(12) STANDARD PATENT (11) Application No. **AU 2018256863 B2**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Microorganisms and methods for producing cannabinoids and cannabinoid derivatives

(51) International Patent Classification(s)
C12N 9/10 (2006.01) ***C12P 7/22*** (2006.01)
C12N 15/52 (2006.01) ***C12P 17/06*** (2006.01)

(21) Application No: **2018256863** (22) Date of Filing: **2018.04.27**

(87) WIPO No: **WO18/200888**

(30) Priority Data

(31) Number	(32) Date	(33) Country
62/491,114	**2017.04.27**	**US**
62/569,532	**2017.10.07**	**US**

(43) Publication Date: **2018.11.01**
(44) Accepted Journal Date: **2024.06.06**

(71) Applicant(s)
Regents of the University of California

(72) Inventor(s)
Keasling, Jay D.;D'espaux, Leo;Wong, Jeff;Luo, Xiaozhou;Reiter, Michael;Denby, Charles;Lechner, Anna

(74) Agent / Attorney
FOUNDRY INTELLECTUAL PROPERTY PTY LTD, PO Box Q1466, Queen Victoria Building, NSW, 1230, AU

(56) Related Art
DATABASE EMBL [online] 16 October 2011 (2011-10-16), "TSA: Cannabis sativa PK15523.1_1.CasaPuKu mRNA sequence.", XP002782462, retrieved from EBI accession no. EM_TSA:JP460119 Database accession no. JP460119

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
01 November 2018 (01.11.2018)

WIPO | PCT

(10) International Publication Number
WO 2018/200888 A1

(51) International Patent Classification:
C12N 9/10 (2006.01) *C12P 7/22* (2006.01)
C12N 15/52 (2006.01) *C12P 17/06* (2006.01)

(21) International Application Number:
PCT/US2018/029668

(22) International Filing Date:
27 April 2018 (27.04.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/491,114 27 April 2017 (27.04.2017) US
62/569,532 07 October 2017 (07.10.2017) US

(71) Applicant: **REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 1111 Franklin Street, Oakland, California 94607 (US).

(72) Inventors: **KEASLING, Jay D.**; 2150 Shattuck Ave., Suite 510, Berkeley, California 94704 (US). **D'ESPAUX, Leo**; 2150 Shattuck Ave., Suite 510, Berkeley, California 94704 (US). **WANG, Jeff**; 2150 Shattuck Ave., Suite 510, Berkeley, California 94704 (US). **LUO, Xiaozhou**; 2150 Shattuck Ave., Suite 510, Berkeley, California 94704 (US). **REITER, Michael**; 2150 Shattuck Ave., Suite 510, Berkeley, California 94704 (US). **DENBY, Charles**; 2150 Shattuck Ave., Suite 510, Berkeley, California 94704 (US). **LECHNER, Anna**; 2150 Shattuck Ave., Suite 510, Berkeley, California 94704 (US).

(74) Agent: **OSMAN, Richard Aron**; 530 Lawrence Expy #332, Sunnyvale, California 94085 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:
— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

WO 2018/200888 A1

(54) Title: MICROORGANISMS AND METHODS FOR PRODUCING CANNABINOIDS AND CANNABINOID DERIVATIVES

(57) Abstract: The present disclosure provides genetically modified host cells that produce a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative. The present disclosure provides methods of synthesizing a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative.

2018256863 11 Apr 2024

MICROORGANISMS AND METHODS FOR PRODUCING CANNABINOIDS AND CANNABINOID DERIVATIVES

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 62/491,114, filed April 27, 2017, and U.S. Provisional Application No. 62/569,532, filed October 7, 2017, the contents of which are incorporated herein by reference in their entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

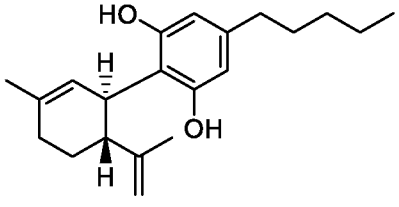
[0002] This invention was made with government support under Grant Numbers 1330914 and 1442724 awarded by the National Science Foundation. The government has certain rights in the invention.

REFERENCE TO A SEQUENCE LISTING

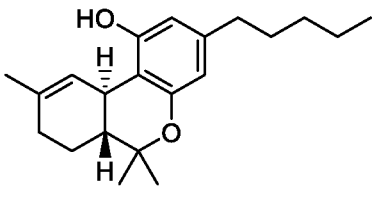
[0003] The Sequence Listing associated with this application is provided in text format in lieu of a paper copy, and is hereby incorporated by reference into the specification. The name of the text file containing the Sequence Listing is DEMT_001_00WOSeqList_ST25_txt. The text file is about 673 KB, was created on April 25, 2018, and is being submitted electronically via EFS-Web.

INTRODUCTION

[0004] Plants from the genus *Cannabis* have been used by humans for their medicinal properties for thousands of years. In modern times, the bioactive effects of *Cannabis* are attributed to a class of compounds termed "cannabinoids," of which there are hundreds of structural analogs including tetrahydrocannabinol (THC) and cannabidiol (CBD). These molecules and preparations of *Cannabis* material have recently found application as therapeutics for chronic pain, multiple sclerosis, cancer-associated nausea and vomiting, weight loss, appetite loss, spasticity, and other conditions.



Cannabidiol / CBD



Tetrahydrocannabinol / THC / Dronabinol / Marinol

2018256863 11 Apr 2024

[0005] The physiological effects of certain cannabinoids are thought to be mediated by their interaction with two cellular receptors found in humans and other animals. Cannabinoid receptor type 1 (CB1) is common in the brain, the reproductive system, and the eye. Cannabinoid receptor type 2 (CB2) is common in the immune system and mediates therapeutic effects related to inflammation in animal models. The discovery of cannabinoid receptors and their interactions with plant-derived cannabinoids predated the identification of endogenous ligands.

[0006] Besides THC and CBD, hundreds of other cannabinoids have been identified in *Cannabis*. However, many of these compounds exist at low levels and alongside more abundant cannabinoids, making it difficult to obtain pure samples from plants to study their therapeutic potential. Similarly, methods of chemically synthesizing these types of products has been cumbersome and costly, and tends to produce insufficient yield. Accordingly, additional methods of making pure cannabinoids, cannabinoid precursors, cannabinoid derivatives, or cannabinoid precursor derivatives are needed.

SUMMARY

[0007] The present disclosure provides methods, polypeptides, nucleic acids encoding said polypeptides, and genetically modified host cells for the production of cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives.

[0008] In a first aspect, the present invention provides a genetically modified yeast cell for producing a cannabinoid or a cannabinoid derivative, the genetically modified yeast cell comprising one or more heterologous nucleic acids integrated into a chromosome of the genetically modified yeast cell and encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:110 or SEQ ID NO:100.

[0009] In a second aspect, the present invention provides a method of producing a cannabinoid or a cannabinoid derivative, the method comprising:

a) culturing a genetically modified yeast cell comprising:

i) one or more heterologous nucleic acids integrated into a chromosome of the genetically modified yeast cell and encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:110 or SEQ ID NO:100; and

ii) one or more heterologous nucleic acids integrated into a chromosome of the genetically modified yeast cell and encoding a tetraketide synthase (TKS) polypeptide comprising the amino acid sequence of SEQ ID NO:11; and/or

iii) one or more heterologous nucleic acids integrated into a chromosome of the genetically modified yeast cell and encoding an olivetolic acid cyclase (OAC) polypeptide comprising the amino acid sequence of SEQ ID NO:10;

in a suitable medium containing a carboxylic acid; and

b) recovering the produced cannabinoid or cannabinoid derivative.

[0010] In a third aspect, the present invention provides a method of producing a cannabinoid or cannabinoid derivative, the method comprising:

a) culturing a genetically modified yeast cell comprising:

i) one or more heterologous nucleic acids integrated into a chromosome of the genetically modified yeast cell and encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:110 or SEQ ID NO:100;

ii) one or more heterologous nucleic acids integrated into a chromosome of the genetically modified yeast cell and encoding a tetraketide synthase (TKS) polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:11;

iii) one or more heterologous nucleic acids integrated into a chromosome of the genetically modified yeast cell and encoding an olivetolic acid cyclase (OAC) polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:10;

iv) one or more heterologous nucleic acids encoding one or more polypeptides that generate geranyl pyrophosphate, wherein the polypeptide that generates geranyl pyrophosphate is a geranyl pyrophosphate synthetase (GPPS) polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:60;

v) one or more heterologous nucleic acids encoding one or more polypeptides that generate an acyl-CoA compound or an acyl-CoA compound derivative wherein the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is an acyl-activating enzyme (AAE) polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:90, SEQ ID NO:92, or SEQ ID NO:149; a fatty acyl-CoA ligase polypeptide comprising an amino acid

sequence having at least 85% sequence identity to SEQ ID NO:145 or SEQ ID NO:147; or a fatty acyl-CoA synthetase (FAA) polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; and

vi) one or more heterologous nucleic acids encoding one or more polypeptides that generate malonyl-CoA, wherein the polypeptide that generates malonyl-CoA is an acetyl-CoA carboxylase-1 (ACC1) polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207;

in a suitable medium; and

b) recovering the produced cannabinoid or cannabinoid derivative,

wherein, optionally, the genetically modified yeast cell comprises one or more heterologous nucleic acids encoding a THCA synthase polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:155 or a CBDA synthase polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:88 or SEQ ID NO:151; and/or

wherein at least one of the heterologous nucleic acids integrated into a chromosome of the genetically modified yeast cell is operably linked to an inducible promoter.

[0011] In a fourth aspect, the present invention provides a method of producing a cannabinoid or a cannabinoid derivative, the method comprising use of a genetically modified yeast cell comprising a heterologous nucleic acid integrated into a chromosome of the genetically modified yeast cell and encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:110 or SEQ ID NO: 100.

[0012] One aspect of the disclosure relates to a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, the genetically modified host cell comprising one or more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase (GOT) polypeptide, wherein said GOT polypeptide catalyzes production of cannabigerolic acid from geranyl pyrophosphate (GPP) and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[0013] Another aspect of the disclosure relates to a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, the genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110.

[0014] One aspect of the disclosure relates to a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, the genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100.

[0015] In certain embodiments of any of the foregoing or following, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a tetraketide synthase (TKS) polypeptide and one or more heterologous nucleic acids encoding an olivetolic acid cyclase (OAC) polypeptide, or one or more heterologous nucleic acids encoding a fusion TKS and OAC polypeptide. In some embodiments, the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76. In some embodiments, the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[0016] In certain embodiments of any of the foregoing or following, the genetically modified host cell further comprises one or more of the following: a) one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; b) one or more heterologous nucleic acids encoding a polypeptide that generates GPP; or c) one or more heterologous nucleic acids encoding a polypeptide that generates malonyl-CoA. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative, wherein the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is an acyl-activating enzyme (AAE) polypeptide. In some embodiments, the AAE polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90. In some embodiments, the AAE polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:92 or SEQ ID NO:149. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative, wherein the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is a fatty acyl-CoA ligase polypeptide. In some embodiments, the fatty acyl-CoA ligase polypeptide comprises an amino acid sequence having at least 50%

sequence identity to SEQ ID NO:145 or SEQ ID NO:147. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative, wherein the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is a fatty acyl-CoA synthetase (FAA) polypeptide. In some embodiments, the FAA polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that generates GPP, wherein the polypeptide that generates GPP is a geranyl pyrophosphate synthetase (GPPS) polypeptide. In some embodiments, the GPPS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that generates malonyl-CoA, wherein the polypeptide that generates malonyl-CoA is an acetyl-CoA carboxylase-1 (ACC1) polypeptide. In some embodiments, the ACC1 polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207.

[0017] In certain embodiments of any of the foregoing or following, the genetically modified host cell further comprises one or more of the following: a) one or more heterologous nucleic acids encoding a HMG-CoA synthase (HMGS) polypeptide; b) one or more heterologous nucleic acids encoding a 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMGR) polypeptide; c) one or more heterologous nucleic acids encoding a mevalonate kinase (MK) polypeptide; d) one or more heterologous nucleic acids encoding a phosphomevalonate kinase (PMK) polypeptide; e) one or more heterologous nucleic acids encoding a mevalonate pyrophosphate decarboxylase (MVD) polypeptide; or f) one or more heterologous nucleic acids encoding a isopentenyl diphosphate isomerase (IDI) polypeptide. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an IDI polypeptide. In some embodiments, the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an HMGR polypeptide. In some embodiments, the HMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:22. In some embodiments, the genetically modified host cell further

comprises one or more heterologous nucleic acids encoding an HMGR polypeptide, wherein the HMGR polypeptide is a truncated HMGR (tHMGR) polypeptide. In some embodiments, the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an HMGS polypeptide. In some embodiments, the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an MK polypeptide. In some embodiments, the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a PMK polypeptide. In some embodiments, the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a MVD polypeptide. In some embodiments, the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66.

[0018] In certain embodiments of any of the foregoing or following, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA. In some embodiments, the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide. In some embodiments, the acetoacetyl-CoA thiolase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25.

[0019] In certain embodiments of any of the foregoing or following, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a pyruvate decarboxylase (PDC) polypeptide. In some embodiments, the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117.

[0020] In certain embodiments of any of the foregoing or following, the genetically modified host cell is a eukaryotic cell. In some embodiments, the eukaryotic cell is a yeast cell. In some embodiments, the yeast cell is *Saccharomyces cerevisiae*. In some

embodiments, the *Saccharomyces cerevisiae* is a protease-deficient strain of *Saccharomyces cerevisiae*. In some embodiments, the genetically modified host cell is a plant cell.

[0021] In certain embodiments of any of the foregoing or following, the genetically modified host cell is a prokaryotic cell.

[0022] In certain embodiments of any of the foregoing or following, at least one of the one or more heterologous nucleic acids is integrated into the chromosome of the genetically modified host cell.

[0023] In certain embodiments of any of the foregoing or following, at least one of the one or more heterologous nucleic acids is maintained extrachromosomally.

[0024] In certain embodiments of any of the foregoing or following, two or more of the one or more heterologous nucleic acids are present in a single expression vector.

[0025] In certain embodiments of any of the foregoing or following, at least one of the heterologous nucleic acids is operably linked to an inducible promoter.

[0026] In certain embodiments of any of the foregoing or following, at least one of the heterologous nucleic acids is operably linked to a constitutive promoter.

[0027] In certain embodiments of any of the foregoing or following, culturing of the genetically modified host cell in a suitable medium provides for synthesis of the cannabinoid or the cannabinoid derivative in an increased amount compared to a non-genetically modified host cell cultured under similar conditions.

[0028] In certain embodiments of any of the foregoing or following, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a cannabinoid synthase polypeptide. In some embodiments, the cannabinoid synthase polypeptide is a tetrahydrocannabinolic acid (THCA) synthase polypeptide. In some embodiments, the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155. In some embodiments, the cannabinoid synthase polypeptide is a cannabidiolic acid (CBDA) synthase polypeptide. In some embodiments, the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

[0029] In certain embodiments of any of the foregoing or following, the cannabinoid is cannabigerolic acid, cannabigerol, Δ^9-tetrahydrocannabinolic acid, Δ^9-tetrahydrocannabinol, Δ^8-tetrahydrocannabinolic acid, Δ^8-tetrahydrocannabinol, cannabidiolic acid, cannabidiol, cannabichromenic acid, cannabichromene, cannabinolic acid, cannabinol, cannabidivarinic acid, cannabidivarin, tetrahydrocannabivarinic acid, tetrahydrocannabivarin,

cannabichromevarinic acid, cannabichromevarin, cannabigerovarinic acid, cannabigerovarin, cannabicyclolic acid, cannabicyclol, cannabielsoinic acid, cannabielsoin, cannabicitranic acid, or cannabicitran.

[0030] One aspect of the disclosure relates to a method of producing a cannabinoid or a cannabinoid derivative in a genetically modified host cell, the method comprising: a) culturing the genetically modified host cell in a suitable medium; and b) recovering the produced cannabinoid or cannabinoid derivative.

[0031] Another aspect of the disclosure relates to a method of producing a cannabinoid or a cannabinoid derivative in a genetically modified host cell, the method comprising: a) culturing the genetically modified host cell in a suitable medium comprising a carboxylic acid; b) recovering the produced cannabinoid or cannabinoid derivative.

[0032] One aspect of the disclosure relates to a method of producing a cannabinoid or a cannabinoid derivative in a genetically modified host cell, the method comprising: a) culturing the genetically modified host cell in a suitable medium comprising olivetolic acid or an olivetolic acid derivative; b) recovering the produced cannabinoid or cannabinoid derivative.

[0033] Another aspect of the disclosure relates to a method of producing a cannabinoid or a cannabinoid derivative in a genetically modified host cell, the method comprising: a) culturing a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide catalyzes production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, in a suitable medium; and b) recovering the produced cannabinoid or cannabinoid derivative.

[0034] One aspect of the disclosure relates to a method of producing a cannabinoid or a cannabinoid derivative in a genetically modified host cell, the method comprising: a) culturing a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110 in a suitable medium; and b) recovering the produced cannabinoid or cannabinoid derivative.

[0035] Another aspect of the disclosure relates to a method of producing a cannabinoid or a cannabinoid derivative in a genetically modified host cell, the method comprising: a) culturing a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65%

sequence identity to SEQ ID NO:100 in a suitable medium; and b) recovering the produced cannabinoid or cannabinoid derivative.

[0036] In certain embodiments of any of the foregoing or following, the suitable medium comprises a fermentable sugar. In some embodiments, the suitable medium comprises a pretreated cellulosic feedstock.

[0037] In certain embodiments of any of the foregoing or following, the suitable medium comprises a non-fermentable carbon source. In some embodiments, the non-fermentable carbon source comprises ethanol.

[0038] One aspect of the disclosure relates to an isolated or purified GOT polypeptide, wherein said GOT polypeptide catalyzes production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[0039] Another aspect of the disclosure relates to an isolated or purified polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110.

[0040] One aspect of the disclosure relates to an isolated or purified polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100.

[0041] Another aspect of the disclosure relates to an isolated or purified nucleic acid encoding a GOT polypeptide, wherein said GOT polypeptide catalyzes production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[0042] One aspect of the disclosure relates to an isolated or purified nucleic acid encoding a polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110.

[0043] Another aspect of the disclosure relates to an isolated or purified nucleic acid encoding a polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100.

[0044] One aspect of the disclosure relates to a vector comprising a nucleic acid encoding a GOT polypeptide, wherein said GOT polypeptide catalyzes production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[0045] Another aspect of the disclosure relates to a vector comprising a nucleic acid encoding a polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110.

[0046] One aspect of the disclosure relates to a vector comprising a nucleic acid encoding a polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100.

[0047] Another aspect of the disclosure relates to a method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide catalyzes production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, into the genetically modified host cell.

[0048] One aspect of the disclosure relates to a method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110 into the genetically modified host cell.

[0049] Another aspect of the disclosure relates to a method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100 into the genetically modified host cell.

[0050] One aspect of the disclosure relates to a method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing a vector comprising a nucleic acid encoding a GOT polypeptide, wherein said GOT polypeptide catalyzes production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82; a vector comprising a nucleic acid encoding a polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110; or a vector comprising a nucleic acid encoding a polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100, into the genetically modified host cell.

BRIEF DESCRIPTION OF THE DRAWINGS

[0051] FIG. 1 provides a schematic diagram of biosynthetic pathways for generating cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives.

[0052] FIG. 2 depicts intracellular olivetolic acid production using pathway 1a and a tetraketide synthase (TKS) polypeptide/olivetolic acid cyclase (OAC) polypeptide.

[0053] FIG. 3 depicts intracellular olivetolic acid production comparing pathway 1a and 1b.

[0054] FIG. 4 provides schematic depictions of 3 expression constructs for olivetolic acid production.

[0055] FIG. 5 provides schematic depictions of 2 expression constructs for olivetolic acid production.

[0056] FIG. 6 provides schematic depictions of 3 expression constructs for geranyl pyrophosphate (GPP) production.

[0057] FIG. 7 provides schematic depictions of 2 expression constructs for GPP production.

[0058] FIG. 8 provides schematic depictions of 3 expression constructs for cannabinoid production.

[0059] FIG. 9 depicts production of olivetolic acid using expression constructs 3 + 4 or expression constructs 3 + 5.

[0060] FIG. 10 depicts production of olivetolic acid using Construct 3 and culturing the cells in medium comprising hexanoate; or using Construct 1.

[0061] FIG. 11 is a schematic depiction of pathways for production of olivetolic acid derivatives by feeding various representative carboxylic acids, where the carboxylic acids are converted to their CoA forms by a promiscuous acyl-activating enzyme polypeptide (e.g., CsAAE1; CsAAE3), generating olivetolic acid derivatives.

[0062] FIG. 12 depicts various representative carboxylic acids with various functional groups that can be used as substrate for the biosynthesis of olivetolic acid or cannabinoid derivatives. FIG. 12 also depicts production of olivetolic acid or cannabinoid derivatives from these carboxylic acids.

[0063] FIG. 13 depicts various representative cannabinoid derivatives that can be generated by feeding different acids and the further derivatization of those derivatives with chemical reactions.

[0064] FIG. 14 depicts cannabinoid biosynthetic pathways utilizing neryl pyrophosphate (NPP) or GPP.

[0065] FIG. 15 depicts generation of cannabigerolic acid (CBGA) using a NphB polypeptide and the substrates olivetolic acid and GPP.

[0066] FIG. 16 depicts an expression construct to produce GPP.

[0067] FIG. 17 depicts an expression construct to produce hexanoyl-CoA and/or hexanoate.

[0068] FIG. 18 depicts an expression construct to produce hexanoyl-CoA and/or hexanoate.

2018256863 11 Apr 2024

[0069] FIG. 19 depicts an expression construct to produce olivetolic acid.

[0070] FIG. 20 depicts an expression construct to produce CBGA.

[0071] FIG. 21 depicts an expression construct to produce cannabidiolic acid (CBDA).

[0072] FIG. 22 depicts an expression construct to produce CBDA.

[0073] FIG. 23 depicts an expression construct to produce CBDA.

[0074] FIG. 24 depicts an expression construct to produce tetrahydrocannabinolic acid (THCA).

[0075] FIG. 25 depicts an expression construct to produce THCA.

[0076] FIGS. 26A, FIG. 26B, and FIG. 26C depict LC-MS traces illustrating the production of CBGA. These figures illustrate an LC-MS trace (m/z = 359.2) for ethyl acetate extraction of the yL444 strain (FIG. 26A), a 10 μM CBGA standard (FIG. 26B), and a mixture of ethyl acetate extraction of yL444 and 10 μM CBGA standard (FIG. 26C). Peaks observed at 9.2 minutes indicated the presence of CBGA in the ethyl acetate extraction of yL444.

[0077] FIG. 27 depicts the production of THCA with a THCA synthase polypeptide with an N-terminal truncation and a ProA signal sequence. The figure illustrates an LC-MS trace (m/z = 357.2144) for ethyl acetate extraction of yXL046 colony 1 (Duplicate 1, Top), yXL046 colony 2 (Duplicate 2, Middle), and a standard containing CBDA and THCA (Standard, Bottom). The peak at 7.9 mins indicated the presence of CBDA and the peak at 9.6 mins indicated the presence of THCA.

[0078] FIG. 28 depicts the production of CBDA with a CBDA synthase polypeptide with an N-terminal truncation and a ProA signal sequence. The figure illustrates an LC-MS trace (m/z = 357.2144) for ethyl acetate extraction of yXL047 colony 1 (Duplicate 1), a yXL047 colony 2 (Duplicate 2), a negative control (Negative) and a standard containing CBDA and THCA (Standard). The peak at 7.9 mins indicated the presence of CBDA and the peak at 9.6 mins indicated the presence of THCA.

[0079] FIGS. 29A and 29B depict expression constructs used in the production of the S21 strain. The expression constructs depicted in FIGS. 29A and 29B are also used in the production of following strains: S29, S31, S34, S35, S37, S38, S39, S41, S42, S43, S44, S45, S46, S47, S49, S50, S51, S78, S80, S81, S82, S83, S84, S85, S86, S87, S88, S89, S90, S91, S94, S95, S97, S104, S108, S112, S114, S115, S116, S118, S123, S147, S164, S165, S166, S167, S168, S169, and S170.

[0080] FIGS. 30A, 30B, and 30C depict expression constructs used in the production of the S31 strain. The expression constructs depicted in FIGS. 30A, 30B, and 30C are also used in the production of following strains: S94, S95, and S97.

2018256863 11 Apr 2024

[0081] FIG. 31 depicts expression constructs used in the production of the S35 strain.

[0082] FIG. 32 depicts expression constructs used in the production of the S37 strain.

[0083] FIG. 33 depicts expression constructs used in the production of the S38 strain.

[0084] FIG. 34 depicts expression constructs used in the production of the S39 strain.

[0085] FIG. 35 depicts expression constructs used in the production of the S41 strain.

[0086] FIG. 36 depicts expression constructs used in the production of the S42 strain.

[0087] FIG. 37 depicts expression constructs used in the production of the S43 strain.

[0088] FIG. 38 depicts expression constructs used in the production of the S44 strain.

[0089] FIG. 39 depicts expression constructs used in the production of the S45 strain.

[0090] FIG. 40 depicts expression constructs used in the production of the S46 strain.

[0091] FIG. 41 depicts expression constructs used in the production of the S47 strain.

[0092] FIGS. 42A, 42B, and 42C depict expression constructs used in the production of the S49 strain.

[0093] FIGS. 43A, 43B, and 43C depict expression constructs used in the production of the S50 strain.

[0094] FIGS. 44A, 44B, and 44C depict expression constructs used in the production of the S51 strain. The expression constructs depicted in FIGS. 44A, 44B, and 44C are also used in the production of following strains: S78, S80, S81, S82, S83, S84, S85, S86, S87, S88, and S89.

[0095] FIG. 45 depicts expression constructs used in the production of the S78 strain.

[0096] FIG. 46 depicts expression constructs used in the production of the S80 strain.

[0097] FIG. 47 depicts expression constructs used in the production of the S81 strain.

[0098] FIG. 48 depicts expression constructs used in the production of the S82 strain.

[0099] FIG. 49 depicts expression constructs used in the production of the S83 strain.

[00100] FIG. 50 depicts expression constructs used in the production of the S84 strain.

[00101] FIG. 51 depicts expression constructs used in the production of the S85 strain.

[00102] FIG. 52 depicts expression constructs used in the production of the S86 strain.

[00103] FIG. 53 depicts expression constructs used in the production of the S87 strain.

[00104] FIG. 54 depicts expression constructs used in the production of the S88 strain.

[00105] FIG. 55 depicts expression constructs used in the production of the S89 strain.

[00106] FIGS. 56A, 56B, and 56C depict expression constructs used in the production of the S90 strain.

[00107] FIGS. 57A, 57B, and 57C depict expression constructs used in the production of the S91 strain.

2018256863 11 Apr 2024

[00108] FIG. 58 depicts expression constructs used in the production of the S94 strain.

[00109] FIG. 59 depicts expression constructs used in the production of the S95 strain.

[00110] FIG. 60 depicts expression constructs used in the production of the S97 strain.

[00111] FIG. 61 depicts expression constructs used in the production of the S104 strain.

[00112] FIG. 62 depicts expression constructs used in the production of the S108 strain.

[00113] FIG. 63 depicts expression constructs used in the production of the S112 strain.

[00114] FIG. 64 depicts expression constructs used in the production of the S114 strain.

[00115] FIG. 65 depicts expression constructs used in the production of the S115 strain.

[00116] FIG. 66 depicts expression constructs used in the production of the S116 strain.

[00117] FIG. 67 depicts expression constructs used in the production of the S118 strain.

[00118] FIG. 68 depicts expression constructs used in the production of the S123 strain.

[00119] FIG. 69 depicts expression constructs used in the production of the S147 strain.

[00120] FIG. 70 depicts expression constructs used in the production of the S164 strain.

[00121] FIG. 71 depicts expression constructs used in the production of the S165 strain.

[00122] FIG. 72 depicts expression constructs used in the production of the S166 strain.

[00123] FIG. 73 depicts expression constructs used in the production of the S167 strain.

[00124] FIG. 74 depicts expression constructs used in the production of the S168 strain.

[00125] FIG. 75 depicts expression constructs used in the production of the S169 strain.

2018256863 11 Apr 2024

[00126] FIG. 76 depicts expression constructs used in the production of the S170 strain.

[00127] FIG. 77 depicts the MS/MS spectrum of the CBGA peak produced from a CsPT4 polypeptide expressing strain (S29).

[00128] FIG. 78 depicts the MS/MS spectrum of an authentic CBGA standard.

[00129] FIG. 79 depicts CBGA produced by a CsGOT polypeptide at 1.06 min (top), CBGA produced by a CsPT4 polypeptide at 1.06 min (middle), and authentic CBGA standard at 1.06 min (bottom).

[00130] FIG. 80 depicts CBGA produced by a CsGOT polypeptide at 1.06 min (scale x 10^2 units).

[00131] FIG. 81 depicts CBGA produced by a CsPT4 polypeptide at 1.06 min (scale x 10^4 units)

[00132] FIG. 82 depicts an authentic CBGA standard at 1.06 min (scale x 10^4 units).

[00133] FIG. 83 depicts CBDA produced by S34 at 1.02 min (top) and an authentic CBDA standard at 1.02 min (bottom).

[00134] FIG. 84 depicts THCA produced from strain D123 at 1.29 min (top) and an authentic THCA standard at 1.29 min (bottom).

[00135] FIG. 85 depicts expression constructs used in the production of the S34 strain.

[00136] FIG. 86 depicts expression constructs used in the production of the S29 strain. The expression constructs depicted in FIG. 86 are also used in the production of following strains: S31, S34, S35, S37, S38, S39, S41, S42, S43, S44, S45, S46, S47, S49, S50, S51, S78, S80, S81, S82, S83, S84, S85, S86, S87, S88, S89, S90, S91, S94, S95, S97, and S123.

DETAILED DESCRIPTION

[00137] The present disclosure provides methods, polypeptides, nucleic acids encoding said polypeptides, and genetically modified host cells for producing cannabinoids, cannabinoid precursors, cannabinoid derivatives (e.g., non-naturally occurring cannabinoids), or cannabinoid precursor derivatives (e.g., non-naturally occurring cannabinoid precursors).

[00138] Geranyl pyrophosphate:olivetolic acid geranyltransferase (GOT, Enzyme Commission Number 2.5.1.102) polypeptides play an important role in the biosynthesis of cannabinoids, but reconstituting their activity in a genetically modified host cell has proven challenging, hampering progress in the production of cannabinoids or cannabinoid derivatives. Herein, novel genes encoding polypeptides of the disclosure that catalyze

production of cannabigerolic acid (CBGA) from GPP and olivetolic acid have been identified, isolated, and characterized. Surprisingly, these polypeptides of the present disclosure can catalyze production of CBGA from GPP and olivetolic acid in an amount at least ten times higher than previously discovered *Cannabis* polypeptides that catalyze production of CBGA from GPP and olivetolic acid (see, for example, U.S. Patent Application Pub. No. US20120144523 and the GOT polypeptide, CsPT1, disclosed therein; SEQ ID NO:82 herein). The new polypeptides of the present disclosure that catalyze production of CBGA from GPP and olivetolic acid are GOT polypeptides (e.g., the CsPT4 polypeptide) and can generate cannabinoids and cannabinoid derivatives *in vivo* (e.g., within a genetically modified host cell) and *in vitro* (e.g., cell-free). These new GOT polypeptides, as well as nucleic acids encoding said GOT polypeptides, are useful in the methods and genetically modified host cells of the disclosure for producing cannabinoids or cannabinoid derivatives.

[00139] The methods of the disclosure may include using microorganisms genetically engineered (e.g., genetically modified host cells) to produce naturally-occurring and non-naturally occurring cannabinoids or cannabinoid precursors. Naturally-occurring cannabinoids and cannabinoid precursors and non-naturally occurring cannabinoids and cannabinoid precursors (e.g., cannabinoid derivatives and cannabinoid precursor derivatives) are challenging to synthesize using chemical synthesis due to their complex structures. The methods of the disclosure enable the construction of metabolic pathways inside living cells to produce bespoke cannabinoids, cannabinoid precursors, cannabinoid derivatives, or cannabinoid precursor derivatives from simple precursors such as sugars and carboxylic acids. One or more heterologous nucleic acids disclosed herein encoding one or more polypeptides disclosed herein can be introduced into host microorganisms allowing for the stepwise conversion of inexpensive feedstocks, e.g., sugar, into final products: cannabinoids, cannabinoid precursors, cannabinoid derivatives, or cannabinoid precursor derivatives. These products can be specified by the choice and construction of expression constructs or vectors comprising one or more heterologous nucleic acids disclosed herein, allowing for the efficient bioproduction of chosen cannabinoid precursors; cannabinoids, such as THC or CBD and less common cannabinoid species found at low levels in *Cannabis*; or cannabinoid derivatives or cannabinoid precursor derivatives. Bioproduction also enables synthesis of cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives with defined stereochemistries, which is challenging to do using chemical synthesis.

[00140] The nucleic acids disclosed herein may include those encoding a polypeptide having at least one activity of a polypeptide present in the cannabinoid biosynthetic pathway, such as a GOT polypeptide (e.g., a CsPT4 polypeptide), responsible for the biosynthesis of the cannabinoid CBGA; a tetraketide synthase (TKS) polypeptide; an olivetolic acid cyclase (OAC) polypeptide; and a CBDA or THCA synthase polypeptide (see FIGS. 1 and 11). Nucleic acids disclosed herein may also include those encoding a polypeptide having at least one activity of a polypeptide involved in the synthesis of cannabinoid precursors. These polypeptides include, but are not limited to, polypeptides having at least one activity of a polypeptide present in the mevalonate pathway; polypeptides that generate acyl-CoA compounds or acyl-CoA compound derivatives (e.g., an acyl-activating enzyme polypeptide, a fatty acyl-CoA synthetase polypeptide, or a fatty acyl-CoA ligase polypeptide); polypeptides that generate GPP; polypeptides that generate malonyl-CoA; polypeptides that condense two molecules of acetyl-CoA to generate acetoacetyl-CoA, or pyruvate decarboxylase polypeptides (see FIGS. 1 and 11).

[00141] The disclosure also provides for generation of cannabinoid precursor derivatives or cannabinoid derivatives, as well as cannabinoids or precursors thereof, with polypeptides that generate acyl-CoA compounds or acyl-CoA compound derivatives. In certain such embodiments, genetically modified host cells disclosed herein are modified with one or more heterologous nucleic acids encoding a polypeptide that generates acyl-CoA compounds or acyl-CoA compound derivatives. These polypeptides may permit production of hexanoyl-CoA, acyl-CoA compounds, derivatives of hexanoyl-CoA, or derivatives of acyl-CoA compounds. In some embodiments, hexanoic acid or carboxylic acids other than hexanoic acid are fed to genetically modified host cells expressing a polypeptide that generates acyl-CoA compounds or acyl-CoA compound derivatives (e.g., are present in the culture medium in which the cells are grown) to generate hexanoyl-CoA, acyl-CoA compounds, derivatives of hexanoyl-CoA, or derivatives of acyl-CoA compounds. These compounds are then converted to cannabinoid derivatives or cannabinoid precursor derivatives, as well as cannabinoids or precursors thereof, via one or more polypeptides having at least one activity of a polypeptide present in the cannabinoid biosynthetic pathway or involved in the synthesis of cannabinoid precursors (see FIGS. 1 and 11).

