



- (51) International Patent Classification:
C12N 9/10 (2006.01) *C12P 33/00* (2006.01)
- (21) International Application Number:
PCT/US2024/015560
- (22) International Filing Date:
13 February 2024 (13.02.2024)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
63/484,926 14 February 2023 (14.02.2023) US
63/586,338 28 September 2023 (28.09.2023) US
- (71) Applicant: **EPIMERON USA, INC.** [US/US]; 1278
Reamwood Ave., Sunnyvale, California 94089 (US).
- (72) Inventors: **ZACHARIA, Vineetha**; c/o Epimeron USA,
Inc., 1278 Reamwood Ave., Sunnyvale, California 94089
(US). **WORKENTINE, Matthew**; c/o Epimeron USA,
Inc., 1278 Reamwood Ave., Sunnyvale, California 94089
(US). **CHOUDHARY, Trish**; c/o Epimeron USA, Inc.,
1278 Reamwood Ave., Sunnyvale, California 94089 (US).
HOBSON, Christian; c/o Epimeron USA, Inc., 1278
Reamwood Ave., Sunnyvale, California 94089 (US).

(74) Agent: **WHITING, Adam K.**; FisherBroyles, LLP, 4100
Redwood Rd 20A-203, Oakland, California 94619 (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, CV, CZ, DE, DJ, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG,
KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY,
MA, MD, MG, MK, MN, MU, 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, ST, SV, SY, TH,
TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS,
ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, CV,
GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, 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).

(54) Title: RECOMBINANT POLYPEPTIDES FOR BIOSYNTHESIS OF URSODEOXYCHOLIC ACID (UDCA)

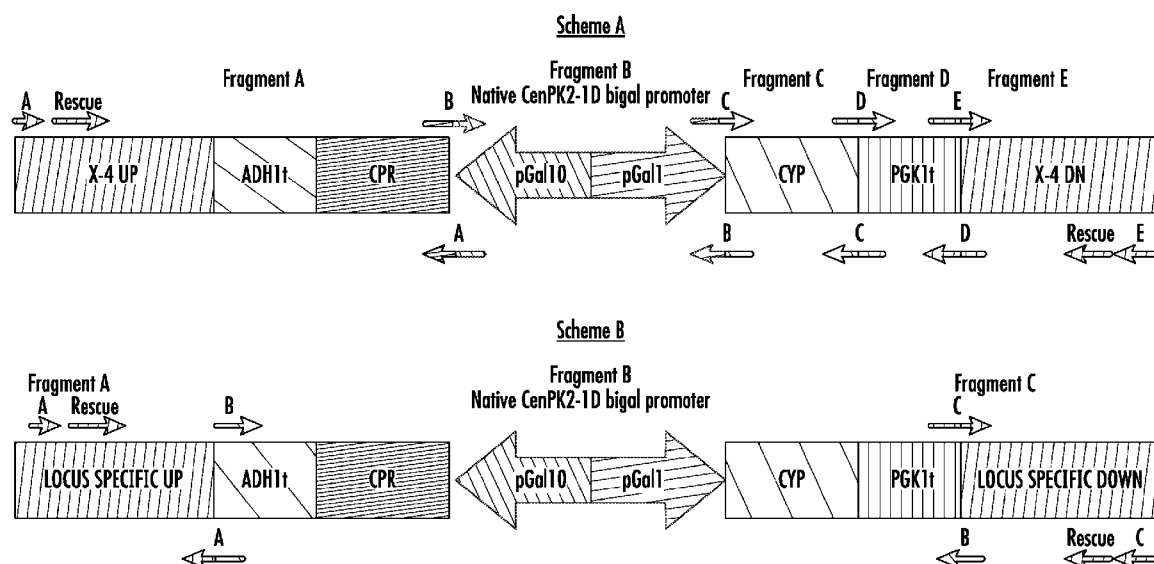


FIG. 1

(57) Abstract: The present disclosure relates to recombinant polypeptides with cytochrome P450 (CYP) activity capable of selectively catalyzing the conversion of precursor compounds, LCA or 3-KCA, to the product compounds, UDCA or 3-KUDCA, and the engineering of these polypeptides for heterologous expression in yeast allowing for enhanced fermentative bioproduction of these product compounds.

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

- *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*
- *with sequence listing part of description (Rule 5.2(a))*

RECOMBINANT POLYPEPTIDES FOR BIOSYNTHESIS OF URSODEOXYCHOLIC ACID (UDCA)

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority of U.S. Provisional Patent Application Number 63/586,338, filed September 28, 2023, and U.S. Provisional Patent Application Number 63/484,926, filed February 14, 2023, the entirety of each of which is hereby incorporated by reference herein.

FIELD

[0002] The present disclosure relates to recombinant polypeptides with cytochrome P450 (CYP) activity capable of selectively catalyzing the conversion of precursor compounds, LCA or 3-KCA, to the product compounds, UDCA or 3-KUDCA, and the engineering of these polypeptides for heterologous expression in yeast allowing for improved fermentative bioproduction of these product compounds.

REFERENCE TO SEQUENCE LISTING

[0003] The official copy of the Sequence Listing is submitted concurrently with the specification via USPTO Patent Center as an WIPO Standard ST.26 formatted XML file with file name "13421-020WO1.xml", a creation date of February 13, 2024, and a size of 1,355,138 bytes. This Sequence Listing filed via USPTO Patent Center is part of the specification and is incorporated in its entirety by reference herein.

BACKGROUND

[0004] Ursodeoxycholic acid (UDCA) is a secondary bile acid that has high therapeutic value in the treatment of cholestatic liver diseases. UDCA also has cytoprotective and anti-inflammatory properties and is being pursued as a potential therapeutic to treat neurological, cardiovascular, and inflammatory bowel diseases. Currently, UDCA is primarily harvested from animal bile (black bears, cattle, poultry). UDCA can also be produced via a costly chemical synthesis route from cholic acid that has 5-steps including a Wolff-Kishner ketone reduction, and an epimerization at C7 to produce UDCA. See e.g., Eggert, et al., *Biotechnol.* 2014, 191, 11-21. A shorter synthesis route based on the biocatalytic epimerization of chenodeoxycholic acid (CDCA) to UDCA has also been reported. See e.g., Zheng, et al., *Process Biochem.* 2015, 50, 598-604. Biosynthesis of UDCA from lithocholic acid (LCA) using an engineered cytochrome P450 monooxygenase from *Streptomyces antibioticus*, has been reported in Grobe, et al., "Engineering Regioselectivity of a P450 Monooxygenase Enables the Synthesis of Ursodeoxycholic Acid via 7 β -Hydroxylation of Lithocholic Acid," *Angew. Chem. Int. Ed.*, 60 (2): 753-757 (January 11, 2021); Pages 753-757 10.1002/anie.202012675. The engineered CYP converts LCA to UDCA preferentially over MDCA (73%:27%), but the bioconversion has very

low productivity ~ 67 mM in 24 hours. WO2022115710A1 describes a recombinant yeast transformed with a naturally occurring cytochrome P450 monooxygenase from *Gibberella zeae* that is capable of carrying out the biosynthesis of UDCA from LCA, or the structurally related bioconversion of 3-KCA to 3-KUDCA.

[0005] There remains a need for a more efficient recombinant cell systems for the commercially viable biosynthetic production UDCA.

SUMMARY

[0006] The present disclosure relates generally to recombinant polypeptides with cytochrome P450 (CYP) activity capable of selectively catalyzing the conversion of precursor compounds, LCA or 3-KCA, to the product compounds, UDCA or 3-KUDCA, and the engineering of these polypeptides for heterologous expression in yeast allowing for improved fermentative bioproduction of these product compounds. This summary is intended to introduce the subject matter of the present disclosure, but does not cover each and every embodiment, combination, or variation that is contemplated and described within the present disclosure. Further embodiments are contemplated and described by the disclosure of the detailed description, drawings, and claims.

[0007] In at least one embodiment, the present disclosure provides a recombinant host cell comprising a heterologous nucleic acid encoding a polypeptide with CYP activity capable of converting LCA to UDCA and/or converting 3-KCA to 3-KUDCA, wherein the polypeptide with CYP activity comprises an amino acid sequence at least 90% sequence identity to SEQ ID NO: 2 and an amino acid difference relative to SEQ ID NO: 2 at one or more positions selected from L7, S31, Q54, V56, K70, K76, I120, A143, L147, A153, N160, H170, G175, E176, V210, V261, T307, N316, Q320, G325, T333, E338, V364, L365, P401, E402, S408, Q424, L444, F452, K479, and M485. In at least one embodiment, the amino acid differences are selected from L7R, S31V, Q54K, V56G, K70L, K70N, K76H, I120A, A143G, A143R, L147R, L147W, A153Q, A153R, N160H, H170V, G175E, E176D, V210R, V261E, T307A, N316S, Q320M, G325E, T333V, E338Q, V364L, L365A, P401D, E402V, S408R, Q424V, L444F, F452D, K479R, and M485T.

[0008] In at least one embodiment of the recombinant host cell, the polypeptide amino acid sequence comprises a combination of amino acid differences selected from:

S31V, Q54K, A143G, T307A, G325E
S31V, Q54K, A143G, T307A, G325E, L365A
S31V, L147R, G325E
S31V, Q54K, A143R, L147R, T307A, G325E, P401D
S31V, V57F, L147W, P401D
S31V, L147W, P401D
S31V, Q54K, A143R, T307A, P401D
S31V, Q54K, P401D
S31V, A143R, L147W

S31V, A143G, G175E, P401D, K498R
S31V, A143G, L147C, G175E, P401D
S31V, P129S, A143R
S31V, G175E, P401D
S31V, A143G
S31V, A143R
S31V, Q54K, A143G, T307A, G325E, V364L
L7R, S31V, Q54K, A143G, T307A, G325E
S31V, Q54K, A143G, A153Q, T307A, G325E
S31V, Q54K, A143G, A153P, T307A, G325E
S31V, Q54K, A143G, T307A, G325E, Q424V
S31V, Q54K, A143G, E176D, T307A, G325E
S31V, Q54K, A143G, T307A, G325E, K479R
S31V, Q54K, K70L, A143G, T307A, G325E
S31V, Q54K, A143G, T307A, G325E, T333V
S31V, Q54K, A143G, T307A, G325E, E338Q
S31V, Q54K, K70N, A143G, T307A, G325E,
S31V, Q54K, A143G, V210R, T307A, G325E,
S31V, Q54K, A143G, T307A, G325E, F452D
S31V, Q54K, A143G, T307A, Q320M, G325E,
S31V, Q54K, A143G, H170V, T307A, G325E,
S31V, Q54K, A143G, N160H, T307A, G325E,
S31V, Q54K, A143G, T307A, G325E, S408R
L7R, S31V, Q54K, A143G, V210R, T307A, G325E, V364L, K479R
L7R, S31V, Q54K, K70L, A143G, N160H, H170V, E176D, T307A, G325E, V364L,
S31V, Q54K, A143G, T307A, G325E, V364L, L444F
S31V, Q54K, A143G, T307A, G325E, V364L, Q320M, E402V
S31V, Q54K, K70L, A143G, T307A, G325E, V364L,
S31V, Q54K, A143G, V261E, T307A, Q320M, G325E, T333V, V364L, K479R
S31V, Q54K, A143G, A153Q, H170V, T307A, G325E, V364L, Q424V
S31V, Q54K, K70L, A143G, T307A, G325E, V364L, E338Q, M485T
S31V, Q54K, A143G, T307A, N316S G325E, V364L,
L7R, K479R S31V, Q54K, A143G, E176D, T307A, G325E, V364L
L7R, S31V, Q54K, A143G, T307A, Q320M, G325E, V364L, K479R
L7R, S31V, Q54K, K70L, A143G, H170V, T307A, G325E, V364L, K479R
L7R, S31V, Q54K, A143G, E176D, T307A, Q320M, G325E, V364L,
S31V, Q54K, A143G, N160H, T307A, Q320M, G325E, V364L,
L7R, S31V, Q54K, A143G, A153P, T307A, G325E, V364L, S408R

[0009] In at least one embodiment of the recombinant host cell, the heterologous nucleic acid encoding the polypeptide further comprises a silent mutation selected from R328 (CGT > CGC), A368 (GCT > GCC), and V476 (GTT > GTC).

[0010] In at least one embodiment of the recombinant host cell, the heterologous nucleic acid comprises: (a) a sequence of at least 80% identity to a sequence selected from the group consisting of SEQ ID NO: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 1185, 1187, 1189, 1191, 1193, 1195, 1197, 1199, 1201, 1203,

1205, 1207, 1209, 1211, 1213, 1215, 1217, 1219, 1243, 1245, 1247, 1249, 1251, 1253, 1255, 1257, 1259, 1261, 1263, 1265, 1267, 1269, 1271, and 1275; or (b) a codon degenerate sequence of a sequence selected from the group consisting of SEQ ID NO: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 1185, 1187, 1189, 1191, 1193, 1195, 1197, 1199, 1201, 1203, 1205, 1207, 1209, 1211, 1213, 1215, 1217, 1219, 1243, 1245, 1247, 1249, 1251, 1253, 1255, 1257, 1259, 1261, 1263, 1265, 1267, 1269, 1271, and 1275.

[0011] In at least one embodiment of the recombinant host cell, the polypeptide with CYP activity comprises an amino acid sequence having at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity to SEQ ID NO: 2.

[0012] In at least one embodiment, the polypeptide with CYP activity comprises an amino acid sequence selected from SEQ ID NO: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 1186, 1188, 1190, 1192, 1194, 1196, 1198, 1200, 1202, 1204, 1206, 1208, 1210, 1212, 1214, 1216, 1218, 1220, 1244, 1246, 1248, 1250, 1252, 1254, 1256, 1258, 1260, 1262, 1264, 1266, 1268, 1270, and 1272; optionally, wherein the polypeptide is truncated by 1-31 amino acids at its N-terminus.

[0013] In at least one embodiment of the recombinant host cell, the heterologous nucleic acid encodes a polypeptide with CPR activity; optionally, wherein the polypeptide with CPR activity comprises an amino acid sequence of at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity to a sequence selected from SEQ ID NO: 4 and 1278.

[0014] In at least one embodiment of the recombinant host cell, the heterologous nucleic acid encodes a polypeptide with CYP activity capable of converting LCA to UDCA and/or converting 3-KCA to 3-KUDCA fused via a linker to a second polypeptide second polypeptide with CPR activity; optionally, wherein the fusion polypeptide comprises an amino acid sequence of SEQ ID NO: 1280

[0015] In at least one embodiment of the recombinant host cell, the heterologous nucleic acid is integrated into a site in the host cell genome. In at least one embodiment, the heterologous nucleic acid is integrated at one or more sites in the genome; optionally, integrated at three or more sites. In at least one embodiment, the heterologous nucleic acid is under the control of a promoter system selected from pGal1/10, and pCAT1:pFDH1.

[0016] In at least one embodiment of the recombinant host cell, the source organism of the recombinant host cell is selected from *Saccharomyces cerevisiae*, *Pichia pastoris*, *Yarrowia lipolytica*, and *Escherichia coli*. In at least one embodiment, the recombinant host cell is

Saccharomyces cerevisiae and the heterologous nucleic acid is integrated at one or more sites in the genome selected from X-4, XI-2, XII-4, and X-2. In at least one embodiment, the heterologous nucleic acid is integrated at three or more sites; optionally, integrated at four or more sites.

[0017] In at least one embodiment of the recombinant host cell, the recombinant host cell is *Pichia pastoris* and the heterologous nucleic acid is integrated at one or more sites in the genome selected from AOX1, Int6, Int15, and HIS4. In at least one embodiment, the heterologous nucleic acid is integrated at three or more sites; optionally, integrated at four or more sites.

[0018] In at least one embodiment, the present disclosure also provides a method for producing UDCA or 3-KUDCA comprising: (a) culturing a recombinant host cell of the present disclosure in a suitable medium comprising LCA and/or 3-KCA; and (b) recovering the produced UDCA and/or 3-KUDCA.

[0019] In at least one embodiment, the present disclosure also provides a recombinant polypeptide with CYP activity capable of converting LCA to UDCA and/or converting 3-KCA to 3-KUDCA, wherein the polypeptide comprises an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 2 and an amino acid difference relative to SEQ ID NO: 2 at one or more positions selected from L7, S31, Q54, V56, K70, K76, I120, A143, L147, A153, N160, H170, G175, E176, V210, V261, T307, N316, Q320, G325, T333, E338, V364, L365, P401, E402, S408, Q424, L444, F452, K479, and M485. In at least one embodiment, the amino acid differences are selected from L7R, S31V, Q54K, V56G, K70L, K70N, K76H, I120A, A143G, A143R, L147R, L147W, A153Q, A153R, N160H, H170V, G175E, E176D, V210R, V261E, T307A, N316S, Q320M, G325E, T333V, E338Q, V364L, L365A, P401D, E402V, S408R, Q424V, L444F, F452D, K479R, and M485T. In at least one embodiment, the polypeptide amino acid sequence comprises a combination of amino acid differences selected from:

S31V, Q54K, A143G, T307A, G325E
S31V, Q54K, A143G, T307A, G325E, L365A
S31V, L147R, G325E
S31V, Q54K, A143R, L147R, T307A, G325E, P401D
S31V, V57F, L147W, P401D
S31V, L147W, P401D
S31V, Q54K, A143R, T307A, P401D
S31V, Q54K, P401D
S31V, A143R, L147W
S31V, A143G, G175E, P401D, K498R
S31V, A143G, L147C, G175E, P401D
S31V, P129S, A143R
S31V, G175E, P401D
S31V, A143G
S31V, A143R
S31V, Q54K, A143G, T307A, G325E, V364L
L7R, S31V, Q54K, A143G, T307A, G325E

S31V, Q54K, A143G, A153Q, T307A, G325E
S31V, Q54K, A143G, A153P, T307A, G325E
S31V, Q54K, A143G, T307A, G325E, Q424V
S31V, Q54K, A143G, E176D, T307A, G325E
S31V, Q54K, A143G, T307A, G325E, K479R
S31V, Q54K, K70L, A143G, T307A, G325E
S31V, Q54K, A143G, T307A, G325E, T333V
S31V, Q54K, A143G, T307A, G325E, E338Q
S31V, Q54K, K70N, A143G, T307A, G325E,
S31V, Q54K, A143G, V210R, T307A, G325E,
S31V, Q54K, A143G, T307A, G325E, F452D
S31V, Q54K, A143G, T307A, Q320M, G325E,
S31V, Q54K, A143G, H170V, T307A, G325E,
S31V, Q54K, A143G, N160H, T307A, G325E,
S31V, Q54K, A143G, T307A, G325E, S408R
L7R, S31V, Q54K, A143G, V210R, T307A, G325E, V364L, K479R
L7R, S31V, Q54K, K70L, A143G, N160H, H170V, E176D, T307A, G325E, V364L,
S31V, Q54K, A143G, T307A, G325E, V364L, L444F
S31V, Q54K, A143G, T307A, G325E, V364L, Q320M, E402V
S31V, Q54K, K70L, A143G, T307A, G325E, V364L,
S31V, Q54K, A143G, V261E, T307A, Q320M, G325E, T333V, V364L, K479R
S31V, Q54K, A143G, A153Q, H170V, T307A, G325E, V364L, Q424V
S31V, Q54K, K70L, A143G, T307A, G325E, V364L, E338Q, M485T
S31V, Q54K, A143G, T307A, N316S G325E, V364L,
L7R, K479R S31V, Q54K, A143G, E176D, T307A, G325E, V364L
L7R, S31V, Q54K, A143G, T307A, Q320M, G325E, V364L, K479R
L7R, S31V, Q54K, K70L, A143G, H170V, T307A, G325E, V364L, K479R
L7R, S31V, Q54K, A143G, E176D, T307A, Q320M, G325E, V364L,
S31V, Q54K, A143G, N160H, T307A, Q320M, G325E, V364L,
L7R, S31V, Q54K, A143G, A153P, T307A, G325E, V364L, S408R

[0020] In at least one embodiment of the polypeptide, the polypeptide comprises an amino acid sequence having at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity to SEQ ID NO: 2.

[0021] In at least one embodiment, the polypeptide comprises an amino acid sequence selected from SEQ ID NO: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 1186, 1188, 1190, 1192, 1194, 1196, 1198, 1200, 1202, 1204, 1206, 1208, 1210, 1212, 1214, 1216, 1218, 1220, 1244, 1246, 1248, 1250, 1252, 1254, 1256, 1258, 1260, 1262, 1264, 1266, 1268, 1270, 1272, and 1276.

[0022] In at least one embodiment, the recombinant polypeptide with CYP activity capable of converting LCA to UDCA and/or converting 3-KCA to 3-KUDCA of the present disclosure comprises an N-terminal truncation of from 1 to 31 amino acids as compared to SEQ ID NO: 2. In at least one embodiment, the polypeptide comprises an N-terminal truncation of from 1 to 31 amino acids of an amino acid sequence selected from SEQ ID NO: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 1186, 1188, 1190,

1192, 1194, 1196, 1198, 1200, 1202, 1204, 1206, 1208, 1210, 1212, 1214, 1216, 1218, 1220, 1244, 1246, 1248, 1250, 1252, 1254, 1256, 1258, 1260, 1262, 1264, 1266, 1268, 1270, and 1272. In at least one embodiment, the polypeptide comprises an amino acid of SEQ ID NO: 1276.

[0023] In at least one embodiment, the recombinant polypeptide with CYP activity capable of converting LCA to UDCA and/or converting 3-KCA to 3-KUDCA of the present disclosure is fused via a linker to a second polypeptide. In at least one embodiment, the second polypeptide has CPR activity. In at least one embodiment, the second polypeptide with CPR activity comprises an amino acid sequence having at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity to a sequence selected from SEQ ID NO: 4 and 1278.

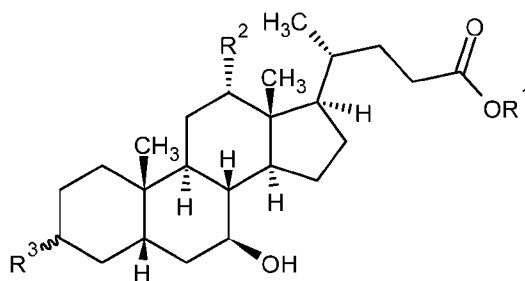
[0024] In at least one embodiment, the recombinant polypeptide with CYP activity capable of converting LCA to UDCA and/or converting 3-KCA to 3-KUDCA is fused via a linker to a second polypeptide second polypeptide with CPR activity; optionally, wherein the fusion polypeptide comprises an amino acid sequence of SEQ ID NO: 1280.

[0025] In at least one embodiment, the present disclosure also provides a polynucleotide encoding a polypeptide of the present disclosure. In at least one embodiment, the polynucleotide sequence comprises: (a) a sequence of at least 80% identity to a sequence selected from the group consisting of SEQ ID NO: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 1185, 1187, 1189, 1191, 1193, 1195, 1197, 1199, 1201, 1203, 1205, 1207, 1209, 1211, 1213, 1215, 1217, 1219, 1243, 1245, 1247, 1249, 1251, 1253, 1255, 1257, 1259, 1261, 1263, 1265, 1267, 1269, 1271, and 1275; or (b) a codon degenerate sequence of a sequence selected from the group consisting of SEQ ID NO: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 1185, 1187, 1189, 1191, 1193, 1195, 1197, 1199, 1201, 1203, 1205, 1207, 1209, 1211, 1213, 1215, 1217, 1219, 1243, 1245, 1247, 1249, 1251, 1253, 1255, 1257, 1259, 1261, 1263, 1265, 1267, 1269, 1271, and 1275.

[0026] In at least one embodiment, the present disclosure also provides an expression vector comprising a polynucleotide of the present disclosure. In at least one embodiment, the expression vector comprises a control sequence.

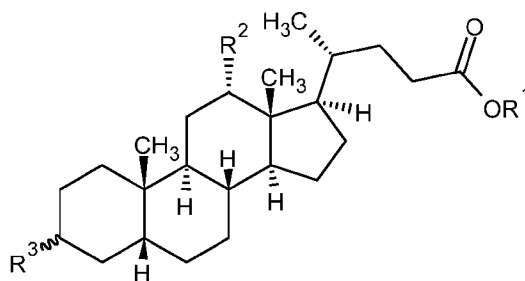
[0027] In at least one embodiment, the present disclosure also provides a host cell comprising a polynucleotide or an expression vector of the present disclosure.

[0028] In another embodiment, the present disclosure provides a method for preparing a compound of structural formula (I)



(I)

wherein, R^1 is hydrogen, a carboxylate salt counterion, an optionally substituted C1-C20 alkyl, or an optionally substituted aryl; R^2 is hydrogen, or a hydroxyl; and R^3 is a hydroxyl, or an oxo group; the method comprising contacting under suitable reactions conditions a compound of structural formula (I)



(II)

wherein, R^1 is hydrogen, a carboxylate salt counterion, an optionally substituted C1-C20 alkyl, or an optionally substituted aryl; R^2 is hydrogen, or a hydroxyl; and R^3 is a hydroxyl, or an oxo group; and a recombinant polypeptide of the present disclosure.

[0029] In at least one embodiment of the method for preparing a compound of structural formula (I), R^2 is hydrogen, and (a) the compound of structure formula (I) is UDCA and the compound of structural formula (II) is LCA; or (b) the compound of structure formula (I) is 3-KUDCA and the compound of structural formula (II) is 3-KCA.

[0030] In at least one embodiment of the method for preparing a compound of structural formula (I), R^2 is a hydroxyl, and the compound of structure formula (I) is UCA and the compound of structural formula (II) is DCA.

BRIEF DESCRIPTION OF THE DRAWINGS

[0031] A better understanding of the novel features and advantages of the present disclosure will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the disclosure are utilized, and the accompanying drawings (also "Figure" and "FIG." herein), of which:

[0032] FIG. 1 depicts Schemes A and B used for PCR primer synthesis of DNA fragments used in building strains of *Saccharomyces cerevisiae* with CPR and CYP genes under control of pGal1/10 bidirectional promoter as described in Example 1.

[0033] FIG. 2 depicts the lineage of the four *Saccharomyces cerevisiae* strains (SH013, SH021, SH020, and SHP025) built as described in Example 1.

[0034] FIG. 3 depicts lineage of *Pichia pastoris* strains built as described in Example 4.

[0035] FIG. 4 depicts lineage of *Pichia pastoris* strains built as described in Examples 7-9.

[0036] FIG. 5 depicts a plot of screening results for *Pichia pastoris* strains built as described in Example 8.

[0037] FIG. 6 depicts plots of screening results for *Pichia pastoris* strains built as described in Example 9.

DETAILED DESCRIPTION

[0038] For the descriptions herein and the appended claims, the singular forms “a”, and “an” include plural referents unless the context clearly indicates otherwise. Thus, for example, reference to “a protein” includes more than one protein, and reference to “a compound” refers to more than one compound. It is further noted that the claims may be drafted to exclude any optional element. As such, this statement is intended to serve as antecedent basis for use of such exclusive terminology as “solely,” “only” and the like in connection with the recitation of claim elements, or use of a “negative” limitation. The use of “comprise,” “comprises,” “comprising” “include,” “includes,” and “including” are interchangeable and not intended to be limiting. It is to be further understood that where descriptions of various embodiments use the term “comprising,” those skilled in the art would understand that in some specific instances, an embodiment can be alternatively described using language “consisting essentially of” or “consisting of.”

[0039] Where a range of values is provided, unless the context clearly dictates otherwise, it is understood that each intervening integer of the value, and each tenth of each intervening integer of the value, unless the context clearly dictates otherwise, between the upper and lower limit of that range, and any other stated or intervening value in that stated range, is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges, and are also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of these limits, ranges excluding (i) either or (ii) both of those included limits are also included in the invention. For example, “1 to 50,” includes “2 to 25,” “5 to 20,” “25 to 50,” “1 to 10,” etc.

[0040] Generally, the nomenclature used herein and the techniques and procedures described herein include those that are well understood and commonly employed by those of ordinary skill in the art, such as the common techniques and methodologies described in e.g., Green and Sambrook, Molecular Cloning: A Laboratory Manual (Fourth Edition), Vols. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 2012 (hereinafter “Sambrook”); and Current Protocols in Molecular Biology, F. M. Ausubel et al., eds., originally published in 1987 in book

form by Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., and regularly supplemented through 2011, and now available in journal format online as Current Protocols in Molecular Biology, Vols. 00 - 130, (1987-2020), published by Wiley & Sons, Inc. in the Wiley Online Library (hereinafter "Ausubel").

[0041] All publications, patents, patent applications, and other documents referenced in this disclosure are hereby incorporated by reference in their entireties for all purposes to the same extent as if each individual publication, patent, patent application or other document were individually indicated to be incorporated by reference herein for all purposes.

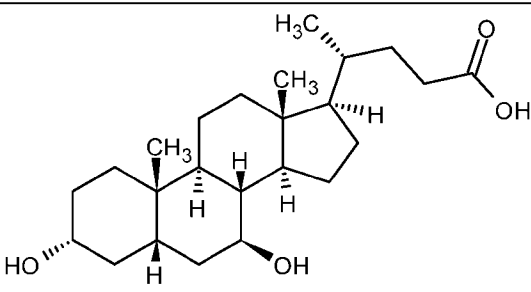
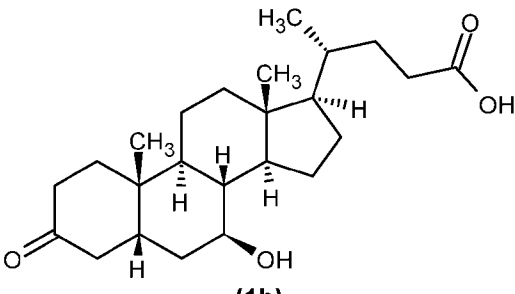
[0042] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the present invention pertains. It is to be understood that the terminology used herein is for describing particular embodiments only and is not intended to be limiting. For purposes of interpreting this disclosure, the following description of terms will apply and, where appropriate, a term used in the singular form will also include the plural form and vice versa.

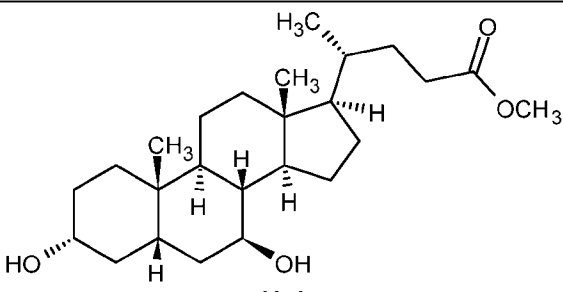
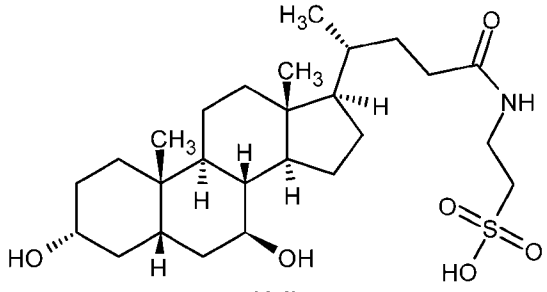
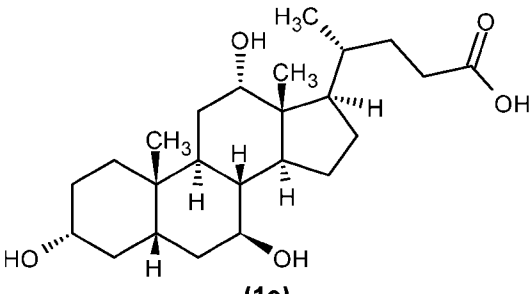
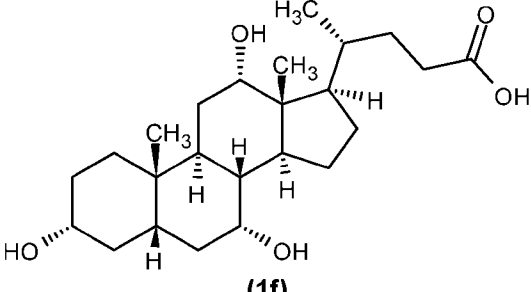
[0043] Definitions

[0044] "UDCA" refers to the compound ursodeoxycholic acid having the chemical structure shown as compound **1a** in Table 1 (below).

[0045] "UDCA derivatives" as referenced in the present disclosure include, but are not limited to, structural analogs of UDCA, such as 7 β -hydroxy-3-oxo-5 β -cholanoic acid (3-KUDCA), compound **1b**, and the other exemplary compounds shown below in Table 1 (below).

[0046] TABLE 1: Exemplary UDCA and UDCA derivative compounds

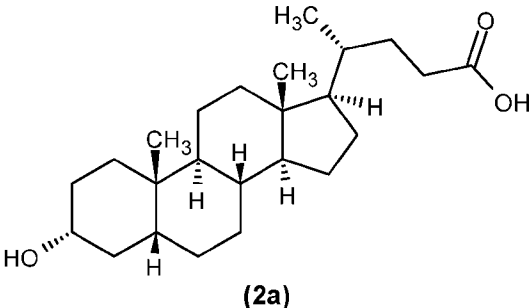
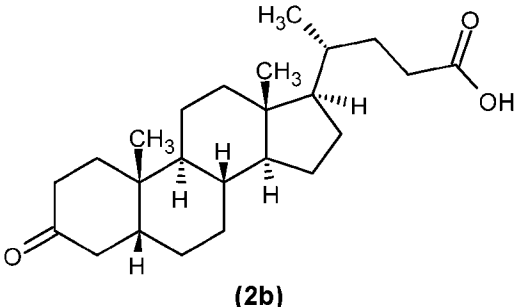
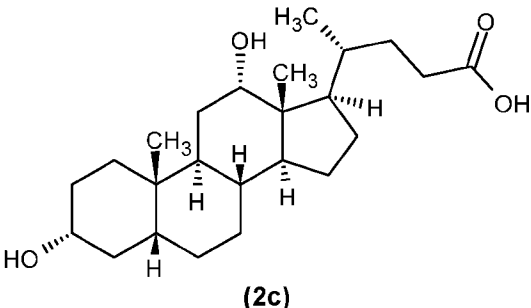
Compound Name	Abbrev. Name	Chemical Structure
Ursodeoxycholic acid	UDCA	 (1a)
7 β -hydroxy-3-oxo-5 β -cholanoic acid	3-KUDCA	 (1b)

Methyl Ursodeoxycholate	Me-UDCA	 (1c)
Ursodeoxycholytaurine	UDCT	 (1d)
Ursocholic acid	UCA	 (1e)
Cholic acid	CA	 (1f)

[0047] “UDCA precursor” as used herein refers to a compound capable of being converted into UDCA, or a UDCA derivative, by an enzyme capable producing UDCA alone, or in combination with another enzyme or a non-enzymatic chemical reaction. UDCA precursors as referenced in the present disclosure include, but are not limited to, the exemplary compounds summarized in Table 2 (below).

[0048] TABLE 2: Exemplary UDCA and UDCA derivative precursor compounds

Compound Name	Abbrev. Name	Chemical Structure
---------------	--------------	--------------------

Lithocholic acid	LCA	 (2a)
3-ketolithocholic acid	3-KCA	 (2b)
Deoxycholic acid	DCA	 (2c)

[0049] “CYP activity” as used herein refers to the catalytic activity of a cytochrome P450 monooxygenase enzyme.

[0050] “CPR activity” as used herein refers to the catalytic activity of a cytochrome P450 reductase enzyme.

[0051] “Conversion” as used herein refers to the enzymatic conversion of a substrate(s) to a corresponding product(s). “Percent conversion” refers to the percent of the substrate that is converted to the product within a period of time under specified conditions. Thus, the “enzymatic activity” or “activity” of an enzymatic conversion can be expressed as “percent conversion” of the substrate to the product.

[0052] “Substrate” as used herein in the context of an enzyme mediated process refers to the compound or molecule acted on by the enzyme.

[0053] “Product” as used herein in the context of an enzyme mediated process refers to the compound or molecule resulting from the activity of the enzyme.

[0054] “Host cell” as used herein refers to a cell capable of being functionally modified with recombinant nucleic acids and functioning to express recombinant products, including polypeptides and compounds produced by activity of the polypeptides.

[0055] “Nucleic acid,” or “polynucleotide” as used herein interchangeably to refer to two or more nucleosides that are covalently linked together. The nucleic acid may be wholly comprised ribonucleosides (e.g., RNA), wholly comprised of 2'-deoxyribonucleotides (e.g., DNA) or mixtures of ribo- and 2'-deoxyribonucleosides. The nucleoside units of the nucleic acid can be linked together via phosphodiester linkages (e.g., as in naturally occurring nucleic acids), or the nucleic acid can include one or more non-natural linkages (e.g., phosphorothioester linkage). Nucleic acid or polynucleotide is intended to include single-stranded or double-stranded molecules, or molecules having both single-stranded regions and double-stranded regions. Nucleic acid or polynucleotide is intended to include molecules composed of the naturally occurring nucleobases (i.e., adenine, guanine, uracil, thymine, and cytosine), or molecules comprising that include one or more modified and/or synthetic nucleobases, such as, for example, inosine, xanthine, hypoxanthine, etc.

[0056] “Protein,” “polypeptide,” and “peptide” are used herein interchangeably to denote a polymer of at least two amino acids covalently linked by an amide bond, regardless of length or post-translational modification (e.g., glycosylation, phosphorylation, lipidation, myristilation, ubiquitination, etc.). As used herein “protein” or “polypeptide” or “peptide” polymer can include D- and L-amino acids, and mixtures of D- and L-amino acids.

[0057] “Naturally-occurring” or “wild-type” as used herein refers to the form as found in nature. For example, a naturally occurring nucleic acid sequence is the sequence present in an organism that can be isolated from a source in nature, and which has not been intentionally modified by human manipulation.

[0058] “Recombinant,” “engineered,” or “non-naturally occurring” when used herein with reference to, e.g., a cell, nucleic acid, or polypeptide, refers to a material, or a material corresponding to the natural or native form of the material, that has been modified in a manner that would not otherwise exist in nature, or is identical thereto but is produced or derived from synthetic materials and/or by manipulation using recombinant techniques. Non-limiting examples include, among others, recombinant cells expressing genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise expressed at a different level.

[0059] “Nucleic acid derived from” as used herein refers to a nucleic acid having a sequence at least substantially identical to a sequence of found in naturally in an organism. For example, cDNA molecules prepared by reverse transcription of mRNA isolated from an organism, or nucleic acid molecules prepared synthetically to have a sequence at least substantially identical to, or which hybridizes to a sequence at least substantially identical to a nucleic sequence found in an organism.

[0060] “Coding sequence” refers to that portion of a nucleic acid (e.g., a gene) that encodes an amino acid sequence of a protein.

[0061] "Heterologous nucleic acid" as used herein refers to any polynucleotide that is introduced into a host cell by laboratory techniques and includes polynucleotides that are removed from a host cell, subjected to laboratory manipulation, and then reintroduced into a host cell.

[0062] "Codon degenerate" describes a nucleotide sequence that has one or more different codons relative to the reference nucleotide sequence, but which encodes a polypeptide that is identical to the polypeptide encoded by a reference nucleotide sequence. The different codons between the nucleotide sequence and the reference nucleotide sequence are called "synonyms" or "synonymous" codons in that they use different triplets of nucleotides to encode the same amino acid in a polypeptide.

[0063] "Codon optimized" refers to changes in the codons of the polynucleotide encoding a protein to those preferentially used in a particular organism such that the encoded protein is efficiently expressed in the organism of interest. Although the genetic code is degenerate in that most amino acids are represented by several different "synonymous" codons, it is well known that codon usage by particular organisms is nonrandom and biased towards particular codon triplets. This codon usage bias may be higher in reference to a given gene, genes of common function or ancestral origin, highly expressed proteins versus low copy number proteins, and the aggregate protein coding regions of an organism's genome. In some embodiments, the polynucleotides encoding the imine reductase enzymes may be codon optimized for optimal production from the host organism selected for expression.

[0064] "Preferred, optimal, high codon usage bias codons" refers to codons that are used at higher frequency in the protein coding regions than other codons that code for the same amino acid. The preferred codons may be determined in relation to codon usage in a single gene, a set of genes of common function or origin, highly expressed genes, the codon frequency in the aggregate protein coding regions of the whole organism, codon frequency in the aggregate protein coding regions of related organisms, or combinations thereof. Codons whose frequency increases with the level of gene expression are typically optimal codons for expression. A variety of methods are known for determining the codon frequency (e.g., codon usage, relative synonymous codon usage) and codon preference in specific organisms, including multivariate analysis, for example, using cluster analysis or correspondence analysis, and the effective number of codons used in a gene (see GCG CodonPreference, Genetics Computer Group Wisconsin Package; CodonW, John Peden, University of Nottingham; McInerney, J. O, 1998, *Bioinformatics* 14:372-73; Stenico et al., 1994, *Nucleic Acids Res.* 22:2437-46; Wright, F., 1990, *Gene* 87:23-29). Codon usage tables are available for a growing list of organisms (see for example, Wada et al., 1992, *Nucleic Acids Res.* 20:2111-2118; Nakamura et al., 2000, *Nucl. Acids Res.* 28:292; Duret, et al., *supra*; Henaut and Danchin, "Escherichia coli and Salmonella," 1996, Neidhardt, et al. Eds., ASM Press, Washington D.C., p. 2047-2066. The data source for obtaining codon usage may rely on any available nucleotide sequence capable of coding for a

protein. These data sets include nucleic acid sequences actually known to encode expressed proteins (e.g., complete protein coding sequences-CDS), expressed sequence tags (ESTS), or predicted coding regions of genomic sequences (see for example, Mount, D., *Bioinformatics: Sequence and Genome Analysis*, Chapter 8, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 2001; Uberbacher, E. C., 1996, *Methods Enzymol.* 266:259-281; Tiwari et al., 1997, *Comput. Appl. Biosci.* 13:263-270).

[0065] "Control sequence" as used herein refers to all sequences, which are necessary or advantageous for the expression of a polynucleotide and/or polypeptide as used in the present disclosure. Each control sequence may be native or foreign to the nucleic acid sequence encoding a polypeptide. Such control sequences include, but are not limited to, a leader, a promoter, a polyadenylation sequence, a pro-peptide sequence, a signal peptide sequence, and a transcription terminator. At a minimum, control sequences typically include a promoter, and transcriptional and translational stop signals. The control sequences may be provided with linkers for the purpose of introducing specific restriction sites facilitating ligation of the control sequences with the coding region of the nucleic acid sequence encoding a polypeptide.

[0066] "Operably linked" as used herein refers to a configuration in which a control sequence is appropriately placed (e.g., in a functional relationship) at a position relative to a polynucleotide sequence or polypeptide sequence of interest such that the control sequence directs or regulates the expression of the sequence of interest.

[0067] "Promoter sequence" refers to a nucleic acid sequence that is recognized by a host cell for expression of a polynucleotide of interest, such as a coding sequence. The promoter sequence contains transcriptional control sequences, which mediate the expression of a polynucleotide of interest. The promoter may be any nucleic acid sequence which shows transcriptional activity in the host cell of choice including mutant, truncated, and hybrid promoters, and may be obtained from genes encoding extracellular or intracellular polypeptides either homologous or heterologous to the host cell.

[0068] "Percentage of sequence identity," "percent sequence identity," "percentage homology," or "percent homology" are used interchangeably herein to refer to values quantifying comparisons of the sequences of polynucleotides or polypeptides, and are determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (or gaps) as compared to the reference sequence for optimal alignment of the two sequences. The percentage values may be calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Alternatively, the percentage may be calculated by determining the number of positions at which either the identical nucleic acid base or amino

acid residue occurs in both sequences or a nucleic acid base or amino acid residue is aligned with a gap to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Those of skill in the art appreciate that there are many established algorithms available to align two sequences. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith and Waterman, 1981, Adv. Appl. Math. 2:482, by the homology alignment algorithm of Needleman and Wunsch, 1970, J. Mol. Biol. 48:443, by the search for similarity method of Pearson and Lipman, 1988, Proc. Natl. Acad. Sci. USA 85:2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the GCG Wisconsin Software Package), or by visual inspection (see generally, Current Protocols in Molecular Biology, F. M. Ausubel et al., eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (1995 Supplement) (Ausubel)). Examples of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al., 1990, J. Mol. Biol. 215: 403-410 and Altschul et al., 1977, Nucleic Acids Res. 3389-3402, respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information website. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as, the neighborhood word score threshold (Altschul et al, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff and Henikoff, 1989, Proc Natl Acad Sci USA 89:10915). Exemplary determination of sequence alignment and % sequence identity can

employ the BESTFIT or GAP programs in the GCG Wisconsin Software package (Accelrys, Madison Wis.), using default parameters provided.

[0069] "Reference sequence" refers to a defined sequence used as a basis for a sequence comparison. A reference sequence may be a subset of a larger sequence, for example, a segment of a full-length nucleic acid or polypeptide sequence. A reference sequence typically is at least 20 nucleotide or amino acid residue units in length but can also be the full length of the nucleic acid or polypeptide. Since two polynucleotides or polypeptides may each (1) comprise a sequence (i.e., a portion of the complete sequence) that is similar between the two sequences, and (2) may further comprise a sequence that is divergent between the two sequences, sequence comparisons between two (or more) polynucleotides or polypeptide are typically performed by comparing sequences of the two polynucleotides or polypeptides over a "comparison window" to identify and compare local regions of sequence similarity.

