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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
9 September 2016 (09.09.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/141207 A1

(51) International Patent Classification:
C12P 19/60 (2006.01) *C12N 9/10* (2006.01)
C12N 9/02 (2006.01) *C12P 17/16* (2006.01)

(21) International Application Number:
PCT/US2016/020729

(22) International Filing Date:
3 March 2016 (03.03.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/127,778 3 March 2015 (03.03.2015) US

(71) Applicant: **THE REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 1111 Franklin Street, 12th Floor, Oakland, CA 94607-5200 (US).

(72) Inventors: **DUEBER, John, Eugene**; C/o UC Berkeley, 2151 Berkeley Way, Room 547, Berkeley, CA 94720 (US). **RUSS, Zachary, Nicholas**; C/o UC Berkeley, 2151 Berkeley Way, Room 547, Berkeley, CA 94720 (US). **HSU, Tammy, Melody**; C/o UC Berkeley, 2151 Berkeley Way, Room 547, Berkeley, CA 94720 (US). **JOHNSON, Terry, Don, Jr.**; C/o UC Berkeley, 2151 Berkeley Way, Room 547, Berkeley, CA 94720 (US). **CERVANTES, Bernardo**; C/o UC Berkeley, 2151 Berkeley Way, Room 547, Berkeley, CA 94720 (US). **PRATHURI, Ramya, Lakshmi**; 1172 Scotland Drive, Cupertino, CA 95014 (US). **BHAKTA, Shyam, Pravin**; C/o UC Berkeley, 2151

[Continued on next page]

(54) Title: PROTECTING GROUP CHEMISTRY FOR CLEAN, REDUCTANT-FREE DYEING

(57) Abstract: The present disclosure relates to the biosynthesis of indigoid dye precursors and their conversion to indigoid dyes. Specifically, the present disclosure relates to methods of using polypeptides to produce indigoid dye precursors from indole feed compounds, and the use of the indigoid dye precursors to produce indigoid dyes.

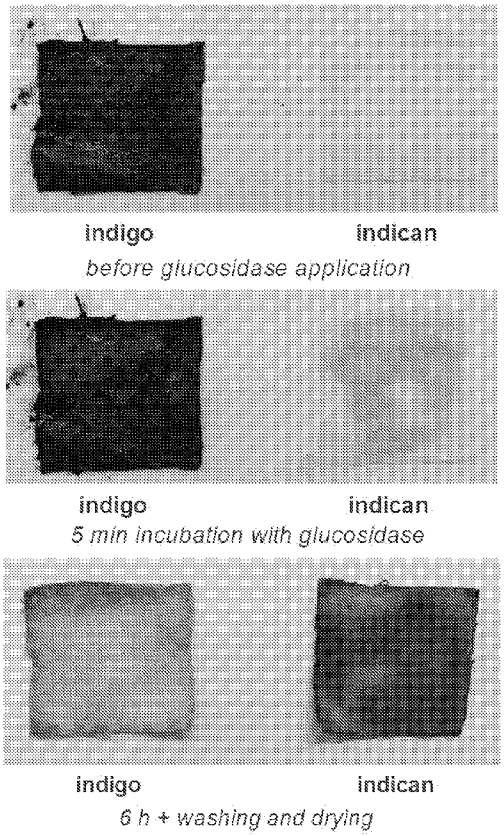


FIG. 5

WO 2016/141207 A1

Berkeley Way, Room 547, Berkeley, CA 94720 (US). **FONG, Arthur, Muir, III**; C/o UC Berkeley, 2151 Berkeley Way, Room 547, Berkeley, CA 94720 (US). **LATIMER, Luke, Nathaniel**; C/o UC Berkeley, 2151 Berkeley Way, Room 547, Berkeley, CA 94720 (US).

(74) **Agents**: **WARD, Michael, R.** et al.; Morrison & Foerster LLP, 425 Market Street, San Francisco, CA 94105-2482 (US).

(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, 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))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

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

PROTECTING GROUP CHEMISTRY FOR CLEAN, REDUCTANT-FREE DYEING

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Prov. App. No. 62/127,778, filed on March 3, 2015, where is hereby incorporated by reference in its entirety.

FIELD

[0002] The present disclosure relates generally to the production of dye precursors, and more specifically to the use of polypeptides to produce indigoid dye precursors.

BACKGROUND

[0003] The blue dye indigo, produced by some plants, is one of the oldest dyes in the world. It is still a widely used textile dye, particularly in the denim clothing industry. Modern indigo is no longer sourced from plants, but is produced synthetically, mostly from petroleum. Indigo itself is a water-insoluble compound. In order to dye denim, indigo must be treated with a reducing agent to produce an unstable, water-soluble intermediate. After application of this intermediate to the fabric, the fabric is exposed to air and the intermediate oxidizes back to indigo, crystallizing within the fabric fibers. The most common reducing agent used for this process is sodium dithionite, which has many limitations. Sodium dithionite is unstable; can over-reduce the indigo, destroying the dye; is required in excess quantities; and its use produces large amounts of sulfate and sulfite which are detrimental to the environment. Because of these shortcomings, substantial amounts of sodium dithionite are lost to byproducts and degradation during the dyeing process.

[0004] The biosynthesis of natural indigo in plants proceeds through a different synthetic pathway. An indigo precursor compound is produced and stored in the plant, then converted to indoxyl, which spontaneously dimerizes to form indigo. However, the quantity and variety of indigo precursors that can be produced through current natural biological pathways is limited.

[0005] Thus, there exists a need for new methods to produce indigoid dye precursors with greater structural variety and which can be used to produce indigoid dyes without a harsh reductant chemical.

BRIEF SUMMARY

[0006] In one aspect, the present disclosure relates to methods of producing indigoid dye precursors from indole feed compounds. In some aspects, the present disclosure relates to the use of polypeptides to produce indigoid dye precursors from indole feed compounds.

[0007] In one aspect, the present disclosure relates to a method of producing a compound of Formula (A), the method comprising:

a) contacting a host cell with a compound of Formula (II); and

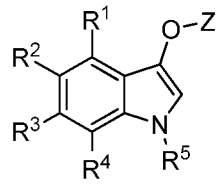
b) culturing the host cell under conditions such that a compound of Formula (A) is produced from at least a portion of the compound of Formula (II);

wherein the host cell comprises a single or multiple recombinant nucleic acid(s) encoding:

i) a polypeptide with oxygenase activity; and

ii) a polypeptide with glycosyltransferase activity or a polypeptide with sulfotransferase activity, or a combination thereof;

wherein the compound of Formula (A) is:



(A),

wherein:

R^1, R^2, R^3, and R^4 are independently selected from H, halo, nitro, sulfate, phosphate, hydroxyl, $–C(O)OR^b$, $–R^aC(O)OR^b$, $–R^aOC(O)R^b$, $–OR^a$, $–NR^bR^c$, and alkyl;

wherein alkyl may be unsubstituted or substituted with one or more substituents independently selected from the group consisting of halo, nitro, sulfate, phosphate, hydroxyl, $–C(O)OR^b$, $–R^aC(O)OR^b$, $–R^aOC(O)R^b$, $–OR^a$, $–NR^bR^c$, and $–R^aOR^d$;

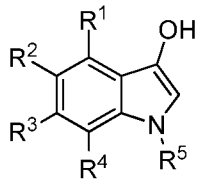
R^5 is H or alkyl, wherein alkyl is unsubstituted or substituted with one or more substituents independently selected from the group consisting of halo, hydroxyl, =O, $-C(O)OR^b$, $-R^aC(O)OR^b$, $-R^aOC(O)R^b$, $-OR^a$, $-NR^bR^c$, and $-R^aOR^d$;

Z is a glycone, $-C(O)R^b$, or $-SO_3^-$;

R^a and R^d are independently alkyl;

R^b and R^c are independently H or alkyl; and

wherein the compound of Formula (II) is:



(II),

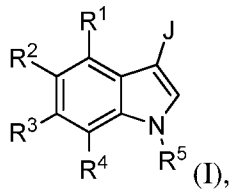
wherein R^1, R^2, R^3, R^4, and R^5 are defined as for Formula (A).

[0008] In some embodiments, the host cell is a bacterial cell. In some embodiments the host cell is selected from *Escherichia coli* and *Corynebacterium glutamicum*. In other embodiments, the host cell is a yeast. In certain embodiments, the host cell is selected from *Saccharomyces cerevisiae*, *Pichia pastoris*, *Kluyveromyces lactis, Kluyveromyces marxianus*, and *Schizosaccharomyces pombe*.

[0009] In some embodiments, the polypeptide with glycosyltransferase activity comprises SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, or SEQ ID NO: 20, or the amino acid sequence of any homologs thereof. In some embodiments, the polypeptide with sulfotransferase activity comprises SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, or SEQ ID NO: 14, or the amino acid sequence of any homologs thereof. In some embodiments, the homolog of a polypeptide with oxygenase activity comprises an amino acid sequence at least 90%, at least 95%, or at least 99% identical to an amino acid sequence selected from SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, and SEQ ID NO: 24. In some embodiments, the homolog of a

polypeptide with glycosyltransferase comprises an amino acid sequence at least 90%, at least 95%, or at least 99% identical to an amino acid sequence selected from SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, and SEQ ID NO: 20. In some embodiments, the homolog of a polypeptide with sulfotransferase comprises an amino acid sequence at least 90%, at least 95%, or at least 99% identical to an amino acid sequence selected from SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, and SEQ ID NO: 14.

[0010] In another aspect, the present disclosure relates to a method of producing a compound of Formula (A) from a compound of Formula (II), further comprising providing a compound of Formula (I):



(I),

wherein:

R^1, R^2, R^3, R^4, and R^5 are defined as for Formula (A);

J is alkyl, $-OC(O)R^b$, or phosphate;

wherein alkyl is unsubstituted or substituted with one or more substituents independently selected from the group consisting of halo, hydroxyl, =O, $-C(O)OR^b$, $-R^aC(O)OR^b$, $-R^aOC(O)R^b$, $-OR^a$, $-NR^bR^c$, and $-R^aOR^d$; and

R^a, R^b, R^c, and R^d are defined as for Formula (A).

[0011] In some embodiments, the titer of the compound of Formula (A) produced according to the methods herein is at least 50 mg/L, 100 mg/L, 1 g/L, 10 g/L, 25 g/L, 50 g/L, 75 g/L, 100 g/L, 125 g/L, 150 g/L, 175 g/L, or 200 g/L.

[0012] In some embodiments, the compound of Formula (A) is secreted by the host cell. In certain embodiments, the methods disclosed herein further comprise isolating the compound of Formula (A) produced.

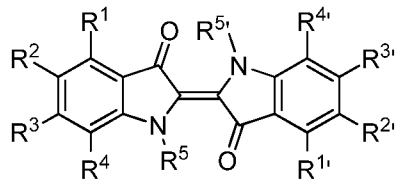
[0013] In yet other embodiments, the methods disclosed herein further comprise converting the compound of Formula (A) to a compound of Formula (II).

[0014] In another aspect, disclosed herein is a method of producing a compound of Formula (B), the method comprising converting a first compound of Formula (A) and a second compound of Formula (A) to a compound of Formula (B);

wherein the first compound of Formula (A) and the second compound of Formula (A) are the same compound of Formula (A) or different compounds of Formula (A);

at least one of the first compound of Formula (A) and the second compound of Formula (A) is produced according to the methods described herein; and

the compound of Formula (B) is:



(B),

wherein R^1, R^2, R^3, R^4, R^5, $R^{1'}$, $R^{2'}$, $R^{3'}$, $R^{4'}$, and $R^{5'}$ are defined as for Formula (A).

[0015] In some embodiments, the compound of Formula (A) is converted to a compound of Formula (II); and the compound of Formula (II) is converted to the compound of Formula (B). In certain embodiments, the compound of Formula (A) is contacted by a hydrolase to convert the compound of Formula (A) to the compound of Formula (II). In some embodiments, the hydrolase is a glucosidase, while in other embodiments, the hydrolase is a sulfatase.

[0016] In some embodiments of the methods described herein, the polypeptide having oxygenase activity and the polypeptide having glycosyltransferase activity are the same

polypeptide. In other embodiments, the polypeptide having oxygenase activity and the polypeptide having sulfotransferase activity are the same polypeptide.

[0017] In another aspect, disclosed herein is a host cell comprising single or multiple recombinant nucleic acid(s) encoding a polypeptide with monooxygenase activity and a polypeptide with glucosyltransferase activity.

[0018] In yet another aspect, disclosed herein is a host cell comprises a single or multiple recombinant nucleic acid(s) encoding a polypeptide with monooxygenase activity and a polypeptide with sulfotransferase activity.

[0019] In a further aspect, the present disclosure relates to a method of producing a compound of Formula (A), the method comprising contacting a compound of Formula (II) with i) a polypeptide with oxygenase activity; and ii) a polypeptide with glycosyltransferase activity or a polypeptide with sulfotransferase activity, or a combination thereof; and producing a compound of Formula (A) from at least a portion of the compound of Formula (II).

[0020] In some embodiments that may be combined with any of the preceding embodiments, the compound of Formula (I) is 1*H*-indol-3-yl and the compound of Formula (A) is 3-(β-D-glucosido)indole. In other embodiments, the compound of Formula (A) is 1*H*-indol-3-yl sulfate.

[0021] In certain embodiments, the present disclosure relates to cDNA encoding the polypeptides described herein, such as cDNAs encoding SEQ ID NOs: 1-29.

DESCRIPTION OF THE FIGURES

[0022] The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

[0023] **FIG. 1** depicts the high performance liquid chromatography/mass spectrometry (HPLC/MS) trace of 3-(β-D-glucosido)indole (*i.e.*, indican) produced by *E. coli* host cells heterologously expressing different recombinant glucosyltransferases.

[0024] **FIG. 2** depicts an image of *E. coli* host cell cultures heterologously expressing different recombinant glucosyltransferases and a heterologous flavin-containing monooxygenase (FMO).

[0025] **FIG. 3** depicts images of an *E. coli* host cell culture heterologously expressing a recombinant glucosyltransferase and a heterologous oxygenase FMO, immediately after the addition of β-glucosidase (left), 75 min after the addition of β-glucosidase (middle), and 20.5 h after the addition of β-glucosidase (right).

[0026] **FIG. 4** depicts a graph comparing the concentration of indican of each *E. coli* host cell culture expressing a different glucosyltransferase and an FMO, and additionally with and without expression of UDP-glucose synthesis enzymes ("UDP-Glc"), and the oligosaccharide transporter CDT1.

[0027] **FIG. 5** depicts the appearance of cloth pieces after incubation with indigo or indican (top), 5 min after application of β-glucosidase (middle), and after six hours of incubation followed by washing and drying (bottom).

[0028] **FIG. 6** depicts images of the conversion of indigoid dye precursors to indigoid dyes in the presence of the sulfatase atsA, from *P. aeruginosa*.

[0029] **FIG. 7** depicts a graph of the concentration of 1*H*-indol-3-yl sulfate produced by *E. coli* host cell cultures heterologously expressing different sulfotransferases.

[0030] **FIG. 8** depicts a photograph of different dyes produced from sulfated dye precursors in the presence of different hydrolases.

[0031] **FIG. 9** depicts halogenated indigoid dyes produced by contacting substituted haloindoles with an *E. coli* host cell expressing a heterologous oxygenase FMO.

[0032] **FIG. 10** depicts the production of indigo in *E. coli* host cells expressing a heterologous oxygenase FMO.

[0033] **FIG. 11** depicts the relative rate of indoxyl acetate hydrolysis by wild-type E. coli strain MG1655 and two knockout strains upon incubation with 1 mM indoxyl acetate.

Hydrolysis was calculated by an initial increase in indoxyl fluorescence when 1 mM indoxyl acetate is mixed with saturated culture (2 OD600 units per mL) in phosphate buffered saline pH 7 and 5% DMSO. **5xKO:** Δaes, ΔyjfP, ΔbioH, ΔyeiG, ΔfrmB. **11xKO**: Δaes, ΔyjfP, ΔbioH, ΔyeiG, ΔfrmB, ΔentH, ΔydiL, ΔtesA, ΔnanS, ΔyqiA, ΔybfF

DETAILED DESCRIPTION

[0034] The following description sets forth exemplary methods, parameters and the like. It should be recognized, however, that such description is not intended as a limitation on the scope of the present disclosure but is instead provided as a description of exemplary embodiments.

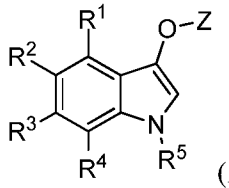
Methods for Producing Compounds of Formula (A)

[0035] The methods described herein provide methods of producing indigoid dye precursors by contacting a feed indole compound with a) a polypeptide with oxygenase activity, and b) a polypeptide with glycosyltransferase activity, a polypeptide with sulfotransferase activity, a polypeptide with acyltransferase activity, or a polypeptide with phosphotransferase activity, or a combination thereof. The polypeptide with oxygenase activity and the polypeptide with glycosyltransferase activity, sulfotransferase activity, acyltransferase activity, or phosphotransferase activity, or a combination thereof, may be the same polypeptide or separate polypeptides. In some embodiments, the feed indole compound is contacted by the polypeptides in a host cell, while in other embodiments, the feed indole compound is contacted by the polypeptides *in vitro*.

[0036] The indigoid dye precursors produced according to the methods described herein may be used to produce indigoid dyes, including, for example, indigo. Thus, in one aspect, provided herein are methods of producing indigoid dyes from the indigoid dye precursors by: a) contacting a feed indole compound with i) a polypeptide with oxygenase activity, and ii) a polypeptide with glycosyltransferase activity, a polypeptide with sulfotransferase activity, a polypeptide with acyltransferase activity, or a polypeptide with phosphotransferase activity, or a combination thereof to produce an indigoid dye precursor; then b) converting the indigoid dye precursor to an indigoid dye.

Compounds of Formula (A): Indigoid Dye Precursors

[0037] In one aspect, provided herein are methods of producing indigoid dye precursors by contacting a host cell with a feed indole compound under conditions such that the indigoid dye precursor is produced from at least a portion of the feed indole compound. In certain embodiments the indigoid dye precursor produced is a compound of Formula (A):



(A),

wherein:

R^1, R^2, R^3, and R^4 are independently selected from H, halo, nitro, sulfate, phosphate, hydroxyl, $-C(O)OR^b$, $-R^aC(O)OR^b$, $-R^aOC(O)R^b$, $-OR^a$, $-NR^bR^c$, and alkyl;

wherein R^a is alkyl; and

R^b and R^c are independently H or alkyl;

R^5 is H or alkyl;

Z is glycosyl, $-C(O)R^b$, $-SO_3^-$, or $-PO_3^{2-}$.

[0038] In some embodiments of Formula (A), the alkyl of R^1, R^2, R^3, and R^4 at each occurrence is independently unsubstituted or substituted with one or more substituents selected from halo, nitro, sulfate, phosphate, hydroxyl, $-C(O)OR^b$, $-R^aC(O)OR^b$, $-R^aOC(O)R^b$, $-OR^a$, $-NR^bR^c$, and $-R^aOR^d$; wherein R^a and R^d are independently alkyl, and R^b and R^c are independently H or alkyl. In certain embodiments, R^5 is independently alkyl, and the alkyl is unsubstituted or substituted with one or more substituents selected from halo, hydroxyl, =O, $-C(O)OR^b$, $-R^aC(O)OR^b$, $-R^aOC(O)R^b$, $-OR^a$, and $-NR^bR^c$; wherein R^a is independently alkyl, and R^b and R^c are independently H or alkyl.

[0039] In some variations of Formula (A), R^1, R^2, R^3, and R^4 are independently H, halo, nitro, hydroxyl, or alkyl. In certain variations, R^1, R^2, R^3, and R^4 are independently H, nitro, or

hydroxyl. In other variations, R^1, R^2, R^3, and R^4 are independently H or nitro. In yet other variations, R^1, R^2, R^3, and R^4 are independently H or hydroxyl.

[0040] In some variations of the compound of Formula (A), R^1, R^2, R^3, and R^4 are independently H, chloro, bromo, iodo, or fluoro. In some variations, R^1, R^2, R^3, and R^4 are independently H, chloro, or bromo. In other variations, R^1, R^2, R^3, and R^4 are independently H or chloro. In yet other variations, R^1, R^2, R^3, and R^4 are independently H or bromo.

[0041] In some variations, R^1, R^2, R^3, and R^4 are independently H or alkyl. In some embodiments, R^1, R^2, R^3, and R^4 are independently H or methyl, ethyl, propyl, butyl, or pentyl. In certain variations R^1, R^2, R^3, and R^4 are independently H or methyl.

[0042] In certain variations, the alkyl of R^1, R^2, R^3, and R^4 at each occurrence is independently unsubstituted or substituted. For example, in certain variations, at least one of R^1, R^2, R^3, and R^4 is alkyl, and the alkyl is substituted with halo. In other variations, the alkyl is substituted with hydroxyl. In yet other variations, the alkyl is substituted with $-NR^bR^c$. In other variations, the alkyl is substituted with $-NR^bR^c$ and hydroxyl.

[0043] "Alkyl" as used herein refers to refers to a linear or branched saturated hydrocarbon chain. Examples of alkyl groups include methyl, ethyl, *n*-propyl, *iso*-propyl, *n*-butyl, *sec*-butyl, *tert*-butyl, *n*-pentyl, 2-pentyl, *iso*-pentyl, *neo*-pentyl, hexyl, 2-hexyl, 3-hexyl, and 3-methylpentyl. When an alkyl residue having a specific number of carbons is named, all geometric isomers having that number of carbons may be encompassed; thus, for example, "butyl" can include *n*-butyl, *sec*-butyl, *iso*-butyl and *tert*-butyl; "propyl" can include *n*-propyl and *iso*-propyl. In some embodiments, alkyl as used herein, such as in compounds of Formulae (A), (B), (I), and (II), has 1 to 30 carbon atoms (*i.e.*, C_{1-30} alkyl), 1 to 20 carbon atoms (*i.e.*, C_{1-20} alkyl), 1 to 15 carbon atoms (*i.e.*, C_{1-15} alkyl), 1 to 9 carbon atoms (*i.e.*, C_{1-9} alkyl), 1 to 8 carbon atoms (*i.e.*, C_{1-8} alkyl), 1 to 7 carbon atoms (*i.e.*, C_{1-7} alkyl), 1 to 6 carbon atoms (*i.e.*, C_{1-6} alkyl), 1 to 5 carbon atoms (*i.e.*, C_{1-5} alkyl), 1 to 4 carbon atoms (*i.e.*, C_{1-4} alkyl), 1 to 3 carbon atoms (*i.e.*, C_{1-3} alkyl), 1 to 2 carbon atoms (*i.e.*, C_{1-2} alkyl), 1 carbon atom (*i.e.*, C_1 alkyl), or 5 to 30 carbon atoms (*i.e.*, C_{5-30} alkyl), or 5 to 20 carbon atoms (*i.e.*, C_{5-20} alkyl).

[0044] R^1, R^2, R^3, and R^4 may be the same or different. In some embodiments of Formula (A), R^1, R^2, R^3, and R^4 are all H. In other embodiments, one of R^1, R^2, R^3, and R^4 is nitro; and the remaining R^1, R^2, R^3, and R^4 are H. In other embodiments, one of R^1, R^2, R^3, and R^4 is halo; and the remaining R^1, R^2, R^3, and R^4 are H. In yet other embodiments, two of R^1, R^2, R^3, and R^4 are independently halo; and the remaining R^1, R^2, R^3, and R^4 are H.

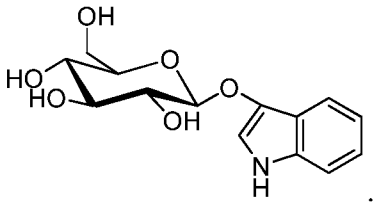
[0045] In certain variations of Formula (A), R^5 is H. In other variations of Formula (A), R^5 is unsubstituted alkyl. For example, in some variations, R^5 is unsubstituted methyl, ethyl, propyl, butyl, or pentyl. In certain variations, R^5 is methyl. In other variations, R^5 is substituted alkyl. For example, in some variations, R^5 is alkyl substituted with halo, hydroxyl, =O, $-C(O)OR^b$, $-R^aC(O)OR^b$, $-R^aOC(O)R^b$, $-OR^a$, or $-NR^bR^c$. In certain variations, R^5 is ethyl substituted with =O. In one variation, R^5 is $-C(O)CH_3$.

[0046] In some embodiments of the compound of Formula (A), Z is glycosyl. The term "glycosyl", as used herein, refers to a sugar residue group bonded from the anomeric carbon. The bond from the anomeric carbon may be in the α orientation or the β orientation. In some embodiments, the glycosyl is glucosyl, fructosyl, glucuronosyl, mannosyl, xylosyl, or galactosyl. In certain variations, the glycosyl is glucosyl in which the bond from the anomeric carbon is in the β orientation. In other variations, the glycosyl is glucosyl in which the bond from the anomeric carbon is in the α orientation. In certain variations, the glycosyl is glucuronosyl in which the bond from the anomeric carbon is in the β orientation. In yet other variations, the glycosyl is glucuronosyl in which the bond from the anomeric carbon is in the α orientation.

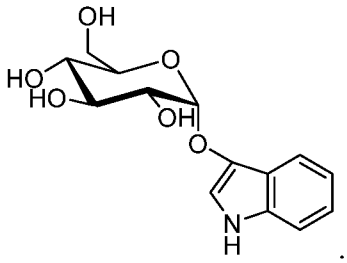
[0047] In other embodiments of the compound of Formula (A), Z is $-C(O)R^b$, wherein R^b is H or alkyl. In some embodiments, Z is $-C(O)R^b$, and R^b is H, methyl, ethyl, propyl, butyl, pentyl, hexyl, heptyl, or octyl. In other embodiments, Z is $-C(O)R^b$, and R^b is H, methyl, ethyl, propyl, or butyl. In certain embodiments, Z is $-C(O)R^b$, and R^b is methyl.

[0048] In some embodiments of the compound of Formula (A), Z is $-SO_3^-$ or $-PO_3^{2-}$.