[00142] Surprisingly, it was found that polypeptides that generate acyl-CoA compounds or acyl-CoA compound derivatives, as well as many polypeptides having at least one activity of a polypeptide present in the cannabinoid biosynthetic pathway, such as TKS polypeptides, OAC polypeptides, GOT polypeptides (e.g., a CsPT4 polypeptide), and CBDA

or THCA synthase polypeptides, have broad substrate specificity. This broad substrate specificity permits generation of not only cannabinoids and cannabinoid precursors, but also cannabinoid derivatives and cannabinoid precursor derivatives that are not naturally occurring, both within a genetically modified host cell or in a cell-free reaction mixture comprising one or more of the polypeptides disclosed herein. Because of this broad substrate specificity, hexanoyl-CoA, acyl-CoA compounds, derivatives of hexanoyl-CoA, or derivatives of acyl-CoA compounds produced in genetically modified host cells by polypeptides that generate acyl-CoA compounds or acyl-CoA compound derivatives can be utilized by TKS and OAC polypeptides to make olivetolic acid or derivatives thereof. The olivetolic acid or derivatives thereof can then be utilized by a GOT polypeptide to afford cannabinoids or cannabinoid derivatives. Alternatively, olivetolic acid or derivatives thereof can be fed to genetically modified host cells comprising a GOT polypeptide to afford cannabinoids or cannabinoid derivatives. These cannabinoids or cannabinoid derivatives can then be converted to THCA or CDBA, or derivatives thereof, via a CBDA or THCA synthase polypeptide.

[00143] Besides allowing for the production of desired cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives, the present disclosure provides a more reliable and economical process than agriculture-based production. Microbial fermentations can be completed in days versus the months necessary for an agricultural crop, are not affected by climate variation or soil contamination (e.g., by heavy metals), and can produce pure products at high titer.

[00144] The present disclosure also provides a platform for the economical production of cannabinoid precursors, or derivatives thereof, and high-value cannabinoids including THC and CBD, as well as derivatives thereof. It also provides for the production of different cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives for which no viable method of production exists.

[00145] Additionally, the disclosure provides methods, genetically modified host cells, polypeptides, and nucleic acids encoding said polypeptides to produce cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives *in vivo* or *in vitro* from simple precursors. Nucleic acids disclosed herein encoding one or more polypeptides disclosed herein can be introduced into microorganisms (e.g., genetically modified host cells), resulting in expression or overexpression of the one or more polypeptides, which can then be utilized *in vitro* or *in vivo* for the production of

2018256863 11 Apr 2024

cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives. In some embodiments, the *in vitro* methods are cell-free.

[00146] To produce cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives, and create biosynthetic pathways within genetically modified host cells, the genetically modified host cells may express or overexpress combinations of the heterologous nucleic acids disclosed herein encoding polypeptides disclosed herein.

Cannabinoid Biosynthesis

[00147] Nucleic acids encoding polypeptides having at least one activity of a polypeptide present in the cannabinoid biosynthesis pathway can be useful in the methods and genetically modified host cells disclosed herein for the synthesis of cannabinoids, cannabinoid precursors, cannabinoid derivatives, or cannabinoid precursor derivatives.

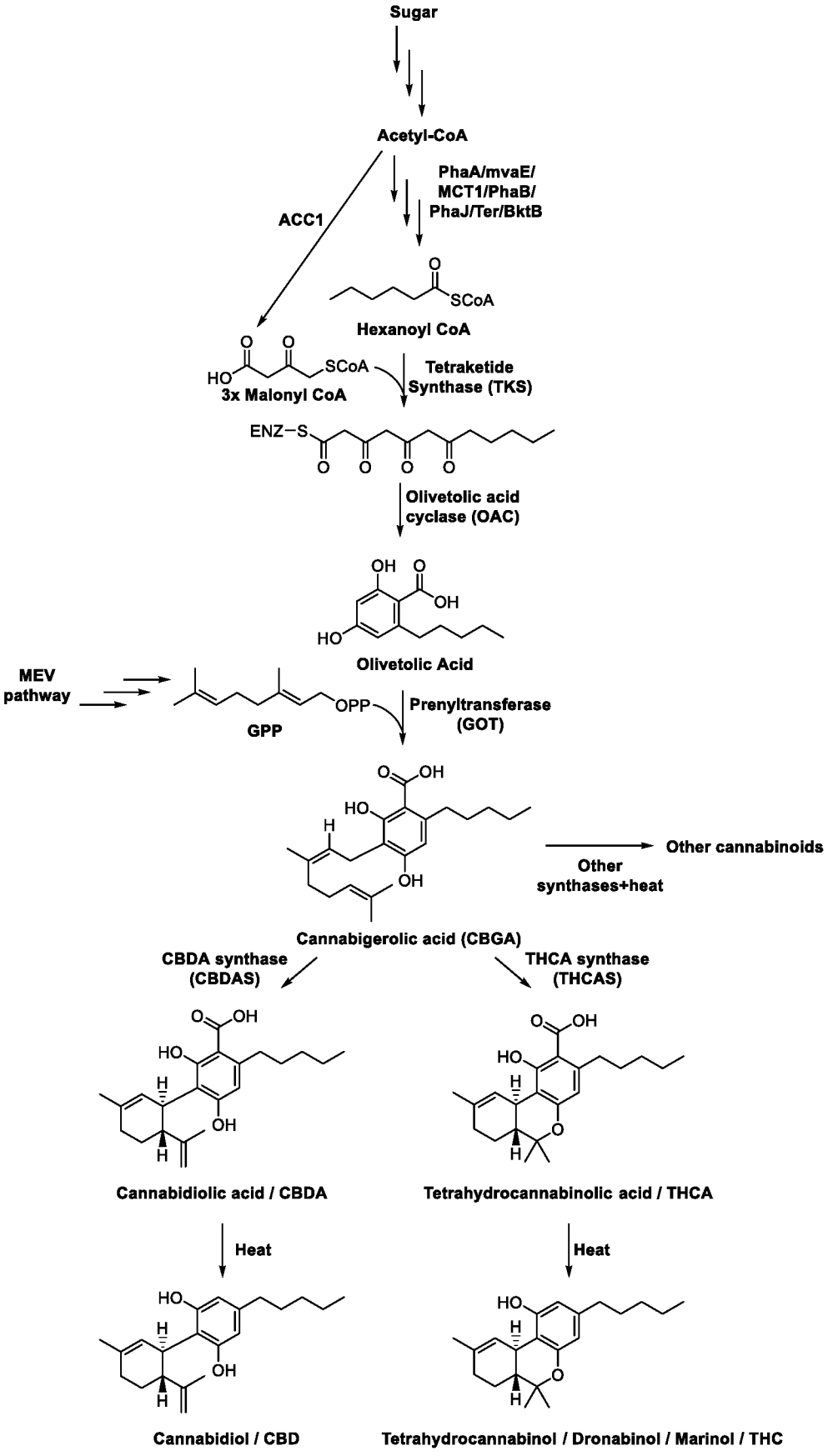
[00148] In *Cannabis*, cannabinoids are produced from the common metabolite precursors geranylpyrophosphate (GPP) and hexanoyl-CoA by the action of three polypeptides so far only identified in *Cannabis*. Hexanoyl-CoA and malonyl-CoA are combined to afford a 12-carbon tetraketide intermediate by a TKS polypeptide. This tetraketide intermediate is then cyclized by an OAC polypeptide to produce olivetolic acid. Olivetolic acid is then prenylated with the common isoprenoid precursor GPP by a GOT polypeptide (e.g., a CsPT4 polypeptide) to produce CBGA, the cannabinoid also known as the "mother cannabinoid." Different synthase polypeptides then convert CBGA into other cannabinoids, e.g., a THCA synthase polypeptide produces THCA, a CBDA synthase polypeptide produces CBDA, etc. In the presence of heat or light, the acidic cannabinoids can undergo decarboxylation, e.g., THCA producing THC or CBDA producing CBD.

[00149] GPP and hexanoyl-CoA can be generated through several pathways (see FIGS. 1 and 11). One or more nucleic acids encoding one or more polypeptides having at least one activity of a polypeptide present in these pathways can be useful in the methods and genetically modified host cells for the synthesis of cannabinoids, cannabinoid precursors, cannabinoid derivatives, or cannabinoid precursor derivatives.

[00150] Polypeptides that generate GPP or are part of a biosynthetic pathway that generates GPP may be one or more polypeptides having at least one activity of a polypeptide present in the mevalonate (MEV) pathway. The term "mevalonate pathway" or "MEV pathway," as used herein, may refer to the biosynthetic pathway that converts acetyl-CoA to isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). The

mevalonate pathway comprises polypeptides that catalyze the following steps: (a) condensing two molecules of acetyl-CoA to generate acetoacetyl-CoA (e.g., by action of an acetoacetyl-CoA thiolase polypeptide); (b) condensing acetoacetyl-CoA with acetyl-CoA to form hydroxymethylglutaryl-CoA (HMG-CoA) (e.g., by action of a HMG-CoA synthase (HMGS) polypeptide); (c) converting HMG-CoA to mevalonate (e.g., by action of a HMG-CoA reductase (HMGR) polypeptide); (d) phosphorylating mevalonate to mevalonate 5-phosphate (e.g., by action of a mevalonate kinase (MK) polypeptide); (e) converting mevalonate 5-phosphate to mevalonate 5-pyrophosphate (e.g., by action of a phosphomevalonate kinase (PMK) polypeptide); (f) converting mevalonate 5-pyrophosphate to isopentenyl pyrophosphate (e.g., by action of a mevalonate pyrophosphate decarboxylase (MVD) polypeptide); and (g) converting isopentenyl pyrophosphate (IPP) to dimethylallyl pyrophosphate (DMAPP) (e.g., by action of an isopentenyl pyrophosphate isomerase (IDI) polypeptide) (FIGS. 1 and 11). A geranyl diphosphate synthase (GPPS) polypeptide then acts on IPP and/or DMAPP to generate GPP. Additionally, polypeptides that generate GPP or are part of a biosynthetic pathway that generates GPP may be one or more polypeptides having at least one activity of a polypeptide present in the deoxyxylulose-5-phosphate (DXP) pathway, instead of those of the MEV pathway (FIG. 1).

[00151] Polypeptides that generate hexanoyl-CoA may include polypeptides that generate acyl-CoA compounds or acyl-CoA compound derivatives (e.g., a hexanoyl-CoA synthase (HCS) polypeptide, an acyl-activating enzyme polypeptide, a fatty acyl-CoA synthetase polypeptide, or a fatty acyl-CoA ligase polypeptide). Hexanoyl-CoA may also be generated through pathways comprising one or more polypeptides that generate malonyl-CoA, such as an acetyl-CoA carboxylase (ACC) polypeptide. Additionally, hexanoyl-CoA may be generated with one or more polypeptides that are part of a biosynthetic pathway that produces hexanoyl-CoA, including, but not limited to: a malonyl CoA-acyl carrier protein transacylase (MCT1) polypeptide, a PaaH1 polypeptide, a Crt polypeptide, a Ter polypeptide, and a BktB polypeptide; a MCT1 polypeptide, a PhaB polypeptide, a PhaJ polypeptide, a Ter polypeptide, and a BktB polypeptide; a short chain fatty acyl-CoA thioesterase (SCFA-TE) polypeptide; or a fatty acid synthase (FAS) polypeptide (see FIGS. 1 and 11). Hexanoyl CoA derivatives, acyl-CoA compounds, or acyl-CoA compound derivatives may also be formed via such pathways and polypeptides.



Biosynthetic pathways to cannabinoids

[00152] GPP and hexanoyl-CoA may also be generated through pathways comprising polypeptides that condense two molecules of acetyl-CoA to generate acetoacetyl-CoA and

2018256863 11 Apr 2024

pyruvate decarboxylase polypeptides that generate acetyl-CoA from pyruvate (see FIGS. 1 and 11). Hexanoyl CoA derivatives, acyl-CoA compounds, or acyl-CoA compound derivatives may also be formed via such pathways.

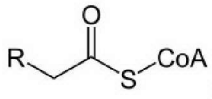
General Information

[00153] The practice of the present disclosure will employ, unless otherwise indicated, conventional techniques of molecular biology (including recombinant techniques), microbiology, cell biology, biochemistry, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature: "*Molecular Cloning: A Laboratory Manual*," second edition (Sambrook et al., 1989); "*Oligonucleotide Synthesis*" (M. J. Gait, ed., 1984); "*Animal Cell Culture*" (R. I. Freshney, ed., 1987); "*Methods in Enzymology*" (Academic Press, Inc.); "*Current Protocols in Molecular Biology*" (F. M. Ausubel et al., eds., 1987, and periodic updates); "*PCR: The Polymerase Chain Reaction*," (Mullis et al., eds., 1994). Singleton et al., *Dictionary of Microbiology and Molecular Biology* 2nd ed., J. Wiley & Sons (New York, N.Y. 1994), and March, *Advanced Organic Chemistry Reactions, Mechanisms and Structure* 4th ed., John Wiley & Sons (New York, N.Y. 1992), provide one skilled in the art with a general guide to many of the terms used in the present application.

[00154] "Cannabinoid" or "cannabinoid compound" as used herein may refer to a member of a class of unique meroterpenoids found until now only in *Cannabis sativa*. Cannabinoids may include, but are not limited to, cannabichromene (CBC) type (e.g. cannabichromenic acid), cannabigerol (CBG) type (e.g. cannabigerolic acid), cannabidiol (CBD) type (e.g. cannabidiolic acid), Δ^9-trans-tetrahydrocannabinol (Δ^9 -THC) type (e.g. Δ^9-tetrahydrocannabinolic acid), Δ^8-trans-tetrahydrocannabinol (Δ^8 -THC) type, cannabicyclol (CBL) type, cannabielsoin (CBE) type, cannabinol (CBN) type, cannabinodiol (CBND) type, cannabitriol (CBT) type, cannabigerolic acid (CBGA), cannabigerolic acid monomethylether (CBGAM), cannabigerol (CBG), cannabigerol monomethylether (CBGM), cannabigerovarinic acid (CBGVA), cannabigerovarin (CBGV), cannabichromenic acid (CBCA), cannabichromene (CBC), cannabichromevarinic acid (CBCVA), cannabichromevarin (CBCV), cannabidiolic acid (CBDA), cannabidiol (CBD), cannabidiol monomethylether (CBDM), cannabidiol-C_4 (CBD-C_4), cannabidivarinic acid (CBDVA), cannabidivarin (CBDV), cannabidiorcol (CBD-C_1), Δ^9 –tetrahydrocannabinolic acid A (THCA-A), Δ^9 –tetrahydrocannabinolic acid B (THCA-B), Δ^9 –tetrahydrocannabinol (THC), Δ^9 –tetrahydrocannabinolic acid-C_4 (THCA-C_4), Δ^9 –tetrahydrocannabinol-C_4 (THC-C_4), Δ^9 –tetrahydrocannabivarinic acid (THCVA), Δ^9 –tetrahydrocannabivarin (THCV), Δ^9 –

tetrahydrocannabiorcolic acid (THCA-C_1), Δ^9 –tetrahydrocannabiorcol (THC-C_1), Δ^7 –cis-iso-tetrahydrocannabivarin, Δ^8 –tetrahydrocannabinolic acid (Δ^8 –THCA), Δ^8 –tetrahydrocannabinol (Δ^8 –THC), cannabicyclolic acid (CBLA), cannabicyclol (CBL), cannabicyclovarin (CBLV), cannabielsoic acid A (CBEA-A), cannabielsoic acid B (CBEA-B), cannabielsoin (CBE), cannabielsoinic acid, cannabicitranic acid, cannabinolic acid (CBNA), cannabinol (CBN), cannabinol methylether (CBNM), cannabinol-C_4, (CBN-C_4), cannabivarin (CBV), cannabinol-C_2 (CNB-C_2), cannabiorcol (CBN-C_1), cannabinodiol (CBND), cannabinodivarin (CBVD), cannabitriol (CBT), 10-ethyoxy-9-hydroxy-delta-6a-tetrahydrocannabinol, 8,9-dihydroxyl-delta-6a-tetrahydrocannabinol, cannabitriolvarin (CBTVE), dehydrocannabifuran (DCBF), cannabifuran (CBF), cannabichromanon (CBCN), cannabicitran (CBT), 10-oxo-delta-6a-tetrahydrocannabinol (OTHC), delta-9-cis-tetrahydrocannabinol (cis-THC), 3,4,5,6-tetrahydro-7-hydroxy-alpha-alpha-2-trimethyl-9-n-propyl-2,6-methano-2H-1-benzoxocin-5-methanol (OH-iso-HHCV), cannabiripsol (CBR), and trihydroxy-delta-9-tetrahydrocannabinol (triOH-THC).

[00155] "Cannabinoid precursor" as used herein may refer to any intermediate present in the cannabinoid biosynthetic pathway before the production of the "mother cannabinoid," cannabigerolic acid (CBGA). Cannabinoid precursors may include, but are not limited to, GPP, olivetolic acid, hexanoyl-CoA, pyruvate, acetoacetyl-CoA, butyryl-CoA, acetyl-CoA, HMG-CoA, mevalonate, mevalonate-5-phosphate, mevalonate diphosphate, and malonyl-CoA.

[00156] An acyl-CoA compound as detailed herein may include compounds with the following structure: 

, wherein R is a fatty acid side chain optionally comprising one or more functional and/or reactive groups as disclosed herein (i.e., an acyl-CoA compound derivative).

[00157] As used herein, a hexanoyl CoA derivative, an acyl-CoA compound derivative, a cannabinoid derivative, or a cannabinoid precursor derivative (e.g., an olivetolic acid derivative) is produced by a genetically modified host cell disclosed herein or in a cell-free reaction mixture comprising one or more of the polypeptides disclosed herein and may refer to hexanoyl CoA, an acyl-CoA compound, a cannabinoid, or a cannabinoid precursor (e.g., olivetolic acid) comprising one or more functional and/or reactive groups. Functional groups may include, but are not limited to, azido, halo (e.g., chloride, bromide, iodide, fluorine), methyl, alkyl (including branched and linear alkyl groups), alkynyl, alkenyl, methoxy, alkoxy, acetyl, amino, carboxyl, carbonyl, oxo, ester, hydroxyl, thio, cyano, aryl,

heteroaryl, cycloalkyl, cycloalkenyl, cycloalkylalkenyl, cycloalkylalkynyl, cycloalkenylalkyl, cycloalkenylalkenyl, cycloalkenylalkynyl, heterocyclylalkenyl, heterocyclylalkynyl, heteroarylalkenyl, heteroarylalkynyl, arylalkenyl, arylalkynyl, heterocyclyl, spirocyclyl, heterospirocyclyl, thioalkyl, sulfone, sulfonyl, sulfoxide, amido, alkylamino, dialkylamino, arylamino, alkylarylamino, diarylamino, N-oxide, imide, enamine, imine, oxime, hydrazone, nitrile, aralkyl, cycloalkylalkyl, haloalkyl, heterocyclylalkyl, heteroarylalkyl, nitro, thioxo, and the like. See, e.g., FIGS. 12 and 13. Suitable reactive groups may include, but are not necessarily limited to, azide, carboxyl, carbonyl, amine, (e.g., alkyl amine (e.g., lower alkyl amine), aryl amine), halide, ester (e.g., alkyl ester (e.g., lower alkyl ester, benzyl ester), aryl ester, substituted aryl ester), cyano, thioester, thioether, sulfonyl halide, alcohol, thiol, succinimidyl ester, isothiocyanate, iodoacetamide, maleimide, hydrazine, alkynyl, alkenyl, and the like. A reactive group may facilitate covalent attachment of a molecule of interest. Suitable molecules of interest may include, but are not limited to, a detectable label; imaging agents; a toxin (including cytotoxins); a linker; a peptide; a drug (e.g., small molecule drugs); a member of a specific binding pair; an epitope tag; ligands for binding by a target receptor; tags to aid in purification; molecules that increase solubility; molecules that enhance bioavailability; molecules that increase *in vivo* half-life; molecules that target to a particular cell type; molecules that target to a particular tissue; molecules that provide for crossing the blood-brain barrier; molecules to facilitate selective attachment to a surface; and the like. Functional and reactive groups may be optionally substituted with one or more additional functional or reactive groups.

[00158] A cannabinoid derivative or cannabinoid precursor derivative produced by a genetically modified host cell disclosed herein or in a cell-free reaction mixture comprising one or more of the polypeptides disclosed herein may also refer a naturally-occurring cannabinoid or naturally-occurring cannabinoid precursor lacking one or more chemical moieties. Such chemical moieties may include, but are not limited to, methyl, alkyl, alkenyl, methoxy, alkoxy, acetyl, carboxyl, carbonyl, oxo, ester, hydroxyl, aryl, heteroaryl, cycloalkyl, cycloalkenyl, cycloalkylalkenyl, cycloalkenylalkyl, cycloalkenylalkenyl, heterocyclylalkenyl, heteroarylalkenyl, arylalkenyl, heterocyclyl, aralkyl, cycloalkylalkyl, heterocyclylalkyl, heteroarylalkyl, and the like. In some embodiments, a cannabinoid derivative or cannabinoid precursor derivative lacking one or more chemical moieties found in a naturally-occurring cannabinoid or naturally-occurring cannabinoid precursor, and produced by a genetically modified host cell disclosed herein or in a cell-free reaction

mixture comprising one or more of the polypeptides disclosed herein, may also comprise one or more of any of the functional and/or reactive groups described herein. Functional and reactive groups may be optionally substituted with one or more additional functional or reactive groups.

[00159] The term "nucleic acid" used herein, may refer to a polymeric form of nucleotides of any length, either ribonucleotides or deoxynucleotides. Thus, this term may include, but is not limited to, single-, double-, or multi-stranded DNA or RNA, genomic DNA, cDNA, genes, synthetic DNA or RNA, DNA-RNA hybrids, or a polymer comprising purine and pyrimidine bases or other naturally-occurring, chemically or biochemically modified, non- naturally-occurring, or derivatized nucleotide bases.

[00160] The terms "peptide," "polypeptide," and "protein" may be used interchangeably herein, and may refer to a polymeric form of amino acids of any length, which can include coded and non-coded amino acids, chemically or biochemically modified or derivatized amino acids, full-length polypeptides, fragments of polypeptides, or polypeptides having modified peptide backbones. The polypeptides disclosed herein may be presented as modified or engineered forms, including truncated or fusion forms, retaining the recited activities. The polypeptides disclosed herein may also be variants differing from a specifically recited "reference" polypeptide (e.g., a wild-type polypeptide) by amino acid insertions, deletions, mutations, and/or substitutions, but retains an activity that is substantially similar to the reference polypeptide.

[00161] As used herein, the term "heterologous" may refer to what is not normally found in nature. The term "heterologous nucleotide sequence" may refer to a nucleotide sequence not normally found in a given cell in nature. As such, a heterologous nucleotide sequence may be: (a) foreign to its host cell (i.e., is "exogenous" to the cell); (b) naturally found in the host cell (i.e., "endogenous") but present at an unnatural quantity in the cell (i.e., greater or lesser quantity than naturally found in the host cell); or (c) be naturally found in the host cell but positioned outside of its natural locus. The term "heterologous enzyme" or "heterologous polypeptide" may refer to an enzyme or polypeptide that is not normally found in a given cell in nature. The term encompasses an enzyme or polypeptide that is: (a) exogenous to a given cell (i.e., encoded by a nucleic acid that is not naturally present in the host cell or not naturally present in a given context in the host cell); and (b) naturally found in the host cell (e.g., the enzyme or polypeptide is encoded by a nucleic acid that is endogenous to the cell) but that is produced in an unnatural amount (e.g., greater or lesser than that naturally found) in the host cell. As such, a heterologous nucleic acid may be: (a)

foreign to its host cell (i.e., is "exogenous" to the cell); (b) naturally found in the host cell (i.e., "endogenous") but present at an unnatural quantity in the cell (i.e., greater or lesser quantity than naturally found in the host cell); or (c) be naturally found in the host cell but positioned outside of its natural locus.

[00162] "Operably linked" may refer to an arrangement of elements wherein the components so described are configured so as to perform their usual function. Thus, control sequences operably linked to a coding sequence are capable of effecting the expression of the coding sequence. The control sequences need not be contiguous with the coding sequence, so long as they function to direct the expression thereof. Thus, for example, intervening untranslated yet transcribed sequences can be present between a promoter sequence and the coding sequence and the promoter sequence can still be considered "operably linked" to the coding sequence.

[00163] "Isolated" may refer to polypeptides or nucleic acids that are substantially or essentially free from components that normally accompany them in their natural state. An isolated polypeptide or nucleic acid may be other than in the form or setting in which it is found in nature. Isolated polypeptides and nucleic acids therefore may be distinguished from the polypeptides and nucleic acids as they exist in natural cells. An isolated nucleic acid or polypeptide may further be purified from one or more other components in a mixture with the isolated nucleic acid or polypeptide, if such components are present.

[00164] A "genetically modified host cell" (also referred to as a "recombinant host cell") is a host cell into which has been introduced a heterologous nucleic acid, e.g., an expression vector or construct. For example, a prokaryotic host cell is a genetically modified prokaryotic host cell (e.g., a bacterium), by virtue of introduction into a suitable prokaryotic host cell of a heterologous nucleic acid, e.g., an exogenous nucleic acid that is foreign to (not normally found in nature in) the prokaryotic host cell, or a recombinant nucleic acid that is not normally found in the prokaryotic host cell; and a eukaryotic host cell is a genetically modified eukaryotic host cell, by virtue of introduction into a suitable eukaryotic host cell of a heterologous nucleic acid, e.g., an exogenous nucleic acid that is foreign to the eukaryotic host cell, or a recombinant nucleic acid that is not normally found in the eukaryotic host cell.

[00165] As used herein, a "cell-free system" may refer to a cell lysate, cell extract or other preparation in which substantially all of the cells in the preparation have been disrupted or otherwise processed so that all or selected cellular components, e.g., organelles, proteins, nucleic acids, the cell membrane itself (or fragments or components thereof), or the

like, are released from the cell or resuspended into an appropriate medium and/or purified from the cellular milieu. Cell-free systems can include reaction mixtures prepared from purified or isolated polypeptides and suitable reagents and buffers.

[00166] In some embodiments, conservative substitutions may be made in the amino acid sequence of a polypeptide without disrupting the three-dimensional structure or function of the polypeptide. Conservative substitutions may be accomplished by the skilled artisan by substituting amino acids with similar hydrophobicity, polarity, and R-chain length for one another. Additionally, by comparing aligned sequences of homologous proteins from different species, conservative substitutions may be identified by locating amino acid residues that have been mutated between species without altering the basic functions of the encoded proteins. The term "conservative amino acid substitution" may refer to the interchangeability in proteins of amino acid residues having similar side chains. For example, a group of amino acids having aliphatic side chains consists of glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains consists of serine and threonine; a group of amino acids having amide containing side chains consisting of asparagine and glutamine; a group of amino acids having aromatic side chains consists of phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains consists of lysine, arginine, and histidine; a group of amino acids having acidic side chains consists of glutamate and aspartate; and a group of amino acids having sulfur containing side chains consists of cysteine and methionine. Exemplary conservative amino acid substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, and asparagine-glutamine.

[00167] A polynucleotide or polypeptide has a certain percent "sequence identity" to another polynucleotide or polypeptide, meaning that, when aligned, that percentage of bases or amino acids are the same, and in the same relative position, when comparing the two sequences. Sequence identity can be determined in a number of different manners. To determine sequence identity, sequences can be aligned using various methods and computer programs (e.g., BLAST, T-COFFEE, MUSCLE, MAFFT, etc.), available over the world wide web at sites including ncbi.nlm.nili.gov/BLAST,ebi.ac.uk/Tools/msa/tcoffee/ebi.ac.uk/Tools/msa/muscle/mafft.cbrc.jp/alignment/software/. See, e.g., Altschul et al. (1990), J. Mol. Biol. 215:403-10.

[00168] Before the present disclosure is further described, it is to be understood that this disclosure is not limited to particular embodiments described, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of

describing particular embodiments only, and is not intended to be limiting, since the scope of the present disclosure will be limited only by the appended claims.

[00169] Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit 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 disclosure. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges, and are also encompassed within the disclosure, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the disclosure.

[00170] 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 this disclosure belongs. Although any methods and materials similar or equivalent to those described herein can also be used in the practice or testing of the present disclosure, the preferred methods and materials are now described. All publications mentioned herein are incorporated herein by reference to disclose and describe the methods and/or materials in connection with which the publications are cited.

[00171] It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" may include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a cannabinoid compound" or "cannabinoid" may include a plurality of such compounds and reference to "the genetically modified host cell" may include reference to one or more genetically modified host cells and equivalents thereof known to those skilled in the art, and so forth. 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.

[00172] It is appreciated that certain features of the disclosure, which are, for clarity, described in the context of separate embodiments, may also be provided in combination in a single embodiment. Conversely, various features of the disclosure, which are, for brevity, described in the context of a single embodiment, may also be provided separately or in any suitable sub-combination. All combinations of the embodiments pertaining to the disclosure are specifically embraced by the present disclosure and are disclosed herein just as if each and every combination was individually and explicitly disclosed. In addition, all sub-combinations of the various embodiments and elements thereof are also specifically

2018256863 11 Apr 2024

embraced by the present disclosure and are disclosed herein just as if each and every such sub-combination was individually and explicitly disclosed herein.

Geranyl Pyrophosphate:Olivetolic Acid Geranyltransferase Polypeptides and Nucleic Acids Encoding Said Polypeptides

[00173] As described herein, novel polypeptides for catalyzing production of cannabigerolic acid from GPP and olivetolic acid have been identified and characterized. Surprisingly, these new polypeptides of the present disclosure can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than previously discovered *Cannabis* polypeptides that catalyze production of cannabigerolic acid from GPP and olivetolic acid (see, for example, U.S. Patent Application Pub. No. US20120144523 and the GOT polypeptide, CsPT1, disclosed therein; SEQ ID NO:82 herein). The new polypeptides of the present disclosure that catalyze production of cannabigerolic acid from GPP and olivetolic acid are new geranyl pyrophosphate:olivetolic acid geranyltransferase (GOT) polypeptides, the CsPT4 polypeptide and truncated versions thereof. These new polypeptides of the present disclosure can generate cannabinoids and cannabinoid derivatives *in vivo* (e.g., within a genetically modified host cell) and *in vitro* (e.g., cell-free).

[00174] These new GOT polypeptides, as well as nucleic acids encoding said GOT polypeptides, are useful in the methods and genetically modified host cells of the disclosure for producing cannabinoids or cannabinoid derivatives. In some embodiments, the GOT polypeptide of the disclosure cannot catalyze production of 5-geranyl olivetolic acid.

[00175] The CsPT4 polypeptide is remarkably different in sequence and activity than the previously identified CsPT1 polypeptide, also a GOT polypeptide. The CsPT1 polypeptide has only 57% homology to the CsPT4 polypeptide. Further, unlike the CsPT1 polypeptide, the activity of the CsPT4 polypeptide, or a truncated version thereof, can be readily reconstituted in a genetically modified host cell of the disclosure, permitting the production of cannabinoids or cannabinoid derivatives by the genetically modified host cells. A truncated version of the CsPT4 polypeptide, the CsPT4t polypeptide, lacking N-terminal amino acids 1-76 of the amino acid sequence set forth in SEQ ID NO:110 (the full-length CsPT4 polypeptide amino acid sequence) was found to readily catalyze the production of cannabigerolic acid from GPP and olivetolic acid, with activity similar to that of the full-length CsPT4 polypeptide. However, other truncated versions of the CsPT4 polypeptide lacking N-terminal amino acids 1-112 (SEQ ID NO:211), 1-131 (SEQ ID NO:213), 1-142

(SEQ ID NO:215), 1-166 (SEQ ID NO:217), or 1-186 (SEQ ID NO:219) were unable to catalyze formation of cannabigerolic acid from GPP and olivetolic acid, suggesting that these truncation polypeptides lacked residues required for catalytic activity.

[00176] Surprisingly, it was found that the CsPT4 polypeptide, or a truncated version thereof, has broad substrate specificity, permitting generation of not only cannabinoids, but also cannabinoid derivatives. Because of this broad specificity, olivetolic acid or derivatives thereof produced in genetically modified host cells disclosed herein by TKS and OAC polypeptides can be utilized by a CsPT4 polypeptide, or a truncated version thereof, to afford cannabinoids and cannabinoid derivatives. Alternatively, olivetolic acid or derivatives thereof can be fed to genetically modified host cells disclosed herein comprising a CsPT4 polypeptide, or a truncated version thereof, to afford cannabinoids and cannabinoid derivatives. The cannabinoids and cannabinoid derivatives can then be converted to other cannabinoids or cannabinoid derivatives via a CBDA or THCA synthase polypeptide.

Isolated or Purified Nucleic Acids Encoding GOT Polypeptides

[00177] Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[00178] Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a truncated CsPT4 polypeptide (CsPT4t polypeptide, lacking N-terminal amino acids 1-76 of the amino acid sequence set forth in SEQ ID NO:110), comprising the amino acid sequence set forth in SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4t polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:100.

[00179] Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence

having at least 95% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof.

[00180] Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a full-length GOT polypeptide, a CsPT4 polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4 polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:110.

[00181] Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4 polypeptide,

comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified nucleic acid encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof.

[00182] Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:111. Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:111, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:111. Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:111. Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:111.

[00183] Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:111. Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:111. Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:111, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:111, or a codon degenerate nucleotide sequence thereof.

[00184] Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:225. Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:225, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:225. Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:225. Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:225.

[00185] Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:225. Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ

ID NO:225. Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:225, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an isolated or purified CsPT4 nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:225, or a codon degenerate nucleotide sequence thereof.

[00186] Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:224. Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:224, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:224. Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:224. Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:224.

[00187] Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:224. Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:224. Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:224, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:224, or a codon degenerate nucleotide sequence thereof.

[00188] Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:221. Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:221, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:221. Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:221. Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:221.

[00189] Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:221. Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:221. Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:221, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an isolated or purified CsPT4t nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:221, or a codon degenerate nucleotide sequence thereof.

[00190] Further included are nucleic acids that hybridize to the nucleic acids disclosed here. Hybridization conditions may be stringent in that hybridization will occur if there is at least a 90%, 95%, or 97% sequence identity with the nucleotide sequence present in the nucleic acid encoding the polypeptides disclosed herein. The stringent conditions may include those used for known Southern hybridizations such as, for example, incubation overnight at 42 °C in a solution having 50% formamide, 5×SSC (150 mM NaCl, 15 mM

2018256863 11 Apr 2024

trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5×Denhardt's solution, 10% dextran sulfate, and 20 micrograms/milliliter denatured, sheared salmon sperm DNA, following by washing the hybridization support in 0.1×SSC at about 65 °C. Other known hybridization conditions are well known and are described in Sambrook et al., Molecular Cloning: A Laboratory Manual, Third Edition, Cold Spring Harbor, N.Y. (2001).

[00191] The length of the nucleic acids disclosed herein may depend on the intended use. For example, if the intended use is as a primer or probe, for example for PCR amplification or for screening a library, the length of the nucleic acid will be less than the full length sequence, for example, 15-50 nucleotides. In certain such embodiments, the primers or probes may be substantially identical to a highly conserved region of the nucleotide sequence or may be substantially identical to either the 5' or 3' end of the nucleotide sequence. In some cases, these primers or probes may use universal bases in some positions so as to be "substantially identical" but still provide flexibility in sequence recognition. It is of note that suitable primer and probe hybridization conditions are well known in the art. Also included are cDNA molecules of the disclosed nucleic acids.

Isolated or Purified GOT Polypeptides

[00192] Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[00193] Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4t polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4t polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:100.

[00194] Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof.

[00195] Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4 polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4 polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:110.

[00196] Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4

polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an isolated or purified GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof.

Vectors Comprising Nucleic Acids Encoding GOT Polypeptides

[00197] Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[00198] Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:100. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:100.

[00199] Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids

encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof.

[00200] Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:110. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:110.

[00201] Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence

having at least 75% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to a vector comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof.

[00202] Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:111. Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:111, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:111. Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:111. Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic

acid comprising a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:111.

[00203] Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:111. Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:111. Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:111, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:111, or a codon degenerate nucleotide sequence thereof.

[00204] Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:225. Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:225, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:225. Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:225. Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:225.

[00205] Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:225. Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:225. Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:225, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to a vector comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:225, or a codon degenerate nucleotide sequence thereof.

[00206] Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:221. Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:221, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:221. Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:221. Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:221.

