



## Use of heterologous expressed polyketide synthase and small molecule foldases to make aromatic and cyclic compounds

Frandsen, Rasmus John Normand; Mortensen, Uffe Hasbro; Coumou, Hilde; Kannangara, Rubini; Madsen, Lars; Nafisi, Majse; Ranberg, Johan Andersen; Kongstad, Kenneth Thermann; T. Okkels, Finn; Khorsand-Jamal, Paiman

*Total number of authors:*  
11

*Publication date:*  
2016

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

[Link back to DTU Orbit](#)

### *Citation (APA):*

Frandsen, R. J. N., Mortensen, U. H., Coumou, H., Kannangara, R., Madsen, L., Nafisi, M., Ranberg, J. A., Kongstad, K. T., T. Okkels, F., Khorsand-Jamal, P., & Stærk, D. (2016). Use of heterologous expressed polyketide synthase and small molecule foldases to make aromatic and cyclic compounds. (Patent No. WO2016198623).

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



## (51) International Patent Classification:

C12P 7/26 (2006.01) C12P 19/60 (2006.01)  
 C12P 7/40 (2006.01) C12N 1/15 (2006.01)  
 C12P 15/00 (2006.01) C12N 1/19 (2006.01)  
 C12P 17/06 (2006.01) C12N 15/82 (2006.01)

## (21) International Application Number:

PCT/EP2016/063331

## (22) International Filing Date:

10 June 2016 (10.06.2016)

## (25) Filing Language:

English

## (26) Publication Language:

English

## (30) Priority Data:

15171430.0 10 June 2015 (10.06.2015) EP

(71) Applicants: DANMARKS TEKNISKE UNIVERSITET [DK/DK]; Anker Englund's Vej 101 A, 2800 Kgs. Lyngby (DK). KØBENHAVNS UNIVERSITET [DK/DK]; Nørregade 10, 1165 Copenhagen K (DK).

(72) Inventors: FRANDSEN, Rasmus John Normand; En-evangen 55, 3450 Allerød (DK). MORTENSEN, Uffe Hasbro; Schleppegrellsgade 6, 4th, 2200 Copenhagen N (DK). COUMOU, Hilde Cornelijne; Masnedøgade 18, 4. th., 2100 Copenhagen Ø (DK). KANNANGARA, Rubini Maya; Dronning Olgas Vej 34, 5. tv., 2000 Frederiksberg (DK). MADSEN, Bjørn; Tordenskjoldsvej 9, 3000

Helsingør (DK). NAFISI, Majse; Byhøjten 28, 1. tv., 2720 Vanløse (DK). ANDERSEN-RANBERG, Johan; Hothers Plads 13, 3. th, 2200 Copenhagen N (DK). KONGSTAD, Kenneth Thermann; Ingerslevsgade 170, 3. sal, 1705 Copenhagen V (DK). OKKELS, Finn Thyge; Kongemarken 11, 4000 Roskilde (DK). KHORSAND-JAMAL, Paiman; Gedvad 3, 1. tv., 2800 Kgs. Lyngby (DK). STÆRK, Dan; Perlestensvej 7, 3540 Lyngø (DK).

(74) Agent: GUARDIAN IP CONSULTING I/S; Diplomvej, Building 381, 2800 Kgs. Lyngby (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, 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,

[Continued on next page]

(54) Title: USE OF HETEROLOGOUS EXPRESSED POLYKETIDE SYNTHASE AND SMALL MOLECULE FOLDASES TO MAKE AROMATIC AND CYCLIC COMPOUNDS

	PKS1	PKS2	PKS3	PKS4	...	PKS <sub>n</sub>
No cyclase						
Cyclase 1 (C1-C6)						
Cyclase 2 (C2-C7)						
Cyclase 3 (C3-C8)						
Cyclase 4 (C4-C9)						
Cyclase 5 (C5-C10)						
...						
Cyclase <sub>n</sub>						

Figure 1

(57) Abstract: A method for producing individual or libraries of tri- to pentadecaketide-derived aromatic compounds of interest by heterologous expression of polyketide synthase and aromatase/cyclase in a recombinant host cell.

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))*

## USE OF HETEROLOGOUS EXPRESSED POLYKETIDE SYNTHASE AND SMALL MOLECULE FOLDASES TO MAKE AROMATIC AND CYCLIC COMPOUNDS

### 5 FIELD OF THE INVENTION

The present invention relates to a method for producing individual and libraries of tri- to pentadecaketide derived aromatic and cyclic compounds of interest by heterologous expression of a polyketide synthase and one or more aromatases/cyclases in a recombinant  
10 host cell.

### BACKGROUND OF THE INVENTION

Small molecules, of biological origin, often include aromatic or cyclic groups that impact  
15 their physiochemical and biological properties. Although nature is rich in aromatic compounds with different carbon skeletons, there is an urgent need for biosynthetic systems capable of producing both natural and new-to-nature aromatic compounds. Areas of specific interest are the formation of carbon skeletons that can be used medicinally (e.g. new antibiotics), or as chemical substitutes, or as food ingredients, or as precursors for the  
20 formation of more complex compounds. Among the top 100 drugs developed, 60% are small molecules (excluding proteins), and of these 82% possess aromatic motifs. Complex aromatic compounds are produced via many different biosynthetic pathways in nature, either as part of primary or secondary metabolism. One of the most versatile biosynthetic schemes for producing aromatic compounds is via the non-reducing polyketide pathways,  
25 wherein two-carbon units (-CH<sub>2</sub>-CO-), referred to as ketides or 'ketide units', are polymerized into linear chains called polyketides, which subsequently can fold into aromatic structures. The formation of polyketides is dependent on an enzymes class known as polyketide synthases (PKSs).

30 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 coupled 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  
35 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 (CL-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, e.g. but not exclusively,  
40 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. If the substrate is the standard starter (acetyl-CoA) and extender (malonyl-CoA) units, then the number of carbon atoms in the resulting polyketide chain will equal two times the number of iterations/'condensation reactions', performed by the PKS enzyme. Thus, a heptaketide  
5 synthase will perform six condensation reactions joining one starter unit (two carbons) with six extender units (six times two carbons), resulting in a polyketide consisting of seven ketide units, made up of a total of fourteen carbon atoms. However, PKSs may use alternative starter and extender units which can alter the number of carbon atoms in the final product, for example a heptaketide synthase could use p-coumaric acid (nine carbons)  
10 as a starter unit and six methyl-malonyl-CoA (six times three carbons) as extender units resulting in a heptaketide with twenty-seven carbon atoms. Each individual PKS, e.g. a heptaketide synthase, displays a different affinity for different starter and extender units, and can hence produce very different compounds which all will be categorized as heptaketides. The substrate availability in the host cell can also affect which product a given  
15 PKS produces as its preferred substrate may only be available in very limited amounts, or not at all, compared to less preferred substrates which then will outcompete the preferred substrate.

The chain length of the polyketide product is thus the result of the number of condensation  
20 reactions the PKS performs, which covalently joins one starter unit with one or more extender units together in a head-to-tail manner. A PKS that performs one iteration/condensation will produce a diketide, one that performs two iterations/condensations will produce a triketide, one that performs three iterations/condensations will produce a tetraketide, and so forth. The number of carbon  
25 atoms in the resulting polyketides will in addition be the result of which starter and extender units the enzyme utilize.

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  
30 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.  
35 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 are further subdivided into modular PKSs and iterative PKSs, where iterative PKSs only possess a single copy of each active site type and re-use these repeatedly until the growing  
40 polyketide chain has reached its predetermined length. Type I iterative PKSs that forms

aromatic and polyaromatic compounds typically rely on endogenous PT and CL-CYC domains to direct folding of the formed non-reduced polyketide chain. Dissected PT domains have been shown to work *in trans* with heterologous KS-AT-ACP fragments from the type I iterative PKSs to form folded polyketide products. The PT domains typically promote the formation of several intramolecular bonds. 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 I modular PKSs that form macrolide (macrocyclic) compounds includes a terminal CL-CYC domain.

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 transiently to form the functional PKS enzyme. The involved enzymes include activities for KS, CLF and ACP. Type II PKSs forms linear non-reduced polyketides that spontaneously folds into aromatic/cyclic compounds via the formation of intra-molecular carbon-carbon and carbon-oxygen bonds.

Types I modular (Im), type I iterative (Ii) and type II (II) are all dependent on an ACP domain(s) which is responsible for tethering the growing polyketide (acyl) chain to the enzyme during synthesis. In the ACP-dependent PKS types, the acyl group is transferred from the incoming Co-enzyme A (CoA) to the ACP domain and is subsequently condensed with another acyl group bound to the KS domain of the enzyme, resulting in a diketide bound to the ACP domain. The formed diketide is subsequently moved back to the KS domain and another ACP bound extender unit, is loaded into the enzyme.

Type III PKSs generally only consist of a KS domain, referred to as a KASIII or Chalcone synthase domain and they lack an ACP domain. Type III PKSs are self-contained enzymes that form homodimers. Their single active site in each monomer catalyzes the priming and extension reactions iteratively to form polyketide products. Type III PKS from bacteria, plant and fungi have been described. Type III PKSs (also known as Chalcone synthase) 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. Formation of flavonoids and stilbenes depends on one p-coumaroyl CoA starter unit and three malonyl-CoA extender units. The products of type III PKSs often spontaneously fold into complex aromatic/cyclic compounds, e.g. flavonoids in plants. Type III PKSs that use acetyl/malonyl-CoA as starter unit and malonyl-CoA as extender units resulting in linear non-reduced polyketides have also been described in plants.

Type III enzymes do not have an 'acyl carrier protein' (ACP) functionality, but instead they rely on Co-enzyme A linking for associating the growing polyketide chain with the enzyme during the multiple catalytic cycles. In type III PKSs, the incoming acyl group remains bound to the Co-enzyme A unit, and the condensation between the two acyl groups results in a diketide bound to the incoming Co-enzyme A. The formed diketide is subsequently moved back to the KS domain and another Co-enzyme A bound extender unit, is loaded into the enzyme.

The above described unique functional and corresponding structural properties of the Type I, Type II or Type III PKS allow members of these three enzyme groups to be distinguished.

The subsequent folding and release of the polyketide chain produced by the different classes of PKS enzymes is either spontaneous, or may be catalyzed by several different enzyme families typically referred to as aromatases and/or cyclases, or by domain(s) within the PKS, such as a PT and/or CL-CYC domains. Herein these are collectively referred to as 'small molecule foldases'. This group of enzymes is characterized by catalyzing the regiospecific formation of intra-molecular carbon-carbon or carbon-oxygen bonds within a polyketide, resulting in the formation of aromatic or cyclic motifs. 'Small molecule foldases', acting on polyketides, are found in bacteria, fungi and plants. Several examples exist where folding of the polyketide is a spontaneous process, e.g. flavonoids in plants. Though 'small molecule foldases' perform similar functions in polyketide biosynthetic pathways they are very different at the primary sequence level, and can hence be categorized based on which structural and primary sequence motifs they contain. The group of 'small molecule foldases' that act on polyketides include enzymes from the 'Cyclase', 'SRPBCC Cyclases/aromatase', 'DABB Cyclase/aromatase', 'Polyketide synthesis cyclase', 'Lactamase\_B/MBL fold metallohydrolase', ketoreductase from Act cluster and 'Cupin\_2' Superfamilies and, in addition, includes dissected PT and CL-CYC domains from type I iterative PKS from filamentous fungi.

Importantly, the Type I, Type II or Type III PKSs are further distinguished by the timing and mechanism by which the formed polyketide chain are folded into complex structures with cyclic and aromatic motifs. In Type I modular PKS, containing a CL-CYC domain, the polyketide chain remains attached to the enzyme's ACP domain, and the CL-CYC domain is both responsible for folding of the chain into a macrolide and its simultaneously release from the ACP domain and thereby also the enzyme. Type I iterative PKSs contain a PT domain and/or CL-CYC domain, that catalyse the cyclization reactions and formation of aromatic groups in the polyketide chain. The PT domain acts on the polyketide that is bound to enzyme's ACP domain, where the ACP domain influences the docking and positioning of the polyketide substrate into the active site of the PT domain and thereby the chains folding pattern. The CL-CYC domains forms cyclic structures and simultaneously releases the ACP bound product from the enzyme.

In the case of type II PKSs, polyketide folding is a post-PKS enzyme guided and catalyzed process. In this case, the KS/CLF/ACP enzyme complex forms a polyketide chain of a predetermined length, which remains bound to the ACP enzyme while it is folded by  
5 aromatase(s) and cyclase(s).

In the case of type III PKSs, the formed linear polyketide chain is released, likely following hydrolysis of the linkage to Co-enzyme A, whereafter the chain undergoes spontaneous folding into a range of sterically stable folds.  
10

### SUMMARY OF THE INVENTION

The problem solved by the present invention relates to the provision of a suitable  
15 biosynthetic pathway that forms aromatic and cyclic compounds (e.g. C<sub>6</sub>-C<sub>31</sub> poly aromatic compounds) and/or libraries of aromatic compounds of interest *in vivo*.

The present invention is based on experimental results disclosed herein, which demonstrate that *in vivo* heterologous co-expression of a Type III polyketide synthase (PKSIII) from  
20 plants/bacteria/fungi and one or more 'small molecule foldases' from fungi/bacteria, wherein the aromatase/cyclase is from a different genus than the PKSIII, in a recombinant host cell (e.g. a yeast cell or bacterial cell), provides a suitable biosynthetic pathway for the production of aromatic compounds. The *in vivo* heterologously-expressed PKSIII produces a non-reduced polyketide which is converted *in vivo* into cyclic or/and aromatic compounds of  
25 interest by the action of the one or more heterologously-expressed 'small molecule foldases'.

Recombinant host cells expressing the PKSIII and one or more 'small molecule foldases' collectively form a programmable system for the formation of aromatic compounds, of any  
30 desirable length and fold. The natural systems do not offer such flexibility and predictability and the present invention therefore represent a major technological advance compared to existing technologies available for the creation of biosynthetic pathways that are not found in nature. The recombinant host cells may be used in a method to produce specific aromatic and cyclic compounds (e.g. C<sub>6</sub>-C<sub>31</sub> poly aromatic compounds) and/or libraries of aromatic  
35 compounds of interest *in vivo*.

Accordingly, a first aspect of the present invention relates to a method of producing a library of polyketide-derived aromatic and/or polyaromatic; cyclic and/or polycyclic compounds; or any combination thereof, wherein the carbon atom chain length of the  
40 polyketide backbone of the compounds is selected from two or more of 6, 7, 8, 9, 10, 11,



12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 and 31 carbon atoms, comprising the steps of:

- a. providing one or more heterogeneous populations of recombinant cells, wherein each cell in the one or more populations comprises:
  - 5 i. a transgene encoding a heterologous type III polyketide synthase capable of forming a linear non-reduced polyketide compound, wherein the carbon atom chain length of the polyketide backbone of the formed compound is selected from 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 and 31 carbon atoms; and
  - 10 ii. a transgene encoding a first heterologous 'small molecule foldase' enzyme capable of catalyzing the formation of one or more region-specific intramolecular carbon-carbon or carbon-oxygen bonds in a linear non-reduced polyketide compound, wherein the carbon atom chain length of the polyketide backbone of the compound is one or more of 6, 7, 8, 9, 10, 11, 12, 13, 14, 15,  
15 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 and 31 carbon atoms, and
  - 20 iii. optionally one or more transgene(s) encoding a second, third and fourth heterologous 'small molecule foldase' enzyme capable of catalyzing the formation of one or more region-specific intramolecular carbon-carbon or carbon-oxygen bonds in a non-linear polyketide,  
25 wherein each of the first, second, third and fourth heterologous 'small molecule foldase' enzyme is a bacterial or fungal enzyme, and wherein the genus from which said bacterial or fungal enzyme is derived is different from the genus from which the PKSIII enzyme is derived,  
30 wherein the one or more populations of recombinant cells comprises cells capable of producing polyketide-derived aromatic and/or polyaromatic; cyclic and/or polycyclic compounds; or any combination thereof, wherein the carbon atom chain length of the polyketide backbone of the compounds is selected from two or more of 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 and 31 carbon atoms; and
- b. incubating and/or culturing the one or more heterogeneous populations of recombinant cells in a culture medium to support synthesis of the library of  
35 compounds.

A second aspect of the present invention relates to a heterogeneous population of

recombinant cells capable of producing a library of polyketide-derived aromatic and/or polyaromatic; cyclic and/or polycyclic compounds; or any combination thereof, according to the method of the invention, wherein each cell in the population comprises:

- 5 a. a transgene encoding a heterologous type III PKS capable of forming a polyketide-derived aromatic, polyaromatic, cyclic or polycyclic compound, wherein the carbon atom chain length of the polyketide backbone of the formed compound is selected from among 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 and 31 carbon atoms; and
- 10 b. a transgene encoding a heterologous 'small molecule foldase' enzyme capable of catalyzing the formation of one or more specific intramolecular carbon-carbon bonds in a polyketide-derived aromatic, polyaromatic, cyclic and polycyclic compound, wherein the carbon atom chain length of the polyketide backbone of the compound is one or more of 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 and 31 carbon atoms, and
- 15 c. optionally one more transgene(s) encoding a second, third and fourth heterologous 'small molecule foldase' enzyme capable of catalyzing the formation of one or more region-specific intramolecular carbon-carbon or carbon-oxygen bonds in a non-linear polyketide,
- 20 wherein each of the first, second, third and fourth heterologous 'small molecule foldase' enzyme is a bacterial or fungal enzyme, and wherein the genus from which said bacterial or fungal enzyme is derived is different from the genus from which the PKSIII enzyme is derived,
- 25 wherein the population of recombinant cells comprises cells capable of producing polyketide-derived aromatic and/or polyaromatic; cyclic and/or polycyclic compounds; or any combination thereof, wherein the carbon atom chain length of the polyketide backbone of the compounds is selected from two or more of 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 and 31 carbon atoms.
- 30 It is envisaged that individual heterologous host cells capable of producing an aromatic compound of interest may be identified as a result of the screening of the library of aromatic compounds produced by the one or more populations of heterologous host cells of the invention. This, or any individual heterologous host cell (or its clonal derivatives) of the invention may be used for the production of an aromatic compound.
- 35 Accordingly, a second aspect of the present invention relates to a method of producing a polyketide-derived aromatic, polyaromatic, cyclic or polycyclic compound, wherein the carbon atom chain length of the polyketide backbone of the compound is selected from 6, 7,

8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 and 31 carbon atoms, comprising the steps of:

a. providing a recombinant cell comprising:

- 5 i. a transgene encoding a heterologous type III polyketide synthase capable of forming a linear non-reduced polyketide compound wherein the carbon atom chain length of the polyketide backbone of the formed compound is selected from 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 and 31 carbon atoms; and
- 10 ii. a transgene encoding a first heterologous 'small molecule foldase' enzyme capable of catalyzing the formation of one or more region-specific intramolecular carbon-carbon or carbon-oxygen bonds in a linear non-reduced polyketide compound, wherein the carbon atom chain length of the polyketide backbone of the compound is one or more of 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 and 31 carbon atoms, and
- 15 iii. optionally one more transgene(s) encoding a second, third and fourth heterologous 'small molecule foldase' enzyme capable of catalyzing the formation of one or more region-specific intramolecular carbon-carbon or carbon-oxygen bonds in a non-linear polyketide compound,
- 20 wherein each of the first, second, third and fourth heterologous 'small molecule foldase' enzyme is a bacterial or fungal enzyme, and wherein the genus from which said bacterial or fungal enzyme is derived is different from the genus from which the PKSIII enzyme is derived,
- 25 wherein the recombinant cell is capable of producing polyketide-derived aromatic, polyaromatic, cyclic or polycyclic compound, wherein the carbon atom chain length of the polyketide backbone of the compound is selected from 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 and 31 carbon atoms; and
- 30 b. incubating and/or culturing the recombinant cell in a culture medium to support synthesis of the polyketide-derived aromatic, polyaromatic, cyclic or polycyclic compound.

