript with flanking gateway recombination attL sites. The synthetic fragments were used as PCR templates with specific primers to generate the corresponding C-terminal StrepII-tagged versions. The six gene constructs (tagged and non-tagged fragments) were cloned into the gateway destination plasmid pYES-DEST52 (Invitrogen)

using LR clonaseII enzyme mix. The six pYES-DEST52 plasmid constructs were transformed separately into the Invsc1 yeast strain (Invitrogen) and positive transformants were verified by PCR. Heterologous protein production was performed according to the instructions of the pYES-DEST52 gateway vector (Invitrogen). Production of heterologous StrepII-tagged protein was verified by western blotting using anti-Strep antibody. A membrane bound protein fraction was prepared from verified yeast transformants as described in (D. Pompon, B. Louerat, A. Bronine, P. Urban, Yeast expression of animal and plant P450s in optimized redox environments, Methods Enzymol. 272 (1996) 51–64) and screened for glucosylation activity towards flavokermesic acid/kermesic acid. The yeast optimized sequence is shown in SEQ ID NO: 3 herein.

LC-MS-MS

LC/MS was performed on an Agilent Q-TOF with the following HPLC system:

Column Kinetix 2.6µ XB-C18 100A (100 x 4.60 mm, Phenomenex). Solvent A (900 ml deionized water, 100 ml methanol and 50 ml formic acid). Solvent B (700 ml methanol, 300 ml deionized water and 50 ml formic acid).

Flow 0.8 ml/min. 35°C.

Gradient elution. 0-1 min 100 % A. Linear gradient to 83 % A 3 min. linear gradient to 63 % A 6 min, linear gradient to 45 % A 9 min, linear gradient to 27 % A 12 min, linear gradient to 10 % A 15 min, linear gradient to 3 % A 17 min, linear gradient to 2 % A 19 min, linear gradient to 0 % A 20 min, 0 %A 22 min, linear gradient to 100 % A 25 min.

Retention times were 7.6 min for carminic acid, 7.8 min for DC II, 13.7 min for flavokermesic acid and 13.9 min for kermesic acid.

Results:

The ability to glycosylate flavokermesic acid/kermesic acid using C14-UDP-glucose as a substrate was detected in homogenized *D. coccus* insects. The activity was shown to be membrane bound and the activity was purified and the purified proteins were submitted to proteomics analysis. It was shown that the enzymatic activity was to come from a polypeptide with a sequence corresponding to our candidate gene DcUGT2.

As discussed above, the herein relevant glycosyltransferase enzyme of SEQ ID NO: 2 may herein be termed "DcUGT2".

The amino acid sequence of DcUGT2 shows less than 45% homology to any known glycosyl transferase.

Knowing that cloning the wildtype sequence into yeast had given no relevant enzyme activity, we redesigned the nucleotide sequence of DcUGT2 to a sequence coding for the same polypeptide but using nucleotide codons optimized for *S. cerevisiae,* a process called codon optimization (the *S. cerevisiae* optimized sequence is shown as SEQ ID No. 3 herein). Subsequently, the codon optimized sequence of DcUGT2 was cloned and expressed in

yeast. The heterologous yeast strain contains a membrane bound enzyme activity capable of glucosylating kermesic acid and flavokermesic acid.

After obtaining peptide mass fingerprinting data from a *Dactylopius coccus* protein fraction enriched with GT activity towards flavokermesic acid/kermesic acid, we matched the peptide masses to the transcriptomic dataset and identified three putative UGTs (DcUGT2, DcUGT4 and DcUGT5).

Heterologous expression of the three candidates in yeast revealed that only one of these UGTs, namely DcUGT2 was responsible for the observed glucosylation activity towards flavokermesic acid/kermesic acid in the *D. coccus* protein fraction.

A viscozyme treatment of the generated C-14 radiolabelled glucoside, showed that it was resistant towards hydrolysis, further suggesting that the DcUGT2 is a C-GT, responsible for producing DCII and carminic acid.

A LC-MS-MS showed formation of products with the same retention time, spectrum, molecular mass and molecular degradation pattern as DcII and carminic acid respectively.

Conclusion

The result of this Example 1 demonstrated that it was not an easy task to isolate/clone the herein relevant glycosyltransferase enzyme of SEQ ID NO: 2, which may herein be termed "DcUGT2" or simply "DcUGT".

For instance, the identified gene sequences of the genome and transcriptome of *D. coccus* insects were analyzed for similarity to herein relevant public known C-glycosyltransferase sequences and the result was negative in the sense that none of the identified gene sequences of the genome/transcriptome showed herein significant similarity to publicly known herein relevant C-glycosyltransferase sequences.

However, even though the bioinformatic sequence similarity analysis could be said to indicate that the genome of *Dactylopius coccus* would not comprise a gene encoding a herein relevant glycosyltransferase – the present inventors continued to investigate the matter and the present inventors identified a *Dactylopius coccus* extract (including extracts of the endosymbionts present in *D. coccus*) with herein relevant GT activity and by a combination of herein relevant purification and testing steps the inventors were finally able to get a relatively pure fraction/composition wherefrom it was possible to obtain several partial amino acid sequences of possible GT enzyme candidates.

The present inventors tested the activity of the herein described isolated/cloned novel glycosyltransferase of SEQ ID NO: 2 (DcUGT2) and found that it was able to conjugate glucose to the aglycons flavokermesic acid (FK) and kermesic acid (KA) – see Figure 1 herein.

Example 2 *Testing KA GT activity* of prior art known UrdGT2

As discussed above, the UrdGT2 is described in the article Baig et al (Angew Chem Int Ed Engl. 2006 Nov 27;45(46):7842-6).

As discussed above, this article describes that UrdGT2 is capable of glycosylating different aglycon molecules that may be considered structurally similar to the herein relevant Kermesic acid (KA) and Flavokermesic acid (FK) aglycons.

A codon optimized synthetic version of UrdGT2 for *E. coli* expression was cloned and recombinantly expressed in *E. coli*. A crude soluble protein extract containing the the recombinant UrdGT2 was obtained, i.e. an extract comprising the UrdGT2

The UrdGT2 GT activity was analyzed *in vitro* using either UDP-glucose or TDP-glucose as a sugar donor and FA/KA as aglycone substrates. No activity was detected towards these aglycons, i.e. no herein relevant GT activity was identified in relation to these aglycons.

However, it was confirmed that the recombinant UrdGT2 was active, as demonstrated by the *in vitro* formation of a C14-radiolabelled glucoside derived from the glucosylation of an unidentified compound in the crude *E. coli* extract.

Example 3 *GT activity in Aloe plant and Haworthia* plant

Isolation and test of GT activity from *Aloe*

1) The plant was washed from soil particles and separated into : A) Root, B) Green leaf tissue and C) the gel material from the leaf
2) 5 g of tissue was frozen immediately in liquid nitrogen and ground in a cold mortar with a pestle to a fine powder.
3) 20 mL of cold extraction buffer [20 mM Tricine-HCl, 10 mM NaCl, 5 mM DTT, 1 mM PMSF, pH 7.9] containing a Complete protease inhibitor without EDTA (Roche), 0.1 % (w/v) proteamine sulfate and 0.5 g of PVPP were added to the powder and vortexed.
4) The homogenate was gently stirred at 4 °C for 10 min and then centrifuged at 12,000 xg at 4 °C for 5 min.
5) Supernatant was isolated and 1 mL of 2% (w/v) proteamine sulfate in 20 mM Tricine-HCl, pH 7.9 was added dropwise over 2 min at 4°C under constant stirring.
6) The supernatant was filtered through 2 pieces of nylon mesh. The filtered supernatant was then centrifuged at at 12,000 xg at 4°C for 5 min.
7) The supernatant was isolated and ultracentrifuged at 110,000 xg at 4 °C for 1 h.

8) The soluble protein fraction (supernatant) was isolated and buffer-exchanged 5 times with 20 mM Tricine-HCl, pH 7.9 containing 5 mM DTT using a Amicon Ultra centrifugal filter device-3K (Millipore)
9) 20 µL soluble protein extract was incubated in a total reaction volume of 60 µL containing UDP-glucose (1.25 mM final conc.) and either FK (50 µM final conc.) , KA (50 µM final conc) or MeO-FK/EtO-FK (50 µM/50 µM final conc) for 2 h at 30 °C, shaking at 650 rpm.
10) Enzyme reactions were terminated with 180 µL cold methanol and filtered through a 0.45 micron filter and subjected to HPLC-MS analysis.

m/z [M-H]⁻ values --- Aloe Soluble protein	475 m/z $[M-H]^-$ FK-monoglc	491 m/z $[M-H]^-$ KA-monoglc	489 m/z $[M-H]^-$ MeOFK-monoglc	503 m/z $[M-H]^-$ EtOFK-monoglc
Leaf	3.73	3.71	5.81	6.63
Gel				
Root		3.71		

Table 1. Glucosides formed in *in vitro* glucosylation assays using enzyme extracts from *Aloe*.

Crude soluble enzyme extracts of three *Aloe* tissues, green leaf material (Leaf), gel material from the leaf (Gel) and Root were tested for glucosylation activity towards flavokermesic acid (FK), kermesic acid (KA), methyl ester of flavokermesic acid (MeOFK) and ethyl ester of flavokermesic acid (EtOFK). Numbers correspond to retention times (min) after HPLC-MS separation of the novel glucosides formed *in vitro* (Table 1).

The m/z values 475 and 491 are the same m/z values as are obtained for DcII and CA, respectively, solubilized in similar solutions. Both m/z values are 162 (m/z value of glucose in a glucoside) higher than the m/z values of the FK and KA indicating that the glucose moiety from UDP-glucose in the reaction buffer has been transferred to the aglycone by a GT in the extract. The m/z [M-H] values 489 and 503 are also 162 higher than the m/z values obtained with MeOFK and EtOFK, respectively, indicating that a glucose unit has been added to both MeOFK and EtOFK by a GT present in the extract.

Isolation and test of GT activity from *Haworthia limifolia*

The procedure was as described for *Aloe* but plant tissue analyzed were following: A) Green leaf tissue, B) Gel material from the leaf, C) Base tissue (pink part between root and stem) and D) Root tissue.

Crude soluble enzyme extracts of four *Haworthia limifolia* tissues, green leaf material (Leaf), gel material from the leaf (Gel), pink tissue between root and stem (Base) and Root were tested for glucosylation activity towards flavokermesic acid (FK), kermesic acid (KA), methyl ester of flavokermesic acid (MeOFK) and ethyl ester of flavokermesic acid (EtOFK). Numbers correspond to retention times (min) after HPLC-MS separation of the novel glucosides formed *in vitro* (Table 2).

m/z $[M-H]^-$ values ---------- - Haworthia Soluble protein	475 m/z $[M-H]^-$ FK-monoglc	491 m/z $[M-H]^-$ KA-monoglc	489 m/z $[M-H]^-$ MeOFK-monoglc	503 m/z $[M-H]^-$ EtOFK-monoglc
Leaf	3.73	3.71	5.81	6.63
Gel				
Base	3.73	3.71	5.81	6.63
Root	3.73	3.71	5.81	6.63

Table 2. Glucosides formed in *in vitro* glucosylation assays using enzyme extracts from *Haworthia limifolia*.

The m/z values 475 and 491 are the same m/z values as are obtained for DcII and CA, respectively, solubilized in similar solutions. Both m/z values are 162 (m/z value of glucose in a glucoside) higher than the m/z values of the FK and KA indicating that the glucose moety from UDP-glucose in the reaction buffer has been transferred to the aglycone by a GT in the extract. The m/z [M-H] values 489 and 503 are also 162 higher than the m/z values obtained with MeOFK and EtOFK, respectively, indicating that a glucose unit has been added to both MeOFK and EtOFK by a GT present in the extract.

Conclusion

The results of this example demonstrate that herein relevant glycosyltransferase (GT) enzymes can be identified in *Aloe* plants and *Haworthia* plants.

Said in other words, *Aloe* plants and *Haworthia* plants comprise a glycosyltransferase which is capable of glycosylating flavokermesic acid in order to produce flavokermesic acid glycoside; and/or capable of glycosylating kermesic acid in order to produce kermesic acid glycoside.

Example 4 GT activity in Sorghum and rice plant

As known the art, *Sorghum* and rice plants comprise glycosyltransferases.
As known in the art, some of the *Sorghum* and rice glycosyltransferases may glycosylate low molecular weight aglycone compounds.

The in the art described glycosyltransferases from *Sorghum* and rice plants have significant less than 70% identity with amino acids 1 to 515 of SEQ ID NO:2 as disclosed herein.

It is not known in the art if glycosyltransferases of *Sorghum* and/or rice plants would be a herein relevant glycosyltransferase, i.e. a glycosyltransferase which is capable of glycosylating flavokermesic acid in order to produce flavokermesic acid glycosides; and/or capable of glycosylating kermesic acid in order to produce kermesic acid glycosides.

The known glycosyltransferases from *Sorghum* (*Sorghum bicolor*), SbUGT85B1, with Genbank ID number AF199453.1 (nucleotide seq.) / AAF17077.1 (polypeptide seq) and rice (*Oryza sativa*), OsCGT, with Genbank ID number FM179712.1 (nucleotide seq.) / CAQ77160.1 (polypeptide seq) were expressed in *E.coli* strain Xjb and crude *E.coli* proteins extracts were prepared and tested for glucosylation activity on the substrates kermesic acid and flavokermisic acid as described by Kannangara *et al*. (2011) and Augustin *et al.* (2012).

Figure 2 in PCT/EP2014/078540 (PCT filing date 18 December 2014) shows in LC-MS analyses of glucosylated products formed in assays containing crude lysate of *E.coli* strain Xjb expressing either SbUGT85B1 or OsCGT, UDP-glucose and flavokermesicc acid (FK) or kermesic acid (KA). As a negative control crude extract from the *E.coli* strain Xjb was used in the assays.

There were identified KA glycosides (491 m/z [M-H] – the m/z[M-H] value of CA) for both glycosyltransferases and FK glycosides (475 m/z [M-H] the m/z[M-H] value of DcII) for OsCGT.