[00207] Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:221. Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:221. Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ

2018256863 11 Apr 2024

ID NO:221, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:221, or a codon degenerate nucleotide sequence thereof.

[00208] Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:224. Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:224, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:224. Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:224. Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:224.

[00209] Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:224. Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:224. Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:224, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to a vector comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:224, or a codon degenerate nucleotide sequence thereof.

Expression Constructs Comprising Nucleic Acids Encoding GOT Polypeptides

[00210] Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[00211] Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:100. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:100.

[00212] Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide,

comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:100. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4t polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof.

[00213] Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:110. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising the amino acid sequence set forth in SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to

SEQ ID NO:110. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:110.

[00214] Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:110. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising one or more

nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising one or more nucleic acids encoding a GOT polypeptide, a CsPT4 polypeptide, comprising an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof.

[00215] Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:111. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:111, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:111. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:111. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:111.

[00216] Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:111. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:111. Some embodiments of the

disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:111, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:111, or a codon degenerate nucleotide sequence thereof.

[00217] Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:225. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:225, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:225. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:225. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:225.

[00218] Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:225. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:225. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:225, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4 nucleic acid comprising a nucleotide sequence

having at least 95% sequence identity to SEQ ID NO:225, or a codon degenerate nucleotide sequence thereof.

[00219] Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:221. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:221, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:221. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:221. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:221.

[00220] Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:221. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:221. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:221, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:221, or a codon degenerate nucleotide sequence thereof.

[00221] Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising the nucleotide sequence set forth in SEQ ID

NO: 224. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising the nucleotide sequence set forth in SEQ ID NO:224, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:224. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:224. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:224.

[00222] Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:224. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:224. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:224, or a codon degenerate nucleotide sequence thereof. Some embodiments of the disclosure relate to an expression construct comprising a CsPT4t nucleic acid comprising a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:224, or a codon degenerate nucleotide sequence thereof.

Polypeptides, Nucleic Acids, and Genetically Modified Host Cells for the Production of Cannabinoids, Cannabinoid Derivatives, Cannabinoid Precursors, or Cannabinoid Precursor Derivatives

[00223] The present disclosure provides genetically modified host cells for producing a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor

derivative. A genetically modified host cell of the present disclosure may be genetically modified with one or more heterologous nucleic acids disclosed herein encoding one or more polypeptides disclosed herein. Culturing of the genetically modified host cell in a suitable medium provides for synthesis of the cannabinoid, the cannabinoid derivative, the cannabinoid precursor, or the cannabinoid precursor derivative in a recoverable amount. In some embodiments, the genetically modified host cell of the disclosure produces a cannabinoid or a cannabinoid derivative.

[00224] The disclosure also provides nucleic acids, which can be introduced into microorganisms (e.g., genetically modified host cells), resulting in expression or overexpression of the one or more polypeptides, which can then be utilized *in vitro* (e.g., cell-free) or *in vivo* for the production of cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives. In certain such embodiments, cannabinoids or cannabinoid derivatives are produced.

[00225] One or more polypeptides which can be utilized for the production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative are disclosed herein, and may include, but are not limited to: one or more polypeptides having at least one activity of a polypeptide present in the cannabinoid biosynthetic pathway, such as, a GOT polypeptide, a CBDA or THCA synthase polypeptide, a TKS polypeptide, and an OAC polypeptide; one or more polypeptides having at least one activity of a polypeptide present in the mevalonate (MEV) pathway; a polypeptide that generates acyl-CoA compounds or acyl-CoA compound derivatives (e.g., an acyl-activating enzyme polypeptide, a fatty acyl-CoA synthetase polypeptide, or a fatty acyl-CoA ligase polypeptide); a polypeptide that generates GPP; a polypeptide that generates malonyl-CoA; a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA, and a pyruvate decarboxylase polypeptide. Additionally, polypeptides which can be utilized for the production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative may be one or more polypeptides having at least one activity of a polypeptide present in the DXP pathway, instead of those of the MEV pathway.

[00226] Polypeptides which can be utilized for the production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative may also include a hexanoyl-CoA synthase (HCS) polypeptide or one or more polypeptides that are part of a biosynthetic pathway that produces hexanoyl-CoA, including, but not limited to: a MCT1 polypeptide, a PaaH1 polypeptide, a Crt polypeptide, a Ter polypeptide, and a BktB polypeptide; a MCT1 polypeptide, a PhaB polypeptide, a PhaJ polypeptide, a Ter

polypeptide, and a BktB polypeptide; a short chain fatty acyl-CoA thioesterase (SCFA-TE) polypeptide; or a fatty acid synthase (FAS) polypeptide. Hexanoyl CoA derivatives, acyl-CoA compounds, or acyl-CoA compound derivatives may also be formed via such pathways and polypeptides.

[00227] Polypeptides which can be utilized for the production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative may also include polypeptides that modulate NADH or NADPH redox balance, polypeptides that generate neryl pyrophosphate, and NphB polypeptides.

[00228] The disclosure also provides nucleic acids encoding said polypeptides which can be utilized for the production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative. The disclosure also provides genetically modified host cells comprising one or more of said nucleic acids and polypeptides which can be utilized for the production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative.

Geranyl Pyrophosphate:Olivetolic Acid Geranyltransferase (GOT) Polypeptides, Nucleic Acids, and Genetically Modified Host Cells Expressing Said Polypeptides

[00229] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase (GOT) polypeptide.

[00230] Exemplary GOT polypeptides disclosed herein may include a full-length GOT polypeptide, a fragment of a GOT polypeptide, a variant of a GOT polypeptide, a truncated GOT polypeptide, or a fusion polypeptide that has at least one activity of a GOT polypeptide. In some embodiments, the GOT polypeptide has aromatic prenyltransferase (PT) activity. In some embodiments, the GOT polypeptide modifies a cannabinoid precursor or a cannabinoid precursor derivative. In certain such embodiments, the GOT polypeptide modifies olivetolic acid or an olivetolic acid derivative. In some embodiments, the GOT polypeptide cannot catalyze the production of 5-geranyl olivetolic acid.

[00231] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or

more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least 200-500 times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least 10, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 200, at least 300, at least 400, or at least 500 times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least 10-50, at least 50-100, at least 100-200, at least 100-300, at least 100-400, at least 200-400, at least 100-500, at least 200-500, or at least 300-500 times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[00232] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:100 or SEQ ID NO:110. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:100 or SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:100 or SEQ ID NO:110.

[00233] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:12, SEQ ID NO:82, SEQ ID NO:98, SEQ ID NO:99, or SEQ ID NO:223. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:12, SEQ ID NO:82, SEQ ID NO:98, SEQ ID NO:99, or SEQ ID NO:223, or a conservatively substituted amino acid

sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:12, SEQ ID NO:82, SEQ ID NO:98, SEQ ID NO:99, or SEQ ID NO:223.

[00234] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:13, SEQ ID NO:101, SEQ ID NO:102, or SEQ ID NO:103. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:13, SEQ ID NO:101, SEQ ID NO:102, or SEQ ID NO:103, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:13, SEQ ID NO:101, SEQ ID NO:102, or SEQ ID NO:103.

[00235] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:211, SEQ ID NO:213, SEQ ID NO:215, SEQ ID NO:217, or SEQ ID NO:219. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:211, SEQ ID NO:213, SEQ ID NO:215, SEQ ID NO:217, or SEQ ID NO:219, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least

2018256863 11 Apr 2024

99.9%, or 100% amino acid sequence identity to SEQ ID NO:211, SEQ ID NO:213, SEQ ID NO:215, SEQ ID NO:217, or SEQ ID NO:219.

[00236] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:12. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:12, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:12. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:12. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:12.

[00237] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:13. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:13, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:13. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:13. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:13.

2018256863 11 Apr 2024

[00238] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT1 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT1 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:82, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT1 polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:82. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT1 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:82. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT1 polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:82.

[00239] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a truncated CsPT1 (CsPT1_t75) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:223. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT1_t75 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:223, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT1_t75 polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:223. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT1_t75 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:223. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT1_t75 polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:223.

[00240] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsGOTt75 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:98. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsGOTt75 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:98, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsGOTt75 polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:98. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsGOTt75 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:98. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsGOTt75 polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:98.

[00241] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsGOTt33 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:99. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsGOTt33 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:99, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsGOTt33 polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:99. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsGOTt33 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:99. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsGOTt33 polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:99.

[00242] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4t polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:100. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4t polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4t polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:100. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4t polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:100. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4t polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:100. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4t polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:100.

[00243] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4t polypeptide and comprises an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:100. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4t polypeptide and comprises an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:100. In some embodiments, the GOT polypeptide

encoded by the one or more heterologous nucleic acids is a CsPT4t polypeptide and comprises an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:100. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4t polypeptide and comprises an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:100. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4t polypeptide and comprises an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4t polypeptide and comprises an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4t polypeptide and comprises an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4t polypeptide and comprises an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof.

[00244] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT7t polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:101. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT7t polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:101, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT7t polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:101. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT7t polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:101. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT7t polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:101.

[00245] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a H1PT1Lt polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:102. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a H1PT1Lt polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:102, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a H1PT1Lt polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:102. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a H1PT1Lt polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:102. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a H1PT1Lt polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:102.

[00246] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a H1PT2Lt polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:103. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a H1PT2Lt polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:103, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a H1PT2Lt polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:103. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a H1PT2Lt polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:103. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a H1PT2Lt polypeptide and comprises an

amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:103.

[00247] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:110. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4 polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:110. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:110. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4 polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:110. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4 polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:110.

[00248] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4 polypeptide and comprises an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:110. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4 polypeptide and comprises an amino acid sequence having at least 75% amino

acid sequence identity to SEQ ID NO:110. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4 polypeptide and comprises an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:110. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4 polypeptide and comprises an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:110. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4 polypeptide and comprises an amino acid sequence having at least 65% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4 polypeptide and comprises an amino acid sequence having at least 75% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4 polypeptide and comprises an amino acid sequence having at least 85% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4 polypeptide and comprises an amino acid sequence having at least 95% amino acid sequence identity to SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof.

[00249] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a truncated CsPT4 (CsPT4_t112) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:211. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t112 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:211, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t112 polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:211. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t112 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:211. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t112 polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:211.

[00250] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a truncated CsPT4 (CsPT4_t131) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:213. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t131 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:213, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t131 polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:213. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t131 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:213. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t131 polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:213.

[00251] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a truncated CsPT4 (CsPT4_t142) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:215. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t142 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:215, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t142 polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:215. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t142 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:215. In

some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t142 polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:215.

[00252] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a truncated CsPT4 (CsPT4_t166) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:217. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t166 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:217, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t166 polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:217. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t166 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:217. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t166 polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:217.

[00253] In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a truncated CsPT4 (CsPT4_t186) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:219. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t186 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:219, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t186 polypeptide and comprises an amino acid sequence having at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:219. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t186

polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:219. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids is a CsPT4_t186 polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:219.

[00254] Exemplary GOT heterologous nucleic acids disclosed herein may include nucleic acids that encode a GOT polypeptide, such as, a full-length GOT polypeptide, a fragment of a GOT polypeptide, a variant of a GOT polypeptide, a truncated GOT polypeptide, or a fusion polypeptide that has at least one activity of a GOT polypeptide.

[00255] In some embodiments, the GOT polypeptide is overexpressed in the genetically modified host cell. Overexpression may be achieved by increasing the copy number of the GOT polypeptide-encoding heterologous nucleic acid, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the GOT polypeptide-encoding heterologous nucleic acid to a strong promoter. In some embodiments, the genetically modified host cell has one copy of a GOT polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has two copies of a GOT polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has three copies of a GOT polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of a GOT polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has five copies of a GOT polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of a GOT polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has six copies of a GOT polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of a GOT polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has seven copies of a GOT polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of a GOT polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has eight copies of a GOT polypeptide-encoding heterologous nucleic acid.

[00256] In some embodiments, the one or more heterologous nucleic acids encoding a GOT polypeptide comprise a nucleotide sequence encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[00257] In some embodiments, the one or more heterologous nucleic acids encoding a GOT polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:111, SEQ ID NO:221, SEQ ID NO:224, or SEQ ID NO:225. In some embodiments, the one or more heterologous nucleic acids encoding a GOT polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:111, SEQ ID NO:221, SEQ ID NO:224, or SEQ ID NO:225, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GOT polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:111, SEQ ID NO:221, SEQ ID NO:224, or SEQ ID NO:225.

[00258] In some embodiments, the one or more heterologous nucleic acids encoding a GOT polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:220 or SEQ ID NO:222. In some embodiments, the one or more heterologous nucleic acids encoding a GOT polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:220 or SEQ ID NO:222, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GOT polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:220 or SEQ ID NO:222.

[00259] In some embodiments, the one or more heterologous nucleic acids encoding a GOT polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:210, SEQ ID NO:212, SEQ ID NO:214, SEQ ID NO:216, or SEQ ID NO:218. In some embodiments, the one or more heterologous nucleic acids encoding a GOT polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:210, SEQ ID NO:212, SEQ ID NO:214, SEQ

ID NO:216, or SEQ ID NO:218, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GOT polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:210, SEQ ID NO:212, SEQ ID NO:214, SEQ ID NO:216, or SEQ ID NO:218.

[00260] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:111. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:111, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:111. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:111. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:111.

[00261] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:111. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:111. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:111, or a codon degenerate nucleotide sequence thereof. In

some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:111, or a codon degenerate nucleotide sequence thereof.

[00262] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:225. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:225, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:225. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:225. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:225.

[00263] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:225. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:225. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:225, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4 polypeptide comprise a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:225, or a codon degenerate nucleotide sequence thereof.

[00264] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:221. In some

embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:221, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:221. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:221. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:221.

[00265] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:221. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:221. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:221, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:221, or a codon degenerate nucleotide sequence thereof.

[00266] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:224. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:224, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:224.

In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:224. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:224.

[00267] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:224. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:224. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise a nucleotide sequence having at least 85% sequence identity to SEQ ID NO:224, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4t polypeptide comprise a nucleotide sequence having at least 95% sequence identity to SEQ ID NO:224, or a codon degenerate nucleotide sequence thereof.

[00268] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t112 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:210. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t112 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:210, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t112 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:210. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t112 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least

99.9%, or 100% sequence identity to SEQ ID NO:210. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t112 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:210.

[00269] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t131 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:212. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t131 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:212, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t131 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:212. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t131 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:212. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t131 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:212.

[00270] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t142 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:214. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t142 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:214, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t142 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:214. In some embodiments, the one or more heterologous

nucleic acids encoding a CsPT4_t142 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:214. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t142 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:214.

[00271] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t166 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:216. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t166 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:216, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t166 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:216. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t166 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:216. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t166 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:216.

[00272] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t186 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:218. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t186 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:218, or a codon

degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t186 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:218. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t186 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:218. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT4_t186 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:218.

[00273] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:220. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:220, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT1 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:220. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT1 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:220. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT1 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:220.

2018256863 11 Apr 2024

[00274] In some embodiments, the one or more heterologous nucleic acids encoding a CsPT1_t75 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:222. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT1_t75 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:222, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT1_t75 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:222. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT1_t75 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:222. In some embodiments, the one or more heterologous nucleic acids encoding a CsPT1_t75 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:222.

Cannabinoid Synthase Polypeptides, Nucleic Acids Comprising Said Polypeptides, and Genetically Modified Host Cells Expressing Said Polypeptides

[00275] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a cannabinoid synthase polypeptide.

[00276] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than one cannabinoid synthase polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than two cannabinoid synthase polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than three cannabinoid synthase polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids

encoding two cannabinoid synthase polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding three cannabinoid synthase polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, 3, or more cannabinoid synthase polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, or 3 cannabinoid synthase polypeptides.

[00277] In some embodiments, a cannabinoid synthase polypeptide is a tetrahydrocannabinolic acid synthase (THCAS) polypeptide. THCAS polypeptides can catalyze the conversion of cannabigerolic acid to THCA. Exemplary THCAS polypeptides disclosed herein may include a fragment of a THCAS polypeptide, a full-length THCAS polypeptide, a variant of a THCAS polypeptide, a truncated THCAS polypeptide, or a fusion polypeptide that has at least one activity of a THCAS polypeptide.

[00278] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a THCAS polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than one THCAS polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than two THCAS polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than three THCAS polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding two THCAS polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding three THCAS polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, 3, or more THCAS polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, or 3 THCAS polypeptides.

[00279] In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:14,

SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155, or a conservatively substituted amino acid sequence thereof. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155.

[00280] In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:14. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:14, or a conservatively substituted amino acid sequence thereof. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:14. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:14. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:14.

[00281] In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:86. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:86, or a conservatively substituted amino acid sequence thereof. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid

sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:86. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:86. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:86.

[00282] In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:155. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:155, or a conservatively substituted amino acid sequence thereof. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:155. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:155. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:155.

[00283] In some embodiments, the THCAS polypeptide may include a modified THCAS polypeptide with an N-terminal truncation to remove the secretion peptide and localize to cytoplasm. For example, in some embodiments, the THCAS polypeptide lacks N-terminal amino acids 1-28 of the amino acid sequence set forth in SEQ ID NO:14, or a corresponding signal peptide of another THCAS polypeptide.

[00284] In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:15. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous

nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:15, or a conservatively substituted amino acid sequence thereof. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:15. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:15. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:15.

[00285] In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO: 104. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO: 104, or a conservatively substituted amino acid sequence thereof. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:104. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:104. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:104.

[00286] In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO: 153. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO: 153, or a conservatively substituted amino acid sequence thereof. In some embodiments, the THCAS

polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:153. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:153. In some embodiments, the THCAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:153.

[00287] Exemplary THCAS heterologous nucleic acids disclosed herein may include nucleic acids that encode a THCAS polypeptide, such as, a fragment of a THCAS polypeptide, a variant of a THCAS polypeptide, a full-length THCAS polypeptide, a truncated THCAS polypeptide, or a fusion polypeptide that has at least one activity of a THCAS polypeptide.

[00288] In some embodiments, the THCAS polypeptide is overexpressed in the genetically modified host cell. Overexpression may be achieved by increasing the copy number of the THCAS polypeptide-encoding heterologous nucleic acid, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the THCAS polypeptide-encoding heterologous nucleic acid to a strong promoter. In some embodiments, the genetically modified host cell has one copy of a THCAS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has two copies of a THCAS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has three copies of a THCAS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of a THCAS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has five copies of a THCAS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has six copies of a THCAS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has seven copies of a THCAS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has eight copies of a THCAS polypeptide-encoding heterologous nucleic acid.

2018256863 11 Apr 2024

[00289] In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:85, SEQ ID NO:154, or SEQ ID NO:156. In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:85, SEQ ID NO:154, or SEQ ID NO:156, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:85, SEQ ID NO:154, or SEQ ID NO:156.

[00290] In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:85. In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:85, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:85. In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:85.

[00291] In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:154. In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:154, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:154. In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:154.

[00292] In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:156. In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:156, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:156. In some embodiments, the one or more heterologous nucleic acids encoding a THCAS polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:156.

[00293] In some embodiments, a cannabinoid synthase polypeptide is cannabidiolic acid synthase (CBDAS) polypeptide. CBDAS polypeptides can catalyze the conversion of cannabigerolic acid to cannabidiolic acid (CBDA). Exemplary CBDAS polypeptides disclosed herein may include a full-length CBDAS polypeptide, a fragment of a CBDAS polypeptide, a variant of a CBDAS polypeptide, a truncated CBDAS polypeptide, or a fusion polypeptide that has at least one activity of a CBDAS polypeptide.

[00294] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a CBDAS polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than one CBDAS polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than two CBDAS polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than three CBDAS polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding two CBDAS polypeptides.

2018256863 11 Apr 2024

In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding three CBDAS polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, 3, or more CBDAS polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, or 3 CDBAS polypeptides.

[00295] In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:88 or SEQ ID NO:151. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:88 or SEQ ID NO:151, or a conservatively substituted amino acid sequence thereof. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

[00296] In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:88. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:88, or a conservatively substituted amino acid sequence thereof. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:88. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:88. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:88.

[00297] In some embodiments, the CBDAS polypeptide may include a modified CBDAS polypeptide with an N-terminal truncation to remove the secretion peptide and localize to cytoplasm. For example, in some embodiments, the CBDAS polypeptide lacks N-terminal amino acids 1-28 of the amino acid sequence set forth in SEQ ID NO:88, or a corresponding signal peptide of another CBDAS polypeptide.

[00298] In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:16. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:16, or a conservatively substituted amino acid sequence thereof. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:16. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:16. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:16.

[00299] In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:105. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:105, or a conservatively substituted amino acid sequence thereof. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:105. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:105. In some embodiments, the CBDAS

polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:105.

[00300] In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:151. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:151, or a conservatively substituted amino acid sequence thereof. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:151. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:151. In some embodiments, the CBDAS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:151.

[00301] Exemplary CBDAS heterologous nucleic acids disclosed herein may include nucleic acids that encode a CBDAS polypeptide, such as, a full-length CBDAS polypeptide, a fragment of a CBDAS polypeptide, a variant of a CBDAS polypeptide, a truncated CBDAS polypeptide, or a fusion polypeptide that has at least one activity of a CBDAS polypeptide.

[00302] In some embodiments, the CBDAS polypeptide is overexpressed in the genetically modified host cell. Overexpression may be achieved by increasing the copy number of the CBDAS polypeptide-encoding heterologous nucleic acid, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the CBDAS polypeptide-encoding heterologous nucleic acid to a strong promoter. In some embodiments, the genetically modified host cell has one copy of a CBDAS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has two copies of a CBDAS polypeptide-encoding

heterologous nucleic acid. In some embodiments, the genetically modified host cell has three copies of a CBDAS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of a CBDAS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has five copies of a CBDAS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has six copies of a CBDAS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has seven copies of a CBDAS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has eight copies of a CBDAS polypeptide-encoding heterologous nucleic acid.

[00303] In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:152 or SEQ ID NO:167. In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:152 or SEQ ID NO:167, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:152 or SEQ ID NO:167.

[00304] In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:87. In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:87, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:87. In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:87.

[00305] In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:152. In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:152, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:152. In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:152.

[00306] In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:167. In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:167, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:167. In some embodiments, the one or more heterologous nucleic acids encoding a CBDAS polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:167.

[00307] In some embodiments, at least one of the heterologous nucleic acids encoding a cannabinoid synthase polypeptide is operably linked to an inducible promoter. In some embodiments, at least one of the heterologous nucleic acids encoding a cannabinoid synthase polypeptide is operably linked to a constitutive promoter. In some embodiments, a signal peptide is linked to the N-terminus of a THCAS or CBDAS polypeptide or other cannabinoid synthase polypeptide.

Polypeptides that Generate Acyl-CoA Compounds or Acyl-CoA Compound Derivatives, Nucleic Acids Comprising Said Polypeptides, and Genetically Modified Host Cells Expressing Said Polypeptides

[00308] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a polypeptide that generates acyl-CoA compounds or acyl-CoA compound derivatives. Such polypeptides may include, but are not limited to, acyl-activating enzyme (AAE) polypeptides, fatty acyl-CoA synthetases (FAA) polypeptides, or fatty acyl-CoA ligase polypeptides.

[00309] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding an AAE, FAA, or fatty acyl-CoA ligase polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than one AAE, FAA, or fatty acyl-CoA ligase polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than two AAE, FAA, or fatty acyl-CoA ligase polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than three AAE, FAA, or fatty acyl-CoA ligase polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding two AAE, FAA, or fatty acyl-CoA ligase polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding three AAE, FAA, or fatty acyl-CoA ligase polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, 3, or more AAE, FAA, or fatty acyl-CoA ligase polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, or 3 AAE, FAA, or fatty acyl-CoA ligase polypeptides.

[00310] AAE polypeptides, FAA polypeptides, and fatty acyl-CoA ligase polypeptides can convert carboxylic acids to their CoA forms and generate acyl-CoA compounds or acyl-CoA compound derivatives. Promiscuous acyl-activating enzyme polypeptides, such as CsAAE1 and CsAAE3, FAA polypeptides, or fatty acyl-CoA ligase

polypeptides, may permit generation of cannabinoid derivatives (e.g., cannabigerolic acid derivatives) or cannabinoid precursor derivatives (e.g., olivetolic acid derivatives), as well as cannabinoids (e.g., cannabigerolic acid) or precursors thereof (e.g., olivetolic acid). In some embodiments, hexanoic acid or carboxylic acids other than hexanoic acid are fed to genetically modified host cells expressing an AAE polypeptide, FAA polypeptide, or fatty acyl-CoA ligase polypeptide (e.g., are present in the culture medium in which the cells are grown) to generate hexanoyl-CoA, acyl-CoA compounds, derivatives of hexanoyl-CoA, or derivatives of acyl-CoA compounds. In certain such embodiments, the cell culture medium comprising the genetically modified host cells comprises hexanoate. In some embodiments, the cell culture medium comprising the genetically modified host cells comprises a carboxylic acid other than hexanoate.

[00311] Exemplary AAE, FAA, or fatty acyl-CoA ligase polypeptides disclosed herein may include a full-length AAE, FAA, or fatty acyl-CoA ligase polypeptide; a fragment of a AAE, FAA, or fatty acyl-CoA ligase polypeptide; a variant of a AAE, FAA, or fatty acyl-CoA ligase polypeptide; a truncated AAE, FAA, or fatty acyl-CoA ligase polypeptide; or a fusion polypeptide that has at least one activity of an AAE, FAA, or fatty acyl-CoA ligase polypeptide.

[00312] In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE1 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:90. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE1 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:90, or a conservatively substituted amino acid sequence thereof. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE1 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:90. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE1 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:90. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE1 polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100%

amino acid sequence identity to SEQ ID NO:90. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE1 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:90.

[00313] In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:92 or SEQ ID NO:149. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:92 or SEQ ID NO:149, or a conservatively substituted amino acid sequence thereof. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:92 or SEQ ID NO:149.

[00314] In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:92. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:92, or a conservatively substituted amino acid sequence thereof. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:92. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:92. In some

embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:92.

[00315] In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:112. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:112, or a conservatively substituted amino acid sequence thereof. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:112. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:112. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:112. In these proceeding embodiments, the CsAAE3 polypeptide lacks the RELIQKVRSNM C-terminal amino acids.

[00316] In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:149. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:149, or a conservatively substituted amino acid sequence thereof. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:149. In some embodiments, the AAE

polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:149. In some embodiments, the AAE polypeptide encoded by the one or more heterologous nucleic acids is a CsAAE3 polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:149. In these proceeding embodiments, the CsAAE3 polypeptide lacks the RRELIQKVRSNM C-terminal amino acids.

[00317] In some embodiments, the fatty acyl-CoA ligase polypeptide encoded by the one or more heterologous nucleic acids is a FADK polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:145 or SEQ ID NO:147. In some embodiments, the fatty acyl-CoA ligase polypeptide encoded by the one or more heterologous nucleic acids is a FADK polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:145 or SEQ ID NO:147, or a conservatively substituted amino acid sequence thereof. In some embodiments, the fatty acyl-CoA ligase polypeptide encoded by the one or more heterologous nucleic acids is a FADK polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:145 or SEQ ID NO:147.

[00318] In some embodiments, the fatty acyl-CoA ligase polypeptide encoded by the one or more heterologous nucleic acids is a FADK polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:145. In some embodiments, the fatty acyl-CoA ligase polypeptide encoded by the one or more heterologous nucleic acids is a FADK polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:145, or a conservatively substituted amino acid sequence thereof. In some embodiments, the fatty acyl-CoA ligase polypeptide encoded by the one or more heterologous nucleic acids is a FADK polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:145. In some embodiments, the fatty acyl-CoA ligase polypeptide encoded by the one or more

heterologous nucleic acids is a FADK polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:145. In some embodiments, the fatty acyl-CoA ligase polypeptide encoded by the one or more heterologous nucleic acids is a FADK polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:145.

[00319] In some embodiments, the fatty acyl-CoA ligase polypeptide encoded by the one or more heterologous nucleic acids is a FADK polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:147. In some embodiments, the fatty acyl-CoA ligase polypeptide encoded by the one or more heterologous nucleic acids is a FADK polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:147, or a conservatively substituted amino acid sequence thereof. In some embodiments, the fatty acyl-CoA ligase polypeptide encoded by the one or more heterologous nucleic acids is a FADK polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:147. In some embodiments, the fatty acyl-CoA ligase polypeptide encoded by the one or more heterologous nucleic acids is a FADK polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:147. In some embodiments, the fatty acyl-CoA ligase polypeptide encoded by the one or more heterologous nucleic acids is a FADK polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:147.

[00320] In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200, or a

conservatively substituted amino acid sequence thereof. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200.

[00321] In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA2 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:169. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA2 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:169, or a conservatively substituted amino acid sequence thereof. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA2 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:169. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA2 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:169. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA2 polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:169. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA2 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:169.

[00322] In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a truncated FAA2 (tFAA2) polypeptide. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a tFAA2 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:194. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a tFAA2 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:194, or a conservatively substituted amino acid sequence thereof. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a tFAA2 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:194. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a tFAA2 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:194. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a tFAA2 polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:194. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a tFAA2 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:194.

[00323] In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a mutated FAA2 (FAA2mut) polypeptide. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA2mut polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:196. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA2mut polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:196, or a conservatively substituted amino acid sequence

thereof. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA2mut polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:196. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA2mut polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:196. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA2mut polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:196. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA2mut polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:196.

[00324] In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA1 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:192. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA1 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:192, or a conservatively substituted amino acid sequence thereof. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA1 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:192. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA1 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:192. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA1 polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:192. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA1 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:192.

[00325] In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA3 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:198. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA3 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:198, or a conservatively substituted amino acid sequence thereof. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA3 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:198. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA3 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:198. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA3 polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:198. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA3 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:198.

[00326] In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA4 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:200. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA4 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:200, or a conservatively substituted amino acid sequence thereof. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA4 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:200. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA4 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:200. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA4 polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:200. In some embodiments, the FAA polypeptide encoded by the one or more heterologous nucleic acids is a FAA4 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:200.

[00327] Exemplary AAE, FAA, or fatty acyl-CoA ligase heterologous nucleic acids disclosed herein may include nucleic acids that encode an AAE, FAA, or fatty acyl-CoA ligase polypeptide, such as, a full-length AAE, FAA, or fatty acyl-CoA ligase polypeptide; a fragment of a AAE, FAA, or fatty acyl-CoA ligase polypeptide; a variant of a AAE, FAA, or fatty acyl-CoA ligase polypeptide; a truncated AAE, FAA, or fatty acyl-CoA ligase polypeptide; or a fusion polypeptide that has at least one activity of an AAE, FAA, or fatty acyl-CoA ligase polypeptide.

[00328] In some embodiments, the AAE, FAA, or fatty acyl-CoA ligase polypeptide is overexpressed in the genetically modified host cell. Overexpression may be achieved by increasing the copy number of the AAE, FAA, or fatty acyl-CoA ligase polypeptide-encoding heterologous nucleic acid, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the AAE, FAA, or fatty acyl-CoA ligase polypeptide-encoding heterologous nucleic acid to a strong promoter. In some embodiments, the genetically modified host cell has one copy of an AAE, FAA, or fatty acyl-CoA ligase polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has two copies of an AAE, FAA, or fatty acyl-CoA ligase polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has three copies of an AAE, FAA, or fatty acyl-CoA ligase polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of an AAE, FAA, or fatty acyl-CoA ligase polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has five copies of an AAE, FAA, or fatty acyl-CoA ligase polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has six copies of an AAE, FAA, or fatty acyl-CoA ligase polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has seven copies of an AAE, FAA, or fatty acyl-CoA ligase polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has eight copies of an AAE, FAA, or fatty acyl-CoA ligase polypeptide-encoding heterologous nucleic acid.

[00329] In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:164 or SEQ ID NO:165. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:164 or SEQ ID NO:165, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:164 or SEQ ID NO:165.

[00330] In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:89. In some

2018256863 11 Apr 2024

embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:89, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:89. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:89.

[00331] In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:164. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:164, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:164. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:164.

[00332] In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:165. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:165, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:165. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE1 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:165.

[00333] In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:150 or SEQ ID NO:166. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:150 or SEQ ID NO:166, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO: 150 or SEQ ID NO:166.

[00334] In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO: 91. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO: 91, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:91. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:91.

[00335] In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:150. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:150, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence

identity to SEQ ID NO:150. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:150.

[00336] In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:166. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:166, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:166. In some embodiments, the one or more heterologous nucleic acids encoding a CsAAE3 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:166.

[00337] In some embodiments, the one or more heterologous nucleic acids encoding a FADK polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:146 or SEQ ID NO:148. In some embodiments, the one or more heterologous nucleic acids encoding a FADK polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:146 or SEQ ID NO:148, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a FADK polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:146 or SEQ ID NO:148.

[00338] In some embodiments, the one or more heterologous nucleic acids encoding a FADK polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:146. In some embodiments, the one or more heterologous nucleic acids encoding a FADK polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:146, or a codon degenerate

nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a FADK polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:146. In some embodiments, the one or more heterologous nucleic acids encoding a FADK polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:146.

[00339] In some embodiments, the one or more heterologous nucleic acids encoding a FADK polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:148. In some embodiments, the one or more heterologous nucleic acids encoding a FADK polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:148, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a FADK polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:148. In some embodiments, the one or more heterologous nucleic acids encoding a FADK polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:148.

[00340] In some embodiments, the one or more heterologous nucleic acids encoding a FAA polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:168, SEQ ID NO:191, SEQ ID NO:193, SEQ ID NO:195, SEQ ID NO:197, or SEQ ID NO:199. In some embodiments, the one or more heterologous nucleic acids encoding a FAA polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:168, SEQ ID NO:191, SEQ ID NO:193, SEQ ID NO:195, SEQ ID NO:197, or SEQ ID NO:199, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a FAA polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to

2018256863 11 Apr 2024

SEQ ID NO:168, SEQ ID NO:191, SEQ ID NO:193, SEQ ID NO:195, SEQ ID NO:197, or SEQ ID NO:199.

[00341] In some embodiments, the one or more heterologous nucleic acids encoding a FAA2 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:168. In some embodiments, the one or more heterologous nucleic acids encoding a FAA2 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:168, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a FAA2 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:168. In some embodiments, the one or more heterologous nucleic acids encoding a FAA2 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:168. In some embodiments, the one or more heterologous nucleic acids encoding a FAA2 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:168.

[00342] In some embodiments, the one or more heterologous nucleic acids encoding a tFAA2 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:193. In some embodiments, the one or more heterologous nucleic acids encoding a tFAA2 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:193, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a tFAA2 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:193. In some embodiments, the one or more heterologous nucleic acids encoding a tFAA2 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:193. In some embodiments, the one or more heterologous nucleic acids

encoding a tFAA2 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:193.

[00343] In some embodiments, the one or more heterologous nucleic acids encoding a FAA2mut polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:195. In some embodiments, the one or more heterologous nucleic acids encoding a FAA2mut polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:195, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a FAA2mut polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:195. In some embodiments, the one or more heterologous nucleic acids encoding a FAA2mut polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:195. In some embodiments, the one or more heterologous nucleic acids encoding a FAA2mut polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:195.

[00344] In some embodiments, the one or more heterologous nucleic acids encoding a FAA1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:191. In some embodiments, the one or more heterologous nucleic acids encoding a FAA1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:191, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a FAA1 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:191. In some embodiments, the one or more heterologous nucleic acids encoding a FAA1 polypeptide comprise a nucleotide sequence having at least 85%, at least 86%, at least 87%,

2018256863 11 Apr 2024

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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:191. In some embodiments, the one or more heterologous nucleic acids encoding a FAA1 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:191.