## 35 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.

5 The term "extender units" relates to the substrates that the PKS III adds to the starter unit and the growing polyketide chain. The extender units are delivered as acyl groups bound to Co-enzyme A, such as, but not exclusively, malonyl-CoA, methylmalonyl-CoA, hydroxyl malonyl-CoA or ethyl-malonyl.

10 The term "heterologous host" is here defined as the situation where a gene is expressed in a recombinant host cell that is taxonomically classified as belonging to a different genus than the organism where the gene of interest was obtained from.

15 The term "heterologous" with respect to an enzyme encoded by a transgene that is expressed in a recombinant cell of the invention, means that the enzyme is expressed in a cell that does not normally express that enzyme; since the gene encoding the enzyme is derived from (and naturally found in) a cell from a different genetic origin (e.g. species) than the cell in which it is expressed.

20 The term "the genus" describes the taxonomic classification of the organism from which a bacterial or fungal 'small molecule foldase' enzyme is derived, which is different from the genus from which the PKSIII enzyme is derived, which means that the 'small molecule foldase' enzyme and the PKSIII enzyme are derived from organisms that are classified to different genera.

25 The term "hybridizes" in relation to a polynucleotide which hybridizes under at least medium stringency conditions with (i) a nucleic acid molecule or (ii) a complementary strand of (i), relates to the nucleotide sequence hybridizing to a labeled nucleic acid probe corresponding to a nucleotide sequence disclosed herein, or its complementary strand under medium to very high stringency conditions. Molecules to which the nucleic acid probe  
30 hybridizes under these conditions can be detected using e.g. X-ray film. Herein relevant hybridization stringency conditions are defined in 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  
35 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  
40 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).

5

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 living cells or a whole living organism as opposed to a partial or dead cell or organism, or an *in vitro* ("within the glass", e.g., in a test tube or petri dish) controlled environment.

The term "ketide" refers to a single acyl unit added during a single condensation reaction step catalyzed by a PKS. If malonyl-CoA or methyl-malonyl are used as an extender units, then the ketide unit will be  $-\text{CH}_2\text{-CO-}$ , and  $-\text{C}(\text{CH}_3)\text{H-CO-}$ , respectively.

20

The term "non-reduced polyketide" denotes a non-reduced polyketide, characterized by the presence of the original ketone groups in the ketides (eg.  $-\text{CH}_2\text{-CO-}$  if malonyl-CoA has been used as the extender unit), originating from the starter or extender units, either as ketones or in the form of carbonyls in phenolic groups ( $-\text{CH}_2\text{-CO-}$  or its tautomeric form  $-\text{CH}=\text{COH-}$ ). In the case of reduced polyketides, a single or all ketones have been reduced to alcohol ( $-\text{CH}_2\text{-CHOH-}$ ) groups by e.g. the KR domain/enzyme, or further to an alkene group ( $-\text{C}=\text{C-}$ ) by e.g. a DH domain/enzyme, or even further to an alkane group ( $-\text{CH}_2\text{-CH}_2\text{-}$ ) 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.

30

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"

40

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 that are necessary or advantageous for the expression of a polynucleotide encoding a polypeptide of the present invention. Each control sequence may be native or  
5 foreign to the nucleotide sequence encoding the polypeptide. Such control sequences include, but are not limited to, a leader, polyadenylation sequence, pro-peptide encoding sequence, promoter, signal peptide encoding 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  
10 introducing specific restriction sites facilitating ligation of the control sequences with the coding region of the nucleotide sequence encoding a polypeptide.

Numbering of the carbon atoms in the polyketides, and the numbering of the individual carbon atoms found in polyketide backbone is counted from the carboxylic acid (-COOH)  
15 end of the molecule. A single or double carbon-carbon bond that links e.g. the 5 and 12 carbon atom, counted from the carboxylic acid end of the polyketide, is represented as C5-C12.

The term "pentadeca" (Greek for "fifteen") denotes a polyketide chain consisting of fifteen  
20 ketide units, meaning that the polyketide backbone consists of 30 carbon atoms.

The term "pfam####" refers to specific motif in the Wellcome Trust Sanger Institute Protein-family (pfam) online database (<http://pfam.xfam.org/>) described in Finn et al. 2014 (R.D. Finn, A. Bateman, J. Clements, P. Coggill, R.Y. Eberhardt, S.R. Eddy, A. Heger, K.  
25 Hetherington, L. Holm, J. Mistry, E.L.L. Sonnhammer, J. Tate, M. Punta. (2014) The Pfam protein families database. Nucleic Acids Research (2014), Database Issue 42:D222-D230), that allows for the identification of conserved functional sequence motifs based on Hidden Markov Models and multiple sequence alignments.

30 The term "starter unit" relates to the first substrate that a PKS selects for incorporation into the growing polyketide chain, and hence the first ketide unit found in the polyketide chain originates from the starter units. The starter unit is delivered as acyl groups bound to Co-enzyme A, such as, but not exclusively, acetyl-CoA, malonyl-CoA, methylmalonyl-CoA, p-coumaroyl-CoA, phenylacetyl-CoA or benzoyl-CoA. Type III PKSs normally uses malonyl-  
35 CoA as extender units, but can use the other starter units.

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  
40 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" is a cell comprising a recombinant polynucleotide (e.g. DNA) molecule and a recombinant host cell will therefore not be understood as covering a natural wildtype cell as such. 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.

10

The term, 'small molecule foldases' relates to enzymes that are capable of catalyzing the formation of intra-molecular carbon-carbon or carbon-oxygen bonds within a molecule, resulting in the formation of aromatic or cyclic motifs within the molecule. These include members of the following enzyme families: pfam04199 (Cyclase superfamily domain), pfam10604 and pfam03364 (SRPBCC Cyclases/aromatases), pfam07876 (DABB Cyclases/aromatases), pfam04673 (Polyketide synthesis cyclase), pfam00753 (Lactamase\_B/MBL fold metallo-hydrolase), ketoreductase from Act cluster, pfam07883 (Cupin\_2) and in addition dissected PT domains from type I iterative PKS from filamentous fungi

20

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 will the "Length of Alignment" 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 "TIGRXXX" denotes a sequence motif in the The Institute of Genomic Research's Protein family database (<http://www.jcvi.org/cgi-bin/tigrfams/Terms.cgi>) that allows for the identification of conserved functional sequence motifs based on Hidden Markov Models and multiple sequence alignments.

The term "a dissected product template domain from type I iterative PKS" denotes an artificially constructed enzyme that only contains the Product Template (PT) portion of a type I non-reducing iterative PKS from fungi. The PT domain can either be identified via the National Center for Biotechnology Information (NCBI) Conserved Domain Database (CDD) and the associated search tool (CD-Search), which is available via <http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi> 0020. In the CDD the PT domain has accession number "TIGR04532: PT\_fungal\_PKS". The artificial enzyme is designed by fusing the coding sequence of the PT domain with a 5' start codon (ATG) and a 3' stop codon (TGA, TAA or TAG).

The term "triketide" (greek for "three") denotes a polyketide chain consisting of three ketide units, meaning that the polyketide backbone consists of 6 carbon atoms. The term "ketide" refers to a –CH<sub>2</sub>-CO- unit.

The term "Type III polyketide synthase (PKS)" is a self-contained enzyme that form homodimers. The single active site in each monomer catalyzes the priming and extension to form polyketide products.

40 **DRAWINGS**



Figure 1: Example of library of aromatic compounds made according to the invention (see working example herein). A population of PKSs (PKS1-PKS<sub>n</sub>) that produces different chain lengths are combined in individual cells with 'small molecule foldases' (Cyc1 to Cyc<sub>n</sub>) that catalyze different folding patterns to form unique products.

Figure 2: The figure shows Extracted Ion Chromatograms (EIC) of the novel compounds synthesized by 5 different GMO strains of *S. cerevisiae* comprising transgenes expressing the following type III PKSs: triketide synthase 2-PS from the plant *Gerbera hybrid* (GH2PS); the pentaketide synthase PCS from the plant *Aloe arborescens* (AaPCS); hexaketide synthase HKS from plant *Drosophyllum lusitanicu* (DluHKS), and the heptaketide synthase PKS3 from the plant *Aloe arborescens* (AaPKS3), and the octaketide synthase OKS from *Aloe arborescens* (AaOKS), as compared to parent control strain lacking these transgenes. Compounds that correspond to the molecular mass of the various detected compounds are:

(A) Triactetic acid at EIC 127.0390 +/- 0.005 (including a triactetic lactone standard-TAL); (B) 5,7-dihydroxy-2-methylchromone (pentaketidepyrone) at EIC 193.0495 +/- 0.005; (C) 6-(2',4'-dihydroxy-6'-methylphenyl)-4-hydroxy-2-pyrone (hexaketidepyrone) at EIC 235.0601 +/-0.005. Retention times, masses and compounds names can be found in table 1.

20

Figure 3: The figure shows Extracted Ion Chromatograms (EIC) of the novel compounds synthesized by 2 different GMO strains of *S. cerevisiae* comprising transgenes expressing the following type III PKSs: the octaketide synthase OKS from *Aloe arborescens* (AaOKS) or the heptaketide synthase PKS3 from the plant *Aloe arborescens* (AaPKS3), as compared to parent control strain lacking these transgenes. Compounds that correspond to molecular mass of the various detected compounds: (A) Heptaketide pyrone (TW93a) at EIC 277.0707 +/-0.005; (B) Aloesone at EIC 233.0808 +/-0.005; (C) The compounds SEK4 / SEK4b at EIC 319.0709 and dehydrated SEK4/SEK4B with a EIC of 319.0812. Retention times, masses and compounds names can be found in table 1.

30

Figure 4: Table showing an example of a library of aromatic compounds synthesized *in vivo* according to the invention (see working example herein). The introduction of a second 'small molecule foldase' (Cyclase  $\alpha$  to Cyclase<sub>n</sub>) into a system that already contain a PKS and a cyclase generates novel compounds.

35

Figure 5: The figure shows Extracted Ion Chromatograms (EIC) of a compound synthesized by 2 different GMO strains of *S. cerevisiae* comprising transgenes co-expressing: hexaketide synthase HKS, a type III PKS from plant *Drosophyllum lusitanicu* (DluHKS), together with a dissected product template domain (small molecule foldase), either: BIK1-PT from *Fusarium graminearum* or mdpG-PT from *Aspergillus nidulans*, as compared to a control GMO strain

40

expressing only the type III PKS, hexaketide synthase HKS from plant *Drosophyllum lusitanicu* (DluHKS). The detected compound #1 corresponds to a molecular mass 225.1120 m/z eluting at 4.89 minutes. A) EIC at 225.1120 m/z for the 'DluHKS+BIK-PT', 'DluHKS+mdpG-PT' and control 'DluHKS' strain. B) UV spectrum for compound #1 eluting at 5 4.89 minutes in the 'DluHKS+mdpG-PT' strain.

Figure 6: The figure shows Extracted Ion Chromatograms (EIC) for a compound synthesized by a GMO strain of *S. cerevisiae* comprising transgenes co-expressing: hexaketide synthase HKS, a type III PKS from plant *Drosophyllum lusitanicu* (DluHKS), together with the cyclase 10 (small molecule foldase) *gra-orf4* from *Streptomyces violaceoruber*, as compared to a control GMO strain expressing only the type III PKS, hexaketide synthase HKS from plant *Drosophyllum lusitanicu* (DluHKS). The compound #2, corresponding to a molecular mass 191.0707 m/z, elutes at 3.95 minutes, that is just detectable in the control strain, is produced in larger amounts in the strain co-expressing the type IIIPKS and the cyclase. A) 15 EIC at 191.0707 m/z for the 'DluHKS+*gra-orf4*' and 'DluHKS' strains. B) UV spectra for compound #2 eluting at 4.89 minutes in the 'DluHKS+*gra-orf4*' strain.

Figure 7: The figure shows Extracted Ion Chromatograms (EIC) for two compounds synthesized by 2 different GMO strains of *S. cerevisiae* comprising transgenes co- 20 expressing: hexaketide synthase HKS, a type III PKS from plant *Drosophyllum lusitanicu* (DluHKS), together with a dissected product template domain (small molecule foldase), either: BIK1-PT from *Fusarium graminearum* or mdpG-PT from *Aspergillus nidulans*, as compared to a control GMO strain expressing only the type III PKS, hexaketide synthase HKS from plant *Drosophyllum lusitanicu* (DluHKS). Detected compound #3, corresponding 25 to a molecular mass (235.0608 m/z) elutes at 2.86 minutes and detected compound #4 (235.0608 m/z) elutes at 3.09 minutes. A) EIC at 235.0606 m/z for the "DluHKS+BIK1-PT", "DluHKS+mdpG-PT" and control "DluHKS" strains. B) UV spectra for compound #3 eluting at 2.86 m minutes in the 'DluHKS+BIK1-PT' strain. C) UV/VIS spectrum compound #4 eluting at 3.8 minutes in the 'DluHKS+BIK1-PT strain'.

30

Figure 8: The figure shows Extracted Ion Chromatograms (EIC) for a compound synthesized by a GMO strain of *S. cerevisiae* comprising transgenes co-expressing: hexaketide synthase HKS, a type III PKS from plant *Drosophyllum lusitanicu* (DluHKS), together with a dissected product template domain (small molecule foldase) mdpG-PT from *Aspergillus nidulans*, as 35 compared to a control GMO strain expressing only the type III PKS, hexaketide synthase HKS from plant *Drosophyllum lusitanicu* (DluHKS). The detected compound #5 corresponding to a molecular mass (237.0757 m/z) elutes at 2.59 minutes, that is just detectable in the control strain, is produced in larger amounts in the strain co-expressing the type III PKS and the mdpG-PT small foldase. A) EIC at 237.0757 m/z for the

'DluHKS+mdpG-PT' and 'DluHKS' strains. B) UV spectra for compound #5 eluting at 2.59 minutes in the 'DluHKS+mdpG-PT' strain.

Figure 9: The figure shows Extracted Ion Chromatograms (EIC) for a compound synthesized by a GMO strain of *S. cerevisiae* comprising transgenes co-expressing: hexaketide synthase HKS, a type III PKS from plant *Drosophyllum lusitanicu* (DluHKS), together with a cyclase ZhuI (small molecule foldase) from the bacterium *Streptomyces* sp. R1128, as compared to a control GMO strain expressing only the type III PKS, hexaketide synthase HKS from plant *Drosophyllum lusitanicu* (DluHKS). The detected compound #6, corresponding to a molecular mass 121.0649 m/z, elutes at 3.57 minutes. A) EIC at 121.0649 m/z of 'DluHKS+ZhuI' and control (DluHKS) strains. B) UV spectra for compound #6 eluting at 3.57 minutes in the 'DluHKS+ZhuI' strain.

Figure 10: The figure shows Extracted Ion Chromatograms (EIC) for compounds synthesized by GMO *Nicotiana benthamiana* lines co-expressing a type III polyketide synthase from *Aloe arborescens* (OKS), together with the cyclases/ketoreductase CYC, CYC\_DH from the actinorhodin biosynthetic gene cluster in *Streptomyces coelicolor* A3 and a ketoreductase (KR) (cyclase superfamily), as compared to a control *N. benthamiana* expressing only the type III PKS (OKS). A) GMO plants expressing only type III PKS (OKS); B) GMO plants expressing type III PKS (OKS) and KR cyclase; C) GMO plants expressing type III PKS (OKS) and the cyclases/ketoreductase CYC, CYC\_DH; D) GMO plants expressing type III PKS (OKS) and the cyclases/ketoreductase CYC and KR; E) GMO plants expressing type III PKS (OKS) and the cyclases/ketoreductase CYC, CYC\_DH and KR.

Figure 11: shows the structure of the heptaketides aloesone, aloesol and O-glucosylated derivatives thereof, synthesized by GMO *N. benthamiana* co-expressing a type III polyketide synthase (HpPKS2) together with several small molecule foldases.

## DETAILED DESCRIPTION OF THE INVENTION

30

### I A method for producing libraries of aromatic compounds

The invention provides a method of producing a library of polyketide-derived aromatic, polyaromatic, cyclic and polycyclic compounds, wherein the carbon atom chain length of the polyketide backbone of the compounds is selected from two or more of 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 and 31 carbon atoms. Alternatively, the carbon atom chain length of the polyketide backbone of the compounds is selected from six, eight, ten, twelve, fourteen, sixteen, eighteen, twenty, twenty-two, and twenty-four or twenty-eight of 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 and 31 carbon atoms. The method employs recombinant cells transformed with different heterologous genes encoding

40

enzymes in a biosynthetic pathway leading to the formation of the library of polyketide-derived aromatic, polyaromatic, cyclic and polycyclic compounds. Surprisingly, the inventors have discovered that a recombinant cell that expresses a heterologous Type III polyketide synthase (PKS) and a heterologous 'small molecule foldase' derived from a fungal/bacterial source, where the aromatase/cyclase and the PKS are derived from a different genus, is capable of producing a non-reduced polyketide which is then converted *in vivo* into an aromatic compound of interest. 'Small molecule foldases' of bacterial or fungal origin are only known to act on polyketides that are bound to ACP within the KS/CLF/ACP enzyme complex of type II PKS or type I PKS. The ability of 'Small molecule foldases' of bacterial or fungal origin, that in nature act on polyketides tethered to PKS I or PKS II, to guide the folding of untethered non-reduced linear polyketides products of PKS III enzymes derived from a different genus was therefore unexpected.

Depending on the specificity of both the PKS III and the small molecule foldase type expressed in a given recombinant cell, a wide range of aromatic compounds of interest can be produced. The inventors have further discovered that a population of heterologous recombinant cells, comprising individual host cells transformed with transgenes encoding different combinations of one type of heterologous Type III polyketide synthase (PKS) and at least one type of heterologous bacterial or fungal 'small molecule foldase', is capable of producing the library of polyketide-derived aromatic, polyaromatic, cyclic and polycyclic compounds.

### **ii Recombinantly expressed heterologous Type III polyketide synthases**

Despite their structural simplicity, type III PKSs are thought to contribute to the biosynthesis of a wide array of compounds in nature, such as chalcones, pyrones, acridones, phloroglucinols, stilbenes, and resorcinolic lipids. The linear non-reduced polyketides produced by type III PKSs are characterized by the presence of ketone groups in the ketides (-CH<sub>2</sub>-CO-), originating from the starter or extender units, either as ketones or in the form of carbonyls in phenolic groups (-CH<sub>2</sub>-CO- or its tautomeric form -CH=COH-). A Type I PKS and/or a Type II PKS may be mutated to remove relevant elements (e.g. active sites) to be converted into a Type III PKS. A PKS, which by the skilled person is functionally considered to be a Type III PKS is herein understood to be a Type III PKS.

Preferably the individual type III PKS used produces products of a single chain length, i.e. only releases products after a fixed number of iterations. This will ensure that the individual recombinant cell in the library only produces one specific product which is desirable as 1) it increases the yields of the the specific product, by reducing the amount of less shunt products, and 2) it eases the identification of the active compound produced by the recombinant cell.

Preferably 80% of the formed polyketides should be of the same chain length, more preferably 90% should be of same chain length, even more preferably 95% should be of the same single chain length and most preferably 99% of the formed product should be of the same chain length.

5

A recombinant cell of the invention comprises a transgene encoding a heterologous Type III PKS, which may be an enzyme that is natively expressed in a bacterial, fungal or plant cell. If the encoded enzyme is of bacterial origin it is preferably selected from *Pseudomonas* or *Streptomyces*.