Conclusion

The result of this example demonstrated that herein relevant glycosyltransferase (GT) enzymes can be identified in *Sorghum* and/or rice plants.

Said in other words, *Sorghum* and/or rice plants comprise a glycosyltransferase which is capable of glycosylating flavokermesic acid in order to produce flavokermesic acid glycoside; and/or capable of glycosylating kermesic acid in order to produce kermesic acid glycoside.

Example 5 Use of endogenous GT gene or GT activity

As known in the art glycosyltransferases able to glycosylate low molecular weight are present in a lot of different organisms. A method to contact the glycosyltransferase of the cells of an organism with a low molecular weight compound is to introduce one or more genes directing the biosynthesis of the low molecular weight compound and thus enabling the cells to glycosylate the low molecular weight compound. The low molecular weight compound may be e.g. flavokermesic acid or kermersic acid or decorated versions of these molecules.

One or more genes directing the biosynthesis of flavokermesic acid or kermesic acid or decorated version of these molecules are introduced into a glycosyltransferase containing organism, e.g. the tobacco plant, *Nicotiana benthamiana*.

When the gene/genes is/are transiently expressed according to the methods described in (D'Aoust et al. (Methods Mol Biol 483 (2009): 41-50) in e.g. plant tissue the low molecular weight compound or compounds is/are produced. Cells stably expressing the gene/genes are produced and selected according to the methods described in Gelvin (Microbiol Mol Biol Rev 67(1) (2003): 16-37)).

In cells containing either stably expressed and/or transiently expressed gene/genes the low molecular weight compounds come into contact with the endogenous glycosyltransferases, resulting in the formation of one or more glycosides of flavokermesic acid, kermesic acid or decorated versions of these molecules.

The presence of the glycosides is demonstrated by the extraction and the analytical methods described in Example 3.

Samples are prepared for LC/MS by the method for extraction described by (Rauwald and Sigler (Phytochemical Analysis 5 (1994):266-270).

Conclusion

The results of this example demonstrate that endogenous glycosyltransferases present in the cells of a recombinant organism can be used to convert flavokermesic acid, kermesic acid or decorated versions of these molecules into glycosides when a gene/genes directing the biosynthesis of the aglycons are introduced into the organism.

Said in other words introduction of a gene or genes directing the biosynthesis of flavokermesic acid, kermesic acid, decorated versions of these molecules, or related low molecular weight compounds is a method to bring the low molecular weight compound in contact with glycosyltransferases and thus a method to produced glycosides of flavokermesic acid, kermesic acid or decorated version of these compounds.

Example 6: ***Stable expression of type III PKSs in Aspergillus nidulans***

Materials and methods

Media and solutions

All solutions were prepared with Milli-Q H_2O and sterilized at 121 °C for 20 min.

Trace element solution: For 1 L mix 0.4 g $CuSO_4 \cdot 5H_2O$; 0.04 g $Na_2B_4O_7 \cdot 10H_2O$; 0.8 g $FeSO_4 \cdot 7H_2O$; 0.8 g $MnSO_4 \cdot 2H_2O$; 0.8 g $Na_2MoO_4 \cdot 2H_2O$; and 8.0 g $ZnSO_4 \cdot 7H_2O$.

Mineral Mix (50x): For 1 L solution mix 26 g KCl; 26 g $MgSO_4 \cdot 7H_2O$; and 76 g KH_2PO_4; and 50 ml Trace element solution.

20X Nitrate salts solution: For 1 L solution: dissolve 120 g $NaNO_3$, 10.4 g KCl, 10.4 g $MgSO_4 * H_2O$, 30.4 g KH_2PO_4 in Milli-Q water.

Trace element solution (1 L stock solution): For 1 L solution: dissolve 0.4 g $CuSO_4 \cdot 5H_2O$, 0.04 g $Na_2B_4O_7 \cdot 10H_2O$, 0.8 g $FeSO_4 \cdot 7H_2O$, 0.8 g $MnSO_4 \cdot 2H_2O$, 0.8 g $Na_2MoO_4 \cdot 2H_2O$, 8 g $ZnSO_4 \cdot 7H_2O$, up to 1 L Milli-Q water.

Thiamine 1%: Final concentration 0.001%.

D-glucose 20% (w/vol): For 1 L solution 200 g D-glucose in Milli-Q water

Aspergillus **protoplastation buffer (APB):** Final concentration 1.1 M $MgSO_4$ and 10 mM Na-phosphate buffer. pH is adjusted with 2 N NaOH to 5.8.

Aspergillus transformation buffer (ATB): Final concentrations: 1.2 M Sorbitol; 50 mM $CaCl_2 \cdot 2 H_2O$; 20 mM Tris; and 0.6 M KCl. pH is adjusted with 2 N HCl to 7.2.

PCT: Final concentration: 50 % w/vol PEG 8000 (4000, 6000 and other PEG can also be used); 50 mM $CaCl_2$; 20 mM Tris; and 0.6 M KCl. pH is adjusted with 2 N HCl to 7.5.

Minimal Media (MM): For 1 L of solution: 1 ml Trace elements; 50 ml nitrate salts (1M); 50 ml 20 %-w/vol glucose; 1mL Thiamine; 20 g Agar (So.Bi gel). For liquid MM Agar is not added.

Supplements to the media: Supplements were added to the different media if necessary, based on the genotype of the *Aspergillus* nidulans strain, in the following amounts: Arginine 4 mM, uracile 10 mM and uridine 10 mM. For counter selection of the *AfpyrG* marker 5-fluororotic acid (5-FOA) is added to a concentration of 1.3 mg/mL.

Solid Transformation Media (TM): For 1 L of solution: 1 ml Trace elements; 50 ml nitrate salts (1M); Sucrose 171.15 g; 1mL Thiamine; 20 g Agar (So.Bi gel). For liquid MM Agar is not added.

Transformation of Aspergillus nidulans

Aspergillus nidulans protoplasts were generated following a standard protocol of the prior art, shortly summarized here: *Aspergillus nidulans* macromedia are geminated overnight, and the resulting biomass was harvested by filtering through a Miracloth filter (Merck Millipore). The mycelium was re-suspended in 10 ml APB buffer containing 40 mg Glucanex/ml (Novozymes A/S). The mixture was incubated on a shaker at 37°C with 150 rpm for 3 hours. APB was added to yield a total volume of 40 ml. An overlay of 5 ml 50%

ATB and 50% MQ-water was applied and the tubes were centrifuged at 3000 RCF, 16 °C, for 12 minutes resulting in a two-phase system with the protoplast in the interphase. The protoplasts were washed using 40 ml of ATB and centrifugation at 3000 RCF, 16°C, for 12 minutes. The resulting pellet was re-suspended in 1 ml ATB.

The resulting protoplasts were used for genetic transformation experiments in aliquots of 50 µl mixed with 1.5-5 µg DNA and 150 µL PCT. The transformation mixture was incubated for 10 minutes. The mixture was then added 250 µL ATB and transferred to a transformation plate with required supplements dispatched with a Drigalski spatula.

Following incubation for 4 days at 37°C, the resulting *Aspergillus nidulans* transformants were isolated and sub-cultured on individual agar plates with a suitable selection regime.

Targeted integration of the expression cassette was analyzed by PCR using the original primers used for amplifying the gene to be expressed, followed by primer pairs with one primer annealing inside the insert and one in the surrounding genome. In the case of AfpyrG based strains, the selection marker was eliminated, following the transformation, by counter selection on 5-FOA containing plates and homologous recombination between short directional repeats surrounding the AfpyrG marker gene in the expression cassette. Following removal of the marker gene, the strain was again verified by diagnostic PCR. Removal of the selection marker gene allowed for a subsequent transformation round with a new target gene, using the same selection marker gene as used during the first transformation.

Construction of the Aspergillus nidulans host strains

The used *Aspergillus nidulans* strains are listed in Table 3 below and Table 4 summarizes the modified loci/genes. Targeted gene deletion (or replacement) was achieved by constructing gene targeting cassettes, consisting of a recyclable selection marker gene surrounded by two ap. 1500 bp sequences identical to the sequences surrounding the locus that should be replaced in the genome. The targeting construct for replacement of the *wA* and *yA* genes were constructed via the split-marker PCR-based method described in (Nielsen M.L., Albertsen L., Lettier G., Nielsen J.B., Mortensen M.H., 2006. Efficient PCR-based gene targeting with a recyclable marker for *Aspergillus nidulans*. Fungal Genetics and Biology, Vol. 43:54-64). First the *wA* gene was targeted for deletion and the required targeting sequences were amplified from genomic DNA using the primers: AN*wA*-dl-Up-F (5'- GGAAGAAGGTCGCATACCA-3') combined with AN*wA*-dl-Up-Rad (5'-gatccccgggaattgccatgGATCAGGAGAAGGAGAGTCAAG-3') and AN*wA*-dl-Dw-Fad (5'-aattccagctgaccaccatgGGCGAAAAGGCAAAGGAGC-3') combined with AN*wA*-dl-Dw-R (5'-GCTAGAAAAGGCAAGGGAGG-3'). The two marker fragments were amplified by combining the M1 primer (5'- catggcaattcccggggatc-GCCGGCAATTCTTTTTAGGTAGC-3') combined with the M2 primer (5'- CCAGAAGCAGTACACGGC-3') and the M3 primer (5'-GTTGTCTGCTTGCGCTTCTTC-3') with the M4 primer (5'-catggtggtcagctggaatTCCTCCGCCATTTCTTATTCCC-3'). Following PCR amplification of the

gene targeting and marker DNA fragments the fragments were fused by PCR, as described in Nielsen *et al.* 2006, the DNA fragments were gel purified and transformed into the recipient *Aspergillus nidulans* strain NID1 (described in Nielsen J.B, Nielsen M.L., Mortensen U.H, 2008, Transient disruption of non-homologous end-joining facilitates targeted genome manipulations in the filamentous fungus *Aspergillus nidulans*. Fungal Genetics and Biology, Vol. 45:165-170). The marker in the resulting strain was eliminated by counter selection on 5-FOA plates to identify transformants where the *AfpyrG* marker gene had looped out spontaneously. Targeted deletion of the *yA* gene was conducted as described for the *wA* gene, but using the following two primer pairs for amplifying the targeting sequences: Del-*yA*-5'-F (5'- GTGGGTTGAACCGCTTACTCAG-3') combined with Del-*yA*-5'-R (5'-gatccccgggaattgccatg-CCCGGAGGAATCAAAATGACGC-3') and Del-*yA*-3'F (5'-aattccagctgaccaccatgGTTTGGGATTCTTAGGTGAGCTC-3') combined with Del-*yA*-3'-R (5'-CCTCCCTGGCGTATACACAAAC-3'). The resulting *AfpyrG* marker free strain is referred to as NID598 in the subsequent description.

Targeted deletion of the asperthecin PKS (*aptA*) was performed in the NID598 background, using the experimental strategy described for the *yA* and *wA* gene deletions. The two targeting DNA fragments were PCR amplified using the primer AnAptA-UP-F (5'-GCTCGAGCTTGCCAGCC-3') combined with AnAptA-UP-R (5'- gatccccgggaattgccatg-GCTGGTGTTGGGACACACG -3') and the AnAptA-Dw-F (5'-aattccagctgaccaccatgGCTTGGAAATCAGTATAGCTTTCTG -3') combined with AnAptA-Dw-R (5'- GCTTGTGGTCTGTCTGAATCG -3'). The gel purified targeting construct was then transformed into the marker free NID598 strain, resulting in the strain NID930.

The *Apt*-cluster, *mdp*-cluster and *stc*-clusters were sequentially deleted in the NID598 background. For this, the required targeting constructs were constructed via directional Uracil-Specific Excision Reagent Cloning (USER) of the respective targeting fragments into the P1(P6) vector, as described in (Hansen B.G., Salomonsen B., Nielsen M.T., Nielsen J.B., Hansen N.B., Nielsen K.F., Regueira T.B., Nielsen J., Patil K.R., Mortensen U.H. 2011. Versatile enzyme expression and characterization system for *Aspergillus nidulans*, with the Penicillium brevicompactum polyketide synthase gene from the mycophenolic acid gene cluster as a test case. Appl Environ Microbiol. Vol. 77(9):3044-51). For deletion of the Apt-cluster the targeting fragments were amplified with the primers ANAPTcluster-Dl-Up-FU (5'-GGGTTTAAdUGAGGAGCAGAGGATGCGG-3') combined with ANAPTcluster-Dl-Up-RU (5'-GGACTTAAdUGTAGTGGTGGTGCTGGTG-3') and ANAPTcluster-Dl-Dw-FU (5'-GGCATTAAdUCGCGTGGAATTTGGAAGAGAG-3') combined with ANAPTcluster-Dl-Dw-RU (5'-GGTCTTAAdUGTGCTCGGGGACGTGAAAG-3'). The used primers each included a 2-deoxyuridin (dU) based to allow for the creation of 3' overhangs. The resulting PCR fragments were gel purified and directionally cloned by USER cloning into the PacI/Nt.BbvCI digested P1(P6) vector. The resulting plasmid was SwaI digested to liberate the targeting fragment, including the *AfpyrG* marker. The digested plasmid was transformed into the marker-free NID598 strain. The *AfpyrG* marker was subsequently removed by 5-

FOA counter selection, as described for *wA* above, to prepare the strain for deletion of the *mdp*-cluster. Targeted deletion of the *mdp* and *stc* clusters were conducted as described for the *apt* cluster. For construction of the targeting sequence for the mdp-cluster was PCR amplified using the ANMDPcluster-DI-Up-FU (5'-GGGTTTAAdUGGTCGTCTGTCAAGGAGTTG-3') primer combined with the ANMDPcluster-DI-Up-RU (5'-GGACTTAAdUGCAGTGCTGTATATGGGTCTTG-3') primer and the ANMDPcluster-DI-Dw-FU (5'-GGCATTAAdUGAGTTTGTGAGATGTTCAGGATGG-3') primer combined with the ANMDPcluster-DI-Dw-RU (5'-GGTCTTAAdUGAGGTGAAGGACACAGCG-3') primer. Moreover, the targeting sequence for the *stc* cluster were PCR amplified with the primers; ANSTCcluster-DI-Up-FU (5'-GGGTTTAAdUCGCAGAGACTAGGACACAAGTG-3') combined with ANSTCcluster-DI-Up-RU (5'-GGACTTAAdUGCGGCGATCTGTGGTAGAG-3') and ANSTCcluster-DI-Dw-FU (5'-GGCATTAAdUGCCAGCATATTCAAACCCAGTC-3') combined with ANSTCcluster-DI-Dw-RU (5'-GGTCTTAAdUCACACAACCAACCTCCGATC-3'). The resulting strain with the deletion of the *apt*, *mdp* and *stc* clusters is referred to as NID_SMA.