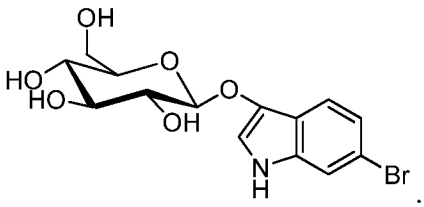
[0049] In one variation, R^1, R^2, R^3, R^4, and R^5 are all H; and Z is glucosyl in which the bond from the anomeric carbon is in the β orientation. Thus, in such a variation, the compound of Formula (A) is:



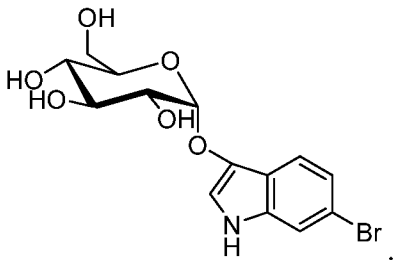
[0050] In one variation, R^1, R^2, R^3, R^4, and R^5 are all H; and Z is glucosyl in which the bond from the anomeric carbon is in the α orientation. Thus, in such a variation, the compound of Formula (A) is:



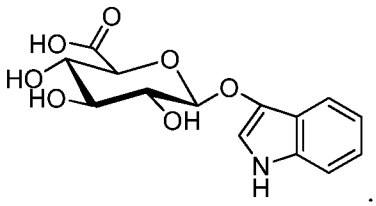
[0051] In another variation, R^1, R^2, R^4, and R^5 are all H; R^3 is bromo; and Z is glucosyl in which the bond from the anomeric carbon is in the β orientation. Thus, in such a variation, the compound of Formula (A) is:



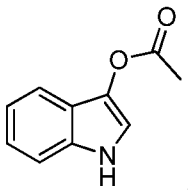
[0052] In another variation, R^1, R^2, R^4, and R^5 are all H; R^3 is bromo; and Z is glucosyl in which the bond from the anomeric carbon is in the α orientation. Thus, in such a variation, the compound of Formula (A) is:



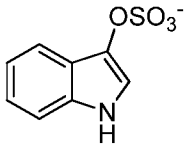
[0053] In yet another variation, R^1, R^2, R^3, R^4, and R^5 are all H; and Z is glucuronosyl in which the attachment from the anomeric carbon is in the β orientation. Thus, in such a variation, the compound of Formula (A) is:



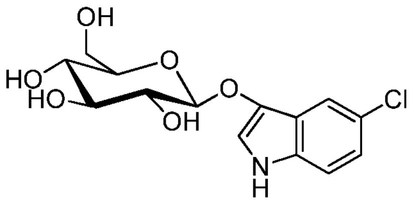
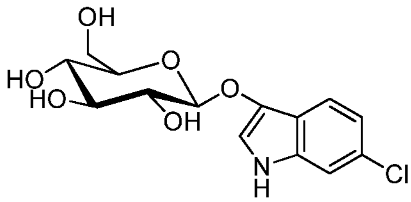
[0054] In another variation, R^1, R^2, R^3, R^4, and R^5 are all H; and Z is $-C(O)R^b$, wherein R^b is methyl. Thus, in such a variation, the compound of Formula (A) is:

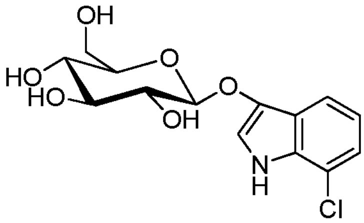
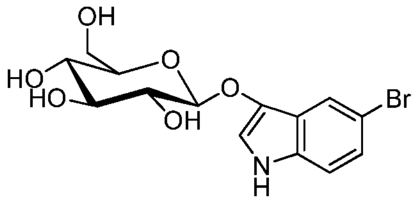


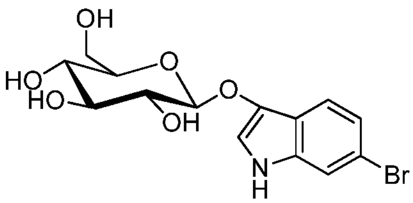
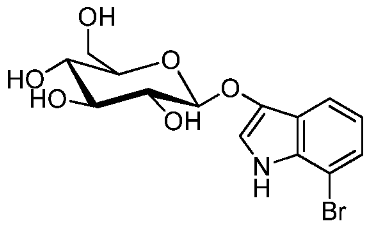
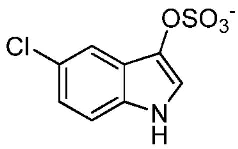
[0055] In another variation, R^1, R^2, R^3, R^4, and R^5 are all H; and Z is $-SO_3^-$. Thus, in such a variation, the compound of Formula (A) is:

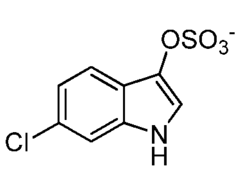
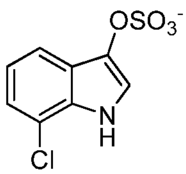
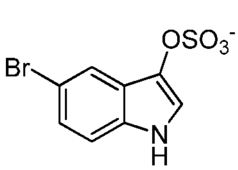
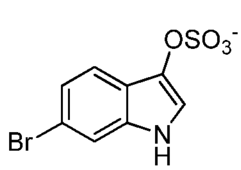
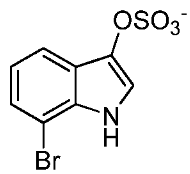


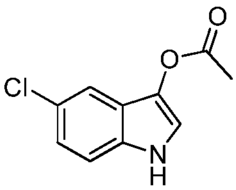
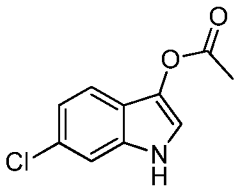
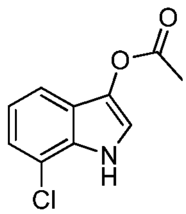
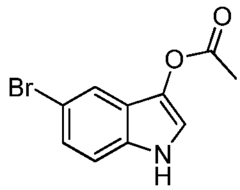
[0056] In other variations, the compound of Formula (A) is selected from:

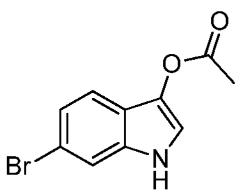
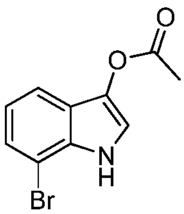
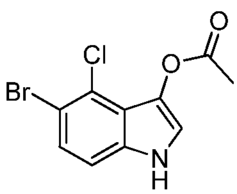
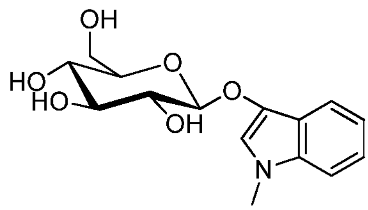
 ,  ,

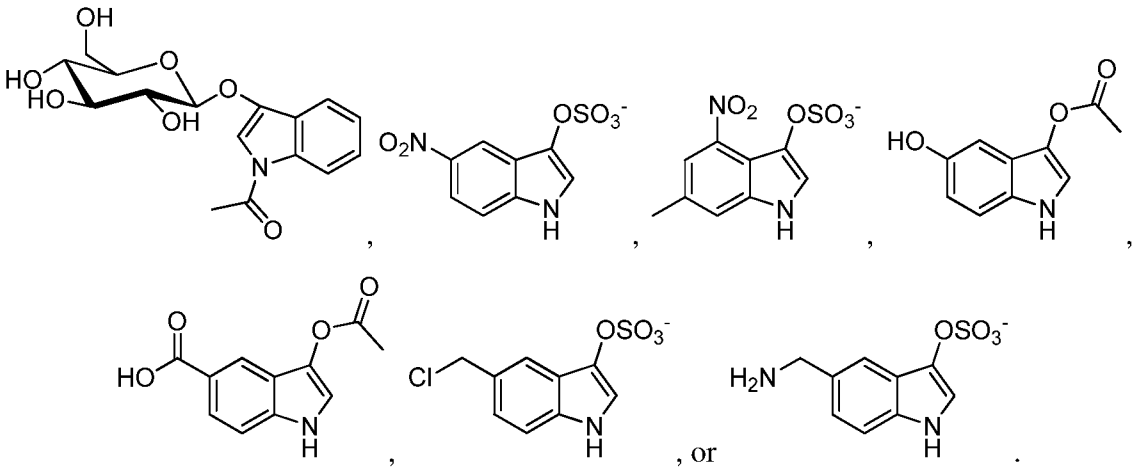
 ,  ,

 ,  ,  ,

 ,  ,  ,  ,  ,

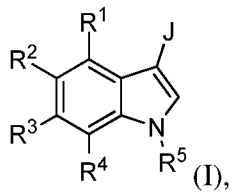
 ,  ,  ,  ,

 ,  ,  ,  ,



Compounds of Formula (I) and (II): Feed Indole Compounds

[0057] In some embodiments, the feed indole compound used in the methods described herein is a compound of Formula (I):



wherein:

R^1, R^2, R^3, and R^4 are independently selected from H, halo, nitro, sulfate, phosphate, hydroxyl, $-C(O)OR^b$, $-R^aC(O)OR^b$, $-R^aOC(O)R^b$, $-OR^a$, $-NR^bR^c$, and alkyl;

wherein R^a is alkyl; and

R^b and R^c are independently H or alkyl;

R^5 is H or alkyl; and

J is alkyl, $-OC(O)R^b$, hydroxyl, or phosphate.

[0058] In some embodiments of Formula (I), the alkyl of R^1, R^2, R^3, and R^4 at each occurrence is independently unsubstituted or substituted with one or more substituents selected from halo, nitro, sulfate, phosphate, hydroxyl, $-C(O)OR^b$, $-R^aC(O)OR^b$, $-R^aOC(O)R^b$, $-OR^a$, –

NR^bR^c, and $-R^aOR^d$; wherein R^a and R^d are independently alkyl, and R^b and R^c are independently H or alkyl. In certain embodiments, R^5 is independently alkyl, and the alkyl is unsubstituted or substituted with one or more substituents selected from halo, hydroxyl, =O, $-C(O)OR^b$, $-R^aC(O)OR^b$, $-R^aOC(O)R^b$, $-OR^a$, $-NR^bR^c$; wherein R^a is independently alkyl, and R^b and R^c are independently H or alkyl.

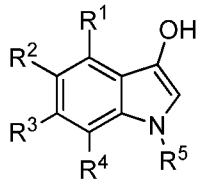
[0059] In some embodiments of Formula (I), J is alkyl, and the alkyl is unsubstituted or substituted with one or more substituents independently selected from the group consisting of halo, hydroxyl, =O, $-C(O)OR^b$, $-R^aC(O)OR^b$, $-R^aOC(O)R^b$, $-OR^a$, $-NR^bR^c$, and $-R^aOR^d$; wherein R^a and R^d are independently alkyl, and R^b and R^c are independently H or alkyl.

[0060] In some variations of Formula (I), J is alkyl. In certain variations, J is unsubstituted alkyl. For example, in certain variations, J is unsubstituted methyl, ethyl, propyl, butyl, pentyl, hexyl, heptyl, oxtyl, nonyl, or decyl. In other variations, J is alkyl substituted one or more groups independently selected from halo, hydroxyl, =O, $-C(O)OR^b$, $-R^aC(O)OR^b$, $-R^aOC(O)R^b$, and $-NR^bR^c$. In certain variations, J is alkyl substituted with $-C(O)OR^b$ and $-NR^bR^c$, wherein R^b and R^c are H. In other variations, J is alkyl substituted with $-NR^bR^c$ and $-R^aOC(O)R^b$. In yet other variations, J is alkyl substituted with halo, $-C(O)OR^b$ and $-NR^bR^c$, wherein R^b and R^c are H.

[0061] In other variations of Formula (I), J is phosphate.

[0062] In yet other variations of Formula (I), J is $-OC(O)R^b$, wherein R^b is H or alkyl. In certain variations, J is $-OC(O)R^b$, and R^b is H, methyl, ethyl, propyl, butyl, pentyl, hexyl, heptyl, octyl, nonyl, or decyl. In one embodiment, J is $-OC(O)R^b$, and R^b is methyl.

[0063] In other variations of Formula (I), J is hydroxyl, and the compound of Formula (I) is a compound of Formula (II):



(II),

wherein R^1, R^2, R^3, R^4, and R^5 are defined as for Formula (I).

[0064] It should generally be understood that variations of Formula (I) detailed throughout, where applicable, apply equally to Formula (II), the same as if each and every variation were specifically and individually listed for Formula (II).

[0065] It should also generally be understood that any of the variations for R^1, R^2, R^3, R^4 and R^5 as described herein for Formula (I) or Formula (II) may be combined the same as if each and every combination of the variables were specifically and individually listed.

[0066] In some variations of Formula (I) or Formula (II), R^1, R^2, R^3, and R^4 are independently H, halo, nitro, hydroxyl, or alkyl. In certain variations, R^1, R^2, R^3, and R^4 are independently H, nitro, or hydroxyl. In other variations, R^1, R^2, R^3, and R^4 are independently H or nitro. In yet other variations, R^1, R^2, R^3, and R^4 are independently H or hydroxyl.

[0067] In some variations of the compound of Formula (I) or Formula (II), R^1, R^2, R^3, and R^4 are independently H, chloro, bromo, iodo, or fluoro. In some variations, R^1, R^2, R^3, and R^4 are independently H, chloro, or bromo. In other variations, R^1, R^2, R^3, and R^4 are independently H or chloro. In yet other variations, R^1, R^2, R^3, and R^4 are independently H or bromo.

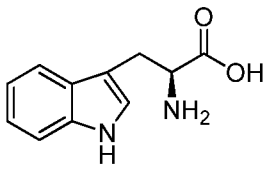
[0068] In some variations, R^1, R^2, R^3, and R^4 are independently H or alkyl. In some embodiments, R^1, R^2, R^3, and R^4 are independently H or methyl, ethyl, propyl, butyl, or pentyl. In certain variations R^1, R^2, R^3, and R^4 are independently H or methyl.

[0069] In certain variations, the alkyl of R^1, R^2, R^3, and R^4 at each occurrence is independently unsubstituted or substituted. For example, in certain variations, at least one of R^1, R^2, R^3, and R^4 is alkyl, and the alkyl is substituted with halo. In other variations, the alkyl is substituted with hydroxyl. In yet other variations, the alkyl is substituted with $-NR^bR^c$. In other variations, the alkyl is substituted with $-NR^bR^c$ and hydroxyl.

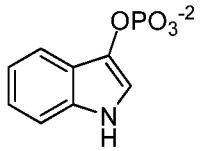
[0070] R^1, R^2, R^3, and R^4 may be the same or different. In some embodiments of Formula (I) or Formula (II), R^1, R^2, R^3, and R^4 are all H. In other embodiments, one of R^1, R^2, R^3, and R^4 is nitro; and the remaining R^1, R^2, R^3, and R^4 are H. In yet other embodiments, one of R^1, R^2, R^3, and R^4 is halo; and the remaining R^1, R^2, R^3, and R^4 are H.

[0071] In certain variations of Formula (I) or Formula (II), R^5 is H. In other variations of Formula (I) or Formula (II), R^5 is unsubstituted alkyl. For example, in some variations, R^5 is unsubstituted methyl, ethyl, propyl, butyl, or pentyl. In certain variations, R^5 is methyl. In other variations, R^5 is substituted alkyl. For example, in some variations, R^5 is alkyl substituted with one or more of halo, hydroxyl, =O, $-C(O)OR^b$, $-R^aC(O)OR^b$, $-R^aOC(O)R^b$, and $-NR^bR^c$. In certain variations, R^5 is ethyl substituted with =O. In one variation, R^5 is $-C(O)CH_3$.

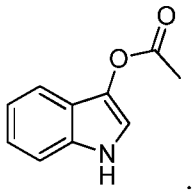
[0072] In some variations of the compound of Formula (I), R^1, R^2, R^3, R^4 and R^5 are H; J is ethyl substituted with $-C(O)OR^b$ and $-NR^bR^c$; R^b in each instance is H, and R^c is H. In one such variation, the compound of Formula (I) is:



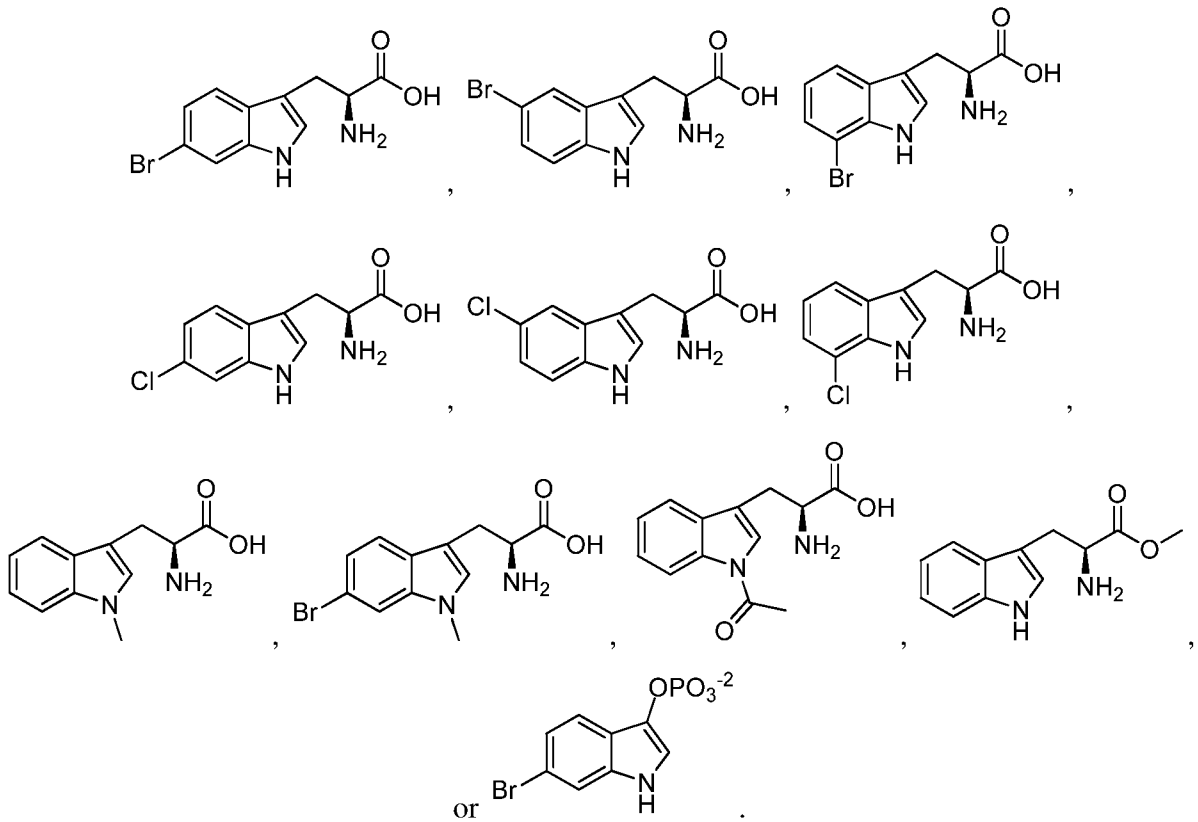
[0073] In another variation, R^1, R^2, R^3, R^4 and R^5 are H; and J is phosphate; and the compound of Formula (I) is:



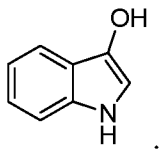
[0074] In yet another variation, R^1, R^2, R^3, R^4 and R^5 are H; J is $-OC(O)R^b$; R^b is methyl; and the compound of Formula (I) is:



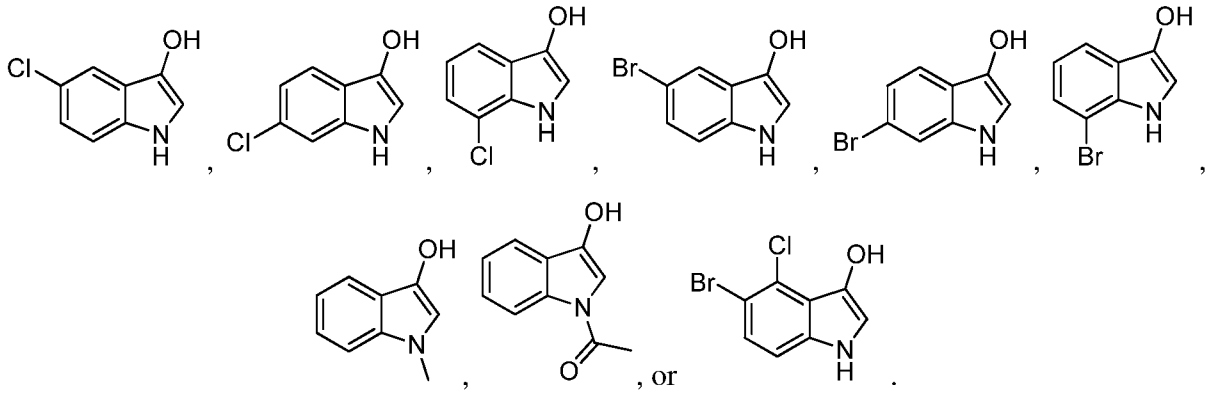
[0075] In yet other variations, the compound of Formula (I) is:



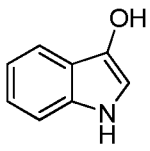
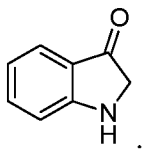
[0076] In one variation of the compound of Formula (II), R^1, R^2, R^3, R^4, and R^5 are all H. In such a variation, the compound of Formula (II) is:



[0077] In other variations, the compound of Formula (II) is:



[0078] One of skill in the art would recognize that compounds of Formula (II) may readily undergo keto-enol tautomerization to form the corresponding ketone. For example, in some embodiments, the compound of Formula (II):

 may tautomerize to form the corresponding ketone  .

[0079] One of skill in the art would recognize conditions that favor one tautomer over the other, including, for example, changes in temperature, the presence of water, and/or the presence of acid. The corresponding ketone tautomers of compounds of Formula (II) may be used to produce indigoid dyes. Thus, in one aspect, provided herein are methods of producing indigoid dyes from the indigoid dye precursors by: a) contacting a feed indole compound with i) a polypeptide with oxygenase activity, and ii) a polypeptide with glycosyltransferase activity, a polypeptide with sulfotransferase activity, a polypeptide with acyltransferase activity, or a combination thereof to produce an indigoid dye precursor; b) tautomerizing the indigoid dye precursor to the corresponding ketone tautomer; and c) converting the indigoid dye precursor ketone tautomer to an indigoid dye.

[0080] In some embodiments, the indigoid dye precursor ketone tautomer may undergo additional steps before being converted to an indigoid dye. For example, in some embodiments, the compound of Formula (I) 1*H*-indol-3-ol is produced according to the methods described herein; the 1*H*-indol-3-ol is tautomerized to form the corresponding ketone indolin-3-one; the

indolin-3-one is oxidized to form indoline-2,3-dione, and the indoline-2,3-dione is converted into the indigoid dye indirubin.

Polypeptides of the Disclosure

[0081] The present disclosure relates to polypeptides which facilitate the production of indigoid dye precursors from feed indole compounds. As used herein, a "polypeptide" is an amino acid sequence including a plurality of consecutive polymerized amino acid residues (*e.g.*, at least about 15 consecutive polymerized amino acid residues). As used herein, "polypeptide" refers to an amino acid sequence, oligopeptide, peptide, protein, or portions thereof, and the terms "polypeptide" and "protein" are used interchangeably.

[0082] In some embodiments, a polypeptide for use in the methods described herein is a polypeptide with oxygenase activity. As used herein, "oxygenase activity" refers to the ability of a polypeptide to oxidize a substrate by transferring at least one atom of oxygen to the substrate. In some embodiments, the polypeptide is a monooxygenase, *i.e.*, transfers one atom of oxygen to the substrate. In other embodiments, the polypeptide is a dioxygenase, *i.e.*, transfers two atoms of oxygen to the substrate. Any suitable polypeptide with oxygenase activity may be used in the methods described herein. For example, in some embodiments, the polypeptide with oxygenase activity is a naphthalene 1,2-dioxygenase; a cytochrome P450; or a flavin-containing monooxygenase.

[0083] In some embodiments, the polypeptide with oxygenase activity for use in the methods described herein is a polypeptide having the amino acid sequence of SEQ ID NO: 1, which encodes the flavin monooxygenase (FMO) from *Methylophaga* sp. strain SK1. In other embodiments, the polypeptide with oxygenase activity has the amino acid sequence of SEQ ID NO: 2, which encodes the cytochrome P450 oxygenase CYP102A1 from *Bacillus megaterium.* In yet other embodiments, the polypeptide with oxygenase activity has the amino acid sequence of SEQ ID NO: 21-24, which encode subunits ndoA, ndoB, ndoC, and ndoR, respectively, of the naphthalene dioxygenase (NDO) from *Pseudomonas putida*. In some embodiments, the polypeptide with oxygenase activity for use in the methods of the present disclosure is a homolog of the polypeptide having the amino acid sequence of any of one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, and/or SEQ ID NO: 24.

Methods for the identification of polypeptides that are homologs of a polypeptide of interest are well-known to one of skill in the art, as described herein. In some embodiments, polypeptides of the present disclosure include polypeptides containing an amino acid sequence having at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identity to the amino acid sequence of any one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, and/or SEQ ID NO: 24. Polypeptides of the disclosure also include polypeptides having at least 10, at least 12, at least 14, at least 16, at least 18, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, or at least 80 consecutive amino acids of the amino acid sequence of any one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, and/or SEQ ID NO: 24.

[0084] In some embodiments, a polypeptide for use in the methods described herein is a polypeptide with glycosyltransferase activity. As used herein, "glycosyltransferase activity" refers to the ability of a polypeptide to transfer a glycosyl group to a substrate. In some embodiments, the polypeptide is a glucosyltransferase, *i.e.*, transfers a glucosyl group to a substrate. In other embodiments, the polypeptide is a glucuronosyltransferase, *i.e.*, transfers a glucuronosyl group to a substrate.

[0085] In some embodiments, the polypeptide with glycosyltransferase activity for use in the methods described herein is a polypeptide having the amino acid sequence of SEQ ID NO: 3, which encodes the UDP-glucosyltransferase isoform 1 protein from *P. tinctorium*. In other embodiments, the polypeptide with glycosyltransferase activity has the amino acid sequence of SEQ ID NO: 4, which encodes the UDP-glucosyltransferase isoform 2 protein from *P. tinctorium*. In other embodiments, the polypeptide with glycosyltransferase activity has the amino acid sequence of SEQ ID NO: 5, which encodes the UDP-glucosyltransferase protein AHZ08761.1 from *N. tabacum*. In yet other embodiments, the polypeptide with glycosyltransferase activity has the amino acid sequence of SEQ ID NO: 6, which encodes the UDP-glucosyltransferase protein UGT72B1 from *A. thaliana*. In yet other embodiments, the polypeptide with glycosyltransferase activity has the amino acid sequence of SEQ ID NO: 7, which encodes the UDP-glucosyltransferase protein UGT72E2 from *A. thaliana*. In yet other

embodiments, the polypeptide with glycosyltransferase activity has the amino acid sequence of SEQ ID NO: 8, which encodes the UDP-glucosyltransferase protein UGT72E3 from *A. thaliana*.

[0086] In yet other embodiments, the polypeptide with glycosyltransferase activity has the amino acid sequence of SEQ ID NO: 15, which encodes the protein UGT1 from *I. tinctoria*. In yet other embodiments, the polypeptide with glycosyltransferase activity has the amino acid sequence of SEQ ID NO: 16, which encodes the protein UGT2 from *I. tinctoria*. In yet other embodiments, the polypeptide with glycosyltransferase activity has the amino acid sequence of SEQ ID NO: 17, which encodes the protein UGT1 from *I. suffruticosa*. In yet other embodiments, the polypeptide with glycosyltransferase activity has the amino acid sequence of SEQ ID NO: 18, which encodes the protein UGT2 from *I. suffruticosa*. In yet other embodiments, the polypeptide with glycosyltransferase activity has the amino acid sequence of SEQ ID NO: 19, which encodes the protein XP_002320190.1 from *P. trichocarpa*. In yet other embodiments, the polypeptide with glycosyltransferase activity has the amino acid sequence of SEQ ID NO: 20, which encodes the protein BAG80556.1 from *L. barbarum*.

[0087] In yet other embodiments, the polypeptide with glycosyltransferase activity has the amino acid sequence of the catalytic core of the amino acid sequence of any one of SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, and/or SEQ ID NO: 20. In some embodiments, the catalytic core of any one of SEQ ID NO: 3 through SEQ ID NO: 8, or SEQ ID NO: 15 through SEQ ID NO: 20 comprises amino acid 1 through amino acid 270 of the corresponding SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, or SEQ ID NO: 20.

[0088] In some embodiments, the polypeptide with glycosyltransferase activity for use in the methods of the present disclosure is a homolog of the polypeptide having the amino acid sequence of any of one of SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, and/or SEQ ID NO: 20. Methods for the identification of polypeptides that are homologs of a polypeptide of interest are well-known to one of skill in the art, as described

herein. In some embodiments, polypeptides of the present disclosure include polypeptides containing an amino acid sequence having at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identity to the amino acid sequence of any one of SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, and/or SEQ ID NO: 20. Polypeptides of the disclosure also include polypeptides having at least 10, at least 12, at least 14, at least 16, at least 18, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, or at least 80 consecutive amino acids of the amino acid sequence of any one of SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, and/or SEQ ID NO: 20.

[0089] In some embodiments, a polypeptide for use in the methods described herein is a polypeptide with sulfotransferase activity. As used herein, "sulfotransferase activity" refers to the ability of a polypeptide to transfer a sulfo group to a substrate.

[0090] In some embodiments, the polypeptide with sulfotransferase activity for use in the methods described herein is a polypeptide having the amino acid sequence of SEQ ID NO: 9, which encodes the sulfotransferase protein variant SULT1A1*1 wild type from *H. sapiens.* In other embodiments, the polypeptide with sulfotransferase activity is a polypeptide having the amino acid sequence of SEQ ID NO: 10, which encodes the sulfotransferase protein variant SULT1A1*1 D249G from *H. sapiens*. In other embodiments, the polypeptide with sulfotransferase activity is a polypeptide having the amino acid sequence of SEQ ID NO: 11, which encodes the sulfotransferase protein variant SULT1A1*1 enh1 from *H. sapiens*. In yet other embodiments, the polypeptide with sulfotransferase activity is a polypeptide having the amino acid sequence of SEQ ID NO: 12, which encodes the sulfotransferase protein variant SULT1A3*1 wild type from *H. sapiens.* In other embodiments, the polypeptide with sulfotransferase activity is a polypeptide having the amino acid sequence of SEQ ID NO: 13, which encodes the sulfotransferase protein variant SULT1A3*1 D249G from *H. sapiens*. In other embodiments, the polypeptide with sulfotransferase activity is a polypeptide having the

amino acid sequence of SEQ ID NO: 14, which encodes the sulfotransferase protein variant SULT1A3*1 enh1 from *H. sapiens*.