"Comparison window" refers to a conceptual segment of at least about 20 contiguous nucleotide positions or amino acids residues wherein a sequence may be compared to a reference sequence of at least 20 contiguous nucleotides or amino acids and wherein the portion of the sequence in the comparison window may comprise additions or deletions (or gaps) of 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences.

[0070] "Substantial identity" or "substantially identical" refers to a polynucleotide or polypeptide sequence that has at least 70% sequence identity, at least 80% sequence identity, at least 85% sequence identity, at least 90% sequence identity, at least 95 % sequence identity, or at least 99% sequence identity, as compared to a reference sequence over a comparison window of at least 20 nucleoside or amino acid residue positions, frequently over a window of at least 30-50 positions, wherein the percentage of sequence identity is calculated by comparing the reference sequence to a sequence that includes deletions or additions which total 20 percent or less of the reference sequence over the window of comparison.

[0071] "Corresponding to," "reference to," or "relative to" when used in the context of the numbering of a given amino acid or polynucleotide sequence refers to the numbering of the residues of a specified reference sequence when the given amino acid or polynucleotide sequence is compared to the reference sequence. In other words, the residue number or residue position of a given polymer is designated with respect to the reference sequence rather than by the actual numerical position of the residue within the given amino acid or polynucleotide sequence. For example, a given amino acid sequence, such as that of an engineered imine reductase, can be aligned to a reference sequence by introducing gaps to optimize residue matches between the two sequences. In these cases, although the gaps are present, the numbering of the residue in the given amino acid or polynucleotide sequence is made with respect to the reference sequence to which it has been aligned.

[0072] “Isolated” as used herein in reference to a molecule means that the molecule (e.g., UDCA, polynucleotide, polypeptide) is substantially separated from other compounds that naturally accompany it, e.g., protein, lipids, and polynucleotides. The term embraces nucleic acids which have been removed or purified from their naturally occurring environment or expression system (e.g., host cell or in vitro synthesis).

[0073] “Substantially pure” refers to a composition in which a desired molecule is the predominant species present (i.e., on a molar or weight basis it is more abundant than any other individual macromolecular species in the composition) and is generally a substantially purified composition when the object species comprises at least about 50 percent of the macromolecular species present by mole or % weight.

[0074] “Recovered” as used herein in relation to an enzyme, protein, or compound (e.g., UDCA), refers to a more or less pure form of the enzyme, protein, or compound.

[0075] Engineered Genes Encoding Recombinant Polypeptides with CYP Activity

[0076] The present disclosure provides engineered genes encoding a polypeptide with cytochrome P450 (CYP) monooxygenase activity capable of converting LCA to UDCA and/or capable of converting 3-KCA to 3-KUDCA. The engineered CYP genes are derived from a naturally occurring parent CYP gene isolated from *Gibberella zeae* that encodes the 513 amino acid CYP polypeptide of SEQ ID NO: 2.

MATDLDLVLGKSQYALFCGITLEFSFFILKYSLGNGGKQYPYINPKKPFELSNQRVVQDFIENARDILTK
GRSLYKDTPTYKAHTDLGDVLPPEFADALKSERQLDFTEVARDDTHGYIPGFEPIGSPFDLVPLVNKYL
TRALAKLTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMRIVSRMSSRIFMGEELCKDDDWLKVSIEYTV
QLFQTADLRLNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQPHIDRRNAVKEAIAEGRPSPFDDSEW
FENEYEGKSDPATEQIKLSLVAIHTTTDLLSETMFNIALQPELLGPLREEIVTVLSTEGLKKTSFYNLKL
MDSVIKESQRLRPVLLGAFRRMALADVTLPLNGDVIKKGTKIICDTHQWNPEYYPDASKFNAYRFLQMRQ
TPGQDKRAHLVSTSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKDYDWKLPEGVVPKSKALGMSLLGD
REAKLMVKRRAAEIDIDTIGSDE (SEQ ID NO: 2)

[0077] In at least one embodiment of the present disclosure, the 513 amino acid CYP polypeptide of SEQ ID NO: 2 can be encoded by the yeast codon-optimized 1542 nucleotide sequence of SEQ ID NO: 1.

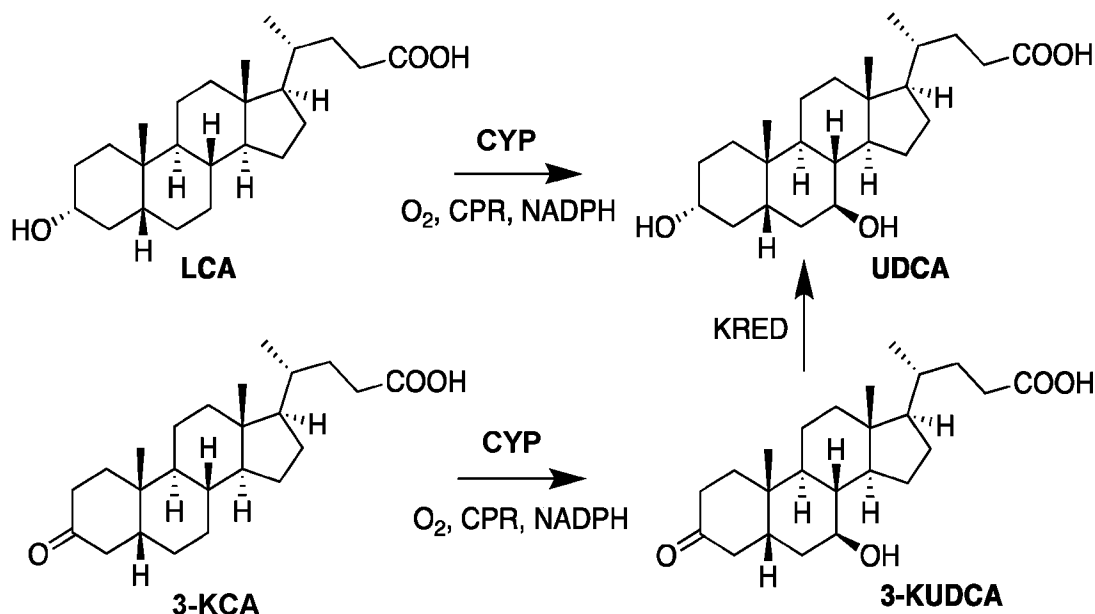
ATGCCACCGATCTAGACCTAGTATTAGGAAAAAGTCAATACGCATTATTTTGTGGCATAACTTTATTTA
GCTTTTTCATACTAAAGTATAGTCTTCTCGGAAACGGGGGCAAGCAATATCCTTATATCAACCCTAAGAA
ACCTTTTGAGCTGTCTGAATCAACGAGTTGTCCAGGATTTTATCGAGAACGCACGAGACATTTTGACTAAG
GGTCGCTCACTTTACAAGGATACGCCCTACAAGGCTCATACCGATTTAGGGGACGTCTTGGTAATCCCGC
CTGAATTTGCCGACGCTCTAAAATCTGAAAGACAGTTAGATTTTACCGAAGTCGCAAGAGACGATACTCA
CGGTTATATTCTTGGATTTCGAGCCAATCGGTTCCCGTTTCGATCTGGTGCCGCTCGTTAACAAGTATCTT
ACAAGGGCTTTGGCAAACTAACAAAACCATTTGTGGGCCGAAGCTTCTTTAGGTGTAAACCATGTTTTAG
GCACTTCTACTGAGTGGCATCTTATTAACCCAGGGGAAGATATCATGAGGATTGTCTCCAGAATGTCATC
CAGAATATTCATGGGTGAGGAACCTTTGTAAAGATGATGATTGGTTGAAAGTTAGTATTGAGTACACTGTG
CAATTGTTTCAAACCGCAGACGAATTACGTAACCTATCCACGTTGGACAAGACCATATATTCACTGGTTTT
TGCTTTCCTGTCAAGGGGTTAGGAGAAAATTGCAGGAAGCGCGTGATTTATTGCAACCCCATATTGATAG
GAGAAATGCTGTGAAGAAAGAAGCTATCGCTGAAGGTAGACCATCACCATTTCGACGATTCAATAGAAATGG
TTTGAAAATGAATACGAGGGCAAATCTGATCCCGCCACTGAACAAATTAACTATCACTGGTGGCGATTTC
ACACAACAACTGATTTACTGTCTGAAACAATGTTTAATATAGCTTTGCAACCAGAACTCCTTGGTCCACT
ACGTGAAGAGATAGTTACGGTTTTATCCACGGAAGGTCTAAAAAAGACGTCGTTTTACAATTTGAAGTTG

ATGGATTTCGGTCATAAAAAGAGTACAGAGACTTCGCCCTGTTTTATTAGGTGCTTTTCAGGAGAATGGCAT
 TGGCTGATGTTACCTTGCCCAATGGCGACGTAATTAAAAAAGGTACCAAGATCATTTGCGACACTACACA
 TCAATGGAATCCAGAATACTATCCAGATGCTAGTAAGTTCAATGCATATAGATTTTTGCAAAATGAGACAG
 ACACCGGGTCAGGACAAAAGAGCACACCTTGTGAGCACAAGCCATGATCAAATGGGATTTCGGACATGGCT
 TGCACGCGTGCCCGAGGAAGATTTTTTCGAGCCAATGAGATTAAGATTGCGCTGTGTCATATGCTATTGAA
 ATATGATTGGAAATTACCAGAAGGTGTTGTACCTAAGTCTAAGGCTTTAGGAATGTCTTTACTTGGTGAC
 CGGGAAGCCAAACTGATGGTGAAAAGGAGAGCAGCCGAAATCGATATAGACACTATTGGTAGTGATGAAT
 AA (SEQ ID NO: 1)

[0078] As described elsewhere herein, the polynucleotide sequence of SEQ ID NO: 1, or alternative codon-optimized sequences that encode the CYP polypeptide of SEQ ID NO: 2, can be used as the parent gene sequence for engineering of the CYP polypeptide for production of UDCA in yeast.

[0079] In the presence of a cytochrome P450 reductase (CPR) enzyme, the cofactor NADPH, and oxygen, the 7-position hydroxylating monooxygenase activity of the recombinant CYP polypeptides encoded by the engineered genes of the present disclosure is capable of catalyzing the conversion of LCA (compound **2a**) to UDCA (compound **1a**) as shown by the upper reaction illustrated in Scheme 1 (below).

Scheme 1

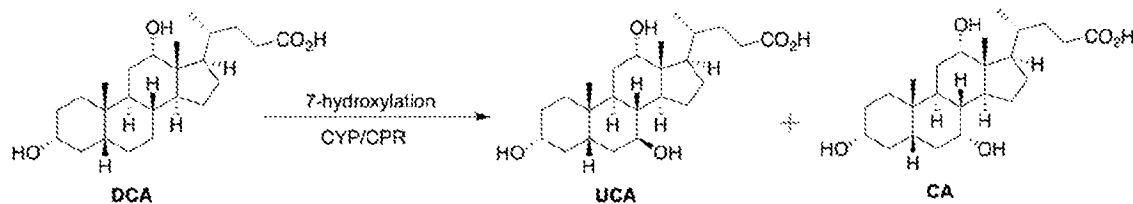


[0080] Similarly, the CYP activity of the polypeptides encoded by the engineered genes of the present disclosure is capable of catalyzing the conversion of 3-KCA (compound **2b**) to 3-KUDCA (compound **1b**) as shown by the lower reaction illustrated in Scheme 1. The 3-KUDCA product can then be converted to UDCA by a further keto reduction reaction carried out by a ketoreductase (KRED) enzyme.

[0081] Additionally, the CYP activity of the polypeptides encoded by the engineered genes of the present disclosure is capable of catalyzing the 7-position hydroxylation of DCA (compound

2c) to UCA (compound 1e) and/or CA (compound 1f) as shown by the reaction illustrated in Scheme 2.

Scheme 2



[0082] A CPR with the reductase activity to facilitate the CYP activity of SEQ ID NO: 2 (or an engineered variant thereof) in carrying out the LCA to UDCA conversion of Scheme 1 (or the DCA to UCA/CA conversion of Scheme 2) can be provided by the 264 amino acid CPR from *Gibberella zeae* of SEQ ID NO: 4.

MALRTSLSRPVLLATLTASAIGVLSILSKMMFSTASAESPSPQKIFSGAFASVKLPLHSSEYESHDTKRL
RFKLPQETAVTGLPLAYLVHIPPSSHQRLDTPDEPGYMDLLVKKYPKGQGSTYLHSLQPGDLSFTSLP
LKPAWKTNFPHITLIAGGCGITPLFNLAQGILRDPAEKTRMTFIFGARSDEVDLLKKELDGFKEFPER
FEVKYTALLEEVLGGVGRDTKVFCGPKEMEALVGGRGVLKEIGFEKSQIHTF (SEQ ID NO: 4)

[0083] In at least one embodiment of the present disclosure, the 264 amino acid CPR polypeptide of SEQ ID NO: 4 can be encoded by a yeast codon-optimized 795 nucleotide sequence of SEQ ID NO: 3.

ATGGCACTACGAACATCACTCTCTCGCCCCGTTCCCTTGTTAGCTACCCTGACTGCCAGCGCAATCGGAG
TATCTATATTGTCTAAATGATGTTTTCAACCGCCAGTGCCGAGAGTCCATCTCCGCAAAAAATTTTTTC
CGGTGCTTTTGCTTCCGTAATACTACCGCTGCATTCAAGTGAATACGAATCCCATGATACAAAGAGGTTG
CGTTTCAAACCTTCCCTCAAGAGACTGCAGTCACGGGTTTACCATTAGCTTACTTGGTTTCACATTCCACCTA
GCCACCATCAAAGAGACTTGACTACGCCGGATGAACCTGGATACATGGACCTGTTGGTGAAGAAATATCC
CAAAGGTCAGGGTTCGACATATCTACACTCGCTCCAGCCAGGTGATACCTTATCATTACATCTCTACCA
TTGAAACCAGCGTGGAACAAATAATTTTCTCATATCACTCTTATTGCTGGAGGGTGTGGGATTACGC
CATTATTCAACTTGGCTCAAGGTATACTTAGAGATCCTGCCGAAAAAACTAGGATGACCTTTATTTTTGG
TGCAAGATCAGACGAGGATGTTTTACTGAAAAAGGAGTTAGATGGCTTTGCAAAAGAGTTCCAGAAAGA
TTCGAAGTGAAATATACAGCGCTTTTGGAAAGATCCTAGGAGGCGTGGGTGCTGATACTAAGGTTTTTG
TCTGCGGTCTTAAGGAAATGGAAAGGCTTTAGTTGGAGGCAGAGGTGTATTAAAGGAAATAGGCTTCGA
AAAGTCTCAAATCCATACTTTTTTAA (SEQ ID NO: 3)

[0084] As described elsewhere herein, the polynucleotide of SEQ ID NO: 3, or an alternative codon-optimized polynucleotide that encodes a CPR polypeptide of SEQ ID NO: 4, can be used in a heterologous nucleic acid along with a gene (such as SEQ ID NO: 1) encoding the CYP polypeptide of SEQ ID NO: 2 to generate a recombinant host cell capable of carrying out the biocatalytic conversion of LCA to UDCA and/or 3-KCA to 3-KUDCA

[0085] WO2022115710A1 describes a yeast strain denoted "SAND122" that is transformed with a gene encoding the CYP of SEQ ID NO: 2 and a gene encoding the CPR of SEQ ID NO: 4. The SAND122 strain is capable of converting LCA to UDCA and/or converting 3-KCA to 3-KUDCA.

[0086] In at least one embodiment, when a host cell (e.g., *S. cerevisiae* or *P. pastoris*) is transformed with a heterologous nucleic acid comprising an engineered CYP gene of the present disclosure, a gene encoding a CPR polypeptide (e.g., SEQ ID NO: 4), and the

recombinant host cell is fed LCA, the product UDCA, is produced by the host cell in greater yield relative to a comparable recombinant host cell integrated with the parent gene encoding the wild-type CYP polypeptide of SEQ ID NO: 2. Without intending to be bound by any particular theory or mechanism, the enhanced yield of the UDCA and/or 3-KUDCA biosynthetic product is correlated with the one or more amino acid residue differences in recombinant polypeptides of the present disclosure, as compared to the amino acid sequence of wild-type CYP from *G. zeae* of SEQ ID NO: 2 from which the engineered polypeptide sequences are derived.

[0087] Exemplary engineered CYP genes and encoded recombinant polypeptides with CYP activity that exhibit the unexpected and surprising technical effect of comparable or increased UDCA yield when integrated in a recombinant host cell are summarized in Table 3 below (as well as in the following Examples and the accompanying Sequence Listing).

[0088] TABLE 3: Engineered CYP polypeptides

CYP Amino Acid Sequence	AA difference ¹	Silent codon change ²	NT SEQ ID NO:	AA SEQ ID NO:
MATDLDLVLGKSQYALFCGITLEFSFFILKYVLLGNG GKQYPYINPKKPFELSNQRVVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRAL AKLTKPLWAEASLGVNHLVLTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSIETYVQLFQTA DELRNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSIWFENEYEGK SDPATEQIKLSLVAIHTTTDLLSETMFNIALQPELL GPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTL PNGDVIKKGT KIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGGFGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE*	S31V		5	6
MATDLDLVLGKSQYALFCGITLEFSFFILKYSLLGNG GKQYPYINPKKPFELSNQRVVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRRL AKLTKPLWAEASLGVNHLVLTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSIETYVQLFQTA DELRNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSIWFENEYEGK SDPATEQIKLSLVAIHTTTDLLSETMFNIALQPELL GPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTL PNGDVIKKGT KIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGGFGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE*	A143R		7	8
MATDLDLVLGKSQYALFCGITLEFSFFILKYSLLGNG GKQYPYINPKKPFELSNQRVVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF	L147W		9	10

TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRAL AKWTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DEL RNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIHTTTDLLSETMFNIALQPELL GPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE*				
MATDLDLVLGKSQYALFCGITLFSFFILKYSLLGNG GKQYPYINPKKPFELSNQRVVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DEL RNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIHTTTDLLSETMFNIALQPELL GPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE*	A143G		11	12
MATDLDLVLGKSQYALFCGITLFSFFILKYSLLGNG GKQYPYINPKKPFELSNQRVVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRAL AKLTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DEL RNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIHTTTDLLSETMFNIALQPELL GPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE*	P401D		13	14
MATDLDLVLGKSQYALFCGITLFSFFILKYSLLGNG GKQYPYINPKKPFELSNQRVVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRAL AKLTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DEL RNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIHTTTDLLSETMFNIALQPELL GPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE*	G175E		15	16

MATDLDLVLGKSQYALFCGITLFSFFILKYSLLGNG GKQYPYINPKKPFELSNKRVVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRAL AKLTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSIEYTVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSIWFENEYEGK SDPATEQIKLSLVAIHTTTDLLSETMFNIALQPELL GPLREEIVTVLSTEGLKKTsfYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTLPNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE*	Q54K		17	18
MATDLDLVLGKSQYALFCGITLFSFFILKYSLLGNG GKQYPYINPKKPFELSNQRVVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRAL AKLTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSIEYTVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSIWFENEYEGK SDPATEQIKLSLVAIHTTTADLLSETMFNIALQPELL GPLREEIVTVLSTEGLKKTsfYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTLPNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE*	T307A		19	20
MATDLDLVLGKSQYALFCGITLFSFFILKYSLLGNG GKQYPYINPKKPFELSNQRVVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRAL AKRTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSIEYTVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSIWFENEYEGK SDPATEQIKLSLVAIHTTTDLLSETMFNIALQPELL GPLREEIVTVLSTEGLKKTsfYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTLPNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE*	L147R		21	22
MATDLDLVLGKSQYALFCGITLFSFFILKYSLLGNG GKQYPYINPKKPFELSNQRGVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRAL AKLTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSIEYTVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSIWFENEYEGK SDPATEQIKLSLVAIHTTTDLLSETMFNIALQPELL GPLREEIVTVLSTEGLKKTsfYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTLPNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE*	V56G		23	24

DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE*				
MATDLDLVLGKSQYALFCGITLEFSFFILKYSLLGNG GKQYPYINPKKPFELSNQRVVQDFIENARDILTKGR SLYHDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRAL AKLTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIHTTTDLSETMFNIALQPELL GPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE*	K76H		25	26
MATDLDLVLGKSQYALFCGITLEFSFFILKYSLLGNG GKQYPYINPKKPFELSNQRVVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYAPGFEPIGSPFDLVPLVNKYLTRAL AKLTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIHTTTDLSETMFNIALQPELL GPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE*	I120A		27	28
MATDLDLVLGKSQYALFCGITLEFSFFILKYSLLGNG GKQYPYINPKKPFELSNQRVVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRAL AKLTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIHTTTDLSETMFNIALQPELL EPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE*	G325E		29	30
MATDLDLVLGKSQYALFCGITLEFSFFILKYVLLGNG GKQYPYINPKKPFELSNKRVVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTL PNGDVIKKGTKIICDTT	S31V, Q54K, A143G, T307A, G325E		31	32

HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE				
MATDLDLVLGKSQYALFCGITLEFSFFILKYVLLGNG GKQYPYINPKKPFELS NQRVVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRAL AKRTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DEL RNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIHTTTDLLSETMFNIALQPELL EPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, L147R, G325E	V476 (GTT > GTC)	33	34
MATDLDLVLGKSQYALFCGITLEFSFFILKYVLLGNG GKQYPYINPKKPFELS NKR VVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRRL AKRTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DEL RNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNDEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, Q54K, A143R, L147R, T307A, G325E, P401D	V476 (GTT > GTC)	35	36
MATDLDLVLGKSQYALFCGITLEFSFFILKYVLLGNG GKQYPYINPKKPFELS NQR VVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRAL AKWTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DEL RNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIHTTTDLLSETMFNIALQPELL GPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNDEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, V57F, L147W, P401D	A368 (GCT > GCC)	37	38
MATDLDLVLGKSQYALFCGITLEFSFFILKYVLLGNG GKQYPYINPKKPFELS NQR VVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRAL AKWTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DEL RNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIHTTTDLLSETMFNIALQPELL	S31V, L147W, P401D	A368 (GCT > GCC)	39	40

GPLREEIVTVLSTEGLKKTSFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTLPNGDVIKKGTKIICDTT HQWNDEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE				
MATDLDLVLGKSQYALFCGITLEFSFFILKYVLLGNG GKQYPYINPKKPFELSNKRVVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRRL AKLTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIH TTADLLSETMFNIALQPELL GPLREEIVTVLSTEGLKKTSFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTLPNGDVIKKGTKIICDTT HQWNDEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, Q54K, A143R, T307A, P401D		41	42
MATDLDLVLGKSQYALFCGITLEFSFFILKYVLLGNG GKQYPYINPKKPFELSNKRVVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRAL AKLTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIH TTDLSETMFNIALQPELL GPLREEIVTVLSTEGLKKTSFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTLPNGDVIKKGTKIICDTT HQWNDEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, Q54K, P401D		43	44
MATDLDLVLGKSQYALFCGITLEFSFFILKYVLLGNG GKQYPYINPKKPFELSNQRVVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRRL AKWTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIH TTDLSETMFNIALQPELL GPLREEIVTVLSTEGLKKTSFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTLPNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, A143R, L147W		45	46
MATDLDLVLGKSQYALFCGITLEFSFFILKYVLLGNG GKQYPYINPKKPFELSNQRVVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLGVNHVLGTSTEWHPINPEEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ	S31V, A143G, G175E, P401D, K498R		47	48

PHIDRRNAVKKKEAIAEGRPSPFDDSIWFENEYEGK SDPATEQIKLSLVAIHTTTDLLSETMFNIALQPELL GPLREEIVTVLSTEGCLKKTSFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTLPNGDVIKKGTKIICDTT HQWNDEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVRRRAEI DIDTIGSDE				
MATDLDLVLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSNQRVVQDFIENARDILTGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKCTKPLWAEASLGVNHVLGTSTEWHPINPEEDIMR IVSRMSSRIFMGEELCKDDDWLKVSIEYTVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSIWFENEYEGK SDPATEQIKLSLVAIHTTTDLLSETMFNIALQPELL GPLREEIVTVLSTEGCLKKTSFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTLPNGDVIKKGTKIICDTT HQWNDEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVRRRAEI DIDTIGSDE	S31V, A143G, L147C, G175E, P401D		49	50
MATDLDLVLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSNQRVVQDFIENARDILTGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSSFDLVPLVNKYLTRRL AKLTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSIEYTVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSIWFENEYEGK SDPATEQIKLSLVAIHTTTDLLSETMFNIALQPELL GPLREEIVTVLSTEGCLKKTSFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTLPNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVRRRAEI DIDTIGSDE	S31V, P129S, A143R		51	52
MATDLDLVLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSNQRVVQDFIENARDILTGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRAL AKLTKPLWAEASLGVNHVLGTSTEWHPINPEEDIMR IVSRMSSRIFMGEELCKDDDWLKVSIEYTVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSIWFENEYEGK SDPATEQIKLSLVAIHTTTDLLSETMFNIALQPELL GPLREEIVTVLSTEGCLKKTSFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTLPNGDVIKKGTKIICDTT HQWNDEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVRRRAEI DIDTIGSDE	S31V, G175E, P401D	R328 (CGT > CGC)	53	54
MATDLDLVLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSNQRVVQDFIENARDILTGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR	S31V, A143G		55	56

IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DEL RNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EWFENEYEGK SDPATEQIKLSLVAIHTTTDLLSETMFNIALQPELL GPLREEIVTVLSTEGCLKKTSFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGGFGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE				
MATDLDLVLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSNQRVVQDFIENARDILT KGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRRL AKLTKPLWAEASLG VNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DEL RNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EWFENEYEGK SDPATEQIKLSLVAIHTTTDLLSETMFNIALQPELL GPLREEIVTVLSTEGCLKKTSFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGGFGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, A143R		57	58
MATDLDLVLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSNKR VVQDFIENARDILT KGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLG VNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DEL RNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EWFENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTEGCLKKTSFYNLKLMDSVIKESQR LRPVALLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGGFGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, Q54K, A143G, T307A, G325E, L365A		1185	1186
MATDLDLVLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSNKR VVQDFIENARDILT KGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLG VNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DEL RNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EWFENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTEGCLKKTSFYNLKLMDSVIKESQR LRPLLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGGFGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, Q54K, A143G, T307A, G325E, V364L		1187	1188

MATDLDLDRVLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSNKRVVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	L7R, S31V, Q54K, A143G, T307A, G325E		1189	1190
MATDLDLVLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSNKRVVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWQEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, Q54K, A143G, A153Q, T307A, G325E		1191	1192
MATDLDLVLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSNKRVVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWPEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, Q54K, A143G, A153P, T307A, G325E		1193	1194
MATDLDLVLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSNKRVVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, Q54K, A143G, T307A, G325E, Q424V		1195	1196

DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE				
MATDLDLVLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSINKRVVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLGVNHVLGTSTEWHPINPGDDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DELRNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, Q54K, A143G, E176D, T307A, G325E		1197	1198
MATDLDLVLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSINKRVVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DELRNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, Q54K, A143G, T307A, G325E, K479R		1199	1200
MATDLDLVLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSINKRVVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DELRNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, Q54K, K70L, A143G, T307A, G325E		1201	1202
MATDLDLVLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSINKRVVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DELRNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, Q54K, A143G, T307A, G325E, T333V		1203	1204

HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE				
MATDLDLVLGKSQYALFCGITLEFSFFILKYVLLGNG GKQYPYINPKKPFELSINKRVVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTQGLKKT SFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, Q54K, A143G, T307A, G325E, E338Q		1205	1206
MATDLDLVLGKSQYALFCGITLEFSFFILKYVLLGNG GKQYPYINPKKPFELSINKRVVQDFIENARDILTNGR SLYKDTPTYKAHTDLGDLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, Q54K, K70N, A143G, T307A, G325E,		1207	1208
MATDLDLVLGKSQYALFCGITLEFSFFILKYVLLGNG GKQYPYINPKKPFELSINKRVVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTRQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, Q54K, A143G, V210R, T307A, G325E,		1209	1210
MATDLDLVLGKSQYALFCGITLEFSFFILKYVLLGNG GKQYPYINPKKPFELSINKRVVQDFIENARDILTKGR SLYKDTPTYKAHTDLGDLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL	S31V, Q54K, A143G, T307A, G325E, F452D		1211	1212

EPLREEIVTVLSTEGCLKKTSFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTLPNGDVIKKGTKIICD'TT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFDAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE				
MATDLDLVLGKSQYALFCGITLEFSFFILKYVLLGNG GKQYPYINPKKPFELSNKR'VVQDFIENARDILTKGR SLYKDT'PYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEP'IGSPFDLVPLVNKYLTRGL AKLT'KPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRI'FMGEELCKDDDWLKVSI'EYTVQLFQTA DEL'RNYP'RWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAV'KKEAIAEGRPS'PFDDSI'EFENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALMPELL EPLREEIVTVLSTEGCLKKTSFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTLPNGDVIKKGTKIICD'TT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, Q54K, A143G, T307A, Q320M, G325E,		1213	1214
MATDLDLVLGKSQYALFCGITLEFSFFILKYVLLGNG GKQYPYINPKKPFELSNKR'VVQDFIENARDILTKGR SLYKDT'PYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEP'IGSPFDLVPLVNKYLTRGL AKLT'KPLWAEASLGVNHVLGTSTEWV'PINPGEDIMR IVSRMSSRI'FMGEELCKDDDWLKVSI'EYTVQLFQTA DEL'RNYP'RWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAV'KKEAIAEGRPS'PFDDSI'EFENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTEGCLKKTSFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTLPNGDVIKKGTKIICD'TT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, Q54K, A143G, H170V, T307A, G325E,		1215	1216
MATDLDLVLGKSQYALFCGITLEFSFFILKYVLLGNG GKQYPYINPKKPFELSNKR'VVQDFIENARDILTKGR SLYKDT'PYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEP'IGSPFDLVPLVNKYLTRGL AKLT'KPLWAEASLGVHHVLGTSTEWHPINPGEDIMR IVSRMSSRI'FMGEELCKDDDWLKVSI'EYTVQLFQTA DEL'RNYP'RWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAV'KKEAIAEGRPS'PFDDSI'EFENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTEGCLKKTSFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTLPNGDVIKKGTKIICD'TT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, Q54K, A143G, N160H, T307A, G325E,		1217	1218
MATDLDLVLGKSQYALFCGITLEFSFFILKYVLLGNG GKQYPYINPKKPFELSNKR'VVQDFIENARDILTKGR SLYKDT'PYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEP'IGSPFDLVPLVNKYLTRGL AKLT'KPLWAEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRI'FMGEELCKDDDWLKVSI'EYTVQLFQTA DEL'RNYP'RWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAV'KKEAIAEGRPS'PFDDSI'EFENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTEGCLKKTSFYNLKLMDSVIKESQR LRPVLLGAFRRMALADVTLPNGDVIKKGTKIICD'TT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, Q54K, A143G, T307A, G325E, S408R		1219	1220

PHIDRRNAVKKKEAIAEGRPSPFDDSIWFENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTEGLKKTsfynlklmDSVIKESQR LRPVLLGAFRRMALADVTLPNGDVIKKGTkIICDTT HQWNPEYYPDARKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE				
MATDLDRVLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSNKRVVQDFIENARDILTGR SLYKDTPTYKAHTDLGDLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLGVNHLVLTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSIETYRQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSIWFENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTEGLKKTsfynlklmDSVIKESQR LRPLLLGAFRRMALADVTLPNGDVIKKGTkIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	L7R, S31V, Q54K, A143G, V210R, T307A, G325E, V364L, K479R		1243	1244
MATDLDRVLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSNKRVVQDFIENARDILTGR SLYKDTPTYKAHTDLGDLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLGVHHVLGTSTEWVPINPGDDIMR IVSRMSSRIFMGEELCKDDDWLKVSIETYVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSIWFENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTEGLKKTsfynlklmDSVIKESQR LRPLLLGAFRRMALADVTLPNGDVIKKGTkIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	L7R, S31V, Q54K, K70L, A143G, N160H, H170V, E176D, T307A, G325E, V364L,		1245	1246
MATDLDLVLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSNKRVVQDFIENARDILTGR SLYKDTPTYKAHTDLGDLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLGVNHLVLTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSIETYVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSIWFENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTEGLKKTsfynlklmDSVIKESQR LRPLLLGAFRRMALADVTLPNGDVIKKGTkIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, Q54K, A143G, T307A, G325E, V364L, L444F		1247	1248
MATDLDLVLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSNKRVVQDFIENARDILTGR SLYKDTPTYKAHTDLGDLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLGVNHLVLTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSIETYVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSIWFENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTEGLKKTsfynlklmDSVIKESQR LRPLLLGAFRRMALADVTLPNGDVIKKGTkIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGFHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, Q54K, A143G, T307A, G325E, V364L, Q320M, E402V		1249	1250

IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DEL RNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EWFENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALMPELL EPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPLLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPVYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE				
MATDLDLVLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSNKR VVQDFIENARDILTGR SLYKDPYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLG VNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DEL RNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EWFENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPLLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, Q54K, K70L, A143G, T307A, G325E, V364L,		1251	1252
MATDLDLVLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSNKR VVQDFIENARDILTGR SLYKDPYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLG VNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DEL RNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAEKKEAIAEGRPSPFDDSI EWFENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALMPELL EPLREEIVVVLSTEGLKKT SFYNLKLMDSVIKESQR LRPLLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, Q54K, A143G, V261E, T307A, Q320M, G325E, T333V, V364L, K479R		1253	1254
MATDLDLVLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSNKR VVQDFIENARDILTGR SLYKDPYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWQEASLG VNHVLGTSTEWV PINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DEL RNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EWFENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALMPELL EPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPLLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, Q54K, A143G, A153Q, H170V, T307A, G325E, V364L, Q424V		1255	1256
MATDLDLVLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSNKR VVQDFIENARDILTGR SLYKDPYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWQEASLG VNHVLGTSTEWV PINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DEL RNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EWFENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPLLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPYYPDASKFNAYRFLQMRQTPGV DKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, Q54K, K70L, A143G, T307A, G325E,		1257	1258

TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLGVNHLVLTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DEL RNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTQGLKKT SFYNLKLMDSVIKESQR LRPLLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGTSLLGDREAKLMVKRRAAEI DIDTIGSDE	V364L, E338Q, M485T			
MATDLDLVLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSNKR VVQDFIENARDILT KGR SLYKDTPYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLGVNHLVLTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DEL RNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFSIALQPELL EPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPLLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, Q54K, A143G, T307A, N316S G325E, V364L,		1259	1260
MATDLDRLVLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSNKR VVQDFIENARDILT KGR SLYKDTPYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLGVNHLVLTSTEWHPINPGDDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DEL RNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPLLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	L7R, K479R S31V, Q54K, A143G, E176D, T307A, G325E, V364L		1261	1262
MATDLDRLVLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSNKR VVQDFIENARDILT KGR SLYKDTPYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLGVNHLVLTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSI EYTVQLFQTA DEL RNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSI EW FENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTEGLKKT SFYNLKLMDSVIKESQR LRPLLLGAFRRMALADVTL PNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	L7R, S31V, Q54K, A143G, T307A, Q320M, G325E, V364L, K479R		1263	1264

MATDLDRLVGLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSNKRVVQDFIENARDILTGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLGVNHVLGTSTEWVPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSIEYTVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSIWFENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTEGLKKTsfynlklmDSVIKESQR LRPLLLGAFRRMALADVTLPNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPRSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	L7R, S31V, Q54K, K70L, A143G, H170V, T307A, G325E, V364L, K479R		1265	1266
MATDLDRLVGLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSNKRVVQDFIENARDILTGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLGVNHVLGTSTEWHPINPGDDIMR IVSRMSSRIFMGEELCKDDDWLKVSIEYTVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSIWFENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALMPELL EPLREEIVTVLSTEGLKKTsfynlklmDSVIKESQR LRPLLLGAFRRMALADVTLPNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	L7R, S31V, Q54K, A143G, E176D, T307A, Q320M, G325E, V364L,		1267	1268
MATDLDLVLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSNKRVVQDFIENARDILTGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWAEASLGVHHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSIEYTVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSIWFENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALMPELL EPLREEIVTVLSTEGLKKTsfynlklmDSVIKESQR LRPLLLGAFRRMALADVTLPNGDVIKKGTKIICDTT HQWNPEYYPDASKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY DWKLPEGVVPSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE	S31V, Q54K, A143G, N160H, T307A, Q320M, G325E, V364L,		1269	1270
MATDLDRLVGLGKSQYALFCGITLFSFFILKYVLLGNG GKQYPYINPKKPFELSNKRVVQDFIENARDILTGR SLYKDTPTYKAHTDLGDVLVIPPEFADALKSERQLDF TEVARDDTHGYIPGFEPIGSPFDLVPLVNKYLTRGL AKLTKPLWPEASLGVNHVLGTSTEWHPINPGEDIMR IVSRMSSRIFMGEELCKDDDWLKVSIEYTVQLFQTA DELARNYPRWTRPYIHWFLPSCQGVRRKLQEARDLLQ PHIDRRNAVKKKEAIAEGRPSPFDDSIWFENEYEGK SDPATEQIKLSLVAIHTTADLLSETMFNIALQPELL EPLREEIVTVLSTEGLKKTsfynlklmDSVIKESQR LRPLLLGAFRRMALADVTLPNGDVIKKGTKIICDTT HQWNPEYYPDARKFNAYRFLQMRQTPGQDKRAHLVS TSHDQMGFGHGLHACPGRFFAANEIKIALCHMLLKY	L7R, S31V, Q54K, A143G, A153P, T307A, G325E, V364L, S408R		1271	1272

DWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEI DIDTIGSDE				
MLGNGGKQYPYINPKKPFELSNKRVVQDFIENARDI LTKGRSLYKDTPTYKAHTDLGDVLVIPPEFADALKSE RQLDFTEVARDDTHGYIPGFEPIGSPFDLVPLVKNKY LTRGLAKLTKPLWAEASLGVNHVLGTSTEWHPINPG EDIMRIVSRMSSRIFMGEELCKDDDWLKVSI EYTVQ LFQTADELARNYPRWTRPYIHWFLPSCQGVRRKLQEA RDLLOPHIDRRNAVKKKEAIAEGRPSPFDDSI EWFE EYEGKSDPATEQIKLSLVAIH TTADLLSETMFNIAL QPELLEPLREEIVTVLSTEGLKKTSFYNLKLMDSVI KESQRLRPLLLGAFRRMALADVTL PNGDVIKKGTKI ICDTTHQWNPEYYPDASKFNAYRFLQMRQTPGQDKR AHLVSTSHDQMGFGHGLHACPRFFAANEIKIALCH MLLKYDWKLPEGVVPKSKALGMSLLGDREAKLMVKR RAAEIDIDTIGSDE	31-aa N-term truncation + Q54K, A143G, T307A, G325E, V364L		1275	1276
¹ Amino acid difference(s) relative to the wild-type CYP sequence of SEQ ID NO: 2 denoted by standard format of single letter amino acid and position number followed by substituted amino acid in single-letter – e.g., “S31V”. ² Silent codon change(s) relative to the codon-optimized nucleotide sequence of SEQ ID NO: 1 as denoted by the amino acid position in SEQ ID NO: 2 and the specific codon change indicated by parentheses.				

[0089] In at least one embodiment, the recombinant polypeptides having CYP activity have one or more residue differences as compared to the reference *G. zeae* CYP polypeptide of SEQ ID NO: 2. In some embodiments, the recombinant polypeptides have one or more amino acid residue differences as compared to SEQ ID NO: 2 at an amino acid position selected from L7, S31, Q54, V56, K70, K76, I120, A143, L147, A153, N160, H170, G175, E176, V210, V261, T307, N316, Q320, G325, T333, E338, V364, L365, P401, E402, S408, Q424, L444, F452, K479, and M485. In some embodiments, the recombinant polypeptides have one or more amino acid residue differences as compared to SEQ ID NO: 2 selected from L7R, S31V, Q54K, V56G, K70L, K70N, K76H, I120A, A143G, A143R, L147R, L147W, A153Q, A153R, N160H, H170V, G175E, E176D, V210R, V261E, T307A, N316S, Q320M, G325E, T333V, E338Q, V364L, L365A, P401D, E402V, S408R, Q424V, L444F, F452D, K479R, and M485T.

[0090] It is contemplated that the residue differences relative to SEQ ID NO: 2 at residue positions associated with increased CYP activity can be used in various combinations to form recombinant CYP polypeptides having desirable functional characteristics when integrated in a recombinant host cell, for example increased yield of a product compound, such as UDCA. Some exemplary combinations of amino acid differences include those combinations found in the exemplary polypeptides of Table 3 and elsewhere herein. For example, the present disclosure provides a recombinant polypeptide having increased CYP activity, a set of amino acid residue differences as compared to SEQ ID NO: 2 selected from:

S31V, Q54K, A143G, T307A, G325E
S31V, Q54K, A143G, T307A, G325E, L365A
S31V, L147R, G325E

S31V, Q54K, A143R, L147R, T307A, G325E, P401D
S31V, V57F, L147W, P401D
S31V, L147W, P401D
S31V, Q54K, A143R, T307A, P401D
S31V, Q54K, P401D
S31V, A143R, L147W
S31V, A143G, G175E, P401D, K498R
S31V, A143G, L147C, G175E, P401D
S31V, P129S, A143R
S31V, G175E, P401D
S31V, A143G
S31V, A143R
S31V, Q54K, A143G, T307A, G325E, V364L
L7R, S31V, Q54K, A143G, T307A, G325E
S31V, Q54K, A143G, A153Q, T307A, G325E
S31V, Q54K, A143G, A153P, T307A, G325E
S31V, Q54K, A143G, T307A, G325E, Q424V
S31V, Q54K, A143G, E176D, T307A, G325E
S31V, Q54K, A143G, T307A, G325E, K479R
S31V, Q54K, K70L, A143G, T307A, G325E
S31V, Q54K, A143G, T307A, G325E, T333V
S31V, Q54K, A143G, T307A, G325E, E338Q
S31V, Q54K, K70N, A143G, T307A, G325E,
S31V, Q54K, A143G, V210R, T307A, G325E,
S31V, Q54K, A143G, T307A, G325E, F452D
S31V, Q54K, A143G, T307A, Q320M, G325E,
S31V, Q54K, A143G, H170V, T307A, G325E,
S31V, Q54K, A143G, N160H, T307A, G325E,
S31V, Q54K, A143G, T307A, G325E, S408R
L7R, S31V, Q54K, A143G, V210R, T307A, G325E, V364L, K479R
L7R, S31V, Q54K, K70L, A143G, N160H, H170V, E176D, T307A, G325E, V364L,
S31V, Q54K, A143G, T307A, G325E, V364L, L444F
S31V, Q54K, A143G, T307A, G325E, V364L, Q320M, E402V
S31V, Q54K, K70L, A143G, T307A, G325E, V364L,
S31V, Q54K, A143G, V261E, T307A, Q320M, G325E, T333V, V364L, K479R
S31V, Q54K, A143G, A153Q, H170V, T307A, G325E, V364L, Q424V
S31V, Q54K, K70L, A143G, T307A, G325E, V364L, E338Q, M485T
S31V, Q54K, A143G, T307A, N316S G325E, V364L,
L7R, K479R S31V, Q54K, A143G, E176D, T307A, G325E, V364L
L7R, S31V, Q54K, A143G, T307A, Q320M, G325E, V364L, K479R
L7R, S31V, Q54K, K70L, A143G, H170V, T307A, G325E, V364L, K479R
L7R, S31V, Q54K, A143G, E176D, T307A, Q320M, G325E, V364L,
S31V, Q54K, A143G, N160H, T307A, Q320M, G325E, V364L,
L7R, S31V, Q54K, A143G, A153P, T307A, G325E, V364L, S408R

[0091] Based on the correlation of recombinant polypeptide functional information provided herein with the sequence information provided in Table 3, the accompanying Sequence Listing, and/or the Examples disclosed herein, one of ordinary skill can recognize that the present disclosure provides a range of recombinant polypeptides having CYP activity, wherein the polypeptide comprises an amino acid sequence comprising one or more of the amino acid differences or sets of amino acid differences (relative to SEQ ID NO: 2) disclosed in any one of SEQ ID NO: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48,

50, 52, 54, 56, 58, 1186, 1188, 1190, 1192, 1194, 1196, 1198, 1200, 1202, 1204, 1206, 1208, 1210, 1212, 1214, 1216, 1218, 1220, 1244, 1246, 1248, 1250, 1252, 1254, 1256, 1258, 1260, 1262, 1264, 1266, 1268, 1270, 1272, and 1276, and otherwise have at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity to a sequence selected from the group consisting of SEQ ID NO: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 1186, 1188, 1190, 1192, 1194, 1196, 1198, 1200, 1202, 1204, 1206, 1208, 1210, 1212, 1214, 1216, 1218, 1220, 1244, 1246, 1248, 1250, 1252, 1254, 1256, 1258, 1260, 1262, 1264, 1266, 1268, 1270, 1272, and 1276.