Table 3: The different Aspergillus nidulans strains used in the study

Strain name	**Genotype**
NID1	*nkuA*Δ *argB2*, *pyrG89*, *veA1*
NID598	*nkuA*Δ, *argB2, pyrG89, veA1, wA*Δ, *yA*Δ
NID930	*nkuA*Δ, *argB2, pyrG89, veA1, wA*Δ, *yA*Δ, *aptA*Δ
NID_SMA	*nkuA*Δ, *argB2, pyrG89, veA1, wA*Δ, *yA*Δ, aptΔ, mdpΔ, stcΔ

Table 4: The affected genes

Gene names and accession numbers in the Aspergillus-genome-database:	
yA = AN6635	Ascospore pigment biosynthesis (Laccase)
wA = AN8209	Ascospore pigment biosynthesis (PKS)
aptA = AN6000	Asperthecin PKS
apt = aptA (AN6000) to aptC (AN6002)	Asperthecin gene cluster
mdp = mdpL (AN10023) to mdpA (AN10021)	Monodictyphenone/emondin gene cluster
stc = stcW (AN7804) to stcA (AN7825)	Sterigmatocystin gene cluster

Expression of type III PKS in Aspergillus nidulans

Synthetic (de novo synthesized) codon optimized versions of *HpPKS2*, *AaOKS* were made for yeast expression, and sequences with the natural codon usage were purchased from Genscript. Codon optimized genes are denoted with a 'ScOpt' suffix, e.g. *HpPKS2*-ScOpt, while genes with the original codon usage from the natural host is denoted with an 'Orig'

suffix, e.g. *HpPKS2*-Orig. The synthetic DNA fragments were used as PCR template for PCR reactions with specific primers (IDT) for the coding sequences of the genes including stop codons. The used primers contained 5' overhangs compatible with Uracil Specific Excision Reagent (USER™) cloning of the resulting PCR amplicons into expression vectors for targeted integration of the expression cassettes into one of seven possible sites in the *Aspergillus nidulans* genome. The plasmids features a fungal selection marker gene (*Aspergillus fumigatus AfpyrG*), a USER cloning site (AsiSI and Nb.btsI) and is flanked by two 1-1.5 kb *Aspergillus nidulans* DNA sequences (named Up and Down) to allow for integration into the *Aspergillus nidulans* genome by targeted homologous recombination (Hansen, B. G., Salomonsen, B., Nielsen, M. T., Nielsen, J. B., Hansen, N. B., Nielsen, K. F., … Mortensen, U. H. (2011)). Versatile enzyme expression and characterization system for *Aspergillus nidulans*, with the *Penicillium brevicompactum* polyketide synthase gene from the mycophenolic acid gene cluster as a test case. Applied and Environmental Microbiology, 77(9), 3044–3051. http://doi.org/10.1128/AEM.01768-10). The PCR amplicons were amplified using PfuX7 DNA polymerase and the resulting DNA fragments were gel purified. The recipient vectors were prepared for USER™ cloning by digesting it with AsiSI and Nb.btsI overnight, followed by gel purification. The PCR amplicons were directionally cloned into the recipient vector by USER™ cloning, combined with transformation into *E. coli*. DH5α cells were thawed on ice, and the USER™ mix was added to 50 µL of cells ($1x10^8$ cfu/µg pUC19). The mixture is placed on ice for 10 minutes, and heat shocked for 90 seconds at 45°C. Transferred to ice again and incubated for 5 minutes. The cells are plated on a LB-plate with selective antibiotic and incubated at 37°C over night. The colonies are verified by PCR and the true transformants are grown in liquid LB-media with selective antibiotic over night for plasmid purification (Taylor, R. G., Walker, D. C., & McInnes, R. R. (1993). E. coli host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. Nucleic Acids Research, 21(7), 1677–8. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=309390&tool=pmcentrez&rendertype=abstract).

Primers used for the construction process, where dU represents 2-deoxyuridin:

An_HpPKS2-ScOpt-F	5'- AGAGCGAdUATGGGTTCCTTAGACAACGGTTC
An_HpPKS2-ScOpt-R	5'- TCTGCGAdUTCACAAAGGAACACTTCTCAAAACC
An_AaOKS-ScOpt-F	5'- AGAGCGAdUATGAGTTCACTCTCCAACGCTTCC
An_AaOKS-ScOpt-R	5'- TCTGCGAdUTCACATGAGAGGCAGGCTGTG
An_HpPKS2-Orig-F	5'- AGAGCGAdUGGGTTCCCTTGACAATGGTT
An_HpPKS2-Orig-R	5'- TCTGCGAdUTTAGAGAGGCACACTTCGGAGAAC
An_AaOKS-Orig-F	5'- AGAGCGAdUATGAGTAGTTTATCAAATGCCAGTC
An_AaOKS-Orig-R	5'- TCTGCGAdUTCACATCAATGGCAAGGAA

The verified expression plasmids were digested with SwaI (with *AfpyrG* marker), to liberate the expression cassette flanked by the up and down targeting sequences. The cassette was introduced into *Aspergillus nidulans* protoplast as specified above.

The following expression plasmids and were constructed:

pIS53(pyrG)::HpPKS2_ScOpt	HpPKS2_ScOpt in IS53 locus
pIS53(pyrG)::HpPKS2_Orig	HpPKS2_Orig in IS53 locus
pIS53(pyrG)::AaOKS_ScOpt	AaOKS_ScOpt in IS53 locus
pIS53(pyrG)::AaOKS_Orig	AaOKS_Orig in IS53 locus

The expression cassettes of the individual plasmids were liberated by SwaI digestion (with *AfpyrG* marker) and subsequently transformed into *Aspergillus nidulans* protoplasts.
Following construction of the individual strain the used *AfpyrG* markers were eliminated by 5-FOA counter selection.

Chemical analysis of the constructed Aspergillus nidulans strains

To access the effects the introduced genes had on the metabolism of the constructed *Aspergillus nidulans* strains, the strain were grown on MM, with appropriate supplements, for 5 to 14 days at 37°C. Metabolites were extracted by micro-scale extraction procedure described by Smedsgaard (Smedsgaard, J. (1997). Micro-scale extraction procedure for standardized screening of fungal metabolite production in cultures. Journal of Chromatography A, 760(2), 264–270. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S0021967396008035). Summarized: Ten 0.4-0.6 mm circular plugs were cut from the culturing plates using a 'cork borer' and placed in a 8 ml glass vial containing 2 ml of extraction solution made of methanol and 1 % (v/v) formic acid. The tubes were ultrasonicated for 60 minutes in a water bath. The supernatant was transferred to a new vial and evaporated under a stream of nitrogen gas and heat (30°C), and the resulting pellet was dissolved in 100 µl of HPLC grade methanol and 1% formic acid (Sigma-Aldrich) by incubating the solution for 20 minutes in an ultrasonic bath. The resulting extraction solution was then filtered through a PTFE 0.45 um, 15 mm Premium Syringe Filters (Agilent Technologies) into HPLC vials. The extracted metabolites were analyzed using Agilent 1200 HPLC coupled to a Bruker micrOTOF-Q II mass spectrometer equipped with an electrospray ionization source. Chromatographic separation was obtained on a Luna C_{18} (2) column (150 × 4.6 mm, 3 µm, 100 Å, Phenomenex) maintained at 40°C. The aqueous eluent (A) consisted of water/acetonitrile (95:5, v/v) and the organic eluent (B) consisted of water/acetonitrile (5:95, v/v); both acidified with 0.1% formic acid According to the purpose of the analysis, two different linear gradient elution profiles has been used:

Method 1: 0 min, 0% B; 20 min, 100% B; 22 min 100% B; 24 min, 0% B. The flowrate was maintained at 0.8 mL/min and 7 min equilibration was used.

Method 2: 0 min, 0% B; 30 min, 100% B; 33 min 100% B; 35 min, 0% B. The flowrate was maintained at 0.5 mL/min and 10 min equilibration was used.

The samples were analyzed both in positive and negative ionization mode. The resulting dataset was analyzed using the following Software MassHunter (Agilent technologies).

The resulting Base Peak Chromatograms (BPC) was inspected to identify changes in the metabolism of the fungus. While Extracted Ion Chromatograms (EIC) for masses fitting the expected intermediates in the biosynthetic pathway was generated to ease identification of the relevant peaks.

Alternatively, the production of metabolites was analyzed in liquid MM cultures. For this 500 mL sterile Erlenmeyer flasks with 100 ml of liquid minimal medium, with appropriate supplements, were inoculated with a solution of macroconidia from the strains to be analyzed. The flasks were incubated for 1-15 days at 37°C in a shake incubator with 150 rpm. The mycelium was poured through Miracloth (Merck Millipore) and separated from culture media. The mycelium was fast frozen with liquid nitrogen. The media was sterile filtered through an FPE-214-250 JET BIOFIL®. The sterile media was then transferred to a 40 mL vial for chemical analysis.

Extraction

Fermentation broth (40 mL) was evaporated to dryness in vacuo and extracted with ethanol + 10% formic (20 mL) acid by means of sonication (2h) at 60°C. The solvent was decanted and dried in vacuo before being dissolved in 1.5 mL ethanol + 10% formic acid using sonication. The supernatant was dried in a speedvac and dissolved in 180 µL 90% methanol + 10% formic acid, centrifuged and analyzed by HPLC-MS/MS

Results:

Analysis of the formed metabolites by UHPLC-HRMS of the strains comprising heterologous *HpPKS2 or AaOKS* showed that expression of the type III PKSs (*HpPKS2 or AaOKS*) resulted in the production of different compounds including the compounds SEK4, SEK4B and FK. Expression in the SMA cluster deletion strain resulted in a higher ratio of FK and SEK4/SEK4B.

Conclusion:

Expression of plant type III PKS (*HpPKS2* or *AaOKS*) resulted in the production of different compounds including the compounds SEK4, SEK4B, FK anthrone, Mutactin and FK *in vivo* in *Aspergillus nidulans*. Expression in the strain with deleted asperthecin, sterigmatocystin and monodictyphenone/emodin gene clusters had a positive impact on production of the metabolites.

Since there in this Example 6 were not inserted heterologous cyclases and/or aromatases into the *Aspergillus* strains and FK compound was identified – it indicates that *Aspergillus*

strains may comprise homologous cyclases and/or aromatases capable of converting in vivo non-reduced octaketide into FK compound.

Example 7: *Stable expression of type III PKSs and cyclases and/or aromatases in Aspergillus nidulans*

Materials and methods

Expression of bacterial cyclases and aromatases in Aspergillus nidulans

Synthetic codon optimized version of *ZhuI*, *ZhuJ* for *Aspergillus*, *S. cerevisiae* and *E. coli* expression were purchased from Genscript. Similarly as described for the type III PKSs the cyclases and aromatases encoding genes were cloned into expression vectors targeting specific sites in the *Aspergillus* genome, different from the site used for expression of the Type III PKS (see Example 6). Said in other words, the *ZhuI* and *ZhuJ* genes were cloned into the *Aspergillus* strains of Example 5 that already comprised the heterologous type III PKS genes (*HpPKS2*, *AaOKS*).

Primers used for the construction process, where dU represents 2-deoxyuridin:

An_ZhuI_EcOpt-F	5'- AGATATACCAdUGCGTCATGTTGAACATACCGT
An_ZhuI_EcOpt-R	5'- ATGGCTGCdUTTATGCGGTAACTGTACCAACACCA
An_ZhuJ_EcOpt-F	5'- ATATACATAdUGAGCGGTCGTAAAACCTTT
An_ZhuJ_EcOpt-R	5'- ATATCCAATdUTTAATCCTCTTCTTCTTGTTC
An_ZhuI_ScOpt-F	5'-AGAGCGAdUGAGACACGTTGAACACA
An_ZhuI_ScOpt-R	5'-TCTGCGAdUTTATGCAGTTACGGTACCA
An_ZhuJ_ScOpt-F	5'-AGAGCGAUGTCCGGTAGAAAGACCTT
An_ZhuJ_ScOpt-R	5'-TCTGCGAUTTAATCTTCTTCTTCTTGTTCG

The PCR amplified coding sequences were cloned into vectors for targeted integration in the *Aspergillus nidulans* genome: *ZhuI* in IS80 and *ZhuJ* in IS82. The verified expression cassettes were transformed into *Aspergillus nidulans*, by sequential targeted integration of the individual expression cassette, as described in Example 6. The *AfpyrG* marker was eliminated following each transformation round to allow for a subsequent transformation round introducing an additional expression cassette.

Chemical analysis was performed as described in Example 6.

Results:

Analysis of the formed metabolites by UHPLC-HRMS showed that expression of ZhuI and ZhuJ did not significantly affect amount of produced metabolites (including the compounds SEK4, SEK4B and FK), i.e. the amounts produced by the strains of this Example 7

(comprising PKS (*HpPKS2 or AaOKS*) + ZhuI and ZhuJ) were similar to the amounts produced by the strains of Example 6 (comprising only PKS (*HpPKS2 or AaOKS*)).

Conclusion:

The results showed that expression of ZhuI and ZhuJ optimized for expression in *E. coli* did not significantly affect the produced metabolites in *Aspergillus nidulans*.

Example 8: *Heterologous expression of type III PKSs and Dactylopius coccus C-glycosyltransferase in Aspergillus nidulans*:

Materials and methods

Construction of plasmids, genetic transformation of *Aspergillus nidulans* and chemical analysis was performed as described in Example 6.