[0091] In some embodiments, the polypeptide with sulfotransferase activity for use in the methods of the present disclosure is a homolog of the polypeptide having the amino acid sequence of any of one of SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, and/or SEQ ID NO: 14. Methods for the identification of polypeptides that are homologs of a polypeptide of interest are well-known to one of skill in the art, as described herein. In some embodiments, polypeptides of the present disclosure include polypeptides containing an amino acid sequence having at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100% identity to the amino acid sequence of any one of SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, and/or SEQ ID NO: 14. Polypeptides of the disclosure also include polypeptides having at least 10, at least 12, at least 14, at least 16, at least 18, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, or at least 80 consecutive amino acids of the amino acid sequence of any one of SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, and/or SEQ ID NO: 14.

[0092] In some embodiments of the methods described herein, the polypeptide with oxygenase activity and the polypeptide with glycosyltransferase activity, sulfotransferase activity, acyltransferase activity, or a combination thereof, are the same polypeptide. For example, in some embodiments, the polypeptide with oxygenase activity and the polypeptide with glycosyltransferase activity, sulfotransferase activity, acyltransferase activity, or a combination thereof is a fusion protein. As used herein, "fusion protein" refers to a single polypeptide that is produced by joining two or more polynucleotides that previously coded for separate polypeptides. In some variations of the methods described herein, the feed indole compound is contacted by a single polypeptide with both oxygenase activity and glycosyltransferase activity to produce the indigoid dye precursor. In other variations, the feed indole compound is contacted by a single polypeptide with both oxygenase activity and sulfotransferase activity. In yet other variations, the feed indole is contacted by a single polypeptide with both oxygenase activity and acyltransferase activity.

[0093] In some embodiments, a polypeptide for use in the methods described herein is a polypeptide with phosphotransferase activity. As used herein, "phosphotransferase activity" refers to the ability of a polypeptide to catalyze a phosphorylation reaction, in which a phosphate group is transferred to a substrate.

[0094] In certain embodiments, the present disclosure relates to cDNA encoding the polypeptides described herein, such as cDNAs encoding SEQ ID NOs: 1-29.

Methods of Identifying Sequence Similarity

[0095] As described above, various polypeptides having similar sequences to the polypeptides used in the methods and compositions of the present disclosure may also be used herein. Various methods are known to those of skill in the art for identifying similar (*e.g.* homologs, orthologs, paralogs, etc.) polypeptide and/or polynucleotide sequences, including phylogenetic methods, sequence similarity analysis, and hybridization methods.

[0096] Phylogenetic trees may be created for a gene family by using a program such as CLUSTAL (Thompson et al. Nucleic Acids Res. 22: 4673-4680 (1994); Higgins et al. *Methods Enzymol* 266: 383-402 (1996)) or MEGA (Tamura et al. *Mol. Biol. & Evo.* 24:1596-1599 (2007)). Once an initial tree for genes from one species is created, potential orthologous sequences can be placed in the phylogenetic tree and their relationships to genes from the species of interest can be determined. Evolutionary relationships may also be inferred using the Neighbor-Joining method (Saitou and Nei, *Mol. Biol.* & Evo. 4:406-425 (1987)). Homologous sequences may also be identified by a reciprocal BLAST strategy. Evolutionary distances may be computed using the Poisson correction method (Zuckerkandl and Pauling, pp. 97-166 in *Evolving Genes and Proteins*, edited by V. Bryson and H.J. Vogel. Academic Press, New York (1965)).

[0097] In addition, evolutionary information may be used to predict gene function. Functional predictions of genes can be greatly improved by focusing on how genes became similar in sequence (*i.e.* by evolutionary processes) rather than on the sequence similarity itself (Eisen, *Genome Res.* 8: 163-167 (1998)). Many specific examples exist in which gene function has been shown to correlate well with gene phylogeny (Eisen, *Genome Res.* 8: 163-167 (1998)).

[0098] When a group of related sequences are analyzed using a phylogenetic program such as CLUSTAL, closely related sequences typically cluster together or in the same clade (a group of similar genes). Groups of similar genes can also be identified with pair-wise BLAST analysis (Feng and Doolittle, *J. Mol. Evol.* 25: 351-360 (1987)). Analysis of groups of similar genes with similar function that fall within one clade can yield sub-sequences that are particular to the clade. These sub-sequences, known as consensus sequences, can not only be used to define the sequences within each clade, but define the functions of these genes; genes within a clade may contain paralogous sequences, or orthologous sequences that share the same function (see also, for example, Mount, Bioinformatics: Sequence and Genome Analysis Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., page 543 (2001)).

[0099] To find sequences that are homologous to a reference sequence, BLAST nucleotide searches can be performed with the BLASTN program, score=100, wordlength=12, to obtain nucleotide sequences homologous to a nucleotide sequence encoding a protein of the disclosure. BLAST protein searches can be performed with the BLASTX program, score=50, wordlength=3, to obtain amino acid sequences homologous to a protein or polypeptide of the disclosure. To obtain gapped alignments for comparison purposes, Gapped BLAST (in BLAST 2.0) can be utilized as described in Altschul et al. (1997) *Nucleic Acids Res.* 25:3389. Alternatively, PSI-BLAST (in BLAST 2.0) can be used to perform an iterated search that detects distant relationships between molecules. See Altschul et al. (1997) supra. When utilizing BLAST, Gapped BLAST, or PSI-BLAST, the default parameters of the respective programs (*e.g.*, BLASTN for nucleotide sequences, BLASTX for proteins) can be used.

[0100] Methods for the alignment of sequences and for the analysis of similarity and identity of polypeptide and polynucleotide sequences are well-known in the art.

[0101] As used herein "sequence identity" and a "sequence at least X% identical to..." refers to the percentage of residues that are identical in the same positions in the sequences being analyzed. As used herein "sequence similarity" refers to the percentage of residues that have similar biophysical / biochemical characteristics in the same positions (*e.g.* charge, size, hydrophobicity) in the sequences being analyzed.

[0102] Methods of alignment of sequences for comparison are well-known in the art, including manual alignment and computer assisted sequence alignment and analysis. This latter approach is a preferred approach in the present disclosure, due to the increased throughput afforded by computer assisted methods. As noted below, a variety of computer programs for performing sequence alignment are available, or can be produced by one of skill.

[0103] The determination of percent sequence identity and/or similarity between any two sequences can be accomplished using a mathematical algorithm. Examples of such mathematical algorithms are the algorithm of Myers and Miller, CABIOS 4:11-17 (1988); the local homology algorithm of Smith et al., Adv. Appl. Math. 2:482 (1981); the homology alignment algorithm of Needleman and Wunsch, *J. Mol. Biol.* 48:443-453 (1970); the search-for-similarity-method of Pearson and Lipman, *Proc. Natl. Acad. Sci.* 85:2444-2448 (1988); the algorithm of Karlin and Altschul, *Proc. Natl. Acad. Sci. USA* 87:2264-2268 (1990), modified as in Karlin and Altschul, *Proc. Natl. Acad. Sci. USA* 90:5873-5877 (1993).

[0104] Computer implementations of these mathematical algorithms can be utilized for comparison of sequences to determine sequence identity and/or similarity. Such implementations include, for example: CLUSTAL in the PC/Gene program (available from Intelligenetics, Mountain View, Calif.); the AlignX program, version10.3.0 (Invitrogen, Carlsbad, CA) and GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Version 8 (available from Genetics Computer Group (GCG), 575 Science Drive, Madison, Wis., USA). Alignments using these programs can be performed using the default parameters. The CLUSTAL program is well described by Higgins et al. *Gene* 73:237-244 (1988); Higgins et al. *CABIOS* 5:151-153 (1989); Corpet et al., *Nucleic Acids Res.* 16:10881-90 (1988); Huang et al. *CABIOS* 8:155-65 (1992); and Pearson et al., *Meth. Mol. Biol.* 24:307-331 (1994). The BLAST programs of Altschul et al. *J. Mol. Biol.* 215:403-410 (1990) are based on the algorithm of Karlin and Altschul (1990) supra.

[0105] Polynucleotides homologous to a reference sequence can be identified by hybridization to each other under stringent or under highly stringent conditions. Single stranded polynucleotides hybridize when they associate based on a variety of well characterized physical-chemical forces, such as hydrogen bonding, solvent exclusion, base stacking and the like. The

stringency of a hybridization reflects the degree of sequence identity of the nucleic acids involved, such that the higher the stringency, the more similar are the two polynucleotide strands. Stringency is influenced by a variety of factors, including temperature, salt concentration and composition, organic and non-organic additives, solvents, etc. present in both the hybridization and wash solutions and incubations (and number thereof), as described in more detail in references cited below (e.g., Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Ed., Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. ("Sambrook") (1989); Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology, vol. 152 Academic Press, Inc., San Diego, Calif. ("Berger and Kimmel") (1987); and Anderson and Young, "Quantitative Filter Hybridisation." In: Hames and Higgins, ed., Nucleic Acid Hybridisation, A Practical Approach. Oxford, TRL Press, 73-111 (1985)).

[0106] Encompassed by the disclosure are polynucleotide sequences that are capable of hybridizing to the disclosed polynucleotide sequences and fragments thereof under various conditions of stringency (see, for example, Wahl and Berger, *Methods Enzymol.* 152: 399-407 (1987); and Kimmel, *Methods Enzymo.* 152: 507-511, (1987)). Full length cDNA, homologs, orthologs, and paralogs of polynucleotides of the present disclosure may be identified and isolated using well-known polynucleotide hybridization methods.

[0107] With regard to hybridization, conditions that are highly stringent, and means for achieving them, are well known in the art. See, for example, Sambrook et al. (1989) (supra); Berger and Kimmel (1987) pp. 467-469 (supra); and Anderson and Young (1985) (supra).

[0108] Hybridization experiments are generally conducted in a buffer of pH between 6.8 to 7.4, although the rate of hybridization is nearly independent of pH at ionic strengths likely to be used in the hybridization buffer (Anderson and Young (1985)(supra)). In addition, one or more of the following may be used to reduce non-specific hybridization: sonicated salmon sperm DNA or another non-complementary DNA, bovine serum albumin, sodium pyrophosphate, sodium dodecylsulfate (SDS), polyvinyl-pyrrolidone, ficoll and Denhardt's solution. Dextran sulfate and polyethylene glycol 6000 act to exclude DNA from solution, thus raising the effective probe DNA concentration and the hybridization signal within a given unit of time. In some instances, conditions of even greater stringency may be desirable or required to reduce non-specific and/or

background hybridization. These conditions may be created with the use of higher temperature, lower ionic strength and higher concentration of a denaturing agent such as formamide.

[0109] Stringency conditions can be adjusted to screen for moderately similar fragments such as homologous sequences from distantly related organisms, or to highly similar fragments such as genes that duplicate functional enzymes from closely related organisms. The stringency can be adjusted either during the hybridization step or in the post-hybridization washes. Salt concentration, formamide concentration, hybridization temperature and probe lengths are variables that can be used to alter stringency. As a general guideline, high stringency is typically performed at T_m-5°C to T_m-20°C, moderate stringency at T_m-20°C to T_m-35°C and low stringency at T_m-35°C to T_m-50° C for duplex >150 base pairs. Hybridization may be performed at low to moderate stringency (25-50°C below T_m), followed by post-hybridization washes at increasing stringencies. Maximum rates of hybridization in solution are determined empirically to occur at T_m-25°C for DNA-DNA duplex and T_m-15°C for RNA-DNA duplex. Optionally, the degree of dissociation may be assessed after each wash step to determine the need for subsequent, higher stringency wash steps.

[0110] High stringency conditions may be used to select for nucleic acid sequences with high degrees of identity to the disclosed sequences. An example of stringent hybridization conditions obtained in a filter-based method such as a Southern or northern blot for hybridization of complementary nucleic acids that have more than 100 complementary residues is about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH.

[0111] Hybridization and wash conditions that may be used to bind and remove polynucleotides with less than the desired homology to the nucleic acid sequences or their complements that encode the present polypeptides include, for example: 6X SSC and 1% SDS at 65°C; 50% formamide, 4X SSC at 42°C; 0.5X SSC to 2.0 X SSC, 0.1% SDS at 50°C to 65°C; or 0.1X SSC to 2X SSC, 0.1% SDS at 50°C - 65°C; with a first wash step of, for example, 10 minutes at about 42°C with about 20% (v/v) formamide in 0.1X SSC, and with, for example, a subsequent wash step with 0.2 X SSC and 0.1% SDS at 65°C for 10, 20 or 30 minutes.

[0112] For identification of less closely related homologs, wash steps may be performed at a lower temperature, *e.g.*, 50° C. An example of a low stringency wash step employs a solution and conditions of at least 25°C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS over 30 min. Greater stringency may be obtained at 42°C in 15 mM NaCl, with 1.5 mM trisodium citrate, and 0.1% SDS over 30 min. Wash procedures will generally employ at least two final wash steps. Additional variations on these conditions will be readily apparent to those skilled in the art (see, for example, US Patent Application No. 20010010913).

[0113] If desired, one may employ wash steps of even greater stringency, including conditions of 65°C -68°C in a solution of 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS, or about 0.2X SSC, 0.1% SDS at 65° C and washing twice, each wash step of 10, 20 or 30 min in duration, or about 0.1 X SSC, 0.1% SDS at 65° C and washing twice for 10, 20 or 30 min. Hybridization stringency may be increased further by using the same conditions as in the hybridization steps, with the wash temperature raised about 3°C to about 5°C, and stringency may be increased even further by using the same conditions except the wash temperature is raised about 6°C to about 9°C.

[0114] Polynucleotide probes may be prepared with any suitable label, including a fluorescent label, a colorimetric label, a radioactive label, or the like. Labeled hybridization probes for detecting related polynucleotide sequences may be produced, for example, by oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide.

Host Cells of the Disclosure

[0115] Host cells of the present disclosure are capable of producing an indigoid dye precursor compound of Formula (A) from a feed indole compound. Host cells of the disclosure express a polypeptide with oxygenase activity; and a polypeptide with glycosyltransferase activity, sulfotransferase activity, acyltransferase activity, or phosphotransferase activity, or a combination thereof. In some embodiments, the polypeptide with oxygenase activity and the polypeptide with glycosyltransferase activity, sulfotransferase activity, acyltransferase activity, or phosphotransferase activity, or a combination thereof, expressed by the host cell are separate polypeptides. In other embodiments, the polypeptide with oxygenase activity and the polypeptide with glycosyltransferase activity, sulfotransferase activity, acyltransferase activity,

or phosphotransferase activity, or a combination thereof, expressed by the host cell are the same polypeptide.

[0116] Host cells of the disclosure may be cultured under conditions such that one or more polypeptides facilitate the production of an indigoid dye precursor compound of Formula (A) from a feed indole compound.

[0117] In some embodiments, host cells contain a recombinant nucleic acid of the present disclosure. In some embodiments, host cells of the present disclosure contain a recombinant nucleic acid encoding SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, and/or SEQ ID NO: 24. In some embodiments, host cells of the present disclosure contain a recombinant nucleic acid encoding a homolog or fragment of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, and/or SEQ ID NO: 24.

[0118] In some embodiments, the host cells contain a recombinant nucleic acid encoding SEQ ID NO: 1 and a recombinant nucleic acid encoding SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, or SEQ ID NO: 20. In other embodiments, the host cells contain a contain a recombinant nucleic acid encoding SEQ ID NO: 2 and a recombinant nucleic acid encoding SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, or SEQ ID NO: 20. In other embodiments, the host cells contain a contain a recombinant nucleic acid encoding SEQ ID NO: 21 and a recombinant nucleic acid encoding SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19,

or SEQ ID NO: 20. In other embodiments, the host cells contain a contain a recombinant nucleic acid encoding SEQ ID NO: 22 and a recombinant nucleic acid encoding SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, or SEQ ID NO: 20. In other embodiments, the host cells contain a contain a recombinant nucleic acid encoding SEQ ID NO: 23 and a recombinant nucleic acid encoding SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, or SEQ ID NO: 20. In other embodiments, the host cells contain a contain a recombinant nucleic acid encoding SEQ ID NO: 24 and a recombinant nucleic acid encoding SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, or SEQ ID NO: 20.

[0119] In yet other embodiments, the host cells contain a contain a recombinant nucleic acid encoding SEQ ID NO: 1 and a recombinant nucleic acid encoding SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, or SEQ ID NO: 14. In still other embodiments, the host cells contain a contain a recombinant nucleic acid encoding SEQ ID NO: 2 and a recombinant nucleic acid encoding SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, or SEQ ID NO: 14. In still other embodiments, the host cells contain a contain a recombinant nucleic acid encoding SEQ ID NO: 21 and a recombinant nucleic acid encoding SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, or SEQ ID NO: 14. In still other embodiments, the host cells contain a contain a recombinant nucleic acid encoding SEQ ID NO: 22 and a recombinant nucleic acid encoding SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, or SEQ ID NO: 14. In still other embodiments, the host cells contain a contain a recombinant nucleic acid encoding SEQ ID NO: 23 and a recombinant nucleic acid encoding SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, or SEQ ID NO: 14. In still other embodiments, the host cells contain a contain a recombinant nucleic acid encoding SEQ ID NO: 24 and a recombinant nucleic acid encoding SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, or SEQ ID NO: 14.

[0120] When the recombinant nucleotide is expressed in the host to produce a polypeptide such as, for example, the polypeptide encoded by SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, and/or SEQ ID NO: 24, the recombinant polypeptide may facilitate the production of a compound of Formula (A) from a compound of Formula (II) when the host cell has been contacted with a compound of Formula (II).

Host Cell Types

[0121] Host cells of the present disclosure may include or be derived from a variety of sources readily apparent to those skilled in the art. Host cells of the present disclosure may be prokaryotic such as, for example, an organism from the kingdom Eubacteria, which includes species of bacteria. In some embodiments, a prokaryotic host cell may include, for example, a cell from the bacterium *Escherichia coli* or the bacterium *Corynebacterium glutamicum*.

[0122] Host cells of the present disclosure may also be eukaryotic and may include, for example, fungal, plant, insect and mammalian cells. In some embodiments, the host cell is from yeast, such as, for example, *Saccharomyces cerevisiae*, *Pichia pastoris*, *Kluyveromyces lactis*, *Kluyveromyces marxianus*, and *Schizosaccharomyces pombe*.

Host Cell Modifications

[0123] Host cells of the present disclosure may also include, for example, host cells that produce uracil-diphosphate glucose (UDP-glucose). Host cells of the present disclosure may be modified to produce excess quantities of UDP-glucose as compared to a corresponding unmodified host cell. The modification may be, for example, genetic modification. Where the modification is a genetic modification, a corresponding unmodified host cell may be, for example, a host cell that lacks the same genetic modification facilitating the production of excess quantities UDP-glucose in the modified host cell.

[0124] Host cells that produce excess quantities of UDP-glucose, as well as methods of making such host cells, are known in the art. For example, in some embodiments,

overexpression of UDP-glucose biosynthesis enzymes (*e.g.*, UDP-glucose pyrophosphorylase) and/or sugar interconversion enzymes (*e.g.*, phosphoglucomutase) in the host cell result in increased production of UDP-glucose as compared to host cells without the corresponding overexpression. In other embodiments, heterologous overexpression of sucrose synthase and sucrose transporter genes by the host cell result in regeneration of UDP-glucose via the breakdown of sucrose. *See* Zichao Mao, Hyun-Dong Shin, and Rachel Ruizhen Chen, Engineering the *E. coli* UDP-Glucose Synthesis Pathway for Oligosaccharide Synthesis, *Biotechnol. Prog.* (2006), 22, pages 369-374; and WO2013022989. In still other embodiments, the expression of genes encoding phosphoglucose isomerase, phosphoglucose mutase, or UDP-glucose-4-epimerase are reduced or eliminated in the host cell, which prevents loss of UDP-glucose to glycolysis and galactose. *See* Dinesh Simkhada, Nagendra Prasad Kurumbang, Hei Chan Lee, and Jae Kyung Sohng, Exploration of Glycosylated Flavonoids from Metabolically Engineered *E. coli*, *Biotechnology and Bioprocess Engineering* (2010), 15, pages 754-760.

[0125] Host cells of the present disclosure may also include, for example, host cells that produce 3'-phosphoadenosine-5'-phosphosulfate (PAPS). Host cells of the present disclosure may be modified to produce excess quantities of PAPS as compared to a corresponding unmodified host cell. The modification may be, for example, genetic modification. Where the modification is a genetic modification, a corresponding unmodified host cell may be, for example, a host cell that lacks the same genetic modification facilitating the production of excess quantities PAPS in the modified host cell.

[0126] Host cells that produce excess quantities of PAPS, as well as methods of making such host cells, are known in the art. In some embodiments, the overexpression of ATP sulfurylase and adenosine-5'-phosphosulfate kinase by the host cell leads to production of excess quantities of PAPS. In other embodiments, the reduction or elimination of expression of genes encoding PAPS reductase in the host cell leads to decreased production of PAPS. *See* Elio Rossi, Sara Motta, Pierluigi Mauri, Paolo Landini, Sulfate assimilation pathway intermediate phosphoadenosine 5′-phosphosulfate acts as a signal molecule affecting production of curli fibres in *Escherichia coli, Microbiology* (2014), 160, 9, pages 1832-1844.

[0127] Host cells of the present disclosure may yet also include, for example, host cells that produce indole. Host cells of the present disclosure may be modified to produce excess quantities of indole as compared to a corresponding unmodified host cell. The modification may be, for example, genetic modification. Where the modification is a genetic modification, a corresponding unmodified host cell may be, for example, a host cell that lacks the same genetic modification facilitating the production of excess quantities indole in the modified host cell.

[0128] Host cells that produce excess quantities of indole, as well as methods of making such host cells, are known in the art. In some embodiments, the reduction or elimination of genes encoding the glucose phosphotransferase system of the host cell and the overexpression of genes encoding a pentose phosphate pathway transketolase of the host cell leads to accumulation of phosphoenolpyruvate and erythrose-4-phosphate, a precursor to aromatic amino acid and indole synthesis. *See* US6962794 In other embodiments, overexpression of genes encoding enzymes of aromatic amino acid (*e.g*, tryptophan) biosynthesis pathways in the host cell lead to increased production of tryptophan, indole, and indole precursors. For example, in certain embodiments, the genes that encode 3-deoxy-D-arabino-heptulosonic acid 7-phosphate synthase and/or anthranilate synthase are overexpressed in the host cell. *See* Berry, A. et al. Application of metabolic engineering to improve both the production and use of biotech indigo, *Journal of Industrial Microbiology & Biotechnology,*(2002), 28, 127, pages 133. In still other embodiments, a gene encoding tryptophan indole-lyase or mutant indole-releasing tryptophan synthase is overexpressed in the host cell, which leads to increased conversion of tryptophan or indole precursors into free indole. *See* US5494816.

[0129] Host cells of the present disclosure may also include, for example, host cells that naturally produce one or more hydrolases, including, for example, glycosidase, sulfatase, and/or esterase. Host cells of the present disclosure may be modified to produce decreased quantities of one or more hydrolases as compared to a corresponding unmodified host cell. The modification may be, for example, genetic modification. Where the modification is a genetic modification, a corresponding unmodified host cell may be, for example, a host cell that lacks the same genetic modification facilitating the production of decreased quantities of one or more hydrolases in the modified host cell.

[0130] Host cells that produce decreased quantities of one or more hydrolases, as well as methods of making such host cells, are known in the art. In some embodiments, the expressionof genes encoding native glycosylhydrolases, sulfatases, and/or esterases is decreased or eliminated in the host cell, which leads to reduced hydrolysis of desired products.

[0131] Host cells of the present disclosure may also include, for example, host cells that possess heterologous genes encoding oligosaccharide transporters, organic anion transporters, and/or multidrug transporters, which may lead to increased export of glycosyl- or sulfate-containing compounds. *See* Sandermann H Jr, beta-D-Galactoside transport in *Escherichia coli:* substrate recognition, *Eur J Biochem* (1977), 80, 2, pages 507-515; Enomoto A, Takeda M, Tojo A, Sekine T, Cha SH, Khamdang S, Takayama F, Aoyama I, Nakamura S, Endou H, Niwa T. Role of organic anion transporters in the tubular transport of indoxyl sulfate and the induction of its nephrotoxicity, *J Am Soc Nephrol*, (2002), 13, 7, pages 1711-1720.

[0132] Host cells of the present disclosure may also include, for example, host cells that possess a heterologous gene for an isatin hydrolase, which may lead to decreased production of the indirubin by-product during production of indoxyl by the host cell. *See* Berry, A. et al., Application of metabolic engineering to improve both the production and use of biotech indigo. *Journal of Industrial Microbiology & Biotechnology*, (2002), 28, pages 127-133.

[0133] Host cells of the present disclosure may also include cells that either naturally exhibit reduced hydrolysis of compounds produced by methods of the present disclosure or are engineered to reduce hydrolysis of compounds produced by methods of the present disclosure, including, for example, indoxyl acetate. Such host cells may naturally, or be engineered such that they do not produce one or more of the following polypeptides: acetyl esterase, esterase yjfP, pimeloyl-[acyl-carrier protein] methyl ester esterase, S-formylglutathione hydrolase YeiG, S-formylglutathione hydrolase FrmB, proofreading thioesterase EntH, uncharacterized protein YdiL, acyl-CoA thioesterase I, probable 9-O-acetyl-N-acetylneuraminic acid deacetylase, esterase YqiA, esterase YbfF, esterase YpfH, acetylornithine deacetylase, esterase FrsA, acyl-CoA thioester hydrolase YbgC, and tryptophanase.

Contacting a Host Cell with a Compound of Formula (I)

[0134] In some embodiments of the methods of the present disclosure, a host cell containing a recombinant polypeptide of the disclosure is contacted with a feed indole compound of Formula (I), and an indigoid dye precursor of Formula (A) is produced from the feed indole compound. In certain embodiments, the compound of Formula (I) is a compound of Formula (II).

[0135] In certain embodiments, to contact a host cell with a feed indole compound, the feed indole compound is added to the growth medium of the host cell. In other embodiments, to contact a host cell with a feed indole compound, the feed indole compound is produced by the host cell. For example, in certain embodiments, the host cell produces 1*H*-indol-3-ol, and converts the 1*H*-indol-3-ol to a compound of Formula (A).

[0136] In yet other embodiments, to contact a host cell with a feed indole compound, the host cell produces a first feed indole compound and converts the first feed indole compound to one or more additional feed indole compounds. For example, in certain embodiments, the host cell produces tryptophan, converts the tryptophan to 1*H*-indol-3-ol, and converts the 1*H*-indol-3-ol to a compound of Formula (A).

[0137] In still other embodiments, to contact a host cell with a feed indole compound, the feed indole compound is produced by other means in the growth medium of the host cell. For example, in certain embodiments, the feed indole compound is produced by another organism in the growth medium of the host cell, and the host cell converts the feed indole compound to a compound of Formula (A). In another variation, for example, the feed indole compound is produced *in vitro* in the growth medium of the host cell, and the host cell converts the feed indole compound to a compound of Formula (A).

Culture Conditions for Host Cells

[0138] In some embodiments, the methods of the present disclosure include contacting a host cell containing a recombinant polypeptide of the disclosure with a feed indole compound, and culturing the host cell under conditions such that an indigoid dye precursor of Formula (A) is produced from the feed indole compound.

[0139] Standard methods of culturing organisms such as, for example, bacteria and yeast, are well-known in the art and are described herein. For example, host cells may be cultured in a standard growth media under standard temperature and pressure conditions, and in an aerobic environment. Standard growth media for various host cells are commercially available and well-known in the art, as are standard conditions for growing various host cells. Suitable conditions for facilitating the production of compounds of Formula (A) from compounds of Formula (I) or (II) by host cells are described herein and will be readily apparent to one of skill in the art in view of the present disclosure.

[0140] In some embodiments, various compounds and/or reagents may be added to the growth medium of a host cell that produces a recombinant polypeptide of the disclosure to enhance or facilitate the production of compounds of Formula (A) from compounds of Formula (II). In some embodiments, the culture medium is supplemented with protecting group precursors such as glucose, galactose, xylose, sucrose, sulfate, or acetate. In some embodiments, the culture medium is supplemented with feed indole compounds, including, for example, tryptophan or indole. The culture media may be supplemented with combinations of various compounds and/or reagents. For example, in some embodiments, the culture media is supplemented with tryptophan and glucose.