[0092] Additionally, in at least one embodiment, a recombinant polypeptide of the present disclosure having CYP activity can have an amino acid sequence comprising one or more of the amino acid differences or sets of amino acid differences (relative to SEQ ID NO: 2) disclosed in any one of SEQ ID NO: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 1186, 1188, 1190, 1192, 1194, 1196, 1198, 1200, 1202, 1204, 1206, 1208, 1210, 1212, 1214, 1216, 1218, 1220, 1244, 1246, 1248, 1250, 1252, 1254, 1256, 1258, 1260, 1262, 1264, 1266, 1268, 1270, 1272, and 1276, and additionally have 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-14, 1-15, 1-16, 1-18, 1-20, 1-22, 1-24, 1-26, 1-30, 1-35, 1-40, 1-45, 1-50, 1-55, or 1-60 residue differences at other residue positions. In some embodiments, the number of differences can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 18, 20, 22, 24, 26, 30, 35, 40, 45, 50, 55, or 60 residue differences at the other residue positions.

[0093] In addition to the residue positions specified above, any of the engineered CYP polypeptides disclosed herein can further comprise other residue differences relative to the reference polypeptide of SEQ ID NO: 2 at other residue positions.

[0094] Residue differences at these other residue positions can provide for additional variations in the amino acid sequence without adversely affecting the ability of the recombinant polypeptide to carry out the desired biocatalytic conversion (e.g., conversion of LCA (compound **2a**) to UDCA (compound **1a**)). In some embodiments, the recombinant polypeptides can have additionally 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-14, 1-15, 1-16, 1-18, 1-20, 1-22, 1-24, 1-26, 1-30, 1-35, 1-40 residue differences at other amino acid residue positions as compared to SEQ ID NO: 2. In some embodiments, the number of differences can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 18, 20, 22, 24, 26, 30, 35, and 40 residue differences at other residue positions. The residue difference at these other positions can include conservative changes or non-conservative changes. In some embodiments, the residue differences can comprise conservative substitutions and non-conservative substitutions as compared to the reference polypeptide of SEQ ID NO: 2.

[0095] Other engineered modifications of the CYP polypeptides contemplated by the present disclosure include modification of the amino acid sequence at either its N- or C- terminus by truncation or fusion. For example, in at least one embodiment, the CYP polypeptides (or CPR polypeptides) of the present disclosure can be further engineered by truncation at the N- and/or C-terminus of the polypeptide chain to provide a truncated version of the polypeptide that retains CYP activity. Methods for preparing and screening truncated versions of CYP polypeptides are known in the art and described in the Examples and elsewhere herein. The present disclosure contemplates that the genes encoding these truncated CYP and CPR enzymes can be used in the recombinant host cell compositions and accompanying biosynthetic methods of the present disclosure. In one embodiment, the engineered CYP polypeptide CYP207 (SEQ ID NO: 1186) of the present disclosure has been prepared in a form that is truncated at the N-terminus by 31 amino acids. This truncated version of CYP207, referred to as trCYP207 (SEQ ID NO: 1276) retains CYP activity a gene encoding it is heterologously expressed in a recombinant *Pichia* host cell. Accordingly, in at least one embodiment any of the engineered CYP polypeptides disclosed herein can include an engineered version of the polypeptide (and the encoding gene), wherein from 1-31 N-terminal amino acids are truncated.

[0096] In at least one embodiment, such a recombinant polypeptide with an N-terminal truncation and CYP activity capable of converting LCA to UDCA and/or converting 3-KCA to 3-KUDCA comprises an N-terminal truncation of 1 to 31 amino acids as compared to SEQ ID NO: 2. It is further contemplated that the polypeptide with CYP activity can comprise an N-terminal truncation of from 1 to 31 amino acids of engineered recombinant polypeptide comprising a sequence of any of SEQ ID NO: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 1186, 1188, 1190, 1192, 1194, 1196, 1198, 1200, 1202, 1204, 1206, 1208, 1210, 1212, 1214, 1216, 1218, 1220, 1244, 1246, 1248, 1250, 1252, 1254, 1256, 1258, 1260, 1262, 1264, 1266, 1268, 1270, and 1272. In at least one embodiment, the N-terminal truncated polypeptide with CYP activity comprises an amino acid of SEQ ID NO: 1276.

[0097] It also is contemplated that the engineered CYP (and CPR) polypeptides of the present disclosure can be modified via fusion to other polypeptides. In at least one embodiment of a fusion, an engineered recombinant gene encoding a polypeptide having CYP activity of the present disclosure is fused to a gene encoding a recombinant polypeptide having CPR activity, whereby the fusion gene when expressed heterologously in a recombinant host cell expression a fusion polypeptide having both CYP activity and CPR activity. As shown in the Examples and elsewhere herein, such engineered CYP-CPR fusion polypeptides can be used in the biosynthetic processes for preparing UDCA.

[0098] In at least one embodiment, the recombinant polypeptide with CYP activity is fused via a polypeptide linker to a second polypeptide with CPR activity. In at least one embodiment of the

fusion, the polypeptide with CYP activity used in the fusion is an 1-31 amino acid N-terminal truncated version of an engineered polypeptide comprising a sequence of at least 80%, 90%, 95%, or 99% identity to any of SEQ ID NO: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 1186, 1188, 1190, 1192, 1194, 1196, 1198, 1200, 1202, 1204, 1206, 1208, 1210, 1212, 1214, 1216, 1218, 1220, 1244, 1246, 1248, 1250, 1252, 1254, 1256, 1258, 1260, 1262, 1264, 1266, 1268, 1270, and 1272. In at least one embodiment of the fusion, the second polypeptide with CPR activity comprises an amino acid sequence having at least 80%, 90%, 95%, or 99% sequence identity to a sequence selected from SEQ ID NO: 4 and 1278.

[0099] In at least one embodiment the recombinant fusion polypeptide with CYP activity and CPR activity is capable of converting LCA to UDCA and/or converting 3-KCA to 3-KUDCA is fused and comprises an amino acid sequence of at least 80%, 90%, 95%, or 99% sequence identity to SEQ ID NO: 1280.

[0100] It is further contemplated that the engineered polypeptides of the present disclosure can be fused to polypeptides such as antibody tags (e.g., myc epitope), purification sequences (e.g., His tags for binding to metals), and cell localization signals (e.g., secretion signals). It is also contemplated that the recombinant polypeptides described herein are not restricted to the genetically encoded amino acids. In addition to the genetically encoded amino acids, the polypeptides described herein may be comprised, either in whole or in part, of naturally occurring and/or synthetic non-encoded amino acids.

[0101] In another aspect, the present disclosure provides polynucleotides encoding the recombinant polypeptides having CYP activity and increased activity and/or yield as described herein. In at least one embodiment, the polynucleotide encoding a recombinant polypeptide having CYP activity comprises an amino acid sequence that is at least about 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more identical to the polypeptide sequence of SEQ ID NO: 2. In some embodiments, the polynucleotide encodes a recombinant polypeptide comprising an amino acid sequence that has the percent identity described above and has one or more amino acid residue differences as compared to SEQ ID NO: 2 described elsewhere herein.

[0102] In at least one embodiment, the polynucleotide has a sequence encoding a recombinant polypeptide which polynucleotide sequence has one or more neutral codon differences relative to SEQ ID NO: 1, which codon differences do not encode an amino acid difference but result in increased yield of the desired product compound (e.g., UDCA) produced by a recombinant host cell in which the polynucleotide sequence is integrated. In at least one embodiment, the polypeptide is encoded by a polynucleotide sequence having at least 80% identity to SEQ ID NO: 1, and at least one neutral codon difference as compared to SEQ ID NO: 1 at a position encoding an amino acid residue selected from R328, A368, and V476; optionally, wherein the

neutral codon difference is selected from: R328 (CGT > CGC), A368 (GCT > GCC), and V476 (GTT > GTC).

[0103] It is also contemplated that the polynucleotides encoding the recombinant polypeptides having CYP activity and increased activity and/or yield as described herein, can include a combination of one or more codon differences relative to SEQ ID NO: 1, wherein at least one of the codon differences encodes an amino acid difference as compared to SEQ ID NO: 2 and at least one codon difference is a neutral codon difference that does not encode an amino acid difference as compared to SEQ ID NO: 2. Accordingly, in at least one embodiment, the present disclosure provides a polynucleotide sequence encoding a recombinant polypeptide having CYP activity, wherein the polynucleotide sequence comprises a combination of a codon differences encoding an amino acid difference and a neutral codon difference selected from: R328 (CGT > CGC), A368 (GCT > GCC), and V476 (GTT > GTC).

[0104] In at least one embodiment, the polynucleotide comprises a sequence encoding an exemplary recombinant polypeptide having CYP activity as disclosed in Table 3 and the accompanying Sequence Listing. In at least one embodiment, the polynucleotide comprises a sequence of at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, or at least 99% identity to a sequence selected from the group consisting of SEQ ID NO: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 1185, 1187, 1189, 1191, 1193, 1195, 1197, 1199, 1201, 1203, 1205, 1207, 1209, 1211, 1213, 1215, 1217, 1219, 1243, 1245, 1247, 1249, 1251, 1253, 1255, 1257, 1259, 1261, 1263, 1265, 1267, 1269, 1271, and 1275. In at least one embodiment, the polynucleotide comprises a codon degenerate sequence of a sequence selected from the group consisting of SEQ ID NO: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 1185, 1187, 1189, 1191, 1193, 1195, 1197, 1199, 1201, 1203, 1205, 1207, 1209, 1211, 1213, 1215, 1217, 1219, 1243, 1245, 1247, 1249, 1251, 1253, 1255, 1257, 1259, 1261, 1263, 1265, 1267, 1269, 1271, and 1275.

[0105] The polynucleotide sequences encoding the recombinant polypeptides of the present disclosure may be operatively linked to one or more heterologous regulatory sequences that control gene expression to create a recombinant polynucleotide capable of expressing the polypeptide. Expression constructs containing a heterologous polynucleotide encoding the recombinant polypeptide can be introduced into appropriate host cells to express the corresponding polypeptide. Because of the knowledge of the codons corresponding to the various amino acids, availability of a protein sequence provides a description of all the polynucleotides capable of encoding the subject. The degeneracy of the genetic code, where the same amino acids are encoded by alternative or synonymous codons allows an extremely large number of nucleic acids to be made, all of which encode the improved transaminase enzymes disclosed herein. Thus, having identified a particular amino acid sequence, those skilled in the art could make any number of different nucleic acids by simply modifying the

sequence of one or more codons in a way which does not change the amino acid sequence of the protein. In this regard, the present disclosure specifically contemplates each and every possible variation of polynucleotides that could be made by selecting combinations based on the possible codon choices, and all such variations are to be considered specifically disclosed for any polypeptide disclosed herein, including the amino acid sequences presented in Table 3 and the accompanying Sequence Listing.

[0106] The codons can be selected to fit the host cell in which the protein is being produced. For example, preferred codons used in bacteria are used to express the gene in bacteria; preferred codons used in yeast are used for expression in yeast; and preferred codons used in mammals are used for expression in mammalian cells. It is contemplated that all codons need not be replaced to optimize the codon usage of the recombinant polypeptide since the natural sequence will comprise preferred codons and because use of preferred codons may not be required for all amino acid residues. Consequently, codon optimized polynucleotides encoding the recombinant polypeptide may contain preferred codons at about 40%, 50%, 60%, 70%, 80%, or greater than 90% of codon positions of the full-length coding region.

[0107] The present disclosure also provides an expression vector comprising a polynucleotide encoding a recombinant polypeptide having CYP activity and increased thermostability, and one or more expression regulating regions such as a promoter, a terminator, a replication origin, or the like, depending on the type of hosts into which they are to be introduced. The various nucleic acid and control sequences described above may be joined together to produce a recombinant expression vector which may include one or more convenient restriction sites to allow for insertion or substitution of the nucleic acid sequence encoding the recombinant polypeptide at such sites. Alternatively, a polynucleotide sequence of the present disclosure may be expressed by inserting the nucleic acid sequence or a nucleic acid construct comprising the sequence into an appropriate vector for expression. In creating the expression vector, the coding sequence is located in the vector so that the coding sequence is operably linked with the appropriate control sequences for expression. The recombinant expression vector may be any vector (e.g., a plasmid or virus), which can be conveniently subjected to recombinant DNA procedures and can bring about the expression of the polynucleotide sequence. The choice of the vector will typically depend on the compatibility of the vector with the host cell into which the vector is to be introduced. The vectors may be linear or closed circular plasmids.

[0108] The expression vector may be an autonomously replicating vector, i.e., a vector that exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, e.g., a plasmid, an extrachromosomal element, a mini-chromosome, or an artificial chromosome. The vector may contain any means for assuring self-replication. Alternatively, the vector may be one which, when introduced into the host cell, is integrated into the genome, and replicated together with the chromosome(s) into which it has been integrated.

Furthermore, a single vector or plasmid or two or more vectors or plasmids which together

contain the total DNA to be introduced into the genome of the host cell, or a transposon may be used. In at least one embodiment, the expression vector further comprises one or more selectable markers, which permit easy selection of transformed cells.

[0109] Use in Recombinant Host Cells

[0110] The engineered genes of the present disclosure that encode recombinant polypeptides with CYP activity (e.g., exemplary polypeptides of Table 3) can be incorporated in recombinant host cells to enable *in vivo* biosynthesis of compounds cholic acid compounds, such as UDCA and UDCA derivative compounds, that require a cytochrome P450 monooxygenase catalyzed reaction. These recombinant host cells comprise a polynucleotide or expression vector that encodes the recombinant polypeptide with CYP activity, wherein the polynucleotide is operatively linked to one or more control sequences for expression of the polypeptide in the host cell. Host cells for use in expressing recombinant genes encoding the polypeptides with CYP activity of the present disclosure are well known in the art and include but are not limited to, bacterial cells, such as *E. coli*, or fungal cells, such as *Saccharomyces cerevisiae* or *Pichia pastoris*, insect cells, such as Drosophila S2 and Spodoptera Sf9, animal cells, such as CHO, COS, BHK, 293, and plant cells. Appropriate mediums and growth conditions for culturing the recombinant host cells so that they express the polypeptide with CYP activity are well known in the art.

[0111] The recombinant host cells can comprise heterologous nucleic acids encoding not only polypeptides with CYP and CPR activity capable of converting LCA to UDCA (or 3-KCA to 3-KUDCA) but also other enzymes capable of producing other precursor compounds, such as precursors for the substrates, LCA or 3-KCA. As described elsewhere herein, nucleic acid sequences encoding pathway enzymes for producing such precursor compounds are known in the art and can readily be used in accordance with the present disclosure. Typically, the nucleic acid sequence encoding the enzymes which form a part of the pathway, further include one or more additional nucleic acid sequences, for example, a nucleic acid sequence controlling expression of the enzymes which form a part of the biosynthetic pathway, and these one or more additional nucleic acid sequences together with the nucleic acid sequence encoding the polypeptides with CPR and/or CYP activity can be considered a heterologous nucleic acid sequence. A variety of techniques and methodologies are available and well known in the art for introducing heterologous nucleic acid sequences, such as nucleic acid sequences encoding the enzymes (e.g., CPR and CYP), into a host cell so as to attain expression the host cell. Such techniques are well known to the skilled artisan and can be found in, for example, Sambrook et al., Molecular Cloning, a Laboratory Manual, Cold Spring Harbor Laboratory Press, 2012, Fourth Ed.

[0112] For example, the introduction of the heterologous nucleic acids can include integration of the nucleic acids into specific loci in the genome of a host cell via CRISPR-Cas9 and other techniques, some of which are demonstrated in the Examples herein. Such techniques are well

known to the skilled artisan and can, for example, be found in Sambrook and other well-known sources. The number of copies of heterologous genes and their locus of integration in a recombinant host cell's genome can result in improved biosynthetic production of a desired product, such as UDCA. For example, a heterologous nucleic acid encoding a polypeptide with CYP activity can be integrated in a host cell's genome in 1, 2, 3, 4, or more copies.

Accordingly, it is contemplated that in the recombinant host cells of the present disclosure, the heterologous nucleic acid encoding the recombinant polypeptide having CYP activity can be integrated in the host cell's genome at one or more loci, including but not limited to the well-known genomic loci in *Saccharomyces cerevisiae* of X-2, X-4, XI-2, XII-4, NDE1, XII-5, Gal80, and ROQ1, or the loci in *Pichia pastoris* of AOX1, Int6, Int15, and HIS4.

[0113] One of ordinary skill will recognize that the heterologous nucleic acids encoding the recombinant enzymes with CPR activity and CYP activity, and any other pathway enzymes will further comprise transcriptional promoters capable of controlling expression of the enzymes in the recombinant host cell. Generally, the transcriptional promoters are selected to be compatible with the host cell, so that promoters obtained from bacterial cells are used when a bacterial host cell is selected in accordance herewith, while a fungal promoter is used when a fungal host cell is selected, a plant promoter is used when a plant cell is selected, and so on. Promoters useful in the recombinant host cells of the present disclosure may be constitutive or inducible, provided such promoters are operable in the host cells. Promoters that may be used to control expression in fungal host cells, such as *Saccharomyces cerevisiae* and *Pichia pastoris*, are well known in the art and include, but are not limited to inducible promoters, such as a Gal1 promoter or Gal10 promoter, a constitutive promoter, such as an alcohol dehydrogenase (ADH) promoter, a glyceraldehyde-3-phosphate dehydrogenase (GPD) promoter, or an *S. pombe* Nmt, or ADH promoter. Exemplary promoters that may be used to control expression in bacterial cells can include the *Escherichia coli* promoters *lac*, *tac*, *trc*, *trp* or the *T7* promoter. Exemplary promoters that may be used to control expression in plant cells include, for example, a Cauliflower Mosaic Virus 35S promoter (Odell *et al.* (1985) Nature 313:810-812), a ubiquitin promoter (U.S. Pat. No. 5,510,474; Christensen *et al.* (1989)), or a rice actin promoter (McElroy *et al.* (1990) Plant Cell 2:163-171). Exemplary promoters that can be used in mammalian cells include, a viral promoter such as an SV40 promoter or a metallothionine promoter. All of these host cell promoters are well known by and readily available to one of ordinary skill in the art. Further nucleic acid control elements useful for controlling expression in a recombinant host cell can include transcriptional terminators, enhancers, and the like, all of which may be used with the heterologous nucleic acids incorporate in the recombinant host cells of the present disclosure.

[0114] A wide variety of techniques are well known in the art for linking transcriptional promoters and other control elements to heterologous nucleic acid sequences encoding pathway genes for biosynthesis of UDCA or 3-KUDCA. Such techniques are described in e.g.,

Sambrook *et al.*, Molecular Cloning, a Laboratory Manual, Cold Spring Harbor Laboratory Press, 2012, Fourth Ed. Accordingly, in at least one embodiment, the heterologous nucleic acid sequences of the present disclosure comprise a promoter capable of controlling expression in a host cell, wherein the promoter is linked to a nucleic acid sequence encoding a recombinant polypeptide having CYP activity of the present disclosure, and as necessary, other enzymes constituting a pathway for production of a UDCA precursor, UDCA, and/or UDCA derivative. This heterologous nucleic acid sequence can be integrated into a recombinant expression vector which ensures good expression in the desired host cell, wherein the expression vector is suitable for expression in a host cell, meaning that the recombinant expression vector comprises the heterologous nucleic acid sequence linked to any genetic elements required to achieve expression in the host cell. Genetic elements that may be included in the expression vector in this regard include a transcriptional termination region, one or more nucleic acid sequences encoding marker genes, one or more origins of replication, and the like. In some embodiments, the expression vector further comprises genetic elements required for the integration of the vector or a portion thereof in the host cell's genome.

[0115] It is also contemplated that in some embodiments an expression vector comprising a heterologous nucleic acid of the present disclosure may further contain a marker gene. Marker genes useful in accordance with the present disclosure include any genes that allow the distinction of transformed cells from non-transformed cells, including all selectable and screenable marker genes. A marker gene may be a resistance marker such as an antibiotic resistance marker against, for example, kanamycin or ampicillin. Screenable markers that may be employed to identify transformants through visual inspection include β -glucuronidase (GUS) (U.S. Pat. Nos. 5,268,463 and 5,599,670) and green fluorescent protein (GFP) (Niedz *et al.*, 1995, Plant Cell Rep., 14: 403).

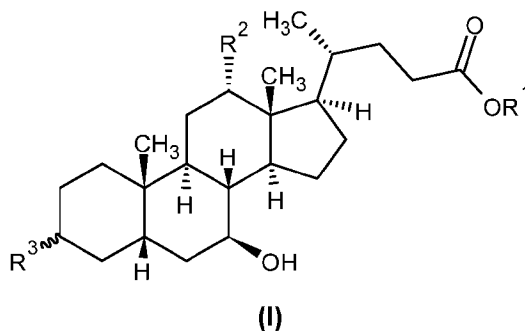
[0116] In at least one embodiment, the present disclosure also provides of a method for producing UDCA or 3-KUDCA, wherein a heterologous nucleic acid encoding a recombinant polypeptide having CYP activity (e.g., an exemplary engineered polypeptide of Table 3) can be introduced into a recombinant host cell. The recombinant host cell can then be used for production of the polypeptide or incorporated in a biocatalytic process that utilized the CYP activity of the recombinant polypeptide expressed by the host cell for the catalytic conversion of a substrate, e.g., the conversion of LCA to UDCA. In at one embodiment, the recombinant host cell can further comprise a pathway of enzymes capable of producing a compound precursor (e.g., LCA) which can act as a substrate for the recombinant polypeptides with CPR and CYP activity. It is contemplated that a recombinant host cell comprising a heterologous nucleic acid encoding a recombinant polypeptide having CYP activity of the present disclosure can provide improved biosynthesis of a desired product compound (e.g., UDCA or UDCA derivative) in terms of titer, yield, and production rate, due to the improved characteristics of the expressed

CYP activity in the cell associated with the amino acid and codon differences engineered in the gene.

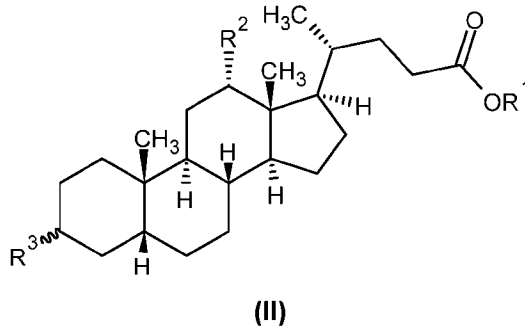
[0117] Accordingly, in at least one embodiment, the present disclosure provides a method for producing UDCA and/or 3-KUDCA comprising: (a) culturing a recombinant host cell of the present disclosure in a suitable medium comprising LCA and/or 3-KCA; and (b) recovering the produced UDCA and/or 3-KUDCA.

[0118] In at least one embodiment, it is contemplated the recombinant polypeptides with CYP activity of the present disclosure can be incorporated in any biosynthesis method requiring a CYP catalyzed biocatalytic step, whether *in vivo* or *in vitro*. Thus, in at least one embodiment, the recombinant polypeptides having CYP activity (e.g., exemplary polypeptides of Table 3) can be used in a method for preparing a compound of structural formula (I)

[0119] In another embodiment, the present disclosure provides a method for preparing a compound of structural formula (I)



wherein, R^1 is hydrogen, a carboxylate salt counterion, an optionally substituted C1-C20 alkyl, or an optionally substituted aryl; R^2 is hydrogen, or a hydroxyl; and R^3 is a hydroxyl, or an oxo group. The method comprises contacting a recombinant polypeptide having CYP activity of the present disclosure (e.g., an exemplary polypeptide of Table 3) under suitable reactions conditions with a compound of structural formula (II)



wherein, R^1 is hydrogen, a carboxylate salt counterion, an optionally substituted C1-C20 alkyl, or an optionally substituted aryl; R^2 is hydrogen, or a hydroxyl; and R^3 is a hydroxyl, or an oxo group.

[0120] Exemplary conversions of UDCA precursor compounds of structural formula (II) to UDCA and related compounds of structural formula (I) that are catalyzed by the recombinant polypeptides having CYP activity of the present disclosure can include: (1) conversion of LCA to UDCA; (2) the conversion of 3-KCA to 3-KUDCA; (3) the conversion of DCA to UCA. Accordingly, in at least one embodiment of the biosynthesis method for conversion a UDCA precursor compound of structural formula (II) to a UDCA compound of structural formula (I), R^2 is hydrogen, the compound of structural formula (II) is LCA, and the compound of structure formula (I) is UDCA. In at least one embodiment, R^2 is hydrogen, the compound of structural formula (II) is 3-KCA and the compound of structure formula (I) is 3-KUDCA. In at least one embodiment, R^2 is a hydroxyl, the compound of structure formula (II) is DCA, and the compound of structural formula (II) is DCA.

[0121] The present disclosure also contemplates that the methods for biocatalytic conversion of a UDCA precursor compound of structural formula (II) to a UDCA compound of structural formula (I) using an recombinant polypeptide having CYP activity of the present disclosure can comprise additional chemical or biocatalytic steps carried out on the product compound of structural formula (II), including steps of product compound work-up, extraction, isolation, purification, and/or crystallization, each of which can be carried out under a range of conditions.

[0122] Suitable reaction conditions for the biosynthesis of compounds such as UDCA, 3-KUDCA, and UCA, are known in the art and can be used with the recombinant polypeptides having CYP activity of the present disclosure. Additionally, suitable reaction conditions for the exemplary polypeptides of the present disclosure can be determined using routine techniques known in the art for optimizing biocatalytic reactions. It is contemplated that various ranges of suitable reaction conditions with the recombinant polypeptides of the present disclosure, including but not limited to ranges of pH, temperature, buffer, solvent system, substrate loading, polypeptide loading, co-substrate or co-factor loading, atmosphere, and reaction time. Suitable reaction conditions can be readily determined and optimized for particular reactions by routine experimentation that includes, but is not limited to, contacting the recombinant polypeptide and substrate under experimental reaction conditions of concentration, pH, temperature, solvent conditions, and detecting the production of the desired compound of structural formula (I). In at least one embodiment, the suitable reaction conditions comprise a reaction solution of ~pH 7-8, a temperature of 25C to 37C; optionally, the reaction conditions comprise a reaction solution of ~ pH 7 and a temperature of ~30C. In at least one embodiment, the reaction solution is allowed to incubate at a temperature of 25C to 37C for a reaction time of at least 1, 6, 12, 24, or 48 hours, before the amount of reaction product is determined.

EXAMPLES

[0123] Various features and embodiments of the disclosure are illustrated in the following representative examples, which are intended to be illustrative, and not limiting. Those skilled in

the art will readily appreciate that the specific examples are only illustrative of the invention as described more fully in the claims which follow thereafter. Every embodiment and feature described in the application should be understood to be interchangeable and combinable with every embodiment contained within.

Example 1: Genomic Integration of Heterologous CYP/CPR Gene Pairs into *Saccharomyces cerevisiae* Host Cells

[0124] This example illustrates the preparation of recombinant yeast host cell strains (*Saccharomyces cerevisiae*), with genomically integrated copies of a pair of heterologous genes encoding a cytochrome P450 (CYP) and a cytochrome reductase (CPR) from *Gibberella zeae*. This example also demonstrates the ability of these engineered strains to carry out the bioconversion of lithocholic acid (LCA) to ursodeoxycholic acid (UDCA) and/or 3-keto-lithocholic acid (3-KCA) to 3-keto-ursodeoxycholic acid (3-KUDCA).

[0125] A recombinant strain of *S. cerevisiae* denoted "SAND122," has been described in International Patent Application publication WO2022115710A1. SAND122 expresses a pair of genes from *Gibberella zeae* that encode a polypeptide with cytochrome P450 (CYP) activity (SEQ ID NO: 2) and a polypeptide with cytochrome P450 reductase (CPR) activity (SEQ ID NO: 4) and has been shown to carry out the conversion of LCA to UDCA and 3-KCA to 3-KUDCA.

[0126] *Material and Methods*

[0127] A. *S. cerevisiae* strain construction

[0128] To improve upon the bioconversion of LCA to UDCA and 3-KCA to 3-KUDCA observed in the strain, SAND122, the wild-type genes encoding the CYP polypeptide from *G. zeae* of SEQ ID NO: 2 and the CPR polypeptide from *G. zeae* of SEQ ID NO: 4 were codon-optimized for optimal expression in *S. cerevisiae* as SEQ ID NO: 1, and SEQ ID NO: 3, respectively. These codon-optimized genes were integrated into the *S. cerevisiae* genome under the bidirectional pGal1/10 promoter system (SEQ ID NO: 59) and ADH1t (SEQ ID NO: 60) and PGKt terminators (SEQ ID NO: 61) to generate single copy strains (SH010 and SH013) as follows.

[0129] Nucleic acids with the codon-optimized CYP gene of SEQ ID NO: 1, and the codon optimized CPR gene of SEQ ID NO: 3 were synthesized as fragments (Twist Bioscience, Inc., South San Francisco, California), which were further PCR amplified to incorporate ~20-30 bp overhang sequences homologous to the bidirectional pGal1/10 promoter and the respective terminator sequences on the 5' or 3' ends of the respective gene sequences. Locus specific PCR amplicons were also amplified with ~20-30 bp homologous overhangs to the terminator regions to enable efficient integration into the *S. cerevisiae* X-4 locus. The PCR primer sequences used in the following *S. cerevisiae* strain building examples by their Primer#, are listed in Table 4 below.

[0130] TABLE 4

Primer #	Primer Sequence	SEQ ID NO:
P1	ACTGCAAAGGCGTGCCCAA	62
P2	GAATTTTGGAAAATTCAATATAAATGGCACTACGAACATCAC	63
P3	GTGATGTTTCGTAGTGCCATTTATATTGAATTTTCAAAAATTC	64
P4	CTAGATCGGTGGCCATTATAGTTTTTCTCCTTG	65
P5	CAAGGAGAAAAAACTATAATGGCCACCGATCTAG	66
P6	ATCGATTTCAATTCAATTCAATTTATTCATCACTACCAATAGTGTC	67
P7	GACACTATTGGTAGTGATGAATAAATTGAATTGAATTGAAATCGAT	68
P8	GTAATTCAAAAAAGCGAATCTTCCCATGACATAGAAATATC	69
P9	GATATTTCTATGTCATGGGAAGATTGCGTTTTTTTTTTGAATTAC	70
P10	ATTGCAACCGACCCCTGGAG	71
P11	CTCTAAACCACCCTACAATGTAGC	72
P12	AGTGATCTTGCTCATCACGGAAC	73
P13	CTTCCAATGCTAATCCGGTCACTGCCACTG	74
P14	GGGCGTTCTATGAGGAAGCTTGATATGAGG	75
P15	GAGATGGGGAGTGATTTGCAAGC	76
P16	ACATAGAAATATCGAATGGGAAAAAACTGCATAAAGG	77
P17	AAAAAGGTTTATATACAATTACAATAGACCCCTT	78
P18	GCCACCGTGCTCAGCAAACGCTTGCAAATCACTCCCCATCTCGAGCTTTACTTG TGGAAG	79
P19	TAATGCCTTTATGCAGTTTTTTTTTCCCATTTCGATATTTCTATGTGGGGTACAC ACGACT	80
P20	CCAGGGAAAAACAAAATTATATAGAAAAGAAG	81
P21	GATAACTTCTGAATACTTTCTAAACTGATCA	82
P22	CTCAGCAAACGCTTGCAAATCACTCCCCATCTCGAAGGAATAGTTCAAAGTTGG CG	83
P23	GCAGTTTTTTTTTCCCATTTCGATATTTCTATGTCAATTTGTACAATTCCCCATT AGAGTC	84
P24	AGACCGTGAAATCTCTTTGCGGTAGTTAAATTACAATATCTATAAATGAGAACT TGACAA	85
P25	TATGTGCGACTATACTAGTATTTTAATATT	86
P26	CTCAGCAAACGCTTGCAAATCACTCCCCATCTCCAAGTTGCCTGCGAGTTTCTC	87
P27	TCGATGCCATCGGCAATAATACG	88
P28	AACCTGTGCCCTTTCTATTTTTATG	89
P29	GGCTGTTCTTCATAGCCCTTT	90
P30	AAATGTAGTTAAGTTTTCGTTAAACTTTTC	91
P31	TTGGATATCAGTAATGAGACGAAAAAGC	92
P32	ATTCTCGCCAAGGCATTACCATC	93
P33	CTTTCACAGGGTTCTCTTTGATCA	94
P34	TGTGGAAATGTAAAGAGCCCCA	95
P35	ACCGGTCGCGTTCCTGAAAC	96
P36	CTTAGATAAGTTATGCCACCAGTGCATACG	97
P37	CGCCAGATAAGGATAATTCAGCCGTCAC	98
P38	CGTAACCACCAAGTTTGGTTTCAGCGCG	99
P39	GCAGTTTTTTTTTCCCATTTCGATATTTCTATGTTCTAGGCCTGCATAATCGGC	100
P40	CTCTCTTAAACACATAGTAACAGCAAAA	101

[0131] With reference to the primer numbers listed in Table 4 and as illustrated in Scheme A of FIG. 1, Fragment A encompassing the X-4 site upstream homology region, ADH1t terminator, and CPR was amplified using the forward primer P1 and the reverse primer P2 as illustrated in

Scheme A of **FIG. 1**. Fragment B encompassing the bidirectional pGal1/10 promoter was amplified using the forward primer P3 and the reverse primer P4. Fragment C containing the CYP gene was amplified using the forward primer P5 and the reverse primer P6. Fragment D containing the PGK1t terminator was amplified using the forward primer P7 and the reverse primer P8. Fragment E containing the X-4 downstream homology region was amplified using the forward primer P9 and the reverse primer P10. The five fragments A-E were assembled by NEBuilder[®] HiFi Assembly and amplified as a single linear DONOR fragment using the following rescue forward primer P11 and the rescue reverse primer P12. The final assembled 4680 bp linear DONOR DNA was gel extracted and integrated as a knock-in using CRISPR-Cas9 into the X-4 locus of the yeast strain CENPK2-1D. This SH010 strain was used as the parent strain to introduce a gal80 knockout to enable induction of the bidirectional promoter pGal1/10 in the absence of glucose. The Gal80 knockout DONOR DNA was amplified using the forward rescue primer P13 and the reverse rescue primer P14 and integrated as a knock-in using CRISPR-Cas9 to disrupt the native gal80 gene in SH010 resulting in the Δ Gal80, single copy CYP-CPR strain SH013.

[0132] PCR primers summarized in Table 4 also were used to amplify the above-mentioned parts as follows and illustrated by Scheme B of **FIG. 1**. Recombinant host cell strains with double (SHP021), triple (SH020), and quadruple (SHP025) copies of the CYP/CPR cassette integrated in the genome were also built. To generate the multi-copy strains, upstream and downstream homology regions for the XI-2, XII-4, and X-2 loci were assembled to the 5' and 3' ends of the central DNA fragment (Fragment B) encompassing the CPR/CYP genes under the bidirectional pGAL1/10 promoter and respective terminators (found in strain SH013). Fragment B was amplified using the forward primer P15 and the reverse primer P16. For the XI-2 locus, the 5' upstream homology arm (Fragment A) was amplified using the forward primer P17 and the reverse primer P18 and the 3' downstream homology arm (Fragment C) was amplified using the forward primer P19 and the reverse primer P20. For the XII-4 locus, 5' upstream homology arm (Fragment A) was amplified using the forward primer P21 and the reverse primer P22 and the 3' downstream homology arm (Fragment C) was amplified using the forward primer P23 and the reverse primer P24. For the X-2 locus, 5' upstream homology arm (Fragment A) was amplified using the forward primer P25 and the reverse primer P26 and the 3' downstream homology arm (Fragment C) was amplified using the forward primer P39 (SEQ ID NO: 100) and the reverse primer P40 (SEQ ID NO: 101). The three fragments A, B, and C were assembled by NEBuilder[®] HiFi Assembly and amplified as a single linear DONOR DNA using the following site-specific primers: forward primer P27 and reverse primer P28 for the XI-2 DONOR, forward primer P29 and reverse primer P30 for the XII-4 DONOR, and forward primer P31 and reverse primer P32 for the X-2 DONOR. The XI-2 DONOR DNA was transformed and integrated into XI-2 locus as a knock-in using CRISPR-Cas9 into the SH013

(single copy) strain to generate the double copy strain SHP021. The XII-4 DONOR was transformed and integrated into XII-4 locus as a knock-in using CRISPR-Cas9 into the SHP021 (double copy) strain to generate the triple copy strain SH020. The X-2 DONOR was transformed and integrated into X-2 locus as a knock-in using CRISPR-Cas9 into the SH020 (triple copy) strain to generate the quadruple copy strain SHP025.

[0133] The lineage for the four above-described strains (SH013, SH021, SH020, and SHP025) constructed in *Saccharomyces cerevisiae* is summarized in **FIG. 2**.

[0134] B. Screening of strains for 3-KCA to 3-KUDCA conversion

[0135] A HTP screening assay of the recombinant host cell strains for the bioconversion of 3-KCA to 3-KUDCA was developed. 3-KCA to 3-KUDCA conversion over LCA to UDCA conversion was used to determine activity in HTP as 3-KCA was found to have better solubility than LCA in the bioconversion reaction buffer. The HTP screening assay was carried out as follows: Individual colonies of the recombinant *S. cerevisiae* strains were picked into 500 mL baffled conical flasks containing 2 x YPD media (100 mL working volume). The flasks were incubated at 30 °C for 48 h with shaking at 250 rpm at 85 % humidity. After this time, a 1% final concentration solution of galactose (2.5 mL of a 40 % stock solution) was added and the flask was further incubated for 4 hours to induce protein production. After this induction period, the biomass was harvested via centrifugation and the supernatant removed. A portion of the resulting biomass (approximately 2 g) was resuspended in 10 mL of bioconversion buffer (295 mg/L 3-KCA, 0.1 M phosphate buffer, pH 9) and the flask was incubated at 30 °C for 24 h with shaking at 250 rpm at 85 % humidity. After this time, an aliquot of the reaction mixture (1 mL) was extracted with a 1:1 ratio of MeOH (1 mL) with shaking at 30 °C, 250 rpm and 85 % humidity for 30 minutes. The biomass was then removed via centrifugation and the supernatant diluted further with MeOH (1000 x final dilution) before analysis using an Agilent 6470 Triple Quadrupole LC/MS.

[0136] LC/MS sample preparation: The bioconversion reaction was extracted and diluted with MeOH for sample preparation as described above. The prepared samples were loaded onto an Agilent 6470 Triple Quadrupole LC/MS and the compounds of interest were detected using MS/MS in SIM mode. The compounds were quantified relative to calibration curves which were prepared using serial dilutions of stock solutions of 3-KCA and 3-KUDCA.

[0137] LC/MS instrumentation and parameters: LC/MS system: Agilent 6470 Triple Quadrupole LC/MS; Column: Agilent EclipsePlusC18 column 1.8 μ m, 3.0 x 50 mm; Mobile phase: Solvent A: H₂O with 0.1% formic acid; Solvent B: Acetonitrile with 0.1% formic acid; Gradient: 0 - 0.2 min isocratic 40% B, 0.2 - 2.2 min gradient to 85% B, 2.2 - 2.8 min isocratic 85% B; Mode: MS/MS in SIM mode (monitoring at 389.3 for 3-KUDCA and 373.2 for 3-KCA); Fragmentor: 90 volts.

[0138] **Results**

[0139] Results of screening assays for the four *S. cerevisiae* strain builds are summarized in Table 5 below.

[0140] TABLE 5

Strain Name	<i>S. cerevisiae</i> Strain Description	3-KCA to 3-KUDCA % Conversion @295mg/L substrate loading
SH013	Single copy of wild-type CYP; Single copy of CPR	33
SHP021	Double copy of wild-type CYP; Double copy of CPR	52
SH020	Triple copy of wild-type CYP; Triple copy of CPR	68
SHP025	Quadruple copy of wild-type CYP; Quadruple copy of CPR	95

Example 2: Optimization of Heterologous CYP Gene in Recombinant *S. cerevisiae* Host Cells by Site Saturation Mutagenesis (SSM)

[0141] This example illustrates the preparation of site saturation mutagenesis (SSM) libraries of engineered polypeptides derived from the parent CYP polypeptide of SEQ ID NO: 2 in the SH013 strain from Example 1 and screening these variant strains for improved activity in the bioconversion of 3-KCA to 3-KUDCA relative to the bioconversion of the parent strain SH013.

[0142] Material and Methods

[0143] A. Site Saturation Mutagenesis library construction

[0144] The heterologous nucleic acid sequence from the single copy strain SH013 build of Example 1, which includes the codon-optimized polynucleotide sequence of SEQ ID NO:1, which encodes the CYP polypeptide of SEQ ID NO: 2, expressed under the pGal1/10 promoter system and PGK1t terminator, was used to design SSM oligonucleotide to generate SSM libraries at positions spanning the entire polypeptide sequence. To generate an appropriate screening strain to evaluate these libraries, the URA3 marker was integrated into the X-4 site (Easy-Clone 2.0) of a yeast strain under the pGal1 promoter within the pGal1/10 promoter system (SEQ ID NO: 59) and PGK1t terminator (SEQ ID NO: 61), along with the codon-optimized polynucleotide sequence of SEQ ID NO: 3, which encodes the CPR polypeptide of SEQ ID NO: 4, under the pGal10 promoter within the pGal1/10 promoter system (SEQ ID NO: 59) and ADH1t terminator (SEQ ID NO: 60). The resulting strain SHEVP003 was used as the negative control and screening host for CYP SSM library integration. Libraries were plated on selective media containing 5'-fluoroorotic acid, which selects for URA3 negative cells, indicating successful integration of DONOR DNA into the SHEVP003 screening strain. SH013 strain was

used as a control strain during screening of the SSM library strains for fold-improvement calculations with respect to conversion of 3-KCA to 3-KUDCA as described below.

[0145] Genomic DNA from the SH013 strain was used as the template to generate two PCR products: (1) a first PCR product (Fragment A), which does not harbor any degenerate codons, and (2) a second PCR product (Fragment B), which has sequence overlap with the Fragment A, and is amplified harboring one NNK degenerate codon only. Primers spanning the full 513 codons of the CYP polypeptide were designed according to standard site-saturation mutagenesis protocols and used for amplification of Fragments A and B and overlap extension.

[0146] Fragment A was amplified using a single forward primer P34 (SEQ ID NO: 95) and a series of 513 reverse primers designed according to the location of the desired mutagenesis site in CYP. The 513 reverse primers for Fragment A listed in Table 6A (below) and are provided in the accompanying Sequence Listing as SEQ ID NOs: 102-614.