10

Alternatively, if the enzyme is of fungal origin it is preferably selected from the group consisting of: *Neurospora*, *Fusarium*, *Aspergillus*, and *Monascus*.

If the encoded enzyme is of plant origin, it is preferably selected from the group consisting of: *Gerbera hybrid*, *Aloe arborescens*, *Drosophyllum lusitanicum*, *Plumbago zeylanica*, *Rheum palmate*, *Hypericum perforatum* and *Plumbago indica*.

Preferably, a recombinant cell of the invention comprises a transgene encoding a heterologous Type III polyketide synthase selected from the members of the groups listed below, or shares high amino acid sequence identity with a member of the group. Preferably the amino acid sequence of the heterologous Type III polyketide synthase shares at least 75, 80, 85, 90, 92, 94, 96, 98, 99 or 100% sequence identity with a member of the group. The GenBank ID numbers identifying the polypeptide sequence and corresponding native nucleotide sequence for each member of the groups of Type III polyketide synthases is given in the lists below. The nucleotide sequence of a transgene encoding any member of the group of Type III polyketide synthases may, however, need to be adapted to correspond to a codon usage required for optimal expression in the host recombinant cell.

Type III polyketide synthases selected for forming triketides are preferably: 2-PS [GenBank ID number Z38097.2 (nucleotide SEQ ID No: 1.) and GenBank ID number P48391.2 (polypeptide SEQ ID No: 2)] from *Gerbera hybrid*.

Type III polyketide synthases selected for forming tetraketides are preferably: PhID [GenBank ID number JN561597.1 position 2882 to 3970 (nucleotide SEQ ID No:3) and GenBank ID number AEW67127.1 (polypeptide SEQ ID No:4)] from *Pseudomonas fluorescens* for forming tetraketides.

Type III polyketide synthases selected for forming pentaketides are preferably: PCS [GenBank ID number AY823626 (nucleotide SEQ ID No:5) and GenBank ID number AAX35541.1 (polypeptide SEQ ID No:6)] from *Aloe arborescens* or ORAS GenBank ID

number XM\_955334.2 position 582 to 1919 (nucleotide SEQ ID No:7) and GenBank ID number EGZ68458 (polypeptide SEQ ID No:8)] from *Neurospora crassa* or 1,3,6,8-tetrahydroxynaphthalene synthase [GenBank ID number CP005080 position 7775934 to 7776986 (nucleotide SEQ ID No:9) and GenBank ID number AGK81780 (polypeptide SEQ ID No:10)] from *Streptomyces fulvissimus*.

Type III polyketide synthases selected for forming hexaketides are preferably: PinPKS [GenBank ID number AB259100 (nucleotide SEQ ID No:11) and GenBank ID number BAF44539 (polypeptide SEQ ID No:12)] from *Plumbago indica*, DluHKS [GenBank ID number EF405822 (nucleotide SEQ ID No:13) and GenBank ID number ABQ59603 (polypeptide SEQ ID No:14)] from *Drosophyllum lusitanicum* or PzPKS [GenBank ID number JQ015381 (nucleotide SEQ ID No:15) and GenBank ID number AEX86944 (polypeptide SEQ ID No:16)] from *Plumbago zeylanica* for forming hexaketides.

Type III polyketide synthases selected for forming heptaketides are preferably: ALS [GenBank ID number AY517486 (nucleotide SEQ ID No:17) and GenBank ID number AAS87170 (polypeptide SEQ ID No:18)] from *Rheum palmatum* or AaPKS3 [GenBank ID number EF537574 (nucleotide SEQ ID No:19) and GenBank ID number ABS72373 (polypeptide SEQ ID No:20)] from *Aloe arborescens* for forming heptaketides.

Type III polyketide synthases selected for forming octaketides are preferably: OKS [GenBank ID number AY567707 (nucleotide SEQ ID No:21) and GenBank ID number AAT48709.1 (polypeptide SEQ ID No:22)] or OKS2 [GenBank ID number FJ536166 (nucleotide SEQ ID No:23) and GenBank ID number ACR19997.1 (polypeptide SEQ ID No:24)] or OKS3 [GenBank ID number FJ536167 (nucleotide SEQ ID No:25) and GenBank ID number ACR19998.1 (polypeptide SEQ ID No:26)] from *Aloe arborescens* or HpPKS2 [GenBank ID number HQ529467 (nucleotide SEQ ID No:27) and GenBank ID number AEE69029 (polypeptide SEQ ID No:28)] from *Hypericum perforatum*.

Type III polyketide synthases selected for forming nonaketides are preferably: PCS F80A/Y82A/M207G, a mutated polypeptide - SEQ ID No:29 (derived from GenBank ID number AAX35541.1), from *Aloe arborescens*, having the specified triple point mutation (F80A/Y82A/M207G), and encoded by a synthetic gene.

Type III polyketide synthases selected for forming decaketides are preferably: OKS N222G a mutated polypeptide SEQ ID No: 30 (derived from GenBank ID number AAT48709.1) from *Aloe arborescens* having the specified point mutation (N222G), and encoded by a synthetic gene.

Type III polyketide synthases selected for forming dodecaketides are preferably: OKS F66L/N222G a mutated polypeptide SEQ ID No: 31 [derived from GenBank ID number AAT48709.1] from *Aloe arborescens* having the specified double point mutations (F66L/N222G), and encoded by a synthetic gene.

5

In one embodiment, the population of heterologous recombinant cells comprises host cells, or their clonal derivatives, where each individual cell comprises a transgene capable of expressing a PKS selected from a triketide synthase, tetraketide synthase, pentaketide synthase, hexaketide synthase, heptaketide synthase, octaketide synthase, nonaketide  
10 synthase, decaketide synthase, undecaketide synthase dodecaketide synthase, trideca synthase, tetradeca synthase, and pentadeca synthase. Preferably the population of heterologous recombinant cells is capable of expressing at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 members of this group.

### 15 **iii Biosynthetic properties of the recombinantly expressed heterologous Type III polyketide synthases**

The Type III polyketide synthase, expressed by the host recombinant cell is capable of converting suitable starter unit and extender units into a non-reduced polyketide under suitable incubation conditions. Suitable starter unit are acetyl-CoA or malonyl-CoA and  
20 suitable extender units are malonyl-CoA or methyl-malonyl-CoA. The biosynthesis of aromatic compounds (spontaneously folded polyketides of different chain length) by the host recombinant cell expressing a heterologous Type III polyketide synthase is exemplified in Example 1.

### 25 **iiii Recombinantly expressed heterologous small molecule foldases**

In bacterial type II PKS systems the 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. These enzyme classes are referred to herein as 'small molecule foldases', a group which includes aromatasases and cyclases. In type II  
30 PKS systems, the formation of compounds with multiple aromatic rings typically relies on the successive action of multiple different 'small molecule foldases'. The 'small molecule foldases' can be divided into two groups based on the substrates they act on: where the first small molecule foldases only acts on linear polyketide chains and catalyze the formation of one or more aromatic/cyclic group, the second group of enzymes only accepts  
35 substrates that already contain an aromatic or cyclic group (= products from the first group of 'small molecule foldases') and catalyze the formation of additional aromatic or cyclic groups.

Surprisingly, the inventors have discovered that a bacterial/fungal 'small molecule foldase'  
40 derived from PKS I enzymes or interacting with PKS II enzymes in nature, when co-

expressed with a Type III PKK in a recombinant cell, is capable of promoting a non-spontaneous fold in a non-reduced linear polyketide synthesized by the Type III PKK, thereby preventing its spontaneous folding/aromatization that it would otherwise undergo *in vivo*. Accordingly, the 'small molecule foldase' enzyme has a trans-acting catalytic activity  
5 that allows *in vivo* conversion of the non-reduced polyketide into an aromatic compound of interest. The 'small molecule foldase' enzyme is heterologous with respect to the host cell in which it is expressed, and is derived from a different genus than from which the PKS III is derived. The biosynthesis of a range of different aromatic compounds by the host recombinant cell co-expressing a heterologous Type III polyketide synthase and a  
10 heterologous bacterial/fungal small molecule foldase (where the genus from which the foldase is derived is different from the genus from which the PKSIII are derived), is exemplified in Example 2, 3 and 4.

Preferably, a recombinant cell of the invention co-expresses a Type III PKS together with a  
15 "small molecule foldase" that is an aromatase/cyclase belonging to a family selected from the group: Cyclase superfamily domain pfam04199; SRPBCC cyclase/aromatase superfamily pfam10604 and/or pfam03364, or DABB cyclase/aromatase superfamily pfam07876; Polyketide synthesis cyclase superfamily pfam04673; Lactamase\_B/MBL fold metallohydrolase superfamily pfam00753; ketoreductase from Act cluster; Cupin\_2 superfamily  
20 pfam07883; and a dissected product template domain from type I iterative PKS originating from filamentous fungi.

Preferably, a recombinant cell of the invention comprises at least one transgene encoding a heterologous 'small molecule foldase' selected from the members of the groups listed  
25 below, or shares high amino acid sequence identity with a member of the group. Preferably the amino acid sequence of the heterologous small molecule foldase shares at least 75, 80, 85, 90, 92, 94, 96, 98, 99 or 100% sequence identity with a member of the group. The GenBank ID numbers identifying the polypeptide sequence and corresponding native nucleotide sequence for each member of the groups of small molecule foldase is given in  
30 the lists below. The nucleotide sequence of a transgene encoding any member of the group of 'small molecule foldase' may, however, need to be adapted to correspond to a codon usage required for optimal expression in the host recombinant cell.

A 'first heterologous small molecule foldase' capable of acting on the linear polyketide  
35 product of the type III PKK to form a first ring (and capable of introducing a fold at the given positions in the chain) is preferably selected from the group consisting of:

- ZhuI (type: SRPBCC) [GenBank ID number AF293442.1 (nucleotide SEQ ID No: 32) and GenBank ID number AAG30197.1 (polypeptide SEQ ID No: 33)] from *Streptomyces* sp. R1128 to form a C7-C12 fold in the linear non-reduced polyketide chain;



- pdmD (type: SRPBCC) [GenBank ID number EF151801.1 Position 23865 to 24326 (nucleotide SEQ ID No: 34) and GenBank ID number ABM21750.1 (polypeptide SEQ ID No: 35)] from *Actinomadura hibisca* to form C9-C14+C7-C16 folds;
- 5 • sanI (type: SRPBCC) [GenBank ID number GU937384.1 position 11996 to 12451 (nucleotide SEQ ID No: 36) and GenBank ID number ADG86318.1 (polypeptide SEQ ID No: 37)] from *Streptomyces* sp. SANK 61196;
- pnxD (type: SRPBCC) [GenBank ID number AB469194.1 position 16730 to 17203 (nucleotide SEQ ID No: 38) and GenBank ID number BAJ52684.1 (polypeptide SEQ ID No: 39)] from *Streptomyces* sp. TA-0256;
- 10 • llpCI (type: SRPBCC) [GenBank ID number AM492533.1 position 8866 to 9333 (nucleotide SEQ ID No: 40) and GenBank ID number CAM34342.1 (polypeptide SEQ ID No: 41)] from *Streptomyces tendae*;
- gra-orf4 (type: 2xSRPBCC) [GenBank ID number AJ011500.1 position 32006 to 32980 (nucleotide SEQ ID No: 42) and GenBank ID number CAA09656.1 (polypeptide SEQ ID No: 43)] from *Streptomyces violaceoruber* to form a C9-C14 fold;
- 15 • schP4/SFUL\_4006 (type: 2xSRPBCC) [GenBank ID number CP005080.1 Position 4477979 to 4478932 (nucleotide SEQ ID No: 44) and GenBank ID number AGK78908.1 (polypeptide SEQ ID No: 45)] from *Streptomyces fulvissimus* DSM 40593 to form C7-C12;
- 20 • Erd4 (bifunc) (type: 2xSRPBCC) [GenBank ID number FJ719113.1 Position 3913 to 4863 (nucleotide SEQ ID No: 46) and GenBank ID number ACX83620.1 (polypeptide SEQ ID No: 47)] from uncultured soil bacterium V167 to form a C7-C12 fold;
- med-ORF19 (type: 2xSRPBCC) [GenBank ID number AB103463.1 Position 13942 to 14898 (nucleotide SEQ ID No: 48) and GenBank ID number BAC79027.1 (polypeptide SEQ ID No: 49)] from *Streptomyces* sp. AM-7161 to form a C7-C12 fold;
- 25 • ssfY1 (type: 2xSRPBCC) [GenBank ID number GQ409537.1 Position 9830 to 10774 (nucleotide SEQ ID No: 50) and GenBank ID number ADE34490.1 (polypeptide SEQ ID No: 51)] from *Streptomyces* sp. SF2575 to form a C7-C12 fold;
- oxyK (type: 2xSRPBCC) [GenBank ID number DQ143963.2 Position 11443 to 12396 (nucleotide SEQ ID No: 52) and GenBank ID number AAZ78334.2 (polypeptide SEQ ID No: 53)] from *Streptomyces rimosus* to form a C7-C12 fold;
- 30 • Act\_ARO-CYC\_actVII (type: 2xSRPBCC) [GenBank ID number AL939122.1 Position 162706 to 163656 (nucleotide SEQ ID No: 54) and GenBank ID number Q02055.1 (polypeptide SEQ ID No: 55)] from *Streptomyces coelicolor* A3(2) to form a C7-C12 fold;
- 35 • wA-PT (type: PT domain) [GenBank ID number None - synthetic (nucleotide SEQ ID No: 58) and GenBank ID number CAA46695 position 1276 to 1651 (polypeptide SEQ ID No: 59)] from *Aspergillus nidulan* to form C7-C12 + C1-C10 folds;

- BIK1-PT (type: PT domain) [GenBank ID number None - synthetic (nucleotide SEQ ID No: 60) and GenBank ID number CAB92399 Position 1252 to 1632 (polypeptide SEQ ID No: 61)] from *Fusarium fujikuroi* to form C7-C12 + C1-C10 + C12-C17 folds;
- 5 • PGL1\_PT (type: PT domain) [GenBank ID number None - synthetic (nucleotide SEQ ID No: 62) and GenBank ID number EYB26831 position 1225 to 1655 (polypeptide SEQ ID No: 63)] from *Fusarium graminearum* to form C4-C9 + C2-C11 folds;
- mpdG\_PT (type: PT domain) [GenBank ID number None - synthetic (nucleotide SEQ ID No: 64) and GenBank ID number XP\_657754.1 position 1335 to 1739 (polypeptide SEQ ID No: 65)] from *Aspergillus nidulans* to form C6-C1 + C4-C13 + C2-C15 folds;
- 10 • ZhuI-1 (type: SRPBCC) [GenBank ID number ANIA\_10642 (nucleotide SEQ ID No: 66) and GenBank ID number CBF80957.1 (polypeptide SEQ ID No: 67)] from *Aspergillus nidulans*;
- ZhuI-2 (type: SRPBCC) [GenBank ID number AN3000.2 (nucleotide SEQ ID No: 68) and GenBank ID number XP\_660604.1 (polypeptide SEQ ID No: 69)] from *Aspergillus*
- 15 *nidulans*;
- AOC-1 (type: Dabb) [GenBank ID number AN8584.2 (nucleotide SEQ ID No: 70) and GenBank ID number XP\_681853.1 (polypeptide SEQ ID No: 71)] from *Aspergillus nidulans*;
- AOC-2 (type: Dabb) [GenBank ID number ANIA\_01204 (nucleotide SEQ ID No: 72) and
- 20 GenBank ID number CBF87939.1 (polypeptide SEQ ID No: 73)] from *Aspergillus nidulans*;
- AOC-3 (type: Dabb) [GenBank ID number ANIA\_10997 (nucleotide SEQ ID No: 74) and GenBank ID number CBF79774.1 (polypeptide SEQ ID No: 75)] from *Aspergillus nidulans*;
- 25 • AOC-4 (type: Dabb) [GenBank ID number ANIA\_11021 (nucleotide SEQ ID No: 76) and GenBank ID number CBF80167.1 (polypeptide SEQ ID No: 77)] from *Aspergillus nidulans*;
- AOC-5 (type: Dabb) [GenBank ID number AN1979.2 (nucleotide SEQ ID No: 78) and GenBank ID number XP\_659583.1 (polypeptide SEQ ID No: 79)] from *Aspergillus*
- 30 *nidulans*.

#### **Iv. Additional populations of heterologous recombinant cells for producing a library of aromatic compounds**

The inventors have further discovered that the diversity of aromatic compounds produced

35 by the heterologous recombinant cells of the invention can be extended by transforming each cell of the first population of heterologous recombinant cells with a second, optionally also a third, and optionally also a fourth transgene, where each of the second, third and fourth transgenes encodes a different heterologous 'small molecule foldase'.

The second 'small molecule foldase' is capable of acting on the aromatic polyketide product of the 'first small foldase' to form an additional aromatic group(s), while the third and fourth 'small molecule foldases' are capable of forming additional aromatic groups in an iterative synthesis (and capable of introducing a fold at the given positions in the chain). The biosynthesis of a range of different aromatic compounds by the host recombinant cell co-expressing a heterologous Type III polyketide synthase and one or more heterologous bacterial/fungal small molecule foldases (where the genus from which the foldase is derived is different from the genus from which the PKSIII are derived), is exemplified in Examples 3 and 4.

10

Preferably, the second, third, and fourth heterologous 'small molecule foldase' is one selected from the members of the groups listed below, or shares high amino acid sequence identity with a member of this group. Preferably the amino acid sequence of the second, third, and fourth heterologous 'small molecule foldase' shares at least 75, 80, 85, 90, 92, 94, 96, 98, 99 or 100% sequence identity with a member of this group. The GenBank ID numbers identifying the polypeptide sequence and corresponding native nucleotide sequence for each member of the groups of 'small molecule foldase' is given in the lists below. The nucleotide sequence of a transgene encoding any member of the group of 'small molecule foldase' may, however, need to be adapted to correspond to a codon usage required for optimal expression in the host recombinant cell are preferably selected from the group consisting of:

- ZhuJ (type: Cyclase) [GenBank ID number AF293442.1 (nucleotide SEQ ID No. 80) and GenBank ID number AAG30196.1 (polypeptide SEQ ID No. 81)] from *Streptomyces* sp. R1128 to form a C5-C14 fold;
- oxyN (type: Cyclase) [GenBank ID number DQ143963.2 position 14855 to 15628 (nucleotide SEQ ID No. 82) and GenBank ID number AAZ78337.1 (polypeptide SEQ ID No. 83)] from *Streptomyces rimosus* to form C5-C14+C3-C16 folds;
- jadI (type: Polyketide synthesis cyclase) [GenBank ID number AAD37852.1 position 2020 to 2349 (nucleotide SEQ ID No. 84) and GenBank ID number AF126429.1 (polypeptide SEQ ID No. 85)] from *Streptomyces venezuelae* to form C4-C17 folds;
- LndF (type: Polyketide synthesis cyclase) [GenBank ID number AY659997.1 (nucleotide SEQ ID No. 86) and GenBank ID number AAU04837.1 (polypeptide SEQ ID No. 87)] from *Streptomyces globisporus* to form C4-C17+C2-C19 folds;
- pgaF (type: Polyketide synthesis cyclase) [GenBank ID number AHGS01000054.1 position 6389 to 6724 (nucleotide SEQ ID No. 88) and GenBank ID number EHN79050.1 (polypeptide SEQ ID No. 89)] from *Streptomyces coelicoflavus* to form C2-C19 folds;
- Act\_CYC (type: Lactamase) [GenBank ID number X63449.1 Position 3830 to 4723 (nucleotide SEQ ID No. 90) and GenBank ID number CAA45047.1 (polypeptide SEQ ID No. 91)] from *Streptomyces coelicolor* A3(2);

- sanE (type: None) [GenBank ID number AF228524.1 position 15 to 584 (nucleotide SEQ ID No. 92) and GenBank ID number AAF61923.1 (polypeptide SEQ ID No. 93)] from *Streptomyces ansochromogenes*;
- pnxK (type: Polyketide synthesis cyclase) [GenBank ID number AB469194.1 position 13057 to 13380 (nucleotide SEQ ID No. 94) and GenBank ID number BAJ52679.1 (polypeptide SEQ ID No. 95)] from *Streptomyces* sp. TA-0256;
- pnxL (type: Cupin\_2 ) [GenBank ID number AB469194.1 position 13377 to 13901 (nucleotide SEQ ID No. 95) and GenBank ID number BAJ52680.1 (polypeptide SEQ ID No. 97)] from *Streptomyces* sp. TA-0256;
- llpCII (type: Cupin\_2 ) [GenBank ID number AM492533.1 position 12120 to 12548 (nucleotide SEQ ID No. 98) and GenBank ID number CAM34346.1 (polypeptide SEQ ID No. 99)] from *Streptomyces tendae*;
- llpCIII (type: Polyketide synthesis cyclase) [GenBank ID number AM492533.1 position 12545 to 12880 (nucleotide SEQ ID No. 100) and GenBank ID number CAM34347.1 (polypeptide SEQ ID No.101)] from *Streptomyces tendae*;
- ZhuJ-1 (type: Cyclase) [GenBank ID number AN5060.2 (nucleotide SEQ ID No.102) and GenBank ID number XP\_662664.1 (polypeptide SEQ ID No.103)] from *Aspergillus nidulans*;
- ZhuJ-2 (type: Cyclase) [GenBank ID number ANIA\_11053 (nucleotide SEQ ID No.104) and GenBank ID number CBF74060.1 (polypeptide SEQ ID No.105)] from *Aspergillus nidulans*;
- ZhuJ-3 (type: Cyclase) [GenBank ID number ANIA\_10146 (nucleotide SEQ ID No.106) and GenBank ID number CBF88175.1 (polypeptide SEQ ID No.107)] from *Aspergillus nidulans*;
- ZhuJ-4 (type: Cyclase) [GenBank ID number AN5068.2 (nucleotide SEQ ID No.108) and GenBank ID number XP\_662672.1 (polypeptide SEQ ID No.109)] from *Aspergillus nidulans*;

#### **Iv Aromatic compounds produced by the recombinant cells of the invention**

- 30 In a preferred embodiment, the library of aromatic compounds may include aromatic compounds in the size range of C<sub>6</sub>-C<sub>31</sub>. The library of aromatic compounds produced by the method of the invention will comprise two to 10<sup>6</sup> different compounds.