Isolation of the Compound of Formula (A)

[0141] In some embodiments, after a compound of Formula (A) has been produced from a feed indole compound, a recovery step may be performed to recover the compound of Formula (A) from the host cell, the growth medium, or the *in vitro* reaction. In some embodiments, the host cells used in the methods described herein may excrete the compounds of Formula (A) produced. In other embodiments, the compounds of Formula (A) produced are not excreted by the host cells. Methods for the recovery of compounds of Formula (A) may include, for example, sonication, centrifugation, precipitation, filtration, chromatography, crystallization, and/or solvent extraction.

[0142] Compounds of Formula (A) that have been recovered from a host cell may be referred to as substantially purified compounds of Formula (A). A substantially purified compound of Formula (A) generally refers to a compound of Formula (A) that is substantially free of

contaminating agents (*e.g.* cellular material and other culture medium components) from the culture medium source where the compound of Formula (A) is produced by the host cell. For example, a substantially purified compound of Formula (A) may be in association with less than 30%, 20%, 10%, and more preferably 5% or less (by weight) contaminating agents. A composition containing a substantially purified compound of Formula (A) preparation may include, for example, a composition where culture medium (and associated contaminating agents) represents less than about 20%, sometimes less than about 10%, and often less than about 5% of the volume of the compound of Formula (A) preparation.

[0143] The titer of the compound of Formula (A) produced may be, for example, at least 1 mg/L, at least 10 mg/L, at least 50 mg/L, at least 100 mg/L, at least 1 g/L, at least 10 g/L, at least 25 g/L, at least 50 g/L, at least 75 g/L, at least 100 g/L, at least 125 g/L, at least 150 g/L, at least 175 g/L, at least 200 g/L, or at least 250 g/L. In certain embodiments, the titer of the compound of Formula (A) produced is between 75 g/L and 150 g/L. In other embodiments, the titer of the compound of Formula (A) produced is at least 100 g/L.

In Vitro Methods for Production of Compounds of Formula (A)

[0144] In some embodiments, the methods of the present disclosure include contacting a feed indole compound with a recombinant polypeptide of the disclosure and incubating the feed indole compound under conditions such that a compound of Formula (A) is produced from the feed indole compound.

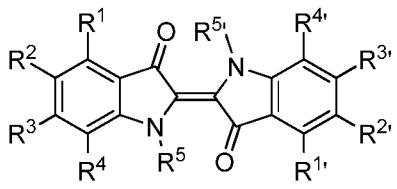
[0145] Standard methods of performing *in vitro* enzymatic reactions are well-known in the art and are described herein. In an exemplary embodiment, a polypeptide of the present disclosure is expressed in a host cell and substantially purified. The substantially purified polypeptide may be added to an *in vitro* reaction platform, such, for example, as a well of a 96-well plate, where the well contains a feed indole compound, such as a compound of Formula (II). The purified polypeptide and the feed indole compound may be incubated together for a period of time to allow production of a compound of Formula (A) from the feed indole compound.

[0146] In some embodiments, various compounds and/or reagents may be added to the *in vitro* platform containing a polypeptide of the disclosure and a feed indole compound to enhance

or facilitate the production of a compound of Formula (A) from a feed indole compound. In some embodiments, the platform is supplemented with redox cofactors, including, for example, nicotinamide adenine dinucleotide (NAD^+) or nicotinamide adenine dinucleotide phosphate ($NADP^+$), or their respective hydrides (NADH or NADPH). The platform may be supplemented with combinations of various compounds and/or reagents, such as, for example, a redox cofactor regeneration system comprising a redox-active enzyme, an enzyme substrate, $NADP^+$, and/or NADPH. In some variations, the redox cofactor regeneration system comprises glucose oxidase, glucose, $NADP^+$, and/or NADPH.

Use of Compounds of Formula (A) to Produce Compounds of Formula (B)

[0147] The indigoid dye precursor compounds of Formula (A) produced according to the methods described herein may be used to produce an indigoid dye compound of Formula (B):



(B),

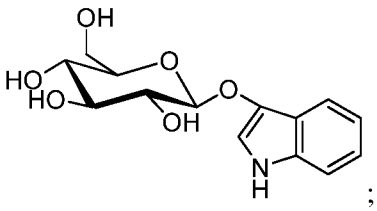
wherein R^1, R^2, R^3, R^4, R^5, $R^{1'}$, $R^{2'}$, $R^{3'}$, $R^{4'}$, and $R^{5'}$ are defined as for Formula (A).

[0148] Thus, in one aspect, provided herein are methods of producing a compound of Formula (B), comprising converting a first compound of Formula (A) and a second compound of Formula (A) to a compound of Formula (B). In some embodiments, the first compound of Formula (A) and the second compound of Formula (A) are the same compound of Formula (A). In other embodiments, the first compound of Formula (A) and the second compound of Formula (A) are different compounds of Formula (A). In some embodiments, at least one of the first compound of Formula (A) and the second compound of Formula (A) is produced according to the methods described herein.

[0149] In some embodiments, one or more compounds of Formula (A) may be used to produce a compound of Formula (B). It should be appreciated by one of skill in the art that the R^1, R^2, R^3, R^4, R^5, $R^{1'}$, $R^{2'}$, $R^{3'}$, $R^{4'}$, and $R^{5'}$ of the one or more compounds of Formula (A) are the same R^1, R^2, R^3, R^4, R^5, $R^{1'}$, $R^{2'}$, $R^{3'}$, $R^{4'}$, and $R^{5'}$ of the compound of Formula (B).

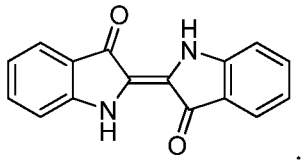
[0150] The one or more compounds of Formula (A) used to produce the compound of Formula (B) may be the same compound of Formula (A).

[0151] For example, in one variation, the compound of Formula (A) is:



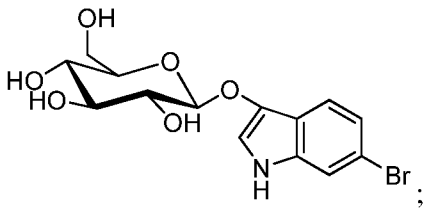
;

and the compound of Formula (B) is:



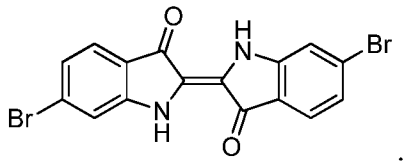
.

[0152] In another variation, the compound of Formula (A) is:



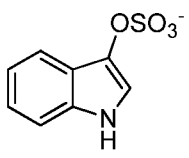
;

and the compound of Formula (B) is:



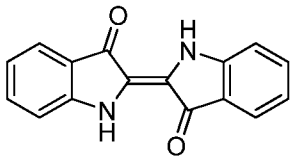
.

[0153] In another variation, the compound of Formula (A) is:



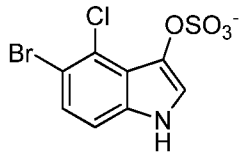
;

and the compound of Formula (B) is:



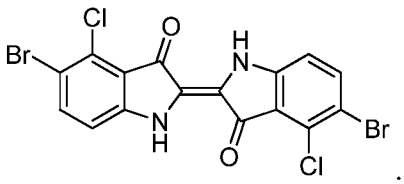
.

[0154] In yet another variation, the compound of Formula (A) is:



;

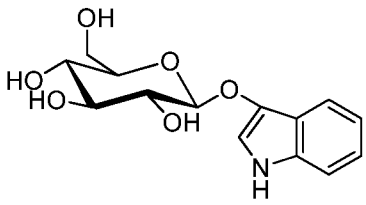
and the compound of Formula (B) is:



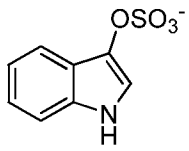
.

[0155] In other embodiments, the one or more compounds of Formula (A) used to produce the compound of Formula (B) may be different compounds of Formula (A).

[0156] For example, in one variation, the compounds of Formula (A) are:

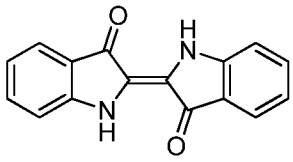


and



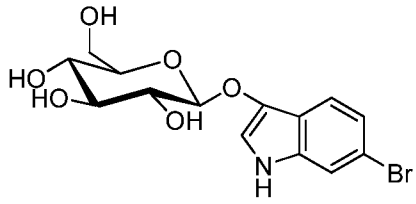
;

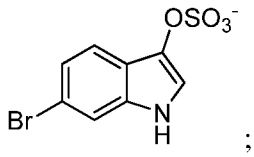
and the compound of Formula (B) is:



.

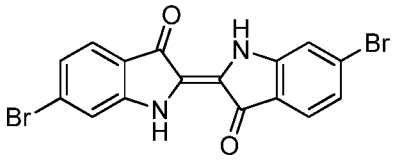
[0157] In another variation, the compounds of Formula (A) are:



and


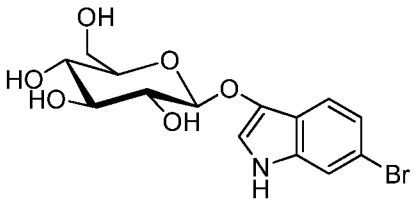
;

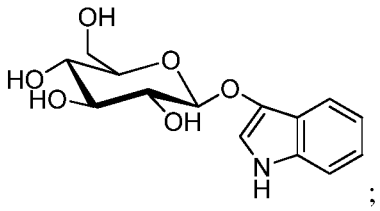
and the compound of Formula (B) is:



.

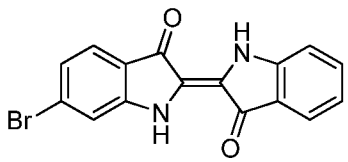
[0158] In yet another variation, the compounds of Formula (A) are:



and


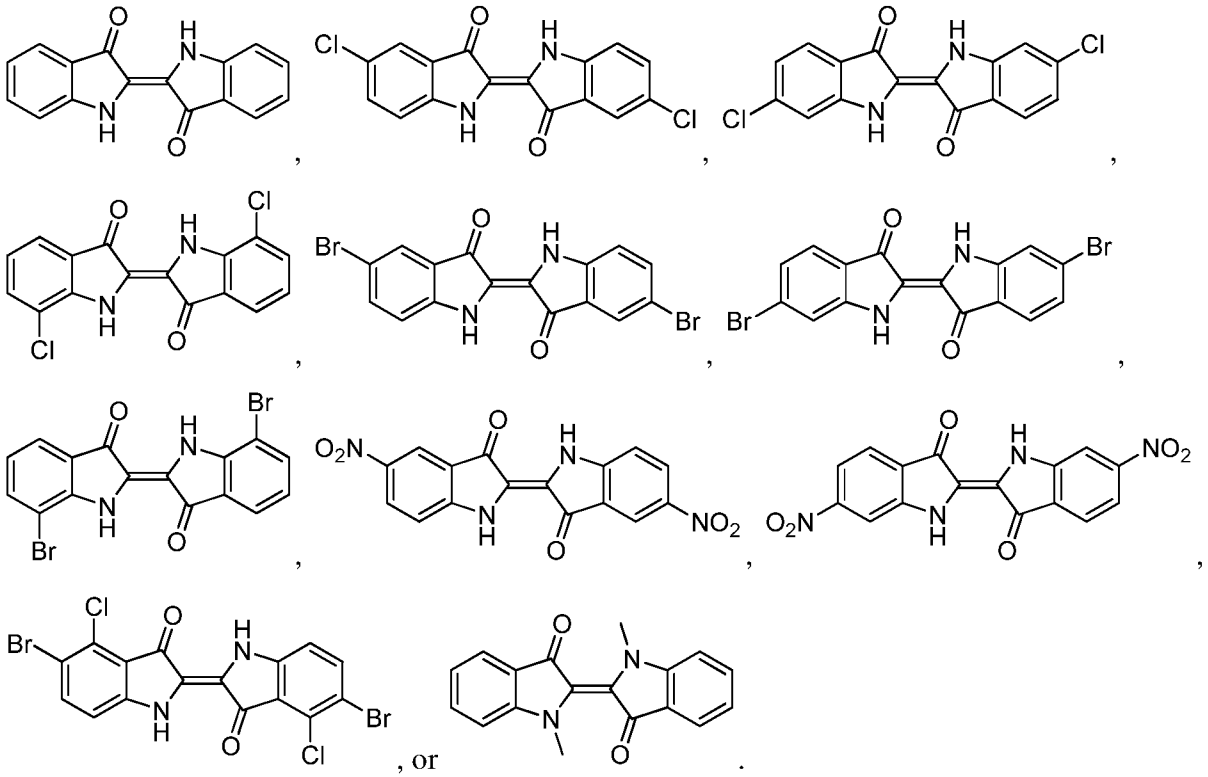
;

and the compound of Formula (B) is:



.

[0159] In certain variations, the compound of Formula (B) is :



[0160] In some embodiments, a compound of Formula (B) is produced from one or more compounds of Formula (A), and each of the one or more compounds of Formula (A) is produced according to the methods described herein.

[0161] In other embodiments, a compound of Formula (B) is produced from one or more compounds of Formula (A), and at least a portion of the one or more compounds of Formula (A) is produced according to the methods described herein, while at least a portion of the one or more compounds of Formula (A) is produced according to other methods. For example, in certain embodiments, at least a portion of the one or more compounds of Formula (A) is isolated from a plant.

[0162] In some embodiments, a compound of Formula (B) may be produced by i) converting one or more compounds of Formula (A) to one or more compounds of Formula (II), then ii) converting the one or more compounds of Formula (II) to a compound of Formula (B). The one or more compounds of Formula (A) may be converted to the one or more compounds of Formula (II) by any suitable methods known in the art.

[0163] For example, in some embodiments, a compound of Formula (A) is contacted by a hydrolase to produce a compound of Formula (II). Any appropriate hydrolase may be used to produce a compound of Formula (II) from a compound of Formula (A). For example, in certain embodiments, a compound of Formula (A) is contacted by a β-glucosidase, α-glucosidase, sulfatase, phosphatase or esterase to produce a compound of Formula (II). In one variation, a compound of Formula (A) is contacted by the hydrolase atsA from *P. aeruginosa* to produce a compound of Formula (II). In another variation, a compound of Formula (A) is contacted by the hydrolase bglA from *Bacillus circulans* to produce a compound of Formula (II).

[0164] In other embodiments, a compound of Formula (A) is converted to a compound of Formula (II) in the presence of an acid catalyst. For example, in certain embodiments, a compound of Formula (A) is converted to a compound of Formula (II) in the presence of hydrochloric acid. In some embodiments, a compound of Formula (A) is converted to a compound of Formula (II) in the presence of 0.001 M hydrochloric acid at a temperature above room temperature. In some embodiments, a compound of Formula (B) is produced by **i)** converting one or more compounds of Formula (A) to one or more compounds of Formula (II), **ii)** combining the one or more compounds of Formula (II) with one or more additional compounds of Formula (II), then **iii)** converting the one or more compounds of Formula (II) to a compound of Formula (B). In some embodiments, the one or more additional compounds of Formula (II) are not produced according to the methods described herein.

[0165] One or more compounds of Formula (II) may be converted to a compound of Formula (B) by any suitable methods known in the art. *See* Lojda, Z. Indigogenic methods for glycosidases, Histochemie (1970), 22, *4*, pages 347-361. For example, in certain variations, one or more compounds of Formula (II) is contacted by oxygen to produce a compound of Formula (B). In some embodiments, one or more compounds of Formula (II) is contacted by atmospheric oxygen to produce a compound of Formula (B). In some embodiments, one or more compounds of Formula (II) is contacted by air to produce a compound of Formula (B).

[0166] In certain embodiments, a compound of Formula (II) is converted to a compound of Formula (B) in the additional presence of an oxidative catalyst, including, for example, $[Fe(CN)_6]^{4-}$ or $[Fe(CN)_6]^{3-}$. In some embodiments, one or more compounds of Formula (II) is

contacted by oxygen in the presence of $[Fe(CN)_6]^{4-}$ or $[Fe(CN)_6]^{3-}$ to produce a compound of Formula (B).

[0167] A compound of Formula (A), Formula (II), and/or Formula (B) produced according to the methods described herein may be used in any suitable application. For example, in some embodiments, a compound of Formula (A), Formula (II), and/or Formula (B) is used in the dyeing of clothes or objects. In some variations, a compound of Formula (A), Formula (II), and/or Formula (B) is used in the dyeing of jeans. In certain variations, a compound of Formula (A) is converted to a compound of Formula (II), the compound of Formula (II) is applied to clothes, and then the compound of Formula (II) is converted to a compound of Formula (B).

EXAMPLES

[0168] The following Examples are offered for illustrative purposes and to aid one of skill in better understanding the various embodiments of the disclosure. The following examples are not intended to limit the scope of the present disclosure in any way.

Example 1

Identification and Isolation of Gene Encoding UDP-glucose:glucosyltransferase isoform 1 from *Polygonum tinctorium*

[0169] This Example demonstrates the identification and isolation of the glycosyltransferase gene encoding UDP-glucose:glucosyltransferase isoform 1 from the plant *Polygonum tinctorium.*

[0170] *P. tinctorium* plants were grown indoors in a greenhouse or laboratory environment. Samples of leaf tissue were taken from live plants at several points during the day and frozen in liquid nitrogen. These samples were mixed, kept frozen with liquid nitrogen, and crushed using 5 mm diameter steel beads in a beadbeater at 30 Hz for 2 min. Total RNA was extracted from the 100 mg of powdered frozen leaf tissue using the Qiagen RNEasy Plant Mini kit obtained from Qiagen GMBH (Hilden, Germany), in accordance with the manufacturer recommended protocol. The mRNA was extracted from the total RNA using magnetic beads coated with oligo $(dT)_{25}$. The mRNA was then sheared to approximately 550 base pairs in length using a Covaris S2 ultrasonicator from Covaris Inc. (Woburn, MA). A cDNA library was generated using the Apollo 324 Next-Gen Library Prep System from Wafergen Biosystems Inc. (Fremont, CA) using

the manufacturer-supplied PrepX RNA-Seq Library Preparation Kit. The cDNA library was then clustered using the cBot from Illumina Inc. (San Diego, CA) and the clustered sample was loaded onto an Illumina HiSeq2500 and sequenced using the Rapid Run reagent kit for 150 base, paired-end reads.

[0171] Paired-end reads received from the Illumina HiSeq2500 sequencer were first trimmed to remove low-quality reads using the Trimmomatic software package in paired-end mode to remove Illumina adapter sequences and using a sliding quality window of 30 or greater, where reads with under 36 acceptable bases are dropped. *See* Bolger AM, Lohse M, Usadel B, Trimmomatic: a flexible trimmer for Illumina sequence data, *Bioinformatics*, (2014), 30(15), pages 2114-20. Overlapping paired-end reads were then merged using the FLASH software package with a minimum overlap size of 15 bases and an expected fragment length of 350 bases. *See* Tanja Magoč and Steven L. Salzberg. FLASH: fast length adjustment of short reads to improve genome assemblies, *Bioinformatics*, (2011), 27(21), pages 2957-2963. The remaining merged and unmerged reads were pooled and digitally normalized to remove redundant data using the khmer software package with options set to paired-end, k-mer size of 19, culling count of 20, and 4 hash tables of 4 GiB each. *See* Brown CT, Howe AC, Zhang Q, Pyrkosz AB, Brom TH, A Reference-Free Algorithm for Computational Normalization of Shotgun Sequencing Data, arXiv:1203.4802 [q-bio.GN], http://arxiv.org/abs/1203.4802, accessed March 2, 2015. Following digital normalization, reads were assembled into transcript scaffolds using the Trinity RNA-seq assembly package in paired-end mode. Scaffolds were then annotated using BLASTX against a library of known plant glycosyltransferases with an E-value threshold of 1e-60 to identify glucosyltransferase candidates. These candidates were then translated into peptides using the Trinity package, Transdecoder, using the default settings. *See* Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A, De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis, *Nat Protoc*, (2013), 8, pages1494-1512.

[0172] The predicted UDP-glucose:glucosyltransferase was purified from *P. tinctorium* following the protocol of Minami et al. *See* Yoshiko Minami, Osamu Nishimura, Ikuko Hara-

Nishimura, Mikio Nishimura, Hiroshi Matsubara, Tissue and Intracellular Localization of Indican and the Purification and Characterization of Indican Synthase from Indigo Plants, *Plant Cell Physiology* (2000), 41, *2*, pages 218-225. 200 grams of fresh leaves were flash frozen with liquid nitrogen and ground into a fine powder using a mortar and pestle. The fine powder was suspended in 400 mL of extraction buffer (2 mM EDTA, 20 mM β-mercaptoethanol, 100 mM H_2KPO_4, 1X cOmplete protease inhibitor, pH 7.0 in water) and centrifuged at 24500 rcf for 30 min while refrigerated at 4°C. The supernatant was transferred to a new tube and precipitated by adding 600 mL of extraction buffer with 50% w/v polyethylene glycol (M_w 6000 Da) to make 1 L total solution with 30% w/v PEG 6000. This solution was centrifuged at 143000 rcf for 30 min at 4°C. The supernatant of this process was mixed with 25 mL DEAE-Sepharose beads. The beads were washed with 125 mL of wash buffer A (1 mM EDTA, 10 mM β-mercaptoethanol, 50 mM HEPES-NaOH pH 7.0), then washed with 50 mL of wash buffer A + 50 mM NaCl, and then protein was eluted with 50 mL of wash buffer A + 100 mM NaCl. This 50 mL elution was loaded on a 10 mL hydroxyapatite column, which was washed with 50 mL buffer A1 (1 mM EDTA, 5 mM dithiothreitol, 10% v/v glycerol, 50 mM HEPES-NaOH pH 7.0) and eluted with 98% A1 / 2% B1 where buffer B1 is buffer A1 + 500 mM H_2KPO_4 pH 7.0. This eluate was loaded onto a Mono Q 5/50 GL column and washed with 5 column volumes of buffer A2 (1 mM EDTA, 5 mM dithiothreitol, 10% v/v glycerol, 10 mM H_2KPO_4, 50 mM HEPES-NaOH pH 7.0). Protein was eluted with a gradient from 0-100% Buffer B2 (Buffer A2 + 200 mM NaCl), at a 1 mL per minute flow rate over 30 min. The protein eluate from 5-9 min was collected and concentrated tenfold using a Millipore 10 kDa molecular weight-cutoff spin column. This concentrate was run on a Superdex 200 size-exclusion column, eluting with 5 mM dithiothreitol, 50 mM HEPES-NaOH pH 7.0. Fractions were lyophilized, resuspended in water, dialyzed in water, and then separated using SDS-PAGE. Proteins trapped in the polyacrylamide gel were electroblotted onto a PVDF membrane and extracted for trypsin digestion and column separation followed by tandem mass spectrometry to identify protein fragments. These protein fragments were correlated to the Transdecoder-predicted sequences using DTAselect and SEQUEST. *See* David L. Tabb, W. Hayes McDonald, John R. Yates III, DTASelect and Contrast: Tools for Assembling and Comparing Protein Identifications from Shotgun Proteomics, *J Proteome Res*, (2002), 1(1): pages 21–26; Eng JK, McCormack AL, Yates JR, An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a

protein database, *J Am Soc Mass Spectrom*, (1994), 5(11), pages 976-989. The sequences with the most complete coverage of matching peptide fragments were selected for further study.

[0173] Another sample of total RNA was prepared as described above, but was converted to adaptor-tagged cDNA using the GeneRacer® Kit with SuperScript® III RT and TOPO TA Cloning® Kit for Sequencing from Life Technologies (Carlsbad, CA). Using gene specific oligonucleotide primers, a polymerase chain reaction was used on the adaptor-tagged cDNA library to clone the sequences of interest identified previously. These genes were then Sanger sequenced to confirm their identity and nucleotide sequence.

Example 2

Identification and Isolation of Gene Encoding UDP-glucose:glucosyltransferase isoform 2 from *Polygonum tinctorium*

[0174] This Example demonstrates the identification and isolation of the glycosyltransferase gene encoding UDP-glucose:glucosyltransferase isoform 2 from the plant *Polygonum tinctorium*.

[0175] The mRNA was extracted from *P. tinctorium* plants and a cDNA library constructed following the procedure as described in Example 1 above.

[0176] The cDNA library was analyzed and the predicted UDP-glucose:glucosyltransferase isoform 2 was purified as described in Example 1 above.

[0177] Another sample of total RNA was prepared from the plants, and sequences of interest identified and Sanger sequenced to confirm their identity and nucleotide sequence as described in Example 1 above.

Example 3

Identification and Isolation of Gene Encoding a UDP-glucose:glucosyltransferase from *Nicotiana tabacum*

[0178] The UDP-glucose:glucosyltransferase of interest from *N. tabacum* was identified to be of interest through BLASTP similarity to other glucosyltransferases of interest. This sequence was cloned via polymerase chain reaction using as template a synthetic piece of DNA

ordered from Integrated DNA Technologies, Inc. (Coralville, IA). The product of the polymerase chain reaction was cloned into an expression vector as described in Example 18.

Example 4

Identification and Isolation of Gene Encoding the UDP-glucose:glucosyltransferase UGT72B1 from *Arabidopsis thaliana*

[0179] The UDP-glucose:glucosyltransferase UGT72B1 from *A. thaliana* was identified to be of interest through BLASTP similarity to other glucosyltransferases of interest. This sequence was cloned via polymerase chain reaction using as template a synthetic piece of DNA ordered from Integrated DNA Technologies, Inc. (Coralville, IA). The product of the polymerase chain reaction was cloned into an expression vector as described in Example 18.

Example 5

Identification and Isolation of Gene Encoding the UDP-glucose:glucosyltransferase UGT72E2 from *Arabidopsis thaliana*

[0180] The UDP-glucose:glucosyltransferase UGT72E2 from *A. thaliana* was identified to be of interest through BLASTP similarity to other glucosyltransferases of interest. This sequence was cloned via polymerase chain reaction using as template a synthetic piece of DNA ordered from Integrated DNA Technologies, Inc. (Coralville, IA). The product of the polymerase chain reaction was cloned into an expression vector as described in Example 18.

Example 6

Identification and Isolation of Gene Encoding the UDP-glucose:glucosyltransferase UGT72E3 from *Arabidopsis thaliana*

[0181] The UDP-glucose:glucosyltransferase UGT72E3 from *A. thaliana* was identified to be of interest through BLASTP similarity to other glucosyltransferases of interest. This sequence was cloned via polymerase chain reaction using as template a synthetic piece of DNA ordered from Integrated DNA Technologies, Inc. (Coralville, IA). The product of the polymerase chain reaction was cloned into an expression vector as described in Example 18.

Example 7

Identification and Isolation of Gene Encoding the Cytochrome P450 BM3 CYP102A1 from *Bacillus megaterium*

[0182] The gene encoding the Cytochrome P450 BM3 CYP102A1 from *B. megaterium* was identified to be of interest by review of the reference source listed in Table **1**. This sequence was cloned via polymerase chain reaction using as template a synthetic piece of DNA ordered from Integrated DNA Technologies, Inc. (Coralville, IA). The product of the polymerase chain reaction was cloned into an expression vector as described in Example 18.

Example 8

Identification and Isolation of Gene Encoding the Flavin Monooxygenase FMO from *Methylophaga* sp. strain SK1

[0183] The gene encoding the Flavin monooxygenase FMO from *Methylophaga* sp. strain SK1 was identified to be of interest by review of the reference source listed in Table **1**. This sequence was cloned via polymerase chain reaction using as template a synthetic piece of DNA ordered from Integrated DNA Technologies, Inc. (Coralville, IA). The product of the polymerase chain reaction was cloned into an expression vector as described in Example 18.

Example 9

Identification and Isolation of Gene Encoding Sulfotransferase SULT1A1*1 Wildtype from *Homo sapiens*

[0184] The gene encoding the sulfotransferase SULT1A1*1 wildtype from *Homo sapiens* was identified to be of interest by review of the reference source listed in Table **1**. This sequence was cloned via polymerase chain reaction using as template a synthetic piece of DNA ordered from Integrated DNA Technologies, Inc. (Coralville, IA). The product of the polymerase chain reaction was cloned into an expression vector as described in Example 18.

Example 10

Identification and Isolation of Gene Encoding Sulfotransferase SULT1A1*1 D249G from *H. sapiens*

[0185] The gene encoding the sulfotransferase SULT1A1*1 D249G from *Homo sapiens* was identified to be of interest by review of the reference source listed in Table **1**. This sequence was cloned via polymerase chain reaction using as template a synthetic piece of DNA ordered from Integrated DNA Technologies, Inc. (Coralville, IA). The product of the polymerase chain reaction was cloned into an expression vector as described in Example 18.