[0147] TABLE 6A: Reverse primers for SSM library Fragment A amplification

Primer Name	Primer Sequence	SEQ ID NO:
Codon86_A1	ATCGGTATGAGCCTTGTAGGGCGTATCCTT	102
Codon108_A2	ATCTAACTGTCTTTTCAGATTTTAGAGCGTC	103
Codon111_A3	TTCGGTAAAATCTAACTGTCTTTCAGATTT	104
Codon112_A4	GACTTCGGTAAAATCTAACTGTCTTTCAGA	105
Codon115_A5	GTCTCTTGCGACTTCGGTAAAATCTAACTG	106
Codon116_A6	ATCGTCTCTTGCGACTTCGGTAAAATCTAA	107
Codon126_A7	TGGCTCGAATCCAGGAATATAACCGTGAGT	108
Codon132_A8	ATCGAACGGGGAACCGATTGGCTCGAATCC	109
Codon133_A9	CAGATCGAACGGGGAACCGATTGGCTCGAA	110
Codon136_A10	GAGCGGCACCAGATCGAACGGGGAACCGAT	111
Codon140_A11	ATACTTGTTAACGAGCGGCACCAGATCGAA	112
Codon186_A12	CATTCTGGAGACAATCCTCATGATATCTTC	113
Codon190_B1	TATTCTGGATGACATTCTGGAGACAATCCT	114
Codon209_B2	GTACTCAATACTAACTTTCAACCAATCATC	115
Codon210_B3	AGTGTACTCAATACTAACTTTCAACCAATC	116
Codon213_B4	CAATTGCACAGTGTACTCAATACTAACTTT	117
Codon214_B5	AAACAATTGCACAGTGTACTCAATACTAAC	118
Codon217_B6	TGCGGTTTGAAACAATTGCACAGTGTACTC	119
Codon220_B7	TAATTCGTCTGCGGTTTGAAACAATTGCAC	120
Codon297_B8	AATTTGTTTCAGTGGCGGGATCAGATTTGCC	121
Codon298_B9	TTTAATTTGTTTCAGTGGCGGGATCAGATTT	122
Codon300_B10	TGATAGTTTAATTTGTTTCAGTGGCGGGATC	123
Codon301_B11	CAGTGATAGTTTAATTTGTTTCAGTGGCGGG	124
Codon302_B12	CACCAGTGATAGTTTAATTTGTTTCAGTGGC	125
Codon303_C1	CGCCACCAGTGATAGTTTAATTTGTTTCAGT	126
Codon304_C2	AATCGCCACCAGTGATAGTTTAATTTGTTTC	127
Codon305_C3	GTGAATCGCCACCAGTGATAGTTTAATTTG	128
Codon306_C4	TGTGTGAATCGCCACCAGTGATAGTTTAAT	129

Codon364_C5	AGGGCGAAGTCTCTGTGACTCTTTTATGAC	130
Codon365_C6	AACAGGGCGAAGTCTCTGTGACTCTTTTAT	131
Codon367_C7	TAATAAAACAGGGCGAAGTCTCTGTGACTC	132
Codon368_C8	ACCTAATAAAACAGGGCGAAGTCTCTGTGA	133
Codon370_C9	GAAAGCACCTAATAAAACAGGGCGAAGTCT	134
Codon440_C10	TCCCATTTGATCATGGCTTGTGCTGACAAG	135
Codon445_C11	CAAGCCATGTCCGAATCCCATTGATCATG	136
Codon446_C12	GTGCAAGCCATGTCCGAATCCCATTGATC	137
Codon447_D1	CGCGTGCAAGCCATGTCCGAATCCCATTG	138
Codon448_D2	GCACGCGTGCAAGCCATGTCCGAATCCCAT	139
Codon449_D3	TGGGCACGCGTGCAAGCCATGTCCGAATCC	140
Codon450_D4	TCCTGGGCACGCGTGCAAGCCATGTCCGAA	141
Codon452_D5	AAATCTTCCTGGGCACGCGTGCAAGCCATG	142
Codon453_D6	GAAAAATCTTCCTGGGCACGCGTGCAAGCC	143
Codon456_D7	ATTGGCTGCGAAAAATCTTCCTGGGCACGC	144
Codon485_D8	TCCTAAAGCCTTAGACTTAGGTACAACACC	145
Codon486_D9	CATTCTTAAGCCTTAGACTTAGGTACAAC	146
Codon488_D10	TAAAGACATTCTTAAGCCTTAGACTTAGG	147
Codon491_D11	GTCACCAAGTAAAGACATTCTTAAGCCTT	148
Codon4_A1	GGTGGCCATGGTTTTTCTCCTTGACGTTA	149
Codon7_A2	GTCTAGATCGGTGGCCATGGTTTTTCTCC	150
Codon8_A3	TAGGTCTAGATCGGTGGCCATGGTTTTTCTC	151
Codon10_A4	TAATACTAGGTCTAGATCGGTGGCCATGGT	152
Codon13_A5	ACTTTTTCTTAATACTAGGTCTAGATCGGT	153
Codon15_A6	GTATTGACTTTTTCTTAATACTAGGTCTAG	154
Codon17_A7	TAATGCGTATTGACTTTTTCTTAATACTAG	155
Codon19_A8	ACAAAATAATGCGTATTGACTTTTTCTTAA	156
Codon24_A9	AAATAAAGTTATGCCACAAAATAATGCGTA	157
Codon37_A10	CCCGTTTCCGAGAAGACTATACTTTAGTAT	158
Codon47_A11	CTTAGGGTTGATATAAGGATATTGCTTGCC	159
Codon66_A12	TCGTGCGTTCTCGATAAAATCCTGGACAAC	160
Codon67_B1	GTCTCGTGCGTTCTCGATAAAATCCTGGAC	161
Codon71_B2	CTTAGTCAAAATGTCTCGTGCGTTCTCGAT	162
Codon74_B3	TGAGCGACCCCTAGTCAAAATGTCTCGTG	163
Codon81_B4	GTAGGGCGTATCCTTGTAAGTGAGCGACC	164
Codon84_B5	ATGAGCCTTGTAGGGCGTATCCTTGTAAG	165
Codon94_B6	CGGGATTACCAAGACGTCCCCTAAATCGGT	166
Codon96_B7	TTCAGGCGGGATTACCAAGACGTCCCCTAA	167
Codon103_B8	AGATTTTAGAGCGTCGGCAAATTCAGGCGG	168
Codon104_B9	TTCAGATTTTAGAGCGTCGGCAAATTCAGG	169
Codon110_B10	GGTAAATCTAACTGTCTTCAGATTTTAG	170
Codon117_B11	AGTATCGTCTCTTGCGACTTCGGTAAAAATC	171
Codon118_B12	GTGAGTATCGTCTCTTGCGACTTCGGTAAA	172
Codon127_C1	GATTGGCTCGAATCCAGGAATATAACCGTG	173
Codon130_C2	CGGGGAACCGATTGGCTCGAATCCAGGAAT	174
Codon131_C3	GAACGGGGAACCGATTGGCTCGAATCCAGG	175

Codon138_C4	GTTAACGAGCGGCACCAGATCGAACGGGGA	176
Codon139_C5	CTTGTTAACGAGCGGCACCAGATCGAACGG	177
Codon146_C6	TGCCAAAGCCCTTGTAAGATACTTGTTAAC	178
Codon160_C7	TACACCTAAAGAAGCTTCGGCCCAATGG	179
Codon163_C8	AACATGGTTTACACCTAAAGAAGCTTCGGC	180
Codon165_C9	GCCTAAACATGGTTTACACCTAAAGAAGC	181
Codon172_C10	AGGATGCCACTCAGTAGAAGTGCCTAAAC	182
Codon174_C11	GTTAATAGGATGCCACTCAGTAGAAGTGCC	183
Codon176_C12	CCCTGGGTAAATAGGATGCCACTCAGTAGA	184
Codon181_D1	CCTCATGATATCTTCCCCTGGGTAAATAGG	185
Codon184_D2	GGAGACAATCCTCATGATATCTTCCCCTGG	186
Codon188_D3	GGATGACATTCTGGAGACAATCCTCATGAT	187
Codon192_D4	CATGAATATTCTGGATGACATTCTGGAGAC	188
Codon199_D5	ATCTTTACAAAGTTCCTCACCCATGAATAT	189
Codon202_D6	CCAATCATCATCTTTACAAAGTTCCTCACC	190
Codon203_D7	CAACCAATCATCATCTTTACAAAGTTCCTC	191
Codon204_D8	TTTCAACCAATCATCATCTTTACAAAGTTC	192
Codon207_D9	AATACTAAGTTTCAACCAATCATCATCTTT	193
Codon219_D10	TTCGTCTGCGGTTTGAAACAATTGCACAGT	194
Codon227_D11	TGTCCAACGTGGATAGTTACGTAATTCGTC	195
Codon231_D12	AATATATGGTCTTGTTCCAACGTGGATAGTT	196
Codon238_E1	ACAGGAAGGCAAAAACAGTGAATATATGG	197
Codon240_E2	CCCTTGACAGGAAGGCAAAAACAGTGAAT	198
Codon260_E3	ATTTCTCCTATCAATATGGGGTTGCAATAA	199
Codon274_E4	TGGTGATGGTCTACCTTCAGCGATAGCTTC	200
Codon275_E5	GAATGGTGATGGTCTACCTTCAGCGATAGC	201
Codon279_E6	TATTGAATCGTCGAATGGTGATGGTCTACC	202
Codon287_E7	CTCGTATTCATTTTCAAACCATTCTATTGA	203
Codon310_E8	TAAATCAGTTGTTGTGTGAATCGCCACCAG	204
Codon314_E9	TGTTTCAGACAGTAAATCAGTTGTTGTGTG	205
Codon323_E10	TTCTGGTTGCAAAGCTATATTAAACATTGT	206
Codon326_E11	ACCAAGGAGTTCTGGTTGCAAAGCTATATT	207
Codon327_E12	TGGACCAAGGAGTTCTGGTTGCAAAGCTAT	208
Codon330_F1	TTCACGTAGTGGACCAAGGAGTTCTGGTTG	209
Codon332_F2	TATCTCTTCACGTAGTGGACCAAGGAGTTC	210
Codon333_F3	AACATCTCTTCACGTAGTGGACCAAGGAG	211
Codon339_F4	TTCCGTGGATAAAACCGTAACATCTCTTC	212
Codon343_F5	CTTTTTTAGACCTTCCGTGGATAAAACCGT	213
Codon345_F6	CGACGTCTTTTTTAGACCTTCCGTGGATAA	214
Codon349_F7	CAAATTGTAAAACGACGTCTTTTTTAGACC	215
Codon352_F8	CATCAACTTCAAATTGTAAAACGACGTCTT	216
Codon353_F9	ATCCATCAACTTCAAATTGTAAAACGACGT	217
Codon357_F10	TTTTATGACCGAATCCATCAACTTCAAATT	218
Codon363_F11	GCGAAGTCTCTGTGACTCTTTTATGACCGA	219
Codon366_F12	TAAAACAGGGCGAAGTCTCTGTGACTCTTT	220
Codon374_G1	TGCCATTCTCCTGAAAGCACCTAATAAAC	221

Codon379_G2	GGTAACATCAGCCAATGCCATTCTCCTGAA	222
Codon387_G3	TTTAATTACGTCGCCATTGGGCAAGGTAAC	223
Codon388_G4	TTTTTTAATTACGTCGCCATTGGGCAAGGT	224
Codon390_G5	GGTACCTTTTTTAATTACGTCGCCATTGGG	225
Codon391_G6	CTTGGTACCTTTTTTAATTACGTCGCCATT	226
Codon392_G7	GATCTTGGTACCTTTTTTAATTACGTCGCC	227
Codon395_G8	GTCGCAAATGATCTTGGTACCTTTTTTAAT	228
Codon397_G9	TGTAGTGTGCGCAAATGATCTTGGTACCTTT	229
Codon405_G10	ATAGTATTCTGGATTCCATTGATGTGTAGT	230
Codon420_G11	TCTCATTTGCAAAAATCTATATGCATTGAA	231
Codon424_G12	ACCCGGTGTCTGTCTCATTTGCAAAAATCT	232
Codon425_H1	CTGACCCGGTGTCTGTCTCATTTGCAAAAA	233
Codon427_H2	TTTGTCTTGACCCGGTGTCTGTCTCATTTG	234
Codon436_H3	ATGGCTTGTGCTGACAAGGTGTGCTCTTTT	235
Codon444_H4	GCCATGTCCGAATCCCATTGATCATGGCT	236
Codon458_H5	AATCTCATTTGGCTGCGAAAAATCTTCCTGG	237
Codon464_H6	ATGACACAGCGCAATCTTAATCTCATTTGGC	238
Codon478_H7	TACAACACCTTCTGGTAATTTCCAATCATA	239
Codon479_H8	AGGTACAACACCTTCTGGTAATTTCCAATC	240
Codon481_H9	AGACTTAGGTACAACACCTTCTGGTAATTT	241
Codon483_H10	AGCCTTAGACTTAGGTACAACACCTTCTGG	242
Codon487_H11	AGACATTCTTAAAGCCTTAGACTTAGGTAC	243
Codon507_H12	TATATCGATTTTCGGCTGCTCTCCTTTTCAC	244
Codon3_A1	GGCCATGGTTTTTTCTCCTTGACGTTAAAG	245
Codon9_A2	TACTAGGTCTAGATCGGTGGCCATGGTTTT	246
Codon18_A3	AAATAATGCGTATTGACTTTTTCTTAATAC	247
Codon22_A4	AGTTATGCCACAAAATAATGCGTATTGACT	248
Codon23_A5	TAAAGTTATGCCACAAAATAATGCGTATTG	249
Codon26_A6	AAAGCTAAATAAAGTTATGCCACAAAATAA	250
Codon31_A7	ATACTTTAGTATGAAAAAGCTAAATAAAGT	251
Codon32_A8	ACTATACTTTAGTATGAAAAAGCTAAATAA	252
Codon33_A9	AAGACTATACTTTAGTATGAAAAAGCTAAA	253
Codon43_A10	ATAAGGATATTGCTTGCCCCGTTTCCGAG	254
Codon51_A11	CTCAAAGGTTTCTTAGGGTTGATATAAGG	255
Codon56_A12	TCGTTGATTGACAGCTCAAAGGTTTCTT	256
Codon59_B1	CTGGACAACCTCGTTGATTGACAGCTCAAA	257
Codon61_B2	AAAATCCTGGACAACCTCGTTGATTGACAG	258
Codon83_B3	AGCCTTGTAGGGCGTATCCTTGTAAGTGA	259
Codon99_B4	GTCGGCAAATTCAGGCGGGATTACCAAGAC	260
Codon100_B5	AGCGTCGGCAAATTCAGGCGGGATTACCAA	261
Codon106_B6	CTGTCTTTCAGATTTTAGAGCGTCGGCAAA	262
Codon109_B7	AAAATCTAACTGTCTTTCAGATTTTAGAGC	263
Codon113_B8	TGCGACTTCGGTAAATCTAACTGTCTTTC	264
Codon119_B9	ACCGTGAGTATCGTCTCTTGCGACTTCGGT	265
Codon120_B10	ATAACCGTGAGTATCGTCTCTTGCGACTTC	266
Codon123_B11	TCCAGGAATATAACCGTGAGTATCGTCTCT	267

Codon124_B12	GAATCCAGGAATATAACCGTGAGTATCGTC	268
Codon125_C1	CTCGAATCCAGGAATATAACCGTGAGTATC	269
Codon141_C2	AAGATACTTGTTAACGAGCGGCACCAGATC	270
Codon142_C3	TGTAAGATACTTGTTAACGAGCGGCACCAG	271
Codon151_C4	TGGTTTTGTTAGTTTTGCCAAAGCCCTTGT	272
Codon153_C5	CCACAATGGTTTTGTTAGTTTTGCCAAAGC	273
Codon155_C6	TTCGGCCCACAATGGTTTTGTTAGTTTTGC	274
Codon159_C7	ACCTAAAGAAGCTTCGGCCCACAATGGTTT	275
Codon168_C8	AGTAGAAGTGCCTAAAACATGGTTTACACC	276
Codon171_C9	ATGCCACTCAGTAGAAGTGCCTAAAACATG	277
Codon175_C10	TGGGTTAATAGGATGCCACTCAGTAGAAGT	278
Codon179_C11	GATATCTTCCCCTGGGTTAATAGGATGCCA	279
Codon189_C12	TCTGGATGACATTCTGGAGACAATCCTCAT	280
Codon194_D1	CTCACCCATGAATATTCTGGATGACATTCT	281
Codon206_D2	ACTAACTTTCAACCAATCATCATCTTTACA	282
Codon223_D3	ATAGTTACGTAATTCGTCTGCGGTTTGAAA	283
Codon226_D4	CCAACGTGGATAGTTACGTAATTCGTCTGC	284
Codon228_D5	TCTTGTCCAACGTGGATAGTTACGTAATTC	285
Codon229_D6	TGGTCTTGTCCAACGTGGATAGTTACGTAA	286
Codon244_D7	TTTTCTCCTAACCCCTTGACAGGAAGGCAA	287
Codon245_D8	CAATTTTCTCCTAACCCCTTGACAGGAAGG	288
Codon246_D9	CTGCAATTTTCTCCTAACCCCTTGACAGGA	289
Codon257_D10	ATCAATATGGGGTTGCAATAAATCACGCGC	290
Codon261_D11	AGCATTTCTCCTATCAATATGGGGTTGCAA	291
Codon262_D12	CACAGCATTTCTCCTATCAATATGGGGTTG	292
Codon264_E1	TTTCTTCACAGCATTTCTCCTATCAATATG	293
Codon266_E2	AGCTTCTTTCTTCACAGCATTTCTCCTATC	294
Codon283_E3	TTCAAACCATTCTATTGAATCGTCGAATGG	295
Codon285_E4	TTCAATTTCAAACCATTCTATTGAATCGTC	296
Codon288_E5	GCCCTCGTATTCAATTTCAAACCATTCAT	297
Codon294_E6	AGTGGCGGGATCAGATTTGCCCTCGTATTC	298
Codon313_E7	TTCAGACAGTAAATCAGTTGTTGTGTGAAT	299
Codon319_E8	AGCTATATTAAACATTGTTTCAGACAGTAA	300
Codon324_E9	GAGTTCTGGTTGCAAAGCTATATTAAACAT	301
Codon328_E10	TAGTGGACCAAGGAGTTCTGGTTGCAAAGC	302
Codon329_E11	ACGTAGTGGACCAAGGAGTTCTGGTTGCAA	303
Codon334_E12	CGTAACTATCTCTTCACGTAGTGGACCAAG	304
Codon336_F1	TAAAACCGTAACTATCTCTTCACGTAGTGG	305
Codon341_F2	TAGACCTTCCGTGGATAAAACCGTAACTAT	306
Codon347_F3	GTAAAACGACGTCTTTTTTAGACCTTCCGT	307
Codon348_F4	ATTGTAAAACGACGTCTTTTTTAGACCTTC	308
Codon351_F5	CAACTTCAAATTGTAAAACGACGTCTTTTT	309
Codon359_F6	TGACTCTTTTATGACCGAATCCATCAACTT	310
Codon360_F7	CTGTGACTCTTTTATGACCGAATCCATCAA	311
Codon362_F8	AAGTCTCTGTGACTCTTTTATGACCGAATC	312
Codon371_F9	CCTGAAAGCACCTAATAAAACAGGGCGAAG	313

Codon373_F10	CATTCTCCTGAAAGCACCTAATAAAACAGG	314
Codon375_F11	CAATGCCATTCTCCTGAAAGCACCTAATAA	315
Codon376_F12	AGCCAATGCCATTCTCCTGAAAGCACCTAA	316
Codon384_G1	GTCGCCATTGGGCAAGGTAACATCAGCCAA	317
Codon389_G2	ACCTTTTTTAATTACGTCGCCATTGGGCAA	318
Codon394_G3	GCAAATGATCTTGGTACCTTTTTTAATTAC	319
Codon400_G4	CCATTGATGTGTAGTGTGCGAAATGATCTT	320
Codon403_G5	TTCTGGATTCCATTGATGTGTAGTGTGCGA	321
Codon407_G6	ATCTGGATAGTATTCTGGATTCCATTGATG	322
Codon411_G7	GAACCTACTAGCATCTGGATAGTATTCTGG	323
Codon422_G8	TGTCTGTCTCATTTGCAAAAATCTATATGC	324
Codon426_G9	GTCCTGACCCGGTGTCTGTCTCATTTGCAA	325
Codon429_G10	TGCTCTTTTGTCTGACCCGGTGTCTGTCT	326
Codon441_G11	GAATCCCATTTGATCATGGCTTGTGCTGAC	327
Codon443_G12	ATGTCCGAATCCCATTTGATCATGGCTTGT	328
Codon463_H1	ACACAGCGCAATCTTAATCTCATTTGGCTGC	329
Codon471_H2	CCAATCATATTTCAATAGCATATGACACAG	330
Codon475_H3	TTCTGGTAATTTCCAATCATATTTCAATAG	331
Codon482_H4	CTTAGACTTAGGTACAACACCTTCTGGTAA	332
Codon490_H5	ACCAAGTAAAGACATTCTAAAGCCTTAGA	333
Codon495_H6	TTTGGCTTCCCGGTCACCAAGTAAAGACAT	334
Codon497_H7	CATCAGTTTGGCTTCCCGGTCACCAAGTAA	335
Codon498_H8	CACCATCAGTTTGGCTTCCCGGTCACCAAG	336
Codon504_H9	TTCGGCTGCTCTCCTTTTCACCATCAGTTT	337
Codon508_H10	GTCTATATCGATTTTCGGCTGCTCTCCTTTT	338
Codon510_H11	AATAGTGTCTATATCGATTTTCGGCTGCTCT	339
Codon512_H12	ACTACCAATAGTGTCTATATCGATTTTCGGC	340
Codon6_A1	TAGATCGGTGGCCATGGTTTTTTCTCCTTG	341
Codon39_A2	CTTGCCCCCGTTTCCGAGAAGACTATACTT	342
Codon42_A3	AGGATATTGCTTGCCCCCGTTTCCGAGAAG	343
Codon48_A4	TTTCTTAGGGTTGATATAAGGATATTGCTT	344
Codon54_A5	ATTGACAGCTCAAAAGGTTTCTTAGGGTT	345
Codon55_A6	TTGATTGACAGCTCAAAAGGTTTCTTAGG	346
Codon57_A7	AACTCGTTGATTGACAGCTCAAAAGGTTT	347
Codon60_A8	ATCCTGGACAACCTCGTTGATTGACAGCTC	348
Codon62_A9	GATAAAATCCTGGACAACCTCGTTGATTCTGA	349
Codon68_A10	AATGTCTCGTGCGTTCTCGATAAAATCCTG	350
Codon69_A11	CAAAATGTCTCGTGCGTTCTCGATAAAATC	351
Codon76_A12	GTAAAGTGAGCGACCCCTAGTCAAAATGTC	352
Codon80_B1	GGGCGTATCCTTGTAAGTGAGCGACCCCTT	353
Codon88_B2	CCCTAAATCGGTATGAGCCTTGAGGGCGT	354
Codon93_B3	GATTACCAAGACGTCCCTTAAATCGGTATG	355
Codon95_B4	AGGCGGGATTACCAAGACGTCCCTTAAATC	356
Codon102_B5	TTTTAGAGCGTCGGCAAAATTCAGGCGGGAT	357
Codon105_B6	TCTTTCAGATTTTAGAGCGTCGGCAAAATTC	358
Codon107_B7	TAACTGTCTTTCAGATTTTAGAGCGTCGGC	359

Codon121_B8	AATATAACCGTGAGTATCGTCTCTTGCGAC	360
Codon128_B9	ACCGATTGGCTCGAATCCAGGAATATAACC	361
Codon129_B10	GGAACCGATTGGCTCGAATCCAGGAATATA	362
Codon143_B11	CCTTGTAAGATACTTGTTAACGAGCGGCAC	363
Codon148_B12	TAGTTTTGCCAAAGCCCTTGTAAGATACTT	364
Codon152_C1	CAATGGTTTTGTTAGTTTTGCCAAAGCCCT	365
Codon154_C2	GGCCCAATGGTTTTGTTAGTTTTGCCAA	366
Codon162_C3	ATGGTTTACACCTAAAGAAGCTTCGGCCCA	367
Codon166_C4	AGTGCCTAAAACATGGTTTACACCTAAAGA	368
Codon169_C5	CTCAGTAGAAGTGCCTAAAACATGGTTTAC	369
Codon170_C6	CCACTCAGTAGAAGTGCCTAAAACATGGTT	370
Codon193_C7	ACCCATGAATATTCTGGATGACATTCTGGA	371
Codon200_C8	ATCATCTTTACAAAGTTCCTCACCCATGAA	372
Codon205_C9	AACTTTCAACCAATCATCATCTTTACAAAG	373
Codon211_C10	CACAGTGTACTCAATACTAACTTTCAACCA	374
Codon218_C11	GTCTGCGGTTTGAAACAATTGCACAGTGTA	375
Codon221_C12	ACGTAATTCGTCTGCGGTTTGAAACAATTG	376
Codon233_D1	CCAGTGAATATATGGTCTTGTCCAACGTGG	377
Codon234_D2	AAACCAGTGAATATATGGTCTTGTCCAACG	378
Codon237_D3	GGAAGGCAAAAACAGTGAATATATGGTCT	379
Codon241_D4	AACCCCTTGACAGGAAGGCAAAAACAGTG	380
Codon242_D5	CCTAACCCCTTGACAGGAAGGCAAAAACCA	381
Codon249_D6	ACGCGCTTCCTGCAATTTTCTCCTAACCCC	382
Codon251_D7	TAAATCACGCGCTTCCTGCAATTTTCTCCT	383
Codon252_D8	CAATAAATCACGCGCTTCCTGCAATTTTCT	384
Codon253_D9	TTGCAATAAATCACGCGCTTCCTGCAATTT	385
Codon256_D10	AATATGGGGTTGCAATAAATCACGCGCTTC	386
Codon258_D11	CCTATCAATATGGGGTTGCAATAAATCACG	387
Codon259_D12	TCTCCTATCAATATGGGGTTGCAATAAATC	388
Codon265_E1	TTCTTTCTTCACAGCATTTCTCCTATCAAT	389
Codon267_E2	GATAGCTTCTTTCTTCACAGCATTTCTCCT	390
Codon269_E3	TTCAGCGATAGCTTCTTTCTTCACAGCATT	391
Codon271_E4	TCTACCTTCAGCGATAGCTTCTTTCTTCAC	392
Codon277_E5	ATCGTCGAATGGTGATGGTCTACCTTCAGC	393
Codon278_E6	TGAATCGTCGAATGGTGATGGTCTACCTTC	394
Codon289_E7	TTTGCCCTCGTATTCATTTTCAAACCATTC	395
Codon292_E8	GGGATCAGATTTGCCCTCGTATTCATTTTC	396
Codon295_E9	TTCAGTGGCGGGATCAGATTTGCCCTCGTA	397
Codon299_E10	TAGTTTAATTTGTTTCAGTGGCGGGATCAGA	398
Codon307_E11	TGTTGTGTGAATCGCCACCAGTGATAGTTT	399
Codon308_E12	AGTTGTTGTGTGAATCGCCACCAGTGATAG	400
Codon315_F1	CATTGTTTCAGACAGTAAATCAGTTGTTGT	401
Codon316_F2	AAACATTGTTTCAGACAGTAAATCAGTTGT	402
Codon318_F3	TATATTAAACATTGTTTCAGACAGTAAATC	403
Codon320_F4	CAAAGCTATATTAAACATTGTTTCAGACAG	404
Codon322_F5	TGGTTGCAAAGCTATATTAAACATTGTTTC	405

Codon325_F6	AAGGAGTTCTGGTTGCAAAGCTATATTAAA	406
Codon340_F7	ACCTTCCGTGGATAAAACCGTAACTATCTC	407
Codon344_F8	CGTCTTTTTTAGACCTTCCGTGGATAAAAC	408
Codon350_F9	CTTCAAATTGTAAAACGACGTCTTTTTTAG	409
Codon372_F10	TCTCCTGAAAGCACCTAATAAAACAGGGCG	410
Codon377_F11	ATCAGCCAATGCCATTCTCCTGAAAGCACC	411
Codon381_F12	GGGCAAGGTAACATCAGCCAATGCCATTCT	412
Codon383_G1	GCCATTGGGCAAGGTAACATCAGCCAATGC	413
Codon399_G2	TTGATGTGTAGTGTGCGAAATGATCTTGGT	414
Codon401_G3	ATTCCATTGATGTGTAGTGTGCGAAATGAT	415
Codon406_G4	TGGATAGTATTCTGGATTCCATTGATGTGT	416
Codon408_G5	AGCATCTGGATAGTATTCTGGATTCCATTG	417
Codon412_G6	ATTGAACTTACTAGCATCTGGATAGTATTC	418
Codon415_G7	TCTATATGCATTGAACTTACTAGCATCTGG	419
Codon423_G8	CGGTGTCTGTCTCATTTGCAAAAATCTATA	420
Codon439_G9	CATTTGATCATGGCTTGTGCTGACAAGGTG	421
Codon460_G10	AATCTTAATCTCATTGGCTGCGAAAAATCT	422
Codon462_G11	CAGCGCAATCTTAATCTCATTGGCTGCGAA	423
Codon466_G12	TAGCATATGACACAGCGCAATCTTAATCTC	424
Codon467_H1	CAATAGCATATGACACAGCGCAATCTTAAT	425
Codon468_H2	TTTCAATAGCATATGACACAGCGCAATCTT	426
Codon469_H3	ATATTTCAATAGCATATGACACAGCGCAAT	427
Codon470_H4	ATCATATTTCAATAGCATATGACACAGCGC	428
Codon472_H5	TTTCCAATCATATTTCAATAGCATATGACA	429
Codon474_H6	TGGTAATTTCCAATCATATTTCAATAGCAT	430
Codon477_H7	AACACCTTCTGGTAATTTCCAATCATATTT	431
Codon480_H8	CTTAGGTACAACACCTTCTGGTAATTTCCA	432
Codon500_H9	CCTTTTCACCATCAGTTTGGCTTCCCCGGTC	433
Codon501_H10	TCTCCTTTTCACCATCAGTTTGGCTTCCCCG	434
Codon503_H11	GGCTGCTCTCCTTTTCACCATCAGTTTGGC	435
Codon506_H12	ATCGATTTTCGGCTGCTCTCCTTTTCACCAT	436
Codon1_A1	GGTTTTTTCTCCTTGACGTAAAGTATAGA	437
Codon11_A2	TCCTAATACTAGGTCTAGATCGGTGGCCAT	438
Codon14_A3	TTGACTTTTTCTTAATACTAGGTCTAGATC	439
Codon16_A4	TGCGTATTGACTTTTTCTTAATACTAGGTCT	440
Codon25_A5	GCTAAATAAAGTTATGCCACAAAATAATGC	441
Codon28_A6	TATGAAAAAGCTAAATAAAGTTATGCCACA	442
Codon30_A7	CTTTAGTATGAAAAAGCTAAATAAAGTTAT	443
Codon34_A8	GAGAAGACTATACTTTAGTATGAAAAAGCT	444
Codon35_A9	TCCGAGAAGACTATACTTTAGTATGAAAAA	445
Codon36_A10	GTTTCCGAGAAGACTATACTTTAGTATGAA	446
Codon38_A11	GCCCCCGTTTCCGAGAAGACTATACTTTAG	447
Codon44_A12	GATATAAGGATATTGCTTGCCCCCGTTTCC	448
Codon45_B1	GTTGATATAAGGATATTGCTTGCCCCCGTT	449
Codon46_B2	AGGGTTGATATAAGGATATTGCTTGCCCCC	450
Codon49_B3	AGGTTTCTTAGGGTTGATATAAGGATATTG	451

Codon50_B4	AAAAGGTTTCTTAGGGTTGATATAAGGATA	452
Codon52_B5	CAGCTCAAAAGGTTTCTTAGGGTTGATATA	453
Codon53_B6	CGACAGCTCAAAAGGTTTCTTAGGGTTGAT	454
Codon64_B7	GTTCTCGATAAAATCCTGGACAACCTCGTTG	455
Codon65_B8	TGCGTTCTCGATAAAATCCTGGACAACCTCG	456
Codon72_B9	ACCCTTAGTCAAAATGTCTCGTGCGTTCTC	457
Codon73_B10	GCGACCCCTTAGTCAAAATGTCTCGTGCGTT	458
Codon78_B11	ATCCTTGTAAGTGAGCGACCCCTAGTCAA	459
Codon89_B12	GTCCCTTAAATCGGTATGAGCCTTGATAGG	460
Codon91_C1	CAAGACGTCCCTTAAATCGGTATGAGCCTT	461
Codon92_C2	TACCAAGACGTCCCTTAAATCGGTATGAGC	462
Codon98_C3	GGCAAATTCAGGCGGGATTACCAAGACGTC	463
Codon114_C4	TCTTGCGACTTCGGTAAATCTAACTGTCT	464
Codon122_C5	AGGAATATAACCGTGAGTATCGTCTCTTGC	465
Codon135_C6	CGGCACCAGATCGAACGGGGAACCGATTGG	466
Codon137_C7	AACGAGCGGCACCAGATCGAACGGGGAACC	467
Codon149_C8	TGTTAGTTTTTGCCAAAGCCCTTGTAAGATA	468
Codon150_C9	TTTTGTTAGTTTTTGCCAAAGCCCTTGTAAG	469
Codon156_C10	AGCTTCGGCCCAATGGTTTTGTTAGTTT	470
Codon158_C11	TAAAGAAGCTTCGGCCCAATGGTTTTGT	471
Codon173_C12	AATAGGATGCCACTCAGTAGAAGTGCCTAA	472
Codon187_D1	TGACATTCTGGAGACAATCCTCATGATATC	473
Codon195_D2	TTCCTCACCCATGAATATTCTGGATGACAT	474
Codon197_D3	ACAAAGTTCCTCACCCATGAATATTCTGGA	475
Codon198_D4	TTTACAAAGTTCCTCACCCATGAATATTCT	476
Codon208_D5	CTCAATACTAACTTTCAACCAATCATCATC	477
Codon212_D6	TTGCACAGTGTACTCAATACTAACTTTCAA	478
Codon222_D7	GTTACGTAATTCGTCTGCGGTTTGAAACAA	479
Codon225_D8	ACGTGGATAGTTACGTAATTCGTCTGCGGT	480
Codon230_D9	ATATGGTCTTGTCACCGTGATAGTTACG	481
Codon232_D10	GTGAATATATGGTCTTGTCACCGTGATAG	482
Codon235_D11	CAAAAACCAAGTGAATATATGGTCTTGTC	483
Codon239_D12	TTGACAGGAAGGCAAAAACCAAGTGAATATA	484
Codon243_E1	TCTCCTAACCCCTTGACAGGAAGGCAAAAA	485
Codon247_E2	TTCCTGCAATTTTCTCCTAACCCCTTGACA	486
Codon248_E3	CGCTTCCTGCAATTTTCTCCTAACCCCTTG	487
Codon250_E4	ATCACGCGCTTCCTGCAATTTTCTCCTAAC	488
Codon254_E5	GGGTTGCAATAAATCACGCGCTTCCTGCAA	489
Codon255_E6	ATGGGGTTGCAATAAATCACGCGCTTCCTG	490
Codon263_E7	CTTACAGCATTTCTCCTATCAATATGGGG	491
Codon272_E8	TGGTCTACCTTCAGCGATAGCTTCTTCTT	492
Codon273_E9	TGATGGTCTACCTTCAGCGATAGCTTCTT	493
Codon282_E10	AAACCATTCTATTGAATCGTCAATGGTGA	494
Codon286_E11	GTATTCATTTTCAAACCATTCTATTGAATC	495
Codon291_E12	ATCAGATTTGCCCTCGTATTCATTTTCAAA	496
Codon296_F1	TTGTTCAAGTGGCGGGATCAGATTTGCCCTC	497

Codon312_F2	AGACAGTAAATCAGTTGTTGTGTGAATCGC	498
Codon317_F3	ATTAAACATTGTTTCAGACAGTAAATCAGT	499
Codon331_F4	CTCTTCACGTAGTGGACCAAGGAGTTCTGG	500
Codon335_F5	AACCGTAACTATCTCTTCACGTAGTGGACC	501
Codon337_F6	GGATAAAACCGTAACTATCTCTTCACGTAG	502
Codon355_F7	GACCGAATCCATCAACTTCAAATTGTAAAA	503
Codon356_F8	TATGACCGAATCCATCAACTTCAAATTGTA	504
Codon358_F9	CTCTTTTATGACCGAATCCATCAACTTCAA	505
Codon361_F10	TCTCTGTGACTCTTTTATGACCGAATCCAT	506
Codon378_F11	AACATCAGCCAATGCCATTCTCCTGAAAGC	507
Codon380_F12	CAAGGTAACATCAGCCAATGCCATTCTCCT	508
Codon382_G1	ATTGGGCAAGGTAACATCAGCCAATGCCAT	509
Codon386_G2	AATTACGTCGCCATTGGGCAAGGTAACATC	510
Codon396_G3	AGTGTCGCAAATGATCTTGGTACCTTTTTT	511
Codon398_G4	ATGTGTAGTGTGCGAAATGATCTTGGTACC	512
Codon402_G5	TGGATTCCATTGATGTGTAGTGTGCGAAAT	513
Codon410_G6	CTTACTAGCATCTGGATAGTATTCTGGATT	514
Codon413_G7	TGCATTGAACTTACTAGCATCTGGATAGTA	515
Codon418_G8	TTGCAAAAATCTATATGCATTGAACTTACT	516
Codon421_G9	CTGTCTCATTTGCAAAAATCTATATGCATT	517
Codon428_G10	TCTTTTGTCTGACCCGGTGTCTGTCTCAT	518
Codon432_G11	GACAAGGTGTGCTCTTTTGTCTGACCCGG	519
Codon433_G12	GCTGACAAGGTGTGCTCTTTTGTCTGACC	520
Codon435_H1	GCTTGTGCTGACAAGGTGTGCTCTTTGTC	521
Codon438_H2	TTGATCATGGCTTGTGCTGACAAGGTGTGC	522
Codon455_H3	GGCTGCGAAAAATCTTCTGGGCACGCGTG	523
Codon457_H4	CTCATTGGCTGCGAAAAATCTTCTGGGCA	524
Codon461_H5	CGCAATCTTAATCTCATTGGCTGCGAAAAA	525
Codon484_H6	TAAAGCCTTAGACTTAGGTACAACACCTTC	526
Codon492_H7	CCGGTCACCAAGTAAAGACATTCTAAAGC	527
Codon493_H8	TTCCCGGTACCAAGTAAAGACATTCTTAA	528
Codon502_H9	TGCTCTCCTTTTCACCATCAGTTTGGCTTC	529
Codon505_H10	GATTTTCGGCTGCTCTCCTTTTCACCATCAG	530
Codon509_H11	AGTGTCTATATCGATTTTCGGCTGCTCTCCT	531
Codon511_H12	ACCAATAGTGTCTATATCGATTTTCGGCTGC	532
Codon2_A1	CATGGTTTTTTCTCCTTGACGTTAAAGTAT	533
Codon5_A2	ATCGGTGGCCATGGTTTTTTCTCCTTGACG	534
Codon12_A3	TTTTCCTAATACTAGGTCTAGATCGGTGGC	535
Codon20_A4	GCCACAAAATAATGCGTATTGACTTTTTTCC	536
Codon21_A5	TATGCCACAAAATAATGCGTATTGACTTTT	537
Codon27_A6	GAAAAAGCTAAATAAAGTTATGCCACAAAA	538
Codon29_A7	TAGTATGAAAAAGCTAAATAAAGTTATGCC	539
Codon40_A8	TTGCTTGCCCCCGTTTCCGAGAAGACTATA	540
Codon41_A9	ATATTGCTTGCCCCCGTTTCCGAGAAGACT	541
Codon58_A10	GACAACTCGTTGATTTCGACAGCTCAAAAAG	542
Codon63_A11	CTCGATAAAATCCTGGACAACCTCGTTGATT	543

Codon70_A12	AGTCAAAATGTCTCGTGCGTTCTCGATAAA	544
Codon75_B1	AAGTGAGCGACCCCTTAGTCAAAATGTCTCG	545
Codon77_B2	CTTGTAAGTGAGCGACCCCTTAGTCAAAAT	546
Codon79_B3	CGTATCCTTGTAAGTGAGCGACCCCTTAGT	547
Codon82_B4	CTTGTAGGGCGTATCCTTGTAAGTGAGCG	548
Codon85_B5	GGTATGAGCCTTGTAGGGCGTATCCTTGTA	549
Codon87_B6	TAAATCGGTATGAGCCTTGTAGGGCGTATC	550
Codon90_B7	GACGTCCCCTAAATCGGTATGAGCCTTGTA	551
Codon97_B8	AAATTCAGGCGGGATTACCAAGACGTCCCC	552
Codon101_B9	TAGAGCGTCGGCAAATTCAGGCGGGATTAC	553
Codon134_B10	CACCAGATCGAACGGGGAACCGATTGGCTC	554
Codon144_B11	AGCCCTTGTAAGATACTTGTTAACGAGCGG	555
Codon145_B12	CAAAGCCCTTGTAAGATACTTGTTAACGAG	556
Codon147_C1	TTTTGCCAAAGCCCTTGTAAGATACTTGTT	557
Codon157_C2	AGAAGCTTCGGCCCACAATGGTTTTGTAG	558
Codon161_C3	GTTTACACCTAAAGAAGCTTCGGCCCACAA	559
Codon164_C4	TAAAACATGGTTTACACCTAAAGAAGCTTC	560
Codon167_C5	AGAAGTGCCTAAAACATGGTTTACACCTAA	561
Codon177_C6	TTCCCCTGGGTAAATAGGATGCCACTCAGT	562
Codon178_C7	ATCTTCCCCTGGGTAAATAGGATGCCACTC	563
Codon180_C8	CATGATATCTTCCCCTGGGTAAATAGGATG	564
Codon182_C9	AATCCTCATGATATCTTCCCCTGGGTAAAT	565
Codon183_C10	GACAATCCTCATGATATCTTCCCCTGGGT	566
Codon185_C11	TCTGGAGACAATCCTCATGATATCTTCCCC	567
Codon191_C12	GAATATTCTGGATGACATTCTGGAGACAAT	568
Codon196_D1	AAGTTCCTCACCATGAATATTCTGGATGA	569
Codon201_D2	ATCATCATCTTTACAAAGTTCTCACCAT	570
Codon215_D3	TTGAAACAATTGCACAGTGACTCAATACT	571
Codon216_D4	GGTTTGAAACAATTGCACAGTGACTCAAT	572
Codon224_D5	TGGATAGTTACGTAATTCGTCTGCGGTTG	573
Codon236_D6	AGGCAAAAACCAGTGAATATATGGTCTTGT	574
Codon268_D7	AGCGATAGCTTCTTTCTTCACAGCATTTCT	575
Codon270_D8	ACCTTCAGCGATAGCTTCTTTCTTCACAGC	576
Codon276_D9	GTCGAATGGTGATGGTCTACCTTCAGCGAT	577
Codon280_D10	TTCTATTGAATCGTCGAATGGTGATGGTCT	578
Codon281_D11	CCATTCTATTGAATCGTCGAATGGTGATGG	579
Codon284_D12	ATTTTCAAACCATTCATTGAATCGTCGAA	580
Codon290_E1	AGATTTGCCCTCGTATTCATTTCAAACCA	581
Codon293_E2	GGCGGGATCAGATTTGCCCTCGTATTCATT	582
Codon309_E3	ATCAGTTGTTGTGTGAATCGCCACCAGTGA	583
Codon311_E4	CAGTAAATCAGTTGTTGTGTGAATCGCCAC	584
Codon321_E5	TTGCAAAGCTATATTAAACATTGTTTCAGA	585
Codon338_E6	CGTGGATAAAACCGTAACTATCTCTTCACG	586
Codon342_E7	TTTTAGACCTTCCGTGGATAAAACCGTAAC	587
Codon346_E8	AAACGACGTCTTTTTTAGACCTTCCGTGGA	588
Codon354_E9	CGAATCCATCAACTTCAAATTGTAAAACGA	589

Codon369_E10	AGCACCTAATAAAACAGGGCGAAGTCTCTG	590
Codon385_E11	TACGTCGCCATTGGGCAAGGTAACATCAGC	591
Codon393_E12	AATGATCTTGGTACCTTTTTTAATTACGTC	592
Codon404_F1	GTATTCTGGATTCCATTGATGTGTAGTGTC	593
Codon409_F2	ACTAGCATCTGGATAGTATTCTGGATTCCA	594
Codon414_F3	ATATGCATTGAACTTACTAGCATCTGGATA	595
Codon416_F4	AAATCTATATGCATTGAACTTACTAGCATC	596
Codon417_F5	CAAAAATCTATATGCATTGAACTTACTAGC	597
Codon419_F6	CATTTGCAAAAATCTATATGCATTGAACTT	598
Codon430_F7	GTGTGCTCTTTTGTCTGACCCGGTGTCTG	599
Codon431_F8	AAGGTGTGCTCTTTTGTCTGACCCGGTGT	600
Codon434_F9	TGTGCTGACAAGGTGTGCTCTTTTGTCTCTG	601
Codon437_F10	ATCATGGCTTGTGCTGACAAGGTGTGCTCT	602
Codon442_F11	TCCGAATCCCATTTGATCATGGCTTGTGCT	603
Codon451_F12	TCTTCCTGGGCACGCGTGCAAGCCATGTCC	604
Codon454_G1	TGCGAAAAATCTTCCTGGGCACGCGTGCAA	605
Codon459_G2	CTTAATCTCATTGGCTGCGAAAAATCTTCC	606
Codon465_G3	CATATGACACAGCGCAATCTTAATCTCATT	607
Codon473_G4	TAATTTCCAATCATATTTCAATAGCATATG	608
Codon476_G5	ACCTTCTGGTAATTTCCAATCATATTTCAA	609
Codon489_G6	AAGTAAAGACATTCTTAAAGCCTTAGACTT	610
Codon494_G7	GGCTTCCCGGTCACCAAGTAAAGACATTCC	611
Codon496_G8	CAGTTTGGCTTCCCGGTCACCAAGTAAAGA	612
Codon499_G9	TTTCACCATCAGTTTGGCTTCCCGGTCACC	613
Codon513_G10	ATCACTACCAATAGTGTCTATATCGATTTC	614

[0148] Fragment B was amplified with a single reverse primer P33 (SEQ ID NO: 94) and a series of 370 forward primers that included a single NNK degenerate codon spanning across successive positions of the CYP encoding gene of SEQ ID NO: 1. The 370 forward primers for Fragment B are listed in Table 6B (below) and are provided in the accompanying Sequence Listing as SEQ ID NOs: 615-984.