[00345] In some embodiments, the one or more heterologous nucleic acids encoding a FAA3 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:197. In some embodiments, the one or more heterologous nucleic acids encoding a FAA3 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:197, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a FAA3 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:197. In some embodiments, the one or more heterologous nucleic acids encoding a FAA3 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:197. In some embodiments, the one or more heterologous nucleic acids encoding a FAA3 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:197.

[00346] In some embodiments, the one or more heterologous nucleic acids encoding a FAA4 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:199. In some embodiments, the one or more heterologous nucleic acids encoding a FAA4 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:199, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic

acids encoding a FAA4 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:199. In some embodiments, the one or more heterologous nucleic acids encoding a FAA4 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:199. In some embodiments, the one or more heterologous nucleic acids encoding a FAA4 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:199.

Polypeptides that Generate or Are Part of a Pathway that Generates Hexanoyl-CoA, Hexanoyl-CoA Derivatives, Acyl-CoA Compounds, or Acyl-CoA Compound Derivatives, Nucleic Acids Comprising Said Polypeptides, and Genetically Modified Host Cells Expressing Said Polypeptides

[00347] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding one or more polypeptides that generate or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives.

[00348] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than one polypeptide that generates or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than two polypeptides that generate or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than

2018256863 11 Apr 2024

three polypeptides that generate or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than four polypeptides that generate or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than five polypeptides that generate or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding two polypeptides that generate or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding three polypeptides that generate or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding four polypeptides that generate or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding five polypeptides that generate or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, 3, 4, 5 or more polypeptides that generate or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, 3, 4, or 5 polypeptides that generate or are part of a biosynthetic pathway that generates hexanoyl-

2018256863 11 Apr 2024

CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives.

[00349] Exemplary polypeptides disclosed herein that generate or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives may include a full-length polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives; a fragment of a polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives; a variant of a polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives; a truncated polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives; or a fusion polypeptide that has at least one activity of a polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives.

[00350] In some embodiments, the one or more polypeptides that generate hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives may include a hexanoyl-CoA synthase (HCS) polypeptide (e.g., as depicted in Box 1a of FIG. 1). In some embodiments, the one or more polypeptides that generate hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives is an HCS polypeptide and the cell culture medium comprising the genetically modified host cell comprises hexanoate. In some embodiments, the one or more polypeptides that generate hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives is an HCS polypeptide and the cell culture medium comprising the genetically modified host cell comprises a carboxylic acid other than hexanoate. In some embodiments, hexanoic acid or carboxylic acids other than hexanoic acid are fed to a genetically modified host cell expressing the HCS polypeptide (e.g., are present in the culture medium in which the cells are grown) to generate hexanoyl-CoA, acyl-CoA compounds, derivatives of hexanoyl-CoA, or derivatives of acyl-CoA compounds.

[00351] In some embodiments, the HCS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:1. In some embodiments, the HCS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:1, or a conservatively

substituted amino acid sequence thereof. In some embodiments, the HCS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:1.

[00352] In some embodiments, the HCS polypeptide encoded by the one or more heterologous nucleic acids is a RevS polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:2. In some embodiments, the HCS polypeptide encoded by the one or more heterologous nucleic acids is a RevS polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:2, or a conservatively substituted amino acid sequence thereof. In some embodiments, the HCS polypeptide encoded by the one or more heterologous nucleic acids is a RevS polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:2.

[00353] In some embodiments, the HCS polypeptide encoded by the one or more heterologous nucleic acids is an AflA polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:3. In some embodiments, the HCS polypeptide encoded by the one or more heterologous nucleic acids is an AflA polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:3, or a conservatively substituted amino acid sequence thereof. In some embodiments, the HCS polypeptide encoded by the one or more heterologous nucleic acids is an AflA polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:3.

[00354] In some embodiments, the HCS polypeptide encoded by the one or more heterologous nucleic acids is an AflB polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:4. In some embodiments, the HCS polypeptide encoded by the one or more heterologous nucleic acids is an AflB polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:4, or a conservatively substituted amino acid sequence thereof. In some embodiments, the HCS polypeptide encoded by the one or more heterologous nucleic acids is an AflB polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:4.

[00355] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with: i) one or more heterologous nucleic acids that encode an AflA polypeptide and ii) one or more heterologous nucleic acids that encode an AflB polypeptide.

[00356] In some embodiments, one or more polypeptides that generate hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives comprise a MCT1 polypeptide, a PaaH1 polypeptide, a Crt polypeptide, a Ter polypeptide, and a BktB polypeptide. See, e.g., Machado et al. (2012) *Metabolic Engineering* 14:504. In some embodiments, the PaaH1 (3-hydroxyacyl-CoA dehydrogenase) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:18 or SEQ ID NO:46. In some embodiments, the PaaH1 polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:18 or SEQ ID NO:46, or a conservatively substituted amino acid sequence thereof. In some embodiments, the PaaH1 polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%,

at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:18 or SEQ ID NO:46. In some embodiments, the Crt (crotonase) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:19 or SEQ ID NO:48. In some embodiments, the Crt polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:19 or SEQ ID NO:48, or a conservatively substituted amino acid sequence thereof. In some embodiments, the Crt polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:19 or SEQ ID NO:48. In some embodiments, the Ter (trans-2-enoyl-CoA reductase) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:20 or SEQ ID NO:50. In some embodiments, the Ter polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:20 or SEQ ID NO:50, or a conservatively substituted amino acid sequence thereof. In some embodiments, the Ter polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:20 or SEQ ID NO:50. In some embodiments, the BktB (β-ketothiolase) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:21 or SEQ ID NO:44. In some embodiments, the BktB polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:21 or SEQ ID NO:44, or a conservatively substituted amino acid sequence thereof. In some embodiments, the BktB polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:21 or SEQ ID NO:44.

[00357] In some embodiments, the one or more polypeptides that generate hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives comprise a MCT1 polypeptide, a PhaB polypeptide, a PhaJ polypeptide, a Ter polypeptide, and a BktB polypeptide. In some embodiments, the PhaB (acetoacetyl-CoA reductase) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:94. In some embodiments, the PhaB polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:94, or a conservatively substituted amino acid sequence thereof. In some embodiments, the PhaB polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:94. In some embodiments, the PhaJ ((R)-specific enoyl-CoA hydratase) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:96. In some embodiments, the PhaJ polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:96, or a conservatively substituted amino acid sequence thereof. In some embodiments, the PhaJ polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:96. In some embodiments, the Ter (trans-2-

enoyl-CoA reductase) and the BktB (β-ketothiolase) polypeptides used are selected from the Ter and BktB polypeptides disclosed herein.

[00358] In some embodiments, the one or more polypeptides that generate hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives or are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives comprise a polypeptide that condenses an acetyl-CoA and a malonyl-CoA to generate acetoacetyl-CoA. Polypeptides that condense an acetyl-CoA and a malonyl-CoA to generate acetoacetyl-CoA may include a malonyl CoA-acyl carrier protein transacylase (MCT1) polypeptide. In some embodiments, the MCT1 polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:42. In some embodiments, the MCT1 polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:42, or a conservatively substituted amino acid sequence thereof. In some embodiments, the MCT1 polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:42. In some embodiments, the host cell is genetically modified with one or more heterologous nucleic acids encoding a polypeptide that condense an acetyl-CoA and a malonyl-CoA to generate acetoacetyl-CoA. In certain such embodiments, the polypeptide that condenses an acetyl-CoA and a malonyl-CoA to generate acetoacetyl-CoA is an MCT1 polypeptide.

[00359] The one or more polypeptides that generate hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives may also include a short chain fatty acyl-CoA thioesterase (SCFA-TE) polypeptide (e.g., as depicted in Box 1c of FIG. 1). In some embodiments, the SCFA-TE polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, or SEQ ID NO:31. In some embodiments, the SCFA-TE polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, or SEQ ID NO:31, or a

conservatively substituted amino acid sequence thereof. In some embodiments, the SCFA-TE polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, or SEQ ID NO:31.

[00360] In some embodiments, the one or more polypeptides that are part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives comprise a fatty acid synthase polypeptide, such as a FAS1 or FAS2 polypeptide. In some embodiments, the FAS1 polypeptide encoded by the one or more heterologous nucleic acids is a FAS1 (I306A, R1834K) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:106. In some embodiments, the FAS1 polypeptide encoded by the one or more heterologous nucleic acids is a FAS1 (I306A, R1834K) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:106, or a conservatively substituted amino acid sequence thereof. In some embodiments, the FAS1 polypeptide encoded by the one or more heterologous nucleic acids is a FAS1 (I306A, R1834K) polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:106. In some embodiments, the FAS2 polypeptide encoded by the one or more heterologous nucleic acids is a FAS2 (G1250S) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:107. In some embodiments, the FAS2 polypeptide encoded by the one or more heterologous nucleic acids is a FAS2 (G1250S) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:107, or a conservatively substituted amino acid sequence thereof. In some embodiments, the FAS2 polypeptide encoded by the one or more heterologous nucleic acids is a FAS2 (G1250S) polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:107.

[00361] Exemplary heterologous nucleic acids disclosed herein may include nucleic acids that encode a polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives, such as, a full-length polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives; a fragment of a polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives; a variant of a polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives; a truncated polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives; or a fusion polypeptide that has at least one activity of a polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives.

[00362] In some embodiments, the polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives is overexpressed in the genetically modified host cell. Overexpression may be achieved by increasing the copy number of the heterologous nucleic acid encoding a polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives encoding heterologous nucleic acid to a strong promoter. In some embodiments, the genetically modified host cell has one copy of a heterologous nucleic acid encoding a polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives. In some embodiments, the genetically modified host cell has two copies of a heterologous

nucleic acid encoding a polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives. In some embodiments, the genetically modified host cell has three copies of a heterologous nucleic acid encoding a polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives. In some embodiments, the genetically modified host cell has four copies of a heterologous nucleic acid encoding a polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives. In some embodiments, the genetically modified host cell has five copies of a heterologous nucleic acid encoding a polypeptide that generates or is part of a biosynthetic pathway that generates hexanoyl-CoA, derivatives of hexanoyl-CoA, acyl-CoA compounds, or acyl-CoA compound derivatives.

[00363] In some embodiments, the one or more heterologous nucleic acids encoding an MCT1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:41. In some embodiments, the one or more heterologous nucleic acids encoding an MCT1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:41, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an MCT1 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:41. In some embodiments, the one or more heterologous nucleic acids encoding a BktB polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:43. In some embodiments, the one or more heterologous nucleic acids encoding a BktB polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:43, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a BktB polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:43. In some embodiments, the one or more heterologous

2018256863 11 Apr 2024

nucleic acids encoding a PaaH1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:45. In some embodiments, the one or more heterologous nucleic acids encoding a PaaH1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:45, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a PaaH1 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:45. In some embodiments, the one or more heterologous nucleic acids encoding a Crt polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:47. In some embodiments, the one or more heterologous nucleic acids encoding a Crt polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:47, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a Crt polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:47. In some embodiments, the one or more heterologous nucleic acids encoding a Ter polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:49. In some embodiments, the one or more heterologous nucleic acids encoding a Ter polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:49, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a Ter polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:49. In some embodiments, the one or more heterologous nucleic acids encoding a PhaB polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:93. In some embodiments, the one or more heterologous nucleic acids encoding a PhaB polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:93, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more

heterologous nucleic acids encoding a PhaB polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:93. In some embodiments, the one or more heterologous nucleic acids encoding a PhaJ polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:95. In some embodiments, the one or more heterologous nucleic acids encoding a PhaJ polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:95, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a PhaJ polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:95.

Polypeptides that Generate Malonyl-CoA, Nucleic Acids Comprising Said Polypeptides, and Genetically Modified Host Cells Expressing Said Polypeptides

[00364] In some embodiments, the host cell is genetically modified with one or more heterologous nucleic acids encoding a polypeptide that generates malonyl-CoA. In some embodiments, the polypeptide that generates malonyl-CoA is an acetyl-CoA carboxylate (ACC) polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding an ACC polypeptide.

[00365] Exemplary ACC polypeptides disclosed herein may include a full-length ACC polypeptide, a fragment of an ACC polypeptide, a variant of an ACC polypeptide, a truncated ACC polypeptide, or a fusion polypeptide that has at least one activity of an ACC polypeptide.

[00366] In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207. In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207, or a conservatively substituted amino

acid sequence thereof. In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207.

[00367] In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:9. In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:9, or a conservatively substituted amino acid sequence thereof. In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:9. In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:9. In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:9.

[00368] In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 (S659A, S1157A) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:97. In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 (S659A, S1157A) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:97, or a conservatively substituted amino acid sequence thereof. In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 (S659A, S1157A) polypeptide and comprises an amino acid sequence having at least 50%, at least

55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:97. In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 (S659A, S1157A) polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:97. In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 (S659A, S1157A) polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:97.

[00369] In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 (S659A, S1157A) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:207. In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 (S659A, S1157A) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:207, or a conservatively substituted amino acid sequence thereof. In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 (S659A, S1157A) polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:207. In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 (S659A, S1157A) polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:207. In some embodiments, the ACC polypeptide encoded by the one or more heterologous nucleic acids is an ACC1 (S659A, S1157A) polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:207.

[00370] Exemplary ACC heterologous nucleic acids disclosed herein may include nucleic acids that encode an ACC polypeptide, such as, a full-length ACC polypeptide, a fragment of an ACC polypeptide, a variant of an ACC polypeptide, a truncated ACC polypeptide, or a fusion polypeptide that has at least one activity of an ACC polypeptide.

2018256863 11 Apr 2024

[00371] In some embodiments, the ACC polypeptide is overexpressed in the genetically modified host cell. See, e.g., Runguphan and Keasling (2014) *Metabolic Engineering* 21:103. Overexpression may be achieved by increasing the copy number of the ACC polypeptide-encoding heterologous nucleic acid, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the ACC polypeptide-encoding heterologous nucleic acid to a strong promoter. In some embodiments, the genetically modified host cell has one copy of an ACC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has two copies of an ACC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has three copies of an ACC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of an ACC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has five copies of an ACC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has six copies of an ACC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has seven copies of an ACC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has eight copies of an ACC polypeptide-encoding heterologous nucleic acid.

[00372] In some embodiments, the one or more heterologous nucleic acids encoding an ACC polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:201. In some embodiments, the one or more heterologous nucleic acids encoding an ACC polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:201, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an ACC polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:201. In some embodiments, the one or more heterologous nucleic acids encoding an ACC polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:201. In some embodiments, the one or more heterologous nucleic acids encoding an ACC polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:201.

Polypeptides that Condense an Acyl-CoA Compound or an Acyl-CoA Compound Derivative with Malonyl-CoA to Generate Olivetolic Acid or Derivatives of Olivetolic Acid, Nucleic Acids Comprising Said Polypeptides, and Genetically Modified Host Cells Expressing Said Polypeptides

[00373] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding one or more polypeptides that condense an acyl-CoA compound, such as hexanoyl-CoA, or an acyl-CoA compound derivative, such as a hexanoyl-CoA derivative, with malonyl-CoA to generate olivetolic acid, or a derivative of olivetolic acid. Polypeptides that react an acyl-CoA compound or an acyl-CoA compound derivative with malonyl-CoA to generate olivetolic acid, or a derivative of olivetolic acid, may include TKS and OAC polypeptides. TKS and OAC polypeptides have been found to have broad substrate specificity, enabling production of cannabinoid derivatives or cannabinoid precursor derivatives, in addition to cannabinoids and cannabinoid precursors.

[00374] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a TKS polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than one TKS polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than two TKS polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than three TKS polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding two TKS polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding three TKS polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, 3, or more TKS polypeptides. In some

2018256863 11 Apr 2024

embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, or 3 TKS polypeptides.

[00375] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding an OAC polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than one OAC polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than two OAC polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than three OAC polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding two OAC polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding three OAC polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, 3, or more OAC polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, or 3 OAC polypeptides.

[00376] Exemplary TKS or OAC polypeptides disclosed herein may include a full-length TKS or OAC polypeptide, a fragment of a TKS or OAC polypeptide, a variant of a TKS or OAC polypeptide, a truncated TKS or OAC polypeptide, or a fusion polypeptide that has at least one activity of a TKS or OAC polypeptide.

[00377] In some embodiments, the TKS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:11 or SEQ ID NO:76. In some embodiments, the TKS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:11 or SEQ ID NO:76, or a conservatively substituted amino acid sequence thereof. In some embodiments, the TKS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least

2018256863 11 Apr 2024

99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:11 or SEQ ID NO:76.

[00378] In some embodiments, the OAC polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:10 or SEQ ID NO:78. In some embodiments, the OAC polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:10 or SEQ ID NO:78, or a conservatively substituted amino acid sequence thereof. In some embodiments, the OAC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00379] In some embodiments, the TKS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:11. In some embodiments, the TKS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:11, or a conservatively substituted amino acid sequence thereof. In some embodiments, the TKS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:11. In some embodiments, the TKS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:11. In some embodiments, the TKS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:11.

[00380] In some embodiments, the TKS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:76. In some embodiments, the TKS polypeptide encoded by the one or more heterologous

nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:76, or a conservatively substituted amino acid sequence thereof. In some embodiments, the TKS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:76. In some embodiments, the TKS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:76. In some embodiments, the TKS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:76.

[00381] In some embodiments, the OAC polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:10. In some embodiments, the OAC polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:10, or a conservatively substituted amino acid sequence thereof. In some embodiments, the OAC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:10. In some embodiments, the OAC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:10. In some embodiments, the OAC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:10.

[00382] In some embodiments, the OAC polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:78. In some embodiments, the OAC polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:78, or a conservatively substituted amino acid sequence thereof. In some embodiments, the OAC

2018256863 11 Apr 2024

polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:78. In some embodiments, the OAC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:78. In some embodiments, the OAC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:78.

[00383] In some embodiments, the TKS and OAC polypeptides are fused into a single polypeptide chain (a TKS/OAC fusion polypeptide). In some embodiments, the TKS/OAC polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:80. In some embodiments, the TKS/OAC polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:80, or a conservatively substituted amino acid sequence thereof. In some embodiments, the TKS/OAC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:80. In some embodiments, the TKS/OAC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:80. In some embodiments, the TKS/OAC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:80. In some embodiments, the TKS/OAC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:80.

[00384] Exemplary TKS or OAC heterologous nucleic acids disclosed herein may include nucleic acids that encode a TKS or OAC polypeptide, such as, a full-length TKS or OAC polypeptide, a fragment of a TKS or OAC polypeptide, a variant of a TKS or OAC polypeptide, a truncated TKS or OAC polypeptide, or a fusion polypeptide that has at least one activity of a TKS or OAC polypeptide.

[00385] In some embodiments, the TKS or OAC polypeptide is overexpressed in the genetically modified host cell. Overexpression may be achieved by increasing the copy number of the TKS and/or OAC polypeptide-encoding heterologous nucleic acid, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the TKS and/or OAC polypeptide-encoding heterologous nucleic acid to a strong promoter. In some embodiments, the genetically modified host cell has one copy of a TKS and/or OAC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has two copies of a TKS and/or OAC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has three copies of a TKS and/or OAC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of a TKS and/or OAC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has five copies of a TKS and/or OAC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has six copies of a TKS and/or OAC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has seven copies of a TKS and/or OAC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has eight copies of a TKS and/or OAC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has nine copies of a TKS and/or OAC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has 10 copies of a TKS and/or OAC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has 11 copies of a TKS and/or OAC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has 12 copies of a TKS and/or OAC polypeptide-encoding heterologous nucleic acid.

[00386] In some embodiments, the one or more heterologous nucleic acids encoding an OAC polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:77 or SEQ

ID NO:163. In some embodiments, the one or more heterologous nucleic acids encoding an OAC polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:77 or SEQ ID NO:163, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an OAC polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:77 or SEQ ID NO:163.

[00387] In some embodiments, the one or more heterologous nucleic acids encoding a TKS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:75. In some embodiments, the one or more heterologous nucleic acids encoding a TKS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:75, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a TKS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:75. In some embodiments, the one or more heterologous nucleic acids encoding a TKS polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:75.

[00388] In some embodiments, the one or more heterologous nucleic acids encoding a TKS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:162. In some embodiments, the one or more heterologous nucleic acids encoding a TKS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:162, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a TKS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:162. In some embodiments, the one or more heterologous nucleic acids encoding a TKS polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to

SEQ ID NO:162. In some embodiments, the one or more heterologous nucleic acids encoding a TKS polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:162.

[00389] In some embodiments, the one or more heterologous nucleic acids encoding an OAC polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:77. In some embodiments, the one or more heterologous nucleic acids encoding an OAC polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:77, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an OAC polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:77. In some embodiments, the one or more heterologous nucleic acids encoding an OAC polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:77.

[00390] In some embodiments, the one or more heterologous nucleic acids encoding an OAC polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:163. In some embodiments, the one or more heterologous nucleic acids encoding an OAC polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:163, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an OAC polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:163. In some embodiments, the one or more heterologous nucleic acids encoding an OAC polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:163.

2018256863 11 Apr 2024

[00391] In some embodiments, the one or more heterologous nucleic acids encoding a TKS/OAC polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:79. In some embodiments, the one or more heterologous nucleic acids encoding a TKS/OAC polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:79, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a TKS/OAC polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:79. In some embodiments, the one or more heterologous nucleic acids encoding a TKS/OAC polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:79. In some embodiments, the one or more heterologous nucleic acids encoding a TKS/OAC polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:79.

Polypeptides that Generate Geranyl Pyrophosphate, Nucleic Acids Comprising Said Polypeptides, and Genetically Modified Host Cells Expressing Said Polypeptides

[00392] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a polypeptide that generates GPP. In some embodiments, the polypeptide that generates GPP is a geranyl diphosphate synthase (GPPS) polypeptide. In some embodiments, the GPPS polypeptide also has a farnesyl diphosphate synthase (FPPS) polypeptide activity. In some embodiments, the GPPS polypeptide is modified such that it has reduced FPPS polypeptide activity (e.g., at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or more than 90%, less FPPS polypeptide activity) than the corresponding wild-type or parental GPPS polypeptide from which the modified GPPS polypeptide is derived. In some embodiments, the GPPS polypeptide is modified such that it has substantially no FPPS polypeptide activity. In some embodiments, a genetically modified host cell of the present disclosure is

2018256863 11 Apr 2024

genetically modified with one or more heterologous nucleic acids encoding a GPPS polypeptide.

[00393] Exemplary GPPS polypeptides disclosed herein may include a full-length GPPS polypeptide, a fragment of a GPPS polypeptide, a variant of a GPPS polypeptide, a truncated GPPS polypeptide, or a fusion polypeptide that has at least one activity of a GPPS polypeptide. In some embodiments, the one or more polypeptides that generate GPP or are part of a biosynthetic pathway that generates GPP are one or more polypeptides having at least one activity of a polypeptide present in the mevalonate (MEV) pathway. In some embodiments, the one or more polypeptides that generate GPP or are part of a biosynthetic pathway that generates GPP are one or more polypeptides having at least one activity of a polypeptide present in the DXP pathway.

[00394] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60.

[00395] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:121, SEQ ID NO:123, SEQ ID NO:125, SEQ ID NO:127, SEQ ID NO:129, SEQ ID NO:131, SEQ ID NO:133, SEQ ID NO:135, SEQ ID NO:137, SEQ ID NO:139, SEQ ID NO:141, SEQ ID NO:143, or SEQ ID NO:203. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:121, SEQ ID NO:123, SEQ ID NO:125, SEQ ID NO:127, SEQ ID NO:129, SEQ ID NO:131, SEQ ID NO:133, SEQ ID NO:135, SEQ ID NO:137, SEQ ID NO:139, SEQ ID NO:141, SEQ ID NO:143, or SEQ ID NO:203, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide

encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:121, SEQ ID NO:123, SEQ ID NO:125, SEQ ID NO:127, SEQ ID NO:129, SEQ ID NO:131, SEQ ID NO:133, SEQ ID NO:135, SEQ ID NO:137, SEQ ID NO:139, SEQ ID NO:141, SEQ ID NO:143, or SEQ ID NO:203.

[00396] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:5 or SEQ ID NO:6. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:5 or SEQ ID NO:6, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:5 or SEQ ID NO:6.

[00397] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with: i) one or more heterologous nucleic acids that encode a GPPS polypeptide comprising an amino acid sequence as set forth in SEQ ID NO:5; and ii) one or more heterologous nucleic acids that encode a GPPS polypeptide comprising an amino acid sequence as set forth in SEQ ID NO:6.

[00398] In some embodiments, the GPPS (Erg20) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:7. In some embodiments, the GPPS (Erg20) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:7, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS (Erg20) polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least

2018256863 11 Apr 2024

70%, or at least 75% amino acid sequence identity to SEQ ID NO:7. In some embodiments, the GPPS (Erg20) polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:7. In some embodiments, the GPPS (Erg20) polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:7.

[00399] In some embodiments, the GPPS (Erg20 (K197G)) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:8. In some embodiments, the GPPS (Erg20 (K197G)) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:8, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS (Erg20 (K197G)) polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:8. In some embodiments, the GPPS (Erg20 (K197G)) polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:8. In some embodiments, the GPPS (Erg20 (K197G)) polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:8. The GPPS (Erg20 (K197G)) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:8 comprises a K197G amino acid substitution relative to the GPPS amino acid sequence set forth in SEQ ID NO:7. This mutation shifts the ratio of GPP to farnesyl diphosphate (FPP), increasing the production of the GPP required to produce CBDA.

[00400] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a

GPPS large subunit polypeptide and a GPPS small subunit polypeptide, where the GPPS large subunit polypeptide and the GPPS small subunit polypeptide together form a heterodimeric GPPS polypeptide. In some embodiments, the GPPS large subunit polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:72. In some embodiments, the GPPS large subunit polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:72, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS large subunit polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:72. In some embodiments, the GPPS small subunit polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:74. In some embodiments, the GPPS small subunit polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:74, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS small subunit polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:74.

[00401] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids is an ERG20mut (F96W, N127W) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:60. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids is an ERG20mut (F96W, N127W) polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:60, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids is an ERG20mut (F96W, N127W) polypeptide and comprises an amino acid sequence having at least 50%, at least

55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:60. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids is an ERG20mut (F96W, N127W) polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:60. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids is an ERG20mut (F96W, N127W) polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:60. This mutation shifts the ratio of GPP to farnesyl diphosphate (FPP), increasing the production of the GPP required to produce CBDA.

[00402] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:121. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:121, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:121. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:121. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:121.

[00403] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:123. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:123, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS

polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:123. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:123. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:123.

[00404] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:125. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:125, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:125. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:125. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:125.

[00405] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:127. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:127, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at

least 75% amino acid sequence identity to SEQ ID NO:127. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:127. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:127.

[00406] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:129. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:129, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:129. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:129. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:129.

[00407] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:131. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:131, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:131. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an

amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:131. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:131.

[00408] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:133. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:133, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:133. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:133. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:133.

[00409] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:135. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:135, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:135. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:135. In some embodiments, the GPPS

polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:135.

[00410] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:137. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:137, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:137. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:137. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:137.

[00411] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:139. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:139, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:139. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:139. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:139.

[00412] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:141. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:141, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:141. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:141. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:141.

[00413] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:143. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:143, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:143. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:143. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:143.

[00414] In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:203. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:203, or a conservatively substituted amino acid sequence thereof. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:203. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:203. In some embodiments, the GPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:203.

[00415] Exemplary GPPS heterologous nucleic acids disclosed herein may include nucleic acids that encode a GPPS polypeptide, such as, a full-length GPPS polypeptide, a fragment of a GPPS polypeptide, a variant of a GPPS polypeptide, a truncated GPPS polypeptide, or a fusion polypeptide that has at least one activity of a GPPS polypeptide.

[00416] In some embodiments, the GPPS polypeptide is overexpressed in the genetically modified host cell. Overexpression may be achieved by increasing the copy number of the GPPS polypeptide-encoding heterologous nucleic acid, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the GPPS polypeptide-encoding heterologous nucleic acid to a strong promoter. In some embodiments, the genetically modified host cell has one copy of a GPPS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has two copies of a GPPS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has three copies of a GPPS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of a GPPS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has five copies of a

GPPS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has six copies of a GPPS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has seven copies of a GPPS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has eight copies of a GPPS polypeptide-encoding heterologous nucleic acid.

[00417] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:122, SEQ ID NO:124, SEQ ID NO:126, SEQ ID NO:128, SEQ ID NO:130, SEQ ID NO:132, SEQ ID NO:134, SEQ ID NO:136, SEQ ID NO:138, SEQ ID NO:140, SEQ ID NO:142, SEQ ID NO:144, or SEQ ID NO:202. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:122, SEQ ID NO:124, SEQ ID NO:126, SEQ ID NO:128, SEQ ID NO:130, SEQ ID NO:132, SEQ ID NO:134, SEQ ID NO:136, SEQ ID NO:138, SEQ ID NO:140, SEQ ID NO:142, SEQ ID NO:144, or SEQ ID NO:202, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:122, SEQ ID NO:124, SEQ ID NO:126, SEQ ID NO:128, SEQ ID NO:130, SEQ ID NO:132, SEQ ID NO:134, SEQ ID NO:136, SEQ ID NO:138, SEQ ID NO:140, SEQ ID NO:142, SEQ ID NO:144, or SEQ ID NO:202.

[00418] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:71 and/or SEQ ID NO:73. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:71 and/or SEQ ID NO:73, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:71 and/or SEQ ID NO:73.

[00419] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide (ERG20mut (F96W, N127W)) comprise the nucleotide sequence set forth in SEQ ID NO:59. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide (ERG20mut (F96W, N127W)) comprise the nucleotide sequence set forth in SEQ ID NO:59, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide (ERG20mut (F96W, N127W)) comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:59. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide (ERG20mut (F96W, N127W)) comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:59.

[00420] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide (ERG20mut (F96W, N127W)) comprise the nucleotide sequence set forth in SEQ ID NO:161. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide (ERG20mut (F96W, N127W)) comprise the nucleotide sequence set forth in SEQ ID NO:161, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide (ERG20mut (F96W, N127W)) comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:161. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide (ERG20mut (F96W, N127W)) comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:161. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide (ERG20mut (F96W, N127W)) comprise a nucleotide sequence having 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

2018256863 11 Apr 2024

96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:161.

[00421] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:122. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:122, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:122. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:122.

[00422] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:124. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:124, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:124. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:124.

[00423] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:126. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:126, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 80%, at

2018256863 11 Apr 2024

least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:126. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:126.

[00424] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:128. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:128, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:128. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:128.

[00425] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:130. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:130, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:130. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:130.

2018256863 11 Apr 2024

[00426] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:132. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:132, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:132. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:132.

[00427] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:134. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:134, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:134. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:134.

[00428] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:136. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:136, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:136. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS

polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:136.

[00429] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:138. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:138, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:138. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:138.

[00430] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:140. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:140, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:140. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:140.

[00431] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:142. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide

2018256863 11 Apr 2024

comprise the nucleotide sequence set forth in SEQ ID NO:142, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:142. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:142.

[00432] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:144. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:144, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:144. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:144.

[00433] In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:202. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:202, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:202. In some embodiments, the one or more heterologous nucleic acids encoding a GPPS polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at

least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:202.

NphB Polypeptides, Nucleic Acids Comprising Said Polypeptides, and Genetically Modified Host Cells Expressing Said Polypeptides

[00434] In some embodiments, a NphB polypeptide is used instead of a GOT polypeptide to generate cannabigerolic acid from GPP and olivetolic acid. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a NphB polypeptide.

[00435] Exemplary NphB polypeptides disclosed herein may include a full-length NphB polypeptide, a fragment of a NphB polypeptide, a variant of a NphB polypeptide, a truncated NphB polypeptide, or a fusion polypeptide that has at least one activity of a NphB polypeptide.

[00436] In some embodiments, the NphB polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:84. In some embodiments, the NphB polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:84, or a conservatively substituted amino acid sequence thereof. In some embodiments, the NphB polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:84.

[00437] Exemplary NphB heterologous nucleic acids disclosed herein may include nucleic acids that encode a NphB polypeptide, such as, a full-length NphB polypeptide, a fragment of a NphB polypeptide, a variant of a NphB polypeptide, a truncated NphB polypeptide, or a fusion polypeptide that has at least one activity of a NphB polypeptide.

[00438] In some embodiments, the NphB polypeptide is overexpressed in the genetically modified host cell. Overexpression may be achieved by increasing the copy number of the NphB polypeptide-encoding heterologous nucleic acid, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the NphB polypeptide-encoding heterologous nucleic acid to a

2018256863 11 Apr 2024

strong promoter. In some embodiments, the genetically modified host cell has one copy of a NphB polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has two copies of a NphB polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has three copies of a NphB polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of a NphB polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has five copies of a NphB polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has six copies of a NphB polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has seven copies of a NphB polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has eight copies of a NphB polypeptide-encoding heterologous nucleic acid.

[00439] In some embodiments, the one or more heterologous nucleic acids encoding a NphB polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:83. In some embodiments, the one or more heterologous nucleic acids encoding a NphB polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:83, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a NphB polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:83.

Polypeptides that Generate Neryl Pyrophosphate or Cannabinerolic Acid, Nucleic Acids Comprising Said Polypeptides, and Genetically Modified Host Cells Expressing Said Polypeptides

[00440] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a neryl pyrophosphate (NPP) synthase (NPPS) polypeptide (FIG. 11). NPP and olivetolic acid may be substrates to generate cannabinerolic acid (CBNRA). In some embodiments, a GOT polypeptide acts on NPP and an olivetolic acid derivative (as described elsewhere herein) to generate a CBNRA derivative. Cannabinerolic acid or derivatives thereof can serve as a

2018256863 11 Apr 2024

substrate for a CBDAS or THCAS polypeptide to generate CBDA or THCA, or derivatives thereof, respectively.

[00441] Exemplary NPPS polypeptides disclosed herein may include a fragment of a NPPS polypeptide, a variant of a NPPS polypeptide, a full-length NPPS polypeptide, a truncated NPPS polypeptide, or a fusion polypeptide that has at least one activity of a NPPS polypeptide.

[00442] In some embodiments, the NPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:70. In some embodiments, the NPPS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:70, or a conservatively substituted amino acid sequence thereof. In some embodiments, the NPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:70. In some embodiments, the NPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:70. In some embodiments, the NPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:70. In some embodiments, the NPPS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:70.

[00443] Exemplary NPPS heterologous nucleic acids disclosed herein may include nucleic acids that encode a NPPS polypeptide, such as, a full-length NPPS polypeptide, a fragment of a NPPS polypeptide, a variant of a NPPS polypeptide, a truncated NPPS polypeptide, or a fusion polypeptide that has at least one activity of a NPPS polypeptide.

2018256863 11 Apr 2024

[00444] In some embodiments, the NPPS polypeptide is overexpressed in the genetically modified host cell. Overexpression may be achieved by increasing the copy number of the NPPS polypeptide-encoding heterologous nucleic acid, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the NPPS polypeptide-encoding heterologous nucleic acid to a strong promoter. In some embodiments, the genetically modified host cell has one copy of an NPPS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has two copies of an NPPS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has three copies of an NPPS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of an NPPS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has five copies of an NPPS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has six copies of an NPPS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has seven copies of an NPPS polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has eight copies of an NPPS polypeptide-encoding heterologous nucleic acid.

[00445] In some embodiments, the one or more heterologous nucleic acids encoding a NPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:69. In some embodiments, the one or more heterologous nucleic acids encoding a NPPS polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:69, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a NPPS polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:69. In some embodiments, the one or more heterologous nucleic acids encoding a NPPS polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:69. In some embodiments, the one or more heterologous nucleic acids encoding a NPPS polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:69.