#### **Ivi A recombinant cell**

- 35 The term "recombinant cell" used in the method of the invention may be a eukaryotic cell [e.g. filamentous fungal cell, a yeast cell or a plant cell] or a prokaryotic cell. Preferably the cell is a yeast cell, that may be selected from the group consisting of Ascomycetes, Basidiomycetes and fungi imperfecti, more preferably an Ascomycete.

Preferably, the Ascomycetes yeast cell is selected from the group consisting of *Ashbya*, *Botryosascus*, *Debaryomyces*, *Hansenula*, *Kluveromyces*, *Lipomyces*, *Saccharomyces* spp e. g. *Saccharomyces cerevisiae*, *Pichia* spp., *Schizosaccharomyces* spp.

- 5 Most preferably, the yeast cell is a yeast cell selected from the group consisting of *Saccharomyces* spp e. g. *Saccharomyces cerevisiae*, and *Pichia* spp.

The recombinant host cell may be a cell selected from the group consisting of a filamentous fungal cell. Filamentous fungi include all filamentous forms of the subdivision Eumycota and  
10 Oomycota (as defined by Hawksworth et al. , 1995, supra). Preferably the filamentous fungal cell is a species of *Acremonium*, *Aspergillus*, *Fusarium*, *Humicola*, *Mucor*, *Myceliophthora*, *Neurospora*, *Penicillium*, *Thielavia*, *Tolypocladium*, and *Trichoderma* or a teleomorph or synonym thereof. For example, the filamentous fungal cell may be an *Aspergillus* cell, in particular *Aspergillus niger*, *Aspergillus oryzae* or *Aspergillus nidulans*.

15

When the recombinant cell is a bacterial cell, it is preferably 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*, *B. amyloliquefaciens* or *B. licheniformis*. A preferred *Streptomyces* cell is *S. setonii* or *S. coelicolor*. A preferred  
20 *Corynebacterium* cell is *C. glutamicum*. A preferred *Pseudomonas* cell is *P. putida* or *P. fluorescens*.

#### **Ivii Production of the library of aromatic compounds by the heterogeneous populations of recombinant cells**

- 25 The one or more heterogeneous populations of recombinant cells are incubated and/or cultivated under conditions that support synthesis of the library of polyketide-derived aromatic, polyaromatic, cyclic and polycyclic compounds. Suitable cultivation conditions depend on the nature of the host recombinant cell. When the host recombinant cell is a yeast, filamentous fungal or bacterial cell, the cultivation medium (aqueous liquid or solid  
30 medium) will comprise nutrients (carbon source, minerals, essential vitamins and substrates for polyketide biosynthesis, e.g. but not exclusively acetate and malonate) necessary for the biosynthetic activity of the host cell and for host cell growth. When the host cell is a plant cell, the cultivation medium may provide a source of water and light.

#### **Iviii Screening the library of aromatic compounds**

- The method of producing a library of polyketide-derived, polyaromatic, cyclic and polycyclic compounds, may include the step of screening the compounds produced by the population of heterologous recombinant cells, wherein each recombinant cell clone present in the one or more heterogeneous population of recombinant cells is grown individually on a solid  
40 support, or individually in a liquid culture. Screening for compounds with antibiotic

properties may be performed by growing the individual member on the recombinant cell library on a surface of bacteria and then observing the formation of clearing zones around the recombinant cells/colonies. Alternatively, the screen may be based on a light or color forming reaction that the formed compound promotes or inhibits. Alternatively the screen  
5 may be performed using in cell assays, build into the recombinant host cells prior to construction of the libraries.

### **Iix Recovery of the library of aromatic compounds**

The method of producing a library of polyketide-derived, polyaromatic, cyclic and polycyclic  
10 compounds, may include the step of recovering the polyketide-derived aromatic, polyaromatic, cyclic and polycyclic compounds produced by the one or more heterogeneous populations of recombinant cells or produced by one or more of the recombinant cell clones present in the one or more heterogeneous populations of recombinant cells. Recovery may be performed by dilution plating or by re-streaking the population onto selective solid  
15 media.

### **II One or more populations of heterologous recombinant cells for production of a library of aromatic compounds**

The invention provides one or more populations of heterologous recombinant cells,  
20 comprising cells capable of producing polyketide-derived aromatic, polyaromatic, cyclic and polycyclic compounds, wherein the carbon atom chain length of the polyketide backbone of the compounds is selected from two or more of 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 and 31 carbon atoms. Maintenance and replication of the individual cells, or clonal derivatives thereof, in the one or more  
25 populations will depend on the nature of the host recombinant cells, and that are known in the art.

### **III A method for the construction of a population of recombinant host cells for production of a library of aromatic compounds**

30 The following method illustrates one way of constructing population(s) of recombinant host cells capable of producing a library of a polyketide-derived aromatic, polyaromatic, cyclic and polycyclic compounds, wherein the carbon atom chain length of the polyketide backbone of the compounds is selected from two or more of 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 and 31 carbon atoms.  
35 Alternatively, the carbon atom chain length of the polyketide backbone of the compounds is selected from six, eight, ten, twelve, fourteen, sixteen, eighteen, twenty, twenty-two, and twenty-four or twenty-eight of 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 and 31 carbon atoms. The method involves transforming each individual member of the host cell population with a transgene encoding a  
40 heterologous type III PKS and one or more transgenes each encoding a different

heterologous 'small molecule foldase(s)', as described in Section I. The method comprises the following steps:

- 5 (i) creating a library of transgenes encoding type III PKSs that is populated by different type III PKSs, where the individual type III PKS is responsible for forming a linear non-reduced polyketide chain of a specific length, wherein the carbon atom chain length of the polyketide backbone of the chain is selected from 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 and 31 carbon atoms.
- 10 (ii) Creating a library of transgenes encoding different types of 'small molecule foldase(s)' that are populated by different foldases that individually catalyze the formation of one or more specific intramolecular carbon-carbon bonds in linear non-reduced polyketides of variable length (from (i)), resulting in the formation of aromatic compounds with different and unique folding patterns.
- 15 (iii) Creating one or more libraries of transgenes encoding different types of 'small molecule foldase(s)' that are populated by different foldases that individually catalyze the formation of one or more specific intramolecular carbon-carbon bonds in non-reduced polyketides of variable length with one or more aromatic groups (from 1(ii)), resulting in the formation of aromatic compounds with different and unique folding patterns.
- 20 (iv) The libraries described in 1(i), 1(ii) and 1(iii) consist of transgenes where the sequences encoding the said genes are operationally linked to regulatory and cis-acting sequences that allows for transcription and translation in a recombinant host cell. The transgenes are preferably cloned into vectors, which can comprise one or more selection marker encoding genes and the vectors may additionally include:
  - 25 i. Sequences that allows for autosomal replication of the vector in the recombinant host cell, or
  - ii. Sequences that allows for targeted integration of the vector into the genome of the recombinant host cell, or
  - 30 iii. Sequence that allows for transfer of the contents of the vectors to another organism by conjugation.
- (v) Randomly combining the PKS type III library described in 1(i) with
  - i. library 1(ii) or
  - ii. library 1(ii) plus library 1(iii) or

- iii. library 1(ii) plus two or three members of library 1(iii).
- (vi) Co-transformation of said libraries into a population of host cells, such that each individual cell comprises at least one transgene from library (i) and (ii) and optionally one or more additional transgene from library (iii).
- 5 (vii) Optionally replicating the heterologous population of transformed cells produced in step (vi); and optionally storing the population, in a manner that each transformed cell produced in step (vi) and its clonal derivatives can be recovered individually.
- (viii) Optionally isolating individual recombinant host cells from the population of host  
10 cells, to establish pure (isogenetic) cultures of the isolated recombinant host cell.

An alternative to the above described method, is as follows: Each library of transgenes described in 1(i), 1(ii) and 1(iii), optionally cloned into vectors, is individually transformed into a population of host cells, such that each individual cell of the library comprises at least one transgene from library (i), or (ii), or library (iii). The transgenes from library (ii), and  
15 optionally library (iii) transformed into the respective populations of host cells, can be transferred to the host cell population comprising library (i) by conjugation, cell-cell fusion or crossing such that the each cell in the resulting population of heterologous host cells comprises at least one transgene encoding a Type III PKS and one or more transgene encoding 'small molecular foldases'.

20

## EXAMPLES

### 25 **Example 1 – Library of PKSs that produce polyketides of different lengths in *S. cerevisiae***

This example aims to show how the expression of different type III PKSs in *S. cerevisiae* result in the formation of a range of different aromatic compounds *in vivo*. This concept is illustrated in Figure 1.

### 30 **Methods**

Five different type III polyketide synthases of variable origin were selected for heterologous expression in *S. cerevisiae*; the triketide synthase 2-PS from the plant *Gerbera hybrida*, the pentaketide synthase PCS from the plant *Aloe arborescens*, hexaketide synthase HKS from the plant *Drosophyllum lusitanicum*, heptaketide synthase PKS3 from the plant *Aloe*  
35 *arborescens*, and the octaketide synthase OKS from *Aloe arborescens*. The genes were codon optimized for expression in *S. cerevisiae* using the GeneArt GeneOptimizer algorithm (LifeTechnologies). The *de novo* synthesized genes were delivered in shuttle vectors, and the coding sequences were amplified by PCR using the primers listed below:



**Primerlist:**

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

5	Sc_Gh_2-PS-F	5'-ATCAACGGGdUAAAAATGGGTTCTACTCTTCTGATGATGTTG-3'	SEQ ID No. 110
	Sc_Gh_2-PS-R	5'-CGTGCGAdUTTAGTTACCATTAGCAACAGCAGCAGTAACTC-3'	SEQ ID No. 111
	Sc_AaOKS-F	5'-ATCAACGGGdUAAAAATGAGTAGTTTATCAAATGCCAGTCAC-3'	SEQ ID No. 112
	Sc_AaOKS-R	5'-CGTGCGADUTTACATCAATGGCAAGGAATGCAATAAG-3'	SEQ ID No. 113
10	Sc_Aa_PCS-F	5'-ATCAACGGGdUAAAAATGTCCTCCTGTCTAATTCCTTGC-3'	SEQ ID No. 114
	Sc_Aa_PCS-R	5'-CGTGCGAdUTTACATCAAAGGCAAAGAATGCA-3'	SEQ ID No. 115
	Sc_DluHKS-F	5'-ATCAACGGGdUAAAAATGGCTTTCGTTGAAGGTATGGGT-3'	SEQ ID No. 116
	Sc_DluHKS-R	5'-CGTGCGAdUTTAGTTGTTGATTGGGAAGGATCTCAAGA-3'	SEQ ID No. 117
	Sc_AaPKS3/ALS-F	5'-ATCAACGGGdUAAAAATGGGTTCTTGTCTGATTCTACTCCA-3'	SEQ ID No. 118
15	Sc_AaPKS3/ALS-R	5'-CGTGCGAdUTTAGACTGGTGGCAAAGAATGCAACA-3'	SEQ ID No. 119
	Promoter-F	5'- ACGTATCGCdUGTGAGTCGTATTACGGATCCTTG -3'	SEQ ID No. 120
	Promoter-R	5'- CGTGCGAdUGCCGCTTGTTTATATTTGTTG -3'	SEQ ID No. 121

**Generation of plasmid constructs for expression in *S. cerevisiae***

20 The used primers included 5' overhangs that allowed for directional cloning into the 2-micron pBOSAL1 vector, by the Uracil-Specific Excision Reagent Cloning (USER) technique, described in Nour-Eldin *et al.* 2006 (Hussam H. Nour-Eldin, Bjarne G. Hansen, Morten H. H. Nørholm, Jacob K. Jensen, and Barbara A. Halkier. Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. *Nucleic Acids Res.* 2006, 34(18):

25 e122.). The *PGK1* promoter was also PCR amplified from the vector pSP-G2, using the primers PGK1-d and PGK\_F, as described in (Mikkelsen MD, Buron LD, Salomonsen B, Olsen CE, Hansen BG, Mortensen UH. Halkier BA. Microbial production of indolyglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. *Metab Eng.* 2012;14:104-111). The PCR amplicons were purified via 1% agarose gel

30 electrophoresis and the Illustra 'GFX PCR DNA and gel band purification kit (GE Healthcare). The recipient vector pCfB257, was digested with AsiSI and Nb.BsmI, and the used restriction enzymes were subsequently heat inactivated. The individual purified coding sequences were combined with the digested recipient vector and the purified promoter element and treated with the USER enzyme mix (NEB) and transformed into chemical

35 competent *E. coli* DH5-alpha cells, as described in Nour-Eldin *et al.* 2006. Directional cloning resulted in the creation of an expression cassette, as described in Mikkelsen *et al.* 2012. Transformants were selected for on Luria-Bertani (LB) agar supplemented with ampicillin. Plasmid DNA from colonies were purified using the GenElute kit (Sigma-Aldrich) and the size and restriction enzyme digestion pattern were analyzed and compared to the

theoretical expected sizes and patterns for the individual plasmid. Final verification of the five constructed plasmids consisted of two overlapping sequencing reactions.

The validated plasmids were digested with NotI to liberate the expression/targeting cassette from each of the five plasmids. The liberated expression cassettes were transformed into  
5 the competent *S. cerevisiae* cells CEN.PK102-5B, mating type a via the lithium acetate/single-stranded carrier DNA/polyethylene glycol transformation method (Gietz, R.D., Schiestl, R.H., 2007. "High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method". Nat. Protoc. 2, 31–34). Transformants were selected for by culturing on SC-Leu agar plates as described in Mikkelsen *et al* 2012. Correct transformants were  
10 identified by colony-PCR using the gene specific primers.

### **Growth of *S. cerevisiae*, metabolite extraction and LC-MS/MS analysis**

The verified *S. cerevisiae* strains, called Sc.CEN.PK::2m::2-PS, Sc.CEN.PK::2m::PCS, Sc.CEN.PK::2m::HKS, Sc.CEN.PK::2m::PKS3 and Sc.CEN.PK::2m::OKS, were cultured in  
15 300 ml Erlenmeyer flasks with either 100 ml liquid SC-Ura or Yeast-Peptone-Dextrose medium (REF). The cultures were allowed to grow for 3 days at 30 °C with 150 rpm orbital shake, after which the cells were harvested by centrifugation. The produced metabolites were extracted from the cells using isopropanol:ethyl acetate (1:3 v/v) with 1% formic acid and from the medium using ethyl-acetate. The solvents were evaporated and the analytes  
20 were resuspended in HPLC grade methanol. The analytes were separated using a Dionex Ultimate 3000 UHPLC equipped with a diode array detector (DAD) system hyphenated to a Q-TOF mass spectrometer. The samples were analyzed with three different injects volumes 1 µl, 5 µl and 10 µl. For separation in the UHPLC system a reversed-phase Kinetex C18 (100 mm, 2,1 mm, 2,6 µm) column was used and the temperature was maintained at 40 °C  
25 and a flow rate of 400 µl/min. The used mobile phases consisted of MilliQ water with 20 mM formic acid (A) and acetonitrile with 20 mM formic acid (B). The analytes were eluted using a gradient starting at 10% solvent B and increased to 100% solvent B over a period of 15 minutes. The column was washed with 100% solvent B for 3 minutes and re-equilibrated for 2.4 minutes with 10% B before the next sample was injected. The analytes were detected  
30 via an online DAD (Dionex Ultimate 3000) detector from 200 to 600 nm and an online maXis 3G Qq-Oa-TOF (Bruker Daltronics GmbH). In the MS the analytes were ionized by electrospray operating in positive mode; capillary voltage at 4.5 kV, nebulizer gas at 2.4 bar, drying gas flow at 12 ml/min and a drying temperature of 220 °C. The MS was used in full scan mode in the mass range of 100-1000 Da. The instrument was calibrated using  
35 sodium formate (HCOONa) (Fluka, analytical grade). The obtained data were processed and handled using Compass DataAnalysis v. 4.0 SP4 Build 281 (Bruker Daltronics). Bruker Daltronics Compass IsotopicPattern was used for calculating isotopic patterns of the pseudo-molecular ion and adducts. An in-house standard of triaceticlactone (spontaneously folded triketide) was run under the same conditions to confirm identity of the produced  
40 triketide. Identification of other aromatic polyketids were performed via detection of the

monoisotopic molecular mass ( $[M+H]^+$ ), supported by the maximal UV absorption wavelengths (nm) for the individual compound as specified in figure 4 in Karppinen *et al.* 2008 Octaketide-producing type III polyketide synthase from *Hypericum perforatum* is expressed in dark glands accumulating hypericins, FEBS 275(17): 4329–4342.

5

### **Results:**

Expression of the five PKSs in *S. cerevisiae* resulted in production of new metabolites not observed in the reference strain not expressing any of the five genes (Table 1, Figure 2, 10 and Figure 3).