[0149] **TABLE 6B:** Forward primers for SSM library Fragment B amplification

Primer Name	Primer Sequence	SEQ ID NO:
Codon4_A1	TAACGTCAAGGAGAAAAAACCATGGCCACCNNKCTAGACCTAGTA	615
Codon7_A2	GGAGAAAAAACCATGGCCACCGATCTAGACNNKGTATTAGGAAAA	616
Codon8_A3	GAAAAAACCATGGCCACCGATCTAGACCTANNKTTAGGAAAAAGT	617
Codon10_A4	ACCATGGCCACCGATCTAGACCTAGTATTANNKAAAAGTCAATAC	618
Codon13_A5	ACCGATCTAGACCTAGTATTAGGAAAAAGTNNKTACGCATTATTT	619
Codon15_A6	CTAGACCTAGTATTAGGAAAAAGTCAATACNNKTTATTTTGTGGC	620
Codon17_A7	CTAGTATTAGGAAAAAGTCAATACGCATTANNKTGTGGCATAACT	621
Codon19_A8	TTAGGAAAAAGTCAATACGCATTATTTTGTNNKATAACTTTATTT	622
Codon24_A9	TACGCATTATTTTGTGGCATAACTTTATTTNNKTTTTCATACTA	623
Codon37_A10	ATACTAAAGTATAGTCTTCTCGGAAACGGGNNKAAGCAATATCCT	624

Codon47_A11	GGCAAGCAATATCCTTATATCAACCCTAAGNNKCCTTTTGAGCTG	625
Codon66_A12	GTTGTCCAGGATTTTATCGAGAACGCACGANNKATTTTGACTAAG	626
Codon67_B1	GTCCAGGATTTTATCGAGAACGCACGAGACNNKTTGACTAAGGGT	627
Codon71_B2	ATCGAGAACGCACGAGACATTTTGACTAAGNNKCGCTCAGTTTAC	628
Codon74_B3	GCACGAGACATTTTGACTAAGGGTCGCTCANNKTACAAGGATACG	629
Codon81_B4	GGTCGCTCACTTTACAAGGATACGCCCTACNNKGCTCATACCGAT	630
Codon84_B5	CTTTACAAGGATACGCCCTACAAGGCTCATNNKGATTTAGGGGAC	631
Codon94_B6	ACCGATTTAGGGGACGTCTTGGTAATCCCGNNKGAATTTGCCGAC	632
Codon96_B7	TTAGGGGACGTCTTGGTAATCCCGCTGAANNKGCCGACGCTCTA	633
Codon103_B8	CCGCCTGAATTTGCCGACGCTCTAAAATCTNNKAGACAGTTAGAT	634
Codon104_B9	CCTGAATTTGCCGACGCTCTAAAATCTGAANNKAGTTAGATTTT	635
Codon110_B10	CTAAAATCTGAAAGACAGTTAGATTTTACCNNKGTCGCAAGAGAC	636
Codon117_B11	GATTTTACCGAAGTCGCAAGAGACGATACTNNKGTTATATTCCT	637
Codon118_B12	TTTACCGAAGTCGCAAGAGACGATACTCACNNKTATATTCCTGGA	638
Codon127_C1	CACGGTTATATTCCTGGATTTCGAGCCAATCNNKTCCTCGTTCGAT	639
Codon130_C2	ATTCCTGGATTTCGAGCCAATCGGTTCCCGNNKGATCTGGTGCCG	640
Codon131_C3	CCTGGATTTCGAGCCAATCGGTTCCCGTTCNNKCTGGTGCCGCTC	641
Codon138_C4	TCCCGTTCGATCTGGTGCCGCTCGTTAACNNKTATCTTACAAGG	642
Codon139_C5	CCGTTTCGATCTGGTGCCGCTCGTTAACAAGNNKCTTACAAGGGCT	643
Codon146_C6	GTTAACAAGTATCTTACAAGGGCTTTGGCANNKCTAACAAAACCA	644
Codon160_C7	CCATTGTGGGCCGAAGCTTCTTTAGGTGTANNKCATGTTTTAGGC	645
Codon163_C8	GCCGAAGCTTCTTTAGGTGTAAACCATGTTNNKGGCACTTCTACT	646
Codon165_C9	GCTTCTTTAGGTGTAAACCATGTTTTAGGCNNKCTACTGAGTGG	647
Codon172_C10	GTTTTAGGCACCTTCTACTGAGTGGCATCCTNNKAACCCAGGGGAA	648
Codon174_C11	GGCACTTCTACTGAGTGGCATCCTATTAACNNKGGGGAAGATATC	649
Codon176_C12	TCTACTGAGTGGCATCCTATTAACCCAGGGNNKGATATCATGAGG	650
Codon181_D1	CCTATTAACCCAGGGGAAGATATCATGAGGNNKGCTCTCCAGAATG	651
Codon184_D2	CCAGGGGAAGATATCATGAGGATTGTCTCCNNKATGTCATCCAGA	652
Codon188_D3	ATCATGAGGATTGTCTCCAGAATGTCATCCNNKATATTCATGGGT	653
Codon192_D4	GTCTCCAGAATGTCATCCAGAATATTCATGNNKGAGGAACTTTGT	654
Codon199_D5	ATATTCATGGGTGAGGAACCTTTGTAAAGATNNKGATTGGTTGAAA	655
Codon202_D6	GGTGAGGAACCTTTGTAAAGATGATGATTGGNNKAAAGTTAGTATT	656
Codon203_D7	GAGGAACCTTTGTAAAGATGATGATTGGTTGNNKGTTAGTATTGAG	657
Codon204_D8	GAACCTTTGTAAAGATGATGATTGGTTGAAAANNKAGTATTGAGTAC	658
Codon207_D9	AAAGATGATGATTGGTTGAAAGTTAGTATTNNKTACACTGTGCAA	659
Codon219_D10	ACTGTGCAATTGTTTCAAACCGCAGACGAANNKCGTAACTATCCA	660
Codon227_D11	GACGAATTACGTAACCTATCCACGTTGGACANNKCCATATATTCAC	661
Codon231_D12	AACTATCCACGTTGGACAAGACCATATATNNKTGGTTTTTGCCT	662
Codon238_E1	CCATATATTCAGTGGTTTTTGCCTTCCTGTNNKGGGGTTAGGAGA	663
Codon240_E2	ATTCAGTGGTTTTTGCCTTCCTGTCAAGGGNNKAGGAGAAAATTG	664
Codon260_E3	TTATTGCAACCCCATATTGATAGGAGAAAATNNKGTGAAGAAAGAA	665
Codon274_E4	GAAGCTATCGCTGAAGGTAGACCATCACCANNKGACGATTCAATA	666
Codon275_E5	GCTATCGCTGAAGGTAGACCATCACCATTNNKGATTCAATAGAA	667
Codon279_E6	GGTAGACCATCACCATTGACGATTCAATANNKTGGTTTGAAAAT	668
Codon287_E7	TCAATAGAATGGTTTGAAAATGAATACGAGNNKAAATCTGATCCC	669
Codon310_E8	CTGGTGGCGATTACACAACAACCTGATTTANNKTCTGAAACAATG	670

Codon314_E9	CACACAACAACCTGATTTACTGTCTGAAACANNKTTTAATATAGCT	671
Codon323_E10	ACAATGTTTAAATATAGCTTTGCAACCAGAANNKCTTGGTCCACTA	672
Codon326_E11	AATATAGCTTTTGCAACCAGAACTCCTTGGTNNKCTACGTGAAGAG	673
Codon327_E12	ATAGCTTTGCAACCAGAACTCCTTGGTCCANNKCGTGAAGAGATA	674
Codon330_F1	CAACCAGAACTCCTTGGTCCACTACGTGAANNKATAGTTACGGTT	675
Codon332_F2	GAACTCCTTGGTCCACTACGTGAAGAGATANNKACGGTTTTATCC	676
Codon333_F3	CTCCTTGGTCCACTACGTGAAGAGATAGTTNNKGTTTTATCCACG	677
Codon339_F4	GAAGAGATAGTTACGGTTTTATCCACGGAANNKCTAAAAAAGACG	678
Codon343_F5	ACGGTTTTATCCACGGAAGGTCTAAAAAAGNNKTCGTTTTACAAT	679
Codon345_F6	TTATCCACGGAAGGTCTAAAAAAGACGTCGNNKTACAATTTGAAG	680
Codon349_F7	GGTCTAAAAAAGACGTCGTTTTACAATTTGNNKTTGATGGATTTCG	681
Codon352_F8	AAGACGTCGTTTTTACAATTTGAAGTTGATGNNKTCGGTCATAAAA	682
Codon353_F9	ACGTCGTTTTTACAATTTGAAGTTGATGGATNNKGTCATAAAAGAG	683
Codon357_F10	AATTTGAAGTTGATGGATTTCGGTCATAAAAAANNKTCACAGAGACTT	684
Codon363_F11	TCGGTCATAAAAGAGTCACAGAGACTTCGCNNKGTTTTATTAGGT	685
Codon366_F12	AAAGAGTCACAGAGACTTCGCCCTGTTTTANNKGGTGCTTTCAGG	686
Codon374_G1	GTTTTATTAGGTGCTTTCAGGAGAATGGCANNKGCTGATGTTACC	687
Codon379_G2	TTCAGGAGAATGGCATTGGCTGATGTTACCNNKCCCAATGGCGAC	688
Codon387_G3	GTTACCTTGCCCAATGGCGACGTAATTAANNKGGTACCAAGATC	689
Codon388_G4	ACCTTGCCCAATGGCGACGTAATTAAAAAANNKACCAAGATCATT	690
Codon390_G5	CCCAATGGCGACGTAATTAAAAAAGGTACCNNKATCATTTGCGAC	691
Codon391_G6	AATGGCGACGTAATTAAAAAAGGTACCAAGNNKATTTGCGACACT	692
Codon392_G7	GGCGACGTAATTAAAAAAGGTACCAAGATCANNKTCGCGACACTACA	693
Codon395_G8	ATTA AAAAAGGTACCAAGATCATTTGCGACNNKACACATCAATGG	694
Codon397_G9	AAAGGTACCAAGATCATTTGCGACACTACANNKCAATGGAATCCA	695
Codon405_G10	ACTACACATCAATGGAATCCAGAATACTATNNKGATGCTAGTAAG	696
Codon420_G11	TTCAATGCATATAGATTTTTTGCAAATGAGANNKACACCGGGTCAG	697
Codon424_G12	AGATTTTTTGCAAATGAGACAGACACCGGGTNNKGACAAAAGAGCA	698
Codon425_H1	TTTTTGC AAATGAGACAGACACCGGGTCAGNNKAAAAGAGCACAC	699
Codon427_H2	CAAATGAGACAGACACCGGGTCAGGACAAAANNKGACACCTTGTC	700
Codon436_H3	AAAAGAGCACACCTTGTCAGCACAAGCCATNNKCAAATGGGATTC	701
Codon444_H4	AGCCATGATCAAATGGGATTTCGGACATGGCANNKACGCGTGCCCA	702
Codon458_H5	CCAGGAAGATTTTTTCGAGCCAATGAGATTNNKATTGCGCTGTGT	703
Codon464_H6	GCCAATGAGATTAAAGATTGCGCTGTGT CATNNKCTATTGAAATAT	704
Codon478_H7	TATGATTGGAAATTACCAGAAGGTGTTGTANNKAAGCTAAGGCT	705
Codon479_H8	GATTGGAAATTACCAGAAGGTGTTGTACCTNNKCTAAGGCTTTA	706
Codon481_H9	AAATTACCAGAAGGTGTTGTACCTAAGTCTNNKGCTTTAGGAATG	707
Codon483_H10	CCAGAAGGTGTTGTACCTAAGTCTAAGGCTNNKGAATGTCTTTA	708
Codon487_H11	GTACCTAAGTCTAAGGCTTTAGGAATGTCTNNKCTTGGTGACCGG	709
Codon507_H12	GTGAAAAGGAGAGCAGCCGAAATCGATATANNKACTATTGGTAGT	710
Codon3_A1	CTTTAACGTCAAGGAGAAAAAACCATGGCCNNKGATCTAGACCTA	711
Codon9_A2	AAAACCATGGCCACCGATCTAGACCTAGTANNKGAAAAAGTCAA	712
Codon18_A3	GTATTAGGAAAAAGTCAATACGCATTATTTNNKGGCATAACTTTA	713
Codon22_A4	AGTCAATACGCATTATTTTGTGGCATAACTNNKTTTAGCTTTTTTC	714
Codon23_A5	CAATACGCATTATTTTGTGGCATAACTTTANNKAGCTTTTTTCATA	715
Codon26_A6	TTATTTTGTGGCATAACTTTATTTAGCTTTNNKATACTAAAGTAT	716

Codon31_A7	ACTTTATTTAGCTTTTTTCATACTAAAGTATNNKCTTCTCGGAAAC	717
Codon32_A8	TTATTTAGCTTTTTTCATACTAAAGTATAGTNNKCTCGGAAACGGG	718
Codon33_A9	TTTAGCTTTTTTCATACTAAAGTATAGTCTTNNKGGAACGGGGGC	719
Codon43_A10	CTCGGAAACGGGGGCAAGCAATATCCTTATNNKAACCTAAGAAA	720
Codon51_A11	CCTTATATCAACCCTAAGAAACCTTTTGAGNNKTCGAATCAACGA	721
Codon56_A12	AAGAAACCTTTTGAGCTGTCTGAATCAACGANNKGTCAGGATTTT	722
Codon59_B1	TTTGAGCTGTCTGAATCAACGAGTTGTCCAGNNKTTTATCGAGAAC	723
Codon61_B2	CTGTCTGAATCAACGAGTTGTCCAGGATTTTNNKGAGAACGCACGA	724
Codon83_B3	TCACTTTACAAGGATACGCCCTACAAGGCTNNKACCGATTTAGGG	725
Codon99_B4	GTCTTGGAATCCCGCCTGAATTTGCCGACNNKCTAAAATCTGAA	726
Codon100_B5	TTGGTAATCCCGCCTGAATTTGCCGACGCTNNKAAATCTGAAAGA	727
Codon106_B6	TTTGCCGACGCTCTAAAATCTGAAAGACAGNNKGATTTTACCGAA	728
Codon109_B7	GCTCTAAAATCTGAAAGACAGTTAGATTTTNNKGAGTCGCAAGA	729
Codon113_B8	GAAAGACAGTTAGATTTTACCGAAGTCGCANNKGACGATACTCAC	730
Codon119_B9	ACCGAAGTCGCAAGAGACGATACTCACGGTNNKATTCCTGGATTTC	731
Codon120_B10	GAAGTCGCAAGAGACGATACTCACGGTTATNNKCCTGGATTTCGAG	732
Codon123_B11	AGAGACGATACTCACGGTTATATTCTTGANNKGAGCCAATCGGT	733
Codon124_B12	GACGATACTCACGGTTATATTCTTGATTTCNNKCCAATCGGTTCC	734
Codon125_C1	GATACTCACGGTTATATTCTTGATTTCGAGNNKATCGGTTCCCCG	735
Codon141_C2	GATCTGGTGCCGCTCGTTAACAAGTATCTTNNKAGGGCTTTGGCA	736
Codon142_C3	CTGGTGCCGCTCGTTAACAAGTATCTTACANNKGCTTTGGCAAAA	737
Codon151_C4	ACAAGGGCTTTGGCAAAACTAACAAAACCAANNKTGGGCCGAAGCT	738
Codon153_C5	GCTTTGGCAAAACTAACAAAACCATTTGTGGNNKGAAGCTTCTTTA	739
Codon155_C6	GCAAAACTAACAAAACCATTTGTGGGCCGAANNKTCTTTAGGTGTA	740
Codon159_C7	AAACCATTGTGGGCCGAAGCTTCTTTAGGTNNKAACCATGTTTTA	741
Codon168_C8	GGTGTAACCATGTTTTAGGCACCTTCTACTNNKTGGCATCCTATT	742
Codon171_C9	CATGTTTTAGGCACCTTCTACTGAGTGGCATNNKATTAACCCAGGG	743
Codon175_C10	ACTTCTACTGAGTGGCATCCTATTAAACCCANNKGAAGATATCATG	744
Codon179_C11	TGGCATCCTATTAAACCCAGGGGAAGATATCNNKAGGATTGTCTCC	745
Codon189_C12	ATGAGGATTGTCTCCAGAATGTCATCCAGANNKTTCATGGGTGAG	746
Codon194_D1	AGAATGTATCCAGAATATTCATGGGTGAGNNKCTTTGTAAAGAT	747
Codon206_D2	TGTAAAGATGATGATTGGTTGAAAGTTAGTNNKGAGTACACTGTG	748
Codon223_D3	TTTCAAACCGCAGACGAATTACGTAACCTATNNKCGTTGGACAAGA	749
Codon226_D4	GCAGACGAATTACGTAACCTATCCACGTTGGNNKAGACCATATATT	750
Codon228_D5	GAATTACGTAACCTATCCACGTTGGACAAGANNKTATATTCACTGG	751
Codon229_D6	TTACGTAACCTATCCACGTTGGACAAGACCANNKATTCACCTGGTTT	752
Codon244_D7	TTGCCTTCCTGTCAAGGGGTTAGGAGAAAAANNKAGGAAGCGCGT	753
Codon245_D8	CCTTCCTGTCAAGGGGTTAGGAGAAAAATTGNNKGAAGCGCGTGAT	754
Codon246_D9	TCCTGTCAAGGGGTTAGGAGAAAAATTGCAGNNKCGCGGTGATTTA	755
Codon257_D10	GCGCGTGATTTATTGCAACCCCATATTGATNNKAGAAATGCTGTG	756
Codon261_D11	TTGCAACCCCATATTGATAGGAGAAATGCTNNKAAGAAAGAAGCT	757
Codon262_D12	CAACCCCATATTGATAGGAGAAATGCTGTGNNKAAAGAAGCTATC	758
Codon264_E1	CATATTGATAGGAGAAATGCTGTGAAGAAANNKGCTATCGCTGAA	759
Codon266_E2	GATAGGAGAAATGCTGTGAAGAAAGAAGCTNNKGCTGAAGGTAGA	760
Codon283_E3	CCATTGACGATTCAATAGAATGGTTTGAANNKGAATACGAGGGC	761
Codon285_E4	GACGATTCAATAGAATGGTTTGAAGAAATGAANNKGAGGGCAAATCT	762

Codon288_E5	ATAGAATGGTTTGAAAATGAATACGAGGGCANNKCTGATCCCGCC	763
Codon294_E6	GAATACGAGGGCAAATCTGATCCCGCCACTNNKCAAATTAACTA	764
Codon313_E7	ATTACACACAACAACTGATTTACTGTCTGAANNKATGTTTAATATA	765
Codon319_E8	TTACTGTCTGAAACAATGTTTAATATAGCTNNKCAACCAGAACTC	766
Codon324_E9	ATGTTTAATATAGCTTTGCAACCAGAACTCANNKGGTCCACTACGT	767
Codon328_E10	GCTTTGCAACCAGAACTCCTTGGTCCACTANNKGAAGAGATAGTT	768
Codon329_E11	TTGCAACCAGAACTCCTTGGTCCACTACGTNNKGAGATAGTTACG	769
Codon334_E12	CTTGGTCCACTACGTGAAGAGATAGTTACGNNKTATCCACGGAA	770
Codon336_F1	CCACTACGTGAAGAGATAGTTACGGTTTTTANNKACGGAAGGTCTA	771
Codon341_F2	ATAGTTACGGTTTTTATCCACGGAAGGTCTANNKAAGACGTCGTTT	772
Codon347_F3	ACGGAAGGTCTAAAAAAGACGTCGTTTTACNNKTGAAGTTGATG	773
Codon348_F4	GAAGGTCTAAAAAAGACGTCGTTTTACAATNNKAAGTTGATGGAT	774
Codon351_F5	AAAAAGACGTCGTTTTTACAATTTGAAGTTGNNKGATTCGGTCATA	775
Codon359_F6	AAGTTGATGGATTTCGGTCATAAAAAGAGTCAANNKAGACTTCGCCCT	776
Codon360_F7	TTGATGGATTTCGGTCATAAAAGAGTCACAGNNKCTTCGCCCTGTT	777
Codon362_F8	GATTCGGTCATAAAAGAGTCACAGAGACTTNNKCCTGTTTTATTA	778
Codon371_F9	CTTCGCCCTGTTTTATTAGGTGCTTTCAGGNNKATGGCATTGGCT	779
Codon373_F10	CCTGTTTTATTAGGTGCTTTCAGGAGAATGNNKTGGCTGATGTT	780
Codon375_F11	TTATTAGGTGCTTTCAGGAGAATGGCATTGNNKGATGTTACCTTG	781
Codon376_F12	TTAGGTGCTTTCAGGAGAATGGCATTGGCTNNKGTTACCTTGCCC	782
Codon384_G1	TTGGCTGATGTTACCTTGCCCAATGGCGACNNKATTAaaaaaAGGT	783
Codon389_G2	TTGCCCAATGGCGACGTAATTAaaaaaAGGTNNKAAGATCATTTGC	784
Codon394_G3	GTAATTAaaaaaAGGTACCAAGATCATTTGCNNKACTACACATCAA	785
Codon400_G4	AAGATCATTTGCGACACTACACATCAATGGNNKCCAGAATACTAT	786
Codon403_G5	TGCGACACTACACATCAATGGAATCCAGAANNKTATCCAGATGCT	787
Codon407_G6	CATCAATGGAATCCAGAATACTATCCAGATNNKAGTAAGTTCAAT	788
Codon411_G7	CCAGAATACTATCCAGATGCTAGTAAGTTCNNKGCATATAGATTT	789
Codon422_G8	GCATATAGATTTTTTGCAAATGAGACAGACANNKGGTCAGGACAAA	790
Codon426_G9	TTGCAAATGAGACAGACACCGGGTCAGGACNNKAGAGCACACCTT	791
Codon429_G10	AGACAGACACCGGGTCAGGACAAAAGAGCANNKCTTGTCAGCACA	792
Codon441_G11	GTCAGCACAAAGCCATGATCAAATGGGATTCNNKATGGCTTGAC	793
Codon443_G12	ACAAGCCATGATCAAATGGGATTCGGACATNNKTGCACGCGTGC	794
Codon463_H1	GCAGCCAATGAGATTAAGATTGCGCTGTGTNNKATGCTATTGAAA	795
Codon471_H2	CTGTGTCATATGCTATTGAAATATGATTGGNNKTACCAGAAGGT	796
Codon475_H3	CTATTGAAATATGATTGGAATTACCAGAANNKGTTGTACCTAAG	797
Codon482_H4	TTACCAGAAGGTGTTGTACCTAAGTCTAAGNNKTAGGAATGTCT	798
Codon490_H5	TCTAAGGCTTTAGGAATGTCTTTACTTGGTNNKCGGAAGCCAAA	799
Codon495_H6	ATGTCTTTACTTGGTGACCGGGAAGCCAAANNKATGGTGAAAAGG	800
Codon497_H7	TTACTTGGTGACCGGGAAGCCAAACTGATGNNKAAAAGGAGAGCA	801
Codon498_H8	CTTGGTGACCGGGAAGCCAAACTGATGGTGNNKAGGAGAGCAGCC	802
Codon504_H9	AAACTGATGGTGAAAAGGAGAGCAGCCGAANNKGATATAGACACT	803
Codon508_H10	AAAAGGAGAGCAGCCGAAATCGATATAGACNNKATTGGTAGTGAT	804
Codon510_H11	AGAGCAGCCGAAATCGATATAGACACTATTNNKAGTGATGAATAA	805
Codon512_H12	GCCGAAATCGATATAGACACTATTGGTAGTNNKGAATAAATTGAA	806
Codon1_A1	TCTATACTTTAACGTCAAGGAGAAAAAACNNKGCCACCGATCTA	807
Codon11_A2	ATGGCCACCGATCTAGACCTAGTATTAGGANNKAGTCAATACGCA	808

Codon14_A3	GATCTAGACCTAGTATTAGGAAAAAGTCAANNKGCATTATTTTGT	809
Codon16_A4	GACCTAGTATTAGGAAAAAGTCAATACGCANNKTTTTGTGGCATA	810
Codon25_A5	GCATTATTTTGTGGCATAACTTTATTTAGCANNKTTCTACTAAAG	811
Codon28_A6	TGTGGCATAACTTTATTTAGCTTTTTCATANNKAAGTATAGTCTT	812
Codon30_A7	ATAACTTTATTTAGCTTTTTCATACTAAAGNNKAGTCTTCTCGGA	813
Codon34_A8	AGCTTTTTTCATACTAAAGTATAGTCTTCTCANNKACGGGGGCAAG	814
Codon35_A9	TTTTTCATACTAAAGTATAGTCTTCTCGGANNKGGGGGCAAGCAA	815
Codon36_A10	TTCATACTAAAGTATAGTCTTCTCGGAAACNNKGGCAAGCAATAT	816
Codon38_A11	CTAAAGTATAGTCTTCTCGGAAACGGGGGCNNKCAATATCCTTAT	817
Codon44_A12	GGAAACGGGGGCAAGCAATATCCTTATATCANNKCCTAAGAAACCT	818
Codon45_B1	AACGGGGGCAAGCAATATCCTTATATCAACNNKAAGAAACCTTTT	819
Codon46_B2	GGGGGCAAGCAATATCCTTATATCAACCCTNNKAAACCTTTTGAG	820
Codon49_B3	CAATATCCTTATATCAACCCTAAGAAACCTNNKGAGCTGTCTGAAT	821
Codon50_B4	TATCCTTATATCAACCCTAAGAAACCTTTTNNKCTGTCTGAATCAA	822
Codon52_B5	TATATCAACCCTAAGAAACCTTTTGAGCTGNNKAATCAACGAGTT	823
Codon53_B6	ATCAACCCTAAGAAACCTTTTGAGCTGTGNNKCAACGAGTTGTC	824
Codon64_B7	CAACGAGTTGTCCAGGATTTTATCGAGAACNNKCGAGACATTTTG	825
Codon65_B8	CGAGTTGTCCAGGATTTTATCGAGAACGCANNKGACATTTTGACT	826
Codon72_B9	GAGAACGCACGAGACATTTTGACTAAGGGTNNKCTCACTTTACAAG	827
Codon73_B10	AACGCACGAGACATTTTGACTAAGGGTGCNNKCTTTACAAGGAT	828
Codon78_B11	TTGACTAAGGGTCGCTCACTTTACAAGGATNNKCCCTACAAGGCT	829
Codon89_B12	CCCTACAAGGCTCATACCGATTTAGGGGACNNKTTGGTAATCCCG	830
Codon91_C1	AAGGCTCATACCGATTTAGGGGACGTCTTGNNKATCCCGCCTGAA	831
Codon92_C2	GCTCATACCGATTTAGGGGACGTCTTGGTANNKCCGCCTGAATTT	832
Codon98_C3	GACGTCTTGGTAATCCCGCCTGAATTTGCCNNKGCTCTAAAATCT	833
Codon114_C4	AGACAGTTAGATTTTACCGAAGTCGCAAGANNKGATACTCACGGT	834
Codon122_C5	GCAAGAGACGATACTCACGGTTATATTCCTNNKTCGAGCCAATC	835
Codon135_C6	CCAATCGGTTCCCCGTTTCGATCTGGTGCCGNNKGTTAACAAGTAT	836
Codon137_C7	GGTTCCCCGTTTCGATCTGGTGCCGCTCGTTNNKAAGTATCTTACA	837
Codon149_C8	TATCTTACAAGGGCTTTGGCAAAACTAACANNKCCATTGTGGGCC	838
Codon150_C9	CTTACAAGGGCTTTGGCAAAACTAACAAAANNKTTGTGGGCCGAA	839
Codon156_C10	AAACTAACAAAACCATTTGTGGGCCGAAGCTNNKTTAGGTGTAAAC	840
Codon158_C11	ACAAAACCATTTGTGGGCCGAAGCTTCTTANNKGTAACCATGTT	841
Codon173_C12	TTAGGCACTTCTACTGAGTGGCATCCTATTNNKCCAGGGGAAGAT	842
Codon187_D1	GATATCATGAGGATTGTCTCCAGAATGTCANNKAGAATATTCATG	843
Codon195_D2	ATGTCATCCAGAATATTCATGGGTGAGGAANNKTGTAAAGATGAT	844
Codon197_D3	TCCAGAATATTCATGGGTGAGGAACCTTTGTNNKGATGATGATTGG	845
Codon198_D4	AGAATATTCATGGGTGAGGAACCTTTGTAAANNKGATGATTGGTTG	846
Codon208_D5	GATGATGATTGGTTGAAAAGTTAGTATTGAGNNKACTGTGCAATTG	847
Codon212_D6	TTGAAAGTTAGTATTGAGTACACTGTGCAANNKTTCAAACCGCA	848
Codon222_D7	TTGTTTCAAACCGCAGACGAATTACGTAAACNNKCCACGTTGGACA	849
Codon225_D8	ACCGCAGACGAATTACGTAACTATCCACGTNNKACAAGACCATAT	850
Codon230_D9	CGTAACTATCCACGTTGGACAAGACCATATNNKCACTGGTTTTTG	851
Codon232_D10	TATCCACGTTGGACAAGACCATATATTCACNNKTTTTTGCCTTCC	852
Codon235_D11	TGGACAAGACCATATATTCACTGGTTTTTGTNNKCTCTGTCAAGGG	853
Codon239_D12	TATATTCACTGGTTTTTGCCTTCTGTCAANNKGTTAGGAGAAAA	854

Codon243_E1	TTTTTGCCCTTCCTGTCAAGGGGTTAGGAGANNKTTGCAGGAAGCG	855
Codon247_E2	TGTCAAGGGGTTAGGAGAAAATTGCAGGAANNKCGTGATTTATTG	856
Codon248_E3	CAAGGGGTTAGGAGAAAATTGCAGGAAGCGNKGATTTATTGCAA	857
Codon250_E4	GTTAGGAGAAAATTGCAGGAAGCGCGTGATNNKTTGCAACCCCAT	858
Codon254_E5	TTGCAGGAAGCGCGTGATTTATTGCAACCCNNKATTGATAGGAGA	859
Codon255_E6	CAGGAAGCGCGTGATTTATTGCAACCCCATNNKGATAGGAGAAAT	860
Codon263_E7	CCCCATATTGATAGGAGAAATGCTGTGAAGNNKGAAGCTATCGCT	861
Codon272_E8	AAGAAAGAAGCTATCGCTGAAGGTAGACCANNKCCATTTCGACGAT	862
Codon273_E9	AAAGAAGCTATCGCTGAAGGTAGACCATCANNKTTTCGACGATTCA	863
Codon282_E10	TCACCATTTCGACGATTCAATAGAATGGTTTNNKAATGAATACGAG	864
Codon286_E11	GATTCAATAGAATGGTTTGAAAATGAATACNNKGGCAAATCTGAT	865
Codon291_E12	TTTGAAAATGAATACGAGGGCAAATCTGATNNKGCCACTGAACAA	866
Codon296_F1	GAGGGCAAATCTGATCCCGCCACTGAACAANNKAAACTATCACTG	867
Codon312_F2	GCGATTTCACACAACAACACTGATTTACTGTCTNNKACAATGTTTAAT	868
Codon317_F3	ACTGATTTACTGTCTGAAACAATGTTTAATNNKGCTTTGCAACCA	869
Codon331_F4	CCAGAACTCCTTGGTCCACTACGTGAAGAGNNKGTTACGGTTTTTA	870
Codon335_F5	GGTCCACTACGTGAAGAGATAGTTACGGTTNNKTCACGGAAGGT	871
Codon337_F6	CTACGTGAAGAGATAGTTACGGTTTTATCCNNKGAAGGTCTAAAA	872
Codon355_F7	TTTTACAATTTGAAGTTGATGGATTTCGGTCNNKAAAAGAGTCACAG	873
Codon356_F8	TACAATTTGAAGTTGATGGATTTCGGTCATANNKGAGTCACAGAGA	874
Codon358_F9	TTGAAGTTGATGGATTTCGGTCATAAAAGAGNNKCAGAGACTTCGC	875
Codon361_F10	ATGGATTTCGGTCATAAAAGAGTCACAGAGANNKCGCCCTGTTTTTA	876
Codon378_F11	GCTTTCAGGAGAATGGCATTGGCTGATGTTNNKTTGCCAATGGC	877
Codon380_F12	AGGAGAATGGCATTGGCTGATGTTACCTTGNNKAATGGCGACGTA	878
Codon382_G1	ATGGCATTGGCTGATGTTACCTTGCCCAATNNKGACGTAATTAAA	879
Codon386_G2	GATGTTACCTTGCCCAATGGCGACGTAATTNNKAAAAGGTACCAAG	880
Codon396_G3	AAAAAAGGTACCAAGATCATTTGCGACACTNNKCATCAATGGAAT	881
Codon398_G4	GGTACCAAGATCATTTGCGACACTACACATNNKTGGAATCCAGAA	882
Codon402_G5	ATTTGCGACACTACACATCAATGGAATCCANNKTACTATCCAGAT	883
Codon410_G6	AATCCAGAATACTATCCAGATGCTAGTAAGNNKAATGCATATAGA	884
Codon413_G7	TACTATCCAGATGCTAGTAAGTTCAATGCANNKAGATTTTTTGCAA	885
Codon418_G8	AGTAAGTTCAATGCATATAGATTTTTTGCAANNKAGACAGACACCG	886
Codon421_G9	AATGCATATAGATTTTTTGCAATGAGACAGNNKCCGGGTCAGGAC	887
Codon428_G10	ATGAGACAGACACCGGGTCAGGACAAAAAGANNKACACCTTGTCAGC	888
Codon432_G11	CCGGGTCAGGACAAAAGAGCACACCTTGTCNNKACAAGCCATGAT	889
Codon433_G12	GGTCAGGACAAAAGAGCACACCTTGTCAGCANNKAGCCATGATCAA	890
Codon435_H1	GACAAAAGAGCACACCTTGTCAGCACAAAGCANNKGATCAAATGGGA	891
Codon438_H2	GCACACCTTGTCAGCACAAAGCCATGATCAANNKGGATTTCGGACAT	892
Codon455_H3	CACGCGTGCCCGAGGAAGATTTTTTCGCAGCCNNKGAGATTAAAGATT	893
Codon457_H4	TGCCCAGGAAGATTTTTTCGCAGCCAATGAGNNKAAGATTGCGCTG	894
Codon461_H5	TTTTTCGCAGCCAATGAGATTAAAGATTGCGNNKGTGCATATGCTA	895
Codon484_H6	GAAGGTGTTGTACCTAAGTCTAAGGCTTTANNKATGTCTTTACTT	896
Codon492_H7	GCTTTAGGAATGTCTTTACTTGGTGACCGGNNKGCCAACTGATG	897
Codon493_H8	TTAGGAATGTCTTTACTTGGTGACCGGGAANNKAAACTGATGGTG	898
Codon502_H9	GAAGCCAACTGATGGTGAAAAGGAGAGCANNKGAAATCGATATA	899
Codon505_H10	CTGATGGTGAAAAGGAGAGAGCCGAAATCANNKATAGACACTATT	900

Codon509_H11	AGGAGAGCAGCCGAAATCGATATAGACACTNNKGGTAGTGATGAA	901
Codon511_H12	GCAGCCGAAATCGATATAGACACTATTGGTNNKGATGAATAAATT	902
Codon2_A1	ATACTTTAACGTCAAGGAGAAAAAACCATGNNKACCGATCTAGAC	903
Codon5_A2	CGTCAAGGAGAAAAAACCATGGCCACCGATNNKGACCTAGTATTA	904
Codon12_A3	GCCACCGATCTAGACCTAGTATTAGGAAAAANNKCAATACGCATTA	905
Codon20_A4	GGAAAAAGTCAATACGCATTATTTTGTGGCANNKACTTTATTTAGC	906
Codon21_A5	AAAAGTCAATACGCATTATTTTGTGGCATANNKTATTTAGCTTTT	907
Codon27_A6	TTTTGTGGCATAACTTTATTTAGCTTTTTTCNNKCTAAAGTATAGT	908
Codon29_A7	GGCATAACTTTATTTAGCTTTTTTCATACTANNKTATAGTCTTCTC	909
Codon40_A8	TATAGTCTTCTCGGAAACGGGGCAAGCAANNKCCTTATATCAAC	910
Codon41_A9	AGTCTTCTCGGAAACGGGGGCAAGCAATATNNKTATATCAACCCT	911
Codon58_A10	CCTTTTGAGCTGTGCAATCAACGAGTTGTCNNKGATTTTATCGAG	912
Codon63_A11	AATCAACGAGTTGTCCAGGATTTTATCGAGNNKGCACGAGACATT	913
Codon70_A12	TTTATCGAGAACGCACGAGACATTTTGACTNNKGGTCGCTCACTT	914
Codon75_B1	CGAGACATTTTGACTAAGGGTCGCTCACTTNNKAAGGATACGCCC	915
Codon77_B2	ATTTTGACTAAGGGTCGCTCACTTTACAAGNNKACGCCCTACAAG	916
Codon79_B3	ACTAAGGGTCGCTCACTTTACAAGGATACGNNKTACAAGGCTCAT	917
Codon82_B4	CGCTCACTTTACAAGGATACGCCCTACAAGNNKCATACCGATTTA	918
Codon85_B5	TACAAGGATACGCCCTACAAGGCTCATACCNNKTAGGGGACGTC	919
Codon87_B6	GATACGCCCTACAAGGCTCATACCGATTTANNKGACGTCTTGGTA	920
Codon90_B7	TACAAGGCTCATACCGATTTAGGGGACGTCNNKGTAATCCCGCCT	921
Codon97_B8	GGGGACGTCTTGGTAATCCCGCCTGAATTTNNKGACGCTCTAAAA	922
Codon101_B9	GTAATCCCGCCTGAATTTGCCGACGCTCTANNKTCTGAAAGACAG	923
Codon134_B10	GAGCCAATCGGTTCCCCGTTTCGATCTGGTGNNKCTCGTTAACAAG	924
Codon144_B11	CCGCTCGTTAACAAGTATCTTACAAGGGCTNNKGCAAAACCTAACA	925
Codon145_B12	CTCGTTAACAAGTATCTTACAAGGGCTTTGNNKAAACTAACAAAA	926
Codon147_C1	AACAAGTATCTTACAAGGGCTTTGGCAAAAANNKACAAAACCATTG	927
Codon157_C2	CTAACAAAACCATTTGTGGGCCGAAGCTTCTNNKGGTGTAACCAT	928
Codon161_C3	TTGTGGGCCGAAGCTTCTTTAGGTGTAAACNNKGTTTTAGGCACT	929
Codon164_C4	GAAGCTTCTTTAGGTGTAAACCATGTTTTANNKACTTCTACTGAG	930
Codon167_C5	TTAGGTGTAAACCATGTTTTAGGCACTTCTNNKGAGTGGCATCCT	931
Codon177_C6	ACTGAGTGGCATCCTATTAACCCAGGGGAANNKATCATGAGGATT	932
Codon178_C7	GAGTGGCATCCTATTAACCCAGGGGAAGATNNKATGAGGATTGTC	933
Codon180_C8	CATCCTATTAACCCAGGGGAAGATATCATGNNKATTGTCTCCAGA	934
Codon182_C9	ATTAACCCAGGGGAAGATATCATGAGGATTNNKTCCAGAATGTCA	935
Codon183_C10	AACCCAGGGGAAGATATCATGAGGATTGTCNNKAGAAATGTCATCC	936
Codon185_C11	GGGGAAGATATCATGAGGATTGTCTCCAGANNKTCATCCAGAATA	937
Codon191_C12	ATTGTCTCCAGAATGTCATCCAGAATATTCNNKGGTGAGGAACTT	938
Codon196_D1	TCATCCAGAATATTCATGGGTGAGGAACTTNNKAAAGATGATGAT	939
Codon201_D2	ATGGGTGAGGAACCTTTGTAAAGATGATGATNNKTGAAAGTTAGT	940
Codon215_D3	AGTATTGAGTACACTGTGCAATTGTTTCAANNKGCAGACGAATTA	941
Codon216_D4	ATTGAGTACACTGTGCAATTGTTTCAAACCNNKGACGAATTACGT	942
Codon224_D5	CAAACCGCAGACGAATTACGTAACTATCCANNKTGGACAAGACCA	943
Codon236_D6	ACAAGACCATATATTCCTGTTTTCCTNNKTGTCAAGGGGTTT	944
Codon268_D7	AGAAATGCTGTGAAGAAAGAAGCTATCGCTNNKGGTAGACCATCA	945
Codon270_D8	GCTGTGAAGAAAGAAGCTATCGCTGAAGGTNNKCCATCACCATTTC	946

Codon276_D9	ATCGCTGAAGGTAGACCATCACCATTTCGACNNKTCAATAGAATGG	947
Codon280_D10	AGACCATCACCATTTCGACGATTCAATAGAAANNKTTTGAAAATGAA	948
Codon281_D11	CCATCACCATTTCGACGATTCAATAGAAATGGNNKGAAAAATGAATAC	949
Codon284_D12	TTCGACGATTCAATAGAAATGGTTTGAAAATNNKTACGAGGGCAAA	950
Codon290_E1	TGGTTTGAAAATGAATACGAGGGCAAATCTNNKCCCGCCACTGAA	951
Codon293_E2	AATGAATACGAGGGCAAATCTGATCCCGCCNNKGAAACAAATTAAA	952
Codon309_E3	TCACTGGTGGCGATTACACACAACACTGATNNKCTGTCTGAAACA	953
Codon311_E4	GTGGCGATTACACACAACACTGATTTACTGNNKGAAACAAATGTTT	954
Codon321_E5	TCTGAAACAATGTTTAAATATAGCTTTGCAANNKGAACTCCTTGGT	955
Codon338_E6	CGTGAAGAGATAGTTACGGTTTTATCCACGNNKGGTCTAAAAAAG	956
Codon342_E7	GTTACGGTTTTATCCACGGAAGGTCTAAAAANNKACGTCGTTTTTAC	957
Codon346_E8	TCCACGGAAGGTCTAAAAAAGACGTCGTTTNNKAATTTGAAGTTG	958
Codon354_E9	TCGTTTTACAATTTGAAGTTGATGGATTCGNNKATAAAAGAGTCA	959
Codon369_E10	CAGAGACTTCGCCCTGTTTTATTAGGTGCTNNKAGGAGAATGGCA	960
Codon385_E11	GCTGATGTTACCTTGCCCAATGGCGACGTANNKAAAAAAGGTACC	961
Codon393_E12	GACGTAATTAAAAAAGGTACCAAGATCATNNKGACACTACACAT	962
Codon404_F1	GACACTACACATCAATGGAATCCAGAATACNNKCCAGATGCTAGT	963
Codon409_F2	TGGAATCCAGAATACTATCCAGATGCTAGTNNKTTCAATGCATAT	964
Codon414_F3	TATCCAGATGCTAGTAAGTTCAATGCATATNNKTTTTTGCAAATG	965
Codon416_F4	GATGCTAGTAAGTTCAATGCATATAGATTTNNKCAAATGAGACAG	966
Codon417_F5	GCTAGTAAGTTCAATGCATATAGATTTTGGNNKATGAGACAGACA	967
Codon419_F6	AAGTTCAATGCATATAGATTTTTCGAAATGNNKCAGACACCGGGT	968
Codon430_F7	CAGACACCGGGTCAGGACAAAAGAGCACACNNKGTGAGCACAAGC	969
Codon431_F8	ACACCGGGTCAGGACAAAAGAGCACACCTTNNKAGCACAAAGCCAT	970
Codon434_F9	CAGGACAAAAGAGCACACCTTGTGAGCACANNKCATGATCAAATG	971
Codon437_F10	AGAGCACACCTTGTGAGCACAAAGCCATGATNNKATGGGATTCGGA	972
Codon442_F11	AGCACAAGCCATGATCAAATGGGATTCGGANNKGGCTTGACGCG	973
Codon451_F12	GGACATGGCTTGACGCGTGCCAGGAAGANNKTTTCGCAGCCAAT	974
Codon454_G1	TTGCACGCGTGCCAGGAAGATTTTTCGCANNKAATGAGATTAAG	975
Codon459_G2	GGAAGATTTTTTCGCAGCCAATGAGATTAAGNNKGGCTGTGTCAT	976
Codon465_G3	AATGAGATTAAGATTGCGCTGTGTCATATGNNKTTGAAATATGAT	977
Codon473_G4	CATATGCTATTGAAATATGATTGGAAATTANNKGAAGGTGTTGTA	978
Codon476_G5	TTGAAATATGATTGGAAATTACCAGAAGGTNNKGTACCTAAGTCT	979
Codon489_G6	AAGTCTAAGGCTTTAGGAATGTCTTTACTTNNKGACCGGGAAGCC	980
Codon494_G7	GGAATGTCTTTACTTGGTGACCGGGAAGCCNNKCTGATGGTGAAA	981
Codon496_G8	TCTTTACTTGGTGACCGGGAAGCCAAACTGNNKGTGAAAAGGAGA	982
Codon499_G9	GGTGACCGGGAAGCCAAACTGATGGTGAAANNKAGAGCAGCCGAA	983
Codon513_G10	GAAATCGATATAGACACTATTGGTAGTGATNNKTAAATTGAATTG	984

[0150] The two fragments A and B were assembled by overlap extension PCR using forward primer P35 (SEQ ID NO: 96) and reverse primer of P36 (SEQ ID NO: 97). The assembled OE-PCR products were then pooled together, and gel purified to provide a saturation mutagenesis library of linear donor DNA. The pooled saturation mutagenesis library linear donor DNA was integrated as a knock-in using CRISPR-Cas9 into the X-4 locus where URA3 was integrated in SHEVP003 which already had the CPR gene integrated.