Example 11

Identification and Isolation of Gene Encoding Sulfotransferase SULT1A1*1 enh1 from *H. sapiens*

[0186] The gene encoding the sulfotransferase SULT1A1*1 enh1 from *Homo sapiens* was identified to be of interest by review of the reference source listed in Table **1**. This sequence was cloned via polymerase chain reaction using as template a synthetic piece of DNA ordered from Integrated DNA Technologies, Inc. (Coralville, IA). The product of the polymerase chain reaction was cloned into an expression vector as described in Example 18.

Example 12

Identification and Isolation of Gene Encoding Sulfotransferase SULT1A3*1 Wildtype from *H. sapiens*

[0187] The gene encoding the sulfotransferase SULT1A3*1 wildtype from *Homo sapiens* was identified to be of interest by review of the reference source listed in Table **1**. This sequence was cloned via polymerase chain reaction using as template a synthetic piece of DNA ordered from Integrated DNA Technologies, Inc. (Coralville, IA). The product of the polymerase chain reaction was cloned into an expression vector as described in Example 18.

Example 13

Identification and Isolation of Gene Encoding Sulfotransferase SULT1A3*1 D249G from *H. sapiens*

[0188] The gene encoding the sulfotransferase SULT1A3*1 D249G from *Homo sapiens* was identified to be of interest by review of the reference source listed in Table **1**. This sequence was cloned via polymerase chain reaction using as template a synthetic piece of DNA ordered

from Integrated DNA Technologies, Inc. (Coralville, IA). The product of the polymerase chain reaction was cloned into an expression vector as described in Example 18.

Example 14

Identification and Isolation of Gene Encoding Sulfotransferase SULT1A3*1 enh1 from *H. sapiens*

[0189] The gene encoding the sulfotransferase SULT1A3*1 enh1 from *Homo sapiens* was identified to be of interest by review of the reference source listed in Table **1**. This sequence was cloned via polymerase chain reaction using as template a synthetic piece of DNA ordered from Integrated DNA Technologies, Inc. (Coralville, IA). The product of the polymerase chain reaction was cloned into an expression vector as described in Example 18.

Example 15

Identification and Isolation of Genes Encoding UDP-glucose:glucosyltransferase from *Indigofera tinctoria* and *Indigofera suffruticosa*

[0190] *Indigofera tinctoria* or *Indigofera suffruticosa* plants were grown indoors in a greenhouse or laboratory environment. Samples of leaf tissue were taken from live plants at several points during the day and frozen in liquid nitrogen. These samples were mixed, kept frozen with liquid nitrogen, and crushed using 5 mm. diameter steel bead in a beadbeater at 30 Hz for 2 minutes. Total RNA was extracted from the 100 mg of powdered frozen leaf tissue using the Qiagen RNEasy Plant Mini kit obtained from Qiagen GMBH (Hilden, Germany), in accordance with the manufacturer recommended protocol. The total RNA was converted to adaptor-tagged cDNA using the GeneRacer® Kit with SuperScript® III RT and TOPO TA Cloning® Kit for Sequencing from Life Technologies (Carlsbad, CA). Using gene specific oligonucleotide primers, a polymerase chain reaction was used on the adaptor-tagged cDNA library to clone the sequences of interest identified through BLASTX similarity to successful sequences found in *Polygonum tinctorium.*

Example 16

Identification and Isolation of Gene Encoding Sulfohydrolase atsA from *Pseudomonas aeruginosa*

[0191] The gene encoding the sulfohydrolase atsA from *P. aeruginosa* was identified to be of interest by review of the reference source listed in Table **1**. This sequence was cloned via polymerase chain reaction using as template a synthetic piece of DNA ordered from Integrated DNA Technologies, Inc. (Coralville, IA). The product of the polymerase chain reaction was cloned into an expression vector as described in Example 18.

Example 17

Identification and Isolation of Genes Encoding the Sulfohydrolase arylsulfatase3 and the Sulfohydrolase Companion anaerobic sulfatase maturase (ansME) from *Providencia stuartii*

[0192] The genes encoding the sulfohydrolase arylsulfatase3 and the sulfohydrolase companion anaerobic sulfatase maturase (ansME) from *P. stuartii* was identified to be of interest by review of the reference source listed in Table **1**. This sequence was cloned via polymerase chain reaction using as template a synthetic piece of DNA ordered from Integrated DNA Technologies, Inc. (Coralville, IA). The product of the polymerase chain reaction was cloned into an expression vector as described in Example 18.

Example 18

Construction of Recombinant DNA Plasmids

[0193] This Example demonstrates the construction of recombinant DNA plasmids for use in transforming host cells.

[0194] Plasmids were constructed using a variant of the "Golden Gate" method. In the design stage, fragments were planned such that they would have flanking BsaI, BsmBI, or BpiI type IIS restriction enzyme recognition sites. These sites were set up such that, upon cleavage with one of these type IIS restriction enzymes, fragments to be assembled together would expose matching overhangs or "sticky ends," allowing multiple distinct fragments to assemble in an ordered fashion. This assembly was achieved by mixing together, in 5 µL of water, from 50-200 ng of each individual plasmid or PCR product to contribute a fragment to the assembly. To this mixture was added 1 µL of 10X NEB ligase buffer from New England Biolabs, Inc. (Ipswich, MA), 3 µL of water, and 0.5 µL each of 3000 kU/mL T7 ligase and 10 kU/mL type IIS

restriction enzyme, from New England Biolabs, Inc. The 10 µL solution was pipetted to mix, and placed in a thermocycler to run the following temperature program: 45 °C for 2 min, then 20 °C for 5 min, repeated 25 times, followed by 60 °C for 10 min to enhance digestion of half-assembled product and 80 °C for 10 min to achieve heat inactivation of enzymes. The assembled plasmid was then transformed into the host cells.

Example 19

Transformation of *E. coli* Host Cells

[0195] This Example demonstrates the transformation of *E. coli* host cells with a plasmid for heterologous gene expression. Plasmids were constructed following the procedure described in Example 18.

[0196] A TSS buffer, 100 mL total, was prepared and filtered with a 0.2 µm filter for sterility, the TSS buffer containing of 85 mL LB media, 10 g polyethylene glycol 3350 in 5 mL H_2O, 5 mL DMSO, 3 mL H_2O, and 2 mL of 1 M $MgCl_2$. A second buffer was prepared, KCM5X, containing 500 mM KCl, 150 mM $CaCl_2$, and 250 mM $MgCl_2$ in water. The host cell *E. coli* strain was inoculated by dilution 1:1000 of 1 mL of saturated culture into 1L LB media, in a 2 L flask. This was kept shaking at 250 rpm and 37 °C until the *E. coli* had grown to an optical density at 600 nm (OD600) of 0.5. Once grown, the culture was centrifuged at 4800 rcf for 5 min, then resuspended in 100 mL of TSS buffer while kept on ice. A 50 µL aliquot of these cells were mixed with 0.5 µL of the appropriate plasmid, and 10 µL of KCM5X buffer was added. The mixture was kept on ice for 2 min, transferred to a 42 °C water bath for 90 sec, incubated at 37 °C for one h, and then the entire mixture was plated onto an LB-agar plate supplemented with the appropriate selective marker. These plates were incubated at 37 °C to allow transformed cell colonies to grow.

Example 20

Transformation of *S. cerevisiae* Host Cells

[0197] This Example demonstrates the transformation of *S. cerevisiae* host cells with a plasmid for heterologous gene expression. Plasmids were constructed following the procedure described in Example 18.

[0198] Growth media (YPD) was prepared by dissolving 20 g of glucose, 10 g of yeast extract, and 20 g of peptone in 1 L of distilled water, and then filtering the resulting mixture with a 0.2 μm filter for sterility. The host cell *S. cerevisiae* strain was inoculated by dilution 1:100 from 50 μL of saturated culture into 5 mL of YPD in a culture tube. This culture was incubated at 250 rpm shaking, 30 °C for approximately 5 hours, until an OD600 of 0.8 was reached. At this point, the cells were transferred to a centrifuge tube and spun down at 4800 rcf for 5 min. The supernatant was discarded and the pellet was resuspended with 5 mL water. The mixture was spun down again at 4800 rcf for 5 min, and the supernatant discarded and the cells resuspended with 5 mL 100 mM lithium acetate in water. The mixture was spun down again at 4800 rcf for 5 min, the supernatant discarded and the cells resuspended with 50 μL of water. To this was added 36 μL of 1 M lithium acetate in water, 240 μL of 50% w/v polyethylene glycol 3350 in water, 25 μL of 2 mg/mL salmon sperm DNA in water, and 1 μL of the plasmid of interest in water. The resulting mixture was vortexed for one minute and transferred to a 42 °C heat bath to heat shock for 30 min. After heat shock, the cell mixture was transferred to ambient temperature and centrifuged at 2000 rcf for 3 min. The supernatant was discarded, the cells were resuspended with 60 μL water, and the cell solution was plated onto SD-dropout agar plates, where one amino acid has been removed from the media to select for cells carrying an auxotrophic marker. *See* Gietz RD, Schiestl RH, High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method, *Nat Protoc* (2007), 2(1), pages 31-4. These plates were then incubated at 30 °C to allow transformed cell colonies to grow.

Example 21

Expression and Purification of Hydrolases in *E. coli* Host Cells

[0199] This Example demonstrates purification of hydrolases which were heterologously expressed in *E. coli* host cells. The *E. coli* host cells were transformed with recombinant plasmids for inducible expression of hexahistidine-tagged proteins following the procedure of Example 19. The strain BL21 DE3 was typically used for expression.

[0200] *Hydrolase expression:* A 2 L flask containing 0.5 L LB media supplemented with 100 μg/mL ampicillin was inoculated with the transformed host cells by the addition of 5 mL of saturated culture, a 1:100 dilution. This was kept shaking at 250 rpm and 37 °C until the *E. coli*

had grown to an OD600 of about 0.5. Once grown, the culture was induced to produce protein by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. The flask was then allowed to continue shaking at 250 rpm and 37 °C for 16 h.

[0201] *Hydrolase purification:* Once incubation was complete, cells were harvested by centrifugation at 4800 rcf for 5 min. The supernatant was discarded and the cells were resuspended in 35 mL of ice-cold lysis buffer containing 150 mM NaCl, 25 mM imidazole, and 25 mM Tris pH 8.0 in water. Cells were kept at 4 °C and sonicated for 30 sec at maximum amplitude to achieve lysis. Lysate was centrifuged at 24500 rcf for 10 min, and the supernatant mixed with 4 mL nickel-nitriloacetic acid agarose beads for 10 min. The beads and supernatant were packed into a column and allowed to flow through by gravity. The column was washed twice with 30 mL of fresh lysis buffer to remove any unbound protein, and the hydrolase was eluted off of the column by running 10 mL of elution buffer (150 mM NaCl, 500 mM imidazole, and 25 mM Tris pH 8.0 in water) through the column. The eluate was concentrated using a Millipore 10 kDa molecular weight-cutoff spin column and glycerol was added to a final concentration of 10% v/v, then aliquots of the hydrolase were frozen at -80 °C.

Example 22

In vivo Production of (3-(β-D-glucosido)indole) in *E. coli* Host Cells Heterologously Expressing Glucosyltransferases

[0202] This Example demonstrates the production of (3-(β-D-glucosido)indole), also known as indican, by different *E. coli* host cells expressing recombinant glucosyltransferases.

[0203] Plasmids were constructed following the procedure described in Example 18. *E. coli* host cells were transformed with recombinant plasmids according to Example 19, to produce a series of host cells expressing both the oxygenase Flavin-containing monooxygenase (FMO) and one glucosyltransferase selected from UGT72E3 from *A. thaliana*, UGT72B1 from *A. thaliana*, AHZ08761.1 from *N. tabacum*, UGT72E2 from *A. thaliana,* UGT isoform 1 from *P. tinctorium*, and UGT isoform 2 from *P. tinctorium.*

[0204] For each host cell culture, a 24-well culture block containing 3 mL per well of EZ-Rich MOPS culture medium obtained from Teknova (Hollister, CA) supplemented with 50

μg/mL spectinomycin and 5 mM L-tryptophan and 2% w/v glucose was inoculated with one recombinant strain of *E. coli* by scraping a single colony from an agar plate. Each block was kept at 37 °C with shaking at 750 rpm for 12 h.

[0205] After incubation, indican was extracted by taking 360 μL of each reaction and adding 40 μL of 50% w/v aqueous NH_4OH to achieve a final volume of 400 μL supplemented with 5% w/v NH_4OH. This mixture was allowed to sit for 5 minutes. Samples were vortexed with 800 μL of methyl *tert*-butyl ether for 10 min, centrifuged at 20000 rcf for 5 min, and the aqueous phase separated by LC/MS on a C18 column. Indican was detected via time-of-flight mass spectrometry using a negative electrospray ion source.

[0206] **FIG. 1** depicts the liquid-chromatography mass spectrometry (LC-MS) traces of each sample quantified. **FIG. 2** depicts a photograph of the host cells grown in LB supplemented with 50 μg/mL spectinomycin expressing, from right to left, UGT72E2, UGT72E3, AHZ08761.1, UGT isoform 1 from *P. tinctorium*, and UGT isoform 2 from *P. tinctorium*. The blue coloration is due to indigo production.

Example 23

Conversion of Indican Secreted from *E. coli* Host Cells to Indigo

[0207] This example demonstrates the conversion of indican (3-(β-D-glucosido)indole) secreted by a recombinant *E. coli* culture to indigo in the presence of β-glucosidase.

[0208] Plasmids were constructed following the procedure described in Example 18. *E. coli* host cells were transformed with recombinant plasmids according to Example 19 to produce host cells expressing the UGT AHZ08761.1 glucosyltransferase *Nicotiana tabacum.* A 4 mg/mL solution of β-glucosidase bglA from *B. circulans* was obtained via protein purification as according to Example 21.

[0209] For each host cell culture, a 24-well culture block containing 3 mL of LB culture medium per well supplemented with 50 μg/mL spectinomycin was inoculated with one recombinant strain of *E. coli* by scraping a single colony from an agar plate. Each block was kept at 37 °C with shaking at 750 rpm for 12 h.

[0210] A 50 μL aliquot of 4 mg/mL β-glucosidase solution was then added to the culture and allowed to sit at ambient temperature for 20.5 h. **FIG. 3** depicts photographs of the culture taken immediately after the addition of β-glucosidase (left), 75 min after the addition of β-glucosidase (middle), and 20.5 h after the addition of β-glucosidase (right). Without wishing to be bound by any theory, this demonstrates that indican is being made and secreted by the recombinant *E. coli*, and the indican remains accessible to a β-glucosidase. The presence and accessibility of indican in the culture media allows for indicates the media may be used to dye objects directly.

Example 24

In vivo Production of 3-(β-D-glucosido)indole in Yeast Strains Heterologously Expressing an Oxygenase and Different Glucosyltransferases

[0211] This example demonstrates the *in vivo* production of 3-(β-D-glucosido)indole, also known as indican, by cultures of *Saccharomyces cerevisiae* heterologously expressing the oxygenase cytochrome P450 CYP102A1 from *Bacillus megaterium* and different glucosyltransferases, when supplemented with indole. This Example also demonstrates the effect that optional coexpression of UDP-glucose synthesis enzymes UGP1 and PGM2 from *S. cerevisiae*, and optional coexpression of the oligosaccharide transporter CDT1 from *Neurospora crassa* have on indican production.

[0212] Plasmids were constructed following the procedure described in Example 18. *S. cerevisiae* host cells were transformed with recombinant plasmids following the procedure described in Example 20, to produce a series of host cells expressing the oxygenase CYP102A1 and one glucosyltransferase selected from UGT72B1 from *A. thaliana*, UGT AHZ08761.1 from *N. tabacum*, UGT isoform 1 from *P. tinctorium*, and UGT isoform 2 (mut) from *P. tinctorium*; optionally overexpressing UGP1 and PGM2 from *S. cerevisiae*; and optionally overexpressing CDT1 *N. crassa*.

[0213] Two control host cell cultures were also produced, one expressing CYP102A1 with no recombinant glucosyltransferase, and one expressing no recombinant CYP102A1 or glucosyltransferase. Indole was obtained from Sigma-Aldrich (St. Louis, MO).

[0214] For each host cell culture, one well in a 24-well culture block containing 3 mL of SD-Uracil dropout medium per well was inoculated with the recombinant strain of *S. cerevisiae* by scraping colony growth off of an agar plate. The block was kept at 30 °C with shaking at 750 rpm overnight. The next day, the 3 mL cell solution was centrifuged at 4800 rcf for 5 min. Cells were resuspended in 600 µL phosphate buffered saline, pH 7.4, supplemented with 1.26 mM glucose; of this, 100 µL was diluted into 3000 µL phosphate buffered saline, pH 7.4, supplemented with 1.26 mM glucose, 1% dimethyl sulfoxide, and 5 mM indole. 500 µL of this mixture was transferred to a 96-well block, which was then incubated at 30 °C with shaking at 750 rpm for 6 h.

[0215] After incubation, indican was extracted by taking 360 µL of each reaction and adding 40 µL of 50% w/v aqueous NH_4OH to achieve a final volume of 400 µL supplemented with 5% w/v NH_4OH. This mixture was allowed to sit for 5 min. Samples were then vortexed with 800 µL of methyl *tert*-butyl ether for 10 min, centrifuged at 20000 rcf for 5 min, and the aqueous phase separated by LC/MS on a C18 column. Indican was detected via time-of-flight mass spectrometry using a negative electrospray ion source.

[0216] **FIG. 4** depicts a graph comparing the concentration of indican of each culture expressing a glucosyltransferase, and additionally with and without expression of both UGP1 and PGM2 ("UDP-Glc"), and the transporter CDT1. These results show that expression of UGP1, PGM2 and CDT1 lead to higher levels of indican production. This indicates that both UGP1 and PGM2, and the oligosaccharide transporter CDT1 contribute to indican production in *S. cerevisiae*.

Example 25

Use of Indigo and Indican in the Presence of *Bacillus circulans* β-glucosidase to Dye Cloth

[0217] This example demonstrates the level of cloth dyeing obtained when using indican (3-(β-D-glucosido)indole) or indigo in the presence of the β-glucosidase bglA from *Bacillus circulans*.

[0218] Indican and indigo were purchased from Sigma-Aldrich (St. Louis, MO). A 4 mg/mL solution of β-glucosidase bglA from *B. circulans* was obtained via protein purification according to Example 21.

[0219] Separate solutions of 10 g/L indigo or indican were prepared by suspension in water; the indigo remained in microscale crystals, while the indican dissolved. A 1 square inch piece of undyed cotton cloth was added to a 5 mL Eppendorf tube containing either the indigo solution or the indican solution. The cloth was allowed to soak in the liquid for 10 min at ambient temperature. The cloth pieces were removed from the solution and laid flat on wax paper. Then, 100 μL of a 4 mg/mL β-glucosidase solution was pipetted unevenly across each cloth piece. The cloth pieces were incubated at ambient temperature for 6 h, then washed repeatedly with water and allowed to dry.

[0220] **FIG. 5** shows the appearance of the cloth pieces after incubation with indigo or indican (top), the cloth pieces 5 min after application of β-glucosidase (middle), and the degree of color remaining after washing each piece of cloth (bottom). These results indicate that enzymatic hydrolysis of indican on fabric can be used to generate indigo dyeing that remains even after repeated washing.

Example 26

In vitro Production of Indigoid Dyes in the Presence of a Sulfatase

[0221] This example demonstrates the *in vitro* conversion of sulfated indigoid dye precursors to indigoid dyes in the presence of the sulfatase atsA from *P. aeruginosa*.

[0222] The sulfated dye precursor 1*H*-indol-3-yl sulfate was obtained from Sigma-Aldrich (St. Louis, MO) and dissolved in water to obtain a stock concentration of 10 mM. The sulfated dye precursor 5-bromo-4-chloro-1*H*-indol-3-yl sulfate was obtained from Sigma-Aldrich (St. Louis, MO) and dissolved in dimethyl sulfoxide (DMSO) to obtain a stock concentration of 10 mM.

[0223] Following the procedure described in Example 19, *E. coli* host cells were transformed with plasmids expressing the hydrolase atsA from *Pseudomonas aeruginosa*. These transformed

cells were grown a tube filled with 5 mL LB medium supplemented with 100 µg/mL ampicillin, and left shaking at 250 rpm, 37 °C, overnight. The next day, the 5 mL *E. coli* saturated culture was centrifuged at 4800 rcf for 5 min. Spent media supernatant was removed and saved for usage in the assay as "supernatant."

[0224] To two wells in a plastic 96-well plate were added 20 µL of a 10 mM solution of one of the sulfated dye precursor substrates. To the 5-bromo-4-chloro-1*H*-indol-3-yl sulfate containing well was added 20 µL of deionized water. To the f 1*H*-indol-3-yl sulfate containing well was added 20 µL of dimethyl sulfoxide (DMSO). To this mixture, 160 µL of supernatant was added for a final concentration of 10% DMSO and 1 mM substrate. During this procedure, the plate was kept cold until all reagents had been added and mixed.The mixtures were then allowed to incubate for 90 min at ambient temperature, and monitored visually for production of indigoid dyes over time. **FIG. 6** depicts images of each well taken over time, showing the production of indigoid dyes from indigoid dye precursors in the presence of the sulfatase atsA.

Example 27

In vivo Production of 1*H*-indol-3-yl Sulfate by *E. coli* Heterologously Expressing Different Sulfatases

[0225] This example demonstrates the production of 1*H*-indol-3-yl sulfate by *E. coli* cultures expressing different recombinant sulfatases and supplemented with 1*H*-indol-3-yl acetate and 3'-phosphoadenosine-5'-phosphosulfate (PAPS).

[0226] The compounds 1*H*-indol-3-yl acetate and 3'-phosphoadenosine-5'-phosphosulfate (PAPS).were obtained from Sigma-Aldrich (St. Louis, MO). *E. coli* BL21 DE3 host cells were transformed with recombinant plasmids according to Example 19 to produce host cells inducibly expressing SULT1A1*1 wildtype from *Homo sapiens*, SULT1A1*1 D249G from *Homo sapiens,* SULT1A1*1 enh1 from *Homo sapiens*, SULT1A3*1 wildtype from *Homo sapiens*, SULT1A3*1 D249G from *Homo sapiens*, or SULT1A3*1 enh1 from *Homo sapiens*.

[0227] For each host cell strain, a well in a 24-well culture block was filled with 3 mL of LB culture medium supplemented with 34 µg/mL chloramphenicol and inoculated by 1:100 dilution of saturated overnight culture with one recombinant strain of *E. coli*. The block was incubated

while shaking at 750 rpm at 37 °C for 3 h, after which point isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 0.5 mM. The block was then incubated for another 16 h, with shaking at 750 rpm and at a temperature of 37 °C. At this point, the block was centrifuged at 4800 rcf for 3 min to collect the cells. The supernatant was discarded and the cell pellets were transferred into an anaerobic environment, where they were each resuspended with 600 µL of LB supplemented with 34 µg/mL chloramphenicol, concentrating the cells five-fold. 58.5 µL of this cell resuspension was supplemented, bringing the total volume to 75 µL and a final concentration of 2 mM 1*H*-indol-3-yl acetate, 2% v/v dimethyl sulfoxide and 2 mM PAPS. This mixture was sealed and incubated for 24 h at 37 °C, to allow native esterase activity of the cells to hydrolyze 1*H*-indol-3-yl acetate to indoxyl. At the end of the incubation, the mixture was exposed to air to oxidize any unprotected indoxyl.

[0228] Following exposure to air, the quantity of 1*H*-indol-3-yl sulfate produced was measured by extraction and LC/MS. To extract the 1*H*-indol-3-yl sulfate, the 75 µL reaction was diluted and aqueous NH_4OH added to achieve a final volume of 400 µL supplemented with 5% w/v NH_4OH. This mixture was allowed to sit for 5 min. Samples were then vortexed with 800 µL of methyl *tert*-butyl ether for 10 min, centrifuged at 20000 rcf for 5 min, and the aqueous phase separated by LC/MS on a C18 column. 1*H*-indol-3-yl sulfate was detected via time-of-flight mass spectrometry using a negative electrospray ion source. **FIG. 7** depicts a graph of the concentration of 1*H*-indol-3-yl sulfate produced by each strain of *E. coli*.

Example 28

Comparison of Different Hydrolases in the *In vitro* Production of Dyes from Sulfated Dye Precursors

[0229] This example demonstrates the use of four different hydrolase enzymes in the *in vitro* conversion of sulfated dye precursors to dyes.

[0230] The sulfated dye precursors 1*H*-indol-3-yl sulfate and 4-nitrophenyl sulfate were obtained from Sigma-Aldrich (St. Louis, MO) and dissolved in water to obtain stock concentrations of 10 mM. The sulfated dye precursor 5-bromo-4-chloro-1*H*-indol-3-yl sulfate was obtained from Sigma-Aldrich (St. Louis, MO) and dissolved in dimethyl sulfoxide (DMSO)

to obtain a stock concentration of 10 mM. The sulfated dye precursor 4-nitrocatechol sulfate was dissolved in water to obtain a stock concentration of 1 mM.

[0231] Following the procedure described in Example 19, *E. coli* host cells were transformed with plasmids to express the hydrolases yidJ from *Escherichia coli*, atsA from *Pseudomonas aeruginosa*, or both arylsulfatase3 and the companion maturase anSME from *Providencia stuartii*. The transformed cells were grown in tubes filled with 5 mL LB medium supplemented with 100 µg/mL ampicillin with shaking at 250 rpm, 37 °C, overnight. The next day, the 5 mL *E. coli* saturated cultures were centrifuged at 4800 rcf for 5 min. Spent media supernatant was removed and saved for usage in the assay as "media." Pellets were resuspended with 3 mL 1.25X phosphate buffered saline (PBS) pH 7.4 and then centrifuged again. This supernatant was discarded and the pellets were once more resuspended with 3 mL 1.25X phosphate buffered saline (PBS) pH 7.4. Of this 3 mL of PBS-washed cells, 1.5 mL was kept for usage as "cell solution." The final remaining 1.5 mL of cells was lysed by adding 160 uL of 10X Bugbuster HT Protein Extraction Reagent obtained from EMD/Millipore-Merck KGaA (Darmstadt, Germany). In this way, the final component, "cell lysate," was produced.

[0232] To a plastic 96-well plate was added 20 µL of one of the prepared sulfated dye precursor substrate solutions. To this solution was added 20 µL of deionized water for the substrate 5-bromo-4-chloro-1*H*-indol-3-yl sulfate, or dimethyl sulfoxide (DMSO) for all other substrates. To this mixture, 160 µL of cell solution, supernatant, or cell lysate was added for a final concentration of 10% DMSO and 0.01 mM of 4-nitrocatechol sulfate, or 1 mM of all other substrates. The plate was incubated for 2 h at room temperature, followed by 17 h at 37 °C, at which point the qualitative production of indigoid dyes was visually determined. **FIG. 8** depicts a photograph of the wells taken at the 19-hour point, with indigoid dye production indicated by blue coloration and other dye production indicated by yellow coloration. These results indicate that *E. coli* strain TG1 naturally lacks significant sulfate hydrolase activity, but the hydrolases yidJ from *Escherichia coli*, atsA from *Pseudomonas aeruginosa*, or arylsulfatase3 and companion maturase anSME from *Providencia stuartii* can cleave sulfated dye precursors.

Example 29

Production of Halogenated Indigoid Dyes by *E. coli* Host Cells Expressing FMO

[0233] This example demonstrates the *in vivo* production of halogenated indigoid dyes by *E. coli* host cells expressing a Flavin-containing monooxygenase (FMO).

[0234] The halogenated indoles 5-chloro-1*H*-indol-3-ol, 6-chloro-1*H*-indol-3-ol, 7-chloro-1*H*-indol-3-ol, 5-bromo-1*H*-indol-3-ol, 6-bromo-1*H*-indol-3-ol, and 7-bromo-1*H*-indol-3-ol were obtained from Sigma-Aldrich (St. Louis, MO). *E. coli* host cells were transformed with recombinant plasmids according to Example 19 to produce host cells expressing the FMO from *Methylophaga* sp. strain SK1.