**Table 1: Products produced from the heterologous expression of type III PKS in *S. cerevisiae*.**

RT [min]	$[M+H]^+$	Mol. form.	Putative compound	GH 2PS	Aa PCS	Dlu HKS	Aa PKS3	Aa OKS
1.3	127.039	C6H6O3	Triacetic lactone	+	+	nd	nd	nd
4.2	193.0495	C10H8O4	Pentaketide pyrone	nd	+	+	+	nd
3.06	235.0601	C12H10O5	Hexaketide pyrone	nd	nd	+	+	+
3.15	277.0707	C14H12O6	Heptaketide pyrone / TW93a	nd	nd	nd	+	+
3.85	233.0808	C13H12O4	Aloesone	nd	nd	nd	+	+
3.3	319.0812	C16H14O7	SEK4	nd	nd	nd	nd	+
3.5	319.0812	C16H14O7	SEK4b	nd	nd	nd	nd	+

RT: retention time;  $[M+H]^+$ : positive molecular ion mass; '+' indicates whether the given compound was 15 detected upon expression of the given PKS. 'nd' indicates that the compound was not detected in the sample.

### **Conclusion:**

Heterologous expression of the five different type III PKS in *S. cerevisiae* resulted in the production of novel compounds, representing spontaneously folded tri-, penta-, hexa-, 20 hepta- and octaketides, in the individual strains. These results demonstrate that it is possible to functionally express type III PKS in *S. cerevisiae* and obtain products similar to those reported in the literature for *in vitro* experiments with purified enzymes. The compounds that have previously been obtained in *in vitro* experiments are the result of spontaneous folding/cyclization of the formed linear non-reduced polyketides. The example 25 shows that *S. cerevisiae* does not express any endogenous enzymes capable of preventing or altering the spontaneous folding/cyclization pattern. This demonstrates that *S. cerevisiae* does not contain any enzymatic activities that will interfere with attempts to control and direct folding of the formed linear non-reduced polyketide by introducing heterologous cyclases/aromatases.

30

**Example 2 – Combining the PKS library with a library of 'small molecule foldases' in *S. cerevisiae***

This example aims to show how different combinations of PKSs and cyclases can result in the formation of a range of different aromatic compounds. This concept is illustrated in Figure 4.

## 5 **Methods**

Four different 'small molecule foldases', including three different bacterial cyclases/aromatases and two product template (PT) domains, dissected from fungal type I iterative polyketide synthases, were selected for heterologous expression in *S. cerevisiae*; ZhuI from the bacterium *Streptomyces* sp. R1128 (C7-C12), gra-orf4 from the bacterium *Streptomyces violaceoruber* (expected C9-C14), BIK1-PT from fungi *Fusarium graminearum* (expected C2-C7) and mdpG-PT from *Aspergillus nidulans* (expected C6-C11).

The genes were codon optimized for expression in *S. cerevisiae* using the GeneArt GeneOptimizer algorithm (LifeTechnologies). The *de novo* synthesized genes were delivered in shuttle vectors, and the coding sequences were amplified by PCR using the primers listed below:

### **Primers used for the construction process, where dU represents 2-deoxyuridine:**

Sc_ZhuI-F	5'-AGCGATACGdUAAAAATGAGACACGTTGAACACACAGTTACCG-3' SEQ ID No.
20 122	
Sc_ZhuI-R	5'-CACGCGAdUTTATTATGCAGTTACGGTACCAACACCAC-3' SEQ ID No. 123
Sc_BIK1-PT-F	5'-AGCGATACGUAAAAATGAGATTGTCCGATTCCGTTTCCACA-3' SEQ ID No. 124
Sc_BIK1-PT-R	5'-CACGCGAUTTAAATCAAACCAGAAGCTGAACCAACTG-3' SEQ ID No. 125
Sc_gra-orf4-F	5'-AGCGATACGdUAAAAATGGCTAGAACTGCTGCTTTGC-3' SEQ ID No. 126
25 Sc_gra-orf4-R	5'-CACGCGAdUTTAACCTGCTTCAGCAGCTTCAGC-3' SEQ ID No. 127
Sc_mdpG-PT-F	5'-AGCGATACGUAAAAATGTCTGGTTTGAGAACTTCCACCG-3' SEQ ID No. 144
Sc_mdpG-PT-F	5'-CACGCGAUTTAGACCAAAGCTTTAGCAGCAACTGAA-3' SEQ ID No. 145

The four 'small molecule foldases' encoding genes were cloned into the pCfB389 vector as described for the five Type III PKS genes in Example 1. The used vector allows for targeted integration into the XI-2 site in the genome of *S. cerevisiae*, as described in Mikkelsen *et al.* 2006. The expression cassettes were transformed into the Sc.CEN.PK 111-61A mating type alpha and selected for on SC-Ura plates. Correct transformants were identified by colony-PCR using the gene specific primers. The obtained verified strains are hereafter referred to as Sc.CEN.PK::XI-2::ZhuI, Sc.CEN.PK::XI-2::gra-orf4, Sc.CEN.PK::XI-2::BIK1-PT, and Sc.CEN.PK::XI-2::mdpG-PT respectively.

The *S. cerevisiae* strains Sc.CEN.PK::2m::HKS and Sc.CEN.PK::2m::OKS, described in Example 1, is in the present example (Example 2) used to exemplify a library of different type III PKSs that produce polyketides of different lengths.

The five foldases were crossed with the type III PKS HKS expressing strains Sc.CEN.PK::2m::HKS, to form diploids yielding five new combinatory strains each containing a PKS and a cyclase/aromatase. The Sc.CEN.PK::2m::OKS strains was crossed  
5 with the Sc.CEN.PK::XI-2::ZhuI. Mating between the PKS carrying strains (mating type a, Leu marker) and the foldase carrying strains (mating type alpha, URA3 marker) was performed by co-inoculating the respective strains combinations on YPD agar plates. The plates were incubated at 30° C for 8 hours, after which the cultures were replica plated onto SC-leu-ura, to select for diploids containing both the selective markers, and incubated at  
10 30° C for four days. Colonies from the double selective plates were streaked onto fresh SC-leu-ura plates to purify them. Single colonies of the diploids containing both the PKS and a foldase were inoculated in shake flasks with 20 mL Delft Synthetic Minimal Medium lacking leucine and uracil, but with added histidine. The cultures were incubated at 30 °C with shake for 4-5 days.

15

The production of novel metabolites was analyzed by UHPLC-HRMS as described in Example 1.

### **Results:**

20 Combining the DluHKS (type III PKS) with the dissected product template domain from mdpG-PT or BIK1-PT resulted in the production of a novel compound with a  $[M+H]^+$  225.1120 m/z which eluted at 4.89 minutes (**Figure 5A**). The UV spectrum of the compound (**Figure 5B**) shows that the compound includes a conjugated bond systems (absorption maxima at 222 nm and 280 nm) similar to what is found in phenolic compounds  
25 with a single aromatic group

Co-expression of DluHKS (type III PKS) and the cyclase gra-orf4 results in the accumulation of increased concentrations (9 times) of a compound with a  $[M+H]^+$  of 191.0707 at 3.95 minutes (**Figure 6A**) compared to when DluHKS is expressed alone. The absorption  
30 maxima of 222 nm and 290 nm (**Figure 6B**) support that the compound includes a conjugated aromatic bond systems characteristic of aromatic compounds.

Expression of DluHKS (type III PKS) with the dissected product template domain (PT) from mdpG or BIK1-PT resulted in a significant increase of the concentrations of two compounds  
35 with a  $[M+H]^+$  of 235.0606 eluting at 2.86 minutes and 3.08 minutes (**Figure 7A**). The compounds at 2.86 minutes has absorption maxima at 212 nm and 302 nm, while the compound at 3.08 minutes absorbs at 220 nm, 250 nm and 294 nm supporting that the two compounds includes aromatic conjugated bond systems (**Figure 7B**).

Combining the DluHKS (type III) with the dissected product template domain (PT) from mdpG resulted in a seven fold increase in the concentration of a compounds with a  $[M+H]^+$  of 237.0757 eluting at 2.58 minutes (**Figure 8A**), The compound absorbs at 218 nm and 276 nm (**Figure 8B**), indicative of a aromatic conjugated bond systems.

5

Co-expression of DluHKS (type III) with the cyclase ZhuI resulted in a six fold increase in the concentration of a compound eluting at 3.57 min and with an  $[M+H]^+$  of 121.0649 (**Figure 9A**). The compound exhibit absorption maxima at 222 nm and 278 nm indicating that the compound includes a aromatic conjugated bond system (**Figure 9B**).

10

### **Conclusion:**

These results show that co-expression of a type III PKS and a heterologous cyclase/aromatase or dissected product template domain from a type I iterative PKS in the host cell *Saccharomyces cerevisiae* results in the formation of novel compounds than what is observed when the PKS is expressed alone. In several cases the co-expression resulted in the significant increase in the formation of aromatic compounds otherwise produced at low concentrations when the PKS is expressed alone. These results surprisingly shows that 'small molecule foldases' originating from bacterial or fungal type I and type II PKS systems, which in nature act on ACP-bound polyketides, can act on free non-reduced linear polyketides produced by type III PKSs.

20

### **Example 3 – Introducing a type III polyketide synthase (OKS) together with cyclases/ketoreductase CYC, CYC\_DH and KR (cyclase superfamily) into *Nicotiana benthamiana* (*N. benthamiana*).**

25

This example illustrates how the introduction of cyclases/ketoreductases, together with a type III polyketide synthase, OKS in *N. benthamiana*, can further increase the compound diversity. This concept is illustrated in Figure 4.

### 30 **Methods**

#### ***Generation of plasmid constructs for expression in N. benthamiana.***

*CYC* (*actIORF5*) and *CYC\_DH* (*actIORF4*) from the actinorhodin biosynthetic gene cluster in *Streptomyces coelicolor* A3 (2) (Genbank accession: X63449.1) were codon optimized for *N. benthamiana* expression, whereas *KR* (Genbank accession: M19536) was codon optimized for *E. coli* expression. All three genes were purchased as synthetic DNA fragments from Genscript together with the native sequence of *OKS* from *Aloe arborescens* (Genbank accession: AY567707). All synthetic fragments were used as PCR templates with compatible deoxyuracil(dU)-containing primers (see table 1) to generate constructs that were cloned into *pEAQ-HT-USER* (Sainsbury et al., 2009) by USER technology. All *pEAQ-HT-USER*

40

plasmid constructs were transformed into the *Agrobacterium tumefaciens* strain, *AGL-1* and infiltrated into leaves 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.).

**Primer sequences for amplification of different gene constructs.**

Gene fragments	Primer sequence
10 OKS-Forward	5'-GGCTTAA/dU/ATGAGTTCACCTCTCCAACGCTTCCCATC-3' SEQ ID No. 130
OKS-Reverse	5'-GGTTTAA/dU/TTACATGAGAGGCAGGCTGTGGAGAAGGATAGT-3' SEQ ID No. 131
ZhuI-Forward	5'-GGCTTAA/dU/ATGAGGCATGTTCGAGCAT-3' SEQ ID No. 132
ZhuI-Reverse	5'-GGTTTAA/dU/TTATGCCGTGACAGTTCGACAC-3' SEQ ID No. 133
15 ZhuJ-Forward	5'-GGCTTAA/dU/ATGTCCGGACGTAAGACG-3' SEQ ID No. 134
ZhuJ-Reverse	5'-GGTTTAA/dU/TTAATCTTCTCCTCCTGTTCAA-3' SEQ ID No. 135
CYC-Forward	5'-GGCTTAA/dU/ATGACTGTTGAAGTTCGT-3' SEQ ID No. 136
CYC-Reverse	5'-GGTTTAA/dU/TTAAGCCAAGCAAGTAGGAAGT-3' SEQ ID No. 137
CYC_DH-Forward	5'-GGCTTAA/dU/ATGTCAAGACCTGGAGAA-3' SEQ ID No. 138
20 CYC_DH-Reverse	5'-GGTTTAA/dU/TTAGCTTGCCGGCCCAGC-3' SEQ ID No. 139
KR-Forward	5'-GGCTTAA/dU/ATGGCAACCCAGGATAGCGAAGTTGCAC-3' SEQ ID No. 140
KR-Reverse	5'-GGTTTAA/dU/TTAATAGTTGCCAGACCACCACAAACATTCAG-3' SEQ ID No. 141
25 HpPKS2-Forward	5'-GGCTTAA/dU/ATGGGTTCCCTTGACAATGGT-3' SEQ ID No. 142
HpPKS2-Reverse	5'-GGTTTAA/dU/TTAGAGAGGCACACTTCGGAGAA-3' SEQ ID No. 143

**Metabolite extraction and LC-MS/MS analysis**

Compounds produced when *OKS* was co-expressed with *CYC*, *CYC\_DH* and *KR* were extracted from discs ( $\varnothing=3\text{cm}$ ) 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<sub>HTS</sub> HV 0.45  $\mu\text{m}$  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<sub>18</sub>(2) column (150  $\times$  4.6 mm, 3  $\mu\text{m}$ , 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.

Linear gradient elution profiles were used: 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.

### Results:

- 5 Introduction and co-expression of *OKS* and *KR* together with either *CYC* and/or *CYC\_DH* in *N. benthamiana*, resulted in production of novel compounds with the masses and retention time shown in the table 2 and Figure 10.

10 **Table 2: Novel compounds produced from the *in vivo* combination of OKS with cyclases/ketoreductases.**

RT [min]	m/z (ESI+)	Molecular formula	B: OKS+KR	C: OKS+KR +CYC_DH	D: OKS+KR+CYC	E: OKS+KR+CYC+CYC_DH
12,32	188,0693	C11H10NO2		+	+	+
13,14	237,0754	C12H12O5	+	+	+	+
13,81	235,0953	C13H14O4	+	+	+	+
16,14	299,0544	C16H11O6				+
16,57	285,0771	C15H12O5		+		
16,66	303,0879	C16H14O6	+			
17,19	235,0956	C13H14O4		+		
17,9	285,0769	C15H12O5		+		
18,23	303,0887	C16H14O6	+		+	
19,4	285,0767	C15H12O5		+		
19,69	303,0885	C16H14O6			+	
19,97	275,212	C17H26N2O			+	+
21,19	285,0768	C15H12O5		+		
32,58	301,1788	C19H24O3	+	+		

- + indicate in which combination the polyketide synthase and foldases produced specific novel polyketide-derived compounds. LC-MS chromatograms in which the novel polyketide-derived compounds were identified from the different combinations (B-E), can be found in Figure 3. RT: retention time and m/z: mass-to-charge ratio and ESI+: positive electrospray ionisation.
- 15

### Conclusion

- 20 The heterologous co-expression, also defined as combinations, of *OKS* from *Aloe arborescens* with foldases (*CYC* and *CYC\_DH*) and *KR* from *Streptomyces coelicolor* A3 (2) gives rise to the production of novel compounds, including polyketides of different chain-length and derivatives thereof in *N. benthamiana*.



**Example 4 – Introducing a type III polyketide synthase (HpPKS2) together with cyclases/ketoreductase ZhuI, ZhuJ, CYC, CYC\_DH and KR (cyclase superfamily) into *N. benthamiana*.**

## 5 **Methods**

### **Generation of plasmid constructs for expression in *N. benthamiana*.**

*CYC* (*actIORF5*) and *CYC\_DH* (*actIORF4*) from the actinorhodin biosynthetic gene cluster in *Streptomyces coelicolor* A3 (2) (Genbank accession: X63449.1), *ZhuI* (Genbank accession: AAG30197) and *ZhuJ* (Genbank accession: AAG30196) were codon optimized for *N. benthamiana* expression, whereas *KR* (Genbank accession: M19536) was codon optimized for *E. coli* expression. All five genes were purchased as synthetic DNA fragments from Genscript together with the native sequence of *HpPKS2* from *Hypericum perforatum* (Genbank accession: HQ529467). All synthetic fragments were used as PCR templates with compatible deoxyuracil(dU)-containing primers (see table 1) to generate constructs that were cloned into *pEAQ-HT-USER* by USER technology. All *pEAQ-HT-USER* plasmid constructs were transformed into the *Agrobacterium tumefaciens* strain, *AGL-1* and infiltrated into leaflets 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.).

### **Metabolite extraction and LC-MS/MS analysis**

Extraction protocol was as described in example 4.

## **Results**

The co-expression of the type III polyketide synthase *HpPKS2* together with either *ZhuI*, *ZhuJ* and/or *KR* in *N. benthamiana*, resulted in the production of novel polyketide-derived compounds. Among these novel compounds the heptaketide aloesone, aloesol and *O*-glucosylated varieties thereof were identified (figure 11).

## **Conclusion**

The heterologous co-expression, also defined as combinations, of *HpPKS2* with foldases (*ZhuI* and *ZhuJ*) and *KR* from *Streptomyces coelicolor* A3 (2) give rise to the production of novel compounds, including polyketides of different chain-lengths and derivatives thereof in *N. benthamiana*.

## 40 **REFERENCES**

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.)

5

Sainsbury, F., Theunemann, EC., Lomonossoff, GP., (2009) pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants, *Plant Biotechnology Journal* 7(7): 682-693.

10

**CLAIMS**

1. A method of producing a polyketide-derived aromatic, polyaromatic, cyclic or polycyclic compound, wherein the carbon atom chain length of the polyketide backbone of the compounds is selected from 6 - 31 carbon atoms, comprising the steps of:
- 5
- a. providing a recombinant cell comprising:
- 10
- i. a transgene encoding a heterologous type III polyketide synthase capable of forming a linear non-reduced polyketide compound, wherein the carbon atom chain length of the polyketide backbone of the formed compound is selected from 6 - 31 carbon atoms; and
- 15
- ii. a transgene encoding a first heterologous 'small molecule foldase' enzyme capable of catalyzing the formation of one or more region-specific intramolecular carbon-carbon or carbon-oxygen bonds in a linear non-reduced polyketide compound, wherein the carbon atom chain length of the polyketide backbone of the compound is one or more of 6 - 31 carbon atoms,
- 20
- wherein the 'small molecule foldase' enzyme is a bacterial or fungal enzyme, and wherein the genus from which said bacterial or fungal enzyme is derived is different from the genus from which said PKSIII enzyme is derived, and
- 25
- wherein the recombinant cell is capable of producing polyketide-derived aromatic, polyaromatic, cyclic or polycyclic compound, wherein the carbon atom chain length of the polyketide backbone of the compound is selected from among 6 - 31 carbon atoms; and
- 30
- b. incubating and/or culturing the recombinant cell in a culture medium to support synthesis of the polyketide-derived aromatic, polyaromatic, cyclic or polycyclic compound.
- 35
2. A method of producing a library of polyketide-derived aromatic, polyaromatic, cyclic, and polycyclic compounds, wherein the carbon atom chain length of the polyketide backbone of the compounds is selected from two or more of 6 - 31 carbon atoms, comprising the steps of:

- a. providing one or more heterogeneous populations of recombinant cells, wherein each cell in the one or more populations comprises:
- 5 i. a transgene encoding a heterologous type III polyketide synthase capable of forming a linear non-reduced polyketide compound , wherein the carbon atom chain length of the polyketide backbone of the formed compound is selected from 6 - 31 carbon atoms; and
  - 10 ii. a transgene encoding a first heterologous 'small molecule foldase' enzyme capable of catalyzing the formation of one or more region-specific intramolecular carbon-carbon or carbon-oxygen bonds in a linear non-reduced polyketide compound , wherein the carbon atom chain length of the polyketide backbone of the compound is one or more of 6 - 31 carbon atoms,
- wherein the 'small molecule foldase' enzyme is a bacterial or fungal enzyme, and wherein the genus from which said bacterial or fungal enzyme is derived is different from the genus from which the PKSIII enzyme is derived, and
- wherein the one or more populations of recombinant cells comprises cells capable of producing polyketide-derived aromatic, polyaromatic, cyclic, and/or polycyclic compounds, wherein the carbon atom chain length of the polyketide backbone of the compounds is selected from two or more of 6 - 31 carbon atoms; and
- 25 b. incubating and/or culturing the one or more heterogeneous populations of recombinant cells in a culture medium to support synthesis of the library of polyketide-derived aromatic, polyaromatic, cyclic, and polycyclic compounds.
3. The method of claim 2, further comprising a step of:
- 30 c. screening the library of polyketide-derived aromatic, polyaromatic, cyclic, and polycyclic compounds, wherein each recombinant cell, or its clonal derivatives, present in the one or more heterogeneous population of recombinant cells is grown individually on a solid support, or individually in a liquid culture.
4. The method of claim 2 or 3, further comprising the step of recovering the polyketide-derived aromatic, polyaromatic, cyclic, and polycyclic compounds produced by the one or more heterogeneous populations of recombinant cells or
- 35

produced by one or more of the recombinant cell clones present in the one or more heterogeneous populations of recombinant cells.