[0151] B. Screening of SSM library

[0152] Screening of the recombinant host cells for bioconversion of 3-KCA to 3-KUDCA was carried out according to a protocol adapted from that described in Example 1. Individual colonies of recombinant *S. cerevisiae* strains were picked into 96-well plates containing 2 x YPD media (300 μ L per well) using a QPix™ 420 colony picking system. The plates were incubated at 30 °C for 24 h with shaking at 250 rpm at 85 % humidity. After this time, the strains were sub-cultured (40 μ L inoculation volume per strain) into a second 96-well plate containing 2 x YPD media (1.2 mL per well) using an Agilent Bravo automated liquid handling platform. The plates were incubated at 30 °C for 48 h with shaking at 250 rpm at 85 % humidity. After this time, a solution of galactose (40 % stock solution) was added using the Bravo (1 % final concentration) and the plates were further incubated for 4 hours to induce protein production. After this time, the biomass was harvested via centrifugation and the supernatant removed. The resulting biomass was resuspended in 300 μ L of bioconversion buffer (295 mg/L 3-KCA, 0.1 M phosphate buffer, pH 9) using the Agilent Bravo automated liquid handling platform and the plates were incubated at 30 °C for 24 h with shaking at 250 rpm at 85 % humidity. After this time, MeOH (300 μ L) was added using the Bravo and the plates were incubated at 30 °C with shaking at 250 rpm at 85 % humidity for a further 30 minutes. The biomass was then separated by centrifugation and the resulting supernatant diluted with MeOH (1000 x total dilution) using the Agilent Bravo automated liquid handling platform before analysis using an Agilent 6470 Triple Quadrupole LC/MS. Sample preparation and analytical methods were identical to those described in Example 1.

[0153] Results

[0154] Data from screening the SSM libraries in terms of fold-improvement in production of 3-KUDCA from 3-KCA relative to the control strain, SH013 are summarized in Table 7 (below).

[0155] TABLE 7

Name	NT SEQ ID NO:	AA SEQ ID NO:	AA Differences	3-KCA to 3- KUDCA % Conversion	FIOC 3-KCA to 3- KUDCA % Conversion
CYP100	5	6	S31V	52	2.85
CYP102	7	8	A143R	40	1.56
	9	10	L147W	38	1.96
	11	12	A143G	28	1.62
CYP101	13	14	P401D	23	2.40
	15	16	G175E	17	1.64
	17	18	Q54K	16	0.63
	19	20	T307A	16	0.70
	21	22	L147R	15	1.18
	23	24	V56G	13	1.32
	25	26	K76H	13	2.42
	27	28	I120A	13	1.25

	29	30	G325E	11	0.40
--	----	----	-------	----	------

Example 3: Further Optimization of Mutant CYP Genes in Recombinant *S. cerevisiae* Host Cells by Combinatorial Mutagenesis

[0156] This example illustrates the preparation of further combinatorial libraries of CYP genes based on the parent strain SHS036 which carries a gene encoding the S31V mutant of CYP (SEQ ID NO: 6) identified through SSM in Example 2 for improved bioconversion of 3-KCA to 3-KUDCA.

[0157] Material and Methods

[0158] A. Combinatorial library construction

[0159] The polynucleotide sequence of SEQ ID NO: 5 encoding the S31V mutant CYP polypeptide (SEQ ID NO: 6) from SHS036 expressed under the pGal1 promoter within the pGal1/10 promoter system (SEQ ID NO: 59) and PGK1t terminator (SEQ ID NO: 61), was used to design oligos to generate combinatorial libraries to randomly incorporate additional identified beneficial amino acid mutations from Example 2 into the S31V CYP backbone sequence. The same screening strain was used to integrate the resulting libraries as in Example 2 (SHEVP003), however the SHS036 strain was used as the parent control strain in order to determine fold-improvement in conversion of 3-KCA to 3-KUDCA as described below.

[0160] A semi-synthetic approach was used to construct the first set of combinatorial libraries. Genomic DNA from the SHS036 strain was used as the template to generate a full-length PCR product using primer pair of P34 (SEQ ID NO: 95) and P37 (SEQ ID NO: 98) while incorporating uracil using a dNTP mix comprising of the following deoxyribonucleotides: dATP, dGTP, dCTP, dTTP, dUTP. The resulting PCR product was gel purified and digested with Uracil-DNA Glycosylase and Endonuclease IV at 37 C for 2 hours, followed by enzyme denaturation at 94 C for two minutes, to generate a pool of fragments in the range of 50-100 bases. These fragments were further combined with differing ratios of pools of the synthesized oligonucleotide primers (each oligo up to 55 bases in length and encoding one or more amino acid change) in several individual assembly PCR reactions using forward primer P5 (SEQ ID NO: 66) and reverse primer P6 (SEQ ID NO: 67) to reassemble the full-length PCR product as Fragment B and incorporate mutagenic amino acid changes within each pool randomly. The synthesized oligonucleotide primers SEQ ID NOs: 985-1001 used in the pools and their encoded mutations are listed in Table 8 below and the accompanying Sequence Listing.

[0161] TABLE 8

Name	Mutation (s)	Length	Primer Sequence	SEQ ID NO:
CYP_CMB1_p1	Q54K-V56G	49	AACCTTTTGAGCTGTCGAATAAACGAGGAGTC CAGGATTTTATCGAGAA	985
CYP_CMB1_p2	Q54K	49	AACCTTTTGAGCTGTCGAATAAACGAGTTGTC CAGGATTTTATCGAGAA	986

CYP_CMB1_p3	V56G	49	AACCTTTTGTGAGCTGTCTGAATCAACGAGGAGTC CAGGATTTTATCGAGAA	987
CYP_CMB1_p4	K76H	43	CTAAGGGTCTGCTCACTTTACCATGATACGCCC TACAAGGCTCA	988
CYP_CMB1_p5	I120A	43	GAGACGATACTCACGGTTATGCCCCCTGGATTTC GAGCCAATCGG	989
CYP_CMB1_p6	A143G-L147R	55	TTAACAAGTATCTTACAAGGGGATTGGCAAAA CGAACAAAACCATTTGTGGGCCGA	990
CYP_CMB1_p7	A143R-L147R	55	TTAACAAGTATCTTACAAGGCGATTGGCAAAA CGAACAAAACCATTTGTGGGCCGA	991
CYP_CMB1_p8	A143G-L147W	55	TTAACAAGTATCTTACAAGGGGATTGGCAAAA TGGACAAAACCATTTGTGGGCCGA	992
CYP_CMB1_p9	A143R-L147W	55	TTAACAAGTATCTTACAAGGCGATTGGCAAAA TGGACAAAACCATTTGTGGGCCGA	993
CYP_CMB1_p10	A143G	55	TTAACAAGTATCTTACAAGGGGATTGGCAAAA CTAACAAAACCATTTGTGGGCCGA	994
CYP_CMB1_p11	A143R	55	TTAACAAGTATCTTACAAGGCGATTGGCAAAA CTAACAAAACCATTTGTGGGCCGA	995
CYP_CMB1_p12	L147R	55	TTAACAAGTATCTTACAAGGGCTTTGGCAAAA CGAACAAAACCATTTGTGGGCCGA	996
CYP_CMB1_p13	L147W	55	TTAACAAGTATCTTACAAGGGCTTTGGCAAAA TGGACAAAACCATTTGTGGGCCGA	997
CYP_CMB1_p14	G175E	43	AGTGGCATCCTATTAACCCAGAGGAAGATATC ATGAGGATTGT	998
CYP_CMB1_p15	T307A	43	TGGTGGCGATTTCACACAACAGCCGATTTACTG TCTGAAACAAT	999
CYP_CMB1_p16	G325E	43	CTTTGCAACCAGAACTCCTTGAGCCACTACGT GAAGAGATAGT	1000
CYP_CMB1_p17	P401D	43	ACACTACACATCAATGGAATGATGAATACTAT CCAGATGCTAG	1001

[0162] For efficient integration of the combinatorial mutant CYP variants, PCR amplicons with homologous 5' and 3' regions to the CYP were generated (Fragment A and Fragment C respectively). Fragment A was amplified using the forward primer P37 (SEQ ID NO: 98) and reverse primer P4 (SEQ ID NO: 65) and Fragment C was amplified using the forward primer P7 (SEQ ID NO: 68) and the reverse primer P38 (SEQ ID NO: 99). Fragments A, B, and C were assembled by NEBuilder[®] HiFi Assembly and amplified as a single linear DONOR using the following forward rescue primer P35 (SEQ ID NO: 96) and reverse primer P36 (SEQ ID NO: 97). The assembled PCR products were then pooled together, and gel purified to provide a combinatorial library of linear donor DNA.

[0163] The pooled combinatorial libraries of linear donor DNA were integrated as a knock-in using CRISPR-Cas9 into the X-4 locus where URA3 was integrated in SHEVP003 which already had the CPR gene integrated as in Examples 1 and 2.

[0164] B. Screening of Combi libraries

[0165] Screening of the recombinant host cells for bioconversion of 3-KCA to 3-KUDCA was carried out using the same protocol described in Example 2 except the 3-KCA loading was increased from 295 mg/L to 590 mg/L.

[0166] **Results**

[0167] Screening of the combinatorial libraries for fold-improvement in production of 3-KUDCA from 3-KCA relative to the control strain, SHS036 which expresses the S31V CYP polypeptide of SEQ ID NO: 6 (also referred to herein as “CYP100”) and the CPR polypeptide of SEQ ID NO: 4 (also referred to herein as “CPR001”), are summarized in Table 9 (below).

[0168] **TABLE 9**

Name	NT SEQ ID NO:	AA SEQ ID NO:	AA Differences [Silent codon mutations]	3-KCA to 3- KUDCA % Conversion ¹	FIOC 3-KCA to 3- KUDCA % Conversion
CYP105	31	32	S31V, Q54K, A143G, T307A, G325E	81	1.9
CYP109	33	34	S31V, L147R, G325E [V476 (GTT > GTC)]	68	1.7
	35	36	S31V, Q54K, A143R, L147R, T307A, G325E, P401D [V476 (GTT > GTC)]	67	1.6
	37	38	S31V, V57F, L147W, P401D [A368 (GCT > GCC)]	62	4.4
CYP107	39	40	S31V, L147W, P401D [A368 (GCT > GCC)]	59	4.5
	41	42	S31V, Q54K, A143R, T307A, P401D	54	4.4
CYP108	43	44	S31V, Q54K, P401D	53	3.6
CYP106	45	46	S31V, A143R, L147W	43	3.3
	47	48	S31V, A143G, G175E, P401D, K498R	29	2.4
	49	50	S31V, A143G, L147C, G175E, P401D	26	1.8
	51	52	S31V, P129S, A143R	18	1.9
	53	54	S31V, G175E, P401D [R328 (CGT > CGC)]	14	1.0
¹ 3-KCA to 3-KUDCA % conversion is calculated as {[3-KUDCA] / ([3-KCA] + [3-KUDCA])} x 100					

Example 4: Genomic Integration of Heterologous CYP/CPR Gene Pairs into *Pichia pastoris* Host Cells

[0169] This example illustrates the preparation of recombinant yeast host cell strains (*Pichia pastoris*) with genomically integrated copies of a pair of heterologous genes encoding a

cytochrome P450 (CYP) and cytochrome reductase (CPR) enzymes from *Gibberella zeae* and demonstration of the ability of these strains to carry out the bioconversion of Lithocholic acid (LCA) to Ursodeoxycholic acid (UDCA) and/or 3-Keto-lithocholic acid (3-KCA) to 3-Keto-ursodeoxycholic acid (3-KUDCA).

[0170] A recombinant strain of *P. pastoris*, denoted as “SAND121” has been described previously in WO2022115710A1. SAND121 expresses a pair of genes encoding a cytochrome P450 (CYP) and a reductase (CPR) from *Gibberella zeae* and has been shown to carry out the conversion of LCA to UDCA and 3-KCA to 3-KUDCA.

[0171] Material and Methods

[0172] *A. P. pastoris* strain building

[0173] To improve upon the bioconversion of LCA to UDCA and 3-KCA to 3-KUDCA observed in the SAND121 strain, yeast codon-optimized polynucleotide sequences of SEQ ID NO: 1 and 3 encoding the CYP polypeptide of SEQ ID NO: 2 and the CPR polypeptide of SEQ ID NO: 4, respectively, were integrated into the *P. pastoris* genome under the bidirectional pCAT1:pFDH1 promoter system (SEQ ID NO: 1002) and tDAS1 and tDAS2 terminators SEQ ID NO: 1003 and SEQ ID NO: 1004, respectively, to generate a single copy strain (SHP026) as follows. A single plasmid (wbplasmid147; SEQ ID NO: 1059) containing a zeocin resistance marker, a pUC origin of replication, a HIS4 homologous sequence (SEQ ID NO: 1005), and a cassette comprising of tDAS1:CPR001:pCAT1:pFDH1:CYP001:tDAS2 was assembled. Sequences were amplified using PCR using the following primers: wboligos5777, wboligos5792, wboligos5793, wboligos5796, wboligos5797, wboligos5798, wboligos5800, wboligos5801, wboligos5802, wboligos5803, wboligos5793, wboligos5794, wboligos5795, and wboligos5776.

[0174] The sequences of the PCR primers described in the following *P. pastoris* strain building examples by their Primer Name are listed in Table 10 below and the accompanying Sequence Listing.

[0175] TABLE 10

Primer Name	Primer Sequence	SEQ ID NO:
wboligos5776	CTGATTCTTTCTGCTGAACTGTGGCAAGTTATTTAAATGAATAATAAC TGTGTATTTTTC	1006
wboligos5777	GAAAAATACACAGTTATTATTCATTTAAATAAAGTGGCACAGTTCAGC AGAAAGAATCAG	1007
wboligos5792	GGGACTCCCAAAGTGTCAGTCACAAGGGTCACTTGAAGTCGGACAGTG AGTGTAAGTCTTG	1008
wboligos5793	CAAGACTACACTCACTGTCCGACTTCAAGTGACCCTTGTGACTGACAC TTTGGGAGTCCC	1009
wboligos5794	ACAAAAAGAAACAAGACATTACTGAAGGATCGACGGGGTTCGTAAACTG GTTCTCCTTTC	1010
wboligos5795	GAAAGGAGGAACCAAGTTTACGAACCCCGTCGATCCTTCAGTAATGTCT TGTTTCTTTTGT	1011

wboligos5796	TTCGAAAAGTCTCAAATCCATACTTTTTAAACGGGAAGTCTTTACAGT TTTAGTTAGGAG	1012
wboligos5797	CTCCTAACTAAAACTGTAAAGACTTCCCCTTTAAAAAGTATGGATTG AGACTTTTCGAA	1013
wboligos5798	CACTTGCTCTAGTCAAGACTTACAATTAAAATGGCACTACGAACATCA CTCTCTCGCCCC	1014
wboligos5800	TCCTAATACTAGGTCTAGATCGGTGGCCATTGTTTAAAGTGGGTGATGT TGGAGGTATTTG	1015
wboligos5801	CAAATACCTCCAACATCACCCACTTAAACAATGGCCACCGATCTAGAC CTAGTATTAGGA	1016
wboligos5802	AACTACTAACCCGTTAGTGGCCAAATCTACTTATTCATCACTACCAAT AGTGTCTATATC	1017
wboligos5803	GATATAGACACTATTGGTAGTGATGAATAAGTAGATTTGGCCACTAAC GGGTTAGTAGTT	1018
wboligos5921	TCCTAATACTAGGTCTAGATCGGTGGCCATTTTCAATAATTAGTTGTT TTTTGATCTTCT	1019
wboligos5922	AGAAGATCAAAAAACAATAATTATTGAAAATGGCCACCGATCTAGAC CTAGTATTAGGA	1020
wboligos5978	CCGGGTGACCCGGCGGGGACGAGGCAAGCTGACGGGGTTTCGTAACTG GTTCTCCTTTC	1021
wboligos5979	GAAAGGAGGAACCAGTTTACGAACCCCGTCAGCTTGCCTCGTCCCCGC CGGGTCACCCGG	1022
wboligos5983	AATATCGGATACGCGGCCCAAGGCTTGCTTCCAGGTGGCACTTTTCGG GGAAATGTGCGC	1023
wboligos5984	GCGCACATTTCCCCGAAAAGTGCCACCTGGAAGCAAGCCTTGGGCCGC GTATCCGATATT	1024
wboligos5985	CAGACACCAGGTCATTGAACTAACGACAAGGTTGGCCCTTTTATCAAC TTTAAAGAGTGC	1025
wboligos5986	GCACTCTTAAAAGTTGATAAAAGGGCCAACCTTGTCGTTAGTTCAATG ACCTGGTGTCTG	1026
wboligos5987	TTCGAAAAGTCTCAAATCCATACTTTTTAAGTACGTAGTTTCGCTTAG TTTAAGACTAAA	1027
wboligos5988	TTTAGTCTTAACTAAGCGAACTACGTACTTAAAAAGTATGGATTG AGACTTTTCGAA	1028
wboligos5989	CATAACTCATTCATTCATCATACGTCGTACTTATTCATCACTACCAAT AGTGTCTATATC	1029
wboligos5990	GATATAGACACTATTGGTAGTGATGAATAAGTACGACGTATGATGAAT GAATGAGTTATG	1030
wboligos5991	GGCGGGGACGAGGCAAGCTCAATTCTCGACTTTCACAGAGCGTCGATC AATACCGCGAAG	1031
wboligos5992	CTTCGCGGTATTGATCGACGCTCTGTGAAAGTCGAGAATTGAGCTTGC CTCGTCCCCGCC	1032
wboligos5993	GCGCACATTTCCCCGAAAAGTGCCACCTGGAGCCAATTTTAGCAAGCC CAATCCAGTTGA	1033
wboligos5994	TCAACTGGATTGGGCTTGCTAAAATTGGCTCCAGGTGGCACTTTTCGG GGAAATGTGCGC	1034

wboligos5995	ACCATGTTTCGTTGGAAGGGAGATGCTGAAAGAGATTGGCTATCCTT AATTGCTCCTTC	1035
wboligos5996	GAAGGAGCAATTAAGGATAGCCAAATCTCTTTCAGCATCTCCCTTCCA ACGAAACATGGT	1036
wboligos5997	ATAGACTTAAAGAATAAAGAGAAATTTTATTTAAAAAGTATGGATTG AGACTTTTCGAA	1037
wboligos5998	TTCGAAAAGTCTCAAATCCATACTTTTTAAATAAAATTTCTCTTTATT CTTTAAGTCTAT	1038
wboligos5999	GATATAGACACTATTGGTAGTGATGAATAAGCCGAATAGTTTGTATAC GTCTTATGTAAT	1039
wboligos6000	ATTACATAAGACGTATACAACTATTCGGCTTATTCATCACTACCAAT AGTGTCTATATC	1040
wboligos6001	GGCGGGGACGAGGCAAGCTCAATTCTCGACTCTTATCTGCCAGAGTCG CTAGATCTGGAA	1041
wboligos6002	TTCCAGATCTAGCGACTCTGGCAGATAAGAGTCGAGAATTGAGCTTGC CTCGTCCCCGCC	1042
wboligos6013	GAAAAATACACAGTTATTATTCAATTTAAATTTACCAATGCTTAATCAG TGAGGCACCTAT	1043
wboligos6014	ATAGGTGCCTCACTGATTAAGCATTGGTAAATTTAAATGAATAATAAC TGTGTATTTTTC	1044
wboligos6015	GTTTCATTCAACCTTTCGTCTTTGGATGTTCCAGGTGGCACTTTTCGG GGAAATGTGCGC	1045
wboligos6016	GCGCACATTTCCCCGAAAAGTGCCACCTGGAACATCCAAGACGAAAG GTTGAATGAAAC	1046
wboligos6089	CACTTGCTCTAGTCAAGACTTACAATTTAAATGGCCACCGATCTAGAC CTAGTATTAGGA	1047
wboligos6090	TCCTAATACTAGGTCTAGATCGGTGGCCATTTTAATTGTAAGTCTTGA CTAGAGCAAGTG	1048
wboligos6091	GATATAGACACTATTGGTAGTGATGAATAAGTACGTAGTTTCGCTTAG TTTAAGACTAAA	1049
wboligos6092	TTTAGTCTTAAACTAAGCGAACTACGTACTTATTCATCACTACCAAT AGTGTCTATATC	1050

[0176] PCR amplification was followed by agarose-gel purification and assembly using NEB's HiFi DNA Assembly Master Mix (catalog no. E2621X) according to manufacturer's instructions. Since the PCR amplicons had overlapping ends, this allowed for whole plasmid assembly. Three microliters of the HiFi assembly were used to transform *E. coli* competent cells and plated on low-salt LB media containing zeocin. Plasmids were extracted from cultures inoculated by colonies using Thermo-Fisher GeneJet Plasmid Miniprep Kit (catalog no. K0503) and sequenced to confirm correct plasmid assembly. One sequence-confirmed plasmid was digested with BamHI and further purified using Zymo Research's DNA Clean and Concentrator Kit (catalog no. D4014). The restriction enzyme cut site found in the HIS4 homologous sequence, linearized the plasmid, allowing for homologous integration at the HIS4 locus in *Pichia* competent cells, which were prepared from the wild-type BG-10 strain using Thermo-

Fisher Pichia EasyComp Transformation Kit (catalog no. K173001) according to manufacturer's instructions. 5 microliters of linearized plasmid were used to transform 50 μ L of BG-10 competent cells and plated on YPDS media plates containing zeocin. After 3 days of growth, individual colonies were confirmed for integration of the linearized plasmid using PCR and genomic DNA isolated from individual colonies.

[0177] A recombinant *Pichia* host cell strain (SHP029) with two copies of a gene encoding CYP (SEQ ID NO: 2) and one copy of a gene encoding CPR (SEQ ID NO: 4) was also constructed as follows. A single plasmid (wbplasmid171; SEQ ID NO: 1061) containing an ampicillin resistance marker, a G418 resistance marker, a pUC origin of replication, promoter pAOX1 (SEQ ID NO: 1051), CYP (SEQ ID NO: 2), and terminator tDAS2 (SEQ ID NO: 1004) was assembled using the method described above. Sequences were amplified using PCR with the following primers: wboligos5979, wboligos6014, wboligos6013, wboligos6015, wboligos6016, wboligos5921, wboligos5922, wboligos5802, wboligos5803, and wboligos5978. Following plasmid recovery and sequence confirmation as described above, a single plasmid was digested with SacI and further purified using Zymo Research's DNA Clean and Concentrator Kit (catalog no. D4014). The restriction enzyme cut site, found in pAOX1, linearized the plasmid, allowing for homologous integration at the AOX1 locus. Five microliters of linearized plasmid were used to transform 50 μ L of SHP026 competent cells and plated on YPD media plates containing G418. After 3 days of growth, individual colonies were confirmed for integration of the linearized plasmid using PCR and genomic DNA from each individual colony.

[0178] A recombinant *Pichia* host cell strain (SHP030) with one copy of a gene encoding the S31V mutant of CYP (SEQ ID NO: 6) and one copy of the gene encoding CPR (SEQ ID NO: 4) was also constructed as follows. A single plasmid (wbplasmid160; SEQ ID NO: 1060) containing a zeocin resistance marker, a pUC origin of replication, a HIS4 homologous sequence (SEQ ID NO: 1005), and a cassette comprising tDAS1:CPR:pCAT1:pFDH1:CYP100:tDAS2 was assembled as described above. As wbplasmid147 shares the same sequence as wbplasmid160 except for the CYP CDS, inverse PCR was used to amplify the vector excluding the CYP CDS using primers wboligos5800 and wboligos5803. CYP (SEQ ID NO: 2) was amplified from SHS036 gDNA using primers wboligos5801 and wboligos5802. Following plasmid recovery and sequence confirmation, a single plasmid was digested with BamHI and further purified using Zymo Research's DNA Clean and Concentrator Kit (catalog no. D4014). The restriction enzyme cut site, found in the HIS4 homologous sequence, linearized the plasmid, allowing for homologous integration at the HIS4 locus. Five microliters of linearized plasmid were used to transform 50 μ L of BG-10 competent cells and plated on YPDS media plates containing zeocin. After 3 days of growth, individual colonies were confirmed for integration of the linearized plasmid using PCR and genomic DNA from each individual colony.

[0179] A recombinant *Pichia* host cell strain (SHP033) with two copies of the gene encoding CYP (SEQ ID NO: 2) and one copy of a gene encoding CPR (SEQ ID NO: 4) was also constructed as follows. A single plasmid (wbplasmid172; SEQ ID NO: 1062) containing an ampicillin resistance marker, a G418 resistance marker, a pUC origin of replication, promoter pAOX1 (SEQ ID NO: 1051), gene encoding CYP100 (SEQ ID NO: 6), and terminator tDAS2 (SEQ ID NO: 1004) was assembled with the method described above. As wbplasmid171 shares the same sequence as wbplasmid172 except for the CYP CDS, inverse PCR was used to amplify the vector excluding the CYP CDS using primers wboligos5921 and wboligos5803. CYP was amplified from SHS036 gDNA using primers wboligos5922 and wboligos5802. Following plasmid recovery and sequence confirmation as described above, a single plasmid was digested with SacI and further purified using Zymo Research's DNA Clean and Concentrator Kit (catalog no. D4014). The restriction enzyme cut site, found in pAOX1, linearized the plasmid, allowing for homologous integration at the AOX1 locus. Five microliters of linearized plasmid were used to transform 50 μ L of SHP030 competent cells and plated on YPD media plates containing G418. After 3 days of growth, individual colonies were confirmed for integration of the linearized plasmid using PCR and genomic DNA from each individual colony.

[0180] A recombinant *Pichia* host cell strain (SHP034) with three copies of a gene encoding the S31V CYP (SEQ ID NO: 6) and two copies of the gene encoding CPR (SEQ ID NO: 4) was also constructed as follows. A single plasmid (wbplasmid173; SEQ ID NO: 1063) containing a hygromycin resistance marker, an ampicillin resistance marker, a pUC origin of replication, an Int6 homologous sequence (SEQ ID NO: 1052), and a cassette comprising terminator tPMP20 (SEQ ID NO: 1053), CPR (SEQ ID NO: 4), bidirectional promoter pCAT1:FDH1 (SEQ ID NO: 1054), CYP100 (SEQ ID NO: 6), and terminator tFLD1 (SEQ ID NO: 1055), was assembled as described above. Sequences were amplified using PCR using the following primers: wboligos5983, wboligos5984, wboligos5985, wboligos5986, wboligos5987, wboligos5988, wboligos5989, wboligos5990, wboligos5991, and wboligos5992. Following plasmid recovery and sequence confirmation as described above, a single plasmid was digested with BsgI and further purified using Zymo Research's DNA Clean and Concentrator Kit (catalog no. D4014). The restriction enzyme cut site, found in the Int6 homologous sequence, linearized the plasmid, allowing for homologous integration at its corresponding locus (an intergenic region). Five microliters of linearized plasmid were used to transform 50 μ L of SHP033 competent cells and plated on YPD media plates containing hygromycin. After 3 days of growth, individual colonies were confirmed for integration of the linearized plasmid using PCR and genomic DNA from each individual colony.

[0181] A recombinant *Pichia* host cell strain (SHP035) with four copies of the gene encoding S31V CYP (SEQ ID NO: 6) and three copies of the gene encoding CPR (SEQ ID NO: 4) was also constructed as follows. A single plasmid (wbplasmid174; SEQ ID NO: 1064) containing a

nourseothricin resistance marker, an ampicillin resistance marker, a pUC origin of replication, an Int15 homologous sequence (SEQ ID NO: 1056), and a cassette comprising of terminator tFBA2 (SEQ ID NO: 1057), gene encoding CPR (SEQ ID NO: 4), bidirectional promoter pCAT1:FDH1 (SEQ ID: 1053), gene encoding CYP100 (SEQ ID NO: 6), and terminator tADH2 (SEQ ID NO: 1058), was assembled as described above. Sequences were amplified using PCR using the following primers: wboligos5993, wboligos5994, wboligos5995, wboligos5996, wboligos5997, wboligos5998, wboligos5999, wboligos6000, wboligos6001, and wboligos6002. Following plasmid recovery and sequence confirmation as described above, a single plasmid was digested with Sall and further purified using Zymo Research's DNA Clean and Concentrator Kit (catalog no. D4014). The restriction enzyme cut site, found in the Int15 homologous sequence, linearized the plasmid, allowing for homologous integration at its corresponding locus (an intergenic region). Five microliters of linearized wbp173 and 5 microliters of linearized wbp174 were used to co-transform 50 µL of SHP033 competent cells and plated on YPD media plates containing hygromycin and nourseothricin. After 3 days of growth, individual colonies were confirmed for integration of the linearized plasmids using PCR and genomic DNA from each individual colony.

[0182] A recombinant *Pichia* host cell strain (SHP038) with one copy of a polynucleotide of SEQ ID NO: 31 encoding the CYP105 mutant (S31V, Q54K, A143G, T307A, G325E) of SEQ ID NO: 32 identified in Example 3 and one copy of the gene of SEQ ID NO: 3 encoding CPR (SEQ ID NO: 4) was also constructed as follows. A single plasmid (wbp193; SEQ ID NO: 1065) containing a zeocin resistance marker, a pUC origin of replication, a HIS4 homologous sequence (SEQ ID NO: 1005), and a cassette comprising of tDAS1:CPR001:pCAT1:pFDH1:CYP105:tDAS2 was assembled as described above. As wbp147 shares the same sequence as wbp193 except for the CYP CDS, inverse PCR was used to amplify the vector minus the CYP CDS using primers wboligos5800 and wboligos5803. CYP105 (SEQ ID NO: 32) was amplified from SHS042 gDNA using primers wboligos5801 and wboligos5802. Following plasmid recovery and sequence confirmation as described above, a single plasmid was digested with Bbvcl and further purified using Zymo Research's DNA Clean and Concentrator Kit (catalog no. D4014). The restriction enzyme cut site, found in the HIS4 homologous sequence, linearized the plasmid, allowing for homologous integration at the HIS4 locus. Five microliters of linearized plasmid were used to transform 50 µL of BG-10 competent cells and plated on YPDS media plates containing zeocin. After 3 days of growth, individual colonies were confirmed for integration of the linearized plasmid using PCR and genomic DNA from each individual colony.

[0183] A recombinant *Pichia* host cell strain (SHP052) with one copy of a gene encoding CYP105 mutant (S31V, Q54K, A143G, T307A, G325E) of SEQ ID NO: 32 identified in Example 3 was also constructed as follows. A single plasmid (wbp202; SEQ ID NO: 1066) containing a geneticin resistance marker, an ampicillin resistance marker, a pUC origin of

replication, and a cassette comprising of pAOX1:CYP105:tDAS2 was assembled as described above. As wbpplasmid202 shares the same sequence as wbpplasmid160 except for the CYP100 CDS, inverse PCR was used to amplify the vector minus the CYP100 CDS using primers wboligos5803 and wboligos5921. CYP105 (SEQ ID NO: 32) was amplified from SHS042 gDNA using primers wboligos5922 and wboligos5802. Following plasmid recovery and sequence confirmation as described above, a single plasmid was digested with SacI and further purified using Zymo Research's DNA Clean and Concentrator Kit (catalog no. D4014). The restriction enzyme cut site, found in the Int6 homologous sequence, linearized the plasmid, allowing for homologous integration at the Int6 locus (intergenic region). Five microliters of linearized plasmid were used to transform 50 μ L of BG-10 competent cells and plated on YPDS media plates containing geneticin. After 3 days of growth, individual colonies were confirmed for integration of the linearized plasmid using PCR and genomic DNA from each individual colony.

[0184] A recombinant *Pichia* host cell strain (SHP053) with two copies of the gene encoding CYP105 (SEQ ID NO: 32) identified in Example 3 and one copy of the gene encoding CPR (SEQ ID NO: 4) was also constructed as follows. Plasmid wbpplasmid202 was assembled, digested, and purified as described above. Five microliters of linearized plasmid were used to transform 50 μ L of SHP038 competent cells and plated on YPDS media plates containing geneticin. After 3 days of growth, individual colonies were confirmed for integration of the linearized plasmid using PCR and genomic DNA from each individual colony.

[0185] The strain build lineage for all nine of the above-described *P. pastoris* strains is summarized in the chart of **FIG. 3**.

[0186] B. Screening of strains for 3-KCA to 3-KUDCA conversion

[0187] Screening of the recombinant host cells for bioconversion of 3-KCA to 3-KUDCA was carried out according to the following assay: individual colonies of recombinant *P. pastoris* strains were picked into 500 mL baffled conical flasks containing BMGY media (100 mL working volume). The flasks were incubated at 30 °C for 72 h with shaking at 250 rpm at 85 % humidity and were supplemented with glycerol (2 % final concentration) every 24 hours. After this time, the biomass was harvested via centrifugation and the supernatant was discarded. The biomass was then resuspended in BMMY and incubated at 30 °C with shaking at 250 rpm at 85 % humidity for 6 hours to induce protein production. After this time, the biomass was harvested via centrifugation and the supernatant removed. A portion of the resulting biomass (2 g) was resuspended in 10 mL of bioconversion buffer (295-10000 mg/L 3-KCA, 0.1 M phosphate buffer, 2 % MeOH, 2 mM 5-ALA, pH 9) and the flask was incubated at 30 °C for 24 h with shaking at 250 rpm at 85 % humidity. The extraction, sample preparation and analytical methods are identical to those described in Example 1.

[0188] **Results**

[0189] Screening results for 3-KCA to 3-KUDCA bioconversion by the *P. pastoris* strain builds are summarized in Table 11 below.

[0190] TABLE 11

Strain Code	<i>P. pastoris</i> Strain Description	3-KCA Loading (mg/L)	% Conversion
SHP026	Single copies of CYP and CPR at HIS4 locus.	295	100
		590	54
		1000	22
SHP030	Two copies of CYP at loci HIS4 and AOX1. Single copy of CPR at HIS4 locus.	590	100
		1000	59
		2000	29
SHP029	Single copies of CYP100 and CPR at HIS4 locus.	1000	40
		2000	18
SHP033	Two copies of CYP100 at loci HIS4 and AOX1. Single copy of CPR at HIS4 locus.	1000	83
		2000	43
		10000	12
SHP034	Three copies of mutant CYP100 at loci HIS4 AOX1, and Int6. Two copies of CPR at loci HIS4 and Int6.	2000	42
SHP035	Four copies of CYP100 at loci HIS4 AOX1, Int6, and Int15. Three copies of CPR at loci HIS4, Int6, and Int15.	2000	68
		10000	12
SHP038	Single copies of CYP105 and CPR at HIS4 locus.	2000	86
		5000	20
SHP054	Three copies of CYP105 at loci HIS4, AOX1, Int6. Two copies of CPR at loci HIS4, Int6.	5000	33
SHP055	Four copies of CYP105 at loci HIS4, AOX1, Int6. Two copies of CPR at loci HIS4, Int6.	5000	30

EXAMPLE 5: Optimization of Heterologous CYP Gene in Recombinant *P. pastoris* Host Cells by Site Saturation Mutagenesis (SSM)

[0191] This example illustrates the preparation of site saturation mutagenesis (SSM) libraries of engineered polypeptides derived from the parent CYP105 polypeptide of SEQ ID NO: 32 from Example 3 expressed in the SHS058 parent strain and screening these variant strains for improved activity in the bioconversion of 3-KCA to 3-KUDCA relative to the bioconversion of the parent strain SHS058.

[0192] Material and Methods

[0193] A. Site Saturation Mutagenesis library (Round 3-SSM1-6 CYP105 backbone libraries) construction

[0194] The heterologous nucleic acid sequence from the single copy strain SHS058 build, which includes the codon-optimized polynucleotide sequence of SEQ ID NO:31 encoding the CYP polypeptide of SEQ ID NO: 32, expressed under the pGal1 promoter within the pGal1/10 promoter system (SEQ ID NO: 59) and PGK1t terminator (SEQ ID NO: 61), was used to design SSM oligonucleotide to generate SSM libraries at positions spanning the entire polypeptide sequence. To generate an appropriate screening strain to evaluate these libraries, the URA3 marker was integrated into the X-4 site (Easy-Clone 2.0) of a yeast strain under the pGal1

promoter within the pGal1/10 promoter system (SEQ ID NO: 59) and PGK1 terminator (SEQ ID NO: 61), along with the codon-optimized polynucleotide sequence of SEQ ID NO: 3, which encodes the CPR polypeptide of SEQ ID NO: 4, under the pGal10 promoter within the pGal1/10 promoter system (SEQ ID NO: 59) and ADH1t terminator (SEQ ID NO: 60). The resulting strain SHEVP003 was used as the negative control and screening host for CYP SSM library integration. Libraries were plated on selective media containing 5'-fluoroorotic acid, which selects for URA3 negative cells, indicating successful integration of DONOR DNA into the SHEVP003 screening strain. SHS058 strain was used as a control strain during screening of the SSM library strains for fold-improvement calculations with respect to conversion of 3-KCA to 3-KUDCA as described below.

[0195] Genomic DNA from the SHS058 strain was used as the template to generate two PCR products: (1) a first PCR product (Fragment A), which does not harbor any degenerate codons, and (2) a second PCR product (Fragment B), which has sequence overlap with the Fragment A, and is amplified harboring one NNK degenerate codon only. Primers spanning the full 513 codons of the CYP polypeptide were designed according to standard site-saturation mutagenesis protocols and used for amplification of Fragments A and B and overlap extension.

[0196] Fragment A was amplified using a single forward primer P34 (SEQ ID NO: 95) and a series of 513 reverse primers designed according to the location of the desired mutagenesis site in CYP. The 513 reverse primers for Fragment A listed in Table 6A (above) and are provided in the accompanying Sequence Listing as SEQ ID NOs: 102-614. Additional reverse primers were designed as to not perturb the codon mutations in the backbone of CYP polypeptide SEQ ID NO: 32. The reverse primers for Fragment A are provided in Table 12A below and are provided in the accompanying Sequence Listing as SEQ ID NOs: 1067-1114.

[0197] TABLE 12A Additional reverse primers for SSM library Fragment A amplification

Primer Name	Primer Sequence	SEQ ID NO:
Codon310_E8	TAAATCGGCTGTTGTGTGAATCGCCACCAG	1067
Codon314_E9	TGTTTCAGACAGTAAATCGGCTGTTGTGTG	1068
Codon326_E11	CTCAAGGAGTTCTGGTTGCAAAGCTATATT	1069
Codon327_E12	TGGCTCAAGGAGTTCTGGTTGCAAAGCTAT	1070
Codon330_F1	TTACGCTAGTGGCTCAAGGAGTTCTGGTTG	1071
Codon332_F2	TATCTCTTCACGTAGTGGCTCAAGGAGTTC	1072
Codon333_F3	AACATCTCTTCACGTAGTGGCTCAAGGAG	1073
Codon32_A8	AACATACTTTAGTATGAAAAAGCTAAATAA	1074
Codon33_A9	AAGAACATACTTTAGTATGAAAAAGCTAAA	1075
Codon56_A12	TCGTTTATTTCGACAGCTCAAAAGGTTTCTT	1076
Codon59_B1	CTGGACAACCTCGTTTATTCGACAGCTCAAA	1077
Codon61_B2	AAAATCCTGGACAACCTCGTTTATTCGACAG	1078
Codon151_C4	TGGTTTTGTTAGTTTTTGCCAATCCCCTTGT	1079
Codon153_C5	CCACAATGGTTTTGTTAGTTTTTGCCAATCC	1080
Codon313_E7	TTCAAGACAGTAAATCGGCTGTTGTGTGAAT	1081
Codon328_E10	TAGTGGCTCAAGGAGTTCTGGTTGCAAAGC	1082
Codon329_E11	ACGTAGTGGCTCAAGGAGTTCTGGTTGCAA	1083
Codon334_E12	CGTAACTATCTCTTCACGTAGTGGCTCAAG	1084

Codon39_A2	CTTGCCCCCGTTTCCGAGAAGAACATACTT	1085
Codon55_A6	TTTATTCGACAGCTCAAAAGGTTTCTTAGG	1086
Codon57_A7	AACTCGTTTATTCGACAGCTCAAAAGGTTT	1087
Codon60_A8	ATCCTGGACAACCTCGTTTATTCGACAGCTC	1088
Codon62_A9	GATAAAATCCTGGACAACCTCGTTTATTCGA	1089
Codon148_B12	TAGTTTGGCCAATCCCCTTGTAAGATACTT	1090
Codon152_C1	CAATGGTTTTGTTAGTTTTGCCAATCCCCT	1091
Codon308_E12	GGCTGTTGTGTGAATCGCCACCAGTGATAG	1092
Codon315_F1	CATTGTTTCAGACAGTAAATCGGCTGTTGT	1093
Codon316_F2	AAACATTGTTTCAGACAGTAAATCGGCTGT	1094
Codon34_A8	GAGAAGAACATACTTTAGTATGAAAAAGCT	1095
Codon35_A9	TCCGAGAAGAACATACTTTAGTATGAAAAA	1096
Codon36_A10	GTTTCCGAGAAGAACATACTTTAGTATGAA	1097
Codon38_A11	GCCCCCGTTTCCGAGAAGAACATACTTTAG	1098
Codon64_B7	GTTCTCGATAAAATCCTGGACAACCTCGTTT	1099
Codon149_C8	TGTTAGTTTTGCCAATCCCCTTGTAAGATA	1100
Codon150_C9	TTTTGTTAGTTTTGCCAATCCCCTTGTAAG	1101
Codon312_F2	AGACAGTAAATCGGCTGTTGTGTGAATCGC	1102
Codon317_F3	ATTAAACATTGTTTCAGACAGTAAATCGGC	1103
Codon331_F4	CTCTTCACGTAGTGGCTCAAGGAGTTCTGG	1104
Codon335_F5	AACCGTAACTATCTCTTCACGTAGTGGCTC	1105
Codon40_A8	TTGCTTGCCCCCGTTTCCGAGAAGAACATA	1106
Codon41_A9	ATATTGCTTGCCCCCGTTTCCGAGAAGAAC	1107
Codon58_A10	GACAACTCGTTTATTCGACAGCTCAAAAGG	1108
Codon63_A11	CTCGATAAAATCCTGGACAACCTCGTTTATT	1109
Codon144_B11	TCCCCTTGTAAGATACTTGTTAACGAGCGG	1110
Codon145_B12	CAATCCCCTTGTAAGATACTTGTTAACGAG	1111
Codon147_C1	TTTTGCCAATCCCCTTGTAAGATACTTGTT	1112
Codon309_E3	ATCGGCTGTTGTGTGAATCGCCACCAGTGA	1113
Codon311_E4	CAGTAAATCGGCTGTTGTGTGAATCGCCAC	1114

[0198] Fragment B was amplified with a single reverse primer P33 (SEQ ID NO: 94) and a series of 513 forward primers that included a single NNK degenerate codon spanning across successive positions of the CYP encoding gene of SEQ ID NO: 31. The 513 forward primers for Fragment B are listed in Table 6B and are provided in the accompanying Sequence Listing as SEQ ID NOs: 615-984. . Additional forward primers were designed as to not perturb the codon mutations in the backbone of CYP polypeptide SEQ ID NO: 32. The forward primers for Fragment B are provided in Table 12B below and are provided in the accompanying Sequence Listing as SEQ ID NOs: 1115-1184.