[0235] For each halogenated indole, one well in a 24-well culture block containing 2 mL of LB culture medium per well supplemented with 100 μg/mL ampicillin was inoculated with the recombinant strain of *E. coli* by 1:1000 dilution of saturated culture. The block was kept at 37 °C with shaking at 750 rpm overnight. The next day, the 2 mL cell solution was centrifuged at 4800 rcf for five minutes. Cells were resuspended in 400 μL phosphate buffered saline, pH 7.4, supplemented with 1.26 mM glucose, 1% dimethyl sulfoxide, and 1 mM of a halogenated indole. Cells were incubated at 37C for 16 h, resuspended, and 200 μL was transferred to 8-well strip tubes. Pellets were collected by centrifugation at 4800 rcf for 5 min, at which point a photograph was taken. **FIG. 9** depicts the color of the indigoid dyes obtained. These color results match reference images of these dyes, demonstrating the presence of halogenated indigoid dyes, indicating the *E. coli* whole-cell catalyzed conversion of precursors to dyes was successful.

Example 30

Production of Indigo Dye by *E. coli* Host Cells Expressing FMO

[0236] This example demonstrates the *in vivo* production of indigo dye by *E. coli* host cells expressing a Flavin-containing monooxygenase (FMO). *E. coli* host cells were transformed with recombinant plasmids according to Example 19 to produce host cells expressing the FMO from *Methylophaga* sp. strain SK1.

[0237] To a 250 mL flask was added 50 mL LB culture medium supplemented with 100 μg/mL ampicillin and 3.7 mM L-tryptophan, and inoculated by 1:1000 dilution of saturated

culture of a recombinant strain of *E. coli*. The recombinant strain of *E. coli* was constitutively expressing the FMO from *Methylophaga* sp. strain SK1. The flask was kept at 37 °C with shaking at 250 rpm for 24 h. **FIG. 10** depicts the color of the indigo dyes obtained. These results demonstrate facile microbial biosynthesis of indigo dye from host cells heterologously expressing an FMO.

Example 31

This Example illustrates that host cells having knockouts in certain genes exhibit lower rates of hydrolysis of indoxyl acetate.

Methods

[0238] 5xKO (Δaes, ΔyjfP, ΔbioH, ΔyeiG, ΔfrmB) and 11xKO (Δaes, ΔyjfP, ΔbioH, ΔyeiG, ΔfrmB, ΔentH, ΔydiL, ΔtesA, ΔnanS, ΔyqiA, Δybf) knockouts of *E. coli* strain MG1655 were generated using the MAGE technique (*See* Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM, Programming cells by multiplex genome engineering and accelerated evolution, *Nature* (2009), 460, pages 894-898.) using oligonucleotide primers designed using the MODEST software package (*See* Bonde MT, Klausen MS, Anderson MV, Wallin AIN, Wang HH, Sommer MOA, MODEST: a web-based design tool for oligonucleotide-mediated genome engineering and recombineering, *Nucl Acids Res*, (2014), 42, pages W408-W415) and validated using polymerase chain reaction.

Gene	**Ordered Locus Name**	**Extended Name**
aes	b0476	Acetyl esterase
yjfP	b4190	Esterase yjfP
bioH	b3412	Pimeloyl-[acyl-carrier protein] methyl ester esterase
yeiG	b2154	S-formylglutathione hydrolase YeiG
frmB	b0355	S-formylglutathione hydrolase FrmB
entH	b0597	Proofreading thioesterase EntH
ydiL	b1689	Uncharacterized protein YdiL
tesA	b0494	Acyl-CoA thioesterase I
nanS	b4309	Probable 9-O-acetyl-N-acetylneuraminic acid deacetylase

yqiA	b3031	Esterase YqiA
ybfF	b0686	Esterase YbfF
ypfH	b2473	Esterase YpfH
argE	b3957	Acetylornithine deacetylase
frsA	b0239	Esterase FrsA
ybgC	b0736	Acyl-CoA thioester hydrolase YbgC
tnaA	b3708	Tryptophanase

[0239] Bacterial colonies (wild type or knockouts) were grown in LB for 16 hours at 37C (200 rpm shaking). Cells were pelleted and resuspended to a concentration of OD600 = 2 in phosphate buffered saline pH 7 with 5% DMSO and 1 mM indoxyl acetate added. Indoxyl acetate hydrolysis rate was then determined by measuring the increase in free indoxyl fluorescence (excitation 410 nm, emission 490 nm) for 30 minutes. Fluorescence increases were observed to be linear.

Results and Discussion

[0240] In **FIG. 11**, 5xKO (Δaes, ΔyjfP, ΔbioH, ΔyeiG, ΔfrmB and 11xKO (Δaes, ΔyjfP, ΔbioH, ΔyeiG, ΔfrmB, ΔentH, ΔydiL, ΔtesA, ΔnanS, ΔyqiA, ΔybfF), exhibit decreased rates of indoxyl acetate hydrolysis as compared to a WT cell. Thus, it appears that knockouts of the following genes enhance stability of compounds containing an acetate moiety (including indoxyl acetate): Δaes, ΔyjfP, ΔbioH, ΔyeiG, ΔfrmB, ΔentH, ΔydiL, ΔtesA, ΔnanS, ΔyqiA, ΔybfF, ΔypfH, ΔargE, ΔfrsA, ΔybgC, ΔtnaA. This enhanced stability can contribute to higher product yields in fermentative processes.

Table 1: Sequences

SEQ ID NO:	Name	Function	Amino Acid Sequence	Reference/ Source
1	*M. sp.* strain SK1 FMO	oxygenase	MATRIAILGAGPSGMAQLRAFQSAQEKGAEIPELV CFEKQADWGGQWNYTWRTGLDENGEPVHSSMY RYLWSNGPKECLEFADYTFDEHFGKPIASYPPREV LWDYIKGRVEKAGVRKYIRFNTAVRHVEFNEDSQT FTVTVQDHTTDTIYSEEFDYVVCCTGHFSTPYVPEF EGFEKFGGRILHAHDFRDALEFKDKTVLLVGSSYS AEDIGSQCYKYGAKKLISCYRTAPMGYKWPENWD ERPNLVRVDTENAYFADGSSEKVDAIILCTGYIHHF PFLNDDLRLVTNNRLWPLNLYKGVVWEDNPKFFYI GMQDQWYSFNMFDAQAWYARDVIMGRLPLPSKE EMKADSMAWREKELTLVTAEEMYTYQGDYIQNLID MTDYPSFDIPATNKTFLEWKHHKKENIMTFRDHSY RSLMTGTMAPKHHTPWIDALDDSLEAYLSDKSEIP VAKEA	Choi HS, Kim JK, Cho EH, Kim YC, Kim JI, Kim SW. A novel flavin-containing monooxygenase from Methylophaga sp strain SK1 and its indigo synthesis in Escherichia coli. Biochem Biophys Res Commun. 2003 Jul 11;306(4):930-6.
2	*B. megaterium* CYP102A1 G4 (P450)	oxygenase	MTIKEMPQPKTFGELKNLPLLNTDKPVQALMKIADE LGEIFKFEAPGLVTRFLSSQRLIKEACDESRFDKNL SQALKFVRDFAGDGLATSWTHEKNWKKAHNILLPS FSQQAMKGYHAMMVDIAVQLVQKWERLNADEHIE VPEDMTRLTLDTIGLCGFNYRFNSFYRDQPHPFITS MVRALDEAMNKLQRANPDDPAYDENKRQFQEDIK VMNDLVDKIIADRKASGEQSDDLLTHMLNGKDPET GEPLDDENIRYQIITFLIAGHETTSGLLSFALYFLVKN PHVLQKAAEEAARVLVDPVPSYKQVKQLKYVGMV LNEALRLWPTLPAFSLYAKEDTVLGGEYPLEKGDE LMVLIPQLHRDKTIWGDDVEEFRPERFENPSAIPQH AFKPFGNGQRACIGQQFALHEATLVLGMMLKHFDF EDHTNYELDIKETLTLKPEGFVVKAKSKKIPLGGIPS PSTEQSAKKVRKKAENAHNTPLLVLYGSNMGTAE GTARDLADIAMSKGFAPQVATLDSHAGNLPREGAV LIVTASYNGHPPDNAKQFVDWLDQASADEVKGVR YSVFGCGDKNWATTYQKVPAFIDETLAAKGAENIA DRGEADASDDFEGTYEEWREHMWSDVAAYFNLDI ENSEDNKSTLSLQFVDSAADMPLAKMHGAFSTNV VASKELQQPGSARSTRHLEIELPKEASYQEGDHLG VIPRNYEGIVNRVTARFGLDASQQIRLEAEEEKLAH LPLAKTVSVEELLQYVELQDPVTRTQLRAMAAKTV CPPHKVELEALLEKQAYKEQVLAKRLTMLELLEKYP ACEMKFSEFIALLPSIRPRYYSISSSPRVDEKQASIT VSVVSGEAWSGYGEYKGIASNYLAELQEGDTITCFI STPQSEFTLPKDPETPLIMVGPGTGVAPFRGFVQA RKQLKEQGQSLGEAHLYFGCRSPHEDYLYQEELE NAQSEGIITLHTAFSRMPNQPKTYVQHVMEQDGKK LIELLDQGAHFYICGDGSQMAPAVEATLMKSYADV HQVSEADARLWLQQLEEKGRYAKDVWAG	Dietrich JA, Yoshikuni Y, Fisher KJ, Woolard FX, Ockey D, McPhee DJ, Renninger NS, Chang MC, Baker D, Keasling JD. A novel semi-biosynthetic route for artemisinin production using engineered substrate-promiscuous P450(BM3). ACS Chem Biol. 2009 Apr 17;4(4):261-7. doi: 10.1021/cb900006h.

3	*P. tinctorium* UGT isoform 1	glucosyl-transferase	MESPAAPPTTAPPPHVIIVPSAGMGHLIPLAEFAKR LLPRFTFTFAVPTSGPPSSSQRDFLSSLPASIDTSF LPEVDLSDAPSDAQIETLMSLMVVRSLPSLRDLIAS YSASGRRVAALVVDLFATDAIDVALELGIRPFIFFPS TAMTLSFFLHLEKLDETVSCEFAELSDPVQIPGCIP VHGKDLIDPVQDRKNDAYKWLLHHSKRYKLAEGVI VNSFEGLEGGPIRELLHPEPGKPRVYPVGPLIQAG SCEKGAAARPECLKWLDQQPRGSVLFVNFGSGGV LSTEQQNELAGVLAHSQQRFLWVVRPPNDGIANA TYFSVDGEIDPLKLLPEGFLEQTAGRGLVLPMWAP QIDVLSHESTGGFLTHCGWNSTLESVFHGVPLITW PLYAEQKMNAVMLTEGLRVGLRPSVGKDGIIRGAEI ARVIGELMEGEEGKRIRSKMQELKRAASAVLSKDG SSTRALEEVAKIWESKV	Example 1
4	*P. tinctorium* UGT isoform 2	glucosyl-transferase	MESPAAPPTTAPPPHVIIMPSAGMGHLIPLAEFAKR LLPRFTFTFAVPTSGPPSSSQRDFLSSLPASIDTSF LPEVDLSDAPSDAQIETLMSLMVVRSLPSLRDLIAS YSASGRRVAALVVDLFATDAIDVALELGIRPFIFFPS TAMTLSFFLHLEKLDETVSCEFAELSDPVQIPGCIP VHGKDLIDPVQDRKNDAYKWLLHHSKRYKLAEGVI VNSFEGLEAGPIRQLLHPEPGKPRVYPVGPLIQAG SCEKGAAARPECLKWLDQQPRGSVLFVNFGSGGV LSTEQQNELAGVLAHSQQRFLWVVRPPNDGIANA TYFSVDGEIDPLKLLPEGFLEQTAGRGLVLPMWAP QIDVLSHESTGGFLTHCGWNSTLESVFHGVPLITW PLYAEQKMNAVMLTEGLRVGLRPSVGKDGIIRGDE IARVIGELMEGEEGKRIRSKMQELKRAASAVLSKD GSSTRALEEVAKIWESKV	Example 2
5	*N. tabacum* GenBank AHZ08761.1	glucosyl-transferase	MAETAIVTKSENPHIVILPSPGMGHLIPLVEFSKRLIS QHQFSVTLILPTDGPISNSQKSFLNSLPSCMDYHLL PPVNFDDLPLDVKIETRISLTVTRSLSSLREVFKTLV DSKKVVAFVVDLFGTDAFDVAIDFNVSPYIFFPSTA MALSLFLYLPKLDATVSCEYRDLPDPIQIPGCIPIHG KDLLDPVQDRKNEAYRWLLHHSKRYRMAEGVVSN SFKELEGGPIKALQEEEPGKPPVYPVGPLIQMDSG SKVDGSGCLTWLDEQPRGSVLYVSYGSGGTLSHE QLIEVASGLEMSEQRFLWVIRCPNDTVANATYFNV QDSTNPLDFLPKGFLERTKGLGLVVPNWAPQAQIL SHGSTGGFLTHCGWNSTLESVVHGVPLIAWPLYA EQKMNAVMLTEDIKVALRPKANENGLVGRLEIAKV VKGLMEGEEGKGVRTRMRDLKDAAAKVLSQDGSS TKALAELATKLKNKVLIN	Example 3

6	*A. thaliana* UGT72B1	glucosyl-transferase	MEESKTPHVAIIPSPGMGHLIPLVEFAKRLVHLHGL TVTFVIAGEGPPSKAQRTVLDSLPSSISSVFLPPVD LTDLSSSTRIESRISLTVTRSNPELRKVFDSFVEGG RLPTALVVDLFGTDAFDVAVEFHVPPYIFYPTTANV LSFFLHLPKLDETVSCEFRELTEPLMLPGCVPVAGK DFLDPAQDRKDDAYKWLLHNTKRYKEAEGILVNTF FELEPNAIKALQEPGLDKPPVYPVGPLVNIGKQEAK QTEESECLKWLDNQPLGSVLYVSFGSGGTLTCEQ LNELALGLADSEQRFLWVIRSPSGIANSSYFDSHS QTDPLTFLPPGFLERTKKRGFVIPFWAPQAQVLAH PSTGGFLTHCGWNSTLESVVSGIPLIAWPLYAEQK MNAVLLSEDIRAALRPRAGDDGLVRREEVARVVKG LMEGEEGKGVRNKMKELKEAACRVLKDDGTSTKA LSLVALKWKAHKKELEQNGNH	Example 4
7	*A. thaliana* UGT72E2	glucosyl-transferase	MHITKPHAAMFSSPGMGHVIPVIELGKRLSANNGF HVTVFVLETDAASAQSKFLNSTGVDIVKLPSPDIYG LVDPDDHVVTKIGVIMRAAVPALRSKIAAMHQKPTA LIVDLFGTDALCLAKEFNMLSYVFIPTNARFLGVSIY YPNLDKDIKEEHTVQRNPLAIPGCEPVRFEDTLDAY LVPDEPVYRDFVRHGLAYPKADGILVNTWEEMEPK SLKSLLNPKLLGRVARVPVYPIGPLCRPIQSSETDH PVLDWLNEQPNESVLYISFGSGGCLSAKQLTELAW GLEQSQQRFVWVVRPPVDGSCCSEYVSANGGGT EDNTPEYLPEGFVSRTSDRGFVVPSWAPQAEILSH RAVGGFLTHCGWSSTLESVVGGVPMIAWPLFAEQ NMNAALLSDELGIAVRLDDPKEDISRWKIEALVRKV MTEKEGEAMRRKVKKLRDSAEMSLSIDGGGLAHE SLCRVTKECQRFLERVVDLSRGA	Example 5
8	*A. thaliana* UGT72E3	glucosyl-transferase	MHITKPHAAMFSSPGMGHVLPVIELAKRLSANHGF HVTVFVLETDAASVQSKLLNSTGVDIVNLPSPDISG LVDPNAHVVTKIGVIMREAVPTLRSKIVAMHQNPTA LIIDLFGTDALCLAAELNMLTYVFIASNARYLGVSIYY PTLDEVIKEEHTVQRKPLTIPGCEPVRFEDIMDAYL VPDEPVYHDLVRHCLAYPKADGILVNTWEEMEPKS LKSLQDPKLLGRVARVPVYPVGPLCRPIQSSTTDH PVFDWLNKQPNESVLYISFGSGGSLTAQQLTELAW GLEESQQRFIWVVRPPVDGSSCSDYFSAKGGVTK DNTPEYLPEGFVTRTCDRGFMIPSWAPQAEILAHQ AVGGFLTHCGWSSTLESVLCGVPMIAWPLFAEQN MNAALLSDELGISVRVDDPKEAISRSKIEAMVRKVM AEDEGEEMRRKVKKLRDTAEMSLSIHGGGSAHES LCRVTKECQRFLECVGDLGRGA	Example 6
9	*H. sapiens* SULT1A1* 1 wild type	sulfotrans-ferase	MELIQDTSRPPLEYVKGVPLIKYFAEALGPLQSFQA RPDDLLISTYPKSGTTWVSQILDMIYQGGDLEKCH RAPIFMRVPFLEFKAPGIPSGMETLKDTPAPRLLKT HLPLALLPQTLLDQKVKVVYVARNAKDVAVSYYHF YHMAKVHPEPGTWDSFLEKFMVGEVSYGSWYQH VQEWWELSRTHPVLYLFYEDMKENPKREIQKILEF VGRSLPEETVDFVVQHTSFKEMKKNPMTNYTTVP QEFMDHSISPFMRKGMAGDWKTTFTVAQNERFDA DYAEKMAGCSLSFRSEL	Banoglu E, King RS. Sulfation of indoxyl by human and rat aryl (phenol) sulfotransferases to form indoxyl sulfate. Eur J Drug Metab Pharmacokine

				t. 2002 Apr-Jun;27(2):135-40.
10	*H. sapiens* SULT1A1* 1 D249G	sulfotrans-ferase	MELIQDTSRPPLEYVKGVPLIKYFAEALGPLQSFQA RPDDLLISTYPKSGTTWVSQILDMIYQGGDLEKCH RAPIFMRVPFLEFKAPGIPSGMETLKDTPAPRLLKT HLPLALLPQTLLDQKVKVVYVARNAKDVAVSYYHF YHMAKVHPEPGTWDSFLEKFMVGEVSYGSWYQH VQEWWELSRTHPVLYLFYEDMKENPKREIQKILEF VGRSLPEETVDFVVQHTSFKEMKKNPMTNYTTVP QEFMGHSISPFMRKGMAGDWKTTFTVAQNERFDA DYAEKMAGCSLSFRSEL	Berger I, Guttman C, Amar D, Zarivach R, Aharoni A (2011) The Molecular Basis for the Broad Substrate Specificity of Human Sulfotransferase 1A1. PLoS ONE 6(11): e26794. doi:10.1371/journal.pone.0026794
11	*H. sapiens* SULT1A1* 1 enh1	sulfotrans-ferase	MELIQDTSRPPLEYVKGVPLIKYFAEALGPLQSFQA RPDDLLISTYPKSGTTWVSEILDMIYQGGDVEKCH RAPIFMRVPFLEFKAPGIPSGMETLKDTPSPRLLKT HLPLALLPQSLLDQKVKVVYVARNAKDVAVSYYHF YHMAKVHPEPGTWDSFLEKFMVGEVSYGSWYQH VQEWWELSRTHPVLYLFYEDMKENPKREIQKILEF VGRSLPEETVDLVVQHTSFKEMKKNPMTNYTTIPQ EFMGHSISPFMRKGMAGDWKTTFTVAQNERFDAD YAEKMAGCSLSFRSEL	Berger I, Guttman C, Amar D, Zarivach R, Aharoni A (2011) The Molecular Basis for the Broad Substrate Specificity of Human Sulfotransferase 1A1. PLoS ONE 6(11): e26794. doi:10.1371/journal.pone.0026794

12	*H. sapiens* SULT1A3* 1 wild type	sulfotrans-ferase	MELIQDTSRPPLEYVKGVPLIKYFAEALGPLQSFQA RPDDLLINTYPKSGTTWVSQILDMIYQGGDLEKCN RAPIYVRVPFLEVNDPGEPSGLETLKDTPPPRLIKS HLPLALLPQTLLDQKVKVVYVARNPKDVAVSYYHF HRMEKAHPEPGTWDSFLEKFMAGEVSYGSWYQH VQEWWELSRTHPVLYLFYEDMKENPKREIQKILEF VGRSLPEETMDFMVQHTSFKEMKKNPMTNYTTVP QELMDHSISPFMRKGMAGDWKTTFTVAQNERFDA DYAEKMAGCSLSFRSEL	Banoglu E, King RS. Sulfation of indoxyl by human and rat aryl (phenol) sulfotransferases to form indoxyl sulfate. Eur J Drug Metab Pharmacokinet. 2002 Apr-Jun;27(2):135-40.
13	*H. sapiens* SULT1A3* 1 D249G	sulfotrans-ferase	MELIQDTSRPPLEYVKGVPLIKYFAEALGPLQSFQA RPDDLLINTYPKSGTTWVSQILDMIYQGGDLEKCN RAPIYVRVPFLEVNDPGEPSGLETLKDTPPPRLIKS HLPLALLPQTLLDQKVKVVYVARNPKDVAVSYYHF HRMEKAHPEPGTWDSFLEKFMAGEVSYGSWYQH VQEWWELSRTHPVLYLFYEDMKENPKREIQKILEF VGRSLPEETMDFMVQHTSFKEMKKNPMTNYTTVP QELMGHSISPFMRKGMAGDWKTTFTVAQNERFDA DYAEKMAGCSLSFRSEL	Berger I, Guttman C, Amar D, Zarivach R, Aharoni A (2011) The Molecular Basis for the Broad Substrate Specificity of Human Sulfotransferase 1A1. PLoS ONE 6(11): e26794. doi:10.1371/journal.pone.0026794
14	*H. sapiens* SULT1A3* 1 enh1	sulfotrans-ferase	MELIQDTSRPPLEYVKGVPLIKYFAEALGPLQSFQA RPDDLLINTYPKSGTTWVSEILDMIYQGGDVEKCN RAPIYVRVPFLEVNDPGEPSGLETLKDTPPPRLIKS HLPLALLPQSLLDQKVKVVYVARNPKDVAVSYYHF HRMEKAHPEPGTWDSFLEKFMAGEVSYGSWYQH VQEWWELSRTHPVLYLFYEDMKENPKREIQKILEF VGRSLPEETMDLMVQHTSFKEMKKNPMTNYTTIPQ ELMGHSISPFMRKGMAGDWKTTFTVAQNERFDAD YAEKMAGCSLSFRSEL	Berger I, Guttman C, Amar D, Zarivach R, Aharoni A (2011) The Molecular Basis for the Broad Substrate Specificity of Human Sulfotransferase 1A1. PLoS ONE 6(11): e26794. doi:10.1371/journal.pone.0026794

15	*I. tinctoria* UGT1	glucosyl-transferase	NLAVTFIIPTDGPPSKAQKTVLHSLPPAISHTFLPPV NLSDVPKDAKIETIISLTVLRSLPSIRDLFRSLTASAL VVDLFGTDAFDVAKEFNVSPYIFFPSTAMALSFFLH LPHLDQEVHSEYRELAEPVKIPGCVPIHGKDLLDPV QDRKNDAYKWVLHHTKRYREAEGIIENSFLELEPG PIKELLKEEPGKPPVYSVGPLVNVETGRAGNGSEC LKWLDEQPPGSVLFVSFGSGGTLSSAQINELALGL EASEQRFLWVVRSPNDKVANASYFSADSQADPFD FLPKEFVKRTKERGLVVSSWAPQTQVLAHGSTGG FLTHCGWNSILESVVNGVPLIAWPL	Example 15
16	*I. tinctoria* UGT2	glucosyl-transferase	ELNILSYLYFPSTAMLLSLCLYSSKLDKEISIEYKDLL EPIKLPGCIPISPSDLPDPLQDRSGESYQQFLEANE RFYLADGILVNSFVEMEGGTIRALQEEESRGIPSVY AIGPFVKMGSCSCDDYEGSEKDNYLTWLDKQEKC SILYVSFGSGGTLFHDQIIELAWGLELSGQKFLWVL RPPSKFGIVADLSAVNLDPLQFLPSGFLERTKGQG LVVPYWATQIEILSHSAIGGYLCHCGWNSILESVVH GVPIIAWPLFAEQKMNAAMLTTGLKVALRPKVSEK GMIEREEIAVVIKNLMVGEEVAKEIRQRMKWLKDAA HDALKEDGSSTRTLTQLAIKWESLAV	Example 15
17	*I. suffruticosa* UGT1	glucosyl-transferase	SVTFIIPTDGPPSKAQKTVLQSLPPAISHTFLPPVNL SDVPKDAMIETIISLTVLRSLPSIRDLFRSLSPSVLVL DLFGTDAFDVAKEFNVSPYIFFPSTAMVLSFFLHLP HLDREVHSEYRELAEPVKIPGCVPVHGKDLLAPVQ DRKNDAYKWVLHHTKRYREAEGIIENSFLELEPGPI KELLKEDSVKPPVYPVGPLVNVETGRAGNGSECLK WLDEQPHGSVLFVSFGSGGTLSXAQXNELALGLE ASEZRFLWVVRSPNDXXANASFFSAXSXADPFDFL PKGFVERTKZRGXXXXSWAPQPQVLAHGSTGGFL THCGWNSILESVVNGVPLIAWPLYAEQKMNAVMLT QDXKVALRXXDXBGXLVXREEIAXVVKXLMXGXEG KKVRXXMKDLK	Example 15
18	*I. suffruticosa* UGT2	glucosyl-transferase	MAKTVHIAVVPSAGFSHLVPVIEFSKRLIKHHPNFH VTCIIPSLESPPQSSKAYLETLPSNIDSIFLPPIKKED LPQGAYTGILIQLTLTYSLPSIHEALKSLNSKAPLAVL VADVFAYQALDFAKEFNSLSYIYVPGSATVLSLVLH MPRLDEEVSGEFKDHKEPIKLPGCVPLMGYDLPNP VQIRSSEAYKQFLERAKRMFDVDGMLINSFLELEP GAIKALEEKGNERRMRFYPVGPITQKGSSNEVDDD SGCLRWLDNQPVGSVLYVSFGSGGTLSQNQIDEL ASGLELSGQRFLWVLRAPSDSSSGAYLGGASEDP LKFLPSGFLERTKEQGLVVPSWAPQIQVLSHESVS GFLSHCGWNSILESVQMGVPLITWPLFAEQRMNA VMLTNGLKVALRPKVNEDGIVKKEEIAKVIRCLMEG EEGKGMRERMEKLKNSAAIALEDGSSTQSLLQLAS DLENLGGGF	Example 15

19	*P. trichocarpa* XP_002320190.1	glucosyl-transferase	MAETDSPPHVAILPSPGMGHLIPLVELAKRLVHQH NLSVTFIIPTDGSPSKAQRSVLGSLPSTIHSVFLPPV NLSDLPEDVKIETLISLTVARSLPSLRDVLSSLVASG TRVVALVVDLFGTDAFDVAREFKASPYIFYPAPAMA LSLFFYLPKLDEMVSCEYSEMQEPVEIPGCLPIHGG ELLDPTRDRKNDAYKWLLHHSKRYRLAEGVMVNS FIDLERGALKALQEVEPGKPPVYPVGPLVNMDSNT SGVEGSECLKWLDDQPLGSVLFVSFGSGGTLSFD QITELALGLEMSEQRFLWVARVPNDKVANATYFSV DNHKDPFDFLPKGFLDRTKGRGLVVPSWAPQAQV LSHGSTGGFLTHCGWNSTLESVVNAVPLIVWPLYA EQKMNAWMLTKDVEVALRPKASENGLIGREEIANI VRGLMEGEEGKRVRNRMKDLKDAAAEVLSEAGSS TKALSEVARKWKNHKCTQDCN	BLASTP similarity to other glucosyltransferases
20	*L. barbarum* BAG80556.1	glucosyl-transferase	MAETPVVTPHIAILPSPGMGHLIPLVEFSKRLIQNHH FSVTLILPTDGPVSNAQKIYLNSLPCSMDYHLLPPV NFDDLPLDTKMETRISLTVTRSLPSLREVFKTLVET KKTVALVVDLFGTDAFDVANDFKVSPYIFYPSTAMA LSLFLYLPKLDETVSCEYTDLPDPVQIPGCIPIHGKD LLDPVQDRKNEAYKWVLHHSKRYRMAEGIVANSF KELEGGAIKALQEEEPGKPPVYPVGPLIQMDSGSG SKADRSECLTWLDEQPRGSVLYISFGSGGTLSHEQ MIELASGLEMSEQRFLWVIRTPNDKMASATYFNVQ DSTNPLDFLPKGFLEKTKGLGLVVPNWAPQAQILG HGSTSGFLTHCGWNSTLESVVHGVPFIAWPLYAE QKMNAVMLSEDIKVALRPKANENGIVGRLEIAKVVK GLMEGEEGKVVRSRMRDLKDAAAKVLSEDGSSTK ALAELATKLKKKVSNN	BLASTP similarity to other glucosyltransferases
21	*P. putida* ndoA	oxygenase	MTVKWIEAVALSDILEGDVLGVTVEGKELALYEVEG EIYATDNLCTHGSARMSDGYLEGREIECPLHQGRF DVCTGKALCAPVTQNIKTYPVKIENLRVMIDLS	Berry, A. et al. Application of metabolic engineering to improve both the production and use of biotech indigo. Journal of Industrial Microbiology & Biotechnology 28, 127 - 133 (2002)