Polypeptides that Generate Acetyl-CoA from Pyruvate, Nucleic Acids Comprising Said Polypeptides, and Genetically Modified Host Cells Expressing Said Polypeptides

[00446] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a polypeptide that generates acetyl-CoA from pyruvate. Polypeptides that generate acetyl-CoA from pyruvate may include a pyruvate decarboxylase (PDC) polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a PDC polypeptide.

[00447] Exemplary PDC polypeptides disclosed herein may include a full-length PDC polypeptide, a fragment of a PDC polypeptide, a variant of a PDC polypeptide, a truncated PDC polypeptide, or a fusion polypeptide that has at least one activity of a PDC polypeptide.

[00448] In some embodiments, the PDC polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:117. In some embodiments, the PDC polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:117, or a conservatively substituted amino acid sequence thereof. In some embodiments, the PDC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:117. In some embodiments, the PDC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:117. In some embodiments, the PDC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:117. In some embodiments, the PDC polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:117.

[00449] Exemplary PDC heterologous nucleic acids disclosed herein may include nucleic acids that encode a PDC polypeptide, such as, a full-length PDC polypeptide, a fragment of a PDC polypeptide, a variant of a PDC polypeptide, a truncated PDC polypeptide, or a fusion polypeptide that has at least one activity of a PDC polypeptide.

[00450] In some embodiments, the PDC polypeptide is overexpressed in the genetically modified host cell. Overexpression may be achieved by increasing the copy number of the PDC polypeptide-encoding heterologous nucleic acid, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the PDC polypeptide-encoding heterologous nucleic acid to a strong promoter. In some embodiments, the genetically modified host cell has one copy of a PDC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has two copies of a PDC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has three copies of a PDC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of a PDC polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has five copies of a PDC polypeptide-encoding heterologous nucleic acid.

[00451] In some embodiments, the one or more heterologous nucleic acids encoding a PDC polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:118. In some embodiments, the one or more heterologous nucleic acids encoding a PDC polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:118, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a PDC polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:118. In some embodiments, the one or more heterologous nucleic acids encoding a PDC polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:118. In some embodiments, the one or more heterologous nucleic acids encoding a PDC polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:118.

Polypeptides that Condense Two Molecules of Acetyl-CoA to Generate Acetoacetyl-CoA, Nucleic Acids Comprising Said Polypeptides, and Genetically Modified Host Cells Expressing Said Polypeptides

[00452] In some embodiments, the host cell is genetically modified with one or more heterologous nucleic acids encoding a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA. In some embodiments, the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase (ERG10p) polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding an acetoacetyl-CoA thiolase polypeptide.

[00453] Exemplary acetoacetyl-CoA thiolase polypeptides disclosed herein may include a full-length acetoacetyl-CoA thiolase polypeptide, a fragment of an acetoacetyl-CoA thiolase polypeptide, a variant of an acetoacetyl-CoA thiolase polypeptide, a truncated acetoacetyl-CoA thiolase polypeptide, or a fusion polypeptide that has at least one activity of an acetoacetyl-CoA thiolase polypeptide.

[00454] In some embodiments, the acetoacetyl-CoA thiolase (ERG10p) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:25. In some embodiments, the acetoacetyl-CoA thiolase (ERG10p) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:25, or a conservatively substituted amino acid sequence thereof. In some embodiments, the acetoacetyl-CoA thiolase (ERG10p) polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:25. In some embodiments, the acetoacetyl-CoA thiolase (ERG10p) polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:25. In some embodiments, the acetoacetyl-CoA thiolase (ERG10p) polypeptide encoded by the one or

more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:25. In some embodiments, the acetoacetyl-CoA thiolase (ERG10p) polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:25.

[00455] Exemplary acetoacetyl-CoA thiolase heterologous nucleic acids disclosed herein may include nucleic acids that encode an acetoacetyl-CoA thiolase polypeptide, such as, a full-length acetoacetyl-CoA thiolase polypeptide, a fragment of an acetoacetyl-CoA thiolase polypeptide, a variant of an acetoacetyl-CoA thiolase polypeptide, a truncated acetoacetyl-CoA thiolase polypeptide, or a fusion polypeptide that has at least one activity of an acetoacetyl-CoA thiolase polypeptide.

[00456] In some embodiments, the acetoacetyl-CoA thiolase polypeptide is overexpressed in the genetically modified host cell. Overexpression may be achieved by increasing the copy number of the acetoacetyl-CoA thiolase polypeptide-encoding heterologous nucleic acid, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the acetoacetyl-CoA thiolase polypeptide-encoding heterologous nucleic acid to a strong promoter. In some embodiments, the genetically modified host cell has one copy of an acetoacetyl-CoA thiolase polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has two copies of an acetoacetyl-CoA thiolase polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has three copies of an acetoacetyl-CoA thiolase polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of an acetoacetyl-CoA thiolase polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has five copies of an acetoacetyl-CoA thiolase polypeptide-encoding heterologous nucleic acid.

[00457] In some embodiments, the one or more heterologous nucleic acids encoding an acetoacetyl-CoA thiolase (ERG10p) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:157. In some embodiments, the one or more heterologous nucleic acids encoding a ERG10p polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:157, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an acetoacetyl-CoA thiolase (ERG10p) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:157. In some embodiments, the one or more heterologous nucleic acids encoding an acetoacetyl-CoA thiolase (ERG10p) polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:157. In some embodiments, the one or more heterologous nucleic acids encoding an acetoacetyl-CoA thiolase (ERG10p) polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:157.

[00458] In some embodiments, the one or more heterologous nucleic acids encoding an acetoacetyl-CoA thiolase (ERG10p) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:209. In some embodiments, the one or more heterologous nucleic acids encoding a ERG10p polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:209, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an acetoacetyl-CoA thiolase (ERG10p) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:209. In some embodiments, the one or more heterologous nucleic acids encoding an acetoacetyl-CoA thiolase (ERG10p) polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:209. In some embodiments, the one or more heterologous nucleic acids

2018256863 11 Apr 2024

encoding an acetoacetyl-CoA thiolase (ERG10p) polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:209.

Mevalonate Pathway Polypeptides, Nucleic Acids Comprising Said Polypeptides, and Genetically Modified Host Cells Expressing Said Polypeptides

[00459] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding one or more polypeptides having at least one activity of a polypeptide present in the mevalonate (MEV) pathway.

[00460] In some embodiments, the one or more polypeptides that generate GPP or are part of a biosynthetic pathway that generates GPP are one or more polypeptides having at least one activity of a polypeptide present in the mevalonate pathway. The mevalonate pathway may comprise polypeptides that catalyze the following steps: (a) condensing two molecules of acetyl-CoA to generate acetoacetyl-CoA (e.g., by action of an acetoacetyl-CoA thiolase polypeptide); (b) condensing acetoacetyl-CoA with acetyl-CoA to form hydroxymethylglutaryl-CoA (HMG-CoA) (e.g., by action of a HMGS polypeptide); (c) converting HMG-CoA to mevalonate (e.g., by action of an HMGR polypeptide); (d) phosphorylating mevalonate to mevalonate 5-phosphate (e.g., by action of a MK polypeptide); (e) converting mevalonate 5-phosphate to mevalonate 5-pyrophosphate (e.g., by action of a PMK polypeptide); (f) converting mevalonate 5-pyrophosphate to isopentenyl pyrophosphate (e.g., by action of a mevalonate pyrophosphate decarboxylase (MPD or MVD) polypeptide); and (g) converting isopentenyl pyrophosphate to dimethylallyl pyrophosphate (e.g., by action of an isopentenyl pyrophosphate isomerase (IDI) polypeptide).

[00461] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a MEV pathway polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than one MEV pathway polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more

heterologous nucleic acids encoding more than two MEV pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than three MEV pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than four MEV pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than five MEV pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than six MEV pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding all MEV pathway polypeptides.

[00462] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding two MEV pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding three MEV pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding four MEV pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding five MEV pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding six MEV pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, 3, 4, 5, 6, or more MEV pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, 3, 4, 5, or 6 MEV pathway polypeptides.

[00463] Exemplary MEV pathway polypeptides disclosed herein may include a full-length MEV pathway polypeptide, a fragment of a MEV pathway polypeptide, a variant of a MEV pathway polypeptide, a truncated MEV pathway polypeptide, or a fusion polypeptide that has at least one activity of a MEV pathway polypeptide. In some embodiments, the one or more MEV pathway polypeptides are selected from the group consisting of an

2018256863 11 Apr 2024

acetoacetyl-CoA thiolase polypeptide, a HMGS polypeptide, an HMGR polypeptide, an MK polypeptide, a PMK polypeptide, an MVD polypeptide, and an IDI polypeptide.

[00464] In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115, or a conservatively substituted amino acid sequence thereof. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115.

[00465] In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is a MvaS polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:23. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is a MvaS polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:23, or a conservatively substituted amino acid sequence thereof. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is a MvaS polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:23. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is a MvaS polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:23. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is a MvaS polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:23.

2018256863 11 Apr 2024

[00466] In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is a MvaS polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:56. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is a MvaS polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:56, or a conservatively substituted amino acid sequence thereof. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is a MvaS polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:56. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is a MvaS polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:56. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is a MvaS polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:56.

[00467] In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is an ERG13 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:24. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is an ERG13 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:24, or a conservatively substituted amino acid sequence thereof. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is an ERG13 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:24. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is an ERG13 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:24. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is an ERG13 polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:24.

[00468] In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is an ERG13 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:115. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is an ERG13 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:115, or a conservatively substituted amino acid sequence thereof. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is an ERG13 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:115. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is an ERG13 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:115. In some embodiments, the HMGS polypeptide encoded by the one or more heterologous nucleic acids is an ERG13 polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:115.

[00469] In some embodiments, the HMGR polypeptide encoded by the one or more heterologous nucleic acids is a MvaE polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:22. In some embodiments, the HMGR polypeptide encoded by the one or more heterologous nucleic acids is a MvaE polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:22, or a conservatively substituted amino acid sequence thereof. In some embodiments, the HMGR polypeptide encoded by the one or more heterologous nucleic acids is a MvaE polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:22. In some embodiments, the HMGR polypeptide encoded by the one or more heterologous nucleic acids is a MvaE polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:22. In some embodiments, the HMGR polypeptide encoded by the one or more heterologous nucleic acids is a MvaE polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:22. In some embodiments, the HMGR polypeptide encoded by the one or more heterologous nucleic acids is a MvaE polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:22.

[00470] In some embodiments, the HMGR polypeptide encoded by the one or more heterologous nucleic acids is a MvaE polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:54. In some embodiments, the HMGR polypeptide encoded by the one or more heterologous nucleic acids is a MvaE polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:54, or a conservatively substituted amino acid sequence thereof. In some embodiments, the HMGR polypeptide encoded by the one or more heterologous nucleic acids is a MvaE polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:54. In some embodiments, the HMGR polypeptide encoded by the one or more heterologous nucleic acids is a MvaE polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:54. In some embodiments, the HMGR polypeptide encoded by the one or more heterologous nucleic acids is a MvaE polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:54.

[00471] In some embodiments, the HMGR polypeptide is a truncated HMGR (tHMGR) polypeptide. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises the

amino acid sequence set forth in SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208, or a conservatively substituted amino acid sequence thereof. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208.

[00472] In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:17. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:17, or a conservatively substituted amino acid sequence thereof. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:17. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:17. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:17.

[00473] In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:52. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:52, or a conservatively substituted amino acid sequence thereof. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:52. In some embodiments, the

tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:52. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:52.

[00474] In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:113. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:113, or a conservatively substituted amino acid sequence thereof. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:113. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:113. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:113.

[00475] In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:208. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:208, or a conservatively substituted amino acid sequence thereof. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:208. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least

84% amino acid sequence identity to SEQ ID NO:208. In some embodiments, the tHMGR polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:208.

[00476] In some embodiments, the MK (ERG12) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:64. In some embodiments, the MK (ERG12) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:64, or a conservatively substituted amino acid sequence thereof. In some embodiments, the MK (ERG12) polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:64. In some embodiments, the MK (ERG12) polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:64. In some embodiments, the MK (ERG12) polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:64. In some embodiments, the MK (ERG12) polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:64.

[00477] In some embodiments, the PMK polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:62 or SEQ ID NO:205. In some embodiments, the PMK polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:62 or

SEQ ID NO:205, or a conservatively substituted amino acid sequence thereof. In some embodiments, the PMK polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:62 or SEQ ID NO:205.

[00478] In some embodiments, the PMK polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:62. In some embodiments, the PMK polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:62, or a conservatively substituted amino acid sequence thereof. In some embodiments, the PMK polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:62. In some embodiments, the PMK polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:62. In some embodiments, the PMK polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:62.

[00479] In some embodiments, the PMK polypeptide encoded by the one or more heterologous nucleic acids is an ERG8 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:205. In some embodiments, the PMK polypeptide encoded by the one or more heterologous nucleic acids is an ERG8 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:205, or a conservatively substituted amino acid sequence thereof. In some embodiments, the PMK polypeptide encoded by the one or more heterologous nucleic acids is an ERG8 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:205. In some embodiments, the PMK

2018256863 11 Apr 2024

polypeptide encoded by the one or more heterologous nucleic acids is an ERG8 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:205. In some embodiments, the PMK polypeptide encoded by the one or more heterologous nucleic acids is an ERG8 polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:205.

[00480] In some embodiments, a PMK polypeptide and MK polypeptide are fused into a single polypeptide chain (a PMK/MK fusion polypeptide). In some embodiments, the PMK/MK polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:68. In some embodiments, the PMK/MK polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:68, or a conservatively substituted amino acid sequence thereof. In some embodiments, the PMK/MK polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:68. In some embodiments, the PMK/MK polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:68. In some embodiments, the PMK/MK polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:68. In some embodiments, the PMK/MK polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:68.

2018256863 11 Apr 2024

[00481] In some embodiments, the MVD polypeptide encoded by the one or more heterologous nucleic acids is an ERG19 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:66. In some embodiments, the MVD polypeptide encoded by the one or more heterologous nucleic acids is an ERG19 polypeptide and comprises the amino acid sequence set forth in SEQ ID NO:66, or a conservatively substituted amino acid sequence thereof. In some embodiments, the MVD polypeptide encoded by the one or more heterologous nucleic acids is an ERG19 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:66. In some embodiments, the MVD polypeptide encoded by the one or more heterologous nucleic acids is an ERG19 polypeptide and comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:66. In some embodiments, the MVD polypeptide encoded by the one or more heterologous nucleic acids is an ERG19 polypeptide and comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:66. In some embodiments, the MVD polypeptide encoded by the one or more heterologous nucleic acids is an ERG19 polypeptide and comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:66.

[00482] In some embodiments, the IDI1 polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:58. In some embodiments, the IDI1 polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:58, or a conservatively substituted amino acid sequence thereof. In some embodiments, the IDI1 polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% amino acid sequence identity to SEQ ID NO:58. In some embodiments, the IDI1 polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid

sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% amino acid sequence identity to SEQ ID NO:58. In some embodiments, the IDI1 polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:58. In some embodiments, the IDI1 polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:58.

[00483] Exemplary MEV pathway heterologous nucleic acids disclosed herein may include nucleic acids that encode a MEV pathway polypeptide, such as, a full-length MEV pathway polypeptide, a fragment of a MEV pathway polypeptide, a variant of a MEV pathway polypeptide, a truncated MEV pathway polypeptide, or a fusion polypeptide that has at least one activity of a polypeptide that is part of the MEV pathway.

[00484] In some embodiments, the MEV pathway polypeptide is overexpressed in the genetically modified host cell. Overexpression may be achieved by increasing the copy number of a MEV pathway polypeptide-encoding heterologous nucleic acid, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the MEV pathway polypeptide-encoding heterologous nucleic acid to a strong promoter. In some embodiments, the genetically modified host cell has one copy of a MEV pathway polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has two copies of a MEV pathway polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has three copies of a MEV pathway polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of a MEV pathway polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has five copies of a MEV pathway polypeptide-encoding heterologous nucleic acid.

2018256863 11 Apr 2024

[00485] In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (mvaS) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:55. In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (mvaS) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:55, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (mvaS) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:55. In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (mvaS) polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:55.

[00486] In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (ERG13) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:116 or SEQ ID NO:120. In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (ERG13) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:116 or SEQ ID NO:120, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (ERG13) polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:116 or SEQ ID NO:120.

[00487] In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (ERG13) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:116. In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (ERG13) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:116, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (ERG13) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:116. In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (ERG13) polypeptide comprise a nucleotide sequence

having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:116.

[00488] In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (ERG13) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:120. In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (ERG13) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:120, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (ERG13) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:120. In some embodiments, the one or more heterologous nucleic acids encoding a HMGS (ERG13) polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:120.

[00489] In some embodiments, the one or more heterologous nucleic acids encoding an HMGR (mvaE) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:53. In some embodiments, the one or more heterologous nucleic acids encoding an HMGR (mvaE) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:53, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an HMGR (mvaE) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:53. In some embodiments, the one or more heterologous nucleic acids encoding an HMGR (mvaE) polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:53.

[00490] In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:51, SEQ ID NO:114, or SEQ ID NO:119. In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise the nucleotide sequence set forth in SEQ ID

2018256863 11 Apr 2024

NO:51, SEQ ID NO:114, or SEQ ID NO:119, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:51, SEQ ID NO:114, or SEQ ID NO:119.

[00491] In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:51. In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:51, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:51. In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:51.

[00492] In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:114. In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:114, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:114. In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:114.

2018256863 11 Apr 2024

[00493] In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:119. In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:119, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:119. In some embodiments, the one or more heterologous nucleic acids encoding a tHMGR polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:119.

[00494] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with two or more heterologous nucleic acids that encode a tHMGR polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with two heterologous nucleic acids that encode a tHMGR polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with two or more heterologous nucleic acids that encode an HMGR polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with two heterologous nucleic acids that encode an HMGR polypeptide.

[00495] In some embodiments, the one or more heterologous nucleic acids encoding an MK (ERG12) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:63 or SEQ ID NO:206. In some embodiments, the one or more heterologous nucleic acids encoding an MK (ERG12) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:63 or SEQ ID NO:206, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an MK (ERG12) polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:63 or SEQ ID NO:206.

2018256863 11 Apr 2024

[00496] In some embodiments, the one or more heterologous nucleic acids encoding an MK (ERG12) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:63. In some embodiments, the one or more heterologous nucleic acids encoding an MK (ERG12) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:63, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an MK (ERG12) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:63. In some embodiments, the one or more heterologous nucleic acids encoding an MK (ERG12) polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:63.

[00497] In some embodiments, the one or more heterologous nucleic acids encoding an MK (ERG12) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:206. In some embodiments, the one or more heterologous nucleic acids encoding an MK (ERG12) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:206, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an MK (ERG12) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:206. In some embodiments, the one or more heterologous nucleic acids encoding an MK (ERG12) polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:206.

[00498] In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:61, SEQ ID NO:160, or SEQ ID NO:204. In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:61, SEQ ID NO:160, or SEQ ID NO:204, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:61, SEQ ID NO:160, or SEQ ID NO:204.

[00499] In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:61. In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:61, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:61. In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:61.

[00500] In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:160. In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:160, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:160. In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:160.

[00501] In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:204. In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:204, or a codon

degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:204. In some embodiments, the one or more heterologous nucleic acids encoding a PMK (ERG8) polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:204.

[00502] In some embodiments, the one or more heterologous nucleic acids encoding a PMK/MK polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:67. In some embodiments, the one or more heterologous nucleic acids encoding a PMK/MK polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:67, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding a PMK/MK polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:67. In some embodiments, the one or more heterologous nucleic acids encoding a PMK/MK polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:67. In some embodiments, the one or more heterologous nucleic acids encoding a PMK/MK polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:67.

[00503] In some embodiments, the one or more heterologous nucleic acids encoding an MVD (ERG19) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:65 or SEQ ID NO:158. In some embodiments, the one or more heterologous nucleic acids encoding an MVD (ERG19) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:65 or SEQ ID NO:158, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an MVD (ERG19)

polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:65 or SEQ ID NO:158.

[00504] In some embodiments, the one or more heterologous nucleic acids encoding an MVD (ERG19) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:65. In some embodiments, the one or more heterologous nucleic acids encoding an MVD (ERG19) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:65, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an MVD (ERG19) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:65. In some embodiments, the one or more heterologous nucleic acids encoding an MVD (ERG19) polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:65.

[00505] In some embodiments, the one or more heterologous nucleic acids encoding an MVD (ERG19) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:158. In some embodiments, the one or more heterologous nucleic acids encoding an MVD (ERG19) polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:158, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an MVD (ERG19) polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:158. In some embodiments, the one or more heterologous nucleic acids encoding an MVD (ERG19) polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:158.

[00506] In some embodiments, the one or more heterologous nucleic acids encoding an IDI1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:57 or SEQ ID

NO:159. In some embodiments, the one or more heterologous nucleic acids encoding an IDI1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:57 or SEQ ID NO:159, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an IDI1 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:57 or SEQ ID NO:159.

[00507] In some embodiments, the one or more heterologous nucleic acids encoding an IDI1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:57. In some embodiments, the one or more heterologous nucleic acids encoding an IDI1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:57, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an IDI1 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:57. In some embodiments, the one or more heterologous nucleic acids encoding an IDI1 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:57.

[00508] In some embodiments, the one or more heterologous nucleic acids encoding an IDI1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:159. In some embodiments, the one or more heterologous nucleic acids encoding an IDI1 polypeptide comprise the nucleotide sequence set forth in SEQ ID NO:159, or a codon degenerate nucleotide sequence thereof. In some embodiments, the one or more heterologous nucleic acids encoding an IDI1 polypeptide comprise a nucleotide sequence having at least 80%, at least 81%, at least 82%, at least 83%, or at least 84% sequence identity to SEQ ID NO:159. In some embodiments, the one or more heterologous nucleic acids encoding an IDI1 polypeptide comprise a nucleotide sequence having 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%, at least 99.5%, at

2018256863 11 Apr 2024

least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% sequence identity to SEQ ID NO:159.

Polypeptides that Modulate NADH or NADPH Redox Balance, Nucleic Acids Comprising Said Polypeptides, and Genetically Modified Host Cells Expressing Said Polypeptides

[00509] In some embodiments, the host cell is genetically modified with one or more heterologous nucleic acids encoding a polypeptide that modulates NADH or NADPH redox balance. GPP production has a redox imbalance in it that can be modulated by changing NADPH-using enzymes to NADH-using enzymes, bringing redox into better balance.

[00510] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a polypeptide that modulates NADH or NADPH redox balance. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than one polypeptide that modulates NADH or NADPH redox balance. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than two polypeptides that modulate NADH or NADPH redox balance. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than three polypeptides that modulate NADH or NADPH redox balance. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding two polypeptides that modulate NADH or NADPH redox balance. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding three polypeptides that modulate NADH or NADPH redox balance. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, 3, or more polypeptides that modulate NADH or NADPH redox balance. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, or 3 polypeptides that modulate NADH or NADPH redox balance.

[00511] Exemplary polypeptides that modulate NADH or NADPH redox balance disclosed herein may include a full-length polypeptide that modulates NADH or NADPH redox balance, a fragment of a polypeptide that modulates NADH or NADPH redox balance,

2018256863 11 Apr 2024

a variant of a polypeptide that modulates NADH or NADPH redox balance, a truncated polypeptide that modulates NADH or NADPH redox balance, or a fusion polypeptide that has at least one activity of a polypeptide that modulates NADH or NADPH redox balance.

[00512] Exemplary heterologous nucleic acids disclosed herein may include nucleic acids that encode a polypeptide that modulates NADH or NADPH redox balance, such as, a full-length polypeptide that modulates NADH or NADPH redox balance, a fragment of a polypeptide that modulates NADH or NADPH redox balance, a variant of a polypeptide that modulates NADH or NADPH redox balance, a truncated polypeptide that modulates NADH or NADPH redox balance, or a fusion polypeptide that has at least one activity of a polypeptide that modulates NADH or NADPH redox balance.

[00513] In some embodiments, the polypeptide that modulates NADH or NADPH redox balance is overexpressed in the genetically modified host cell. Overexpression may be achieved by increasing the copy number of the heterologous nucleic acid encoding a polypeptide that modulates NADH or NADPH redox balance, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the polypeptide that modulates NADH or NADPH redox balance encoding heterologous nucleic acid to a strong promoter. In some embodiments, the genetically modified host cell has one copy of a heterologous nucleic acid encoding a polypeptide that modulates NADH or NADPH redox balance. In some embodiments, the genetically modified host cell has two copies of a heterologous nucleic acid encoding a polypeptide that modulates NADH or NADPH redox balance. In some embodiments, the genetically modified host cell has three copies of a heterologous nucleic acid encoding a polypeptide that modulates NADH or NADPH redox balance. In some embodiments, the genetically modified host cell has four copies of a heterologous nucleic acid encoding a polypeptide that modulates NADH or NADPH redox balance. In some embodiments, the genetically modified host cell has five copies of a heterologous nucleic acid encoding a polypeptide that modulates NADH or NADPH redox balance.

DXP Pathway Polypeptides, Nucleic Acids Comprising Said Polypeptides, and Genetically Modified Host Cells Expressing Said Polypeptides

[00514] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding one or more polypeptides having at least one activity of a polypeptide present in the deoxyxylulose-5-phosphate (DXP) pathway.

2018256863 11 Apr 2024

[00515] In some embodiments, the one or more polypeptides that generate GPP or are part of a biosynthetic pathway that generates GPP are polypeptides of the DXP pathway. The term "1-deoxy-D-xylulose 5-diphosphate pathway" or "DXP pathway" as used herein may refer to the pathway that converts glyceraldehyde-3-phosphate and pyruvate to IPP and DMAPP through a DXP pathway intermediate.

[00516] In the DXP pathway, pyruvate and D-glyceraldehyde-3-phosphate are converted via a series of reactions to IPP and DMAPP. The pathway involves action of the following polypeptides: a 1-deoxy-D-xylulose-5-phosphate synthase (Dxs) polypeptide, a 1-deoxy-D-xylulose-5-phosphate reductoisomerase (IspC) polypeptide, a 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase (IspD) polypeptide, a 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (IspE) polypeptide, a 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate synthase (IspF) polypeptide, a 1-hydroxy-2-methyl-2-(*E*)-butenyl 4-diphosphate synthase (IspG) polypeptide, and an isopentenyl diphosphate isomerase (IspH) polypeptide.

[00517] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding a DXP pathway polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than one DXP pathway polypeptide. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than two DXP pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than three DXP pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than four DXP pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than five DXP pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding more than six DXP pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding all DXP pathway polypeptides.

[00518] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding two

DXP pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding three DXP pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding four DXP pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding five DXP pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding six DXP pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, 3, 4, 5, 6, or more DXP pathway polypeptides. In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with one or more heterologous nucleic acids encoding 1, 2, 3, 4, 5, or 6 DXP pathway polypeptides.

[00519] Exemplary polypeptides disclosed herein that are part of the DXP pathway may include a full-length DXP pathway polypeptide, a fragment of a DXP pathway polypeptide, a variant of a DXP pathway polypeptide, a truncated DXP pathway polypeptide, or a fusion polypeptide that has at least one activity of a polypeptide that is part of the DXP pathway.

[00520] Examples of polypeptides of the DXP pathway are set forth in SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40.

[00521] In some embodiments, the 1-deoxy-D-xylulose-5-phosphate synthase (Dxs) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:32. In some embodiments, the 1-deoxy-D-xylulose-5-phosphate synthase (Dxs) polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:32, or a conservatively substituted amino acid sequence thereof. In some embodiments, the 1-deoxy-D-xylulose-5-phosphate synthase (Dxs) polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:32.

[00522] In some embodiments, the 1-deoxy-D-xylulose 5-phosphate reductoisomerase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:33. In some embodiments, the 1-deoxy-D-xylulose 5-phosphate reductoisomerase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:33, or a conservatively substituted amino acid sequence thereof. In some embodiments, the 1-deoxy-D-xylulose 5-phosphate reductoisomerase polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:33.

[00523] In some embodiments, the 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:34. In some embodiments, the 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:34, or a conservatively substituted amino acid sequence thereof. In some embodiments, the 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:34.

[00524] In some embodiments, the 4-diphosphocytidyl-2-C-methylerythritol kinase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:35. In some embodiments, the 4-diphosphocytidyl-2-C-methylerythritol kinase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:35, or a conservatively

substituted amino acid sequence thereof. In some embodiments, the 4-diphosphocytidyl-2-C-methylerythritol kinase polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:35.

[00525] In some embodiments, the 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:36. In some embodiments, the 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:36, or a conservatively substituted amino acid sequence thereof. In some embodiments, the 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:36.

[00526] In some embodiments, the 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:37. In some embodiments, the 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:37, or a conservatively substituted amino acid sequence thereof. In some embodiments, the 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least

2018256863 11 Apr 2024

99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:37.

[00527] In some embodiments, the 4-hydroxy-3-methylbut-2-enyl diphosphate reductase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:38. In some embodiments, the 4-hydroxy-3-methylbut-2-enyl diphosphate reductase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:38, or a conservatively substituted amino acid sequence thereof. In some embodiments, the 4-hydroxy-3-methylbut-2-enyl diphosphate reductase polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:38.

[00528] In some embodiments, the isopentenyl diphosphate (IPP) isomerase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:39. In some embodiments, the isopentenyl diphosphate (IPP) isomerase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:39, or a conservatively substituted amino acid sequence thereof. In some embodiments, the isopentenyl diphosphate (IPP) isomerase polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:39.

[00529] In some embodiments, the DXP pathway polypeptide is a mutated FPP synthase polypeptide. In some embodiments, the mutated FPP synthase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:40. In some embodiments, the mutated FPP synthase polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth

in SEQ ID NO:40, or a conservatively substituted amino acid sequence thereof. In some embodiments, the mutated FPP synthase isomerase polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:40.

[00530] Exemplary DXP pathway heterologous nucleic acids disclosed herein may include nucleic acids that encode a DXP pathway polypeptide, such as, a full-length DXP pathway polypeptide, a fragment of a DXP pathway polypeptide, a variant of a DXP pathway polypeptide, a truncated DXP pathway polypeptide, or a fusion polypeptide that has at least one activity of a polypeptide that is part of the DXP pathway.

[00531] In some embodiments, the DXP pathway polypeptide is overexpressed in the genetically modified host cell. Overexpression may be achieved by increasing the copy number of the DXP pathway polypeptide-encoding heterologous nucleic acid, e.g., through use of a high copy number expression vector (e.g., a plasmid that exists at 10-40 copies per cell) and/or by operably linking the DXP pathway polypeptide-encoding heterologous nucleic acid to a strong promoter. In some embodiments, the genetically modified host cell has one copy of a DXP pathway polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has two copies of a DXP pathway polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has three copies of a DXP pathway polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has four copies of a DXP pathway polypeptide-encoding heterologous nucleic acid. In some embodiments, the genetically modified host cell has five copies of a DXP pathway polypeptide-encoding heterologous nucleic acid.

Genetically Modified Host Cells to Generate Cannabinoids, Cannabinoid Derivatives, Cannabinoid Precursors, or Cannabinoid Precursor Derivatives

[00532] The disclosure provides for genetically modified host cells for producing cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives. For producing cannabinoids, cannabinoid derivatives, cannabinoid precursors,

or cannabinoid precursor derivatives, genetically modified host cells disclosed herein may be genetically modified to express or overexpress one or more heterologous nucleic acids disclosed herein that encode one or more polypeptides disclosed herein. In some embodiments, the genetically modified host cell of the disclosure produces a cannabinoid or a cannabinoid derivative. The disclosure also provides genetically modified host cells genetically modified to express or overexpress one or more heterologous nucleic acids disclosed herein that encode one or more polypeptides disclosed herein.

[00533] In some embodiments, to produce cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives, expression or overexpression of one or more heterologous nucleic acids disclosed herein that encode one or more polypeptides disclosed herein in a genetically modified host cell may be done in combination with expression or overexpression by the genetically modified host cell of one or more other heterologous nucleic acids disclosed herein that encode one or more polypeptides disclosed herein. In certain such embodiments, the genetically modified host cell produces a cannabinoid or a cannabinoid derivative.

Exemplary Genetically Modified Host Cells Expressing a GOT Polypeptide, Wherein Said GOT Polypeptide Can Catalyze Production of Cannabigerolic Acid from Geranyl Pyrophosphate and Olivetolic Acid in an Amount at Least Ten Times Higher than a Polypeptide Comprising an Amino Acid Sequence Set Forth in SEQ ID NO:82

[00534] Some embodiments of the disclosure relate to a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, the genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[00535] The disclosure also provides genetically modified host cells genetically modified to express or overexpress one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[00536] In some embodiments, the genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in

2018256863 11 Apr 2024

an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, further comprises one or more heterologous nucleic acids encoding a TKS polypeptide and one or more heterologous nucleic acids encoding an OAC polypeptide. In certain such embodiments, the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76, and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00537] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, further comprises one or more of the following: a) one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; b) one or more heterologous nucleic acids encoding a polypeptide that generates GPP; or c) one or more heterologous nucleic acids encoding a polypeptide that generates malonyl-CoA. In certain such embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is an acyl-activating enzyme (AAE) polypeptide. In certain such embodiments, the AAE polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90. In some embodiments, the AAE polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:92 or SEQ ID NO:149. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is a fatty acyl-CoA ligase polypeptide. In certain such embodiments, the fatty acyl-CoA ligase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:145 or SEQ ID NO:147. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is a fatty acyl-CoA synthetase (FAA) polypeptide. In certain such embodiments, the FAA polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide. In certain such embodiments, the GPPS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60. In some embodiments, the polypeptide that generates malonyl-CoA is an ACC polypeptide. In

2018256863 11 Apr 2024

certain such embodiments, the ACC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207.

[00538] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, further comprises one or more of the following: a) one or more heterologous nucleic acids encoding a HMGS polypeptide; b) one or more heterologous nucleic acids encoding a 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMGR) polypeptide; c) one or more heterologous nucleic acids encoding a MK polypeptide; d) a PMK polypeptide; e) one or more heterologous nucleic acids encoding a MVD polypeptide; or f) one or more heterologous nucleic acids encoding an IDI polypeptide. In certain such embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an IDI polypeptide. In certain such embodiments, the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an HMGR polypeptide. In certain such embodiments, the HMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:22. In some embodiments, the HMGR polypeptide is a truncated HMGR (tHMGR) polypeptide. In certain such embodiments, the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an HMGS polypeptide. In certain such embodiments, the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an MK polypeptide. In certain such embodiments, the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a PMK polypeptide. In certain such embodiments, the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a MVD polypeptide. In certain such embodiments, the

MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66.

[00539] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, further comprises one or more heterologous nucleic acids encoding a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA. In certain such embodiments, the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide. In certain such embodiments, the acetoacetyl-CoA thiolase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25.

[00540] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, further comprises one or more heterologous nucleic acids encoding a PDC polypeptide. In certain such embodiments, the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117.

[00541] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, further comprises one or more heterologous nucleic acids encoding a cannabinoid synthase polypeptide. In certain such embodiments, the cannabinoid synthase polypeptide is a THCA synthase polypeptide. In certain such embodiments, the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155. In some embodiments, the cannabinoid synthase polypeptide is a CBDA synthase polypeptide. In certain such embodiments, the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

2018256863 11 Apr 2024

Exemplary Genetically Modified Host Cells Expressing a Polypeptide Comprising an Amino Acid Sequence Having Sequence Identity to SEQ ID NO:110

[00542] Some embodiments of the disclosure relate to a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, the genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110. In certain such embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:110. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:110. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof.

[00543] The disclosure also provides genetically modified host cells genetically modified to express or overexpress one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110. In certain such embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:110. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:110. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:110, or a conservatively substituted amino acid sequence thereof.