5. The method according to any of claims 1-4, wherein the heterologous type III polyketide synthase is selected from the group consisting of:
- 5           a. Triketide synthase polypeptide, wherein the amino acid sequence of the polypeptide has at least 70% sequence identity to 2-PS (SEQ ID No.2) from *Gerbera hybrid*;
- b. Tetraketide synthase polypeptide, wherein the amino acid sequence of the polypeptide has at least 70% sequence identity to PhID (SEQ ID No.4) from  
10           *Pseudomonas fluorescens*;
- c. Pentaketide synthase polypeptide, wherein the amino acid sequence of the polypeptide has at least 70% sequence identity to a sequence selected from the group consisting of PCS (SEQ ID No.6) from *Aloe arborescens*, ORAS (SEQ ID No.8) from *Neurospora crassa*, and 1,3,6,8-  
15           tetrahydroxynaphthalene synthase (SEQ ID No.10) from *Streptomyces fulvissimus*;
- d. Hexaketide synthase polypeptide, wherein the amino acid sequence of the polypeptide has at least 70% sequence identity to a sequence selected from the group consisting of PinPKS (SEQ ID No.12) from *Plumbago indica*,  
20           DluHKS (SEQ ID No.14) from *Drosophyllum lusitanicum*, and PzPKS (SEQ ID No.16) from *Plumbago zeylanica*;
- e. Heptaketide synthase polypeptide, wherein the amino acid sequence of the polypeptide has at least 70% sequence identity to ALS (SEQ ID No.18) from *Rheum palmatum* or AaPKS3 (SEQ ID No.20) from *Aloe arborescens*;
- 25           f. Octaketide synthase polypeptide, wherein the amino acid sequence of the polypeptide has at least 70% sequence identity to a sequence selected from the group consisting of OKS (SEQ ID No.22), OKS2 (SEQ ID No.24), OKS3 (SEQ ID No.26) from *Aloe arborescens* or HpPKS2 (SEQ ID No.28) from *Hypericum perforatum*;
- 30           g. Nonaketide synthase polypeptide, wherein the amino acid sequence of the polypeptide has at least 70% sequence identity to PCS F80A/Y82A/M207G (SEQ ID No.29) from *Aloe arborescens*;

- h. Decaketide synthase polypeptide, wherein the amino acid sequence of the polypeptide has at least 70% sequence identity to OKS N222G (SEQ ID No.30) from *Aloe arborescens*; and
- 5 i. Dodecaketide synthase polypeptide, wherein the amino acid sequence of the polypeptide has at least 70% sequence identity to OKS F66L/N222G (SEQ ID No.31) from *Aloe arborescens*.
6. The method of any one of claims 1-5, wherein the cell comprises one or more transgene encoding a second, third and fourth heterologous 'small molecule foldase' enzyme capable of catalyzing the formation of one or more region-specific
- 10 intramolecular carbon-carbon or carbon-oxygen bonds in a non-linear polyketide compound, and
- wherein the second, third and fourth 'small molecule foldase' enzymes are bacterial or fungal enzymes, and
- wherein the genus from which said bacterial or fungal enzymes is derived is different
- 15 from the genus from which the PKSIII enzyme is derived.
7. The method of any one of claims 1-6, wherein one or more of said first, second, third and fourth heterologous 'small molecule foldase' enzyme has cyclase or aromatase catalytic activity and a corresponding structural domain selected from the group consisting of:
- 20 a. a pfam04199 cyclase superfamily domain;
- b. a pfam10604 or pfam03364 SRPBCC superfamily domain;
- c. a pfam07876 Dabb superfamily domain;
- d. a pfam04673 Polyketide synthesis cyclase superfamily domain;
- 25 e. a pfam00753 Lactamase\_B/MBL fold metallo-hydrolase superfamily domain;
- f. a pfam07883 Cupin-2 superfamily domain;
- g. Dissected Product template (TIGR04532) domains from type I iterative PKS from filamentous fungi.
8. The method of any one of claims 1-7, wherein said first heterologous 'small molecule foldase' enzyme is selected from one or more of the groups consisting of:
- 30 a. SRPBCC Foldase, wherein the amino acid sequence of the polypeptide has at least 70% sequence identity to a sequence selected from the group

- 5 consisting of ZhuI (SEQ ID No.33) from *Streptomyces* sp. R1128, pdmD (SEQ ID No.35) from *Actinomadura hibisca*, sanI (SEQ ID No.37) from *Streptomyces* sp., SANK 61196; pnxD (SEQ ID No.39) from *Streptomyces* sp. TA-0256, IlpCI (SEQ ID No.41) from *Streptomyces tendae*; ZhuI-1 (SEQ ID No.66) from *Aspergillus nidulans* or ZhuI-2 (SEQ ID No.69) from *Aspergillus nidulans*;
- 10 b. 2xSRPBCC foldase, wherein the amino acid sequence of the polypeptide has at least 70% sequence identity to a sequence selected from the group consisting of gra-orf4 (SEQ ID No.43) from *Streptomyces violaceoruber*, schP4 (SEQ ID No.45) from *Streptomyces fulvissimus* DSM 40593, Erd4 (SEQ ID No.47) from uncultured soil bacterium V167, med-ORF19 (SEQ ID No.49) from *Streptomyces* sp. AM-7161, ssfY1 (SEQ ID No.51) from *Streptomyces* sp. SF2575, oxyK (SEQ ID No.53) from *Streptomyces rimosus*, Act\_ARO-CYC (SEQ ID No.55) from *Streptomyces coelicolor* A3(2);
- 15 c. Dabb foldase, wherein the amino acid sequence of the polypeptide has at least 70% sequence identity to a sequence selected from the group consisting of AOC-1 (SEQ ID No.71) from *Aspergillus nidulans*, AOC-2 (SEQ ID No.73) from *Aspergillus nidulans*, AOC-3 (SEQ ID No.75) from *Aspergillus nidulans*, AOC-4 (SEQ ID No.77) from *Aspergillus nidulans*, or AOC-5 (SEQ ID No.79) from *Aspergillus nidulans*; and
- 20 d. Dissected PT domain, wherein the amino acid sequence of the polypeptide has at least 70% sequence identity to a sequence selected from the group consisting of wA-PT (SEQ ID No.59) from *Aspergillus nidulans* to form C7-C12 + C1-C10, BIK1-PT (SEQ ID No.6) from *Fusarium fujikuroi*, PGL1\_PT (SEQ ID No.63) from *Fusarium graminearum*, mpdG\_PT (SEQ ID No.65) from *Aspergillus nidulans* or curs2-PT (GenBank AGC95321.1 position 1270 to 1613) from *Aspergillus*.
- 25
9. The method of any one of claims 6 - 8, wherein one or more of said second, third or fourth heterologous 'small molecule foldase' enzymes is selected from one or more of the groups consisting of:
- 30
- a. Cyclase foldase, wherein the amino acid sequence of the polypeptide has at least 70% sequence identity to a sequence selected from the group consisting of ZhuJ (SEQ ID No.81) from *Streptomyces* sp. R1128, oxyN (SEQ ID No.83) from *Streptomyces rimosus*, jadI (SEQ ID No.85) from *Streptomyces venezuelae*, LndF (SEQ ID No.86) from *Streptomyces globisporus*, pgaF (SEQ ID No.89) from *Streptomyces coelicoflavus*, pnxK
- 35

(SEQ ID No.95) from *Streptomyces* sp., IlpCIII (SEQ ID No.101) from *Streptomyces tendae*, Act\_CYC (SEQ ID No.91) from *Streptomyces coelicolor* A3(2), sanE (SEQ ID No.93) from *Streptomyces ansochromogenes*;

- 5 b. Cupin foldase, wherein the amino acid sequence of the polypeptide has at least 70% sequence identity to a sequence selected from the group consisting of pnxL (SEQ ID No.95) from *Streptomyces* sp. TA-0256, IlpCII (SEQ ID No.99) from *Streptomyces tendae*,
- 10 c. Cyclase foldase, wherein the amino acid sequence of the polypeptide has at least 70% sequence identity to a sequence selected from the group consisting of ZhuJ-1 (SEQ ID No.103) from *Aspergillus nidulans*, ZhuJ-2 (SEQ ID No.105) from *Aspergillus nidulans*, ZhuJ-3 (SEQ ID No.107) from *Aspergillus nidulans*, ZhuJ-4 (SEQ ID No.109) from *Aspergillus nidulans*;
- 15 10. The method of any one of claim 1-9, wherein the recombinant cell or the recombinant cells in the one or more heterogeneous populations, is selected from among a bacterial cell, a filamentous fungal cell, a yeast cell and a plant cell.
- 20 11. The method according to claim 10, wherein the yeast cell is an Ascomycete selected from the group consisting of *Ashbya*, *Botryoaescus*, *Debaryomyces*, *Hansenula*, *Kluyveromyces*, *Lipomyces*, *Saccharomyces* spp and the filamentous fungal cell is selected from the group consisting of *Acremonium*, *Aspergillus*, *Fusarium*, *Humicola*, *Mucor*, *Myceliophthora*, *Neurospora*, *Penicillium*, *Thielavia*, *Tolyposcladium*, and *Trichoderma*.
- 25 12. The method according to claim 10, wherein the bacterial cell is selected from the group consisting of: *Bacillus*, *Streptomyces*, *Corynebacterium*, *Pseudomonas*, lactic acid bacteria and an *E. coli* cell.
- 30 13. The method according to claim 10, wherein the recombinant host cell is a *Nicotiana benthamiana* or *Arabidopsis thaliana* plant cell.
14. A heterogeneous population of recombinant cells capable of producing a library of polyketide-derived aromatic, polyaromatic, cyclic, and/or polycyclic compounds, according to the method of any one of claims 2 - 5, 7, 8, 10-13, wherein each cell in the population comprises:
- a. a transgene encoding a heterologous type III PKS capable of forming a polyketide-derived aromatic, polyaromatic, cyclic, and/or polycyclic compound, wherein the carbon atom chain length of the polyketide backbone of the formed compound is selected from 6 - 31 carbon atoms; and



5 b. a transgene encoding a first heterologous 'small molecule foldase' enzyme capable of catalyzing the formation of one or more specific intramolecular carbon-carbon bonds in a polyketide-derived aromatic, polyaromatic, cyclic and polycyclic compound, wherein the carbon atom chain length of the polyketide backbone of the compound is one or more of 6 - 31 carbon atoms,

10 wherein the 'small molecule foldase' enzyme is a bacterial or fungal enzyme, and wherein the genus from which said bacterial or fungal enzyme is derived is different from the genus from which the PKSIII enzyme is derived,

15 wherein the population of recombinant cells comprises cells capable of producing polyketide-derived aromatic, polyaromatic, cyclic, and/or polycyclic compounds, wherein the carbon atom chain length of the polyketide backbone of the compounds is selected from two or more of 6 - 31 carbon atoms.

20 15. The heterogeneous population of recombinant cells of claim 14, wherein each cell in the population further comprises one or more transgene encoding a second, third and fourth heterologous 'small molecule foldase' enzyme capable of catalyzing the formation of one or more region-specific intramolecular carbon-carbon or carbon-oxygen bonds in a non-linear polyketide compound, and wherein the second, third and fourth 'small molecule foldase' enzymes are bacterial or fungal enzymes, and wherein the genus from which said bacterial or fungal enzymes is derived is different from the genus from which the PKSIII enzyme is derived.

25

Figure 1

	PKS1	PKS2	PKS3	PKS4	...	PKS <sub>n</sub>
<b>No cyclase</b>						
<b>Cyclase 1</b> (C1-C6)						
<b>Cyclase 2</b> (C2-C7)						
<b>Cyclase 3</b> (C3-C8)						
<b>Cyclase 4</b> (C4-C9)						
<b>Cyclase 5</b> (C5-C10)						
...						
<b>Cyclase<sub>n</sub></b>						

Figure 2 A Triacetic lactone (TAL)

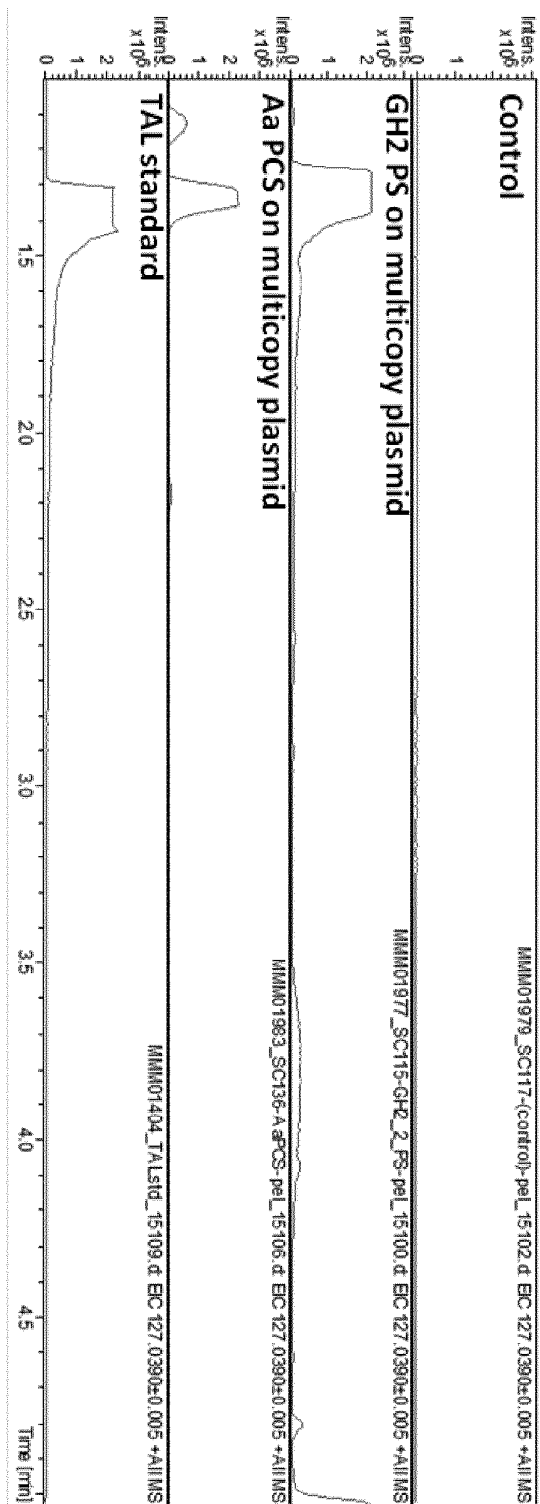


Figure 2 B 5,7-dihydroxy-2-methylchromone (pentaketidepyrone)

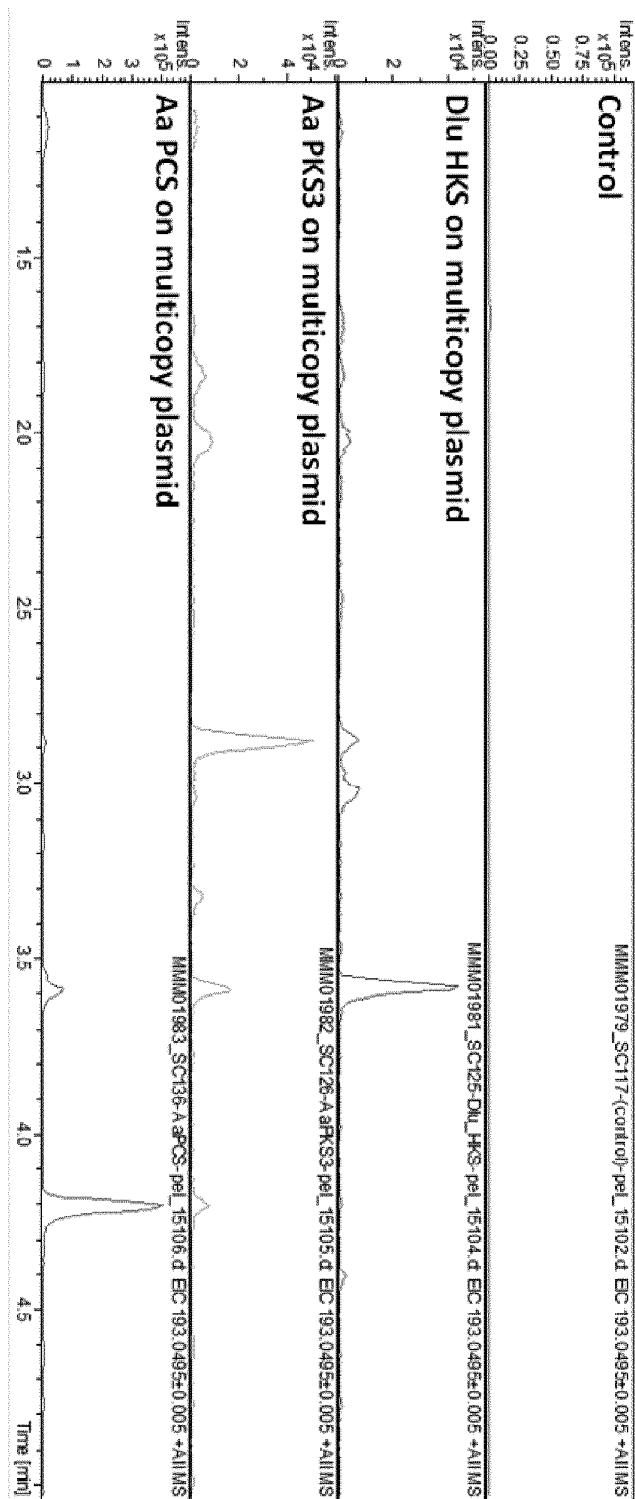


Figure 2 C 6-(2',4'-dihydroxy-6'-methylpheynyl)-4-hydroxy-2-pyrone (hexaketidepyrone)