22	*P. putida* ndoB	oxygenase	MNYNNKILVSESGLSQKHLIHGDEELFQHELKTIFA RNWLFLTHDSLIPAPGDYVTAKMGIDEVIVSRQND GSIRAFLNVCRHRGKTLVSVEAGNAKGFVCSYHG WGFGSNGELQSVPFEKDLYGESLNKKCLGLKEVA RVESFHGFIYGCFDQEAPPLMDYLGDAAWYLEPM FKHSGGLELVGPPGKVVIKANWKAPAENFVGDAY HVGWTHASSLRSGESIFSSLAGNAALPPEGAGLQ MTSKYGSGMGVLWDGYSGVHSADLVPELMAFGG AKQERLNKEIGDVRARIYRSHLNCTVFPNNSMLTC SGVFKVWNPIDANTTEVWTYAIVEKDMPEDLKRRL ADSVQRTFGPAGFWESDDNDNMETASQNGKKYQ SRDSDLLSNLGFGEDVYGDAVYPGVVGKSAIGETS YRGFYRAYQAHVSSSNWAEFEHASSTWHTELTKT TDR	Berry, A. et al. Application of metabolic engineering to improve both the production and use of biotech indigo. Journal of Industrial Microbiology & Biotechnology 28, 127 - 133 (2002)
23	*P. putida* ndoC	oxygenase	MMINIQEDKLVSAHDAEEILRFFNCHDSALQQEATT LLTQEAHLLDIQAYRAWLEHCVGSEVQYQVISREL RAASERRYKLNEAMNVYNENFQQLKVRVEHQLDP QNWGNSPKLRFTRFITNVQAAMDVNDKELLHIRSN VILHRARRGNQVDVFYAAREDKWKRGEGGVRKLV QRFVDYPERILQTHNLMVFL	Berry, A. et al. Application of metabolic engineering to improve both the production and use of biotech indigo. Journal of Industrial Microbiology & Biotechnology 28, 127 - 133 (2002)
24	*P. putida* ndoR	oxygenase	MELLIQPNNRIIPFSAGANLLEVLRENGVAISYSCLS GRCGTCRCRVIDGSVIDSGAENGQSNLTDKQYVLA CQSVLTGNCAIEVPEADEIVTHPARIIKGTVVAVESP THDIRRLRVRLSKPFEFSPGQYATLQFSPEHARPY SMAGLPDDQEMEFHIRKVPGGRVTEYVFEHVREG TSIKLSGPLGTAYLRQKHTGPMLCVGGGTGLAPVL SIVRGALKSGMTNPILLYFGVRSQQDLYDAERLHKL AADHPQLTVHTVIATGPINEGQRAGLITDVIEKDILS LAGWRAYLCGAPAMVEALCTVTKHLGISPEHIYAD AFYPGGI	Berry, A. et al. Application of metabolic engineering to improve both the production and use of biotech indigo. Journal of Industrial Microbiology & Biotechnology 28, 127 - 133 (2002)

25	*B. circulans* bglA	G-hydrolase	MSIHMFPSDFKWGVATAAYQIEGAYNEDGRGMSI WDTFAHTPGKVKNGDNGNVACDSYHRVEEDVQLL KDLGVKVYRFSISWPRVLPQGTGEVNRAGLDYYH RLVDELLANGIEPFCTLYHWDLPQALQDQGGWGS RITIDAFAEYAELMFKELGGKIKQWITFNEPWCMAF LSNYLGVHAPGNKDLQLAIDVSHHLLVAHGRAVTL FRELGISGEIGIAPNTSWAVPYRRTKEDMEACLRV NGWSGDWYLDPIYFGEYPKFMLDWYENLGYKPPI VDGDMELIHQPIDFIGINYYTSSMNRYNPGEAGGM LSSEAISMGAPKTDIGWEIYAEGLYDLLRYTADKYG NPTLYITENGACYNDGLSLDGRIHDQRRIDYLAMHL IQASRAIEDGINLKGYMEWSLMDNFEWAEGYGMR FGLVHVDYDTLVRTPKDSFYWYKGVISRGWLDL	S Paavilainen, J Hellman, T Korpela. Purification, characterization, gene cloning, and sequencing of a new beta-glucosidase from Bacillus circulans subsp. alkalophilus. Appl. Environ. Microbiol. March 1993 vol. 59 no. 3 927-932
26	*P. aeruginosa* atsA	S-hydrolase	MSKRPNFLVIVADDLGFSDIGAFGGEIATPNLDALAI AGLRLTDFHTASTCSPTRSMLLTGTDHHIAGIGTMA EALTPELEGKPGYEGHLNERVVALPELLREAGYQT LMAGKWHLGLKPEQTPHARGFERSFSLLPGAANH YGFEPPYDESTPRILKGTPALYVEDERYLDTLPEGF YSSDAFGDKLLQYLKERDQSRPFFAYLPFSAPHWP LQAPREIVEKYRGRYDAGPEALRQERLARLKELGL VEADVEAHPVLALTREWEALEDEERAKSARAMEV YAAMVERMDWNIGRVVDYLRRQGELDNTFVLFMS DNGAEGALLEAFPKFGPDLLGFLDRHYDNSLENIG RANSYVWYGPRWAQAATAPSRLYKAFTTQGGIRV PALVRYPRLSRQGAISHAFATVMDVTPTLLDLAGV RHPGKRWRGREIAEPRGRSWLGWLSGETEAAHD ENTVTGWELFGMRAIRQGDWKAVYLPAPVGPATW QLYDLARDPGEIHDLADSQPGKLAELIEHWKRYVS ETGVVEGASPFLVR	Beil S, Kehrli H, James P, Staudenmann W, Cook AM, Leisinger T, Kertesz MA. Purification and characterization of the arylsulfatase synthesized by Pseudomonas aeruginosa PAO during growth in sulfate-free medium and cloning of the arylsulfatase gene (atsA). Eur J Biochem. 1995 Apr 15;229(2):385-94.

27	*E. coli* yidJ	S-hydrolase	MKRPNFLFVMTDTQATNMVGCYSGKPLNTQNIDSL AAEGIRFNSAYTCSPVCTPARAGLFTGIYANQSGP WTNNVAPGKNISTMGRYFKDAGYHTCYIGKWHLD GHDYFGTGECPPEWDADYWFDGANYLSELTEKEI SLWRNGLNSVEDLQANHIDETFTWAHRISNRAVDF LQQPARADEPFLMVVSYDEPHHPFTCPVEYLEKYA DFYYELGEKAQDDLANKPEHHRLWAQAMPSPVGD DGLYHHPLYFACNDFVDDQIGRVINALTPEQRENT WVIYTSDHGEMMGAHKLISKGAAMYDDITRIPLIIRS PQGERRQVDTPVSHIDLLPTMMALADIEKPEILPGE NILAVKEPRGVMVEFNRYEIEHDSFGGFIPVRCWV TDDFKLVLNLFTSDELYDRRNDPNEMHNLIDDIRFA DVRSKMHDALLDYMDKIRDPFRSYQWSLRPWRKD ARPRWMGAFRPRPQDGYSPVVRDYDTGLPTQGV KVEEKKQKF	BLASTP similarity to other S-hydrolases
28	*P. stuartii* arylsulfatase 3	S-hydrolase	MKKTLLAIALSSVMSGVALGEVDDRPNVLIIIADDM GYSDISPFGGEIPTPNLQKMAEQGVRMSQYYTSP MSAPARSMLMTGATNQQAGMGGMWWYENTVGK PGYELRLTDRVVTMAERFQDAGYNTLMSGKWHLG YTKGARPTDRGFNQAFAFMGGGTSHFDDAKPLGT VESFHTYYTLNGEKVSLPSDFYSSKNYAQQLEQWI KQTPSDQPIFAYLAFTAPHDPIQAPDDWIRKFDGKY DEGFGKIYRQRINRLKELGIINDKTPMPKLNLDKEW EQLTPEEKRYAAKTMQVYAAMIAYMDDQIGGVINT LKETGRDKNTIIIFATDNGANPASGFYYESDPEYWK QFDNSYENLGRKNSFVSVGPQWANVSNAPYANY HKTTSAQGGINTDLIITGPGIGKAGSIDKTPMAVYDI APTLYEFAGIDANKQIKNIHPLPMLGTSFKSHFLGK STVNPRQLFGVELHNQAALVEGDWKLRRLVKASP KAEMAPWQLFNLKEDPLETRDLAAKHPEIVQKLQK KYEQFAKTGMIIEAKGEAIDYIGVDESTGNYIGIDPK TNKRIEPAKVK	Dealler SF, Hawkey PM, Millar MR. Enzymatic degradation of urinary indoxyl sulfate by Providencia stuartii and Klebsiella pneumoniae causes the purple urine bag syndrome. J Clin Microbiol. 1988 Oct;26(10):2152-6.
29	*P. stuartii* anaerobic sulfatase maturating enzyme (anSME)	S-hydrolase companion	MKISFYDPPRLQGKSLKSAIPFHILLKPVGSGCNLK CDYCYYPQHNEQKAAPMLKAMLEPFIKNYIAAQPA YTKEINFVWQGGEPLLAGLDFYKRAIALQQKYAPH GVRIINTLQTNATLLTPSWCRFLKQHDFVIGVSLDG PESIHDQYRHDRRGNSGSYASVIKGIALLQQFDIEF NILTVVHDGVAHLGKEIYLHFVQLGIRYIQFQPLMLE GDAIHQGFTLSANNWGLFLSSVYQQWQASGHIGR VFVMNIEQVYSQYFTQVSSTCVHSERCGTNMMME TQGEIYACDHQANQSHYLGQFNGQQGFSDFVEAS ISLPFGQNKSRRKECQQCSVKMVCQGGCPAHLNQ FGRNQLCEGYFAFFSLVLAPIRQYQRNAQGVQHW RNAFLKNAVA	Dealler SF, Hawkey PM, Millar MR. Enzymatic degradation of urinary indoxyl sulfate by Providencia stuartii and Klebsiella pneumoniae causes the purple urine bag syndrome. J Clin Microbiol. 1988 Oct;26(10):2152-6.

30	*N. crassa* CDT1	transporter	MSSHGSHDGASTEKHLATHDIAPTHDAIKIVPKGH GQTATKPGAQEKEVRNAALFAAIKESNIKPWSKESI HLYFAIFVAFCCACANGYDGSLMTGIIAMDKFQNQF HTGDTGPKVSVIFSLYTVGAMVGAPFAAILSDRFG RKKGMFIGGIFIIVGSIIVASSSKLAQFVVGRFVLGLG IAIMTVAAPAYSIEIAPPHWRGRCTGFYNCGWFGG SIPAACITYGCYFIKSNWSWRIPLILQAFTCLIVMSS VFFLPESPRFLFANGRDAEAVAFLVKYHGNGDPNS KLVLLETEEMRDGIRTDGVDKVWWDYRPLFMTHS GRWRMAQVLMISIFGQFSGNGLGYFNTVIFKNIGVT STSQQLAYNILNSVISAIGALTAVSMTDRMPRRAVLI IGTFMCAAALATNSGLSATLDKQTQRGTQINLNQG MNEQDAKDNAYLHVDSNYAKGALAAYFLFNVIFSF TYTPLQGVIPTEALETTIRGKGLALSGFIVNAMGFIN QFAGPIALHNIGYKYIFVFVGWDLIETVAWYFFGVE SQGRTLEQLEWVYDQPNPVKASLKVEKVVVQADG HVSEAIVA	Kim H, Lee WH, Galazka JM, Cate JH, Jin YS. Analysis of cellodextrin transporters from Neurospora crassa in Saccharomyces cerevisiae for cellobiose fermentation. Appl Microbiol Biotechnol. 2014 Feb;98(3):1087-94. doi: 10.1007/s00253-013-5339-2. Epub 2013 Nov 5.

CLAIMS

What is claimed is:

1. A method of producing a compound of Formula (A), the method comprising:

a) contacting a host cell with a compound of Formula (II); and

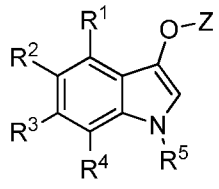
b) culturing the host cell under conditions such that a compound of Formula (A) is produced from at least a portion of the compound of Formula (II);

wherein the host cell comprises a single or multiple recombinant nucleic acid(s) encoding:

i) a polypeptide with oxygenase activity; and

ii) a polypeptide with glycosyltransferase activity or a polypeptide with sulfotransferase activity, or a combination thereof;

wherein the compound of Formula (A) is:


(A),

wherein:

R^1, R^2, R^3, and R^4 are independently selected from the group consisting of H, halo, nitro, sulfate, phosphate, hydroxyl, $-C(O)OR^b$, $-R^aC(O)OR^b$, $-R^aOC(O)R^b$, $-OR^a$, $-NR^bR^c$, and alkyl;

wherein alkyl may be unsubstituted or substituted with one or more substituents independently selected from the group consisting of halo, nitro, sulfate, phosphate, hydroxyl, $-C(O)OR^b$, $-R^aC(O)OR^b$, $-R^aOC(O)R^b$, $-OR^a$, $-NR^bR^c$, and $-R^aOR^d$;

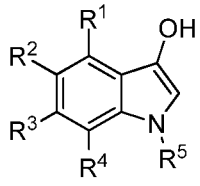
R^5 is H or alkyl, wherein alkyl is unsubstituted or substituted with one or more substituents independently selected from the group consisting of halo, hydroxyl, =O, $-C(O)OR^b$, $-R^aC(O)OR^b$, $-R^aOC(O)R^b$, $-OR^a$, and $-NR^bR^c$;

R^a and R^d are independently alkyl;

R^b and R^c are independently H or alkyl;

Z is a glycone or $-SO_3^-$; and

wherein the compound of Formula (II) is:



(II),

wherein R^1, R^2, R^3, R^4, and R^5 are defined as for Formula (A).

2. The method of claim 1, wherein the host cell is a bacterial cell.

3. The method of claim 2, wherein the host cell is selected from the group consisting of *E. coli* and *Corynebacterium glutamicum.*

4. The method of claim 1, wherein the host cell is a eukaryotic cell.

5. The method of claim 4, wherein the host cell is a yeast.

6. The method of claim 4 or 5, wherein the host cell is selected from the group consisting of *Saccharomyces cerevisiae*, *Pichia pastoris*, *Kluyveromyces lactis, Kluyveromyces marxianus*, and *Schizosaccharomyces pombe*.

7. The method of any one of claims 1 to 6, wherein the polypeptide with oxygenase activity comprises SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, or SEQ ID NO: 24, or the amino acid sequence of any homologs thereof.

8. The method of any one of claims 1 to 7, wherein the polypeptide with glycosyltransferase activity comprises SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, or SEQ ID NO: 20, or the amino acid sequence of any homologs thereof.

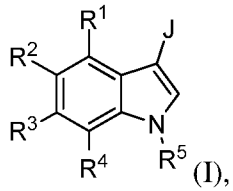
9. The method of any one of claims 1 to 7, wherein the polypeptide with sulfotransferase activity comprises SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, or SEQ ID NO: 14, or the amino acid sequence of any homologs thereof.

10. The method of claim 7, wherein the homolog of a polypeptide with oxygenase activity comprises an amino acid sequence at least 90%, at least 95%, or at least 99% identical to an amino acid sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, and SEQ ID NO: 24.

11. The method of claim 8, wherein the homolog of a polypeptide with glycosyltransferase comprises an amino acid sequence at least 90%, at least 95%, or at least 99% identical to an amino acid sequence selected from the group consisting of SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, and SEQ ID NO: 20.

12. The method of claim 9, wherein the homolog of a polypeptide with sulfotransferase comprises an amino acid sequence at least 90%, at least 95%, or at least 99% identical to an amino acid sequence selected from the group consisting of SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, and SEQ ID NO: 14.

13. The method of any one of claims 1 to 12, further comprising providing a compound of Formula (I):



(I),

wherein:

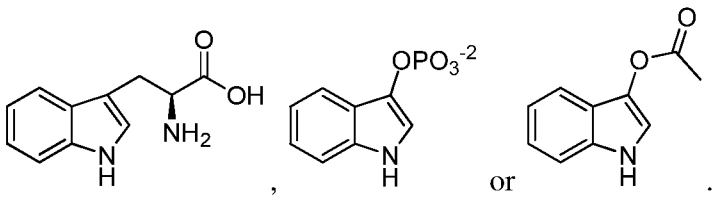
R^1, R^2, R^3, R^4, and R^5 are defined as for Formula (A);

J is alkyl, $-OC(O)R^b$, or phosphate;

wherein alkyl is unsubstituted or substituted with one or more substituents independently selected from the group consisting of halo, hydroxyl, =O, –C(O)OR^b, –R^{a}C(O)OR^b, –R^{a}OC(O)R^b, –OR^a, –NR^{b}R^c, and –R^{a}OR^d; and

R^a, R^b, R^c, and R^d are defined as for Formula (A).

14. The method of claim 13, wherein the compound of Formula (I) is:



.

15. The method of any one of claims 1 to 14, wherein the titer of the compound of Formula (A) produced is at least 50 mg/L, 100 mg/L, 1 g/L, 10 g/L, 25 g/L, 50 g/L, 75 g/L, 100 g/L, 125 g/L, 150 g/L, 175 g/L, or 200 g/L.

16. The method of any one of claims 1 to 15, wherein the compound of Formula (A) is secreted by the host cell.

17. The method of any one of claims 1 to 16, further comprising isolating the compound of Formula (A) produced.

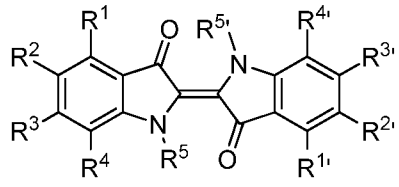
18. The method of any one of claims 1 to 17, further comprising converting the compound of Formula (A) to a compound of Formula (II).

19. A method of producing a compound of Formula (B), the method comprising converting a first compound of Formula (A) and a second compound of Formula (A) to a compound of Formula (B);

wherein the first compound of Formula (A) and the second compound of Formula (A) are the same compound of Formula (A) or different compounds of Formula (A);

at least one of the first compound of Formula (A) and the second compound of Formula (A) is produced according to the method of any of the above claims; and

the compound of Formula (B) is:


(B),

wherein R^1, R^2, R^3, R^4, R^5, $R^{1'}$, $R^{2'}$, $R^{3'}$, $R^{4'}$, and $R^{5'}$ are defined as for Formula (A).

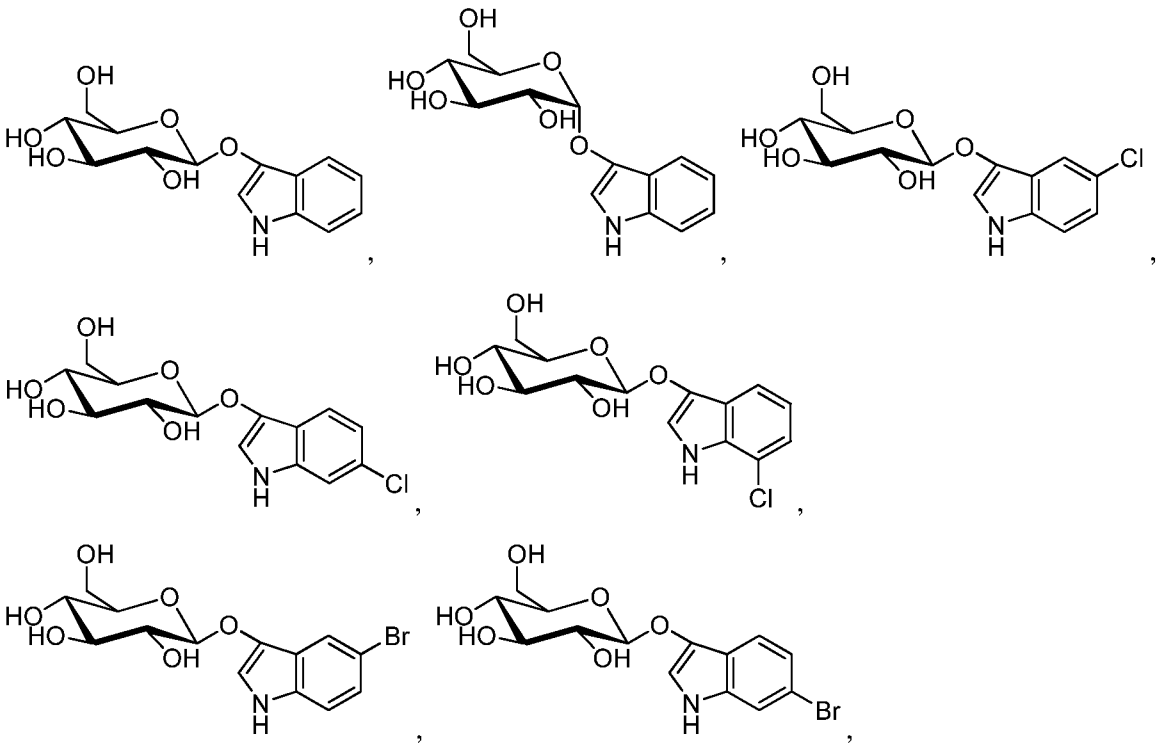
20. The method of claim 19, wherein:

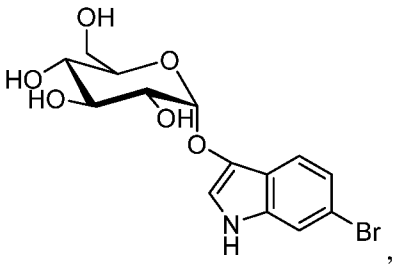
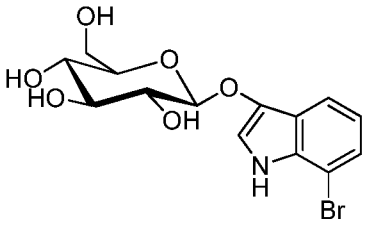
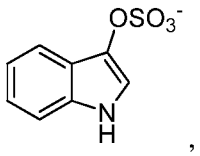
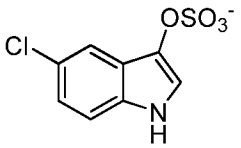
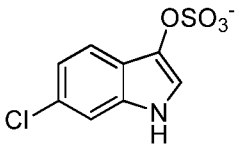
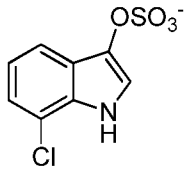
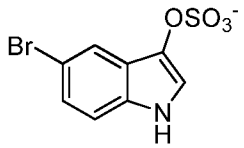
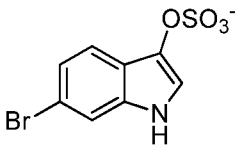
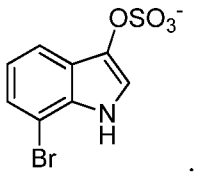
a) the compound of Formula (A) is converted to a compound of Formula (II); and

b) the compound of Formula (II) is converted to the compound of Formula (B).

21. The method of claim 18 or 20, wherein the compound of Formula (A) is contacted by a hydrolase to convert the compound of Formula (A) to the compound of Formula (II).

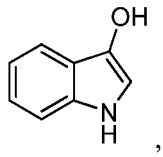
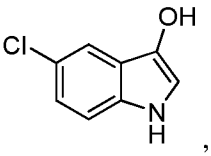
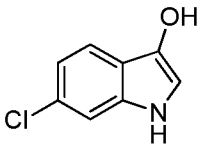
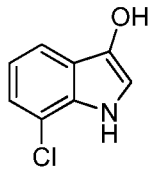
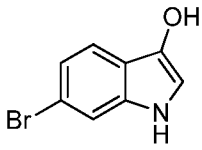
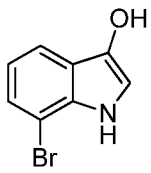
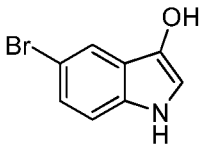
22. The method of any one of claims 1 to 21, wherein the compound of Formula (A) is:



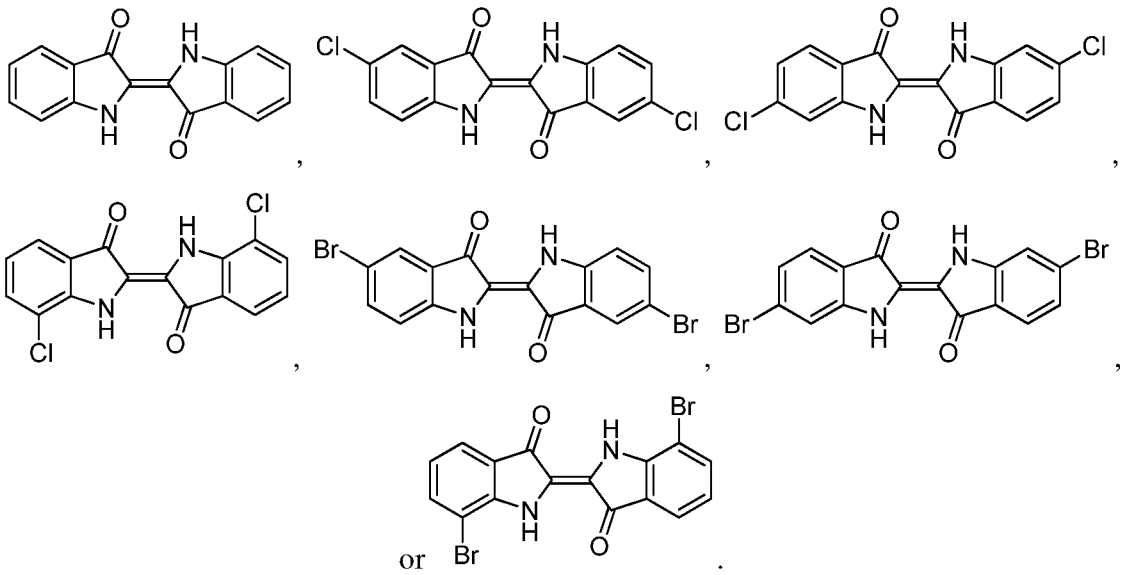
, or .

23. The method of any one of claims 1 to 22, where in the compound of Formula (II) is:

, or .

24. The method of claims 19 or 20, wherein the compound of Formula (B) is:



25. The method of claims 19 or 20, wherein the compound of Formula (A) is indican, the compound of Formula (II) is indoxyl, and the compound of Formula (B) is indigo.

26. The method of any one of claim 21, wherein the hydrolase is a glucosidase.

27. The method of any one of claim 21, wherein the hydrolase is a sulfatase.

28. The method of any of the above claims, wherein the polypeptide having oxygenase activity and the polypeptide having glycosyltransferase activity are the same polypeptide.

29. The method of any one of claims 1 to 27, wherein the polypeptide having oxygenase activity and the polypeptide having sulfotransferase activity are the same polypeptide.

30. A host cell comprising a single or multiple recombinant nucleic acid(s) encoding a polypeptide with monooxygenase activity and a polypeptide with glucosyltransferase activity.

31. A host cell comprising a single or multiple recombinant nucleic acid(s) encoding a polypeptide with monooxygenase activity and a polypeptide with sulfotransferase activity.

32. The method of any one of claims 1 to 18, wherein the host cell does not produce a polypeptide selected from the group consisting of acetyl esterase, esterase yjfP, pimeloyl-[acyl-carrier protein] methyl ester esterase, S-formylglutathione hydrolase YeiG, S-formylglutathione

hydrolase FrmB, proofreading thioesterase EntH, uncharacterized protein YdiL, acyl-CoA thioesterase I, probable 9-O-acetyl-N-acetylneuraminic acid deacetylase, esterase YqiA, esterase YbfF, esterase YpfH, acetylornithine deacetylase, esterase FrsA, acyl-CoA thioester hydrolase YbgC, and tryptophanase.