[0199] **TABLE 12B:** Additional forward primers for SSM library Fragment B amplification

Primer Name	Primer Sequence	SEQ ID NO:
Codon140_A11	TTCGATCTGGTGCCGCTCGTTAACAAGTATNNKACAAGGGGATTG	1115
Codon303_C1	ACTGAACAAATTAACTATCACTGGTGGCGNNKACACAACAGCC	1116
Codon304_C2	GAACAAATTAACTATCACTGGTGGCGATTNNKACAACAGCCGAT	1117
Codon305_C3	CAAATTAACTATCACTGGTGGCGATTACANNKACAGCCGATTTA	1118
Codon306_C4	ATTAACTATCACTGGTGGCGATTACACANNKGCCGATTACTG	1119
Codon37_A10	ATACTAAAGTATGTTCTTCTCGGAAACGGGNNKAAGCAATATCCT	1120
Codon139_C5	CCGTTTCGATCTGGTGCCGCTCGTTAACAAGNNKCTTACAAGGGGA	1121
Codon146_C6	GTTAACAAGTATCTTACAAGGGGATTGGCANNKCTAACAAAACCA	1122

Codon310_E8	CTGGTGGCGATTACACAAACAGCCGATTTANNKTCTGAAACAATG	1123
Codon314_E9	CACACAACAGCCGATTTACTGTCTGAAACANNKTTTAAATATAGCT	1124
Codon323_E10	ACAATGTTTAAATATAGCTTTGCAACCAGAANNKCTTGAGCCACTA	1125
Codon326_E11	AATATAGCTTTGCAACCAGAACTCCTTGAGNNKCTACGTGAAGAG	1126
Codon327_E12	ATAGCTTTGCAACCAGAACTCCTTGAGCCANNKCGTGAAGAGATA	1127
Codon330_F1	CAACCAGAACTCCTTGAGCCACTACGTGAANNKATAGTTACGGTT	1128
Codon332_F2	GAAGTCCTTGAGCCACTACGTGAAGAGATANNKACGGTTTTATCC	1129
Codon333_F3	CTCCTTGAGCCACTACGTGAAGAGATAGTTNNKGTTTTATCCACG	1130
Codon32_A8	TTATTTAGCTTTTTTCACTAAAGTATGTTNNKCTCGGAAACGGG	1131
Codon33_A9	TTTAGCTTTTTTCACTAAAGTATGTTCTNNKGAAACGGGGGC	1132
Codon51_A11	CCTTATATCAACCCTAAGAAACCTTTTGAGNNKTCGAATAAACGA	1133
Codon56_A12	AAGAAACCTTTTGAGCTGTCTGAATAAACGANNKGTCAGGATTTT	1134
Codon59_B1	TTTGAGCTGTCTGAATAAACGAGTTGTCCAGNNKTTTATCGAGAAC	1135
Codon61_B2	CTGTCTGAATAAACGAGTTGTCCAGGATTTNNKGAGAACGCACGA	1136
Codon141_C2	GATCTGGTGCCGCTCGTTAACAAGTATCTNNKAGGGGATTGGCA	1137
Codon142_C3	CTGGTGCCGCTCGTTAACAAGTATCTTACANNKGATTGGCAAAA	1138
Codon151_C4	ACAAGGGGATTGGCAAACTAACAAAACANNKTGGGCCGAAGCT	1139
Codon153_C5	GGATTGGCAAACTAACAAAACCATTTGTGGNNKGAAGCTTCTTTA	1140
Codon313_E7	ATTCACACAACAGCCGATTTACTGTCTGAANNKATGTTTAAATATA	1141
Codon324_E9	ATGTTTAAATATAGCTTTGCAACCAGAACTCANNKGAGCCACTACGT	1142
Codon328_E10	GCTTTGCAACCAGAACTCCTTGAGCCACTANNKGAAGAGATAGTT	1143
Codon329_E11	TTGCAACCAGAACTCCTTGAGCCACTACGTNNKGAGATAGTTACG	1144
Codon334_E12	CTTGAGCCACTACGTGAAGAGATAGTTACGNNKTTATCCACGGAA	1145
Codon39_A2	AAGTATGTTCTTCTCGGAAACGGGGGCAAGNNKTATCCTTATATC	1146
Codon55_A6	CCTAAGAAACCTTTTGAGCTGTCTGAATAAANNKGTTGTCCAGGAT	1147
Codon57_A7	AAACCTTTTGAGCTGTCTGAATAAACGAGTTNNKCAGGATTTTATC	1148
Codon60_A8	GAGCTGTCTGAATAAACGAGTTGTCCAGGATNNKATCGAGAACGCA	1149
Codon62_A9	TCTGAATAAACGAGTTGTCCAGGATTTATCANNKAACGCACGAGAC	1150
Codon148_B12	AAGTATCTTACAAGGGGATTGGCAAACTANNKAAACCATTGTGG	1151
Codon152_C1	AGGGGATTGGCAAACTAACAAAACCATTTGNNKGCCGAAGCTTCT	1152
Codon308_E12	CTATCACTGGTGGCGATTACACAACAGCCNNKTACTGTCTGAA	1153
Codon315_F1	ACAACAGCCGATTTACTGTCTGAAACAATGNNKAAATATAGCTTTG	1154
Codon316_F2	ACAGCCGATTTACTGTCTGAAACAATGTTNNKATAGCTTTGCAA	1155
Codon322_F5	GAAACAATGTTTAAATATAGCTTTGCAACCANNKCTCCTTGAGCCA	1156
Codon28_A6	TGTGGCATAACTTTATTTAGCTTTTTTCATANNKAAAGTATGTTCTT	1157
Codon30_A7	ATAACTTTATTTAGCTTTTTTCACTAAAGNNKGTTCTTCTCGGA	1158
Codon34_A8	AGCTTTTTTCACTAAAGTATGTTCTTCTCANNKAACGGGGGCAAG	1159
Codon35_A9	TTTTTCACTAAAGTATGTTCTTCTCGANNKGGGGGCAAGCAA	1160
Codon36_A10	TTCACTAAAGTATGTTCTTCTCGAAACNNKGCAAGCAATAT	1161
Codon38_A11	CTAAAGTATGTTCTTCTCGAAACGGGGGCNNKCAATATCCTTAT	1162
Codon50_B4	TATCCTTATATCAACCCTAAGAAACCTTTTNNKCTGTCTGAATAAA	1163
Codon52_B5	TATATCAACCCTAAGAAACCTTTTGAGCTGNNKAAATAAACGAGTT	1164
Codon53_B6	ATCAACCCTAAGAAACCTTTTGAGCTGTGNNKAAACGAGTTGTC	1165
Codon64_B7	AAACGAGTTGTCCAGGATTTTATCGAGAACNNKCGAGACATTTTG	1166
Codon149_C8	TATCTTACAAGGGGATTGGCAAACTAACANNKCCATTGTGGGCC	1167
Codon150_C9	CTTACAAGGGGATTGGCAAACTAACAAAANNKTGTGGGGCCGAA	1168
Codon312_F2	GCGATTACACAACAGCCGATTTACTGTCTNNKACAATGTTTAAAT	1169
Codon317_F3	GCCGATTTACTGTCTGAAACAATGTTTAAATNNKGCTTTGCAACCA	1170
Codon331_F4	CCAGAAGTCTTGAGCCACTACGTGAAGAGNNKGTTACGGTTTTTA	1171
Codon335_F5	GAGCCACTACGTGAAGAGATAGTTACGGTTNNKTCACGGAAGGT	1172
Codon27_A6	TTTTGTGGCATAACTTTATTTAGCTTTTTTCNNKCTAAAGTATGTT	1173
Codon29_A7	GGCATAACTTTATTTAGCTTTTTTCACTANNKTATGTTCTTCTC	1174
Codon40_A8	TATGTTCTTCTCGGAAACGGGGGCAAGCAANNKCCTTATATCAAC	1175
Codon41_A9	GTTCTTCTCGGAAACGGGGGCAAGCAATATNNKTATATCAACCCT	1176
Codon58_A10	CCTTTTGAGCTGTCTGAATAAACGAGTTGTCTNNKGATTTTATCGAG	1177

Codon63_A11	AATAAACGAGTTGTCCAGGATTTTATCGAGNNKGCACGAGACATT	1178
Codon144_B11	CCGCTCGTTAACAAGTATCTTACAAGGGGANNKGCAAAACCTAACA	1179
Codon145_B12	CTCGTTAACAAGTATCTTACAAGGGGATTGNNKAACTAACAAAA	1180
Codon147_C1	AACAAGTATCTTACAAGGGGATTGGCAAAANNKACAAAACCATTTG	1181
Codon309_E3	TCACTGGTGGCGATTACACACAACAGCCGATNNKCTGTCTGAAACA	1182
Codon311_E4	GTGGCGATTACACACAACAGCCGATTTACTGNNKGAAACAATGTTT	1183
Codon321_E5	TCTGAAACAATGTTTAATATAGCTTTGCAANNKGAACCTCCTTGAG	1184

[0200] The two fragments A and B were assembled by overlap extension PCR using forward primer P35 (SEQ ID NO: 96) and reverse primer of P36 (SEQ ID NO: 97). The assembled OE-PCR products were then pooled together, and gel purified to provide a saturation mutagenesis library of linear donor DNA. The pooled saturation mutagenesis library linear donor DNA was integrated as a knock-in using CRISPR-Cas9 into the X-4 locus where URA3 was integrated in SHEVP003 which already had the CPR gene integrated.

[0201] B. Screening of SSM library (R3)

[0202] Individual colonies of recombinant *S. cerevisiae* strains were picked into 96-well plates containing 2 x YPD media (300 μ L per well) using a QPix™ 420 colony picking system. The plates were incubated at 30 °C for 24 h with shaking at 250 rpm at 85 % humidity. After this time, the strains were sub-cultured (40 μ L inoculation volume per strain) into a second 96-well plate containing 1 x YPD media (300 μ L per well) using an Agilent Bravo automated liquid handling platform. The plates were incubated at 30 °C for 48 h with shaking at 250 rpm at 85 % humidity. After this time, a solution of galactose (40 % stock solution) was added using the Bravo (1 % final concentration) and the plates were further incubated for 4 hours to induce protein production. After this time, the biomass was harvested via centrifugation and the supernatant removed. The resulting biomass was resuspended in 300 μ L of bioconversion buffer (590 mg/L 3-KCA, 0.1 M phosphate buffer, pH 9) using the Agilent Bravo automated liquid handling platform and the plates were incubated at 30 °C for 24 h with shaking at 250 rpm at 85 % humidity. After this time, MeOH (300 μ L) was added using the Bravo and the plates were incubated at 30 °C with shaking at 250 rpm at 85 % humidity for a further 30 minutes. The biomass was then separated by centrifugation and the resulting supernatant diluted with MeOH (1000 x total dilution) using the Agilent Bravo automated liquid handling platform before analysis using an Agilent 6470 Triple Quadrupole LC/MS.

[0203] LC/MS sample preparation: The bioconversion reaction was extracted and diluted with MeOH for sample preparation as described above. The prepared samples were loaded onto an Agilent 6470 Triple Quadrupole LC/MS and the compounds of interest were detected using MS/MS in SIM mode. The compounds were quantified relative to calibration curves which were prepared using serial dilutions of stock solutions of 3-KCA and 3-KUDCA.

[0204] LC/MS instrumentation and parameters: LC/MS system: Agilent 6470 Triple Quadrupole LC/MS; Column: Agilent EclipsePlusC18 column 1.8 μ m, 3.0 x 50 mm; Mobile phase: Solvent A: H₂O with 0.1% formic acid; Solvent B: Acetonitrile with 0.1% formic acid;

Gradient: 0 - 0.2 min isocratic 40% B, 0.2 - 2.2 min gradient to 85% B, 2.2 - 2.8 min isocratic 85% B; Mode: MS/MS in SIM mode (monitoring at 389.3 for 3-KUDCA and 373.2 for 3-KCA); Fragmentor: 90 volts.

[0205] Results

[0206] Data from screening the SSM libraries in terms of fold-improvement in production of 3-KUDCA from 3-KCA relative to the control strain, SHS058 expressing CYP105 are summarized in Table 13 (below).

[0207] TABLE 13

Name	NT SEQ ID NO:	AA SEQ ID NO:	AA differences relative to CYP105 (SEQ ID NO: 32)	3-KCA to 3-KUDCA % Conversion ¹	FIOC 3-KCA to 3-KUDCA % Conversion
CYP110	1185	1186	L365A	63.9	1.96
CYP207	1187	1188	V364L	64.9	8.1
	1189	1190	L7R	63.4	7.9
	1191	1192	A153Q	66.1	7.2
	1193	1194	A153P	52.0	5.7
	1195	1196	Q424V	44.8	3.7
	1197	1198	E176D	46.3	3.1
	1199	1200	K479R	36.5	3.0
	1201	1202	K70L	3.6	2.8
	1203	1204	T333V	12.1	2.4
	1205	1206	E338Q	3.1	2.4
	1207	1208	K70N	48.4	2.0
	1209	1210	V210R	15.4	1.9
	1211	1212	F452D	23.1	1.9
	1213	1214	Q320M	33.8	1.8
	1215	1216	H170V	30.5	1.6
	1217	1218	N160H	22.8	1.5
	1219	1220	S408R	23.9	1.2
¹ 3-KCA to 3-KUDCA % conversion is calculated as $\{[3\text{-KUDCA}] / ([3\text{-KCA}] + [3\text{-KUDCA}])\} \times 100$					

Example 6: Further Optimization of Mutant CYP Genes in Recombinant *P. pastoris* Host Cells by Combinatorial Mutagenesis

[0208] This example illustrates the preparation of further combinatorial libraries of CYP genes based on the parent strain SHS088 which carries a gene of SEQ ID NO: 1187 encoding the CYP207 (SEQ ID NO: 1188) which has the 6 amino acid differences S31V, Q54K, A143G, T307A, G325E, V364L relative to the wild-type CYP of SEQ ID NO: 2, identified through SSM optimization of CYP105 for improved bioconversion of 3-KCA to 3-KUDCA described in Example 5.

[0209] Material and Methods

[0210] A. Combinatorial library construction (Round 4)

[0211] The polynucleotide sequence of SEQ ID NO: 1187, encoding polypeptide SEQ ID NO: 1188 (also referred to herein as “CYP207”) was used to design oligos to generate combinatorial libraries to randomly incorporate additional identified beneficial amino acid mutations from Example 5 into the CYP207 sequence. The same screening strain from Example 2, SHEVP003) was used as a host for integration of the resulting libraries. SHS088, which expresses the CYP207 polypeptide of SEQ ID NO: 1188 and the CPR001 polypeptide of SEQ ID NO: 4, was used as the control strain.

[0212] A semi-synthetic approach was used to construct the first set of combinatorial libraries. Genomic DNA from the SHS088 strain was used as the template to generate a full-length PCR product using primer pair of P34 (SEQ ID NO: 95) and P37 (SEQ ID NO: 98), while incorporating uracil using a dNTP mix comprising of the following deoxyribonucleotides: dATP, dGTP, dCTP, dTTP, dUTP. The resulting PCR product was gel purified and digested with Uracil-DNA Glycosylase and Endonuclease IV at 37 C for 2 hours, followed by enzyme denaturation at 94 C for two minutes, to generate a pool of fragments in the range of 50-100 bases. These fragments were further combined with differing ratios of pools of the synthesized oligonucleotide primers (each oligo up to 55 bases in length and encoding one or more amino acid change) in several individual assembly PCR reactions using forward primer P5 (SEQ ID NO: 66) and reverse primer P6 (SEQ ID NO: 67) to reassemble the full-length PCR product as Fragment B and incorporate mutagenic amino acid changes within each pool randomly. The synthesized oligonucleotide primers SEQ ID NOs: 1221-1242 used in the pools and their encoded mutations are listed in Table 14 below and the accompanying Sequence Listing.

[0213] TABLE 14

Primer Name	Primer Sequence	SEQ ID NO:
ELITE2_L7R	ATGGCCACCGATCTAGACCGTGTATTAGGAAAAAGTCAATA	1221
ELITE2_K70N	ACGCACGAGACATTTTGACTAACGGTCGCTCACTTTACAAGGA	1222
ELITE2_K70L	ACGCACGAGACATTTTGACTCTAGGTCGCTCACTTTACAAGGA	1223
ELITE2_A153P	TAACAAAACCATTTGTGGCCGGAAGCTTCTTTAGGTGTAAACCATGTT TTAGGCACTTC	1224
ELITE2_A153Q	TAACAAAACCATTTGTGGCAAGAAGCTTCTTTAGGTGTAAACCATGTT TTAGGCACTTC	1225
ELITE2_N160H	TAACAAAACCATTTGTGGGCCGAAGCTTCTTTAGGTGTACATCATGTT TTAGGCACTTC	1226
ELITE2_A153P -N160H	TAACAAAACCATTTGTGGCCGGAAGCTTCTTTAGGTGTACATCATGTT TTAGGCACTTC	1227
ELITE2_A153Q -N160H	TAACAAAACCATTTGTGGCAAGAAGCTTCTTTAGGTGTACATCATGTT TTAGGCACTTC	1228
ELITE2_H170V	AGGCACTTCTACTGAGTGGGTCCCTATTAAACCAGGGGAAGATATCA TGAGGATTGTCT	1229
ELITE2_E176D	AGGCACTTCTACTGAGTGGCATCCTATTAAACCAGGGGACGATATCA TGAGGATTGTCT	1230
ELITE2_H170V -E176D	AGGCACTTCTACTGAGTGGGTCCCTATTAAACCAGGGGACGATATCA TGAGGATTGTCT	1231
ELITE2_V210R	AAGTTAGTATTGAGTACACTCGTCAATTGTTTCAAACCGCAGA	1232
ELITE2_Q320M	CAATGTTTAAATATAGCTTTGATGCCAGAACTCCTTGAGCCACT	1233

ELITE2_T333V	CACTACGTGAAGAGATAGTTGTCGTTTTATCCACGGAAGGTCTAAAA AAGACGTCGTT	1234
ELITE2_E338Q	CACTACGTGAAGAGATAGTTACGGTTTTATCCACGCAAGGTCTAAAA AAGACGTCGTT	1235
ELITE2_T333V- E338Q	CACTACGTGAAGAGATAGTTGTCGTTTTATCCACGCAAGGTCTAAAA AAGACGTCGTT	1236
ELITE2_L365A	CACAGAGACTTCGCCCTTTGGCGTTAGGTGCTTTCAGGAGAAT	1237
ELITE2_L365S	CACAGAGACTTCGCCCTTTGTCTTTAGGTGCTTTCAGGAGAAT	1238
ELITE2_S408R	CAGAATACTATCCAGATGCTCGTAAGTTCAATGCATATAGATT	1239
ELITE2_Q424V	AAATGAGACAGACACCGGGTGTGACACAAAAGAGCACACCTTGT	1240
ELITE2_F452D	ACGCGTGCCCAGGAAGATTTGACGCAGCCAATGAGATTAAAGAT	1241
ELITE2_K479R	TACCAGAAGGTGTTGTACCTCGTTCTAAGGCTTTAGGAATGTC	1242

[0214] For efficient integration of the combinatorial mutant CYP variants, PCR amplicons with homologous 5' and 3' regions to the CYP were generated (Fragment A and Fragment C respectively). Fragment A was amplified using the forward primer P37 (SEQ ID NO: 98) and reverse primer P4 (SEQ ID NO: 65) and Fragment C was amplified using the forward primer P7 (SEQ ID NO: 68) and the reverse primer P38 (SEQ ID NO: 99). Fragments A, B, and C were assembled by NEBuilder[®] HiFi Assembly and amplified as a single linear DONOR using the following forward rescue primer P35 (SEQ ID NO: 96) and reverse primer P36 (SEQ ID NO: 97). The assembled PCR products were then pooled together, and gel purified to provide a combinatorial library of linear donor DNA.

[0215] The pooled combinatorial libraries of linear donor DNA were integrated as a knock-in using CRISPR-Cas9 into the X-4 locus where URA3 was integrated in SHEVP003 which already had the CPR gene integrated as in Examples 1 and 2.

[0216] B. Screening of Combi library

[0217] Screening of the recombinant host cells for bioconversion of 3-KCA to 3-KUDCA was carried out using the same protocol described for screening the SSM library except 0.5 x YPD was used for growth and 3-KCA loading was increased from 590 mg/L to 800 mg/L.

[0218] *Results*

[0219] Screening of the combinatorial libraries for fold-improvement in production of 3-KUDCA from 3-KCA relative to the control strain, SHS088, are summarized in Table 15 (below).

[0220] TABLE 15

NT SEQ ID NO:	AA SEQ ID NO:	AA differences relative to CYP207 (SEQ ID NO: 1188)	3-KCA to 3- KUDCA % Conversion	FIOC 3-KCA to 3-KUDCA % Conversion
1243	1244	L7R, V210R, K479R	33.4	2.0
1245	1246	L7R, K70L, N160H, H170V, E176D	31.9	1.9
1247	1248	L444F	30.9	1.4
1249	1250	Q320M, E402V	22.0	1.3
1251	1252	K70L	25.9	1.2
1253	1254	V261E, Q320M, T333V, K479R	19.4	1.2
1255	1256	A153Q, H170V, Q424V	18.7	1.1

1257	1258	K70L, E338Q, M485T	19.5	0.9
1259	1260	N316S	19.5	0.9
1261	1262	L7R, E176D, K479R	14.6	0.9
1263	1264	L7R, Q320M, K479R	14.4	0.9
1265	1266	L7R, K70L, H170V, K479R	14.6	0.7
1267	1268	L7R, E176D, Q320M	13.4	0.6
1269	1270	N160H, Q320M	6.7	0.4
1271	1272	L7R, A153P, S408R	1.4	0.06

Example 7: Genomic Integration of Heterologous CYP/CPR Gene Pairs into *Pichia pastoris* Host Cells

[0221] This example illustrates the preparation of recombinant yeast host cell strains (*Pichia pastoris*) with genomically integrated copies of a pair of heterologous genes encoding a cytochrome P450 (CYP) and cytochrome reductase (CPR) from *Gibberella zeae* and demonstration of the ability of these strains to carry out the bioconversion of Lithocholic acid (LCA) to Ursodeoxycholic acid (UDCA) and/or 3-Keto-lithocholic acid (3-KCA) to 3-Keto-ursodeoxycholic acid (3-KUDCA).

[0222] Material and Methods

[0223] *A. P. pastoris* strain building

[0224] A recombinant *Pichia* host cell strain (SHP070) with one copy of the gene encoding CYP110 (SEQ ID NO: 1185) was also constructed as follows. A single plasmid, wbp1asmid226 (SEQ ID NO: 1273) containing a geneticin resistance marker, an ampicillin resistance marker, a pUC origin of replication, and a cassette comprising of pAOX1:CYP110:tDAS2 was assembled as described above. The wbp1asmid226 shares the same sequence as wbp1asmid202 (SEQ ID NO: 1066) except for the CYP105 coding sequence (SEQ ID NO: 31) was replaced with the CYP110 coding sequence of SEQ ID NO: 1185. Inverse PCR was used to amplify the vector minus the CYP100 coding sequence using primers wboligos5803 and wboligos5921. The gene encoding CYP110 (SEQ ID NO: 1185) was amplified from SHS062 gDNA using primers wboligos5922 and wboligos5802. CYP110 in SHS062 was expressed under the pGal1 promoter within the pGal1/10 promoter system (SEQ ID NO: 59) and PGK1t terminator (SEQ ID NO: 61). Following plasmid recovery and sequence confirmation as described above, a single plasmid was digested with SacI and further purified using Zymo Research's DNA Clean and Concentrator Kit (catalog no. D4014). The restriction enzyme cut site, found in pAOX1, linearized the plasmid, allowing for homologous integration at AOX1 locus. Five microliters of linearized plasmid were used to transform 50 μ L of BG-11 competent cells and plated on YPD media plates containing geneticin. After 3 days of growth, individual colonies were confirmed for integration of the linearized plasmid using PCR and genomic DNA from each individual colony.

[0225] A recombinant *Pichia* host cell strain (SHP079) with one copy of the gene encoding CYP207 (SEQ ID NO: 1187) was also constructed as follows. A single plasmid, wbp1asmid231 (SEQ ID NO: 1274), containing a geneticin resistance marker, an ampicillin resistance marker,

a pUC origin of replication, and a cassette comprising of pAOX1:CYP207:tDAS2 was assembled as described above. The wbp1asmid231 shares the same sequence as wbp1asmid202 (SEQ ID NO: 1066) except for the CYP105 coding sequence (SEQ ID NO: 31) was replaced with the CYP207 coding sequence of SEQ ID NO: 1187. Inverse PCR was used to amplify the vector minus the CYP100 coding sequence using primers wboligos5803 and wboligos5921. The gene encoding CYP207 (SEQ ID NO: 1187) was amplified from SHS088 gDNA using primers wboligos5922 and wboligos5802. Following plasmid recovery and sequence confirmation as described above, a single plasmid was digested with SacI and further purified using Zymo Research's DNA Clean and Concentrator Kit (catalog no. D4014). The restriction enzyme cut site, found in pAOX1, linearized the plasmid, allowing for homologous integration at AOX1 locus. Five microliters of linearized plasmid were used to transform 50 µL of BG-11 competent cells and plated on YPD media plates containing geneticin. After 3 days of growth, individual colonies were confirmed for integration of the linearized plasmid using PCR and genomic DNA from each individual colony.

[0226] A recombinant *Pichia* host cell strain (SHP083) with one copy of the gene encoding CYP110 (SEQ ID NO: 1185) was also constructed as follows. A single plasmid, wbp1asmid226, was digested with SacI and further purified using Zymo Research's DNA Clean and Concentrator Kit (catalog no. D4014). The restriction enzyme cut site, found in pAOX1, linearized the plasmid, allowing for homologous integration at AOX1 locus. Five microliters of linearized plasmid were used to transform 50 µL of BG-45 competent cells (a double knockout host strain; aox1-KO, yps1-KO) and plated on YPD media plates containing geneticin. After 3 days of growth, individual colonies were confirmed for integration of the linearized plasmid using PCR and genomic DNA from each individual colony. The strain build lineage for the above-described SHP083 *P. pastoris* strain is summarized in the chart of **FIG. 4**.

[0227] B. Screening of *P. pastoris* strains for 3-KCA to 3-KUDCA conversion

[0228] Screening of the recombinant host cells for bioconversion of 3-KCA to 3-KUDCA (as shown in Scheme 1) was carried out according to the following assay: individual colonies of recombinant *P. pastoris* strains were picked into 500 mL baffled conical flasks containing BMGY media (100 mL working volume). The flasks were incubated at 30 °C for 48 h with shaking at 250 rpm at 85 % humidity and were supplemented with glycerol (1 % final concentration) every 24 hours. After this time, the biomass was harvested via centrifugation and the supernatant was discarded. A portion of the resulting biomass (0.5 g) was resuspended in 2.5 mL of bioconversion buffer (10-20 g/L 3-KCA, 0.1 M phosphate buffer, 2 % MeOH pH 9) and the vial was incubated at 30 °C for 72 h with shaking at 250 rpm at 85 % humidity. One volume (2.5 mL) of methanol was then added to the vial for extraction with shaking at 30 °C for 30 mins. The supernatant was then collected via centrifugation and diluted/analyzed as described in previous examples.

[0229] Screening of *P. pastoris* strains for DCA to UCA and CA conversion

[0230] Screening of the recombinant host cells for bioconversion of DCA to UCA/CA (as shown in Scheme 2) was carried out according to the following assay: individual colonies of recombinant SHP070 were picked into 500 mL baffled conical flasks containing BMGY media (100 mL working volume). The flasks were incubated at 30 °C for 48 h with shaking at 250 rpm at 85 % humidity and were supplemented with glycerol (1 % final concentration) every 24 hours. After this time, the biomass was harvested via centrifugation and the supernatant was discarded. A portion of the resulting biomass (0.5 g) was resuspended in 2.5 mL of bioconversion buffer (2 g/L DCA, 0.1 M phosphate buffer, 2 % MeOH pH 9) and the vial was incubated at 30 °C for 72 h with shaking at 250 rpm at 85 % humidity. One volume (2.5 mL) of methanol was then added to the vial for extraction with shaking at 30 °C for 30 mins. The supernatant was then collected via centrifugation and diluted as described in previous examples. The below LC/MS method was used to confirm the production of UCA (23 % conversion) and CA (0.1 % conversion).

[0231] LC/MS instrumentation and parameters: LC/MS system: Agilent 6470 Triple Quadrupole LC/MS; Column: Agilent EclipsePlusC18 column 1.8 μ m, 3.0 x 50 mm; Mobile phase: Solvent A: H₂O with 0.1% formic acid; Solvent B: Acetonitrile with 0.1% formic acid; Gradient: 0 - 0.2 min isocratic 40% B, 0.2 - 2.2 min gradient to 85% B, 2.2 - 2.8 min isocratic 85% B; Mode: MS/MS in SIM mode (monitoring at 391.5 for DCA and 407.5 for UCA and CA); Fragmentor: 90 volts.

[0232] Results

[0233] Screening results for 3-KCA to 3-KUDCA bioconversion by the *P. pastoris* strain builds are summarized in Table 16 below.

[0234] TABLE 16

Strain Code	<i>P. pastoris</i> Strain Description	3-KCA Loading (mg/L)	% Conversion
SHP070	Single copy of CYP110 (SEQ ID NO: 1185) at AOX1 locus	10000	49
SHP079	Single copy of CYP207 (SEQ ID NO: 1187) at AOX1 locus	10000	36
SHP083	Single copy of CYP110 (SEQ ID NO: 1185) at AOX1 locus with $\Delta yps2$	10000	65

[0235] Screening results for DCA to UCA/CA bioconversion by the *P. pastoris* strain builds are summarized in Table 17 below.

[0236] TABLE 17

Strain Code	<i>P. pastoris</i> Strain Description	DCA Loading (mg/L)	% Conversion to CA	% Conversion to UCA
SHP038	Single copies of CYP105 and CPR at HIS4 locus.	2000	0.1	23

Example 8: Design and Expression of a Heterologous CYP-CPR Fusion Polypeptide in *Pichia pastoris* Host Cells

[0237] This example illustrates the design and expression of a heterologous gene encoding a fusion of a cytochrome P450 (CYP) from *Gibberella zeae* and an endogenous cytochrome reductase (CPR) from *Pichia*. The resulting cytosolic fused enzyme with CYP and CPR activities demonstrated an ability to carry out the bioconversion of Lithocholic acid (LCA) to Ursodeoxycholic acid (UDCA) and/or 3-Keto-lithocholic acid (3-KCA) to 3-Keto-ursodeoxycholic acid (3-KUDCA).

[0238] Materials and Methods

[0239] A. Design CYP207-CPR fusion

[0240] A polynucleotide of SEQ ID NO: 1275 was synthesized that encodes an N-terminal 31 amino acid truncation of CYP207 named trCYP207 (SEQ ID NO:1276) fused via short polypeptide linker (aa sequence: ARA) to the N-terminus of a truncated version of the *Pichia pastoris* endogenous reductase, trNCP1 (SEQ ID NO: 1278). The resulting cytosolic fused enzyme encoded by the fusion gene of SEQ ID NO: 1279 has the amino acid sequence of SEQ ID NO: 1280. The fusion and its various constituent polynucleotide and polypeptide sequences are provided in Table 18 below and the accompanying Sequence Listing.

[0241] TABLE 18

Description	SEQ ID NO:	Sequence
trCYP207 gene	1275	ATGCTCGGAAACGGGGGCAAGCAATATCCTTATATCAACCCTAAGAAA CCTTTTGAGCTGTCTGAATAAACGAGTTGTCCAGGATTTTATCGAGAAC GCACGAGACATTTTGACTAAGGGTCGCTCACTTTACAAGGATACGCCC TACAAGGCTCATACCGATTTAGGGGACGTCTTGGAATCCCGCCTGAA TTTGCCGACGCTCTAAAACTGAAAGACAGTTAGATTTTACCGAAGTC GCAAGAGACGATACTCACGGTTATATTCCTGGATTCGAGCCAATCGGT TCCCCGTTTCGATCTGGTGCCGCTCGTTAACAAGTATCTTACAAGGGGA TTGGCAAACTAACAAAACCATTTGTGGGCCGAAGCTTCTTTAGGTGTA AACCATGTTTTAGGCATTCTACTGAGTGGCATCCTATTAACCCAGGG GAAGATATCATGAGGATTGTCTCCAGAATGTCATCCAGAATATTCATG GGTGAGGAACTTTGTAAGATGATGATTGGTTGAAAGTTAGTATTGAG TACACTGTGCAATTGTTTCAAACCGCAGACGAATTACGTAACCTATCCA CGTTGGACAAGACCATATATTCACCTGGTTTTTGCCTTCCTGTCAAGGG GTTAGGAGAAAAATTGCAGGAAGCGCGTGATTTATTGCAACCCCATATT GATAGGAGAAATGCTGTGAAGAAAGAAGCTATCGCTGAAGGTAGACCA TCACCATTTCGACGATTCAATAGAATGGTTTGAAATGAATACGAGGGC AAATCTGATCCCGCCACTGAACAAATTAACTATCACTGGTGGCGATT CACACAACAGCCGATTTACTGTCTGAAACAATGTTTAAATATAGCTTTG CAACCAGAACTCCTTGAGCCACTACGTGAAGAGATAGTTACGGTTTTTA TCCACGGAAGGTCTAAAAAAGACGTCGTTTTACAATTTGAAGTTGATG GATTCGGTTCATAAAAGAGTCACAGAGACTTCGCCCTTTGTTATTAGGT GCTTTCAGGAGAATGGCATTGGCTGATGTTACCTTGCCCAATGGCGAC GTAATTAAAAAAGGTACCAAGATCATTTGCGACACTACACATCAATGG AATCCAGAATACTATCCAGATGCTAGTAAGTTCAATGCATATAGATTT TTGCAAAATGAGACAGACACCGGGTCAGGACAAAAGAGCACACCTTGTC AGCACAAGCCATGATCAAATGGGATTCGGACATGGCTTGACGCGTGC CCAGGAAGATTTTCGCAGCCAATGAGATTAAGATTGCGCTGTGTCAT

		ATGCTATTGAAATATGATTGGAAATTACCAGAAGGTGTTGTACCTAAG TCTAAGGCTTTAGGAATGTCTTTACTTGGTGACCGGGAAGCCAAACTG ATGGTGAAAAAGGAGAGCAGCCGAAATCGATATAGACACTATTGGTAGT GATGAA
trCYP207 polypeptide	1276	MLGNGGKQYPYINPKKPFELSNKRVVQDFIENARDILTKGRSLYKDT YKAHTDLGDVLVIPPEFADALKSERQLDFTEVARDDTHGYIPGFEP SPFDLVPLVNKYLTRGLAKLTKPLWAEASLGVNHVLGTSTEWHPINPG EDIMRIVSRMSSRIFMGEELCKDDDLKVSIEYTVQLFQTADLRNYP RWTRPYIHWFLPSCQGVRRKLQEARDLLQPHIDRRNAVKKKEAIAEGRP SPFDDSIWFENEYEGKSDPATEQIKLSLVAIHTTADLLSETMFNIAL QPELLEPLREEIVTVLSTEGLKKTsfynlklmDSVIKESQRLRPLLLG AFRRMALADVTLPNGDVIKKGTKIICDTHQWNPEYYPDASKFNAYRF LQMRQTPGQDKRAHLVSTSHDQMGFGHGLHACPGRFFAANEIKIALCH MLLK YDWKLPEGVVPKSKALGMSLLGDREAKLMVKRRAAEIDIDTIGS DE
trNCP1 gene	1277	GGCAAAGAGGATGATAACTCCGTCCACGGAGTAGCTGGAGGATTTTCAG ACCAGAGACCTGGTTGAGATTTTGAACCTCTACCAACAAGAAGGCACTC GTCCTTTACGGATCACAAACAGGTACATCTGAGGATTATGCTCATAAG TACGCTCGTGAACCTGCAATCCAAGTTCAGCATTTCCACCCTCTGTGGA GACTTGTGAGAGTTTGATTTTGACAATCTGAACGATATCCCAGAGCAG GTCGAAGGGTTCACCTTTCATCACTTTCTTCATGGCTACGTACGGCGAA GGCGAGCCTACTGACAACGCCGTAGAGTTTATCGAGTTTCTTAAGAAC GACGCCGAGGATCTGTCTAATCTGAAGTATACTGTGTTTGGGTGGGT AACTCTACTTACGAGTTCTATAATCAAATGGGTAAAGACCACTAACAAG CGTTTCTCAGAACTTGGTGCTCAATTGGTTGGAACCTTTGGCGAGGGC GATGATGGTCAGGCAACTATGGACGAGGACTTCCTAGCATGGAAGAT TCCCTGTTTGATACAATCAAGAAGGATTTGCATTTGGAGGAACATGAA GTTGTCTACCAGCCAGGTCTAAAAGTAAAGGAGAACACAGCGCTTACA ACATCTTCTCCTAATGTCTCTGTTGGTGAACCTAATAAGGCTTATGTT CTTCGTGAAGATGAGAATCTGCTGCAATACGGTCCCTTCGACCACACT CATCCTTACATTGCGCCCATCTCCAGCTCCAGAGAACTCTGTTCTGAA ACTTCGGAGAGAACTGTATCCATCTAGAATTTGATCTCAGCAACACC AACTTGC GTTATTCCACTGGCGATCATTGGCCGTGTGGCCAAGTAAT GCTAATGAGCATGTCGAGAGCTTTTTGAAAGTGTTCAACCTCACAGAC AAAAGATCTTCTGTGTTTGATATTGAGTTCCTGGACCCTACTGTCACA GTTCACTTTCCGTTTCCAACCACATACGAAGCAGTTGTTTCGTATCAC CTGGAAATTTCTGGACCAATCTCTAGACAGACATTGAAGCAATTCATT CCTTATGCTCCCGATCAGTCTACTAAACAGGAGGTAATCCGATTATCG GAGAGCAAAGATGTGTTCCACAACGAAGTCACAGCAAAATACTATAAC CTTGCTGACCTCTGTGTTCAAAGTCTCAAAAGAACTCCATGGAATGTT CCATTCAACTTCCTGATCGAGACCATGCCTAATCTGCAGCATCGTTAC TATTCTATCTCATCATCTTCATTAAGTGAAGCAAACTATTCTATATC ACCGCAATGATTGAGGCCTTCACGCCTACTGGAAGTGACCATATAGTG ACAGGAGTCACCACTAATTTGCTATGGAACATTGAGCTGAATCAGGAC AAATCTACTGTTAAGGCGCCTGTCTCCTATGATTTGAACGGCCCAAGA AATTTGTTTTCTCCATATAAACTTCCAGTTCATATCAGAAGGTCCACT TTCAAACTTCCAAGCAACCCGGCCCTGCCAGTGATCATGATTGGACCT GGTACTGGGGTAGCTCCATTGAGAGGCTTTATCAGAGAAAGATGCCAG CAAGTTGACAATGGAACCTCCTAATATTGGTCAATCTATCCTCTATTAT GGCTGTAGAAATCTGAGCAGGATTTCTTTTACCGTGACGAATGGCCA ACATATTGGAAGAAGTTAGGAGATAAGTTCAAGATGTACACTGCGTTC AGCAGAGAAAATTCACACAAAGTTTATGTTTACGATAGGTTATTGGAA AACTCAAGGGAATTTATCGAACTGATGGACCAAGGTGCTTTTATCTAC GTATGTGGAGATGCTGGTAAAAATGGCTAAGGATGTAAATAAGGCGATT GTTGAAATCCTCATTAAGAGAAAGGATTGTCAGAGGAAGATGCTACA GAAAGTATTAGAGAGTTCAAACTTCCAACAGATATCAGGAAGACGTT TGGTAA

trNCP1 polypeptide	1278	GKEDDNSVHGVAGGFQTRDLVEILNSTNKKALVLYGSQTGTSEDYAHK YARELQSKFSIPTLCGDLSEFDFDNLNDIPEQVEGFTFITFFMATYGE GEPTDNAVEFIEFLKND AEDLSNLKYTVFGLGNSTYEFYNQMGKTTNK RFSELGAQLVGTFTGEGDDGQATMDEDFLAWKDSLFDTIKKDLHLEEHE VVYQPGLKVKENTALTTSSPNVSVGEPNKAYVLREDENLLQYGPFDHT HPYIAPISSSRELCSSETSERNCIHLEFDLSNTNLRYSTGDHLAVWPSN ANEHVESFLKVFNLTDKRSSVFDIEFLDPTVTVHFPPFTTYEAVVRHH LEISGPISRQTLKQFIPYAPDQSTKQEVIRLSESKDVFNHNEVTAKYYN LADLLFKVSKETPWNVPFNFLIETMPNLQHRYYSISSSSLSEKQTIHI TAMIEAFTPTGSDHIVTGVTTNLLWNIQLNQDKSTVKAPVSYDLNGPR NLFSPYKLPVHIRRSTFKLPSNPALPVIMIGPGTGVAPFRGFRIRERCQ QVDNGTPNIGQSILYYGCRNSEQDFLYRDEWPTYSKKLGDKFKMYTAF SRENSHKVYVQHRLLENSREFIELMDQGAFIYVCGDAGKMAKDVNKAI VEILIKEKGLSEEDATESIREFKTSNRYQEDVW
linker nt sequence	n/a	GCTAGAGCT
linker aa sequence	n/a	ARA
trCYP207- trNCP1 fusion gene	1279	ATGCTCGGAAACGGGGGCAAGCAATATCCTTATATCAACCCTAAGAAA CCTTTTGAGCTGTGCAATAAACGAGTTGTCCAGGATTTTATCGAGAAC GCACGAGACATTTTGACTAAGGGTCGCTCACTTTACAAGGATACGCCC TACAAGGCTCATACCGATTTAGGGGACGCTCTTGGAATCCCGCCTGAA TTTGCCGACGCTCTAAAACTGAAAGACAGTTAGATTTTACCGAAGTC GCAAGAGACGATACTCACGGTTATATTCCTGGATTTCGAGCCAATCGGT TCCCCGTTTCGATCTGGTGCCGCTCGTTAACAAAGTATCTTACAAGGGGA TTGGCAAACTAACAAAACCATTTGTGGGCCGAAGCTTCTTTAGGTGTA AACCATGTTTTAGGCACTTCTACTGAGTGGCATCCTATTAACCCAGGG GAAGATATCATGAGGATTGTCTCCAGAATGTCATCCAGAATATTCATG GGTGAGGAACCTTTGTAAAGATGATGATTGGTTGAAAGTTAGTATTGAG TACACTGTGCAATTGTTTCAAACCGCAGACGAATTACGTAACCTATCCA CGTTGGACAAGACCATATATTCACCTGGTTTTTGCCTTCCTGTCAAGGG GTTAGGAGAAAATTGCAGGAAGCGCGTGATTTATTGCAACCCCATATT GATAGGAGAAAATGCTGTGAAGAAAGAAGCTATCGCTGAAGGTAGACCA TCACCATTTCGACGATTCAATAGAATGGTTTGAAAATGAATACGAGGGC AAATCTGATCCCGCCACTGAACAAATTAACTATCACTGGTGGCGATT CACACAACAGCCGATTTACTGTCTGAAACAATGTTTAATATAGCTTTG CAACCAGAACTCCTTGAGCCACTACGTGAAGAGATAGTTACGGTTTTTA TCCACGGAAAGGTCTAAAAAAGACGTCGTTTTTACAATTTGAAGTTGATG GATTCGGTCATAAAAGAGTCACAGAGACTTCGCCCTTTGTTATTAGGT GCTTTCAGGAGAATGGCATTGGCTGATGTTACCTTGCCCAATGGCGAC GTAATTAAAAAAGGTACCAAGATCATTTGCGACACTACACATCAATGG AATCCAGAATACTATCCAGATGCTAGTAAGTTCAATGCATATAGATTT TTGCAAATGAGACAGACACCGGGTCAGGACAAAAGAGCACACCTTGTC AGCACAAAGCCATGATCAAATGGGATTTCGGACATGGCTTGACGCGTGC CCAGGAAGATTTTTCGACGCCAATGAGATTAAGATTGCGCTGTGTAT ATGCTATTGAAATATGATTGGAAATTACCAGAAGGTGTTGTACCTAAG TCTAAGGCTTTAGGAATGTCTTTACTTGGTGACCGGGAAGCCAACTG ATGGTGAAAAGGAGAGCAGCCGAAATCGATATAGACACTATTGGTAGT GATGAAGCTAGAGCTGGCAAAGAGGATGATAACTCCGTCCACGGAGTA GCTGGAGGATTTGAGACCAGAGACCTGGTTGAGATTTTGAACCTTACC AACAAGAAGGCACTCGTCCTTTACGGATCACAAACAGGTACATCTGAG GATTATGCTCATAAGTACGCTCGTGAACGCAATCCAAGTTCAGCATT CCCACCTCTGTGGAGACTTGTGAGATTTGATTTTGACAATCTGAAC GATATCCCAGAGCAGGTCGAAGGGTTCATTTTCATCACTTTCTTCATG GCTACGTACGGCGAAGGCGAGCCTACTGACAACGCCGTAGAGTTTATC GAGTTTCTTAAGAACGACGCCGAGGATCTGTCTAATCTGAAGTATACT GTGTTTGGGTTGGGTAACTCTACTTACGAGTTCTATAATCAAATGGGT AAGACCACTAACAAAGCGTTTCTCAGAACTTGGTGCTCAATTGGTTGGA

		ACCTTTGGCGAGGGCGATGATGGTCAGGCAACTATGGACGAGGACTTC CTAGCATGGAAAAGATTCCCTGTTTGATACAATCAAGAAGGATTTGCAT TTGGAGGAACATGAAGTTGTCTACCAGCCAGGTCTAAAAGTAAAGGAG AACACAGCGCTTACAACATCTTCTCCTAATGTCTCTGTTGGTGAACCT AATAAGGCTTATGTTCTTCGTGAAGATGAGAATCTGCTGCAATACGGT CCCTTCGACCACACTCATCCTTACATTGCGCCCATCTCCAGCTCCAGA GAACTCTGTTCTGAACTTCGGAGAGAACTGTATCCATCTAGAATTT GATCTCAGCAACACCAACTTGCCTTATTCCACTGGCGATCATTGGCC GTGTGGCCAAGTAATGCTAATGAGCATGTCGAGAGCTTTTTGAAAGTG TTCAACCTCACAGACAAAAGATCTTCTGTGTTTGATATTGAGTTCCTG GACCCTACTGTCACAGTTCACCTTCCGTTTCCAACCACATACGAAGCA GTTGTTTCGTATCACCTGGAAAATTTCTGGACCAATCTCTAGACAGACA TTGAAGCAATTCATTCCCTTATGCTCCCGATCAGTCTACTAAACAGGAG GTAATCCGATTATCGGAGAGCAAAGATGTGTTCCACAACGAAGTCACA GCAAAATACTATAACCTTGCTGACCTCTTGTTCAAAGTCTCAAAAGAA ACTCCATGGAATGTTCATTCAACTTCCTGATCGAGACCATGCCTAAT CTGCAGCATCGTTACTATTCTATCTCATCATCTTCATTAAGTAAAAAG CAAATATTATATCACCGCAATGATTGAGGCCTTCACGCCTACTGGA AGTGACCATATAGTGACAGGAGTCACCACTAATTTGCTATGGAACATT CAGCTGAATCAGGACAAATCTACTGTTAAGGCGCCTGTCTCCTATGAT TTGAACGGCCCAAGAAATTTGTTTTCTCCATATAAACTTCCAGTTTAT ATCAGAAGGTCCACTTTCAAACCTTCCAAGCAACCCGGCCCTGCCAGTG ATCATGATTGGACCTGGTACTGGGGTAGCTCCATTCAGAGGCTTTATC AGAGAAAAGATGCCAGCAAGTTGACAATGGAACCTCCTAATATTGGTCAA TCTATCCTCTATTATGGCTGTAGAAATTTCTGAGCAGGATTTCTTTTAC CGTGACGAATGGCCAACATATTGAAGAAGTTAGGAGATAAGTTCAAG ATGTACACTGCGTTCAGCAGAGAAAATTCACACAAAGTTTATGTTTAC CATAGGTTATTGAAAACTCAAGGGAATTTATCGAACTGATGGACCAA GGTGCTTTTATCTACGTATGTGGAGATGCTGGTAAAATGGCTAAGGAT GTAAATAAGGCGATTGTTGAAATCCTCATTAAGAGAAAGGATTGTCA GAGGAAGATGCTACAGAAAAGTATTAGAGAGTTCAAAACTTCCAACAGA TATCAGGAAGACGTTTTGGTAA
trCYP207- trNCP1 fusion polypeptide	1280	MLNGGKQYPYINPKPFELSNKRVVQDFIENARDILTKGRSLYKDP YKAHTDLGDVLVIPPEFADALKSERQLDFTEVARDDTHGYIPGFEP IGSPFDLVPLVNKYLTRGLAKLTKPLWAEASLGVNHVLGTSTEWHPIN PGEDIMRIVSRMSSRIFMGEELCKDDDLKVSIEYTVQLFQTADEL RNPYRWTRPYIHWFLPSCQGVRRKLQEARDLLQPHIDRRNAVKK EIAIEGRPSPFDDSIWFENEYEGKSDPATEQIKLSLVAIHTTADLL SETMFNIALQPELLEPLREEIVTVLSTEGLKKTsfynLKLMDSVIK ESQRLRPLLLGAFRRMALADVTLPNGDVIKKGTKIICDTHQWNPEY YPDASKFNAYRFLQMRQTPGQDKRAHLVSTSHDQMGFGHGLHACPR GFFAANEIKIALCHMLLKYDWKLPEGVVPKSKALGMSLLGDREAK LMVKRRAAEIDIDTIGSDEARAGKEDDNSVHGVAGGFQTRDLVEI LNSTNKKALVLYGSQTGTSEDYAHKYARELQSKFSIPTLCGDLSE FDFDNLNDIPEQVEGFTFITFFMATYGEGETDNAVFEIEFLKND AEDLSNLKYTVFGLGNSTYEFYNQMGKTTNKRFSSELGAQLVG TFGEGDDGQATMDEDFLAWKDSLFDTIKKDLHLEEHEVVYQPLK VKENTALTSSPNVSVGEPNKAYVLREDENLLQYGPFDHTHPYIA PISSSRELCSSETSERNCIHLEFDLSNTNLRYSTGDHLAVWPSN ANEHVESFLKVFNLTDRSSVFDIEFLDPTVTVHFPTTYEAVVRH HLEISGPISRQTLKQFIPYAPDQSTKQEVIRLSESKDVFHNEVT AKYYNLADLLFKVSKETPWNVPFNFLIETMPNLQHRYYSISSSS LSEKQTIHITAMIEAFTPTGSDHIVTGVTNLLWNIQLNQDKSTV KAPVSYDLNGPRNLFSPYKLPVHIRRSTFKLPSNPALPVIMIGP GTGVAPFRGFI RERCQQVDNGTPNIGQSILYYGCRNSEQDFLYR DEWPTYSKKLGDKFKMYTAFSRENSHKVYVQHRLLENSREFIE LMDQGAFIYVCGDAGKMAKD VNKAIVEILIKEGLSEEDATESI REFKTSNRYQEDVW

[0242] The trCYP207-trNCP1 fusion gene of SEQ ID NO: 1280 was constructed for transformation in *Pichia* under the pAOX1 promoter and DAS2 terminator sequences expressed at the pAOX1 region of *Pichia pastoris* strain BG-45 (Δ yps, Δ aox1). Transformation of *Pichia* was carried library style and named libSHP153.