[00544] In some embodiments, the genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid

sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110 further comprises one or more heterologous nucleic acids encoding a TKS polypeptide and one or more heterologous nucleic acids encoding an OAC polypeptide. In certain such embodiments, the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76, and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00545] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110 further comprises one or more of the following: a) one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; b) one or more heterologous nucleic acids encoding a polypeptide that generates GPP; or c) one or more heterologous nucleic acids encoding a polypeptide that generates malonyl-CoA. In certain such embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is an acyl-activating enzyme (AAE) polypeptide. In certain such embodiments, the AAE polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90. In some embodiments, the AAE polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:92 or SEQ ID NO:149. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is a fatty acyl-CoA ligase polypeptide. In certain such embodiments, the fatty acyl-CoA ligase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:145 or SEQ ID NO:147. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is a fatty acyl-CoA synthetase (FAA) polypeptide. In certain such embodiments, the FAA polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide. In certain such embodiments, the GPPS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID

NO:8, or SEQ ID NO:60. In some embodiments, the polypeptide that generates malonyl-CoA is an ACC polypeptide. In certain such embodiments, the ACC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207.

[00546] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110 further comprises one or more of the following: a) one or more heterologous nucleic acids encoding a HMGS polypeptide; b) one or more heterologous nucleic acids encoding a 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMGR) polypeptide; c) one or more heterologous nucleic acids encoding a MK polypeptide; d) one or more heterologous nucleic acids encoding a PMK polypeptide; e) one or more heterologous nucleic acids encoding a MVD polypeptide; or f) one or more heterologous nucleic acids encoding an IDI polypeptide. In certain such embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an IDI polypeptide. In certain such embodiments, the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an HMGR polypeptide. In certain such embodiments, the HMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:22. In some embodiments, the HMGR polypeptide is a truncated HMGR (tHMGR) polypeptide. In certain such embodiments, the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an HMGS polypeptide. In certain such embodiments, the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an MK polypeptide. In certain such embodiments, the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a PMK polypeptide. In certain such embodiments, the PMK polypeptide comprises an amino

acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a MVD polypeptide. In certain such embodiments, the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66.

[00547] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110 further comprises one or more heterologous nucleic acids encoding a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA. In certain such embodiments, the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide. In certain such embodiments, the acetoacetyl-CoA thiolase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25.

[00548] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110 further comprises one or more heterologous nucleic acids encoding a PDC polypeptide. In certain such embodiments, the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117.

[00549] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110 further comprises one or more heterologous nucleic acids encoding a cannabinoid synthase polypeptide. In certain such embodiments, the cannabinoid synthase polypeptide is a THCA synthase polypeptide. In certain such embodiments, the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155. In some

embodiments, the cannabinoid synthase polypeptide is a CBDA synthase polypeptide. In certain such embodiments, the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

Exemplary Genetically Modified Host Cells Expressing a Polypeptide Comprising an Amino Acid Sequence Having Sequence Identity to SEQ ID NO:100

[00550] The disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, the genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100. In certain such embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:100. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:100. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof.

[00551] The disclosure also provides genetically modified host cells genetically modified to express or overexpress one or more heterologous nucleic acids encoding GOT a polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100. In certain such embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65%, at least 70%, at least 75%, 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%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, or 100% amino acid sequence identity to SEQ ID NO:100. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino acid sequence set forth in SEQ ID NO:100. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises the amino

acid sequence set forth in SEQ ID NO:100, or a conservatively substituted amino acid sequence thereof.

[00552] In some embodiments, the genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 further comprises one or more heterologous nucleic acids encoding a TKS polypeptide and one or more heterologous nucleic acids encoding an OAC polypeptide. In certain such embodiments, the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76, and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00553] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 further comprises one or more of the following: a) one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; b) one or more heterologous nucleic acids encoding a polypeptide that generates GPP; or c) one or more heterologous nucleic acids encoding a polypeptide that generates malonyl-CoA. In certain such embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is an acyl-activating enzyme (AAE) polypeptide. In certain such embodiments, the AAE polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90. In some embodiments, the AAE polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:92 or SEQ ID NO:149. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is a fatty acyl-CoA ligase polypeptide. In certain such embodiments, the fatty acyl-CoA ligase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:145 or SEQ ID NO:147. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is a fatty acyl-CoA synthetase (FAA) polypeptide. In certain such embodiments, the FAA polypeptide comprises an amino acid sequence having at least 50%

sequence identity to SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide. In certain such embodiments, the GPPS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60. In some embodiments, the polypeptide that generates malonyl-CoA is an ACC polypeptide. In certain such embodiments, the ACC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207.

[00554] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 further comprises one or more of the following: a) one or more heterologous nucleic acids encoding a HMGS polypeptide; b) one or more heterologous nucleic acids encoding a 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMGR) polypeptide; c) one or more heterologous nucleic acids encoding a MK polypeptide; d) one or more heterologous nucleic acids encoding a PMK polypeptide; e) one or more heterologous nucleic acids encoding a MVD polypeptide; or f) one or more heterologous nucleic acids encoding an IDI polypeptide. In certain such embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an IDI polypeptide. In certain such embodiments, the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an HMGR polypeptide. In certain such embodiments, the HMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:22. In some embodiments, the HMGR polypeptide is a truncated HMGR (tHMGR) polypeptide. In certain such embodiments, the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an HMGS polypeptide. In certain such embodiments, the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an MK polypeptide.

In certain such embodiments, the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a PMK polypeptide. In certain such embodiments, the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205. In some embodiments, the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a MVD polypeptide. In certain such embodiments, the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66.

[00555] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 further comprises one or more heterologous nucleic acids encoding a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA. In certain such embodiments, the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide. In certain such embodiments, the acetoacetyl-CoA thiolase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25.

[00556] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 further comprises one or more heterologous nucleic acids encoding a PDC polypeptide. In certain such embodiments, the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117.

[00557] In some embodiments, a genetically modified host cell comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 further comprises one or more heterologous nucleic acids encoding a cannabinoid synthase

polypeptide. In certain such embodiments, the cannabinoid synthase polypeptide is a THCA synthase polypeptide. In certain such embodiments, the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155. In some embodiments, the cannabinoid synthase polypeptide is a CBDA synthase polypeptide. In certain such embodiments, the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

Exemplary Genetically Modified Host Cells Expressing GOT Polypeptides

[00558] The present disclosure provides a genetically modified host cell that produces a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide and b) one or more heterologous nucleic acids that encode a THCA synthase polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155.

[00559] The present disclosure provides a genetically modified host cell that produces a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide and b) one or more heterologous nucleic acids that encode a CBDA synthase polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some

embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

[00560] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide and b) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00561] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide, b) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide, and c) one or more

heterologous nucleic acids that encode a THCA synthase polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78; and the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155.

[00562] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide, b) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide, and c) one or more heterologous nucleic acids that encode a CBDA synthase polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the TKS polypeptide comprises an

amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78; and the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

[00563] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide and b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60.

[00564] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide, b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP, and c) one or more heterologous nucleic acids that encode a THCA synthase polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%,

99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; and the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155.

[00565] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide, b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP, and c) one or more heterologous nucleic acids that encode a CBDA synthase polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; and the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

[00566] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that

generates an acyl-CoA compound or an acyl-CoA compound derivative; and c) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00567] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; c) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; and d) one or more heterologous nucleic acids that encode a THCA synthase polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%,

99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78; and the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155.

[00568] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; c) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; and d) one or more heterologous nucleic acids that encode a CBDA synthase polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA

compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78; and the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

[00569] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; c) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; and d) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78; and the polypeptide that generates malonyl-CoA is an

ACC polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207.

[00570] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; c) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; d) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA; and e) one or more heterologous nucleic acids that encode a THCA synthase polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78; the polypeptide that generates malonyl-CoA is an ACC polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207; and the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155.

2018256863 11 Apr 2024

[00571] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; c) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; d) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA; and e) one or more heterologous nucleic acids that encode a CBDA synthase polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78; the polypeptide that generates malonyl-CoA is an ACC polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207; and the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

[00572] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT

polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; c) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; and d) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00573] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; c) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; d) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; and e) one or more heterologous nucleic acids that encode a THCA synthase polypeptide. In certain such embodiments, the GOT

polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78; the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; and the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155.

[00574] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; c) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; d) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; and e) one or more heterologous nucleic acids that encode a CBDA synthase polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an

amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78; the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; and the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

[00575] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; c) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; d) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; and e) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide

comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78; the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; and the polypeptide that generates malonyl-CoA is an ACC polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207.

[00576] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; c) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; d) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA; e) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; and f) one or more heterologous nucleic acids that encode a THCA synthase polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%,

99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78; the polypeptide that generates malonyl-CoA is an ACC polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207; the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; and the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155.

[00577] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; c) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; d) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA; e) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; and f) one or more heterologous nucleic acids that encode a CBDA synthase polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%,

99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78; the polypeptide that generates malonyl-CoA is an ACC polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207; the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; and the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

[00578] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; d) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; and e) one or more heterologous nucleic acids that encode a PDC polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%,

99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; and the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117.

[00579] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode a THCA synthase polypeptide; d) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; e) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; and f) one or more heterologous nucleic acids that encode a PDC polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65%

(e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117; and the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155.

[00580] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode a CBDA synthase polypeptide; d) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; e) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; and f) one or more heterologous nucleic acids that encode a PDC polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65%

(e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117; and the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

[00581] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; d) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; e) one or more heterologous nucleic acids that encode a PDC polypeptide; and f) one or more heterologous nucleic acids that encode a polypeptide that

generates an acyl-CoA compound or an acyl-CoA compound derivative. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; and the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117.

[00582] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT

polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode a THCA synthase polypeptide; d) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; e) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; f) one or more heterologous nucleic acids that encode a PDC polypeptide; and g) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the PDC polypeptide comprises an amino acid sequence having at least 50%

sequence identity to SEQ ID NO:117; the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; and the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155.

[00583] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode a CBDA synthase polypeptide; d) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; e) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; f) one or more heterologous nucleic acids that encode a PDC polypeptide; and g) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence

identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117; the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; and the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

[00584] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; d) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; e) one or more heterologous nucleic acids that encode a PDC polypeptide; f) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; and g) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%,

85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00585] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode a THCA synthase

polypeptide; d) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; e) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; f) one or more heterologous nucleic acids that encode a PDC polypeptide; g) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; and h) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID

NO:117; the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00586] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode a CBDA synthase polypeptide; d) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; e) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; f) one or more heterologous nucleic acids that encode a PDC polypeptide; g) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; and h) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In

some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117; the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00587] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; d) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; e) one or more heterologous nucleic acids that encode a PDC polypeptide; and f) one or more heterologous nucleic acids that encode a TKS polypeptide

and one or more heterologous nucleic acids that encode an OAC polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00588] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that

generates GPP; c) one or more heterologous nucleic acids that encode a THCA synthase polypeptide; d) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; e) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; f) a PDC polypeptide; and g) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117; the THCA synthase polypeptide comprises an amino acid sequence having at least

50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00589] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode a CBDA synthase polypeptide; d) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; e) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; f) one or more heterologous nucleic acids that encode a PDC polypeptide; and g) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity

to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117; the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00590] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; d) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; e) one or more heterologous nucleic acids that encode a PDC polypeptide; f) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; g) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA; and h) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino

acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117; the polypeptide that generates malonyl-CoA is an ACC polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00591] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode a THCA synthase polypeptide; d) one or more heterologous nucleic acids that encode one or more of the

following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; e) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; f) one or more heterologous nucleic acids that encode a PDC polypeptide; g) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; h) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA; and i) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID

NO:117; the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155; the polypeptide that generates malonyl-CoA is an ACC polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00592] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode a CBDA synthase polypeptide; d) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; e) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; f) one or more heterologous nucleic acids that encode a PDC polypeptide; g) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide; h) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA; and i) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino

acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117; the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151; the polypeptide that generates malonyl-CoA is an ACC polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00593] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode one or more of the

following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; d) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; e) one or more heterologous nucleic acids that encode a PDC polypeptide; f) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA; and g) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117; the polypeptide that generates malonyl-CoA is an ACC polypeptide and comprises an amino acid sequence having at least 50% sequence

identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00594] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode a THCA synthase polypeptide; d) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; e) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; f) one or more heterologous nucleic acids that encode a PDC polypeptide; g) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA; and h) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity

to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117; the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155; the polypeptide that generates malonyl-CoA is an ACC polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00595] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode a CBDA synthase polypeptide; d) one or more heterologous nucleic acids that encode one or more of the following polypeptides: a HMGS polypeptide, a tHMGR polypeptide, a MK polypeptide, a PMK polypeptide, a MVD polypeptide, or an IDI polypeptide; e) one or more heterologous nucleic acids that encode a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA; f) one or more heterologous nucleic acids that encode a PDC polypeptide; g) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA; and h) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids that encode an OAC polypeptide. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%,

94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110. In some embodiments, the polypeptide that generates GPP is a GPPS polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60; the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58; the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208; the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115; the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64; the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205; the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66; the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25; the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117; the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151; the polypeptide that generates malonyl-CoA is an ACC polypeptide and comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00596] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA;

2018256863 11 Apr 2024

d) one or more heterologous nucleic acids that encode one or more polypeptides that condense an acyl-CoA compound or an acyl-CoA compound derivative with malonyl-CoA to generate olivetolic acid or derivatives of olivetolic acid; e) one or more heterologous nucleic acids that encode a geranyl pyrophosphate:olivetolic acid transferase (GOT) polypeptide or an aromatic prenyltransferase polypeptide such as a NphB polypeptide; and f) one or more heterologous nucleic acids that encode a cannabinoid synthase polypeptide.

[00597] The present disclosure also provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; b) one or more heterologous nucleic acids that encode a polypeptide that generates neryl pyrophosphate (NPP); c) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA; d) one or more heterologous nucleic acids that encode one or more polypeptides that condense an acyl-CoA compound or an acyl-CoA compound derivative and malonyl-CoA to generate olivetolic acid or derivatives of olivetolic acid; e) one or more heterologous nucleic acids that encode a GOT polypeptide or a NphB polypeptide; and f) one or more heterologous nucleic acids that encode a cannabinoid synthase polypeptide. In certain such embodiments, culturing of the genetically modified host cell in a suitable medium provides for synthesis of the cannabinoid or the cannabinoid derivative in a recoverable amount.

[00598] The present disclosure provides a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; b) one or more heterologous nucleic acids that encode a TKS/OAC fusion polypeptide; and c) one or more heterologous nucleic acids that encode a GOT polypeptide or a NphB polypeptide. In certain such embodiments, culturing the genetically modified host cell in a medium comprising a carboxylic acid provides for synthesis of a cannabinoid derivative or cannabinoid in a recoverable amount. In some embodiments, the genetically modified host cell is further genetically modified with one or more heterologous nucleic acids that encode a THCAS or CBDAS polypeptide. In certain such embodiments, culturing the genetically modified host cell in a medium comprising a carboxylic acid provides for synthesis of a cannabinoid derivative or a cannabinoid in a recoverable amount.

2018256863 11 Apr 2024

Exemplary Genetically Modified Host Cells Expressing NPPS Polypeptides

[00599] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with: a) one or more heterologous nucleic acids encoding a polypeptide that condenses 2 molecules of acetyl-CoA to generate acetoacetyl-CoA; b) one or more heterologous nucleic acids encoding one or more mevalonate pathway polypeptides; and c) one or more heterologous nucleic acids encoding a NPPS polypeptide.

[00600] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with: a) one or more heterologous nucleic acids encoding a polypeptide that condenses 2 molecules of acetyl-CoA to generate acetoacetyl-CoA; b) one or more heterologous nucleic acids encoding one or more mevalonate pathway polypeptides; c) one or more heterologous nucleic acids encoding a NPPS polypeptide; and d) one or more heterologous nucleic acids encoding a TKS polypeptide and one or more heterologous nucleic acids encoding an OAC polypeptide.

[00601] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with: a) one or more heterologous nucleic acids encoding a polypeptide that condenses 2 molecules of acetyl-CoA to generate acetoacetyl-CoA; b) one or more heterologous nucleic acids encoding one or more mevalonate pathway polypeptides; c) one or more heterologous nucleic acids encoding a NPPS polypeptide; d) one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; and e) one or more heterologous nucleic acids encoding a TKS polypeptide and one or more heterologous nucleic acids encoding an OAC polypeptide.

[00602] A GOT polypeptide, using NPP and olivetolic acid as substrates, can generate cannabinerolic acid (CBNRA). In some embodiments, a GOT polypeptide acts on NPP and an olivetolic acid derivative (as described elsewhere herein) to generate a CBNRA derivative. Cannabinerolic acid or derivatives thereof can serve as a substrate for a CBDAS or THCAS polypeptide to generate CBDA or THCA, or derivatives thereof, respectively.

[00603] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with: a) one or more heterologous nucleic acids encoding a polypeptide that condenses one molecule of acetyl-CoA and one molecule of malonyl-CoA to generate acetoacetyl-CoA; b) one or more heterologous nucleic acids encoding one or more mevalonate pathway polypeptides; c) one or more heterologous nucleic acids encoding a NPPS polypeptide; and d) one or more heterologous nucleic acids encoding a GOT polypeptide. In certain such embodiments, the GOT polypeptide encoded by one or more

heterologous nucleic acids can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:110.

[00604] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with: a) one or more heterologous nucleic acids encoding a polypeptide that condenses 2 molecules of acetyl-CoA to generate acetoacetyl-CoA; b) one or more heterologous nucleic acids encoding one or more mevalonate pathway polypeptides; c) one or more heterologous nucleic acids encoding a NPPS polypeptide; and d) one or more heterologous nucleic acids encoding a GOT polypeptide or a NphB polypeptide. In certain such embodiments, the GOT polypeptide encoded by one or more heterologous nucleic acids can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:110.

[00605] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with: a) one or more heterologous nucleic acids encoding a polypeptide that condenses 2 molecules of acetyl-CoA to generate acetoacetyl-CoA; b) one

or more heterologous nucleic acids encoding one or more mevalonate pathway polypeptides; c) one or more heterologous nucleic acids encoding a NPPS polypeptide; d) one or more heterologous nucleic acids encoding a GOT polypeptide or a NphB polypeptide; and e) one or more heterologous nucleic acids encoding a CBDAS or THCAS polypeptide. In certain such embodiments, the GOT polypeptide encoded by one or more heterologous nucleic acids can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:110.

[00606] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with: a) one or more heterologous nucleic acids encoding a polypeptide that condenses one molecule of acetyl-CoA and one molecule of malonyl-CoA to generate acetoacetyl-CoA; b) one or more heterologous nucleic acids encoding one or more mevalonate pathway polypeptides; c) one or more heterologous nucleic acids encoding a NPPS polypeptide; d) one or more heterologous nucleic acids encoding a TKS polypeptide and one or more heterologous nucleic acids encoding an OAC polypeptide; and e) one or more heterologous nucleic acids encoding a GOT polypeptide. In certain such embodiments, the GOT polypeptide encoded by one or more heterologous nucleic acids can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids

comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:110.

[00607] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with: a) one or more heterologous nucleic acids encoding a polypeptide that condenses 2 molecules of acetyl-CoA to generate acetoacetyl-CoA; b) one or more heterologous nucleic acids encoding one or more mevalonate pathway polypeptides; c) one or more heterologous nucleic acids encoding a NPPS polypeptide; d) one or more heterologous nucleic acids encoding a TKS polypeptide and one or more heterologous nucleic acids encoding an OAC polypeptide; and e) one or more heterologous nucleic acids encoding a GOT polypeptide or a NphB polypeptide. In certain such embodiments, the GOT polypeptide encoded by one or more heterologous nucleic acids can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:110.

[00608] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with: a) one or more heterologous nucleic acids encoding a polypeptide that condenses 2 molecules of acetyl-CoA to generate acetoacetyl-CoA; b) one or more heterologous nucleic acids encoding one or more mevalonate pathway polypeptides; c) one or more heterologous nucleic acids encoding a NPPS polypeptide; d) one or more heterologous nucleic acids encoding a TKS polypeptide and one or more heterologous nucleic acids encoding an OAC polypeptide; e) one or more heterologous nucleic acids encoding a GOT polypeptide or a NphB polypeptide; and f) one or more heterologous nucleic acids encoding a CBDAS or THCAS polypeptide. In certain such embodiments, the

GOT polypeptide encoded by one or more heterologous nucleic acids can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:110.

[00609] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with: a) one or more heterologous nucleic acids encoding a polypeptide that condenses one molecule of acetyl-CoA and one molecule of malonyl-CoA to generate acetoacetyl-CoA; b) one or more heterologous nucleic acids encoding one or more mevalonate pathway polypeptides; c) one or more heterologous nucleic acids encoding a NPPS polypeptide; d) one or more heterologous nucleic acids encoding a TKS polypeptide and one or more heterologous nucleic acids encoding an OAC polypeptide; e) one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; and f) one or more heterologous nucleic acids encoding a GOT polypeptide. In certain such embodiments, the GOT polypeptide encoded by one or more heterologous nucleic acids can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%,

89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:110.

[00610] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified: a) with one or more heterologous nucleic acids encoding a polypeptide that condenses 2 molecules of acetyl-CoA to generate acetoacetyl-CoA; b) with one or more heterologous nucleic acids encoding one or more mevalonate pathway polypeptides; c) with one or more heterologous nucleic acids encoding a NPPS polypeptide; d) with one or more heterologous nucleic acids encoding a TKS polypeptide and with one or more heterologous nucleic acids encoding an OAC polypeptide; e) with one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; and f) with one or more heterologous nucleic acids encoding a GOT polypeptide or a NphB polypeptide. In certain such embodiments, the GOT polypeptide encoded by one or more heterologous nucleic acids can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:110.

[00611] In some embodiments, a genetically modified host cell of the present disclosure is genetically modified with: a) one or more heterologous nucleic acids encoding a polypeptide that condenses 2 molecules of acetyl-CoA to generate acetoacetyl-CoA; b) one or more heterologous nucleic acids encoding one or more mevalonate pathway polypeptides; c) one or more heterologous nucleic acids encoding a NPPS polypeptide; d) one or more heterologous nucleic acids encoding a TKS polypeptide and one or more heterologous nucleic acids encoding an OAC polypeptide; e) one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; f) one or more heterologous nucleic acids encoding a GOT polypeptide or a

NphB polypeptide; and g) one or more heterologous nucleic acids encoding a CBDAS or THCAS polypeptide. In certain such embodiments, the GOT polypeptide encoded by one or more heterologous nucleic acids can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In certain such embodiments, the GOT polypeptide encoded by the one or more heterologous nucleic acids comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) amino acid sequence identity to SEQ ID NO:110.

Exemplary Genetically Modified Host Cells for Making Olivetolic Acid or Olivetolic Acid Derivatives

[00612] The present disclosure provides a genetically modified host cell for producing olivetolic acid or an olivetolic acid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative and b) one or more heterologous nucleic acids that encode a TKS polypeptide and one or more heterologous nucleic acids encoding an OAC polypeptide. In certain such embodiments, the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76; and the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78. In certain such embodiments, culturing the genetically modified host cell disclosed herein in a medium comprising a carboxylic acid provides for synthesis of an olivetolic acid or olivetolic acid derivative in a recoverable amount.

2018256863 11 Apr 2024

[00613] The present disclosure provides a genetically modified host cell for producing olivetolic acid or an olivetolic acid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; and b) one or more heterologous nucleic acids that encode a TKS/OAC fusion polypeptide. In certain such embodiments, culturing the genetically modified host cell in a medium comprising a carboxylic acid provides for synthesis of the olivetolic acid or olivetolic acid derivative in a recoverable amount.

Suitable Host Cells

[00614] Parent host cells that are suitable for use in generating a genetically modified host cell of the present disclosure may include prokaryotic cells and eukaryotic cells. In some embodiments, the eukaryotic cells are yeast cells. In some embodiments, the eukaryotic cells are plant cells.

[00615] Host cells (including parent host cells and genetically modified host cells) are in some embodiments unicellular organisms, or are grown in culture as single cells. In some embodiments, the host cell is a eukaryotic cell. Suitable eukaryotic host cells may include, but are not limited to, yeast cells, insect cells, plant cells, fungal cells, and algal cells. Suitable eukaryotic host cells may include, but are not limited to, *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia methanolica*, *Pichia* sp., *Saccharomyces cerevisiae*, *Saccharomyces* sp., *Hansenula polymorpha* (now known as *Pichia angusta*), *Kluyveromyces* sp., *Kluyveromyces lactis, Kluyveromyces marxianus*, *Schizosaccharomyces pombe*, *Dekkera bruxellensis*, *Arxula adeninivorans*, *Candida albicans*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Trichoderma reesei*, *Chrysosporium lucknowense*, *Fusarium* sp., *Fusarium gramineum*, *Fusarium venenatum*, *Neurospora crassa*, *Chlamydomonas reinhardtii*, and the like. In some embodiments, the host cell is a protease-deficient strain of *Saccharomyces cerevisiae*. In some embodiments, the host cell is a eukaryotic cell other than a plant cell. In some embodiments, the eukaryotic cell is a plant cell. In some embodiments, the eukaryotic cell is a plant cell, where the plant cell is one that does not normally produce a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative. In some embodiments, the host cell is *Saccharomyces*

2018256863 11 Apr 2024

cerevisiae. In some embodiments, the genetically modified host cell disclosed herein is cultured *in vitro*.

[00616] In some embodiments, the host cell is a prokaryotic cell. Suitable prokaryotic cells may include, but are not limited to, any of a variety of laboratory strains of *Escherichia coli*, *Lactobacillus* sp., *Salmonella* sp., *Shigella* sp., and the like. See, e.g., Carrier et al. (1992) *J. Immunol.* 148:1176-1181; U.S. Patent No. 6,447,784; and Sizemore et al. (1995) *Science* 270:299-302. Examples of Salmonella strains which can be employed may include, but are not limited to, *Salmonella typhi* and *S. typhimurium*. Suitable Shigella strains may include, but are not limited to, *Shigella flexneri*, *Shigella sonnei*, and *Shigella disenteriae*. Typically, the laboratory strain is one that is non-pathogenic. Non-limiting examples of other suitable bacteria may include, but are not limited to, *Bacillus subtilis*, *Pseudomonas putida*, *Pseudomonas aeruginosa*, *Pseudomonas mevalonii*, *Rhodobacter sphaeroides*, *Rhodobacter capsulatus*, *Rhodospirillum rubrum*, *Rhodococcus* sp., and the like. In some embodiments, the host cell is *Escherichia coli*.

Genetic Modification of Host Cells

[00617] The present disclosure provides for a method of making a genetically modified host cell for producing a cannabinoid, a cannabinoid derivative, a cannabinoid precursor derivative, or a cannabinoid precursor, comprising introducing into the genetically modified host cell one or more heterologous nucleic acids disclosed herein. In some embodiments, the genetically modified host cell produces a cannabinoid or a cannabinoid derivative. The disclosure also provides a method for making a genetically modified host cell genetically modified to express or overexpress one or more heterologous nucleic acids disclosed herein that encode one or more polypeptides disclosed herein, comprising introducing into the genetically modified host cell one or more heterologous nucleic acids disclosed herein.

[00618] In some embodiments, the disclosure provides for a method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, into the genetically modified host cell. In some embodiments, the present disclosure provides for a method of making a genetically modified host cell genetically modified to express or overexpress one or more

heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, comprising introducing one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, into the genetically modified host cell.

[00619] In some embodiments, the disclosure provides for a method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110 into the genetically modified host cell. In some embodiments, the present disclosure provides for a method of making a genetically modified host cell genetically modified to express or overexpress one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110 comprising introducing one or more heterologous nucleic acids encoding the GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110 into the genetically modified host cell.

[00620] The disclosure provides for a method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 into the genetically modified host cell. In some embodiments, the present disclosure provides for a method of making a genetically modified host cell genetically modified to express or

2018256863 11 Apr 2024

overexpress one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 comprising introducing one or more heterologous nucleic acids encoding the GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 into the genetically modified host cell.

[00621] To genetically modify a parent host cell to produce a genetically modified host cell of the present disclosure, one or more heterologous nucleic acids disclosed herein is introduced stably or transiently into a host cell, using established techniques. Such techniques may include, but are not limited to, electroporation, calcium phosphate precipitation, DEAE-dextran mediated transfection, liposome-mediated transfection, and the like. For stable transformation, a heterologous nucleic acid will generally further include a selectable marker, e.g., any of several well-known selectable markers such as neomycin resistance, ampicillin resistance, tetracycline resistance, chloramphenicol resistance, kanamycin resistance, and the like. In some embodiments, a parent host cell is genetically modified to produce a genetically modified host cell of the present disclosure using a CRISPR/Cas9 system to genetically modify a parent host cell with one or more heterologous nucleic acids disclosed herein.

[00622] One or more nucleic acids disclosed herein can be present in an expression vector or construct. Suitable expression vectors may include, but are not limited to, baculovirus vectors, bacteriophage vectors, plasmids, phagemids, cosmids, fosmids, bacterial artificial chromosomes, viral vectors (e.g. viral vectors based on vaccinia virus, poliovirus, adenovirus, adeno-associated virus, SV40, herpes simplex virus, and the like), P1-based artificial chromosomes, yeast plasmids, yeast artificial chromosomes, and any other vectors specific for specific hosts of interest (such as *E. coli* and yeast). Thus, for example, one or more nucleic acids encoding a mevalonate pathway gene product(s) is included in any one of a variety of expression vectors for expressing the mevalonate pathway gene product(s). Such vectors may include chromosomal, non-chromosomal, and synthetic DNA sequences.

[00623] Numerous additional suitable expression vectors are known to those of skill in the art, and many are commercially available. The following vectors are provided by way

2018256863 11 Apr 2024

of example; for bacterial host cells: pQE vectors (Qiagen), pBluescript plasmids, pNH vectors, lambda-ZAP vectors (Stratagene); pTrc99a, pKK223-3, pDR540, and pRIT2T (Pharmacia); for eukaryotic host cells: pXT1, pSG5 (Stratagene), pSVK3, pBPV, pMSG, and pSVLSV40 (Pharmacia). However, any other plasmid or other vector may be used so long as it is compatible with the host cell.

[00624] In some embodiments, one or more of the nucleic acids disclosed herein are present in a single expression vector. In some embodiments, two or more of the nucleic acids disclosed herein are present in a single expression vector. In some embodiments, three or more of the nucleic acids disclosed herein are present in a single expression vector. In some embodiments, four or more of the nucleic acids disclosed herein are present in a single expression vector. In some embodiments, five or more of the nucleic acids disclosed herein are present in a single expression vector. In some embodiments, six or more of the nucleic acids disclosed herein are present in a single expression vector. In some embodiments, seven or more of the nucleic acids disclosed herein are present in a single expression vector.

[00625] In some embodiments, two or more nucleic acids disclosed herein are in separate expression vectors. In some embodiments, three or more nucleic acids disclosed herein are in separate expression vectors. In some embodiments, four or more nucleic acids disclosed herein are in separate expression vectors. In some embodiments, five or more nucleic acids disclosed herein are in separate expression vectors. In some embodiments, six or more nucleic acids disclosed herein are in separate expression vectors. In some embodiments, seven or more nucleic acids disclosed herein are in separate expression vectors. In some embodiments, eight or more nucleic acids disclosed herein are in separate expression vectors. In some embodiments, nine or more nucleic acids disclosed herein are in separate expression vectors. In some embodiments, ten or more nucleic acids disclosed herein are in separate expression vectors.

[00626] In some embodiments, one or more of the nucleic acids disclosed herein are present in a single expression construct. In some embodiments, two or more of the nucleic acids disclosed herein are present in a single expression construct. In some embodiments, three or more of the nucleic acids disclosed herein are present in a single expression construct. In some embodiments, four or more of the nucleic acids disclosed herein are present in a single expression construct. In some embodiments, five or more of the nucleic acids disclosed herein are present in a single expression construct. In some embodiments, six or more of the nucleic acids disclosed herein are present in a single expression construct.

In some embodiments, seven or more of the nucleic acids disclosed herein are present in a single expression construct.

[00627] In some embodiments, two or more nucleic acids disclosed herein are in separate expression constructs. In some embodiments, three or more nucleic acids disclosed herein are in separate expression constructs. In some embodiments, four or more nucleic acids disclosed herein are in separate expression constructs. In some embodiments, five or more nucleic acids disclosed herein are in separate expression constructs. In some embodiments, six or more nucleic acids disclosed herein are in separate expression constructs. In some embodiments, seven or more nucleic acids disclosed herein are in separate expression constructs. In some embodiments, eight or more nucleic acids disclosed herein are in separate expression constructs. In some embodiments, nine or more nucleic acids disclosed herein are in separate expression constructs. In some embodiments, ten or more nucleic acids disclosed herein are in separate expression constructs.

[00628] The disclosure provides a method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing a vector comprising one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, into the genetically modified host cell. The disclosure provides a method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing a vector comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110 into the genetically modified host cell. The disclosure also provides a method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing a vector comprising one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 into the genetically modified host cell.

[00629] The disclosure also provides a method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing a

vector comprising a CsPT4 heterologous nucleic acid comprising a nucleotide sequence having at least 80% sequence identity to SEQ ID NO:111 into the genetically modified host cell. The disclosure also provides a method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing a vector comprising a CsPT4 heterologous nucleic acid comprising a nucleotide sequence having at least 80% sequence identity to SEQ ID NO:225 into the genetically modified host cell. The disclosure also provides a method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing a vector comprising a CsPT4t heterologous nucleic acid comprising a nucleotide sequence having at least 80% sequence identity to SEQ ID NO:221 into the genetically modified host cell. The disclosure also provides a method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing a vector comprising a CsPT4t heterologous nucleic acid comprising a nucleotide sequence having at least 80% sequence identity to SEQ ID NO:224 into the genetically modified host cell.

[00630] The disclosure provides a method of making a genetically modified host cell genetically modified to express or overexpress one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, comprising introducing a vector comprising one or more heterologous nucleic acids encoding a GOT polypeptide, wherein said GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, into the genetically modified host cell. The disclosure provides a method of making a genetically modified host cell genetically modified to express or overexpress one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110, comprising introducing a vector comprising one or more heterologous nucleic acids encoding the GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:110 into the genetically modified host cell. The disclosure provides a method of making a genetically

modified host cell genetically modified to express or overexpress one or more heterologous nucleic acids encoding a GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100, comprising introducing a vector comprising one or more heterologous nucleic acids encoding the GOT polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 into the genetically modified host cell.

[00631] The disclosure also provides a method of making a genetically modified host cell genetically modified to express or overexpress a CsPT4 heterologous nucleic acid comprising a nucleotide sequence having at least 80% sequence identity to SEQ ID NO:111, comprising introducing a vector comprising a CsPT4 heterologous nucleic acid comprising a nucleotide sequence having at least 80% sequence identity to SEQ ID NO:111 into the genetically modified host cell. The disclosure also provides a method of making a genetically modified host cell genetically modified to express or overexpress a CsPT4 heterologous nucleic acid comprising a nucleotide sequence having at least 80% sequence identity to SEQ ID NO:225, comprising introducing a vector comprising a CsPT4 heterologous nucleic acid comprising a nucleotide sequence having at least 80% sequence identity to SEQ ID NO:225 into the genetically modified host cell. The disclosure also provides a method of making a genetically modified host cell genetically modified to express or overexpress a CsPT4t heterologous nucleic acid comprising a nucleotide sequence having at least 80% sequence identity to SEQ ID NO:221, comprising introducing a vector comprising a CsPT4t heterologous nucleic acid comprising a nucleotide sequence having at least 80% sequence identity to SEQ ID NO:221 into the genetically modified host cell. The disclosure also provides a method of making a genetically modified host cell genetically modified to express or overexpress a CsPT4t heterologous nucleic acid comprising a nucleotide sequence having at least 80% sequence identity to SEQ ID NO:224, comprising introducing a vector comprising a CsPT4t heterologous nucleic acid comprising a nucleotide sequence having at least 80% sequence identity to SEQ ID NO:224 into the genetically modified host cell.