Construction of vectors for expression of the Dactylopius coccus C-glycosyltransferase in Aspergillus nidulans

The *DcUGT2* gene was cloned/inserted into PKS (*AaOKS*) *Aspergillus nidulans* strain of Example 6.

The full length *DcUGT2* gene from *Dactylopius coccus* was codon optimized for expression in *Aspergillus nidulans* and purchased from GenScript as synthetic DNA. For PCR based amplification of the *DcUGT2* coding sequence the following primers were used: DcUGT2_AnOpt-F: 5'-AGAGCGAdUATGGAGTTTCGCTTGCTTATCCT and DcUGT2_AnOpt-R: 5'-TCTGCGAdUTTAATTCTTCTTCAACTTTTCCGACTTAG. The resulting PCR amplicon was cloned into an *Aspergillus nidulans* expression vector as described in Example 6. The used expression vector targeted the IS52 site in the genome of *Aspergillus nidulans*.

Results:

Analysis of the formed metabolites the *Aspergillus nidulans* strains comprising heterologous *AaOKS* and *DcUGT2* showed that expression of resulted in the production of different compounds including the compounds CA, DCII, KA. The compounds FK, SEK4 and SEK4b were also identified, and co-expression of the DcUGT2 gave a higher yield of the before mentioned compounds.

Conclusions

Co-expression of the heterologous PKS (*AaOKS)* and glycosyltransferase *(DcUGT2)* in *Aspergillus nidulans* resulted in production of CA, DCII, KA. The compounds FK, SEK4 and SEK4b were also identified, and co-expression of the DcUGT2 gave a higher yield of the before mentioned compounds.

Example 9: *Expression of AaOKS in N. benthamiana and in vitro activity test*

Expression of AaOKS in *N. benthamiana*

AaOKS (Gene bank accession AY567707) (Abe I, Oguro S, Utsumi Y, Sano Y, Noguchi H (2005b). Engineered biosynthesis of plant polyketides: chain length control in an octaketide-producing plant type III polyketide synthase. J Am Chem Soc 127: 12709-12716.; Abe I, Utsumi Y, Oguro S, Morita H, Sano Y, Noguchi H (2005a). A plant type III polyketide synthase that produces pentaketide chromone. J Am Chem Soc 127:1362-1363) was synthesized by Genescript and amplified with primers containing a C-terminal his-tag and USER overhang for cloning into the USER-compatible pEAQ vector (Sainsbury F, Thuenemann EC, Lomonossoff GP (2009) pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. Plant Biotechnol J 7: 682-693) with the USER-method (Nour-Eldin HH, Hansen BG, Nørholm MH, Jensen JK, Halkier BA (2006). Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. Nucleic Acids Res 34: e122.) to give the resulting plasmid pEAQ-AaOKS. The plasmid was propagated in *E. coli* TOP10 cells and verified by sequencing. The plasmid was transformed into *Agrobacterium tumefaciens* by electroporation and positive clones were selected on Luria-Bertani (LB) agar supplemented with rifampicin (12.5 µg/mL.), kanamycin (50 µg/mL), and ampicillin (25 µg/mL) followed by PCR to verify the presence of AaOKS. For expression, *A. tumefaciens* harboring pEAQ-AaOKS or pEAQ was cultured overnight in liquid LB supplemented with rifampicin (12.5 µg/mL.), kanamycin (50 µg/mL) and ampicillin (25 µg/mL). The cells were pelleted by centrifugation and resuspended in infiltration medium (10 mM MES, 100 µM acetosyringone, 10 mM $MgCl_2$) to a final OD_{600}: 0.5 and incubated at room temperature for 1 hr. *A. tumefaciens* was infiltrated into the leaves of 3-4 weeks old *N. benthamiana* plants by a needle-less syringe and the plants grew for a further 5 days in the green house.

Extraction of soluble protein from *N. benthamiana*

300 mg of plant material was grinded in a cold mortar with pestle and the addition of 0.5 g PVPP and 10 mL cold extraction buffer (50 mM sodium phosphate, 400 mM sucrose, 4 mM DTT, pH 7.2 containing Complete proteinase inhibitor (Roche)). The homogenate was filtered through nylon mesh into cold centrifuge tubes and centrifuged at 10.000 xg at 4 °C for 15 min. The supernatant was isolated and centrifuged at 100.000 xg at 4 °C for 60 min. The soluble fraction (supernatant) was isolated. Production of his-tagged AaOKS was verified by western blotting using anti-his antibody. His-tagged AaOKS was purified from the soluble fraction using His Spin Trap columns (GE health care biosciences) according to the manufacturer's instructions.

Polyketide synthase assay

Crude soluble enzyme extract of *N. benthamiana* leaves and affinity purified his-AaOKS were tested for polyketide synthase activity. The standard assay contain 100 µl soluble protein or 20 µl affinity purified his-AaOKS in a total reaction volume of 250 µl containing 50 mM sodium phosphate, 500 mM sucrose, 1 mM DTT, 40 µM malonyl-CoA, 20 µM acetyl-coA, pH 7. The standard assay was supplemented with [2-^{14}C]-malonyl-CoA (1.8 mCi/mmol) for TLC analysis or [^{13}C3]-malonyl-CoA for structural analysis. The reactions were incubated at 30 °C for 90 min, terminated by the addition of 25 µl 20% HCl and extracted twice with 500 µl ethyl acetate, 1% (w/v) acetic acid. The organic phase was dried in vacuo and resuspended in 20 µl methanol, 1% (w/v) acetic acid and analyzed by TLC or LC-MS.

Results - In vitro enzyme activity of AaOKS:

Crude enzyme extract from *N. benthamiana* expressing AaOKS were tested for octaketide synthase activity using malonyl-CoA as starter substrate. When ^{14}C-malonyl-CoA was incubated with crude enzyme extract or affinity-purified AaOKS the substrate was metabolized, as observed by the appearance of radioactive band on thin layer chromatogram, whereas no products were observed when using crude enzyme extract from tobacco plants infiltrated with plasmid alone. For structural elucidation ^{13}C-malonyl-coA was used as the substrate and the labelled products were identified by LC-MS as SEK4 and SEK4B. There was no difference in product formation when AaOKS was tagged with a C-terminal his-tag and this shows that the HIS-tag did not interfere with enzymatic activity. The results showed that AaOKS heterologously expressed in *N. benthamiana* has octaketide synthase activity and that the affinity purified enzyme *in vitro* produces the octaketide-derived shunt products SEK4 and SEK4B.

Conclusion:

The results showed that *AaOKS* heterologously expressed in *N. benthamiana* has octaketide synthase activity and that the affinity purified enzyme *in vitro* produces the octaketide-derived shunt products SEK4 and SEK4B.

Example 10: ***Cloning truncated versions of DcUGT2 in S. cerevisiae and test of their in vitro activity towards FK and KA***

Materials and methods

Expression truncated DcUGT2 gene constructs in S. cerevisiae

Truncated forms of *DcUGT2* were generated lacking both the signal sequence and membrane anchor (*ΔSP_DcUGT2ΔMD-Strep*) or the membrane anchor (*DcUGT2ΔMD-Strep*)

alone. The fragments encoding *ΔSP_DcUGT2ΔMD-Strep* and *DcUGT2ΔMD-Strep* were amplified separately from the *pYES-DEST52-DcUGT2-Strep* plasmid using specific PCR primers to incorporate a C-terminal StrepII-tag (see table below). Gateway recombination sites, attB1 and attB2, were introduced into the generated fragments in a following PCR using forward primer: 5'-ggggacaagtttgtacaaaaaagcaggct-3' and reverse primer: 5'-ggggaccactttgtacaagaaagctgggt-3'. *ΔSP_DcUGT2ΔMD-Strep* and *DcUGT2ΔMD-Strep flanked with attB sites* were cloned into *pDONR207* plasmid (Invitrogen) and transferred into destination plasmid, *pYES-DEST52* (Invitrogen) by using the Gateway technology system. The two *pYES-DEST52* plasmid constructs were transformed separately into the Invsc1 yeast strain (Invitrogen) and positive transformants were verified by PCR. Heterologous protein production was performed according to the instructions of the *pYES-DEST52* gateway plasmid (Invitrogen). Production of heterologous StrepII-tagged protein was verified by western blotting using anti-Strep antibody. A membrane bound and a soluble protein fraction were prepared from verified yeast transformants as described in (Pompon, D., Louerat, B., Bronine , A., Urban, P. (1996). Yeast expression of animal and plant P450s in optimized redox environments. Methods Enzymol. 272:51–64) and screened for glucosylation activity towards flavokermesic acid/kermesic acid as described above.

Example 11: *In planta production of carminic acid and DcII by heterologous expression of genes encoding PKS, cyclase and full length DcUGT enzymes in Nicotiana benthamiana.*

Transient expression of gene constructs in Nicotiana benthamiana

Synthetic DNA fragments encoding *ZhuI* (Genbank accession: AAG30197) and *ZhuJ* (Genbank accession: AAG30196) codon optimized for *N. benthamiana expression* and *AaOKS* (Genbank accession: AY567707) were purchased from Genscript. All synthetic fragments and the herein described *pYES-DEST52-DcUGT2-Strep* plasmid were used as PCR templates with compatible deoxyuracil(dU)-containing primers (see Table 5 below) to generate constructs that were cloned into *pEAQ-HT-USER* by USER technology. The truncated *DcUGT2* version, *DcUGT2ΔMD-Strep*, was transferred from the *pDONR207* plasmid (Invitrogen) into destination plasmid, *pEAQ-HT-DEST1* (Sainsbury, F., Saxena, P., Geisler, K., Osbourn, A., Lomonossoff, G.P. (2012). Using a Virus-Derived System to Manipulate Plant Natural Product Biosynthetic Pathways. Methods Enzymol. 517:185–202), using the Gateway technology system. All *pEAQ-HT* plasmid constructs were transformed into the *Agrobacterium tumefaciens* strain, *AGL-1* and infiltrated into leafs of *N. benthamiana* plants as described in (Bach, S.S., Bassard, J.É., Andersen-Ranberg, J., Møldrup, M.E., Simonsen, H.T., Hamberger, B. (2014). High-Throughput Testing of Terpenoid Biosynthesis Candidate Genes Using Transient Expression in Nicotiana

benthamiana. In M Rodríguez Concepción, ed, Plant Isoprenoids, Methods in Molecular Biology, Vol. 1153. Humana Press, New York).

Table 5: Primer sequences for amplification of different gene constructs

Gene fragments	**Primer sequence**	
AaOKS	Forward	5′-GGCTTAA/dU/ATGAGTTCACTCTCCAACGCTTCCCATC-3′
	Reverse	5′-GGTTTAA/dU/TTACATGAGAGGCAGGCTGTGGAGAAGGATAGT-3′
ZhuI	Forward	5′-GGCTTAA/dU/ATGAGGCATGTCGAGCAT-3′
	Reverse	5′-GGTTTAA/dU/TTATGCCGTGACAGTTCCGACAC-3′
ZhuJ	Forward	5′-GGCTTAA/dU/ATGTCCGGACGTAAGACG-3′
	Reverse	5′-GGTTTAA/dU/TTAATCTTCCTCCTCCTGTTCAA-3′
DcUGT2 -Strep	Forward	5′-GGCTTAA/dU/ATGGAATTCAGATTGTTGATATTGGCCT-3′
	Reverse	5′-GGTTTAA/dU/TTATTTTTCGAATTGTGGATGAGACCAAGCAGA-3′
DcUGT2 ΔMD-Strep	Forward (attB1)	5′-ggggacaagtttgtacaaaaaagcaggct-3′
	Reverse	5′-TTATTTTTCGAATTGTGGATGAGACCAAGCAGAGTGCAAAAAGGCACCTGCAGT-3′

Metabolite extraction and LC-MS/MS analysis

Metabolites were extracted from discs (Ø=3cm) of agroinfiltrated *N. benthamiana* leaves. Leaf discs, excised with a cork borer, were flash frozen in liquid nitrogen. 0.5 ml of extraction buffer (85 % (v/v) methanol, 0.1 % (v/v) formic acid), equilibrated to 50 °C, were added to each frozen leaf disc followed by incubation for 1 hour at 50 °C, agitating at 600 rpm. The supernatant was isolated and passed through a Multiscreen$_{HTS}$ HV 0.45 µm filter plate (Merck Milipore). The filtered supernatant was subjected to LC-MS/MS analysis which was performed on an Agilent 1200 HPLC coupled to a Bruker micrOTOF-Q II mass spectrometer equipped with an electrospray ionization source. Chromatographic separation was obtained on a Luna C_{18}(2) column (150 × 4.6 mm, 3 µm, 100 Å, Phenomenex) maintained at 40 °C. The aqueous eluent (A) consisted of water/acetonitrile (95:5, v/v) and the organic eluent (B) consisted of water/acetonitrile (5:95, v/v); both acidified with 0.1% formic acid. According to the purpose of the analysis, two different linear gradient elution profiles were used:

Method 1: 0 min, 0% B; 30 min, 100% B; 33 min 100% B; 35 min, 0% B. The flow rate was maintained at 0.5 mL/min and 10 min equilibration was used. Retention times were 15.5 min for dcII, 15.6 min for carminic acid (CA), 17.1 min for flavokermesic-*O*-Glc 2 (FK-*O*-Glc 2), 17.2 min for SEK4, 17.7 min for SEK4B, 22.7 min for flavokermesic acid (FK) and 22.7 min for kermesic acid (KA).

Method 2: 0 min, 10% B; 25 min, 20% B; 27 min 100% B; 35 min, 100% B; 36 min, 10% B. The flow rate was maintained at 0.5 mL/min and 10 min equilibration was used.

Retention times were 16.1 min for flavokermesic-*O*-Glc 1 (FK-*O*-Glc 1), 17.0 min for dcII, 18.2 min for carminic acid (CA), 24.1 min for SEK4, 25.5 min for flavokermesic-*O*-Glc 2 (FK-*O*-Glc 2), 26.8 min for SEK4B, 35.5 min for flavokermesic acid (FK) and 36.0 min for kermesic acid (KA).

O-glucosides were identified with viscozyme L- treatment and by the neutral loss of 162 Da in the MS/MS spectrum.