33. The method of any one of claims 1 to 18, wherein the host cell does not produce acetyl esterase, esterase yjfP, pimeloyl-[acyl-carrier protein] methyl ester esterase, S-formylglutathione hydrolase YeiG, and S-formylglutathione hydrolase FrmB.

34. The method of any one of claims 1 to 18, wherein the host cell does not produce acetyl esterase, esterase yjfP, pimeloyl-[acyl-carrier protein] methyl ester esterase, S-formylglutathione hydrolase YeiG, S-formylglutathione hydrolase FrmB, proofreading thioesterase EntH, uncharacterized protein YdiL, acyl-CoA thioesterase I, probable 9-O-acetyl-N-acetylneuraminic acid deacetylase, esterase YqiA, esterase YbfF, esterase YpfH, acetylornithine deacetylase, esterase FrsA, acyl-CoA thioester hydrolase YbgC, and tryptophanase.

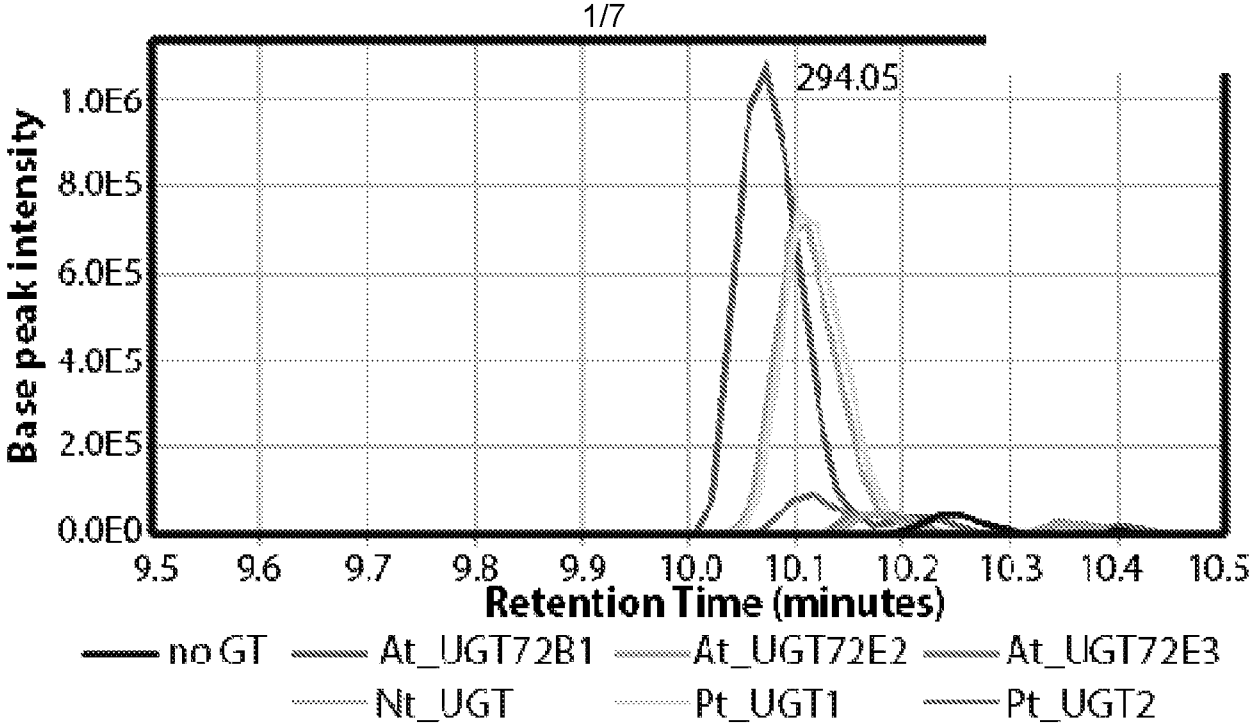

294.05
Base peak intensity
1.0E6
8.0E5
6.0E5
4.0E5
2.0E5
0.0E0
9.5
9.6
9.7
9.8
9.9
10.0
10.1
10.2
10.3
10.4
10.5
Retention Time (minutes)
no GT
At_UGT72B1
At_UGT72E2
At_UGT72E3
Nt_UGT
Pt_UGT1
Pt_UGT2

FIG. 1

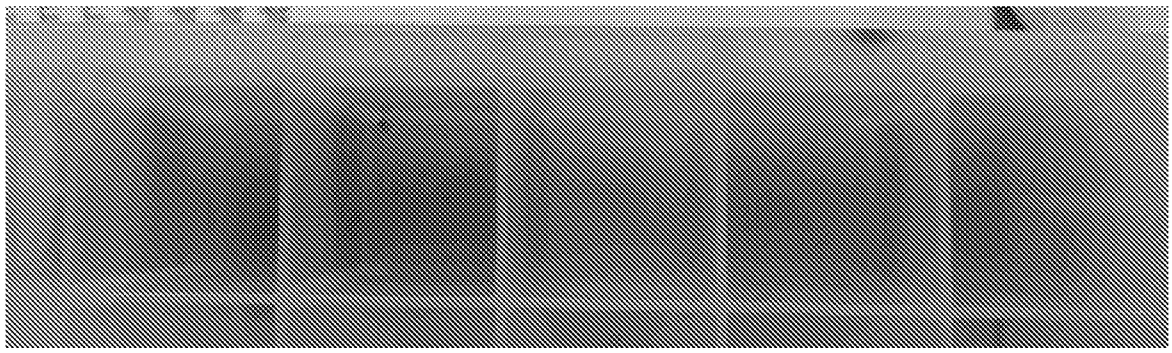


FIG. 2

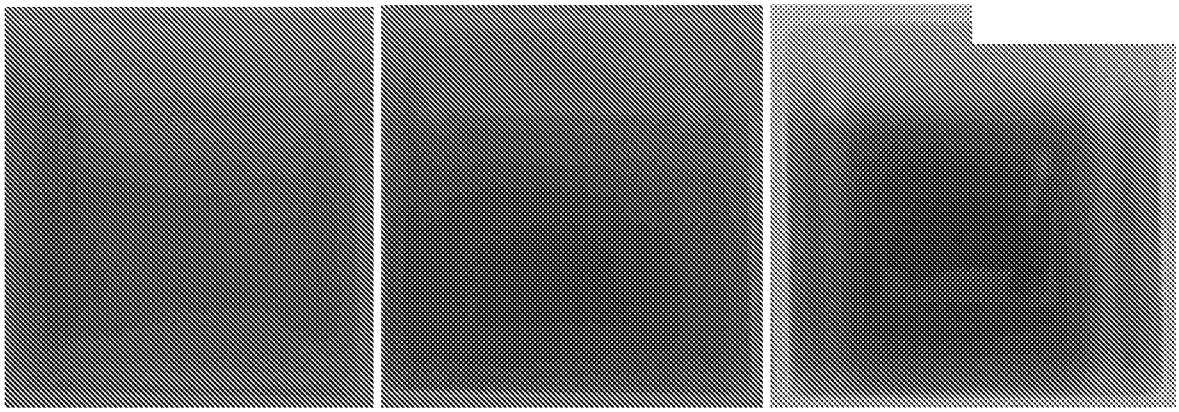


FIG. 3

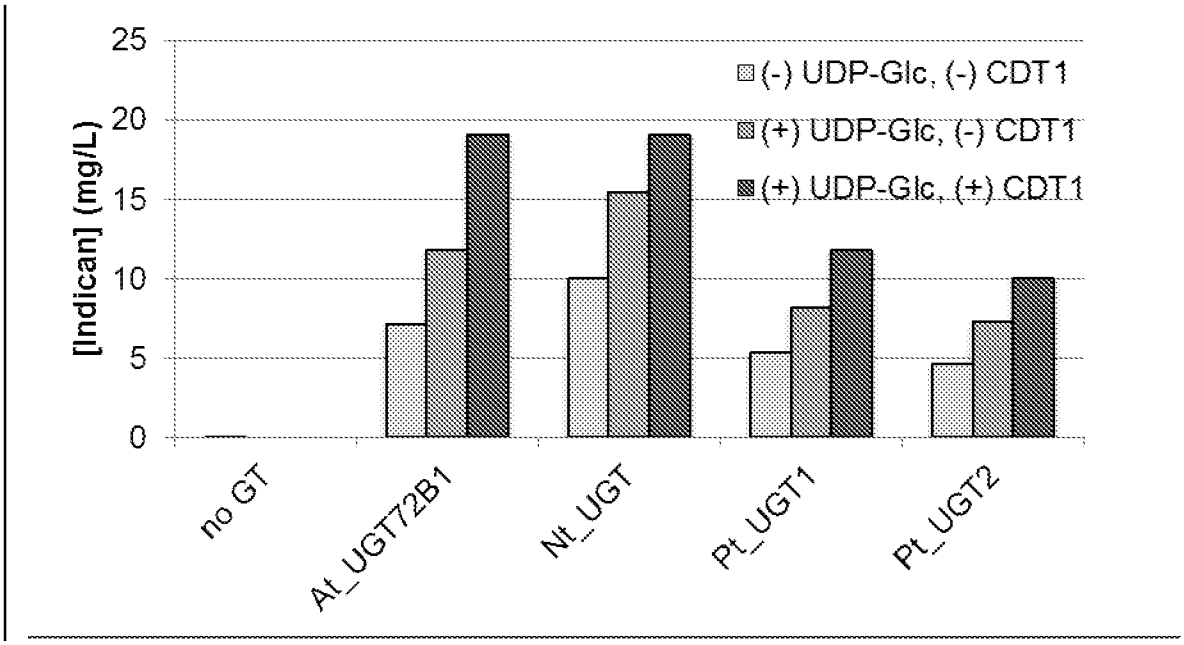
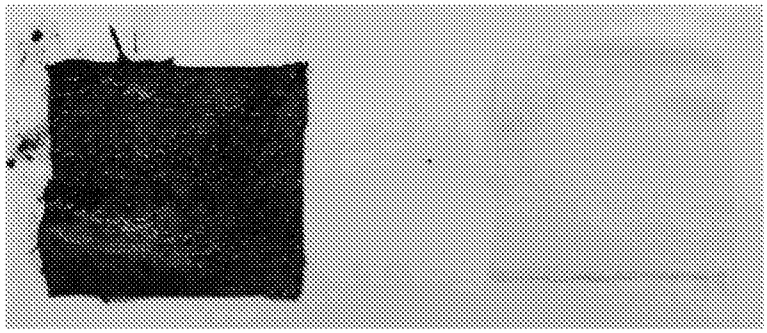
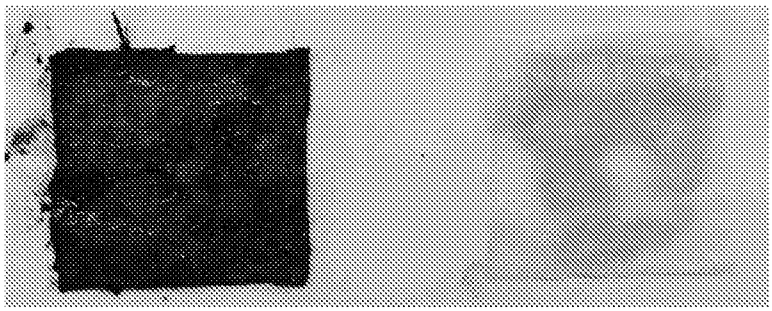

25
20
15
10
5
0
[Indican] (mg/L)
(-) UDP-Glc, (-) CDT1
(+) UDP-Glc, (-) CDT1
(+) UDP-Glc, (+) CDT1
no GT
At_UGT72B1
Nt_UGT
Pt_UGT1
Pt_UGT2

FIG. 4



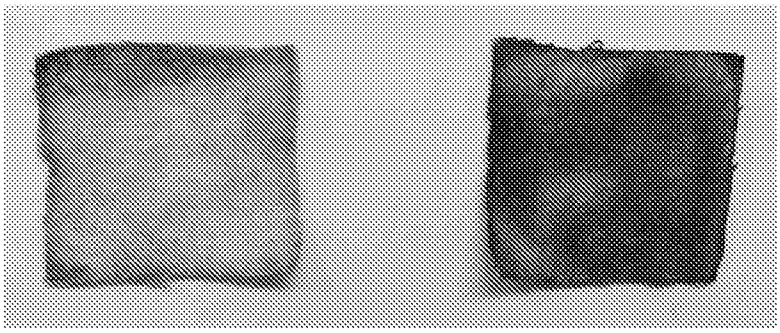
indigo indican

before glucosidase application



indigo indican

5 min incubation with glucosidase



indigo indican

6 h + washing and drying

FIG. 5

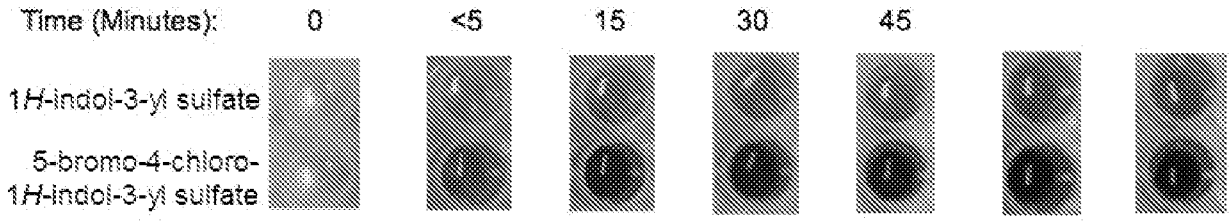

Time (Minutes):
0
<5
15
30
45
1H-indol-3-yl sulfate
5-bromo-4-chloro-
1H-indol-3-yl sulfate

FIG. 6

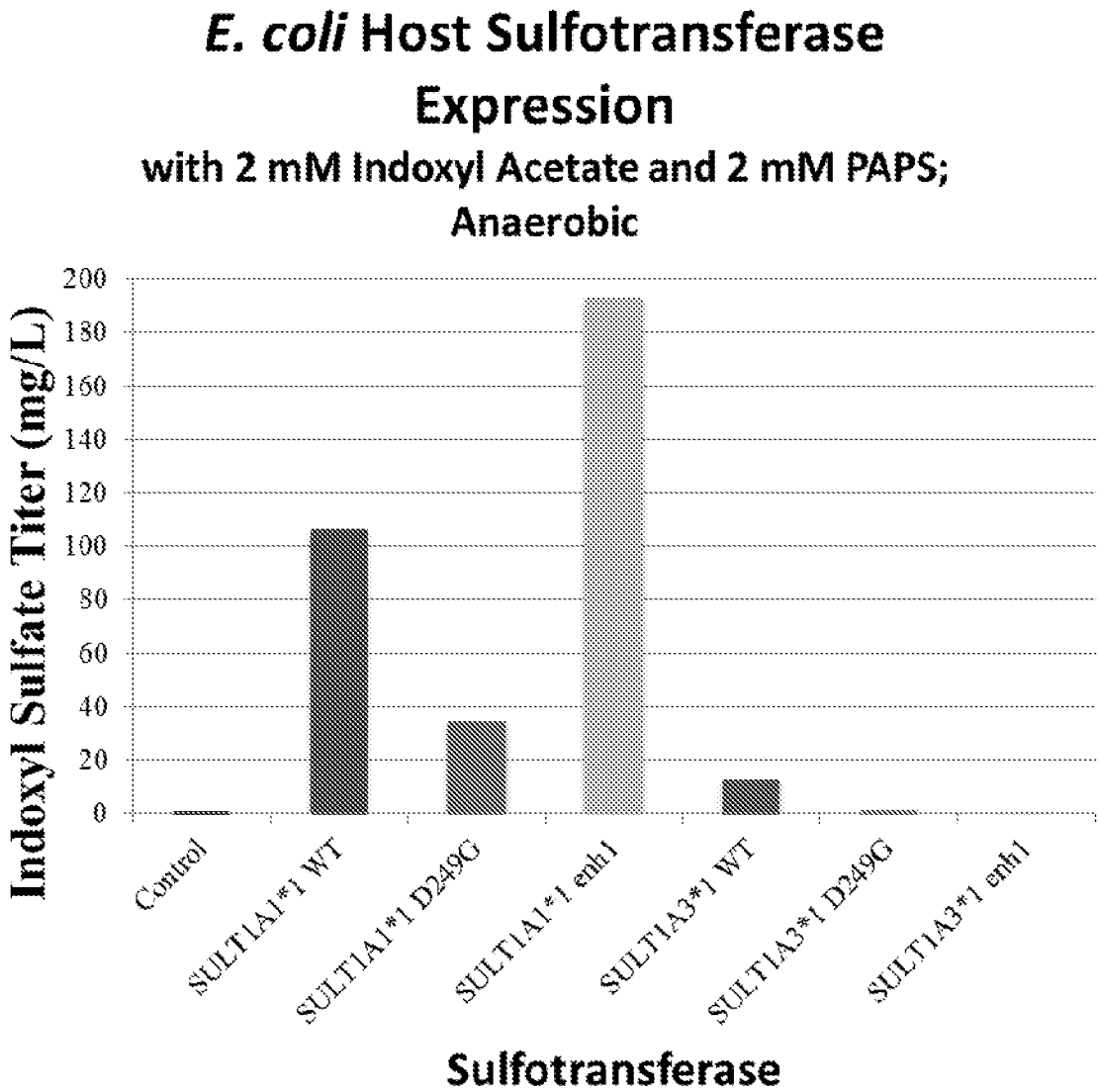
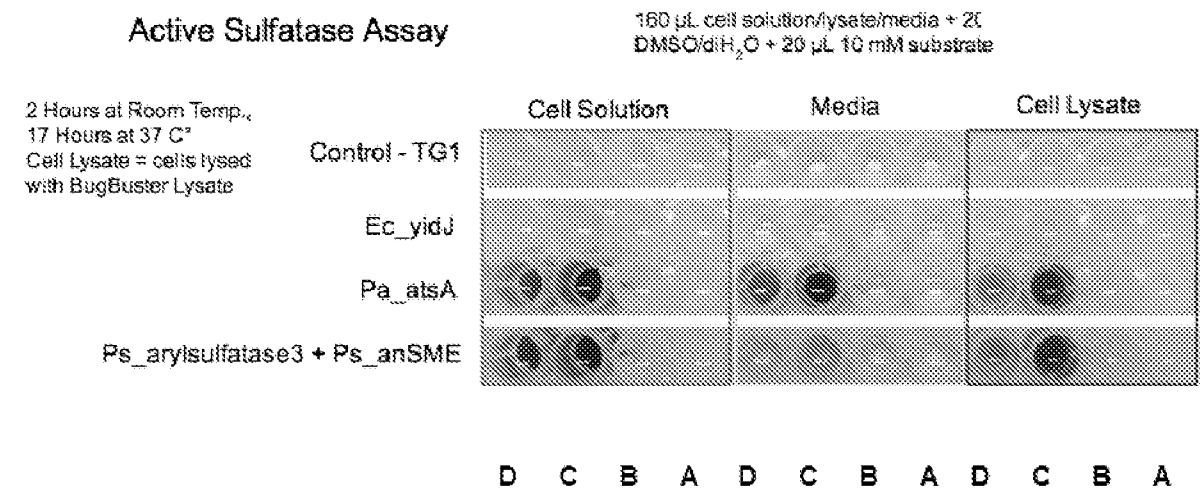
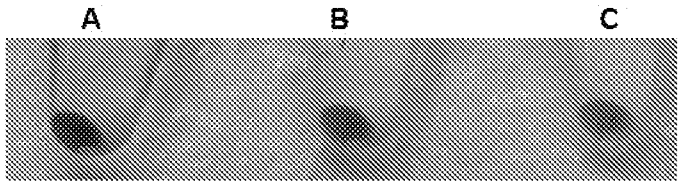

E. coli Host Sulfotransferase Expression
with 2 mM Indoxyl Acetate and 2 mM PAPS; Anaerobic
Indoxyl Sulfate Titer (mg/L)
200
180
160
140
120
100
80
60
40
20
0
Control
SULT1A1*1 WT
SULT1A1*1 D249G
SULT1A1*1 enh1
SULT1A3*1 WT
SULT1A3*1 D249G
SULT1A3*1 enh1
Sulfotransferase

FIG. 7

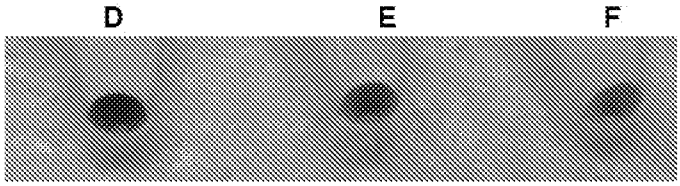


A: 4-nitrophenyl sulfate C: 5-bromo-4-chloro-1*H*-indol-3-yl sulfate
B: 4-nitrocatechol sulfate D: 1*H*-indol-3-yl sulfate

FIG. 8



A: 5,5'-dichloro-[2,2'-biindolinylidene]-3,3'-dione
B: 6,6'-dichloro-[2,2'-biindolinylidene]-3,3'-dione
C: 7,7'-dichloro-[2,2'-biindolinylidene]-3,3'-dione



D: 5,5'-dibromo-[2,2'-biindolinylidene]-3,3'-dione
E: 6,6'-dibromo-[2,2'-biindolinylidene]-3,3'-dione
F: 7,7'-dibromo-[2,2'-biindolinylidene]-3,3'-dione

FIG. 9

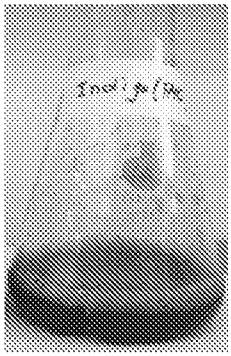


FIG. 10

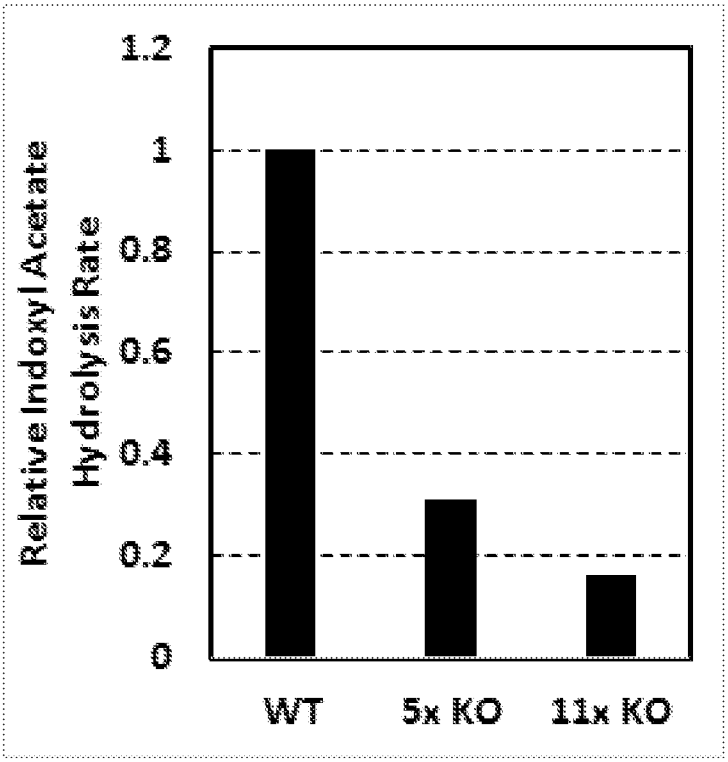

Relative Indoxyl Acetate Hydrolysis Rate
1.2
1
0.8
0.6
0.4
0.2
0
WT
5x KO
11x KO

FIG. 11

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2016/020729

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12P19/60 C12N9/02 C12N9/10 C12P17/16
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12P 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, WPI Data, BIOSIS, COMPENDEX, EMBASE, FSTA

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	HERIBERT WARZECHA ET AL: "Formation of the indigo precursor indican in genetically engineered tobacco plants and cell cultures", PLANT BIOTECHNOLOGY JOURNAL, vol. 5, no. 1, 1 January 2007 (2007-01-01) , pages 185-191, XP055271107, GB ISSN: 1467-7644, DOI: 10.1111/j.1467-7652.2006.00231.x abstract discussion; pages 188-190 ----- -/--	1-8,10, 11,15-18

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
12 May 2016	01/08/2016

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

2

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

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	MARCINEK H ET AL: "Indoxyl-UDPG-glucosyltransferase from Baphicacanthus cusia", PHYTOCHEMISTRY, PERGAMON PRESS, GB, vol. 53, no. 2, 1 January 2000 (2000-01-01), pages 201-207, XP004291276, ISSN: 0031-9422, DOI: 10.1016/S0031-9422(99)00430-6 abstract -----	1,8,11, 22,23
A	Ramya Prathuri: "oleD glucosyltransferase; S. antibiotics (characterized)", Registry of Standard Biological Parts, Part:BBa K1131006, 28 October 2013 (2013-10-28), pages 1-2, XP055272259, Retrieved from the Internet: URL:http://parts.igem.org/Part:BBa_K1131006 [retrieved on 2016-05-12] the whole document -----	1,8,11, 22,23
A	ELIZABETH M. J. GILLAM ET AL: "Exploiting the Versatility of Human Cytochrome P450 Enzymes: The Promise of Blue Roses From Biotechnology", IUBMB LIFE, vol. 52, no. 6, 1 December 2001 (2001-12-01), pages 271-277, XP055272223, GB ISSN: 1521-6543, DOI: 10.1080/152165401317291110 figure 2 Potential Applications of indole metabolism by P450s; page 275 -----	1-29, 32-34
A	DATABASE WPI Week 201204 Thomson Scientific, London, GB; AN 2011-E15079 XP002757646, & KR 2011 0037687 A (UNIV CHONNAM NAT IND FOUND) 13 April 2011 (2011-04-13) abstract -----	19-26

-/--

2

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	HIDEYUKI SAIGO ET AL: "Meclofenamate elicits a nephropreventing effect in a rat model of ischemic acute kidney injury by suppressing indoxyl sulfate production and restoring renal organic anion transporters", DRUG DESIGN, DEVELOPMENT AND THERAPY, 1 August 2014 (2014-08-01), page 1073, XP055272038, DOI: 10.2147/DDDT.S67456 abstract page 1074, left-hand column, last paragraph	1-7,9, 10,12-18
A	LI-JUAN YUAN ET AL: "Biooxidation of indole and characteristics of the responsible enzymes", AFRICAN JOURNAL OF BIOTECHNOLOGY, , vol. 10, no. 86 1 January 2011 (2011-01-01), pages 19855-19863, XP002698100, ISSN: 1684-5315, DOI: 10.5897/AJBX11.008 Retrieved from the Internet: URL:http://www.academicjournals.org/AJB/abstracts/abs2011/30Dec%20Special%20Review/Yuan%20et%20al.htm figures 1, 2 page 19856 - page 19861	1-8,10, 11,15-26
A	CHOI H S ET AL: "A novel flavin-containing monooxygenase from Methylophaga sp. strain SK1 and its indigo synthesis in Escherichia coli", BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, ACADEMIC PRESS INC. ORLANDO, FL, US, vol. 306, no. 4, 11 July 2003 (2003-07-11) , pages 930-936, XP004432109, ISSN: 0006-291X, DOI: 10.1016/S0006-291X(03)01087-8 abstract page 935, left-hand column, paragraph 2-4	1-29, 32-34
A,P	DATABASE BIOSIS [Online] BIOSCIENCES INFORMATION SERVICE, PHILADELPHIA, PA, US; August 2015 (2015-08), MINAMI YOSHIKO ET AL: "Transcriptome analysis for identification of indigo biosynthesis pathway genes in Polygonum tinctorium", XP002757647, Database accession no. PREV201500761707 abstract -/--	1-8,10, 11,15-18

2

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	& BIOLOGIA (BRATISLAVA), vol. 70, no. 8, August 2015 (2015-08), pages 1026-1032, ISSN: 0006-3088(print), DOI: 10.1515/BIOLOG-2015-0131	

L	CN 103 509 819 A (SHANGHAI ACAD AGRI SCIENCES; SHANGHAI RUIFENG AGRICULTURAL TECHNOLOGY) 15 January 2014 (2014-01-15) abstract	1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2016/020729

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-29, 32-34

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-29, 32-34

   Method of producing an indol-3-ol derivative of formula (A)

   ---

2. claim: 30

   Host cell comprising a single or multiple recombinant nucleic acid(s) encoding a monooxygenase and a glucosyltransferase

   ---

3. claim: 31

   Host cell comprising a single or multiple recombinant nucleic acid(s) encoding a monooxygenase and a sulfotransferase

   ---

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2016/020729

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
KR 20110037687 A	13-04-2011	NONE	
CN 103509819 A	15-01-2014	NONE	