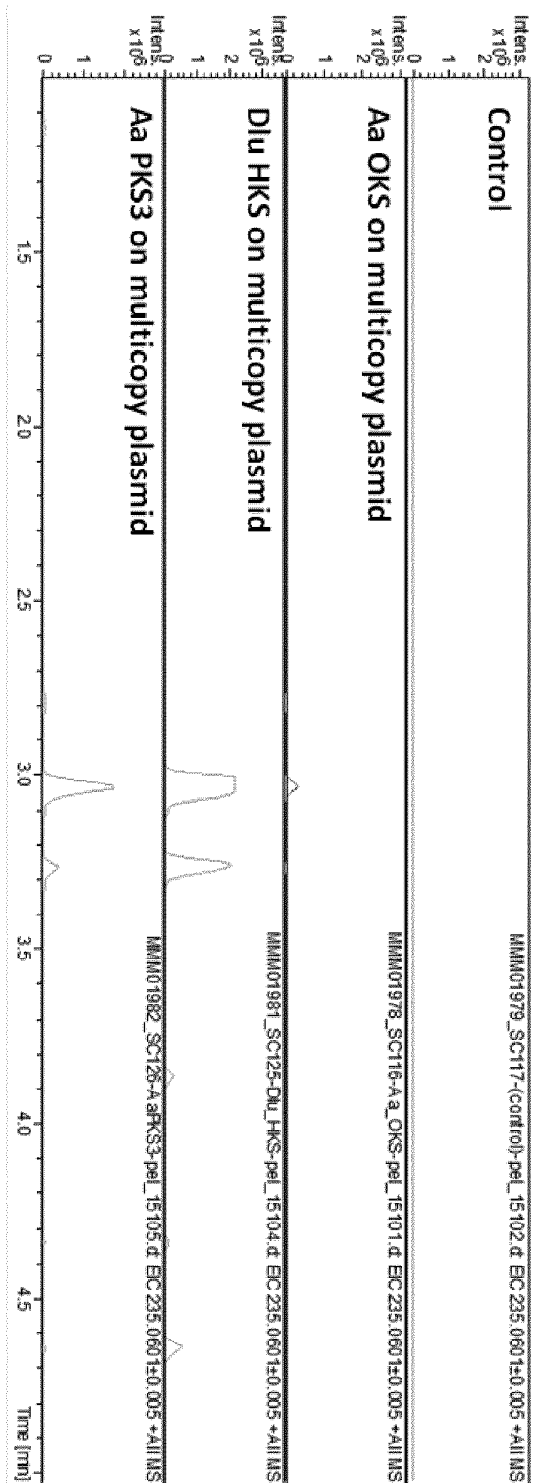


Figure 3 A Heptaketide pyrone or TW93a

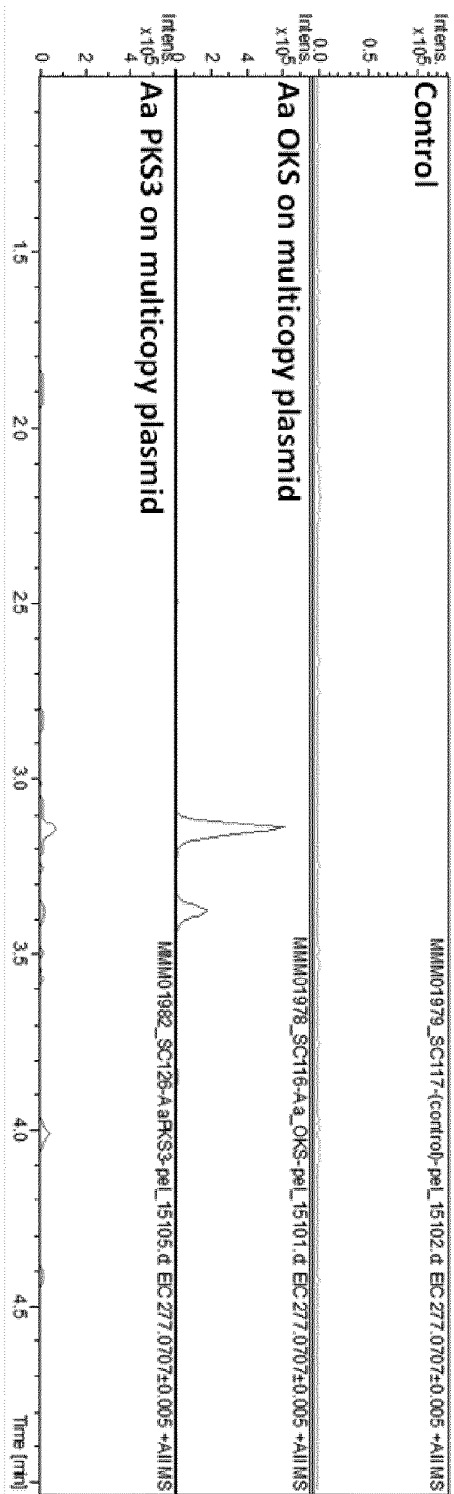


Figure 3 B Aloesone

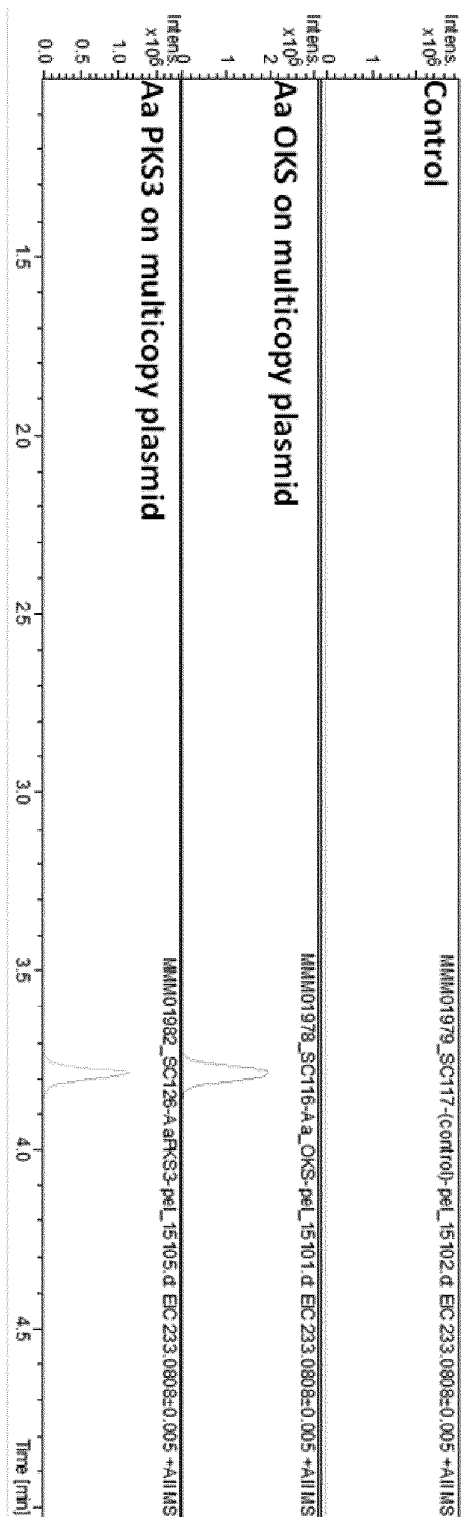


Figure 3 C SEK4/SEK4b

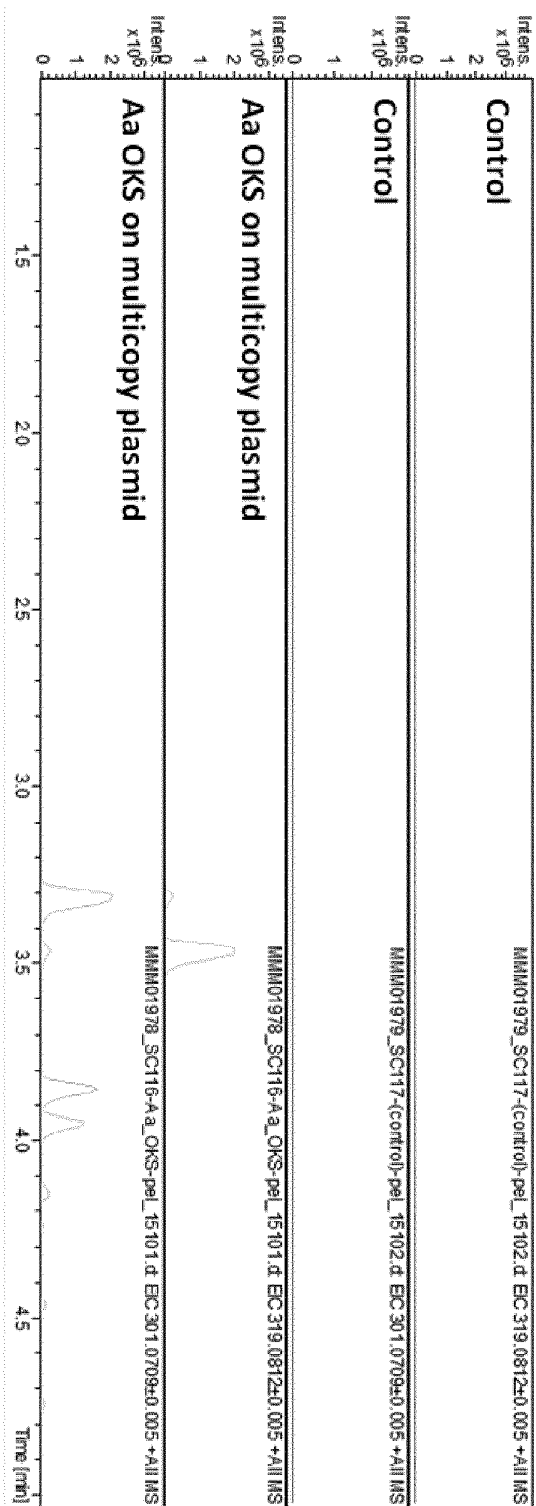




Figure 4

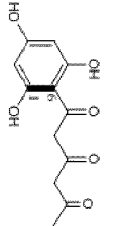
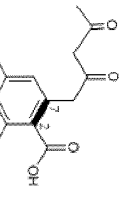
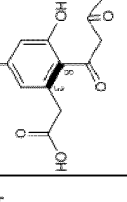
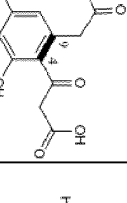
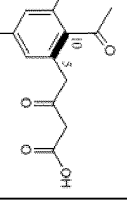
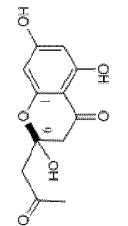
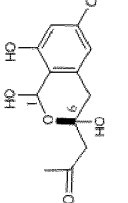
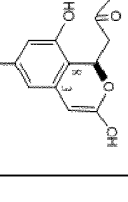
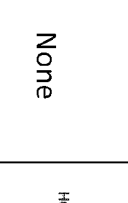
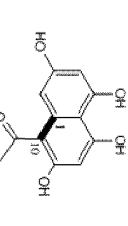
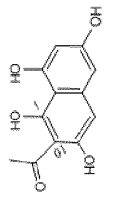
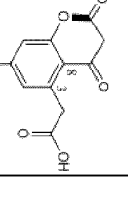
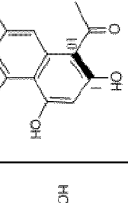
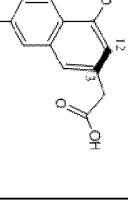
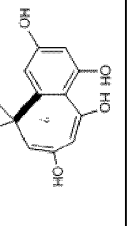
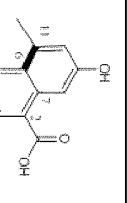
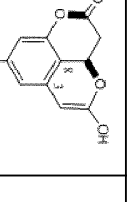
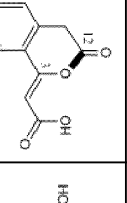
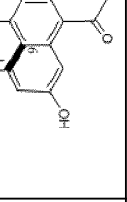
		<b>PKS4 + Cylc1</b>	<b>PKS4 + Cylc2</b>	<b>PKS4 + Cylc3</b>	<b>PKS4 + Cylc4</b>	<b>PKS4 + Cylc5</b>	...	<b>PKS<sub>n</sub></b>
<b>None</b>								
<b>Cyclase α</b>				<b>None</b>				
<b>Cyclase β</b>								
<b>Cyclase χ</b>								
...								
<b>Cyclase<sub>n</sub></b>								