[00632] In some embodiments, one or more of the heterologous nucleic acids disclosed herein is present in a high copy number plasmid, e.g., a plasmid that exists in about

2018256863 11 Apr 2024

10-50 copies per cell, or more than 50 copies per cell. In some embodiments, one or more of the heterologous nucleic acids disclosed herein is present in a low copy number plasmid. In some embodiments, one or more of the heterologous nucleic acids disclosed herein is present in a medium copy number plasmid.

[00633] Depending on the host/vector or host/construct system utilized, any of a number of suitable transcription and translation control elements, including constitutive and inducible promoters, transcription enhancer elements, transcription terminators, etc. may be used in the expression vector or construct (see e.g., Bitter et al. (1987) *Methods in Enzymology*, 153:516-544).

[00634] In some embodiments, heterologous nucleic acids disclosed herein are operably linked to a promoter. In some embodiments, the promoter is a constitutive promoter. In some embodiments, the promoter is an inducible promoter. In some embodiments, the promoter is functional in a prokaryotic cell. In some embodiments, the promoter is functional in a eukaryotic cell. In some embodiments, the promoter can be a strong driver of expression. In some embodiments, the promoter can be a weak driver of expression. In some embodiments, the promoter can be a medium driver of expression.

[00635] Suitable promoters for use in prokaryotic host cells may include, but are not limited to, a bacteriophage T7 RNA polymerase promoter; a trp promoter; a lac operon promoter; a hybrid promoter, e.g., a lac/tac hybrid promoter, a tac/trc hybrid promoter, a trp/lac promoter, a T7/lac promoter; a trc promoter; a tac promoter, and the like; an araBAD promoter; *in vivo* regulated promoters, such as an *ssaG* promoter or a related promoter (*see*, e.g., U.S. Patent Publication No. 20040131637), a *pagC* promoter (Pulkkinen and Miller, *J. Bacteriol*., 1991: 173(1): 86-93; Alpuche-Aranda et al., PNAS, 1992; 89(21): 10079-83), a *nirB* promoter (Harborne et al. (1992) *Mol. Micro*. 6:2805-2813), and the like (*see*, e.g., Dunstan et al. (1999) *Infect. Immun.* 67:5133-5141; McKelvie et al. (2004) *Vaccine* 22:3243-3255; and Chatfield et al. (1992) *Biotechnol.* 10:888-892); a sigma70 promoter, e.g., a consensus sigma70 promoter (see, e.g., GenBank Accession Nos. AX798980, AX798961, and AX798183); a stationary phase promoter, e.g., a *dps* promoter, a *spv* promoter, and the like; a promoter derived from the pathogenicity island SPI-2 (*see*, e.g., WO96/17951); an *actA* promoter (*see*, e.g., Shetron-Rama et al. (2002) *Infect. Immun.* 70:1087-1096); an *rpsM* promoter (*see*, e.g., Valdivia and Falkow (1996). *Mol. Microbiol.* 22:367–378); a tet promoter (*see*, e.g., Hillen, W. and Wissmann, A. (1989) In Saenger, W. and Heinemann, U. (eds), *Topics in Molecular and Structural Biology*, *Protein–Nucleic Acid*

2018256863 11 Apr 2024

Interaction. Macmillan, London, UK, Vol. 10, pp. 143–162); an SP6 promoter (*see*, e.g., Melton et al. (1984) *Nucl. Acids Res.* 12:7035); and the like.

[00636] Suitable constitutive promoters for use in prokaryotic cells are known in the art and may also include, but are not limited to, a sigma70 promoter, e.g., a consensus sigma70 promoter.

[00637] Non-limiting examples of suitable eukaryotic promoters may include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector, construct, and promoter is well within the level of ordinary skill in the art. The expression vector or construct may also contain a ribosome binding site for translation initiation and a transcription terminator. The expression vector or construct may also include appropriate sequences for amplifying expression.

[00638] In yeast, a number of vectors or constructs containing constitutive or inducible promoters may be used. For a review see, Current Protocols in Molecular Biology, Vol. 2, 1988, Ed. Ausubel, et al., Greene Publish. Assoc. & Wiley Interscience, Ch. 13; Grant, et al., 1987, Expression and Secretion Vectors for Yeast, in Methods in Enzymology, Eds. Wu & Grossman, 31987, Acad. Press, N.Y., Vol. 153, pp.516-544; Glover, 1986, DNA Cloning, Vol. II, IRL Press, Wash., D.C., Ch. 3; and Bitter, 1987, Heterologous Gene Expression in Yeast, Methods in Enzymology, Eds. Berger & Kimmel, Acad. Press, N.Y., Vol. 152, pp. 673-684; and The Molecular Biology of the Yeast Saccharomyces, 1982, Eds. Strathern et al., Cold Spring Harbor Press, Vols. I and II. A constitutive yeast promoter such as ADH or LEU2 or an inducible promoter such as GAL may be used (Cloning in Yeast, Ch. 3, R. Rothstein In: DNA Cloning Vol. 11, A Practical Approach, Ed. DM Glover, 1986, IRL Press, Wash., D.C.). Alternatively, vectors may be used which promote integration of foreign DNA sequences into the yeast chromosome.

[00639] Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of *E. coli*, the *S. cerevisiae* TRP1 gene, etc.; and a promoter derived from a highly-expressed gene to direct transcription of the coding sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), α-factor, acid phosphatase, or heat shock proteins, among others.

[00640] Inducible promoters are well known in the art. Suitable inducible promoters may include, but are not limited to, the pL of bacteriophage λ; Plac; Ptrp; Ptac (Ptrp-lac hybrid promoter); an isopropyl-beta-D-thiogalactopyranoside (IPTG)-inducible promoter,

2018256863 11 Apr 2024

e.g., a *lacZ* promoter; a tetracycline-inducible promoter; an arabinose inducible promoter, e.g., P_{BAD} (see, e.g., Guzman et al. (1995) *J. Bacteriol.* 177:4121-4130); a xylose-inducible promoter, e.g., Pxyl (see, e.g., Kim et al. (1996) *Gene* 181:71-76); a GAL1 promoter; a tryptophan promoter; a lac promoter; an alcohol-inducible promoter, e.g., a methanol-inducible promoter, an ethanol-inducible promoter; a raffinose-inducible promoter; a heat-inducible promoter, e.g., heat inducible lambda P_L promoter, a promoter controlled by a heat-sensitive repressor (e.g., CI857-repressed lambda-based expression vectors; *see*, e.g., Hoffmann et al. (1999) *FEMS Microbiol Lett.* 177(2):327-34); and the like.

[00641] In addition, the expression vectors or constructs will in many embodiments contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in prokaryotic host cells such as *E. coli*.

[00642] In some embodiments, one or more heterologous nucleic acids disclosed herein is integrated into the genome of the genetically modified host cell disclosed herein. In some embodiments, one or more heterologous nucleic acids disclosed herein remains episomal (i.e., is not integrated into the genome of the genetically modified host cell). In some embodiments, at least one of the one or more heterologous nucleic acids disclosed herein is maintained extrachromosomally.

[00643] In some embodiments, a subject heterologous nucleic acid or a subject recombinant expression vector or construct comprises a promoter or other regulatory element(s) for expression in a plant cell. Non-limiting examples of suitable constitutive promoters that are functional in a plant cell is the cauliflower mosaic virus 35S promoter, a tandem 35S promoter (Kay et al., *Science* 236:1299 (1987)), a cauliflower mosaic virus 19S promoter, a nopaline synthase gene promoter (Singer et al., *Plant Mol. Biol.* 14:433 (1990); An, *Plant Physiol.* 81:86 (1986), an octopine synthase gene promoter, and a ubiquitin promoter. Suitable inducible promoters that are functional in a plant cell may include, but are not limited to, a phenylalanine ammonia-lyase gene promoter, a chalcone synthase gene promoter, a pathogenesis-related protein gene promoter, a copper-inducible regulatory element (Mett et al., *Proc. Natl. Acad. Sci. USA* 90:4567-4571 (1993); Furst et al., Cell 55:705-717 (1988)); tetracycline and chlor-tetracycline-inducible regulatory elements (Gatz et al., *Plant J.* 2:397-404 (1992); Röder et al., *Mol. Gen. Genet.* 243:32-38 (1994); Gatz, *Meth. Cell Biol.* 50:411-424 (1995)); ecdysone inducible regulatory elements (Christopherson et al., *Proc. Natl. Acad. Sci. USA* 89:6314-6318 (1992); Kreutzweiser et al.,

2018256863 11 Apr 2024

Ecotoxicol. Environ. Safety 28:14-24 (1994)); heat shock inducible regulatory elements (Takahashi et al., *Plant Physiol.* 99:383-390 (1992); Yabe et al., *Plant Cell Physiol.* 35:1207-1219 (1994); Ueda et al., *Mol. Gen. Genet.* 250:533-539 (1996)); and lac operon elements, which are used in combination with a constitutively expressed lac repressor to confer, for example, IPTG-inducible expression (Wilde et al., *EMBO J.* 11:1251-1259 (1992); a nitrate-inducible promoter derived from the spinach nitrite reductase gene (Back et al., *Plant Mol. Biol.* 17:9 (1991)); a light-inducible promoter, such as that associated with the small subunit of RuBP carboxylase or the LHCP gene families (Feinbaum et al., *Mol. Gen. Genet.* 226:449 (1991); Lam and Chua, *Science* 248:471 (1990)); a light-responsive regulatory element as described in U.S. Patent Publication No. 20040038400; a salicylic acid inducible regulatory elements (Uknes et al., *Plant Cell* 5:159-169 (1993); Bi et al., *Plant J.* 8:235-245 (1995)); plant hormone-inducible regulatory elements (Yamaguchi-Shinozaki et al., *Plant Mol. Biol.* 15:905 (1990); Kares et al., *Plant Mol. Biol.* 15:225 (1990)); and human hormone-inducible regulatory elements such as the human glucocorticoid response element (Schena et al., *Proc. Natl. Acad. Sci. USA* 88:10421 (1991).

[00644] Plant tissue-selective regulatory elements also can be included in a subject heterologous nucleic acid or a subject vector or construct. Suitable tissue-selective regulatory elements, which can be used to ectopically express a heterologous nucleic acid in a single tissue or in a limited number of tissues, may include, but are not limited to, a xylem-selective regulatory element, a tracheid-selective regulatory element, a fiber-selective regulatory element, a trichome-selective regulatory element (see, e.g., Wang et al. (2002) *J. Exp. Botany* 53:1891-1897), a glandular trichome-selective regulatory element, and the like.

[00645] Vectors that are suitable for use in plant cells are known in the art, and any such vector can be used to introduce a subject heterologous nucleic acid into a plant host cell. Suitable vectors may include, e.g., a Ti plasmid of *Agrobacterium tumefaciens* or an Ri_1 plasmid of *A. rhizogenes*. The Ti or Ri_1 plasmid is transmitted to plant cells on infection by *Agrobacterium* and is stably integrated into the plant genome. J. Schell, *Science,* 237:1176-83 (1987). Also suitable for use is a plant artificial chromosome, as described in, e.g., U.S. Patent No. 6,900,012.

[00646] As will be appreciated by the skilled artisan, slight changes in nucleotide sequence do not necessarily alter the amino acid sequence of the encoded polypeptide. It will be appreciated by persons skilled in the art that changes in the identities of nucleotides in a specific gene sequence that change the amino acid sequence of the encoded polypeptide may result in reduced or enhanced effectiveness of the genes and that, in some applications (e.g.,

2018256863 11 Apr 2024

anti-sense, co-suppression, or RNAi), partial sequences often work as effectively as full length versions. The ways in which the nucleotide sequence can be varied or shortened are well known to persons skilled in the art, as are ways of testing the effectiveness of the altered genes. In certain embodiments, effectiveness may easily be tested by, for example, conventional gas chromatography. All such variations of the genes are therefore included as part of the present disclosure.

Codon Usage

[00647] As is well known to those of skill in the art, it is possible to improve the expression of a heterologous nucleic acid in a host organism by replacing the nucleotide sequences coding for a particular amino acid (i.e., a codon) with another codon which is better expressed in the host organism (i.e., codon optimization). One reason that this effect arises due to the fact that different organisms show preferences for different codons. In some embodiments, a heterologous nucleic acid disclosed herein is modified or optimized such that the nucleotide sequence reflects the codon preference for the particular host cell. For example, the nucleotide sequence will in some embodiments be modified or optimized for yeast codon preference. See, e.g., Bennetzen and Hall (1982) *J. Biol. Chem.* 257(6): 3026-3031. As another non-limiting example, the nucleotide sequence will in some embodiments be modified or optimized for *E. coli* codon preference. See, e.g., Gouy and Gautier (1982) *Nucleic Acids Res.* 10(22):7055-7074; Eyre-Walker (1996) *Mol. Biol. Evol.* 13(6):864-872. See also Nakamura et al. (2000) *Nucleic Acids Res.* 28(1):292.

[00648] Statistical methods have been generated to analyze codon usage bias in various organisms and many computer algorithms have been developed to implement these statistical analyses in the design of codon optimized gene sequences (Lithwick G, Margalit H (2003) Hierarchy of sequence-dependent features associated with prokaryotic translation. Genome Research 13: 2665-73). Other modifications in codon usage to increase protein expression that are not dependent on codon bias have also been described (Welch et al. (2009). Design parameters to control synthetic gene expression in Escherichia coli. PLoS ONE 4: e7002).

[00649] In some embodiments, the codon usage of a coding sequence is modified such that the level of translation of the encoded mRNA is decreased. Reducing the level of translation of an mRNA by modifying codon usage is achieved by modifying the sequence to include codons that are rare or not commonly used by the host cell. Codon usage tables for many organisms are available that summarize the percentage of time a specific organism

2018256863 11 Apr 2024

uses a specific codon to encode for an amino acid. Certain codons are used more often than other, "rare" codons. The use of "rare" codons in a sequence generally decreases its rate of translation. Thus, e.g., the coding sequence is modified by introducing one or more rare codons, which affect the rate of translation, but not the amino acid sequence of the polypeptide translated. For example, there are 6 codons that encode for arginine: CGT, CGC, CGA, CGG, AGA, and AGG. In *E. coli* the codons CGT and CGC are used far more often (encoding approximately 40% of the arginines in *E. coli* each) than the codon AGG (encoding approximately 2% of the arginines in *E. coli*). Modifying a CGT codon within the sequence of a gene to an AGG codon would not change the sequence of the polypeptide, but would likely decrease the gene's rate of translation.

[00650] Further, it will be appreciated that this disclosure embraces the degeneracy of codon usage as would be understood by one of ordinary skill in the art and illustrated in the following table.

Codon Degeneracies

Amino Acid	Codons
Ala/A	GCT, GCC, GCA, GCG
Arg/R	CGT, CGC, CGA, CGG, AGA, AGG
Asn/N	AAT, AAC
Asp/D	GAT, GAC
Cys/C	TGT, TGC
Gln/Q	CAA, CAG
Glu/E	GAA, GAG
Gly/G	GGT, GGC, GGA, GGG
His/H	CAT, CAC
Ile/I	ATT, ATC, ATA
Leu/L	TTA, TTG, CTT, CTC, CTA, CTG
Lys/K	AAA, AAG
Met/M	ATG
Phe/F	TTT, TTC
Pro/P	CCT, CCC, CCA, CCG
Ser/S	TCT, TCC, TCA, TCG, AGT, AGC
Thr/T	ACT, ACC, ACA, ACG
Trp/W	TGG
Tyr/Y	TAT, TAC
Val/V	GTT, GTC, GTA, GTG

2018256863 11 Apr 2024

START	ATG
STOP	TAG, TGA, TAA

Genetically Modified Plants

[00651] The present disclosure provides genetically modified plants, where the genetically modified plants are genetically modified with one or more heterologous nucleic acids disclosed herein to generate a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative. In some embodiments, the genetically modified plant is a plant of a genus other than *Cannabis*.

[00652] The present disclosure provides a genetically modified plant, wherein the genetically modified plant is genetically modified with: a) one or more heterologous nucleic acids that encode a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP; c) one or more heterologous nucleic acids that encode a polypeptide that generates malonyl-CoA; d) one or more heterologous nucleic acids that encode one or more polypeptides that condense an acyl-CoA compound or an acyl-CoA compound derivative and malonyl-CoA to generate olivetolic acid or derivatives of olivetolic acid; e) one or more heterologous nucleic acids that encode a polypeptide that condenses GPP and olivetolic acid to generate cannabigerolic acid or derivatives thereof; or f) one or more heterologous nucleic acids that encode a cannabinoid synthase polypeptide, wherein the polypeptides are produced in the plant, and wherein production of the polypeptides results in production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative by the genetically modified plant. In some embodiments, the plant is a monocot. In some embodiments, the plant is a dicot. In some embodiments, one or more of the polypeptide-encoding heterologous nucleic acids is operably linked to a constitutive promoter. In some embodiments, one or more of the polypeptide-encoding heterologous nucleic acids is operably linked to an inducible promoter. In some embodiments, one or more of the polypeptide-encoding heterologous nucleic acids is operably linked to a tissue-specific promoter. In some embodiments, the tissue-specific promoter is a trichome-specific promoter.

[00653] The present disclosure provides a method of producing a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative, the method comprising maintaining a transgenic plant under conditions that favor production of the encoded one or more polypeptides, wherein production of the encoded one or more

2018256863 11 Apr 2024

polypeptides results in production of the cannabinoid, the cannabinoid derivative, the cannabinoid precursor, or the cannabinoid precursor derivative.

[00654] In some embodiments, the genome of the transgenic plant comprises a subject heterologous nucleic acid. In some embodiments, the transgenic plant is homozygous for the genetic modification. In some embodiments, the transgenic plant is heterozygous for the genetic modification.

[00655] In some embodiments, a subject transgenic plant produces one or more transgene-encoded polypeptides disclosed herein that result in the production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in an amount that is at least about 50%, at least about 2-fold, at least about 5-fold, at least about 10-fold, at least about 25-fold, at least about 50-fold, or at least about 100-fold, or higher, than the amount of the cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative produced by a control plant, e.g., a non-transgenic plant (a plant that does not include the transgene encoding the one or more polypeptides) of the same species.

[00656] In some embodiments, a subject transgenic plant is a transgenic version of a control, non-transgenic plant that normally produces a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative that is generated by, or is a downstream product of, transgene-encoded one or more polypeptides that produce a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative; where the transgenic plant produces the cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in an amount that is at least about 50%, at least about 2-fold, at least about 5-fold, at least about 10-fold, at least about 25-fold, at least about 50-fold, or at least about 100-fold, or higher, than the amount of the cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative produced by the control plant, e.g., a non-transgenic plant (a plant that does not include the transgene encoding the one or more polypeptides) of the same species.

[00657] Methods of introducing exogenous nucleic acids into plant cells are well known in the art. Such plant cells are considered "transformed." Suitable methods may include viral infection (such as double stranded DNA viruses), transfection, conjugation, protoplast fusion, electroporation, particle gun technology, calcium phosphate precipitation, direct microinjection, silicon carbide whiskers technology, *Agrobacterium*-mediated transformation, CRISPR/Cas9-mediated genome editing, and the like. The choice of method

2018256863 11 Apr 2024

is generally dependent on the type of cell being transformed and the circumstances under which the transformation is taking place (e.g., *in vitro*, *ex vivo*, or *in vivo*).

[00658] A CRISPR/Cas9 system can be used to generate a transgenic (genetically modified) plant of the present disclosure. CRISPR/Cas9 systems and methods are known in the art. See, e.g., Bortesi and Fischer (2015) *Biotechnol. Advances* 33:41; and Fan et al. (2015) *Sci. Reports* 5:12217.

[00659] Transformation methods based upon the soil bacterium *Agrobacterium tumefaciens* can be used for introducing an exogenous nucleic acid into a vascular plant. The wild type form of *Agrobacterium* contains a Ti (tumor-inducing) plasmid that directs production of tumorigenic crown gall growth on host plants. Transfer of the tumor-inducing T-DNA region of the Ti plasmid to a plant genome requires the Ti plasmid-encoded virulence genes as well as T-DNA borders, which are a set of direct DNA repeats that delineate the region to be transferred. An *Agrobacterium*-based vector is a modified form of a Ti plasmid, in which the tumor inducing functions are replaced by the nucleotide sequence of interest to be introduced into the plant host.

[00660] *Agrobacterium*-mediated transformation generally employs cointegrate vectors or, e.g., binary vector systems, in which the components of the Ti plasmid are divided between a helper vector, which resides permanently in the *Agrobacterium* host and carries the virulence genes, and a shuttle vector, which contains the gene of interest bounded by T-DNA sequences. A variety of binary vectors are well known in the art and are commercially available, for example, from Clontech (Palo Alto, Calif.). Methods of coculturing *Agrobacterium* with cultured plant cells or wounded tissue such as leaf tissue, root explants, hypocotyledons, stem pieces or tubers, for example, also are well known in the art. See., e.g., Glick and Thompson, (eds.), *Methods in Plant Molecular Biology and Biotechnology*, Boca Raton, Fla.: CRC Press (1993).

[00661] *Agrobacterium*-mediated transformation is useful for producing a variety of transgenic vascular plants (Wang et al., supra, 1995) including at least one species of *Eucalyptus* and forage legumes such as alfalfa (lucerne); birdsfoot trefoil, white clover, *Stylosanthes, Lotononis bainessii* and sainfoin.

[00662] Microprojectile-mediated transformation also can be used to produce a subject transgenic plant. This method, first described by Klein et al. (*Nature* 327:70-73 (1987)), relies on microprojectiles such as gold or tungsten that are coated with the desired heterologous nucleic acid by precipitation with calcium chloride, spermidine or polyethylene

2018256863 11 Apr 2024

glycol. The microprojectile particles are accelerated at high speed into an angiosperm tissue using a device such as the BIOLISTIC PD-1000 (Biorad; Hercules Calif.).

[00663] A subject heterologous nucleic acid may be introduced into a plant in a manner such that the heterologous nucleic acid is able to enter a plant cell(s), e.g., via an *in vivo* or *ex vivo* protocol. By "*in vivo*," it may mean that the heterologous nucleic acid is administered to a living body of a plant e.g. infiltration. By "*ex vivo*" it may mean that cells or explants are modified outside of the plant, and then such cells or organs are regenerated to a plant. A number of vectors or constructs suitable for stable transformation of plant cells or for the establishment of transgenic plants have been described, including those described in Weissbach and Weissbach, (1989) Methods for Plant Molecular Biology Academic Press, and Gelvin et al., (1990) Plant Molecular Biology Manual, Kluwer Academic Publishers. Specific examples may include those derived from a Ti plasmid of *Agrobacterium tumefaciens*, as well as those disclosed by Herrera-Estrella et al. (1983) Nature 303: 209, Bevan (1984) Nucl Acid Res. 12: 8711-8721, Klee (1985) Bio/Technolo 3: 637-642. Alternatively, non-Ti vectors can be used to transfer the DNA into plants and cells by using free DNA delivery techniques. By using these methods transgenic plants such as wheat, rice (Christou (1991) Bio/Technology 9:957-962) and corn (Gordon-Kamm (1990) Plant Cell 2: 603-618) can be produced. An immature embryo can also be a good target tissue for monocots for direct DNA delivery techniques by using the particle gun (Weeks et al. (1993) Plant Physiol 102: 1077-1084; Vasil (1993) Bio/Technolo 10: 667-674; Wan and Lemeaux (1994) Plant Physiol 104: 37-48 and for Agrobacterium-mediated DNA transfer (Ishida et al. (1996) Nature Biotech 14: 745-750). Exemplary methods for introduction of DNA into chloroplasts are biolistic bombardment, polyethylene glycol transformation of protoplasts, and microinjection (Danieli et al Nat. Biotechnol 16:345-348, 1998; Staub et al Nat. Biotechnol 18: 333-338, 2000; O'Neill et al Plant J. 3:729-738, 1993; Knoblauch et al Nat. Biotechnol 17: 906-909; U.S. Pat. Nos. 5,451,513, 5,545,817, 5,545,818, and 5,576,198; in Intl. Application No. WO 95/16783; and in Boynton et al., Methods in Enzymology 217: 510-536 (1993), Svab et al., Proc. Natl. Acad. Sci. USA 90: 913-917 (1993), and McBride et al., Proc. Natl. Acad. Sci. USA 91: 7301-7305 (1994)). Any vector suitable for the methods of biolistic bombardment, polyethylene glycol transformation of protoplasts and microinjection will be suitable as a targeting vector for chloroplast transformation. Any double stranded DNA vector may be used as a transformation vector, especially when the method of introduction does not utilize *Agrobacterium*.

2018256863 11 Apr 2024

[00664] Plants which can be genetically modified may include grains, forage crops, fruits, vegetables, oil seed crops, palms, forestry, and vines. Specific examples of plants which can be modified may include, but are not limited to, maize, banana, peanut, field peas, sunflower, tobacco, tomato, canola, tobacco, wheat, barley, oats, potato, soybeans, cotton, sorghum, lupin, and rice. Plants which can be genetically modified may include *Theobroma cacao*.

[00665] Also provided by the present disclosure are transformed plant cells, tissues, plants and products that contain the transformed plant cells. A feature of the subject transformed cells, and tissues and products that include the same is the presence of a subject heterologous nucleic acid integrated into the genome, and production by plant cells of one or more polypeptides that are utilized to generate a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative. Recombinant plant cells of the present disclosure are useful as populations of recombinant cells, or as a tissue, seed, whole plant, stem, fruit, leaf, root, flower, stem, tuber, grain, animal feed, a field of plants, and the like.

[00666] Also provided by the present disclosure is reproductive material of a subject transgenic plant, where reproductive material may include seeds, progeny plants and clonal material, where such material can give rise to a plant that produces a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative.

Methods of Producing a Cannabinoid, a Cannabinoid Precursor, a Cannabinoid Derivative, or a Cannabinoid Precursor Derivative

[00667] The present disclosure provides methods of producing a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative. The methods may involve culturing a genetically modified host cell of the present disclosure in a suitable medium and recovering the produced cannabinoid, the cannabinoid precursor, the cannabinoid precursor derivative, or the cannabinoid derivative. The methods may also involve cell-free production of cannabinoids, cannabinoid precursors, cannabinoid precursor derivatives, or cannabinoid derivatives using one or more polypeptides disclosed herein expressed or overexpressed by a genetically modified host cell of the disclosure.

[00668] The present disclosure provides methods of producing a cannabinoid or a cannabinoid derivative. The methods may involve culturing a genetically modified host cell of the present disclosure in a suitable medium and recovering the produced cannabinoid or cannabinoid derivative. The methods may also involve cell-free production of cannabinoids

2018256863 11 Apr 2024

or cannabinoid derivatives using one or more polypeptides disclosed herein expressed or overexpressed by a genetically modified host cell of the disclosure.

[00669] Cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives that can be produced with the methods or genetically modified host cells of the present disclosure may include, but are not limited to, cannabichromene (CBC) type (e.g. cannabichromenic acid), cannabigerol (CBG) type (e.g. cannabigerolic acid), cannabidiol (CBD) type (e.g. cannabidiolic acid), Δ^9-trans-tetrahydrocannabinol (Δ^9 -THC) type (e.g. Δ^9-tetrahydrocannabinolic acid), Δ^8-trans-tetrahydrocannabinol (Δ^8 -THC) type, cannabicyclol (CBL) type, cannabielsoin (CBE) type, cannabinol (CBN) type, cannabinodiol (CBND) type, cannabitriol (CBT) type, olivetolic acid, GPP, derivatives of any of the foregoing, and others as listed in Elsohly M.A. and Slade D., Life Sci. 2005 Dec 22;78(5):539-48. Epub 2005 Sep 30.

[00670] Cannabinoids or cannabinoid derivatives that can be produced with the methods or genetically modified host cells of the present disclosure may also include, but are not limited to, cannabigerolic acid (CBGA), cannabigerolic acid monomethylether (CBGAM), cannabigerol (CBG), cannabigerol monomethylether (CBGM), cannabigerovarinic acid (CBGVA), cannabigerovarin (CBGV), cannabichromenic acid (CBCA), cannabichromene (CBC), cannabichromevarinic acid (CBCVA), cannabichromevarin (CBCV), cannabidiolic acid (CBDA), cannabidiol (CBD), cannabidiol monomethylether (CBDM), cannabidiol-C_4 (CBD-C_4), cannabidivarinic acid (CBDVA), cannabidivarin (CBDV), cannabidiorcol (CBD-C_1), Δ^9 –tetrahydrocannabinolic acid A (THCA-A), Δ^9 –tetrahydrocannabinolic acid B (THCA-B), Δ^9 –tetrahydrocannabinol (THC), Δ^9 –tetrahydrocannabinolic acid-C_4 (THCA-C_4), Δ^9 –tetrahydrocannabinol-C_4 (THC-C_4), Δ^9 –tetrahydrocannabivarinic acid (THCVA), Δ^9 –tetrahydrocannabivarin (THCV), Δ^9 –tetrahydrocannabiorcolic acid (THCA-C_1), Δ^9 –tetrahydrocannabiorcol (THC-C_1), Δ^7 –cis-iso-tetrahydrocannabivarin, Δ^8 –tetrahydrocannabinolic acid (Δ^8 –THCA), Δ^8 –tetrahydrocannabinol (Δ^8 –THC), cannabicyclolic acid (CBLA), cannabicyclol (CBL), cannabicyclovarin (CBLV), cannabielsoic acid A (CBEA-A), cannabielsoic acid B (CBEA-B), cannabielsoin (CBE), cannabielsoinic acid, cannabicitranic acid, cannabinolic acid (CBNA), cannabinol (CBN), cannabinol methylether (CBNM), cannabinol-C_4, (CBN-C_4), cannabivarin (CBV), cannabinol-C_2 (CNB-C_2), cannabiorcol (CBN-C_1), cannabinodiol (CBND), cannabinodivarin (CBVD), cannabitriol (CBT), 10-ethyoxy-9-hydroxy-delta-6a-tetrahydrocannabinol, 8,9-dihydroxyl-delta-6a-tetrahydrocannabinol, cannabitriolvarin (CBTVE), dehydrocannabifuran (DCBF), cannabifuran (CBF), cannabichromanon (CBCN),

2018256863 11 Apr 2024

cannabicitran (CBT), 10-oxo-delta-6a-tetrahydrocannabinol (OTHC), delta-9-cis-tetrahydrocannabinol (cis-THC), 3,4,5,6-tetrahydro-7-hydroxy-alpha-alpha-2-trimethyl-9-n-propyl-2,6-methano-2H-1-benzoxocin-5-methanol (OH-iso-HHCV), cannabiripsol (CBR), trihydroxy-delta-9-tetrahydrocannabinol (triOH-THC), and derivatives of any of the foregoing.

[00671] Additional cannabinoid derivatives that can be produced with the methods or genetically modified host cells of the present disclosure may also include, but are not limited to, 2-geranyl-5-pentyl-resorcylic acid, 2-geranyl-5-(4-pentynyl)-resorcylic acid, 2-geranyl-5-(trans-2-pentenyl)-resorcylic acid, 2-geranyl-5-(4-methylhexyl)-resorcylic acid, 2-geranyl-5-(5-hexynyl) resorcylic acid, 2-geranyl-5-(trans-2-hexenyl)-resorcylic acid, 2-geranyl-5-(5-hexenyl)-resorcylic acid, 2-geranyl-5-heptyl-resorcylic acid, 2-geranyl-5-(6-heptynoic)-resorcylic acid, 2-geranyl-5-octyl-resorcylic acid, 2-geranyl-5-(trans-2-octenyl)-resorcylic acid, 2-geranyl-5-nonyl-resorcylic acid, 2-geranyl-5-(trans-2-nonenyl) resorcylic acid, 2-geranyl-5-decyl-resorcylic acid, 2-geranyl-5-(4-phenylbutyl)-resorcylic acid, 2-geranyl-5-(5-phenylpentyl)-resorcylic acid, 2-geranyl-5-(6-phenylhexyl)-resorcylic acid, 2-geranyl-5-(7-phenylheptyl)-resorcylic acid, (6aR,10aR)-1-hydroxy-6,6,9-trimethyl-3-propyl-6a,7,8,10a-tetrahydro-6H-dibenzo[b,d]pyran-2-carboxylic acid, (6aR,10aR)-1-hydroxy-6,6,9-trimethyl-3-(4-methylhexyl)-6a,7,8,10a-tetrahydro-6H-dibenzo[b,d]pyran-2-carboxylic acid, (6aR,10aR)-1-hydroxy-6,6,9-trimethyl-3-(5-hexenyl)-6a,7,8,10a-tetrahydro-6H-dibenzo[b,d]pyran-2-carboxylic acid, (6aR,10aR)-1-hydroxy-6,6,9-trimethyl-3-(5-hexenyl)-6a,7,8,10a-tetrahydro-6H-dibenzo[b,d]pyran-2-carboxylic acid, (6aR,10aR)-1-hydroxy-6,6,9-trimethyl-3-(6-heptynyl)-6a,7,8,10a-tetrahydro-6H-dibenzo[b,d]pyran-2-carboxylic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-6-(hexan-2-yl)-2,4-dihydroxybenzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-(2-methylpentyl)benzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-(3-methylpentyl)benzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-(4-methylpentyl)benzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-[(1E)-pent-1-en-1-yl]benzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-[(2E)-pent-2-en-1-yl]benzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-[(2E)-pent-3-en-1-yl]benzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-(pent-4-en-1-yl)benzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-propylbenzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-butylbenzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-hexylbenzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-heptylbenzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-

2018256863 11 Apr 2024

6-octylbenzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-nonanylbenzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-decanylbenzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-undecanylbenzoic acid, 6-(4-chlorobutyl)-3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxybenzoic acid, 3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-[4-(methylsulfanyl)butyl]benzoic acid, and others as listed in Bow, E. W. and Rimoldi, J. M., "The Structure–Function Relationships of Classical Cannabinoids: CB1/CB2 Modulation," *Perspectives in Medicinal Chemistry* 2016:8 17–39 doi: 10.4137/PMC.S32171, incorporated by reference herein.

[00672] Cannabinoid precursor derivatives that can be produced with the methods or genetically modified host cells of the present disclosure may also include, but are not limited to, divarinolic acid, 5-pentyl-resorcylic acid, 5-(4-pentynyl)-resorcylic acid, 5-(trans-2-pentenyl)-resorcylic acid, 5-(4-methylhexyl)-resorcylic acid, 5-(5-hexynyl)-resorcylic acid, 5-(trans-2-hexenyl)-resorcylic acid, 5-(5-hexenyl)-resorcylic acid, 5-heptyl-resorcylic acid, 5-(6-heptynoic)-resorcylic acid, 5-octyl-resorcylic acid, 5-(trans-2-octenyl)-resorcylic acid, 5-nonyl-resorcylic acid, 5-(trans-2-nonenyl)-resorcylic acid, 5-decyl-resorcylic acid, 5-(4-phenylbutyl)-resorcylic acid, 5-(5-phenylpentyl)-resorcylic acid, 5-(6-phenylhexyl)-resorcylic acid, and 5-(7-phenylheptyl)-resorcylic acid.

[00673] Cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives that can be produced with the methods or genetically modified host cells of the present disclosure may also include, but are not limited to, polyketides or polyketide derivatives.