Results:

The AaOKS of *Aloe arborescens* has previously been characterized to be an octaketide synthase belonging to the type III PKS enzyme class. To investigate the possibility of using such a type III PKS enzyme to generate the octaketide precursor required for carminic acid production, the *AaOKS* gene was transiently expressed in *N. benthamiana*. The *AaOKS* gene, when compared to empty plasmid control, results in formation of SEK4 and SEK4B after agroinfiltration in to leaves of *N. benthamiana*. This demonstrates that AaOKS can function as an active octaketide synthase *in vivo* in *N. benthamiana*. Furthermore since no flavokermesic acid (FK) anthrone or FK could be observed when *AaOKS* is agroinfiltrated alone, *N. benthamiana* may lack endogenous enzymes to further metabolize SEK4 and SEK4B into these compounds. Thus the cyclase genes, *ZhuI* and *ZhuJ,* were co-agroinfiltrated with *AaOKS*. ZhuI and ZhuJ originate from the R1128 antibiotic biosynthetic pathway and have previously been combined *in vivo* with the actinorhodin minimal PKS in *Streptomyces coelicolor* to produce FK (Tang, Y., Lee, T.S., Khosla, C. (2004). Engineered biosynthesis of regioselectively modified aromatic polyketides using bimodular polyketide synthases. PLOS Biol. 2(2):E31). In *N. benthamiana*, when *ZhuJ* is co-agroinfiltrated with *AaOKS*, accumulation of FK and flavokermesic acid-*O*-glucoside, FK-*O*-Glc 2 is observed and this is not the case when *AaOKS* is co-agroinfiltrated with *ZhuI*. The production of FK-*O*-Glc 2 suggests that one or several endogenous *N. benthamiana* UGT(s) are capable of efficiently using FK as substrate. Additionally, the co-agroinfiltration of *ZhuI* with *AaOKS* reduces the production of SEK4B by promoting the formation of SEK4 which is in accordance with the previous finding that ZhuI directs the initial C7-C12 cyclization (rather than the C10-C15 cyclization which results in SEK4B formation) of the linear octaketide (Ames, B.D., Lee, M.Y., Moody, C., Zhang, W., Tang, Y., Tsai, S.C. (2011). Structural and biochemical characterization of ZhuI aromatase/cyclase from the R1128 polyketide pathway. Biochemistry. 39: 8392-8406). This initial C7-C12 cyclization is also required for

the biosynthesis of the anthraquinone backbone of FK/CA and when *ZhuI* is co-agroinfiltrated with *AaOKS* and *ZhuJ*, the total pool of FK and flavokermesic acid-*O*-glucoside is indeed markedly increased as compared to when it is absent.

In planta production of DCII and CA was detected when full length *DcUGT2_Strep* was co-expressed with *AaOKS*, *ZhuI* and *ZhuJ* in *N. benthamiana*. The production of CA is only found in trace amounts when compared to levels of DCII. This CA production is likely due to an unspecific endogenous monooxygenase activity in *N. benthamiana* that either is capable of hydroxylating FK to kermesic acid (KA) or DCII to CA. However, it is noteworthy that KA is not detectable in any of the agroinfiltration studies, indicating that this compound is either produced at levels below our detection limit or it may be highly toxic for *N. benthamiana* and therefore rapidly metabolized into unknown products.

Conclusions:

The results of this Example demonstrated:

(i): The *AaOKS* gene from *Aloe arborescens* was transiently expressed in *N. benthamiana* and *in vivo* this resulted in formation of SEK4 and SEK4B, which demonstrated that AaOKS can function as an active octaketide synthase *in vivo* in *N. benthamiana*;

(ii): Since no flavokermesic acid (FK) anthrone or FK could be observed when *AaOKS* was agroinfiltrated alone, *N. benthamiana* may lack endogenous enzymes to further metabolize SEK4 and SEK4B into these compounds.

(iii): The cyclase genes *ZhuI* and *ZhuJ,* from *Streptomyces* sp. R1128, were co-agroinfiltrated with *AaOKS* and when *ZhuJ* was co-agroinfiltrated with *AaOKS*, accumulation of FK and flavokermesic acid-*O*-glucoside, FK-*O*-Glc 2 was observed. When *ZhuI* was co-agroinfiltrated with *AaOKS* and *ZhuJ*, the total pool of FK and flavokermesic acid-*O*-glucoside was indeed markedly increased as compared to when it is absent. Accordingly, heterologous expression of *Streptomyces* sp cyclase genes resulted in *in vivo* the production of different compounds including FK;

(iv): *In vivo* production of DCII and CA was detected when full-length *DcUGT2_Strep* was co-expressed with *AaOKS*, *ZhuI* and *ZhuJ* in *N. benthamiana*.

Example 12: *Heterologous expression of a truncated version of DcUGT2 and test of its in planta activity in N. benthamiana.*

To investigate whether the DcUGT2 protein, could be solubilized and still retain glucosylation activity towards KA and FK, the membrane anchor was deleted. DcUGT2 is predicted via bioinformatics to be located to the ER with the C-terminal anchored in the ER membrane and the N-terminal active site facing the lumen. Therefore it was speculated whether the protein might be *N*-glycosylated and, if so this posttranslational modification might be required for the activity of the DcUGT2 enzyme. An *N*-glycosylation bioinformatics

prediction analysis was carried out, identifying 3 putative *N*-glycosylation sites. To investigate whether glycosylation and the ER targeting play a role for the activity of the DcUGT2 enzyme, a truncated DcUGT2 form was generated where the signal peptide was intact but the membrane anchor deleted (DcUGT2ΔMD-Strep). This DcUGT2ΔMD-Strep protein is active *in vivo* and produced DCII when co-expressed with *AaOKS* and *ZhuJ* in *N. benthamiana.* The production of DCII was lower in these plants compared to the DCII levels in plants co-agroinfiltrated with the *DcUGT2*. This could imply that although the DcUGT2ΔMD-Strep protein is active, the level of the enzyme activity may be compromised by the deletion of the membrane anchor. The fact that DcUGT2ΔMD-Strep still possesses some glucosylation activity towards FK and/or KA indicates further that *N*-glycosylation and (or) ER-targeting of the protein might be crucial for the activity of the enzyme. Additionally it should be pointed out that the full-length DcUGT2 appears to compete efficiently with the *N. benthamiana* FK *O*-glucosylation activity, thereby significantly reducing the pool of flavokermesic acid-*O*-Glc (FK-*O*-Glc 1 and 2) and this is not observed for the truncated DcUGT2ΔMD-Strep version.

Conclusions

The results of this Example demonstrated that *DcUGT2ΔMD-Strep* (lacking the membrane anchor - amino acids 1 to 468 of SEQ ID NO:2) was active *in vivo* in *N. benthamiana* plant. The activity of *DcUGT2ΔMD-Strep* (lacking the membrane anchor) was less than full-length DcUGT2 enzyme – however, the fact that it was active may indicate that it could be particular useful for heterologous production in e.g. prokaryotic organisms.

REFERENCES

1: US5424421 (European Colour, published 1995)

2: WO2006/056585A1 (Chr. Hansen A/S)

3: Stathopoulou et al (Analytica Chimica Acta 804 (2013) 264– 272)

4: Zagrobelny et al (Cyanogenic glucosides and plant-insect interactions; Phytochemistry. 2004 Feb;65(3):293-306)

5: Geuder et al (Journal of Chemical Ecology, Vol. 23, No. 5, 1997)

6: Genta et al, (Potential role for gut microbiota in cell wall digestion and glucoside detoxification in Tenebrio molitor larvae), Journal of Insect Physiology 52 (2006) 593–601

7: WO2004/111254A1 (Poalis A/S)

8: Gutmann et al (Pure Appl. Chem, 2013-07-09)

9: Baig et al (Angew Chem Int Ed Engl. 2006 Nov 27;45(46):7842-6)

10: Radominska-Pandya A, Bratton SM, Redinbo MR, Miley MJ. Drug Metab Rev. 2010 Feb;42(1):133-44)

11: Plant Physiology, November 2008, Vol. 148, pp. 1295–1308

12: Esben Halkjaer Hansen et al. Phytochemistry 70(4): 473-482

13: Tang, Y., Lee, T. S., and Khosla, C. (2004) Engineered biosynthesis of regioselectively modified aromatic polyketides using bimodular polyketide synthases. PLoS Biol. 2, e31

14. Mizuuchi Y, Shi SP, Wanibuchi K, Kojima A, Morita H, Noguchi H, Abe I. Novel type III polyketide synthases from *Aloe arborescens*. FEBS J. 2009 Apr;276(8):2391-2401.

15. Karppinen K, Hokkanen J, Mattila S, Neubauer P, Hohtola A. Octaketide-producing type III polyketide synthase from *Hypericum perforatum* is expressed in dark glands accumulating hypericins. FEBS J. 2008 Sep;275(17):4329-4342.

16. Dayu Yu, Fuchao Xu, Jia Zeng, Jixun Zhan. Type III Polyketide Synthases in Natural Product Biosynthesis. IUBMB Life, 64(4): 285–295, April 2012

17. Supriya Jadhav, Prasad Phapale, Hirekodathakallu V. Thulasiram, Sujata Bhargava. Polyketide synthesis in tobacco plants transformed with a *Plumbago zeylanica* type III hexaketide synthase. Phytochemistry 98 (2014) 92–100.

CLAIMS

1. A method for producing an octaketide derived aromatic compound of interest, wherein the method comprises the following steps:

(I): contacting *in vivo* in a recombinant host cell comprising a recombinantly introduced Type III polyketide synthase (PKS) gene encoding an octaketide synthase (OKS) and wherein the OKS is of a different genus than the host cell:

(i): suitable starter unit and extender units with the recombinantly introduced OKS capable of converting the starter and extender units into a non-reduced octaketide under suitable conditions wherein there *in vivo* is produced the non-reduced octaketide; and

(II): converting *in vivo* within the recombinant host cell the non-reduced octaketide of step (I) into a C_{14}-C_{34} aromatic aglycon compound of interest, wherein the aromatic aglycon compound of interest is not SEK4 and/or SEK4B; and

(III): optionally, the recombinant host cell also comprises a glycosyltransferase gene encoding a glycosyltransferase and the in step (II) produced aromatic aglycon compound of interest is *in vivo* glycosylated by the glycosyltransferase to produce a C_{14}-C_{34} aromatic glycoside compound of interest; and

(IV): isolating the aromatic aglycon compound of interest of step (II) and/or isolating the aromatic glycoside compound of interest of step (III) to get a composition, wherein the composition comprises less than 1% w/w dry matter of recombinant host cell material.

2. The method of claim 1, wherein the recombinant host cell in claim 1 is a growing recombinant host cell and step (I) and step (II) of claims 1 are:

(I): contacting *in vivo* in a growing recombinant host cell comprising a recombinantly introduced Type III polyketide synthase (PKS) gene encoding an octaketide synthase (OKS) and wherein the OKS is of a different genus than the host cell:

(i): suitable starter unit and extender units with the recombinantly introduced OKS capable of converting the starter and extender units into a non-reduced octaketide under suitable growth conditions wherein there *in vivo* is produced the non-reduced octaketide; and

(II): converting *in vivo* within the growing recombinant host cell the non-reduced octaketide of step (I) into a C_{14}-C_{34} aromatic aglycon compound of interest, wherein the aromatic aglycon compound of interest is not SEK4 and/or SEK4B.

3. The method according to any of the preceding claims, wherein the recombinant host cell is a plant cell and wherein the plant cell is *Nicotiana sp.*

4. The method according to any of claims 1 to 2, wherein the recombinant host cell is a filamentous fungal cell and wherein the filamentous fungal cell is an *Aspergillus* cell.

5. The method according to any of the preceding claims, wherein the Type III polyketide synthase (PKS) gene of step (I) is a PKS gene from a plant and wherein the plant is a plant selected from the group consisting of: *Aloe spp.* and *Hypericum spp.*

6. The method of claim 5, wherein the octaketide synthase (OKS) in step (I) of claim 1 is a OKS comprising an amino acid sequence which has at least 70% (preferably at least 80%, more preferably at least 90% and even more preferably at least 98%) identity with amino acids 1 to 403 of SEQ ID NO:7 (*AaOKS)*.

7. The method of claim 5, wherein the octaketide synthase (OKS) in step (I) of claim 1 is a OKS comprising an amino acid sequence which has at least 70% (preferably at least 80%, more preferably at least 90% and even more preferably at least 98%) identity with amino acids 1 to 393 of SEQ ID NO:15 (*HpPKS2)*.

8. The method according to any of the preceding claims, wherein the suitable starter unit is acetyl-CoA and/or malonyl-CoA.

9. The method according to any of the preceding claims, wherein the:
- recombinant host cell is a host cell selected from the group consisting of: *Aspergillus* (preferably *Aspergillus nidulans*) and *Nicotiana sp.* (preferably *Nicotiana benthamiana*); and
- the OKS of a different genus than the host cell is an OKS selected from the group consisting of: OKS from *Aloe spp.* (preferably *Aloe arborescens*) and *Hypericum spp.* (preferably *Hypericum perforatum*).

10. The method according to any of the preceding claims, wherein the aromatic aglycon compound of interest is a C_{16} aromatic aglycon compound of interest.

11. The method according to claim 28, wherein the C_{16} aromatic aglycon compound of interest is flavokermesic acid (FK) or kermesic acid (KA).

12. The method according to any of the preceding claims, wherein the conversion *in vivo* in step (II) of claim 1 of the non-reduced octaketide into a C_{14}-C_{34} aromatic aglycon compound of interest is done via involvement of at least one *in trans* acting (independent of the PKS enzyme) aromatase/cyclase.

13. The method of claim 12,
wherein the aromatase/cyclase is a aromatase/cyclase comprising an amino acid sequence which has at least 70% (preferably at least 80%, more preferably at least 90% and even more preferably at least 98%) identity with amino acids 1 to 169 of SEQ ID NO:17 (*Streptomyces* ZhuI); and/or
wherein the aromatase/cyclase is a aromatase/cyclase comprising an amino acid sequence which has at least 70% (preferably at least 80%, more preferably at least 90% and even more preferably at least 98%) identity with amino acids 1 to 256 of SEQ ID NO:19 (*Streptomyces* ZhuJ).