[0243] Screening of the libSHP153 library for recombinant host cells capable of the bioconversion of 3-KCA to 3-KUDCA was carried out as described in Example 7. SHP089 (CYP207 @AOX1) was used as control strain and showed 53.6 % (+/-4.5) bioconversion of 3-KCA to 3-KUDCA.

[0244] **Results**

[0245] As shown by the plot of screening results depicted in FIG. 5, libSHP153 clone_E3 showed the highest conversion, 76.9 % or 1.4 FIOPC (SHP089) and promoted to strain SHP132. A genotype summary of strain **SHP132** is depicted in FIG. 4.

Example 9: Design of Recombinant *Pichia pastoris* Host Cells that Overexpress Endogenous NCP1 Gene with CPR Activity

[0246] This example illustrates the preparation of a recombinant *Pichia* host cell that heterologously expresses the engineered CYP110 polypeptide (SEQ ID NO: 1186) with CYP activity and overexpresses the endogenous cytochrome reductase NCP1 gene with CPR activity from *Pichia*. The resulting recombinant host cells overexpresses the endogenous CPR activity and demonstrated an ability to carry out the bioconversion of Lithocholic acid (LCA) to Ursodeoxycholic acid (UDCA) and/or 3-Keto-lithocholic acid (3-KCA) to 3-Keto-ursodeoxycholic acid (3-KUDCA).

[0247] **Materials and Methods**

[0248] A. Design of engineered *Pichia* strain

[0249] The SHP083 strain of Example 7 was used as the parent strain to integrate an extra copy of the PpNCP1 gene with CPR activity into the His4 locus under the control of pAOX1 and tDAS1 terminator. The transformation was carried out library style and named libSHP155. See Figure 2 for a summary of the results.

[0250] Screening of the libSHP155 library for recombinant host cells capable of the bioconversion of 3-KCA to 3-KUDCA was carried out as described in Example 7. SHP083 was used as control strain and showed around 55 % bioconversion of 3-KCA to 3-KUDCA.

[0251] **Results**

[0252] As shown by the plot of screening results depicted in FIG. 6, the libSHP155_C3 clone showed the highest conversion, 100% (or 1.5 FIOPC). The top four clones were promoted to strains SHP143, SHP144, SHP145 and SHP146. A genotype summary of these four strains is shown in FIG. 4.

[0253] While the foregoing disclosure of the present invention has been described in some detail by way of example and illustration for purposes of clarity and understanding, this disclosure including the examples, descriptions, and embodiments described herein are for illustrative purposes, are intended to be exemplary, and should not be construed as limiting the present disclosure. It will be clear to one skilled in the art that various modifications or changes to the examples, descriptions, and embodiments described herein can be made and are to be included within the spirit and purview of this disclosure and the appended claims. Further, one of skill in the art will recognize a number of equivalent methods and procedure to those described herein. All such equivalents are to be understood to be within the scope of the present disclosure and are covered by the appended claims.

[0254] Additional embodiments of the invention are set forth in the following claims.

[0255] The disclosures of all publications, patent applications, patents, or other documents mentioned herein are expressly incorporated by reference in their entirety for all purposes to the same extent as if each such individual publication, patent, patent application or other document were individually specifically indicated to be incorporated by reference herein in its entirety for all purposes and were set forth in its entirety herein. In case of conflict, the present specification, including specified terms, will control.

CLAIMS

1. A recombinant host cell comprising a heterologous nucleic acid encoding a polypeptide with CYP activity capable of converting LCA to UDCA and/or converting 3-KCA to 3-KUDCA, wherein the polypeptide with CYP activity comprises an amino acid sequence at least 80% sequence identity to SEQ ID NO: 2 and an amino acid difference relative to SEQ ID NO: 2 at one or more positions selected from L7, S31, Q54, V56, K70, K76, I120, A143, L147, A153, N160, H170, G175, E176, V210, V261, T307, N316, Q320, G325, T333, E338, V364, L365, P401, E402, S408, Q424, L444, F452, K479, and M485.
2. The cell of claim 1, wherein the amino acid differences are selected from L7R, S31V, Q54K, V56G, K70L, K70N, K76H, I120A, A143G, A143R, L147R, L147W, A153Q, A153R, N160H, H170V, G175E, E176D, V210R, V261E, T307A, N316S, Q320M, G325E, T333V, E338Q, V364L, L365A, P401D, E402V, S408R, Q424V, L444F, F452D, K479R, and M485T.
3. The cell of any one of claims 1-2, wherein the polypeptide amino acid sequence comprises a combination of amino acid differences selected from:

S31V, Q54K, A143G, T307A, G325E
S31V, Q54K, A143G, T307A, G325E, L365A
S31V, L147R, G325E
S31V, Q54K, A143R, L147R, T307A, G325E, P401D
S31V, V57F, L147W, P401D
S31V, L147W, P401D
S31V, Q54K, A143R, T307A, P401D
S31V, Q54K, P401D
S31V, A143R, L147W
S31V, A143G, G175E, P401D, K498R
S31V, A143G, L147C, G175E, P401D
S31V, P129S, A143R
S31V, G175E, P401D
S31V, A143G
S31V, A143R
S31V, Q54K, A143G, T307A, G325E, V364L
L7R, S31V, Q54K, A143G, T307A, G325E
S31V, Q54K, A143G, A153Q, T307A, G325E
S31V, Q54K, A143G, A153P, T307A, G325E
S31V, Q54K, A143G, T307A, G325E, Q424V
S31V, Q54K, A143G, E176D, T307A, G325E
S31V, Q54K, A143G, T307A, G325E, K479R
S31V, Q54K, K70L, A143G, T307A, G325E
S31V, Q54K, A143G, T307A, G325E, T333V
S31V, Q54K, A143G, T307A, G325E, E338Q
S31V, Q54K, K70N, A143G, T307A, G325E,
S31V, Q54K, A143G, V210R, T307A, G325E,
S31V, Q54K, A143G, T307A, G325E, F452D
S31V, Q54K, A143G, T307A, Q320M, G325E,
S31V, Q54K, A143G, H170V, T307A, G325E,
S31V, Q54K, A143G, N160H, T307A, G325E,
S31V, Q54K, A143G, T307A, G325E, S408R

L7R, S31V, Q54K, A143G, V210R, T307A, G325E, V364L, K479R
L7R, S31V, Q54K, K70L, A143G, N160H, H170V, E176D, T307A, G325E, V364L,
S31V, Q54K, A143G, T307A, G325E, V364L, L444F
S31V, Q54K, A143G, T307A, G325E, V364L, Q320M, E402V
S31V, Q54K, K70L, A143G, T307A, G325E, V364L,
S31V, Q54K, A143G, V261E, T307A, Q320M, G325E, T333V, V364L, K479R
S31V, Q54K, A143G, A153Q, H170V, T307A, G325E, V364L, Q424V
S31V, Q54K, K70L, A143G, T307A, G325E, V364L, E338Q, M485T
S31V, Q54K, A143G, T307A, N316S G325E, V364L,
L7R, K479R S31V, Q54K, A143G, E176D, T307A, G325E, V364L
L7R, S31V, Q54K, A143G, T307A, Q320M, G325E, V364L, K479R
L7R, S31V, Q54K, K70L, A143G, H170V, T307A, G325E, V364L, K479R
L7R, S31V, Q54K, A143G, E176D, T307A, Q320M, G325E, V364L,
S31V, Q54K, A143G, N160H, T307A, Q320M, G325E, V364L,
L7R, S31V, Q54K, A143G, A153P, T307A, G325E, V364L, S408R

4. The cell of any one of claims 1-3, wherein the heterologous nucleic acid encoding the polypeptide further comprises a silent mutation selected from R328 (CGT > CGC), A368 (GCT > GCC), and V476 (GTT > GTC).
5. The cell of any one of claims 1-4, wherein the heterologous nucleic acid comprises:
 - (a) a sequence of at least 80% identity to a sequence selected from the group consisting of SEQ ID NO: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 1185, 1187, 1189, 1191, 1193, 1195, 1197, 1199, 1201, 1203, 1205, 1207, 1209, 1211, 1213, 1215, 1217, 1219, 1243, 1245, 1247, 1249, 1251, 1253, 1255, 1257, 1259, 1261, 1263, 1265, 1267, 1269, 1271, and 1275; or
 - (b) a codon degenerate sequence of a sequence selected from the group consisting of SEQ ID NO: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 1185, 1187, 1189, 1191, 1193, 1195, 1197, 1199, 1201, 1203, 1205, 1207, 1209, 1211, 1213, 1215, 1217, 1219, 1243, 1245, 1247, 1249, 1251, 1253, 1255, 1257, 1259, 1261, 1263, 1265, 1267, 1269, 1271, and 1275.
6. The cell of any one of claims 1-5, wherein the polypeptide with CYP activity comprises an amino acid sequence having at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity to SEQ ID NO: 2.
7. The cell of any one of claims 1-6, wherein the polypeptide with CYP activity comprises an amino acid sequence selected from SEQ ID NO: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 1186, 1188, 1190, 1192, 1194, 1196, 1198, 1200, 1202, 1204, 1206, 1208, 1210, 1212, 1214, 1216, 1218, 1220, 1244,

- 1246, 1248, 1250, 1252, 1254, 1256, 1258, 1260, 1262, 1264, 1266, 1268, 1270, and 1272; optionally, wherein the polypeptide is truncated by 1-31 amino acids at its N-terminus.
8. The cell of any one of claims 1-7, wherein the heterologous nucleic acid encodes a polypeptide with CPR activity; optionally, wherein the polypeptide with CPR activity comprises an amino acid sequence of at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity to a sequence selected from SEQ ID NO: 4 and 1278.
 9. The cell of any one of claims 1-8, wherein the heterologous nucleic acid encodes a polypeptide with CYP activity fused via a linker to a second polypeptide with CPR activity; optionally, wherein the heterologous nucleic acid encodes a polypeptide comprising an amino acid sequence of at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity to SEQ ID NO: 1280.
 10. The cell of any one of claims 1-9, wherein the heterologous nucleic acid is integrated into a site in the host cell genome; optionally integrated at two or more sites in the genome.
 11. The cell of any one of claims 1-10, wherein the heterologous nucleic acid is under the control of a promoter system selected from pGal1/10, and pCAT1:pFDH1.
 12. The cell of any one of claims 1-11, wherein the source organism of the recombinant host cell is selected from *Saccharomyces cerevisiae*, *Pichia pastoris*, *Yarrowia lipolytica*, and *Escherichia coli*.
 13. The cell of any one of claims 1-12, wherein the recombinant host cell is *Saccharomyces cerevisiae* and the heterologous nucleic acid is integrated at one or more sites in the genome selected from X-4, XI-2, XII-4, and X-2.
 14. The cell of claim 13, wherein the heterologous nucleic acid is integrated at three or more sites; optionally, integrated at four or more sites.
 15. The cell of any one of claims 1-12, wherein the recombinant host cell is *Pichia pastoris* and the heterologous nucleic acid is integrated at one or more sites in the genome selected from AOX1, Int6, Int15, and HIS4.

16. The cell of claim 15, wherein the heterologous nucleic acid is integrated at three or more sites; optionally, integrated at four or more sites.
17. A method for producing UDCA or 3-KUDCA comprising:
- (a) culturing a recombinant host cell of any one of claims 1-16 in a suitable medium comprising LCA and/or 3-KCA; and
 - (b) recovering the produced UDCA and/or 3-KUDCA.
18. A recombinant polypeptide with CYP activity capable of converting LCA to UDCA and/or converting 3-KCA to 3-KUDCA, wherein the polypeptide comprises an amino acid sequence having at least 80% sequence identity to SEQ ID NO: 2 and an amino acid difference relative to SEQ ID NO: 2 at one or more positions selected from L7, S31, Q54, V56, K70, K76, I120, A143, L147, A153, N160, H170, G175, E176, V210, V261, T307, N316, Q320, G325, T333, E338, V364, L365, P401, E402, S408, Q424, L444, F452, K479, and M485.
19. The polypeptide of claim 18, wherein the amino acid differences are selected from L7R, S31V, Q54K, V56G, K70L, K70N, K76H, I120A, A143G, A143R, L147R, L147W, A153Q, A153R, N160H, H170V, G175E, E176D, V210R, V261E, T307A, N316S, Q320M, G325E, T333V, E338Q, V364L, L365A, P401D, E402V, S408R, Q424V, L444F, F452D, K479R, and M485T.
20. The polypeptide of any one of claims 18-19, wherein the polypeptide amino acid sequence comprises a combination of amino acid differences selected from:

S31V, Q54K, A143G, T307A, G325E
S31V, Q54K, A143G, T307A, G325E, L365A
S31V, L147R, G325E
S31V, Q54K, A143R, L147R, T307A, G325E, P401D
S31V, V57F, L147W, P401D
S31V, L147W, P401D
S31V, Q54K, A143R, T307A, P401D
S31V, Q54K, P401D
S31V, A143R, L147W
S31V, A143G, G175E, P401D, K498R
S31V, A143G, L147C, G175E, P401D
S31V, P129S, A143R
S31V, G175E, P401D
S31V, A143G
S31V, A143R
S31V, Q54K, A143G, T307A, G325E, V364L
L7R, S31V, Q54K, A143G, T307A, G325E
S31V, Q54K, A143G, A153Q, T307A, G325E
S31V, Q54K, A143G, A153P, T307A, G325E
S31V, Q54K, A143G, T307A, G325E, Q424V
S31V, Q54K, A143G, E176D, T307A, G325E
S31V, Q54K, A143G, T307A, G325E, K479R
S31V, Q54K, K70L, A143G, T307A, G325E
S31V, Q54K, A143G, T307A, G325E, T333V

S31V, Q54K, A143G, T307A, G325E, E338Q
S31V, Q54K, K70N, A143G, T307A, G325E,
S31V, Q54K, A143G, V210R, T307A, G325E,
S31V, Q54K, A143G, T307A, G325E, F452D
S31V, Q54K, A143G, T307A, Q320M, G325E,
S31V, Q54K, A143G, H170V, T307A, G325E,
S31V, Q54K, A143G, N160H, T307A, G325E,
S31V, Q54K, A143G, T307A, G325E, S408R
L7R, S31V, Q54K, A143G, V210R, T307A, G325E, V364L, K479R
L7R, S31V, Q54K, K70L, A143G, N160H, H170V, E176D, T307A, G325E, V364L
S31V, Q54K, A143G, T307A, G325E, V364L, L444F
S31V, Q54K, A143G, T307A, G325E, V364L, Q320M, E402V
S31V, Q54K, K70L, A143G, T307A, G325E, V364L,
S31V, Q54K, A143G, V261E, T307A, Q320M, G325E, T333V, V364L, K479R
S31V, Q54K, A143G, A153Q, H170V, T307A, G325E, V364L, Q424V
S31V, Q54K, K70L, A143G, T307A, G325E, V364L, E338Q, M485T
S31V, Q54K, A143G, T307A, N316S G325E, V364L,
L7R, K479R S31V, Q54K, A143G, E176D, T307A, G325E, V364L
L7R, S31V, Q54K, A143G, T307A, Q320M, G325E, V364L, K479R
L7R, S31V, Q54K, K70L, A143G, H170V, T307A, G325E, V364L, K479R
L7R, S31V, Q54K, A143G, E176D, T307A, Q320M, G325E, V364L,
S31V, Q54K, A143G, N160H, T307A, Q320M, G325E, V364L,
L7R, S31V, Q54K, A143G, A153P, T307A, G325E, V364L, S408R

21. The polypeptide of any one of claims 18-20, wherein the polypeptide comprises an amino acid sequence having at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity to SEQ ID NO: 2.
22. The polypeptide of any one of claims 18-21, wherein the polypeptide comprises an amino acid sequence selected from SEQ ID NO: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 1186, 1188, 1190, 1192, 1194, 1196, 1198, 1200, 1202, 1204, 1206, 1208, 1210, 1212, 1214, 1216, 1218, 1220, 1244, 1246, 1248, 1250, 1252, 1254, 1256, 1258, 1260, 1262, 1264, 1266, 1268, 1270, 1272, and 1276.
23. The polypeptide of any one of claims 18-22, wherein the polypeptide comprises an N-terminal truncation of from 1 to 31 amino acids as compared to SEQ ID NO: 2; optionally, wherein the polypeptide comprises an amino acid of SEQ ID NO: 1276.
24. The polypeptide of any one of claims 18-23, wherein the polypeptide is fused via a linker to a second polypeptide; optionally, wherein the second polypeptide has CPR activity.
25. The polypeptide of claim 24, wherein the second polypeptide with CPR activity comprises an amino acid sequence having at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least

91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity to a sequence selected from SEQ ID NO: 4 and 1278; optionally, wherein the polypeptide comprises an amino acid sequence of SEQ ID NO: 1280.

26. A polynucleotide encoding the polypeptide of any one of claims 18-22.

27. The polynucleotide of claim 26 in which the polynucleotide sequence comprises:

(a) a sequence of at least 80% identity to a sequence selected from the group consisting of SEQ ID NO: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 1185, 1187, 1189, 1191, 1193, 1195, 1197, 1199, 1201, 1203, 1205, 1207, 1209, 1211, 1213, 1215, 1217, 1219, 1243, 1245, 1247, 1249, 1251, 1253, 1255, 1257, 1259, 1261, 1263, 1265, 1267, 1269, 1271, and 1275; or

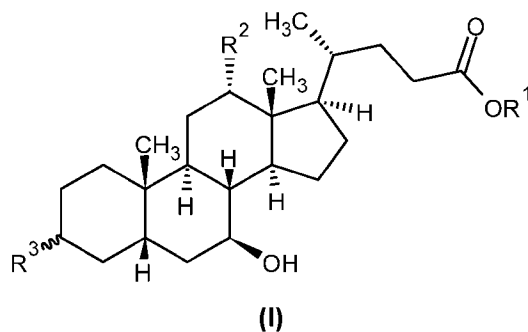
(b) a codon degenerate sequence of a sequence selected from the group consisting of SEQ ID NO: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 1185, 1187, 1189, 1191, 1193, 1195, 1197, 1199, 1201, 1203, 1205, 1207, 1209, 1211, 1213, 1215, 1217, 1219, 1243, 1245, 1247, 1249, 1251, 1253, 1255, 1257, 1259, 1261, 1263, 1265, 1267, 1269, 1271, and 1275.

28. An expression vector comprising the polynucleotide of any one of claims 26-27.

29. The expression vector of claim 28 comprising a control sequence.

30. A host cell comprising the polynucleotide of any one of claims 26-27 or the expression vector of any one of claims 28-29.

31. A method for preparing a compound of structural formula (I)



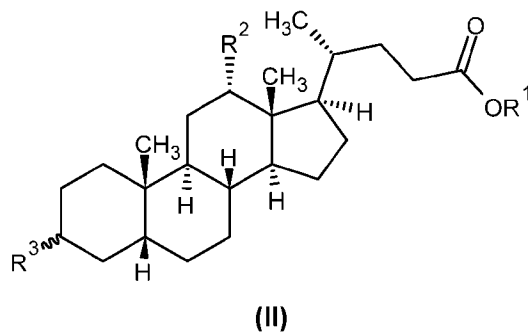
wherein,

R¹ is hydrogen, a carboxylate salt counterion, an optionally substituted C1-C20 alkyl, or an optionally substituted aryl;

R² is hydrogen, or a hydroxyl;

R³ is a hydroxyl, or an oxo group;

the method comprising contacting under suitable reactions conditions a compound of structural formula (II)



wherein,

R¹ is hydrogen, a carboxylate salt counterion, an optionally substituted C1-C20 alkyl, or an optionally substituted aryl;

R² is hydrogen, or a hydroxyl;

R³ is a hydroxyl, or an oxo group;

and a recombinant polypeptide of any one of claims 18-22.

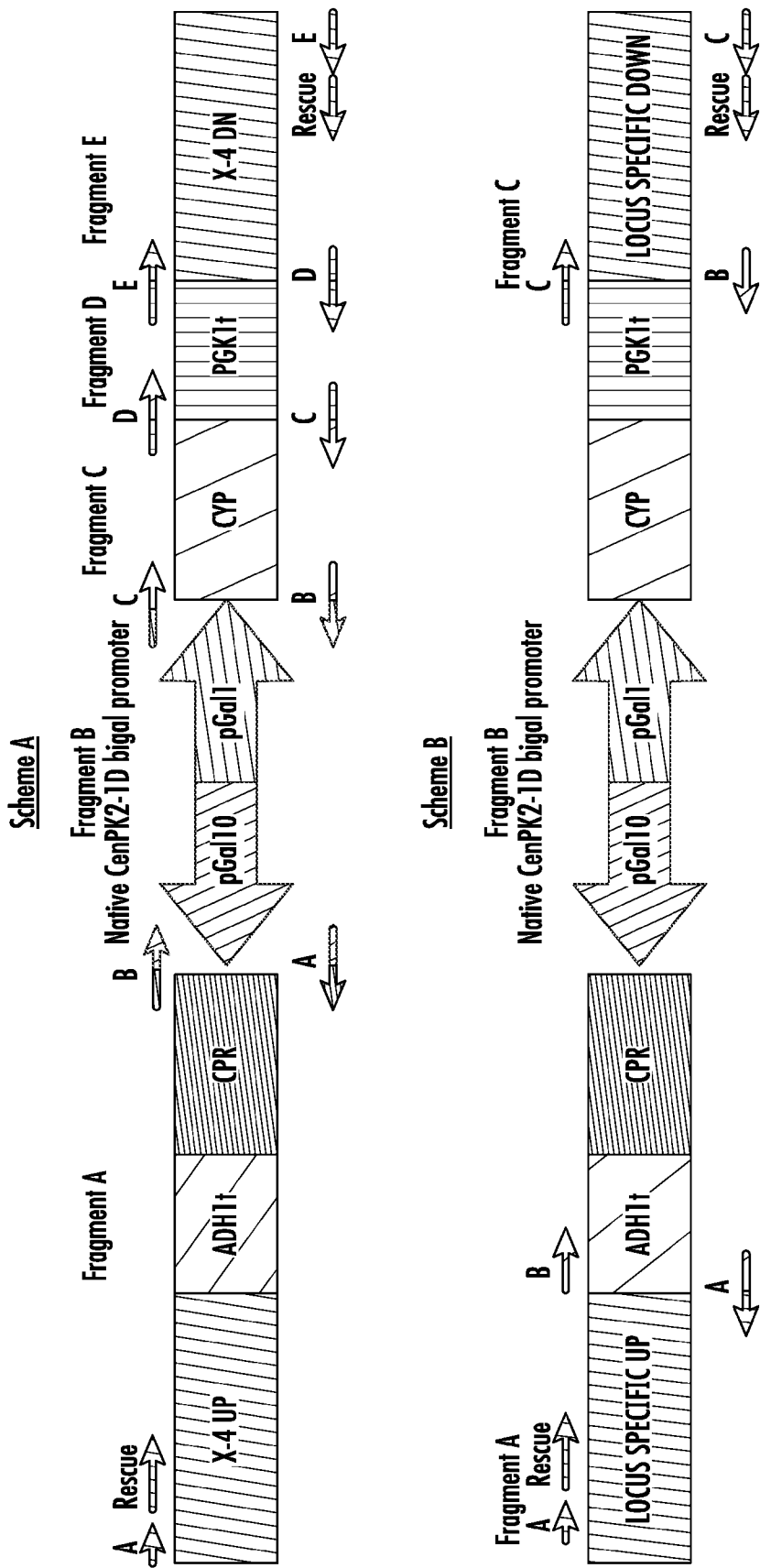
32. The method of claim 31, wherein R² is hydrogen.

33. The method of claim 32, wherein:

- (a) the compound of structure formula (I) is UDCA and the compound of structural formula (II) is LCA; or
- (b) the compound of structure formula (I) is 3-KUDCA and the compound of structural formula (II) is 3-KCA.

34. The method of claim 31, wherein R² is a hydroxyl.

35. The method of claim 34, wherein the compound of structure formula (I) is UCA and the compound of structural formula (II) is DCA.



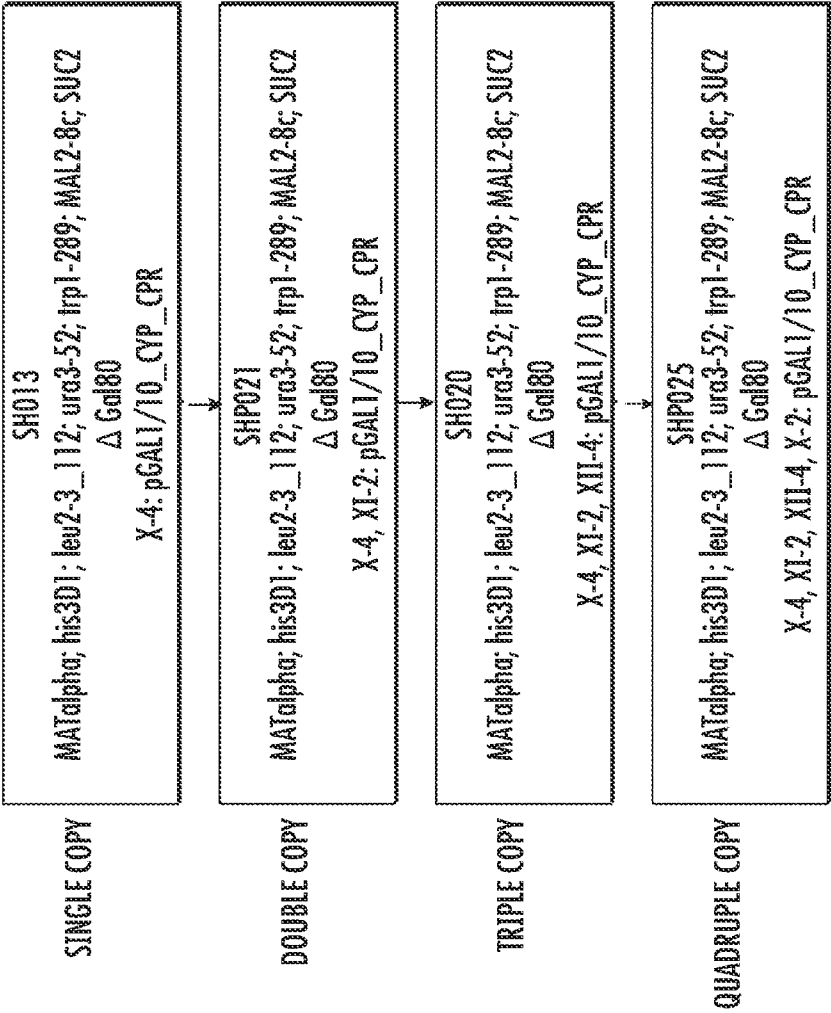


FIG. 2

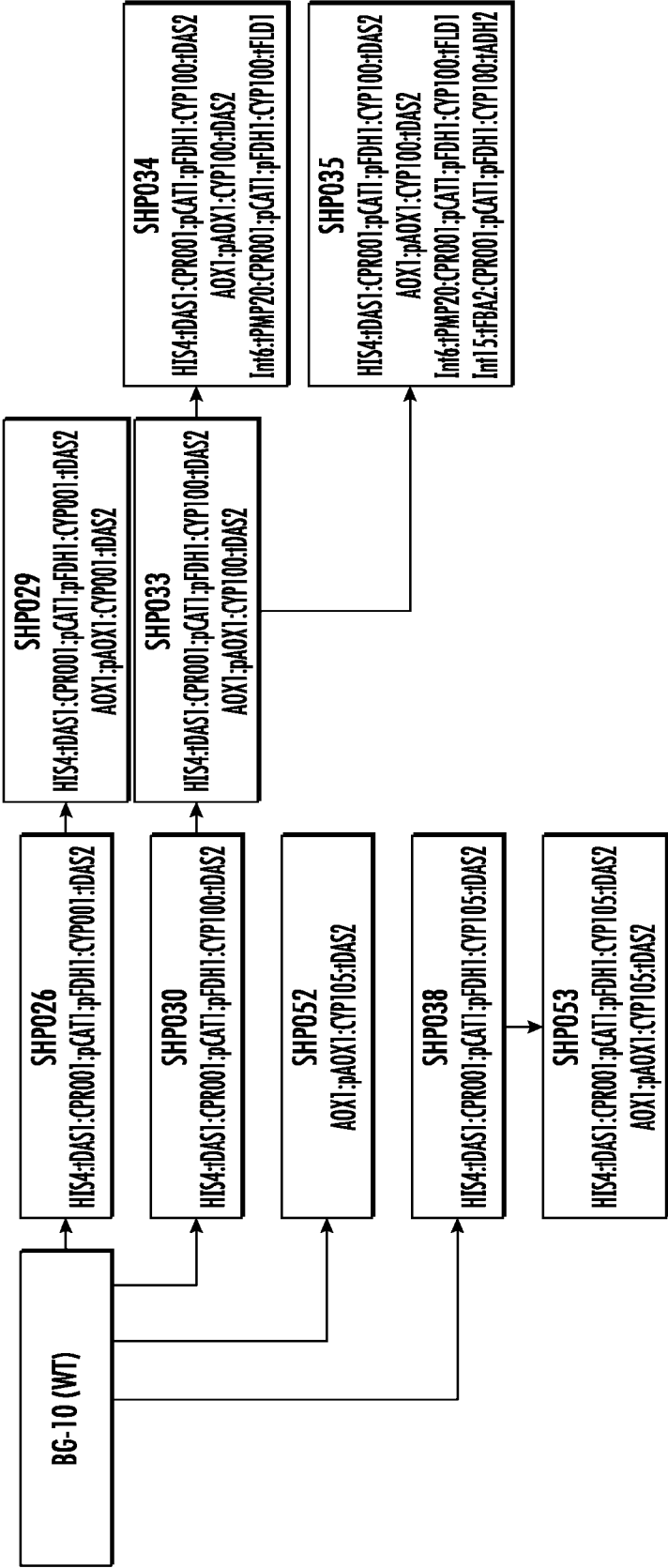


FIG. 3

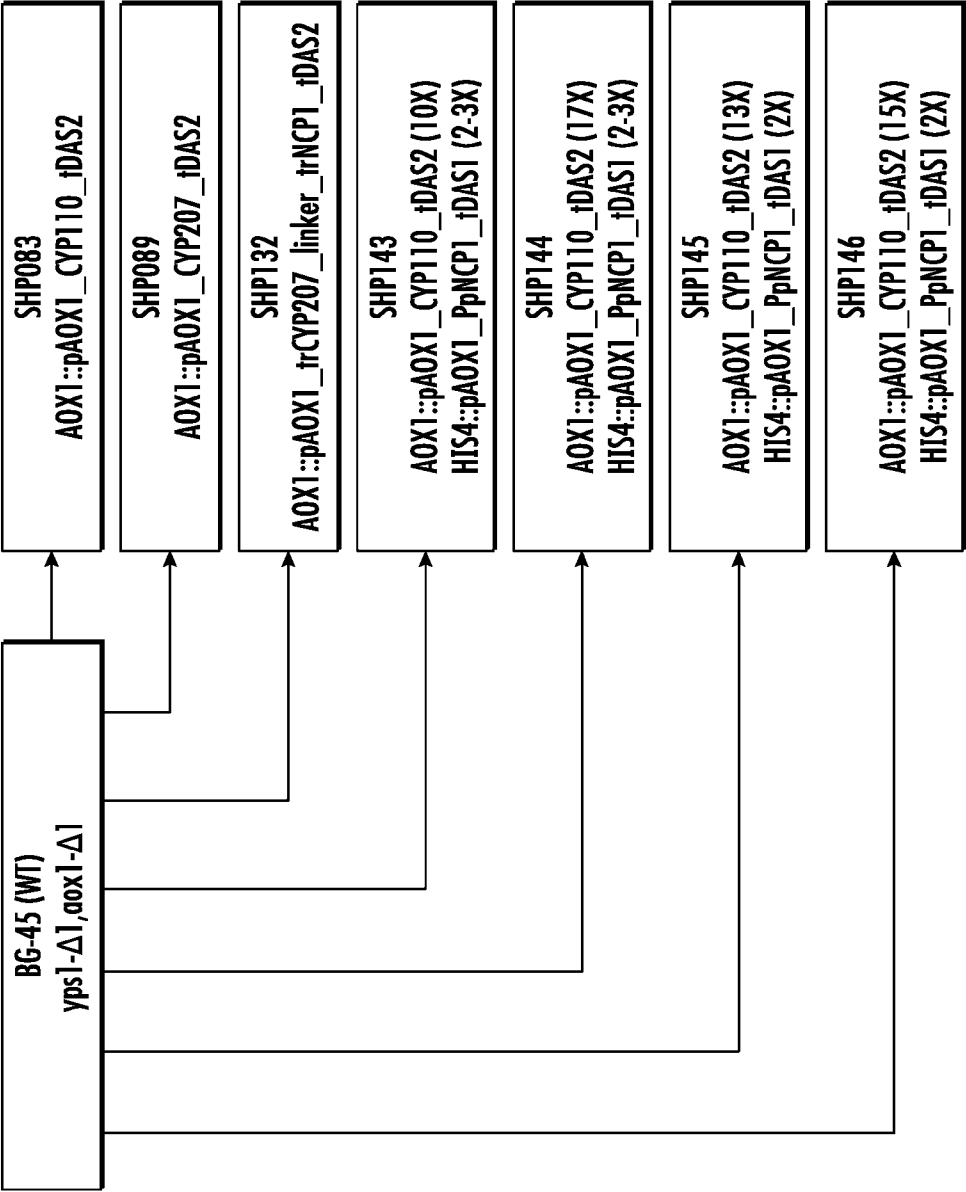


FIG. 4

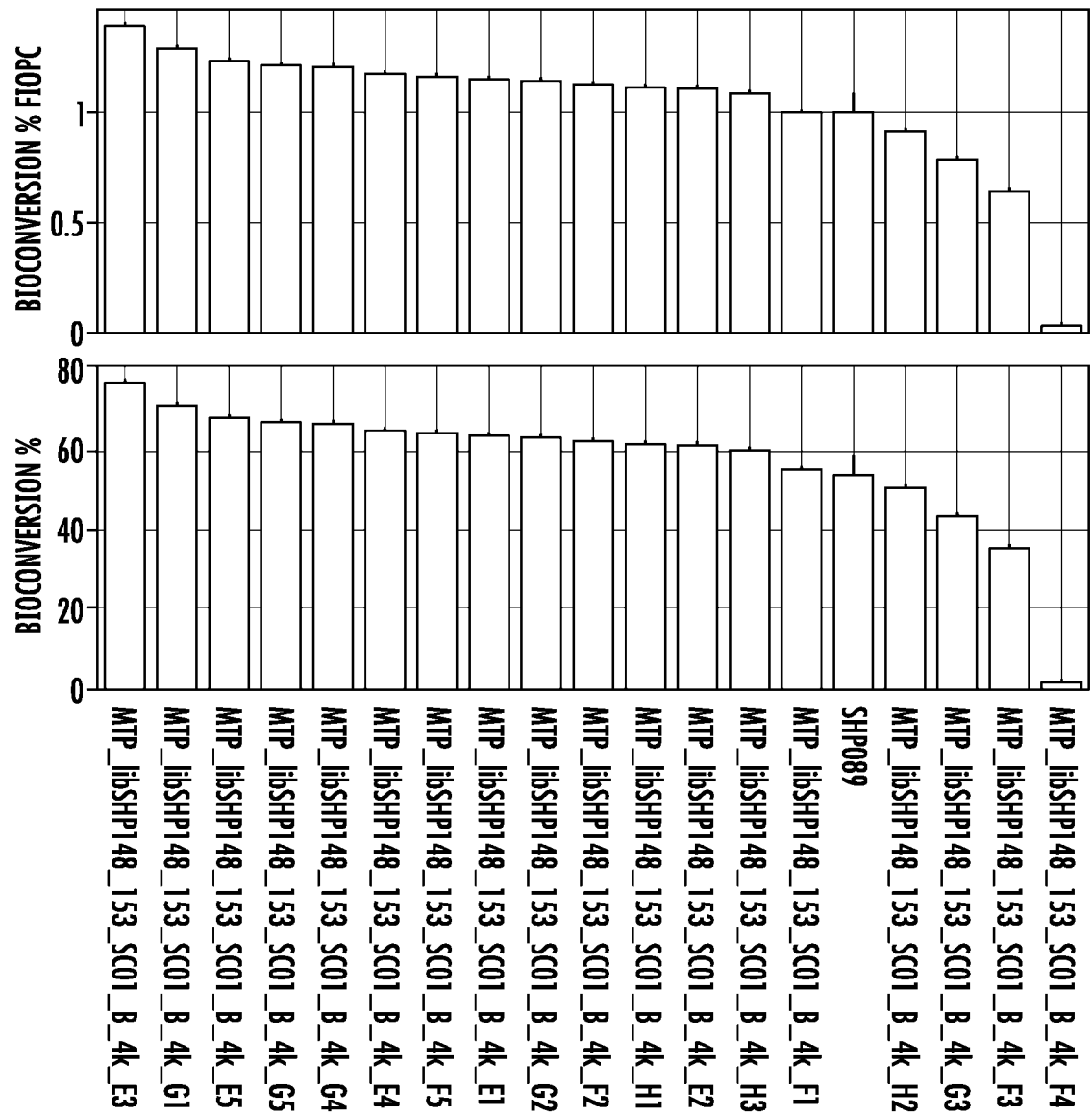


FIG. 5

6/6

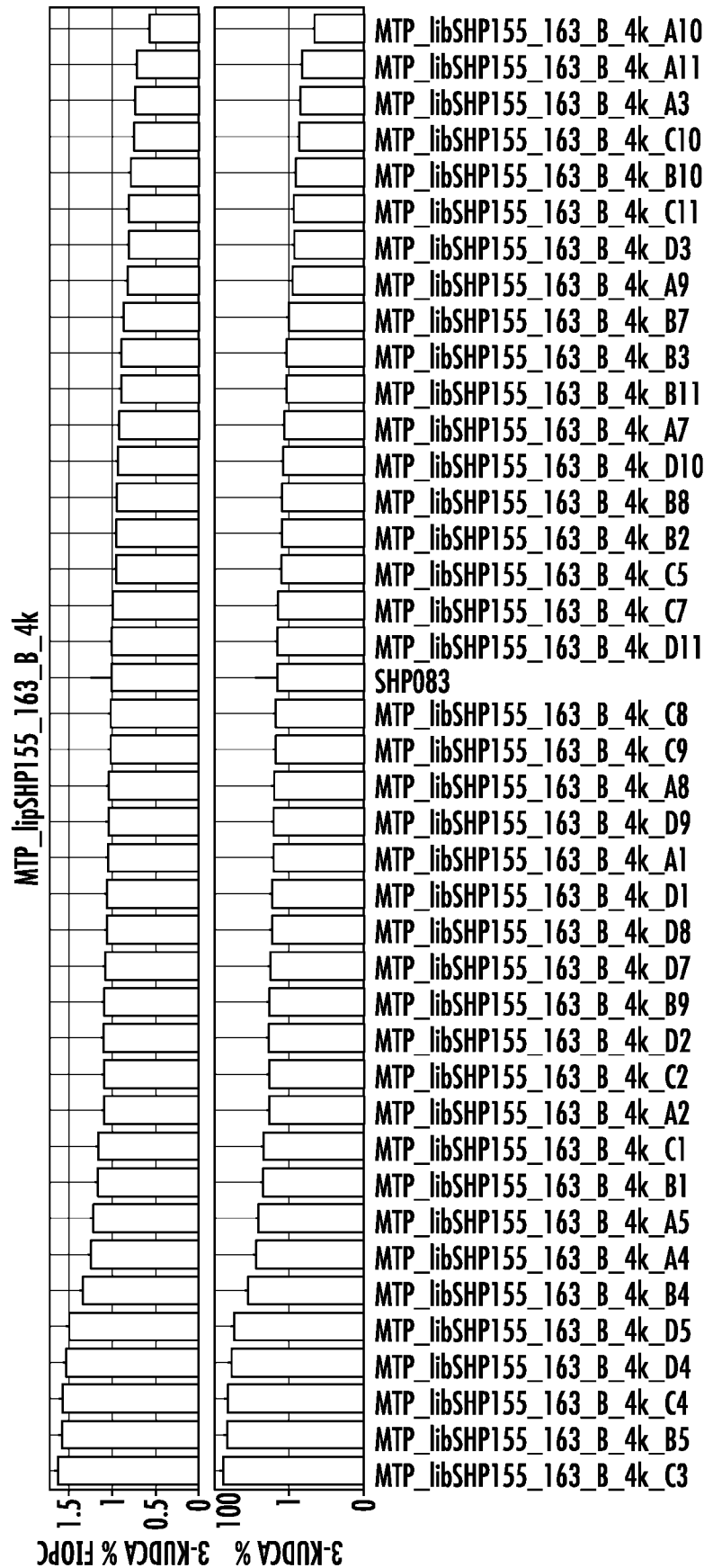


FIG. 6