Figure 5

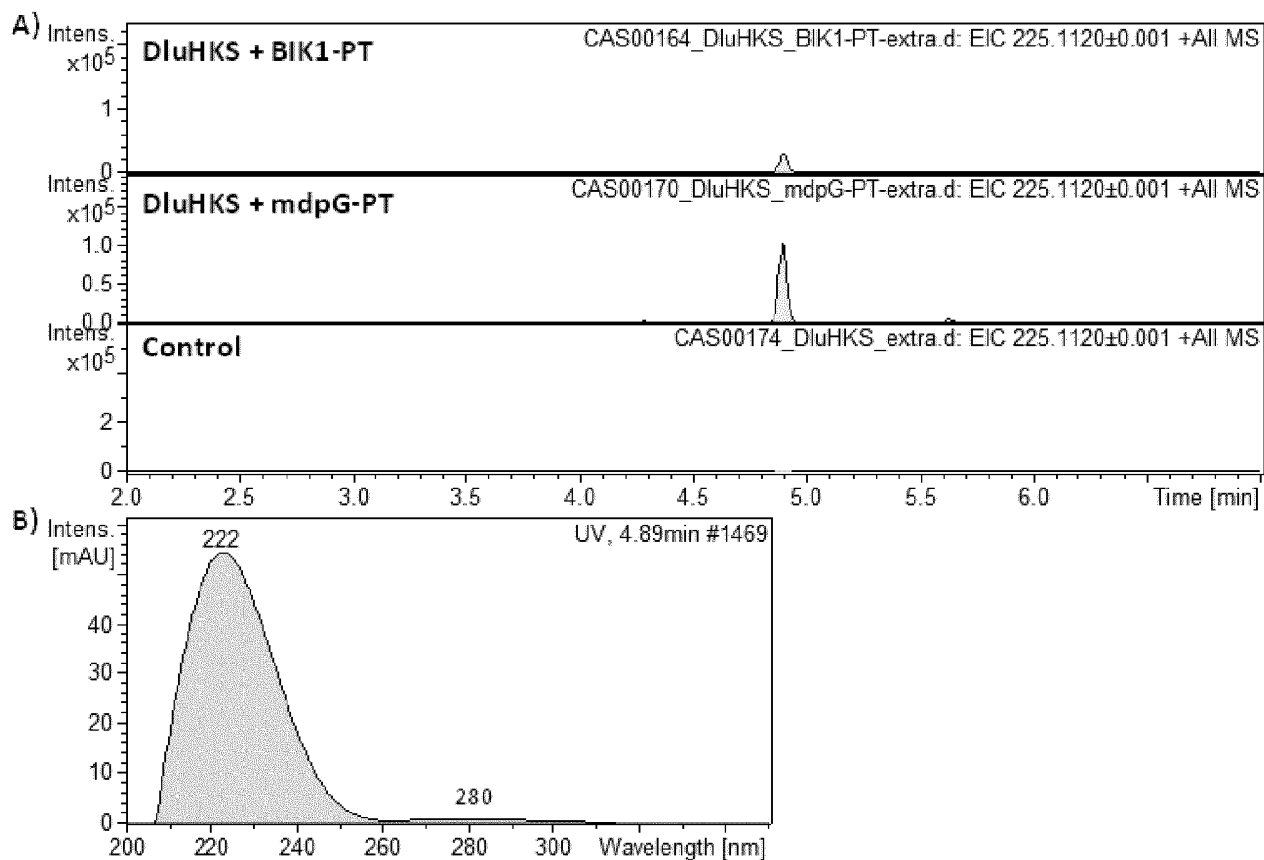


Figure 6

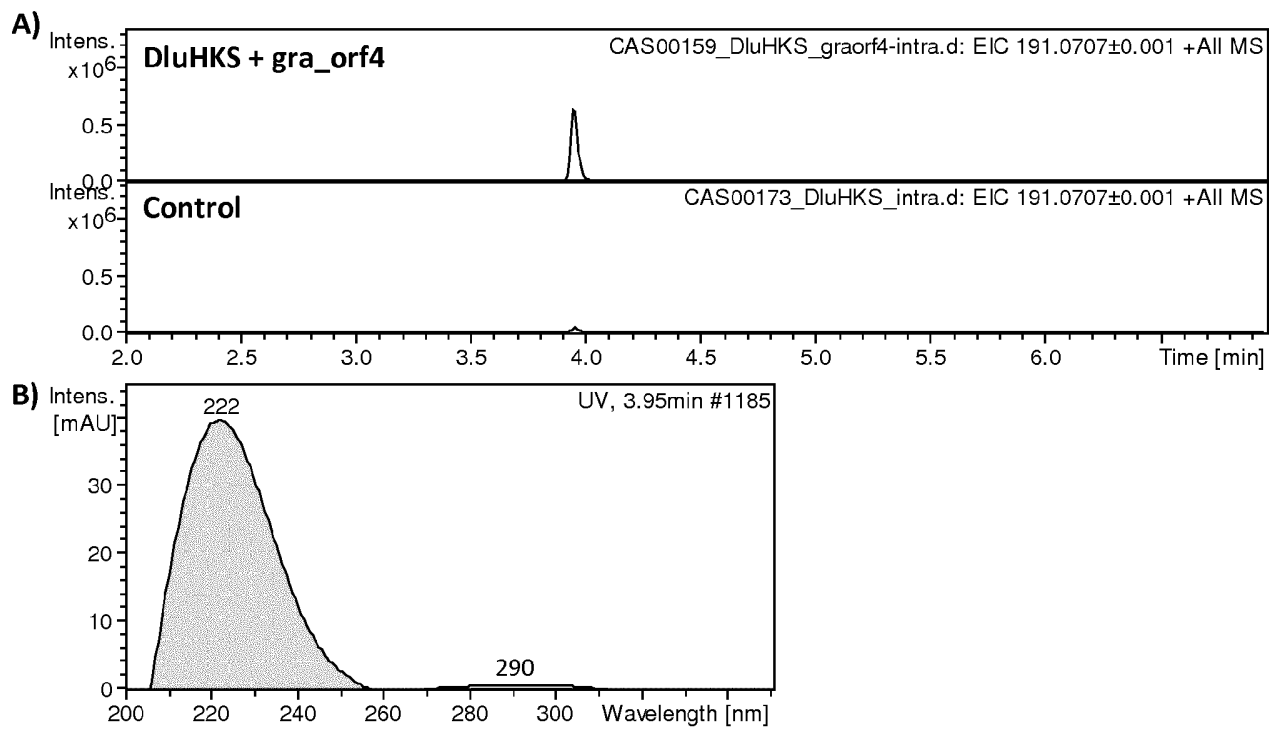


Figure 7

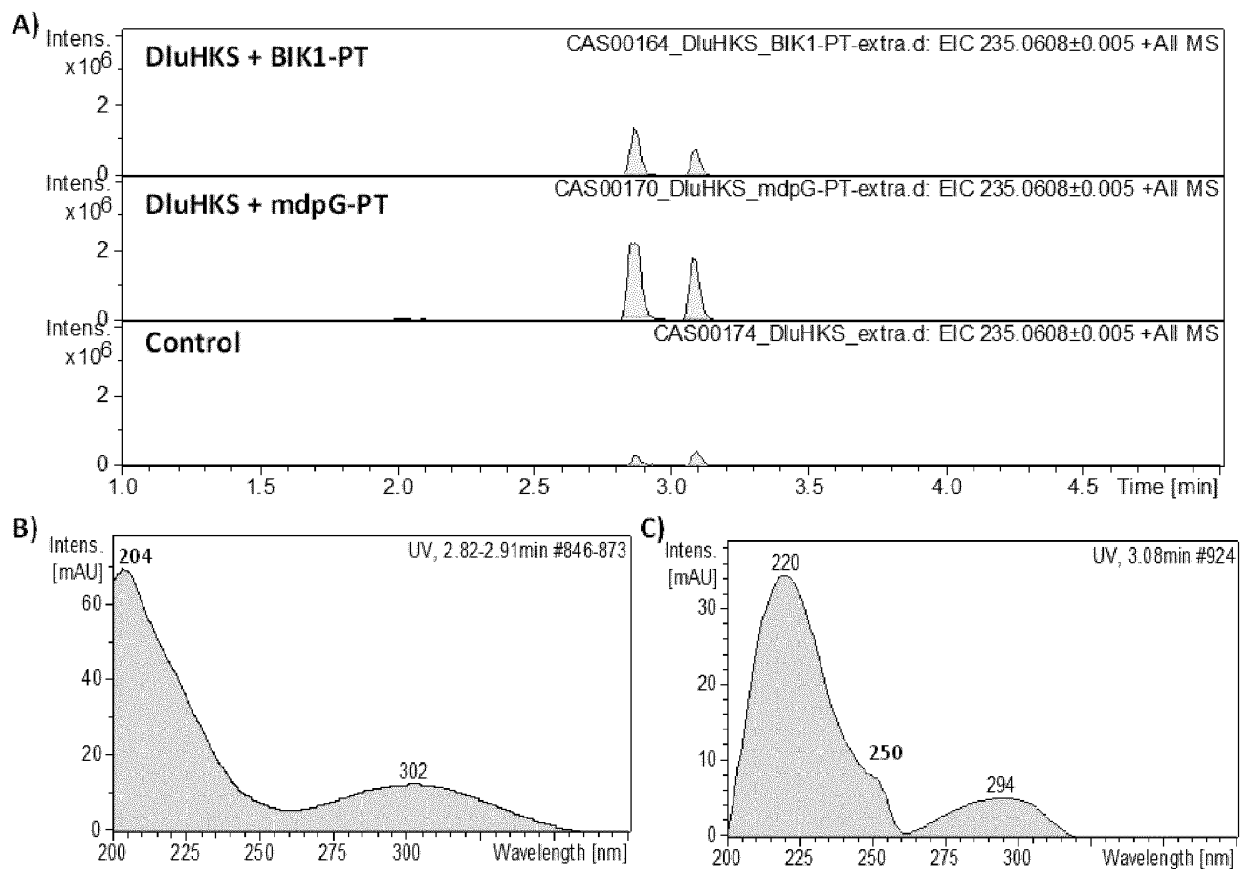


Figure 8

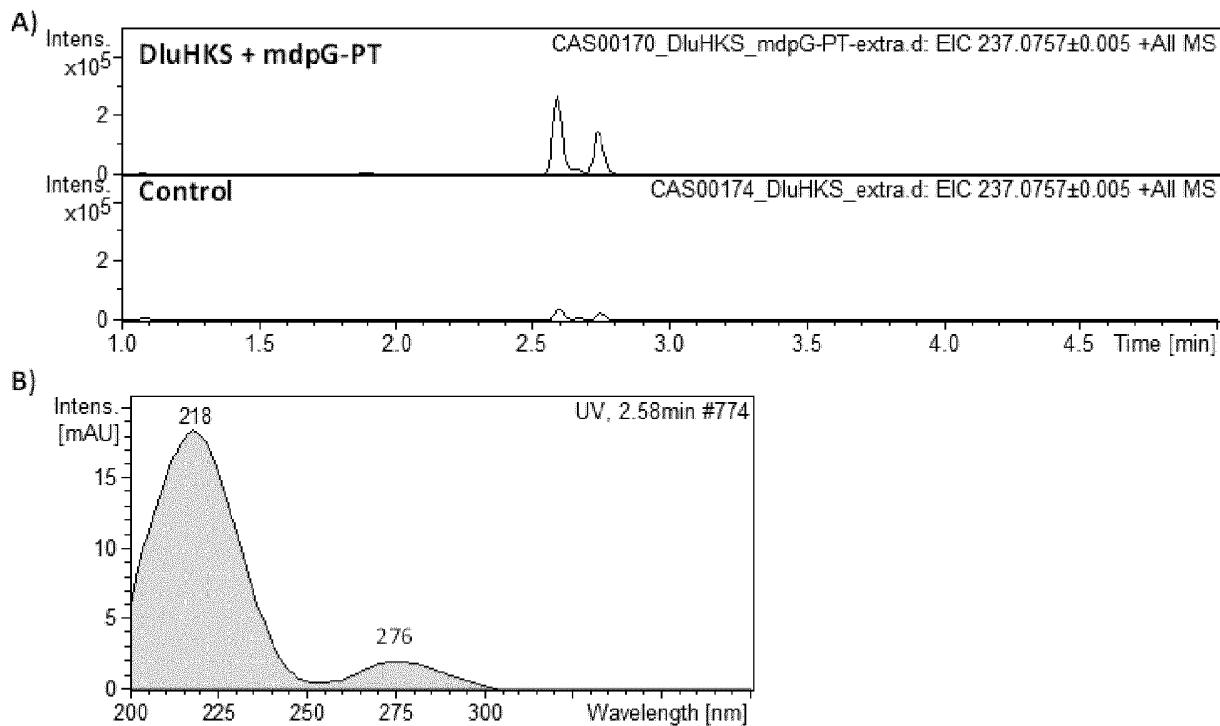


Figure 9

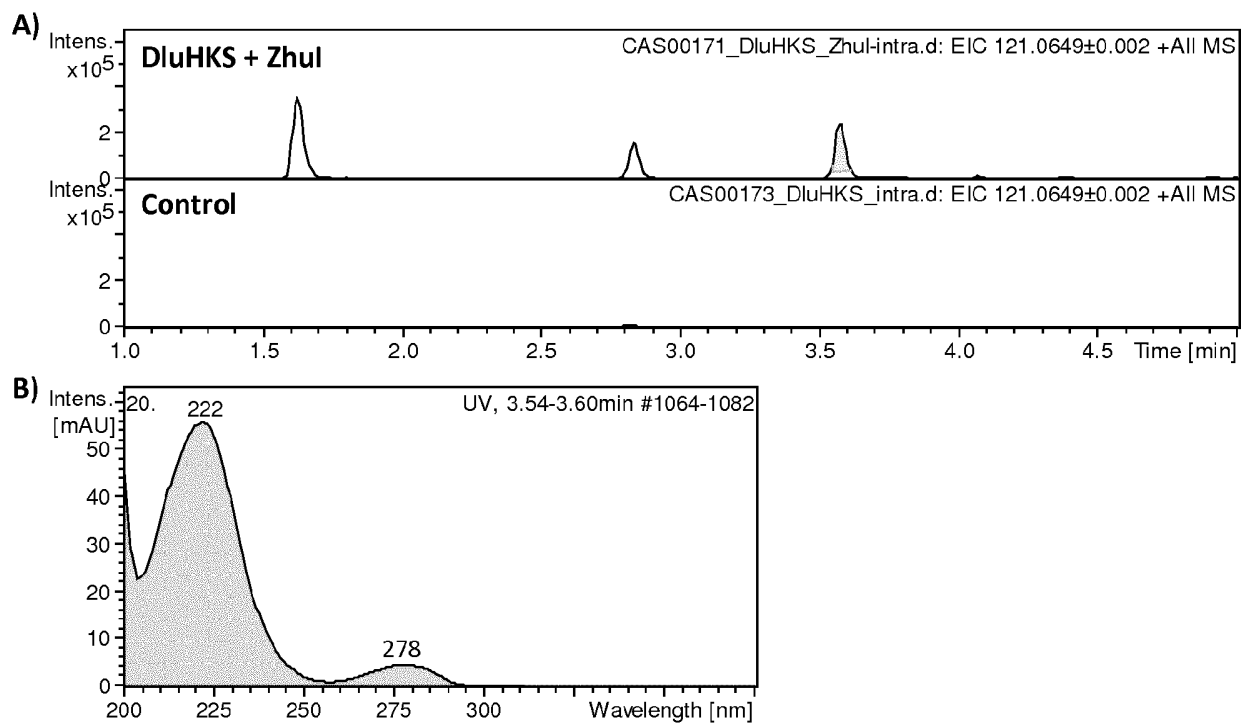


Figure 10

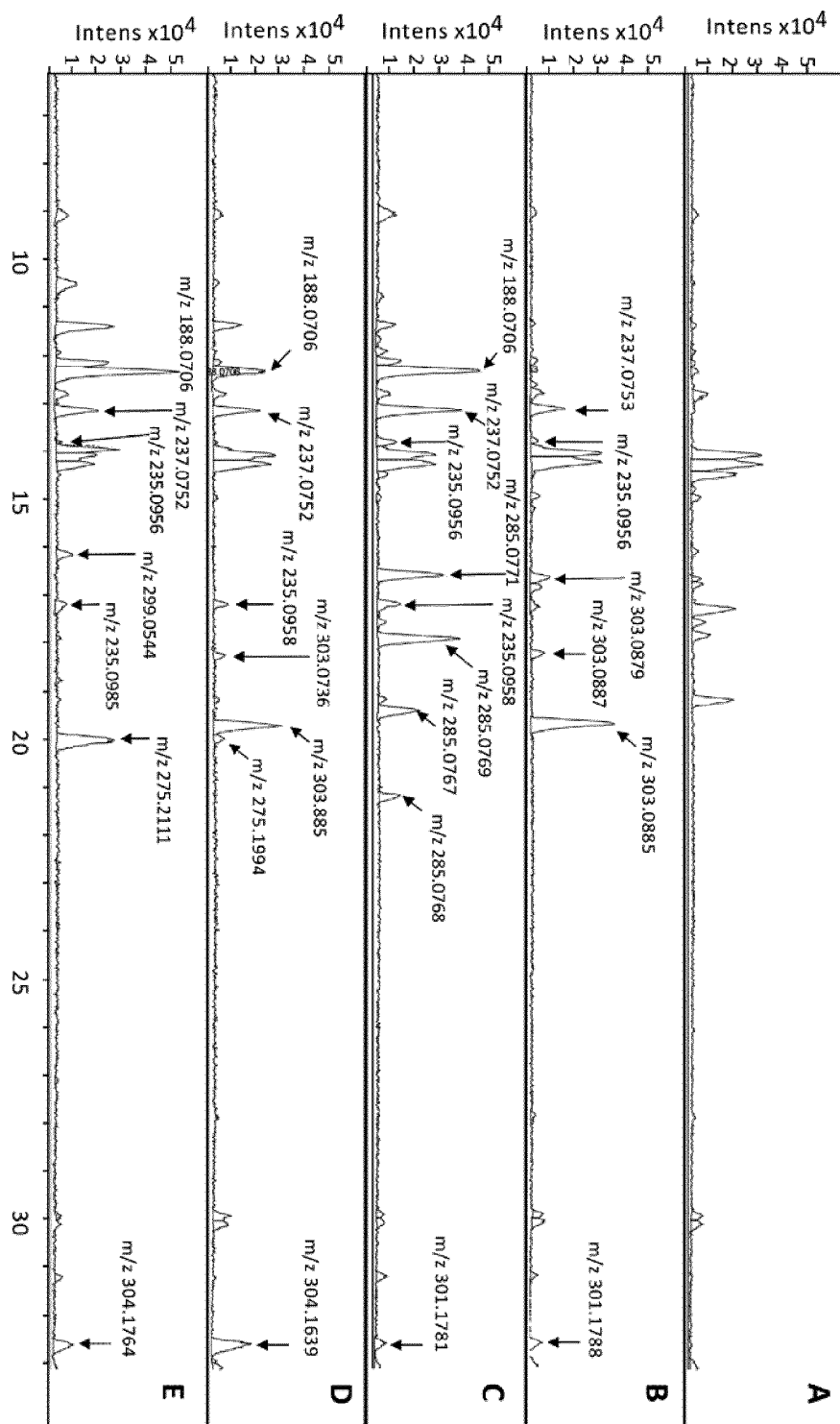
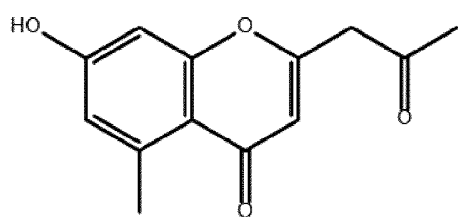
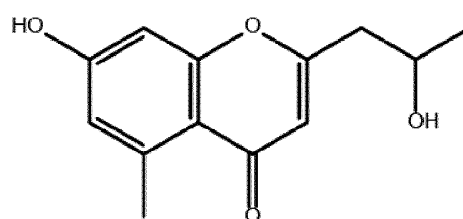


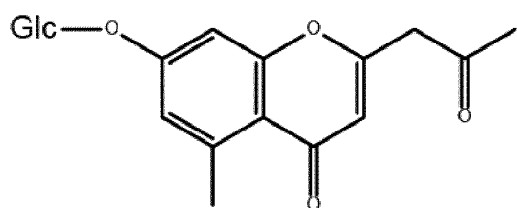
Figure 11



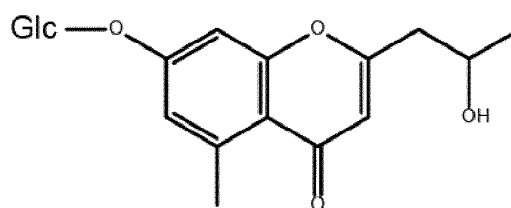
Aloesone



Aloesol



Aloesone-O-glc



Aloesol-O-glc



# INTERNATIONAL SEARCH REPORT

International application No PCT/EP2016/063331
---

<b>A. CLASSIFICATION OF SUBJECT MATTER</b>				
INV. C12P7/26	C12P7/40	C12P15/00		
C12N1/15	C12N1/19	C12N15/82		
ADD.				
According to International Patent Classification (IPC) or to both national classification and IPC				
<b>B. FIELDS SEARCHED</b>				
Minimum documentation searched (classification system followed by classification symbols) C12P C12N				
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, FSTA, BIOSIS, EMBASE, COMPENDEX				
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>				
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.		
Y	<p>GAGNE ET AL: "Identification of olivetolic acid cyclase from Cannabis sativa reveals a unique catalytic route to plant polyketides", PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, U.S.A., vol. 109, no. 31, 31 July 2012 (2012-07-31), pages 12811-12816, XP002747517, * See page 12811 (Abstract/type III PKS + DABB = pfam07876), page 12812 (Figure 1), and page 12813 (Table 1) *</p> <p style="text-align: center;">----- -/--</p>	1-4,6-9, 14,15		
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <span style="margin-left: 200px;"><input type="checkbox"/> See patent family annex.</span>				
<p>* Special categories of cited documents :</p> <table style="width: 100%; border: none;"> <tr> <td style="width: 50%; border: none; vertical-align: top;"> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier application or patent but published on or after the international filing date</p> <p>"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)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> </td> <td style="width: 50%; border: none; vertical-align: top;"> <p>"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</p> <p>"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</p> <p>"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</p> <p>"&amp;" document member of the same patent family</p> </td> </tr> </table>			<p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier application or patent but published on or after the international filing date</p> <p>"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)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p>	<p>"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</p> <p>"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</p> <p>"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</p> <p>"&amp;" document member of the same patent family</p>
<p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier application or patent but published on or after the international filing date</p> <p>"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)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p>	<p>"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</p> <p>"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</p> <p>"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</p> <p>"&amp;" document member of the same patent family</p>			
Date of the actual completion of the international search	Date of mailing of the international search report			
20 July 2016	01/08/2016			
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer  Korsner, Sven-Erik			

## INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2016/063331

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>AMES ET AL: "Structural and biochemical characterization of ZhuI aromatase/cyclase from the R1128 polyketide pathway", BIOCHEMISTRY, vol. 50, 26 August 2011 (2011-08-26), pages 8392-8406, XP002747579, * See page 8393 (Figure 1) *</p> <p style="text-align: center;">-----</p>	1-4,6-9, 14,15
Y	<p>WENJUN ZHANG ET AL: "Investigation of early tailoring reactions in the oxytetracycline biosynthetic pathway", THE JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 282, 31 August 2007 (2007-08-31), pages 25717-25725, XP002500433, * See page 25717 (Abstract/OxyK) *</p> <p style="text-align: center;">-----</p>	1-4,6-9, 14,15
Y	<p>ZHOU ET AL: "Cyclization of aromatic polyketides from bacteria and fungi", NATURAL PRODUCT REPORTS, vol. 27, 2014, pages 839-868, XP002747553, * See pages 841-843(sections 2.1.1-2.1.2/cyclase) as well as the compounds 155 and 169/ compare Figure 1 of the Application *</p> <p style="text-align: center;">-----</p>	1-4,6-9, 14,15
Y	<p>LE ZHANG ET AL: "Synthesis of unnatural small molecules by plant specific polyketide synthases", CHINESE JOURNAL OF ORGANIC CHEMISTRY, vol. 33, 2013, pages 2469-2484, XP002747593, * See page 2472 (section 1.2), page 2478 (Scheme 11/HpPKS2), and page 2482 (Figure 2/library) *</p> <p style="text-align: center;">-----</p>	1-4,6-9, 14,15
A	<p>JADHAV ET AL: "Polyketide synthesis in tobacco plants transformed with a Plumbago zeylanica type III hexaketide synthase", PHYTOCHEMISTRY, vol. 98, 2014, pages 92-100, XP002747554, * See page 92 (Abstract/tobacco-&gt;nicotiana) *</p> <p style="text-align: center;">-----</p>	13
A	<p>KHO GO ET AL: "Synthetic polyketide enzymology: platform for biosynthesis of antimicrobial polyketides", ACS CATALYSIS, vol. 5, 27 May 2015 (2015-05-27), pages 4033-4042, XP008177920, * See page 4033 (Abstract/library) *</p> <p style="text-align: center;">-----</p> <p style="text-align: center;">-/--</p>	2,14

## INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2016/063331

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	HARTUNG JÖRGENSEN: "Fusarium graminearum PKS14 is involved in orsellinic acid and orcinol synthesis", FUNGAL GENETICS AND BIOLOGY, vol. 70, 8 July 2014 (2014-07-08), pages 24-31, XP002747518, * See page 26 (Figure 1B) * -----	1-15
A	HASHIMOTO ET AL: "Fungal type III polyketide synthases", NATURAL PRODUCT REPORTS, vol. 31, 2014, pages 1306-1317, XP002747580, * See page 1309 (Scheme 1) * -----	1-15
A	DAYU YU ET AL: "Type III polyketide synthases in natural product biosynthesis", IUBMB LIFE, vol. 64, 2012, pages 285-295, XP002747467, * Overview of type III PKS * -----	1-15
A	IMAN A.M. ABDEL-RAHMAN ET AL: "In vitro formation of the anthranoid scaffold by cell-free extracts from yeast-extract-treated <i>Cassia bicapsularis</i> cell cultures", PHYTOCHEMISTRY, vol. 88, 8 February 2013 (2013-02-08), pages 15-24, XP002747594, * See page 21 (the paragraph bridging the two columns) * -----	1-15
A	SARAH ELIZABETH POSETH: "Cyclization modes in type III polyketide synthases and the synthetic compounds used as probes for mechanistic studies", UNIVERSITY OF REGINA, SASKATCHEWAN THESIS, 2012, pages Cover page, i-xii-1-88, XP002747581, Retrieved from the Internet: URL: <a href="http://ourspace.uregina.ca:8080/xmlui/bitstream/handle/10294/3819/Posehn_Sarah_200231265_MSC_BIOC_Spring2013.pdf?sequence=1">http://ourspace.uregina.ca:8080/xmlui/bitstream/handle/10294/3819/Posehn_Sarah_200231265_MSC_BIOC_Spring2013.pdf?sequence=1</a> 1 [retrieved on 2015-10-20] * See page 5 (Figure 2) and page 6 (Figure 3) * ----- -/--	1-15

## INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2016/063331

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>VAGSTAD ET AL: "Characterization of a fungal thioesterase having Claisen cyclase and deacetylase activities in melanin biosynthesis",            CHEMISTRY &amp; BIOLOGY,            vol. 19, 2012, pages 1525-1534,            XP002760112,            * See page 1525 (Abstract) and page 1528 (ATHN/compare Figure 4 of the Application)*</p> <p style="text-align: center;">-----</p>	1-15
A	<p>FANG YANG ET AL: "Biosynthesis of phloroglucinol compounds in microorganisms",            APPLIED MICROBIOLOGY AND BIOTECHNOLOGY,            vol. 93, 2012, pages 487-495, XP035001278,            * See page 488 (Figure 1a-b/compare Figure 1 of the Application) *</p> <p style="text-align: center;">-----</p>	1-15
A	<p>KARPPINEN ET AL: "Octaketide-producing type III polyketide synthase from Hypericum perforatum is expressed in dark glands accumulating hypericins",            THE FEBS JOURNAL,            vol. 275, 2008, pages 4329-4342,            XP002752481,            * See page 4333 (Figure 4/B6) *</p> <p style="text-align: center;">-----</p>	1-15