14. The method according to any of the preceding claims, wherein step (III) is fulfilled (i.e. not optional) – i.e. wherein step (III) is: the recombinant host cell also comprises a glycosyltransferase gene encoding a glycosyltransferase and the in step (II) produced aromatic aglycon compound of interest is *in vivo* glycosylated by the glycosyltransferase to produce a C_{14}-C_{34} aromatic glycoside compound of interest.

15. The method of claim 14, wherein the aromatic glycoside compound of interest is a C_{16} aromatic glycoside compound of interest and wherein the C_{16} aromatic glycoside compound of interest is a flavokermesic acid glycoside or a kermesic acid glycoside.

16. The method of claim 15, wherein the C_{16} aromatic glycoside compound of interest is a flavokermesic acid glycoside and the flavokermesic acid glycoside is DcII.

17. The method of claim 15, wherein the C_{16} aromatic glycoside compound of interest is a kermesic acid glycoside and the kermesic acid glycoside is carminic acid (CA).

Figure 1

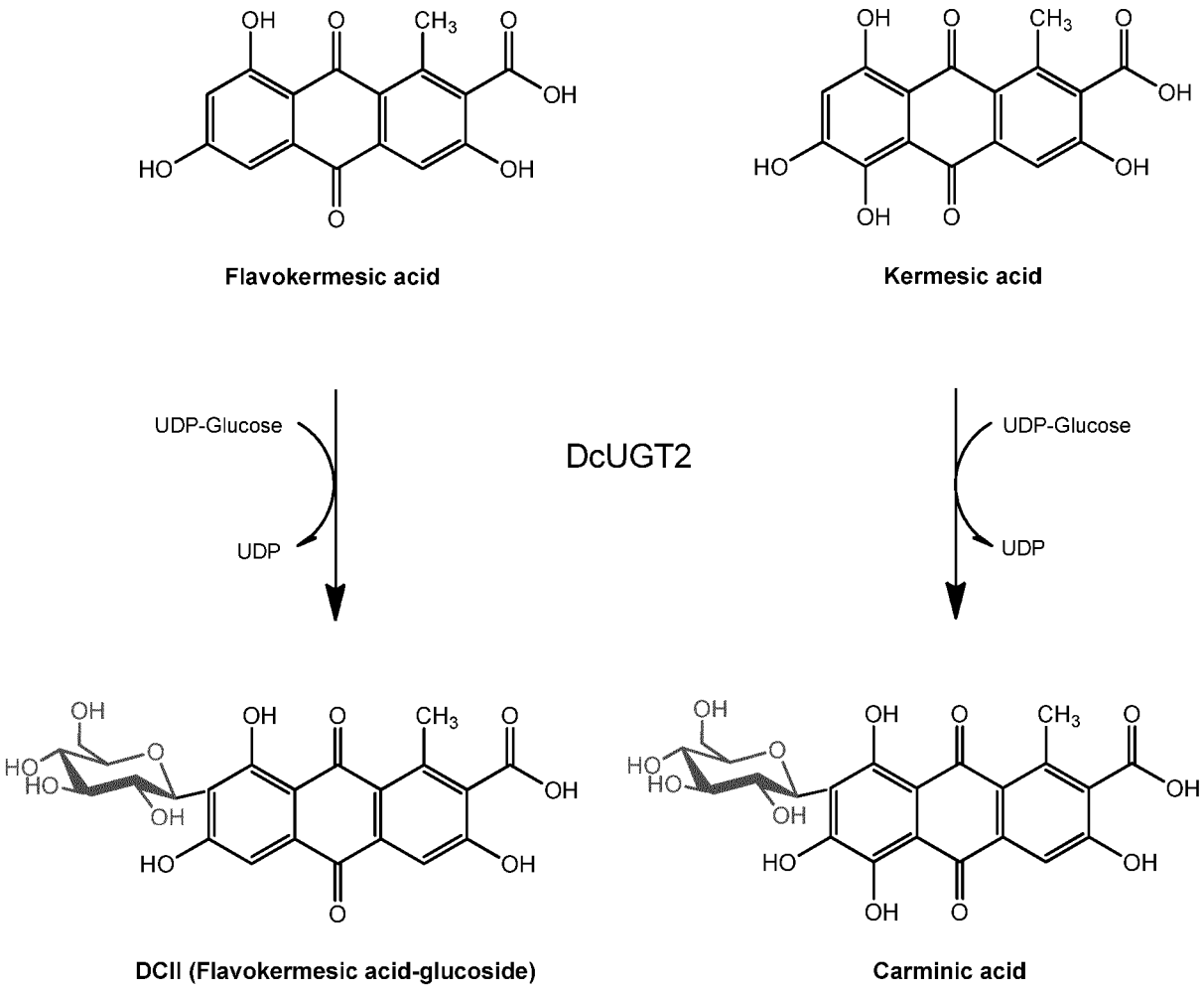

OH
O
CH3
O
OH
HO
OH
O
Flavokermesic acid
OH
O
CH3
O
OH
HO
OH
OH
O
Kermesic acid
UDP-Glucose
UDP
DcUGT2
UDP-Glucose
UDP
OH
HO
HO
O
OH
OH
O
CH3
O
OH
HO
OH
DCII (Flavokermesic acid-glucoside)
OH
HO
HO
O
OH
OH
O
CH3
O
OH
HO
OH
OH
Carminic acid

Figure 2

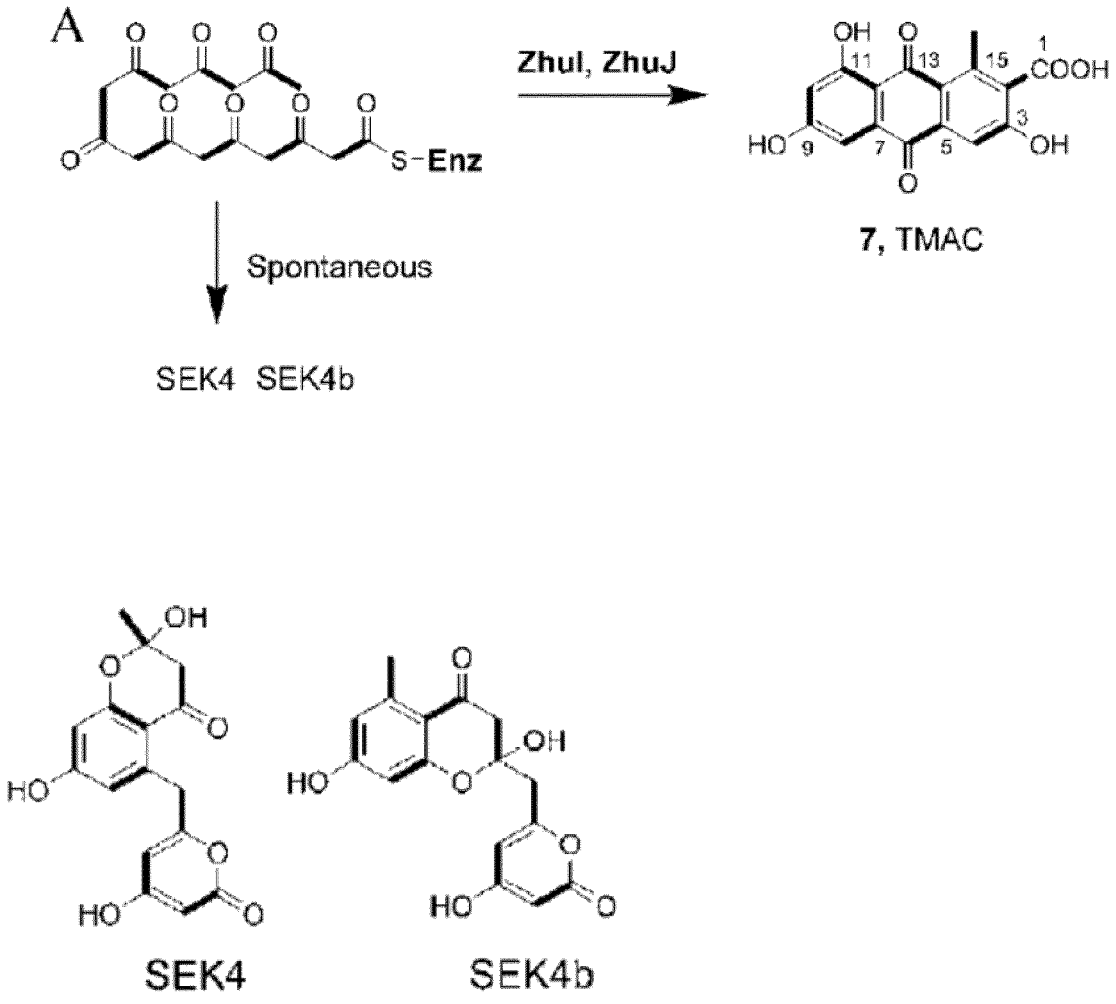


Figure 3

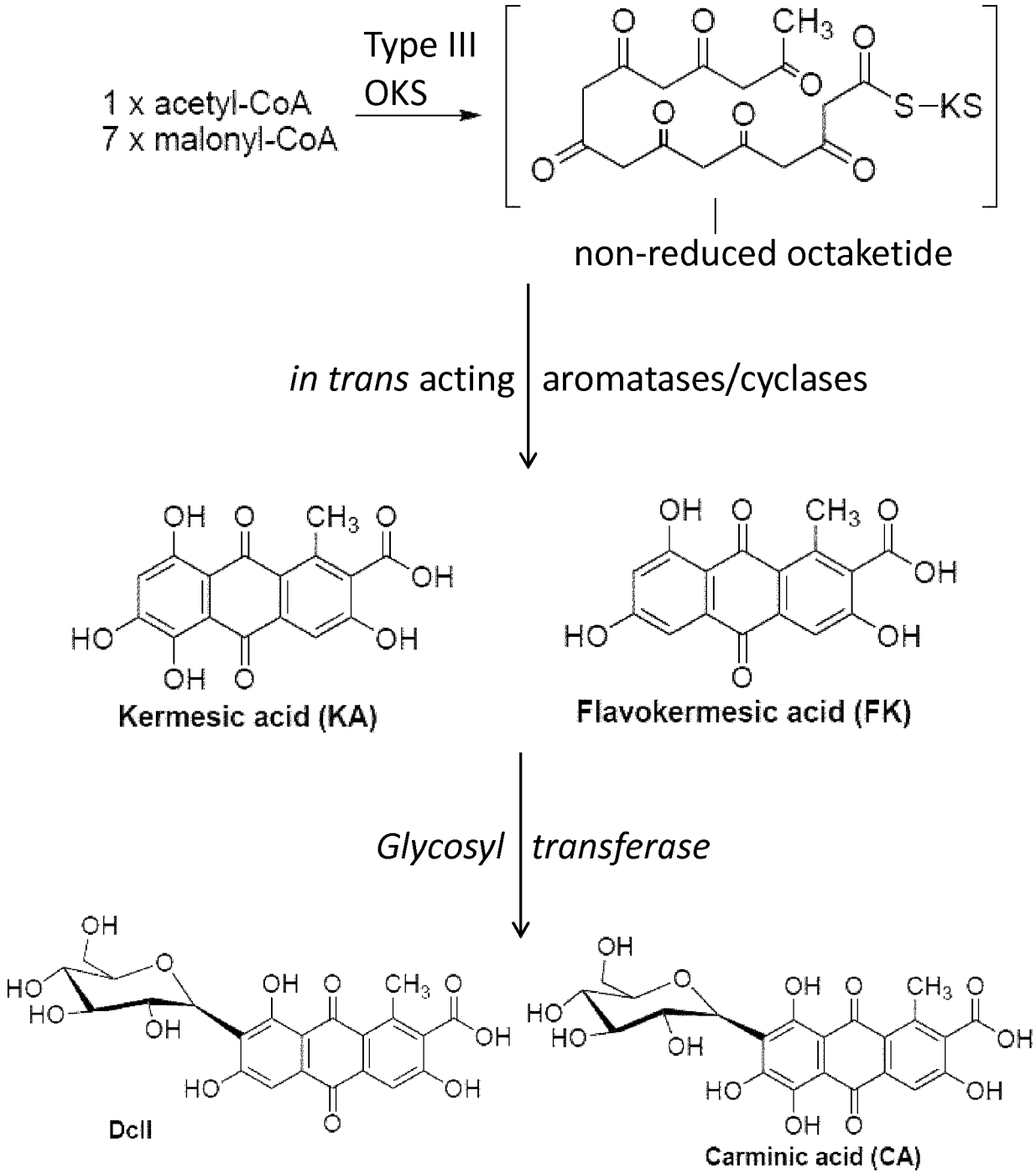
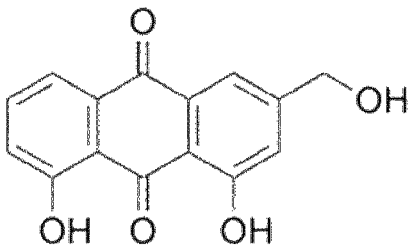
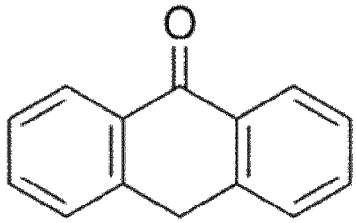

1 x acetyl-CoA
7 x malonyl-CoA
Type III
OKS
S–KS
non-reduced octaketide
in trans acting
aromatases/cyclases
Kermesic acid (KA)
Flavokermesic acid (FK)
Glycosyl
transferase
DcII
Carminic acid (CA)

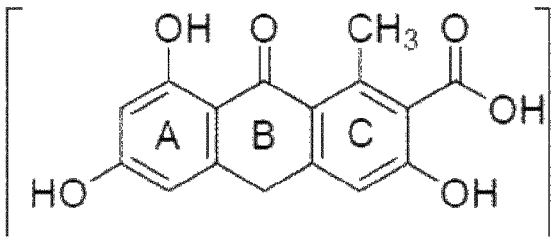
Figure 4



9,10-anthraquinone skeleton



Anthrone



FK antrone

Figure 5 (A)

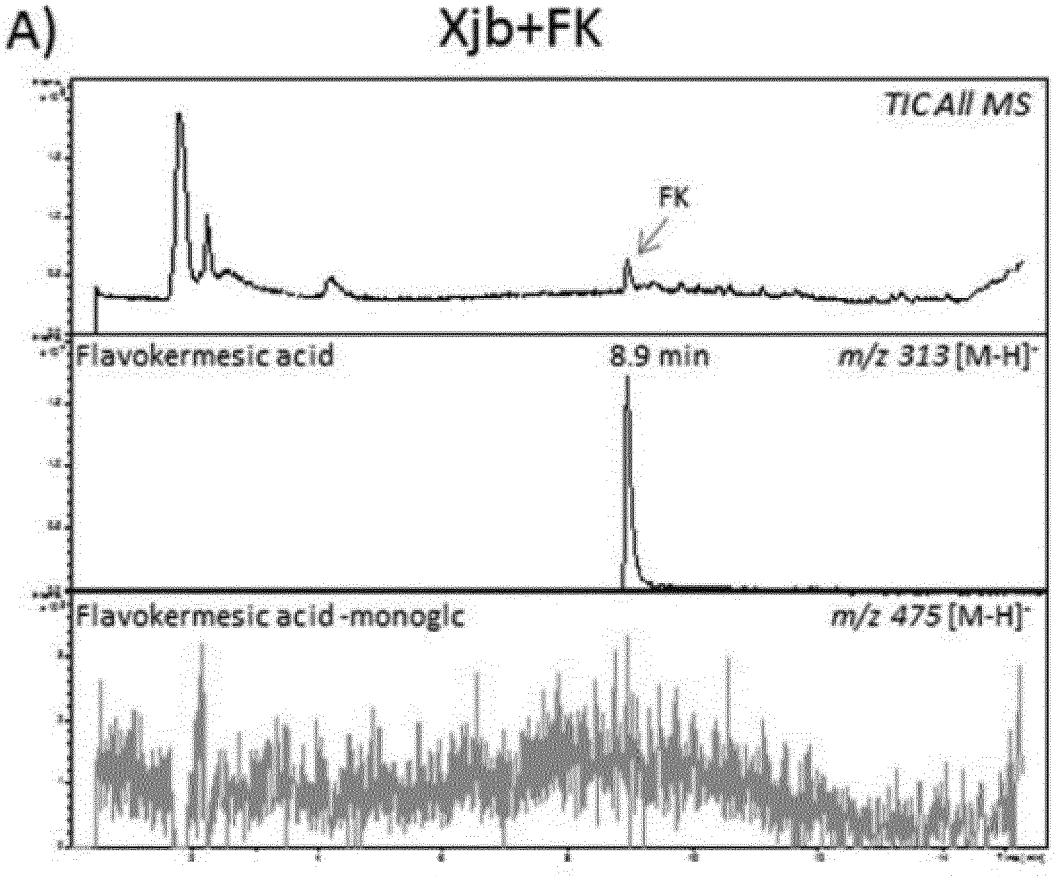

A)
Xjb+FK
TIC All MS
FK
Flavokermesic acid
8.9 min
m/z 313 [M-H]-
Flavokermesic acid -monoglc
m/z 475 [M-H]-

Figure 5 (B)

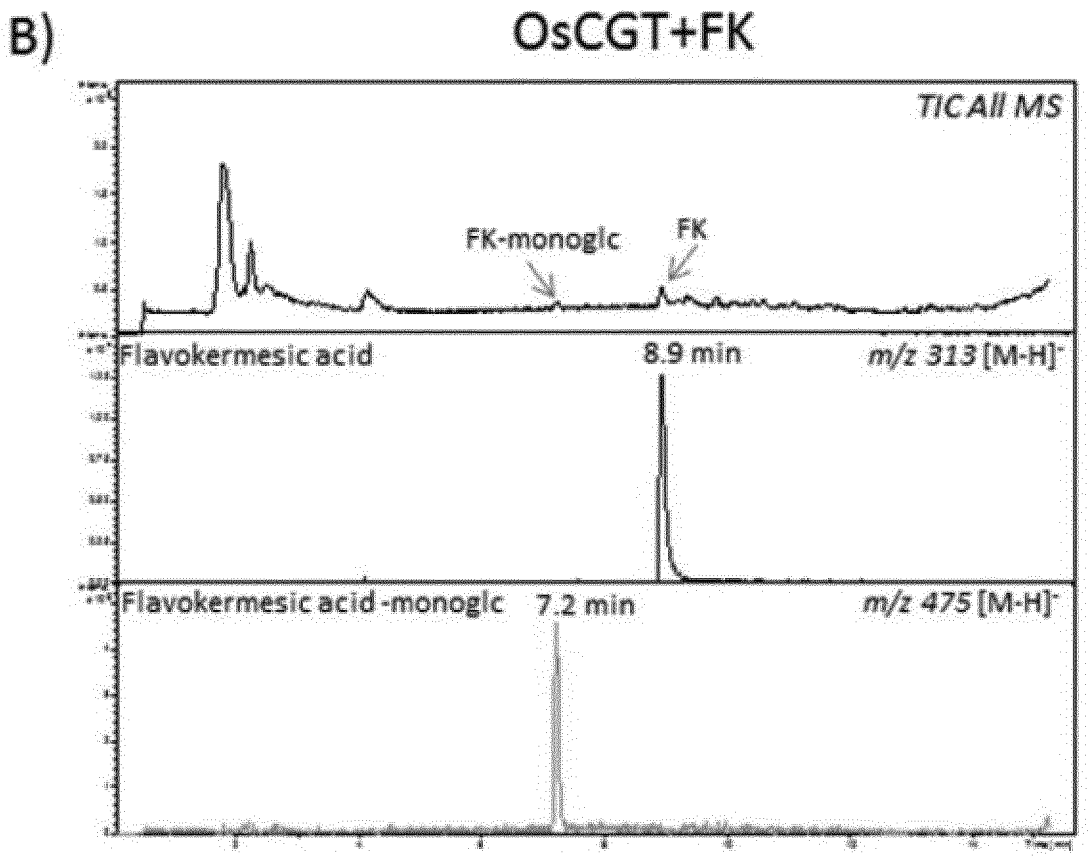

B)
OsCGT+FK
TIC All MS
FK-monoglc
FK
Flavokermesic acid
8.9 min
m/z 313 [M-H]-
Flavokermesic acid -monoglc
7.2 min
m/z 475 [M-H]-

Figure 5 (C)

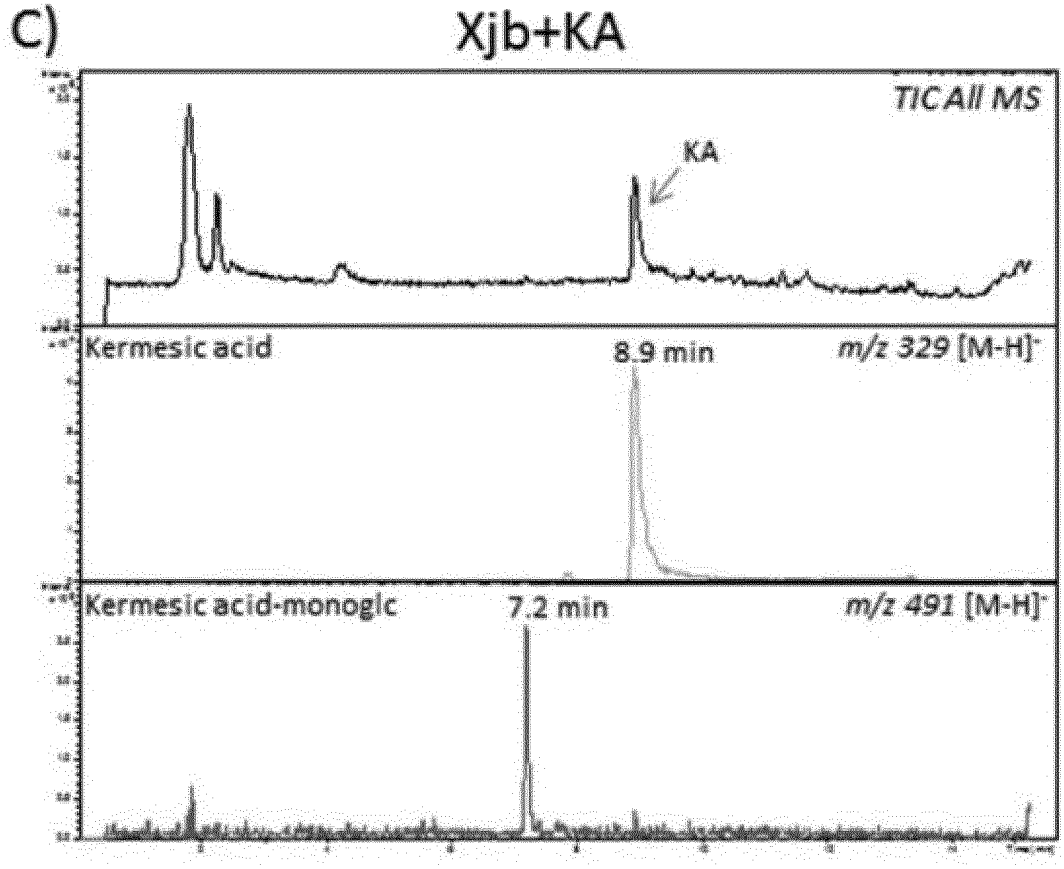

C)
Xjb+KA
TIC All MS
KA
Kermesic acid
8.9 min
m/z 329 [M-H]-
Kermesic acid-monoglc
7.2 min
m/z 491 [M-H]-

Figure 5 (D)

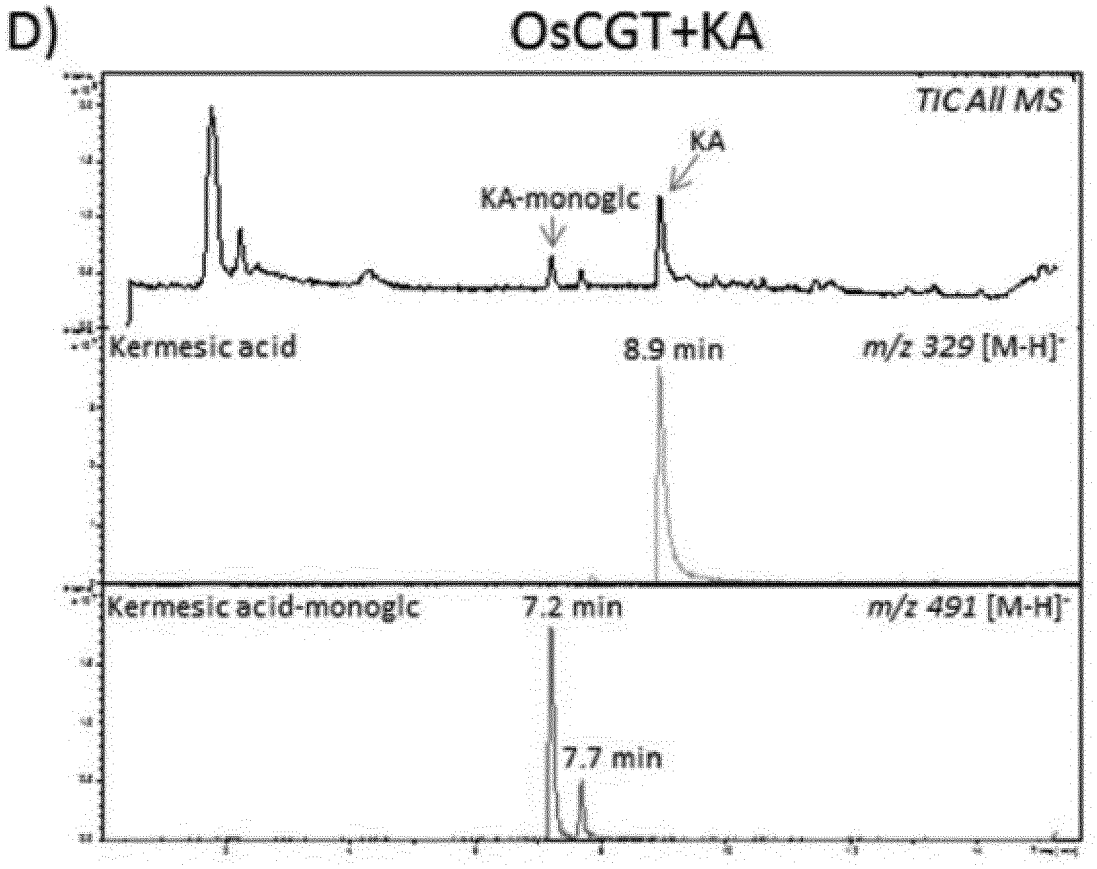

D)
OsCGT+KA
TIC All MS
KA
KA-monoglc
Kermesic acid
8.9 min
m/z 329 [M-H]-
Kermesic acid-monoglc
7.2 min
m/z 491 [M-H]-
7.7 min

Figure 5 (E)

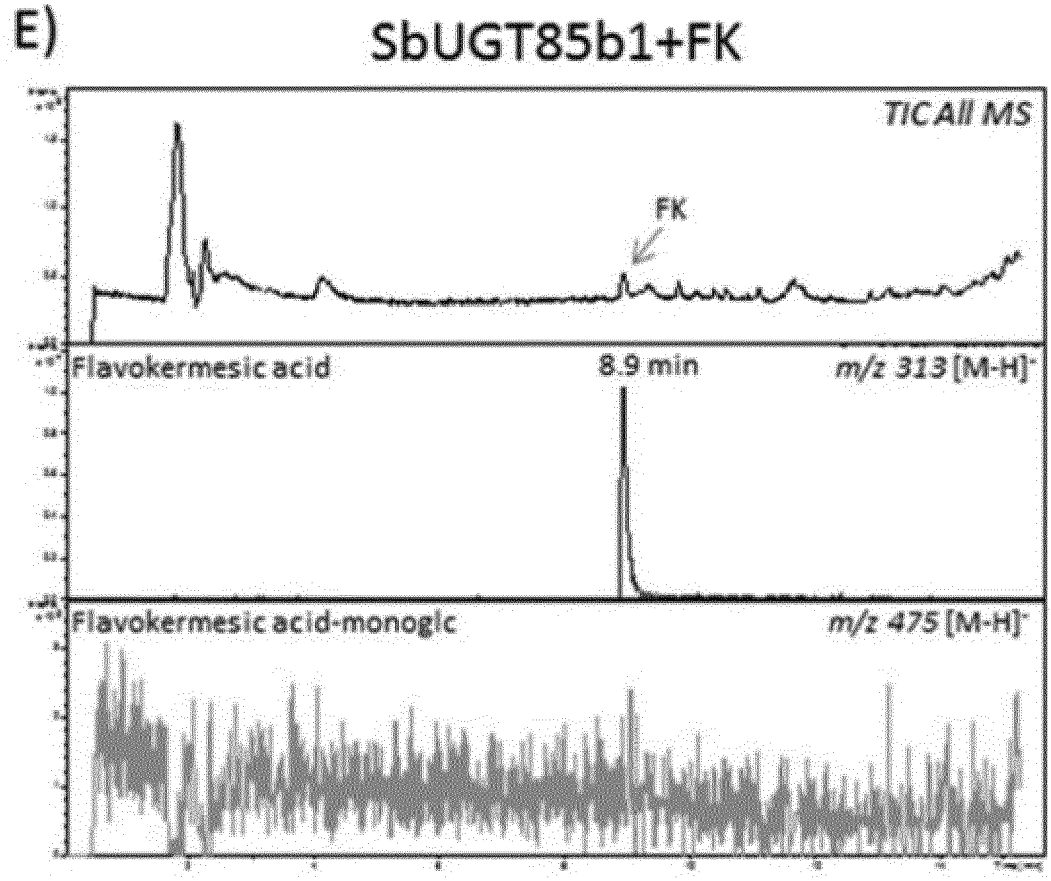

E)
SbUGT85b1+FK
TIC All MS
FK
Flavokermesic acid
8.9 min
m/z 313 [M-H]-
Flavokermesic acid-monogic
m/z 475 [M-H]-

Figure 5 (F)

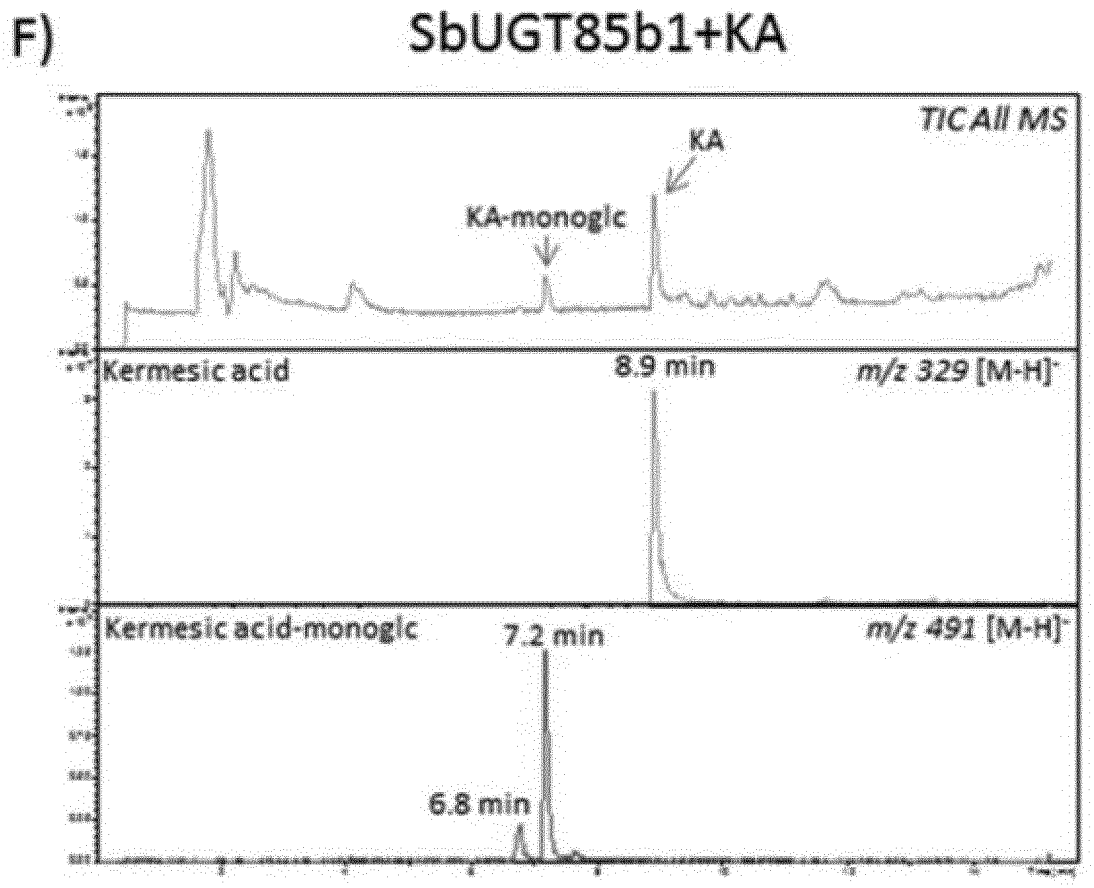

F)
SbUGT85b1+KA
TIC All MS
KA
KA-monoglc
Kermesic acid
8.9 min
m/z 329 [M-H]-
Kermesic acid-monoglc
7.2 min
m/z 491 [M-H]-
6.8 min

Figure 5 (G)

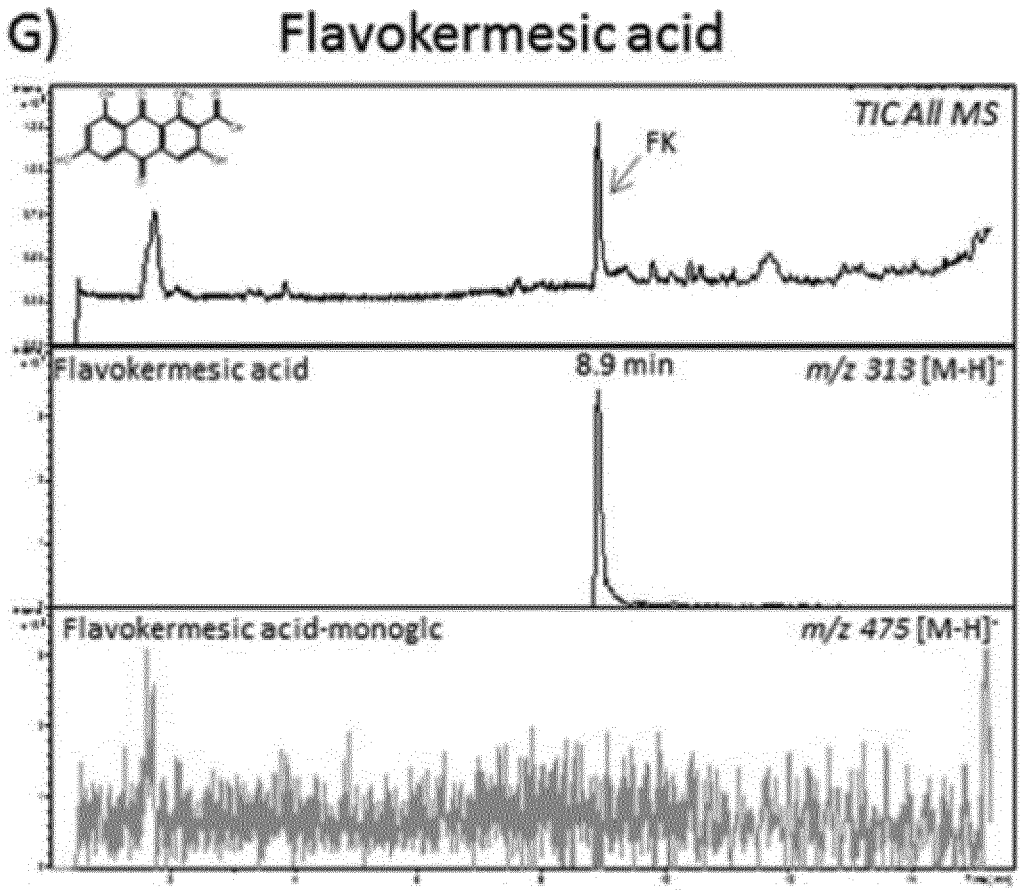
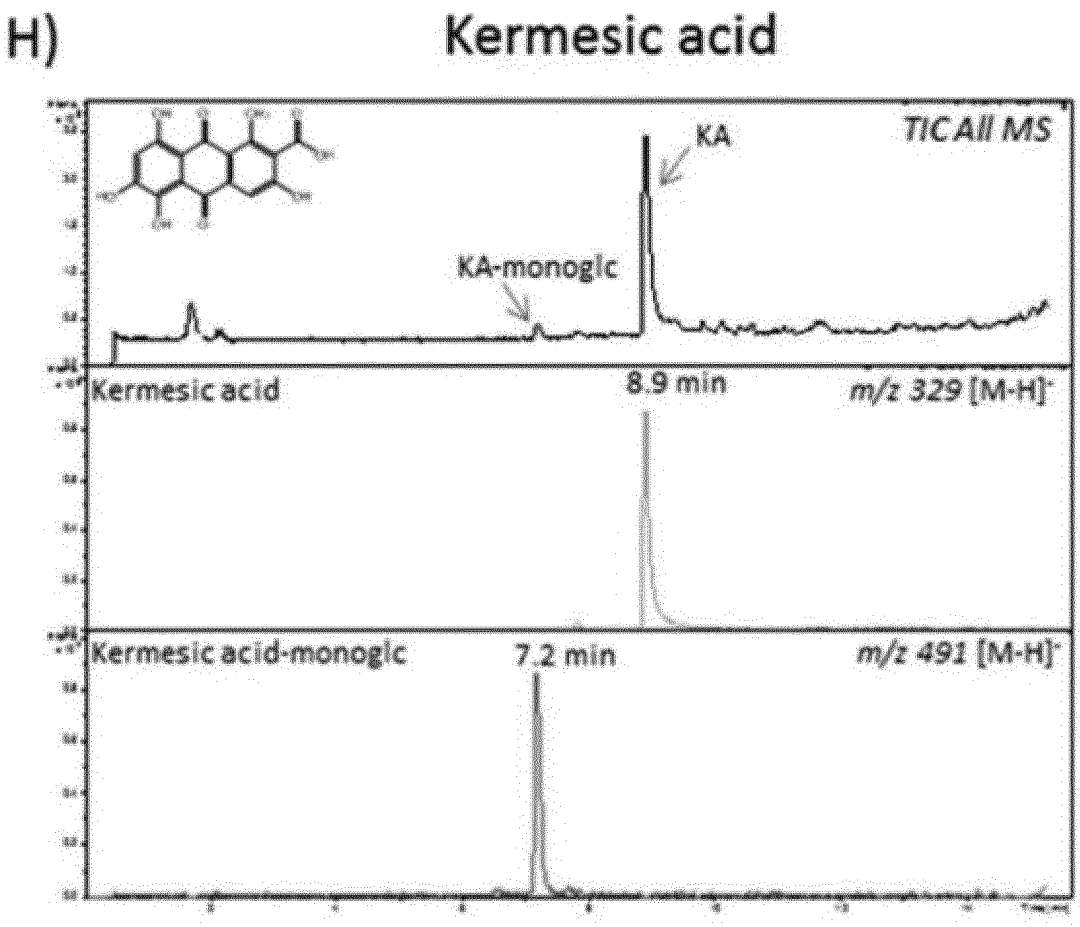

G)
Flavokermesic acid
TIC All MS
FK
Flavokermesic acid
8.9 min
m/z 313 [M-H]-
Flavokermesic acid-monoglc
m/z 475 [M-H]-

Figure 5 (H)


H)
Kermesic acid
KA
TIC All MS
KA-monoglc
Kermesic acid
8.9 min
m/z 329 [M-H]-
Kermesic acid-monoglc
7.2 min
m/z 491 [M-H]-

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2016/063242

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12P7/26 C12P7/66 C12P17/06
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, BIOSIS, COMPENDEX, EMBASE, FSTA, IBM-TDB, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	TANG, Y. ET AL.: "Engineered Biosynthesis of Regioselectively Modified Aromatic Polyketides Using Bimodular Polyketide Synthases", PLOS BIOLOGY, vol. 2, no. 2, February 2004 (2004-02), pages 0227-0238, XP002752479, cited in the application page 0230, column 2, line 30 - page 0232, column 1, line 8; figure 5A; table 1; compound 7 page 0236, column 2, line 16 - page 0237, column 1, line 16 ----- -/--	1-17

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
29 July 2016	09/08/2016

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Fuchs, Ulrike

1

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2016/063242

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. [X] forming part of the international application as filed:

      [X] in the form of an Annex C/ST.25 text file.

      [ ] on paper or in the form of an image file.

   b. [ ] furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. [ ] furnished subsequent to the international filing date for the purposes of international search only:

      [ ] in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      [ ] on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. [ ] In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	ABE, I. ET AL.: "Engineered Biosynthesis of Plant Polyketides: Chain Length Control in an Octaketide-Producing Plant Type III Polyketide Synthase", JOURNAL OF THE AMERICAN CHEMICAL SOCIETY, vol. 127, no. 36, 17 August 2005 (2005-08-17), pages 12709-12716, XP002752480, abstract page 12710, column 1, line 18 - column 2, line 4; figure 1D page 12711; figure 2B page 12713, column 1, lines 6-32 page 12715; figure 6 -----	1-17
A	KARPPINEN, K. ET AL.: "Octaketide-producing type III polyketide synthase from Hypericum perforatum is expressed in dark glands accumulating hypericins", FEBS JOURNAL, vol. 275, no. 17, 21 July 2008 (2008-07-21), pages 4329-4342, XP002752481, abstract page 4330; figure 1 page 4332, column 1, line 22 - column 2, line 9; compounds A4, A7 page 4333; figure 4 page 4336, column 2, lines 1-26 -----	1

1