[00674] A cannabinoid derivative or cannabinoid precursor derivative may lack one or more chemical moieties found in a naturally-occurring cannabinoid or naturally-occurring cannabinoid precursor. Such chemical moieties may include, but are not limited to, methyl, alkyl, alkenyl, methoxy, alkoxy, acetyl, carboxyl, carbonyl, oxo, ester, hydroxyl, aryl, heteroaryl, cycloalkyl, cycloalkenyl, cycloalkylalkenyl, cycloalkenylalkyl, cycloalkenylalkenyl, heterocyclylalkenyl, heteroarylalkenyl, arylalkenyl, heterocyclyl, aralkyl, cycloalkylalkyl, heterocyclylalkyl, heteroarylalkyl, and the like. In some embodiments, a cannabinoid derivative or cannabinoid precursor derivative lacking one or more chemical moieties found in a naturally-occurring cannabinoid or naturally-occurring cannabinoid precursor, and produced by a genetically modified host cell disclosed herein or in a cell-free reaction mixture comprising one or more of the polypeptides disclosed herein, may also comprise one or more of any of the functional and/or reactive groups described

2018256863 11 Apr 2024

herein. Functional and reactive groups may be optionally substituted with one or more additional functional or reactive groups.

[00675] A cannabinoid derivative or cannabinoid precursor derivative may be a cannabinoid or cannabinoid precursor comprising one or more functional and/or reactive groups and is produced by a genetically modified host cell disclosed herein or in a cell-free reaction mixture comprising one or more of the polypeptides disclosed herein. Functional groups may include, but are not limited to, azido, halo (e.g., chloride, bromide, iodide, fluorine), methyl, alkyl, alkynyl, alkenyl, methoxy, alkoxy, acetyl, amino, carboxyl, carbonyl, oxo, ester, hydroxyl, thio, cyano, aryl, heteroaryl, cycloalkyl, cycloalkenyl, cycloalkylalkenyl, cycloalkylalkynyl, cycloalkenylalkyl, cycloalkenylalkenyl, cycloalkenylalkynyl, heterocyclylalkenyl, heterocyclylalkynyl, heteroarylalkenyl, heteroarylalkynyl, arylalkenyl, arylalkynyl, spirocyclyl, heterospirocyclyl, heterocyclyl, thioalkyl, sulfone, sulfonyl, sulfoxide, amido, alkylamino, dialkylamino, arylamino, alkylarylamino, diarylamino, N-oxide, imide, enamine, imine, oxime, hydrazone, nitrile, aralkyl, cycloalkylalkyl, haloalkyl, heterocyclylalkyl, heteroarylalkyl, nitro, thioxo, and the like. See, e.g., FIGS. 12 and 13. Suitable reactive groups may include, but are not necessarily limited to, azide, carboxyl, carbonyl, amine, (e.g., alkyl amine (e.g., lower alkyl amine), aryl amine), halide, ester (e.g., alkyl ester (e.g., lower alkyl ester, benzyl ester), aryl ester, substituted aryl ester), cyano, thioester, thioether, sulfonyl halide, alcohol, thiol, succinimidyl ester, isothiocyanate, iodoacetamide, maleimide, hydrazine, alkynyl, alkenyl, acetyl, and the like. In some embodiments, the reactive group is selected from a carboxyl, a carbonyl, an amine, an ester, a thioester, a thioether, a sulfonyl halide, an alcohol, a thiol, an alkyne, alkene, an azide, a succinimidyl ester, an isothiocyanate, an iodoacetamide, a maleimide, and a hydrazine. Functional and reactive groups may be optionally substituted with one or more additional functional or reactive groups.

[00676] A reactive group may facilitate covalent attachment of a molecule of interest. Suitable molecules of interest may include, but are not limited to, a detectable label; imaging agents; a toxin (including cytotoxins); a linker; a peptide; a drug (e.g., small molecule drugs); a member of a specific binding pair; an epitope tag; ligands for binding by a target receptor; tags to aid in purification; molecules that increase solubility; and the like. A linker may be a peptide linker or a non-peptide linker.

[00677] In some embodiments, a cannabinoid derivative or a cannabinoid precursor derivative comprising an azide may be reacted with a compound comprising an alkyne group via “click chemistry” to generate a product comprising a heterocycle, also known as an

2018256863 11 Apr 2024

azide-alkyne cycloaddition. In some embodiments, a cannabinoid derivative or a cannabinoid precursor derivative comprising an alkyne may be reacted with a compound comprising an azide group via click chemistry to generate a product comprising a heterocycle.

[00678] Additional molecules that may be desirable for attachment to a cannabinoid derivative or cannabinoid precursor derivative may include, but are not necessarily limited to, detectable labels (e.g., spin labels, fluorescence resonance energy transfer (FRET)-type dyes, e.g., for studying structure of biomolecules *in vivo*); small molecule drugs; cytotoxic molecules (e.g., drugs); imaging agents; ligands for binding by a target receptor; tags to aid in purification by, for example, affinity chromatography (e.g., attachment of a FLAG epitope); molecules that increase solubility (e.g., poly(ethylene glycol); molecules that enhance bioavailability; molecules that increase *in vivo* half-life; molecules that target to a particular cell type (e.g., an antibody specific for an epitope on a target cell); molecules that target to a particular tissue; molecules that provide for crossing the blood-brain barrier; and molecules to facilitate selective attachment to a surface, and the like.

[00679] In some embodiments, a molecule of interest comprises an imaging agent. Suitable imaging agents may include positive contrast agents and negative contrast agents. Suitable positive contrast agents may include, but are not limited to, gadolinium tetraazacyclododecanetetraacetic acid (Gd-DOTA); gadolinium-diethylenetriaminepentaacetic acid (Gd-DTPA); gadolinium-1,4,7-tris(carbonylmethyl)-10-(2'-hydroxypropyl)-1,4,7,10-tetraazacyclododecane (Gd-HP-DO3A); Manganese(II)-dipyridoxal diphosphate (Mn-DPDP); Gd-diethylenetriaminepentaacetate-bis(methylamide) (Gd-DTPA-BMA); and the like. Suitable negative contrast agents may include, but are not limited to, a superparamagnetic iron oxide (SPIO) imaging agent; and a perfluorocarbon, where suitable perfluorocarbons may include, but are not limited to, fluoroheptanes, fluorocycloheptanes, fluoromethylcycloheptanes, fluorohexanes, fluorocyclohexanes, fluoropentanes, fluorocyclopentanes, fluoromethylcyclopentanes, fluorodimethylcyclopentanes, fluoromethylcyclobutanes, fluorodimethylcyclobutanes, fluorotrimethylcyclobutanes, fluorobutanes, fluorocyclobutanse, fluoropropanes, fluoroethers, fluoropolyethers, fluorotriethylamines, perfluorohexanes, perfluoropentanes, perfluorobutanes, perfluoropropanes, sulfur hexafluoride, and the like.

[00680] Additional cannabinoid derivatives and cannabinoid precursor derivatives that can be produced with a method or genetically modified host cell of the present disclosure may include derivatives that have been modified via organic synthesis or an enzymatic route

2018256863 11 Apr 2024

to modify drug metabolism and pharmacokinetics (e.g. solubility, bioavailability, absorption, distribution, plasma half-life and metabolic clearance). Modification examples may include, but are not limited to, halogenation, acetylation and methylation.

[00681] The cannabinoids, cannabinoid derivatives, cannabinoid precursors, and cannabinoid precursor derivatives described herein further include all pharmaceutically acceptable isotopically labeled cannabinoids, cannabinoid derivatives, cannabinoid precursors, and cannabinoid precursor derivatives. An "isotopically-" or "radio-labeled" compound is a compound where one or more atoms are replaced or substituted by an atom having an atomic mass or mass number different from the atomic mass or mass number typically found in nature (i.e., naturally occurring). For example, in some embodiments, in the cannabinoids, cannabinoid derivatives, cannabinoid precursors, and cannabinoid precursor derivatives described herein hydrogen atoms are replaced or substituted by one or more deuterium or tritium. Certain isotopically labeled cannabinoids, cannabinoid derivatives, cannabinoid precursors, and cannabinoid precursor derivatives of this disclosure, for example, those incorporating a radioactive isotope, are useful in drug and/or substrate tissue distribution studies. The radioactive isotopes tritium, i.e., ^{3}H, and carbon 14, i.e., ^{14}C, are particularly useful for this purpose in view of their ease of incorporation and ready means of detection. Substitution with heavier isotopes such as deuterium, i.e., ^{2}H, may afford certain therapeutic advantages resulting from greater metabolic stability, for example, increased *in vivo* half-life or reduced dosage requirements, and hence may be preferred in some circumstances. Suitable isotopes that may be incorporated in cannabinoids, cannabinoid derivatives, cannabinoid precursors, and cannabinoid precursor derivatives described herein include but are not limited to ^{2}H (also written as D for deuterium), ^{3}H (also written as T for tritium), ^{11}C, ^{13}C, ^{14}C, ^{13}N, ^{15}N, ^{15}O, ^{17}O, ^{18}O, ^{18}F, ^{35}S, ^{36}Cl , ^{82}Br, ^{75}Br, ^{76}Br, ^{77}Br, ^{123}I, ^{124}I, ^{125}I, and ^{131}I. Substitution with positron emitting isotopes, such as ^{11}C, ^{18}F, ^{15}O, and ^{13}N, can be useful in Positron Emission Topography (PET) studies.

[00682] The methods of bioproduction disclosed herein enable synthesis of cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives with defined stereochemistries, which is challenging to do using chemical synthesis. Cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives disclosed herein may be enantiomers or disastereomers. The term "enantiomers" may refer to a pair of stereoisomers which are non-superimposable mirror images of one another. In some embodiments the cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives may be the (*S*)-enantiomer. In

some embodiments the cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives may be the (*R*)-enantiomer. In some embodiments, the cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives may be the (+) or (-) enantiomers. The term "diastereomers" may refer to the set of stereoisomers which cannot be made superimposable by rotation around single bonds. For example, cis- and trans- double bonds, endo- and exo- substitution on bicyclic ring systems, and compounds containing multiple stereogenic centers with different relative configurations may be considered to be diastereomers. The term "diastereomer" may refer to any member of this set of compounds. Cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives disclosed herein may include a double bond or a fused ring. In certain such embodiments, the double bond or fused ring may be cis or trans, unless the configuration is specifically defined. If the cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative contains a double bond, the substituent may be in the E or Z configuration, unless the configuration is specifically defined.

[00683] In some embodiments when the cannabinoid, cannabinoid precursor, cannabinoid precursor derivative, or cannabinoid derivative is recovered from the cell lysate, from the culture medium, from both the cell lysate and the culture medium, or from a cell-free reaction mixture comprising one or more polypeptides disclosed herein, the recovered cannabinoid, cannabinoid precursor, cannabinoid precursor derivative, or cannabinoid derivative is in the form of a salt. In certain such embodiments, the salt is a pharmaceutically acceptable salt. In some embodiments, the salt of the recovered cannabinoid, cannabinoid precursor, cannabinoid precursor derivative, or cannabinoid derivative is then purified as disclosed herein.

[00684] The disclosure includes pharmaceutically acceptable salts of the cannabinoids, cannabinoid derivatives, cannabinoid precursors, and cannabinoid precursor derivatives described herein. "Pharmaceutically acceptable salts" refer to those salts which retain the biological effectiveness and properties of the free bases, which are not biologically or otherwise undesirable. Representative pharmaceutically acceptable salts include, but are not limited to, e.g., water-soluble and water-insoluble salts, such as the acetate, amsonate (4,4-diaminostilbene-2,2-disulfonate), benzenesulfonate, benzonate, bicarbonate, bisulfate, bitartrate, borate, bromide, butyrate, calcium, calcium edetate, camsylate, carbonate, chloride, citrate, clavulariate, dihydrochloride, edetate, edisylate, estolate, esylate, fiunarate, gluceptate, gluconate, glutamate, glycollylarsanilate, hexafluorophosphate, hexylresorcinate,

hydrabamine, hydrobromide, hydrochloride, hydroxynaphthoate, iodide, sethionate, lactate, lactobionate, laurate, magnesium, malate, maleate, mandelate, mesylate, methylbromide, methylnitrate, methylsulfate, mucate, napsylate, nitrate, N-methylglucamine ammonium salt, 3-hydroxy-2-naphthoate, oleate, oxalate, palmitate, pamoate (1,1-methene-bis-2-hydroxy-3-naphthoate, einbonate), pantothenate, phosphate/diphosphate, picrate, polygalacturonate, propionate, p-toluenesulfonate, salicylate, stearate, subacetate, succinate, sulfate, sulfosalicylate, suramate, tannate, tartrate, teoclate, tosylate, triethiodide, and valerate salts.

[00685] "Pharmaceutically acceptable salt" also includes both acid and base addition salts. "Pharmaceutically acceptable acid addition salt" refers to those salts which retain the biological effectiveness and properties of the free bases, which are not biologically or otherwise undesirable, and which are formed with inorganic acids such as, but are not limited to, hydrochloric acid, hydrobromic acid, sulfuric acid, nitric acid, phosphoric acid and the like, and organic acids such as, but not limited to, acetic acid, 2,2-dichloroacetic acid, adipic acid, alginic acid, ascorbic acid, aspartic acid, benzenesulfonic acid, benzoic acid, 4-acetamidobenzoic acid, camphoric acid, camphor-10-sulfonic acid, capric acid, caproic acid, caprylic acid, carbonic acid, cinnamic acid, citric acid, cyclamic acid, dodecylsulfuric acid, ethane-1,2-disulfonic acid, ethanesulfonic acid, 2-hydroxyethanesulfonic acid, formic acid, fumaric acid, galactaric acid, gentisic acid, glucoheptonic acid, gluconic acid, glucuronic acid, glutamic acid, glutaric acid, 2-oxo-glutaric acid, glycerophosphoric acid, glycolic acid, hippuric acid, isobutyric acid, lactic acid, lactobionic acid, lauric acid, maleic acid, malic acid, malonic acid, mandelic acid, methanesulfonic acid, mucic acid, naphthalene-l,5-disulfonic acid, naphthalene-2-sulfonic acid, l-hydroxy-2-naphthoic acid, nicotinic acid, oleic acid, orotic acid, oxalic acid, palmitic acid, pamoic acid, propionic acid, pyroglutamic acid, pyruvic acid, salicylic acid, 4-aminosalicylic acid, sebacic acid, stearic acid, succinic acid, tartaric acid, thiocyanic acid, p-toluenesulfonic acid, trifluoroacetic acid, undecylenic acid, and the like.

[00686] "Pharmaceutically acceptable base addition salt" refers to those salts which retain the biological effectiveness and properties of the free acids, which are not biologically or otherwise undesirable. These salts are prepared from addition of an inorganic base or an organic base to the free acid. Salts derived from inorganic bases include, but are not limited to, the sodium, potassium, lithium, ammonium, calcium, magnesium, iron, zinc, copper, manganese, aluminum salts and the like. For example, inorganic salts include, but are not limited to, ammonium, sodium, potassium, calcium, and magnesium salts. Salts derived from organic bases include, but are not limited to, salts of primary, secondary, and tertiary amines,

substituted amines including naturally occurring substituted amines, cyclic amines and basic ion exchange resins, such as ammonia, isopropylamine, trimethylamine, diethylamine, triethylamine, tripropylamine, diethanolamine, ethanolamine, deanol, 2-dimethylaminoethanol, 2-diethylaminoethanol, dicyclohexylamine, lysine, arginine, histidine, caffeine, procaine, hydrabamine, choline, betaine, benethamine, benzathine, ethylenediamine, glucosamine, methylglucamine, theobromine, triethanolamine, tromethamine, purines, piperazine, piperidine, N-ethylpiperidine, polyamine resins and the like.

Methods of Using Host Cells to Generate Cannabinoids, Cannabinoid Precursors, Cannabinoid Derivatives, or Cannabinoid Precursor Derivatives

[00687] The disclosure provides methods of producing a cannabinoid, a cannabinoid precursor, a cannabinoid precursor derivative, or a cannabinoid derivative in a genetically modified host cell, the method comprising: culturing a genetically modified host cell of the disclosure in a suitable medium and recovering the produced cannabinoid, cannabinoid precursor, cannabinoid precursor derivative, or cannabinoid derivative. In certain such embodiments, the produced cannabinoid, cannabinoid precursor, cannabinoid precursor derivative, or cannabinoid derivative is then purified as disclosed herein.

[00688] In some embodiments, culturing of the genetically modified host cells of the disclosure in a suitable medium provides for synthesis of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in an increased amount compared to a non-genetically modified host cell cultured under similar conditions.

[00689] The disclosure provides methods of producing a cannabinoid, a cannabinoid precursor, a cannabinoid precursor derivative, or a cannabinoid derivative, the method comprising: culturing a genetically modified host cell of the disclosure in a suitable medium comprising a carboxylic acid and recovering the produced cannabinoid, cannabinoid precursor, cannabinoid precursor derivative, or cannabinoid derivative. In certain such embodiments, the produced cannabinoid, cannabinoid precursor, cannabinoid precursor derivative, or cannabinoid derivative is then purified as disclosed herein.

[00690] In some embodiments, the cannabinoid, cannabinoid precursor, cannabinoid precursor derivative, or cannabinoid derivative is recovered from the cell lysate, from the culture medium, or from both the cell lysate and the culture medium. In certain such embodiments, the recovered cannabinoid, cannabinoid precursor, cannabinoid precursor derivative, or cannabinoid derivative is then purified as disclosed herein.

[00691] The disclosure provides methods of producing a cannabinoid or a cannabinoid derivative in a genetically modified host cell, the method comprising: culturing a genetically modified host cell of the disclosure in a suitable medium and recovering the produced cannabinoid or cannabinoid derivative. In certain such embodiments, the produced cannabinoid or cannabinoid derivative is then purified as disclosed herein.

[00692] In some embodiments, culturing of the genetically modified host cells of the disclosure in a suitable medium provides for synthesis of a cannabinoid or a cannabinoid derivative in an increased amount compared to a non-genetically modified host cell cultured under similar conditions.

[00693] The disclosure provides methods of producing a cannabinoid or a cannabinoid derivative, the method comprising: culturing a genetically modified host cell of the disclosure in a suitable medium comprising a carboxylic acid and recovering the produced cannabinoid or cannabinoid derivative. In certain such embodiments, the produced cannabinoid or cannabinoid derivative is then purified as disclosed herein.

[00694] In some embodiments, the cannabinoid or cannabinoid derivative is recovered from the cell lysate, from the culture medium, or from both the cell lysate and the culture medium. In certain such embodiments, the recovered cannabinoid or cannabinoid derivative is then purified as disclosed herein.

[00695] In some embodiments, the genetically modified host cell of the present disclosure is cultured in a suitable medium comprising a carboxylic acid to generate an acyl-CoA compound or an acyl-CoA compound derivative. In certain such embodiments, the genetically modified host cell is genetically modified with one or more heterologous nucleic acids encoding an AAE polypeptide, a FAA polypeptide, or a fatty acyl-CoA ligase polypeptide, as described herein. In some embodiments, the genetically modified host cells of the present disclosure may further convert an acyl-CoA compound or an acyl-CoA compound derivative to cannabinoids, cannabinoid precursors, cannabinoid precursor derivatives, or cannabinoid derivatives.

[00696] Carboxylic acids may include, but are not limited to, C_3-C_{18} fatty acids, butyric acid, isobutyric acid, valeric acid, hexanoic acid, heptanoic acid, octanoic acid, nonanoic acid, decanoic acid, undecanoic acid, lauric acid, myristic acid, C_{15}-C_{18} fatty acids, fumaric acid, itaconic acid, malic acid, succinic acid, maleic acid, malonic acid, glutaric acid, glucaric acid, oxalic acid, adipic acid, pimelic acid, suberic acid, azelaic acid, sebacic acid, dodecanedioic acid, glutaconic acid, ortho-phthalic acid, isophthalic acid, terephthalic acid, citric acid, isocitric acid, aconitic acid, tricarballylic acid, and

2018256863 11 Apr 2024

trimesic acid. Carboxylic acids may include C_4-C_{10} carboxylic acids. In some embodiments, the carboxylic acid is a C_4 carboxylic acid. In some embodiments, the carboxylic acid is a C_5 carboxylic acid. In some embodiments, the carboxylic acid is a C_6 carboxylic acid. In some embodiments, the carboxylic acid is a C_7 carboxylic acid. In some embodiments, the carboxylic acid is a C_8 carboxylic acid. In some embodiments, the carboxylic acid is a C_9 carboxylic acid. In some embodiments, the carboxylic acid is a C_{10} carboxylic acid. In some embodiments, the carboxylic acid is butyric acid. In some embodiments, the carboxylic acid is valeric acid. In some embodiments, the carboxylic acid is hexanoic acid. In some embodiments, the carboxylic acid is heptanoic acid. In some embodiments, the carboxylic acid is octanoic acid. In some embodiments, the carboxylic acid is nonanoic acid. In some embodiments, the carboxylic acid is decanoic acid. See, e.g., FIG. 12.

[00697] In some embodiments, the carboxylic acid comprises one or more functional and/or reactive groups to generate derivatives of hexanoyl-CoA or derivatives of acyl-CoA compounds. Functional groups may include, but are not limited to, azido, halo (e.g., chloride, bromide, iodide, fluorine), methyl, alkyl, alkynyl, alkenyl, methoxy, alkoxy, acetyl, amino, carboxyl, carbonyl, oxo, ester, hydroxyl, thio, cyano, aryl, heteroaryl, cycloalkyl, cycloalkenyl, cycloalkylalkenyl, cycloalkylalkynyl, cycloalkenylalkyl, cycloalkenylalkenyl, cycloalkenylalkynyl, heterocyclylalkenyl, heterocyclylalkynyl, heteroarylalkenyl, heteroarylalkynyl, arylalkenyl, arylalkynyl, spirocyclyl, heterospirocyclyl, heterocyclyl, thioalkyl, sulfone, sulfonyl, sulfoxide, amido, alkylamino, dialkylamino, arylamino, alkylarylamino, diarylamino, N-oxide, imide, enamine, imine, oxime, hydrazone, nitrile, aralkyl, cycloalkylalkyl, haloalkyl, heterocyclylalkyl, heteroarylalkyl, nitro, thioxo, and the like. See, e.g., FIGS. 12 and 13. Reactive groups may include, but are not necessarily limited to, azide, halogen, carboxyl, carbonyl, amine, (e.g., alkyl amine (e.g., lower alkyl amine), aryl amine), ester (e.g., alkyl ester (e.g., lower alkyl ester, benzyl ester), aryl ester, substituted aryl ester), cyano, thioester, thioether, sulfonyl halide, alcohol, thiol, succinimidyl ester, isothiocyanate, iodoacetamide, maleimide, hydrazine, alkynyl, alkenyl, and the like. In some embodiments, the reactive group is selected from a carboxyl, a carbonyl, an amine, an ester, thioester, thioether, a sulfonyl halide, an alcohol, a thiol, a succinimidyl ester, an isothiocyanate, an iodoacetamide, a maleimide, an azide, an alkyne, an alkene, and a hydrazine. Functional and reactive groups may be optionally substituted with one or more additional functional or reactive groups.

[00698] In some embodiments, the carboxylic acid is isotopically- or radio-labeled. In some embodiments, the carboxylic acid may be an enantiomer or disastereomer. In some

embodiments the carboxylic acid may be the (*S*)-enantiomer. In some embodiments the carboxylic acid may be the (*R*)-enantiomer. In some embodiments, the carboxylic acid may be the (+) or (-) enantiomer. In some embodiments, the carboxylic acid may include a double bond or a fused ring. In certain such embodiments, the double bond or fused ring may be cis or trans, unless the configuration is specifically defined. If the carboxylic acid contains a double bond, the substituent may be in the E or Z configuration, unless the configuration is specifically defined.

[00699] In some embodiments, the carboxylic acid comprises a C=C group. In some embodiments, the carboxylic acid comprises an alkyne group. In some embodiments, the carboxylic acid comprises an N_3 group. In some embodiments, the carboxylic acid comprises a halogen. In some embodiments, the carboxylic acid comprises a CN group. In some embodiments, the carboxylic acid comprises an iodide. In some embodiments, the carboxylic acid comprises a bromide. In some embodiments, the carboxylic acid comprises chloride. In some embodiments, the carboxylic acid comprises fluoride. In some embodiments, the carboxylic acid comprises a carbonyl. In some embodiments, the carboxylic acid comprises an acetyl. In some embodiments, the carboxylic acid comprises an alkyl group.

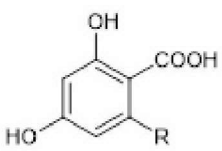
[00700] Carboxylic acids may include, but are not limited to, 2-methylhexanoic acid, 3-methylhexanoic acid, 4-methylhexanoic acid, 5-methylhexanoic acid, 2-hexenoic acid, 3-hexenoic acid, 4-hexenoic acid, 5-hexenoic acid, 5-chlorovaleric acid, 5-aminovaleric acid, 5-cyanovaleric acid, 5-(methylsulfanyl)valeric acid, 5-hydroxyvaleric acid, 5-phenylvaleric acid, 2,3-dimethylhexanoic acid, d_3-hexanoic acid, 5-chloropentanoic acid, 5-(methylsulfanyl)pentanoic acid, 4-pentynoic acid, trans-2-pentenoic acid, 5-hexynoic acid, trans-2-hexenoic acid, 6-heptynoic acid, trans-2-octenoic acid, trans-2-nonenoic acid, 4-phenylbutyric acid, 6-phenylhexanoic acid, 7-phenylyheptanoic acid, and the like. In some embodiments, the carboxylic acid is 2-methylhexanoic acid. In some embodiments, the carboxylic acid is 3-methylhexanoic acid. In some embodiments, the carboxylic acid is 4-methylhexanoic acid. In some embodiments, the carboxylic acid is 5-methylhexanoic acid. In some embodiments, the carboxylic acid is 2-hexenoic acid. In some embodiments, the carboxylic acid is 3-hexenoic acid. In some embodiments, the carboxylic acid is 4-hexenoic acid. In some embodiments, the carboxylic acid is 5-hexenoic acid. In some embodiments, the carboxylic acid is 5-chlorovaleric acid. In some embodiments, the carboxylic acid is 5-aminovaleric acid. In some embodiments, the carboxylic acid is 5-cyanovaleric acid. In some embodiments, the carboxylic acid is 5-(methylsulfanyl)valeric acid. In some embodiments, the carboxylic acid is 5-hydroxyvaleric acid. In some embodiments, the

2018256863 11 Apr 2024

carboxylic acid is 5-phenylvaleric acid. In some embodiments, the carboxylic acid is 2,3-dimethylhexanoic acid. In some embodiments, the carboxylic acid is d_3-hexanoic acid. In some embodiments, the carboxylic acid is 5-chloropentanoic acid. In some embodiments, the carboxylic acid is 5-(methylsulfanyl)pentanoic acid. In some embodiments, the carboxylic acid is 4-pentynoic acid. In some embodiments, the carboxylic acid is trans-2-pentenoic acid. In some embodiments, the carboxylic acid is 5-hexynoic acid. In some embodiments, the carboxylic acid is trans-2-hexenoic acid. In some embodiments, the carboxylic acid is 6-heptynoic acid. In some embodiments, the carboxylic acid is trans-2-octenoic acid. In some embodiments, the carboxylic acid is trans-2-nonenoic acid. In some embodiments, the carboxylic acid is 4-phenylbutyric acid. In some embodiments, the carboxylic acid is 6-phenylhexanoic acid. In some embodiments, the carboxylic acid is 7-phenylheptanoic acid.

[00701] The disclosure also provides methods of producing the following cannabinoid precursor or precursor derivatives: olivetolic acid or olivetolic acid derivatives. In certain such embodiments, the method comprises: culturing a genetically modified host cell of the disclosure in a suitable medium comprising a carboxylic acid, and recovering the produced olivetolic acid or olivetolic acid derivative. In certain such embodiments, the genetically modified host cell of the disclosure is genetically modified with: a) one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative, such as an AAE polypeptide, a FAA polypeptide, or a fatty acyl-CoA ligase polypeptide; b) one or more heterologous nucleic acids encoding a TKS polypeptide; and c) one or more heterologous nucleic acids encoding an OAC polypeptide. In some embodiments, the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative, such as an AAE polypeptide, a FAA polypeptide, or a fatty acyl-CoA ligase polypeptide; and b) one or more heterologous nucleic acids encoding a TKS/OAC fusion polypeptide.

[00702] In some embodiments, the olivetolic acid derivative produced by the methods or genetically modified host cells of the disclosure has the following formula:



where R is alkyl (e.g., C_1-C_{10} alkyl), substituted alkyl, alkyl ester, or alkyl-X, where X is a reactive or functional group, as disclosed herein. In some embodiments, the olivetolic acid or

2018256863 11 Apr 2024

olivetolic acid derivative is recovered from the cell lysate, from the culture medium, or from both the cell lysate and the culture medium. In certain such embodiments, the recovered olivetolic acid or olivetolic acid derivative is then purified as disclosed herein. In some embodiments, the olivetolic acid or olivetolic acid derivative is further converted by the genetically modified host cell to a cannabinoid derivative or a cannabinoid.

[00703] The disclosure also provides methods of producing a cannabinoid or a cannabinoid derivative, the method comprising: culturing a genetically modified host cell of the disclosure in a suitable medium comprising olivetolic acid or an olivetolic acid derivative and recovering the produced cannabinoid or cannabinoid derivative. In certain such embodiments, the produced cannabinoid or cannabinoid derivative is then purified as disclosed herein. The disclosure also provides methods of producing a cannabinoid derivative, the method comprising: culturing a genetically modified host cell of the disclosure in a suitable medium comprising olivetolic acid or an olivetolic acid derivative and recovering the produced cannabinoid derivative. In certain such embodiments, the produced cannabinoid derivative is then purified as disclosed herein.

[00704] In some embodiments, the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode a GOT polypeptide and b) one or more heterologous nucleic acids that encode a polypeptide that generates GPP (e.g., a GPPS polypeptide). In some embodiments, the olivetolic acid or olivetolic acid derivative is further converted to a cannabinoid derivative or a cannabinoid.

[00705] Olivetolic acid derivatives used herein may comprise one or more reactive and/or functional groups as disclosed herein. In some embodiments when the suitable medium comprises an olivetolic acid derivative, the olivetolic acid derivative is orsellinic acid. In some embodiments when the suitable medium comprises an olivetolic acid derivative, the olivetolic acid derivative is divarinic acid.

[00706] In some embodiments, a method of producing a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative may involve growing a transgenic (genetically modified) plant of the present disclosure under conditions that favor production of the cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative. The cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative can be purified from the plant, or a part of the plant. The present disclosure provides food products made from a transgenic (genetically modified) plant of the present disclosure.

2018256863 11 Apr 2024

Exemplary Cell Culture Conditions

[00707] Suitable media may include standard culture media (e.g., Luria-Bertani broth, optionally supplemented with one or more additional agents, such as an inducer (e.g., where heterologous nucleic acids disclosed herein is under the control of an inducible promoter, etc.); standard yeast culture media; and the like). In some embodiments, the culture medium can be supplemented with a fermentable sugar (e.g., a hexose sugar, e.g., glucose, xylose, and the like). In some embodiments, the culture medium can be supplemented with hexanoate, carboxylic acids other than hexanoate, olivetolic acid, or olivetolic acid derivatives. In some embodiments, the culture medium can be supplemented with pretreated cellulosic feedstock (e.g., wheat grass, wheat straw, barley straw, sorghum, rice grass, sugarcane straw, bagasse, switchgrass, corn stover, corn fiber, grains, or any combination thereof). In some embodiments, the culture medium can be supplemented with oleic acid. In some embodiments, the suitable medium comprises a non-fermentable carbon source. In certain such embodiments, the non-fermentable carbon source comprises ethanol. In some embodiments, the suitable media comprises an inducer. In certain such embodiments, the inducer comprises galactose.

[00708] The carbon source in the suitable media can vary significantly, from simple sugars like glucose to more complex hydrolysates of other biomass, such as yeast extract. The addition of salts generally provide essential elements such as magnesium, nitrogen, phosphorus, and sulfur to allow the cells to synthesize polypeptides and nucleic acids. The suitable media can also be supplemented with selective agents, such as antibiotics, to select for the maintenance of certain plasmids and the like. For example, if a microorganism is resistant to a certain antibiotic, such as ampicillin or tetracycline, then that antibiotic can be added to the medium in order to prevent cells lacking the resistance from growing. The suitable media can be supplemented with other compounds as necessary to select for desired physiological or biochemical characteristics, such as particular amino acids and the like.

[00709] In some embodiments, genetically modified host cells disclosed herein are grown in minimal medium. As used herein, the terms "minimal medium" or "minimal media" may refer to growth medium containing the minimum nutrients possible for cell growth, generally, but not always, without the presence of one or more amino acids (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more amino acids). Minimal medium typically contains: (1) a carbon source for cellular (e.g. bacterial or yeast) growth; (2) various salts, which can vary among cellular (e.g. bacterial or yeast) species and growing conditions; and (3) water.

[00710] In some embodiments, genetically modified host cells disclosed herein are grown in rich medium or rich media. In certain such embodiments, the rich medium or rich media comprises yeast extract peptone dextrose (YPD) media comprising water, 10 g/L yeast extract, 20 g/L Bacto peptone, and 20 g/L dextrose (glucose). In some embodiments, the rich medium or rich media comprises YP + 20 g/L galactose and 1 g/L glucose. In some embodiments, the rich medium or rich media further comprises a carboxylic acid (e.g., 1 mM olivetolic acid, 1 mM olivetolic acid derivative, 2 mM hexanoic acid, or 2 mM of a carboxylic acid other than hexanoic acid). In some embodiments, rich medium or rich media affords more rapid cell growth compared to minimal media or minimal medium.

[00711] Materials and methods suitable for the maintenance and growth of the recombinant cells of the disclosure are described herein, e.g., in the Examples section. Other materials and methods suitable for the maintenance and growth of cell (e.g. bacterial or yeast) cultures are well known in the art. Exemplary techniques can be found in International Publication No. WO2009/076676, U.S. Patent Application No. 12/335,071 (U.S. Publ. No. 2009/0203102), WO 2010/003007, US Publ. No. 2010/0048964, WO 2009/132220, US Publ. No. 2010/0003716, Manual of Methods for General Bacteriology Gerhardt et al, eds), American Society for Microbiology, Washington, D.C. (1994) or Brock in Biotechnology: A Textbook of Industrial Microbiology, Second Edition (1989) Sinauer Associates, Inc., Sunderland, MA.

[00712] Standard cell culture conditions can be used to culture the genetically modified host cells disclosed herein (see, for example, WO 2004/033646 and references cited therein). In some embodiments, cells are grown and maintained at an appropriate temperature, gas mixture, and pH (such as at about 20 °C to about 37 °C, at about 0.04% to about 84% CO_2, at about 0% to about 100% dissolved oxygen, and at a pH between about 2 to about 9). In some embodiments, genetically modified host cells disclosed herein are grown at about 34 °C in a suitable cell culture medium. In some embodiments, genetically modified host cells disclosed herein are grown at about 20 °C to about 37 °C in a suitable cell culture medium. In some embodiments, genetically modified host cells disclosed herein are grown at about 20 °C, about 21 °C, about 22 °C, about 23 °C, about 24 °C, about 25 °C, about 26 °C, about 27 °C, about 28 °C, about 29 °C, about 30 °C, about 31 °C, about 32 °C, about 33 °C, about 34 °C, about 35 °C, about 36 °C, or about 37 °C in a suitable cell culture medium. In some embodiments, the pH ranges for fermentation are between about pH 3.0 to about pH 9.0 (such as about pH 3.0, about pH 3.5, about pH 4.0, about pH 4.5, about pH 5.0, about pH 5.5, about pH 6.0, about pH 6.5, about pH 7.0, about pH 7.5, about pH 8.0, about

pH 8.5, about pH 6.0 to about pH 8.0 or about 6.5 to about 7.0). In some embodiments, the pH ranges for fermentation are between about pH 4.5 to about pH 5.5. In some embodiments, the pH ranges for fermentation are between about pH 4.0 to about pH 6.0. In some embodiments, the pH ranges for fermentation are between about pH 3.0 to about pH 6.0. In some embodiments, the pH ranges for fermentation are between about pH 3.0 to about pH 5.5. In some embodiments, the pH ranges for fermentation are between about pH 3.0 to about pH 5.0. In some embodiments, the dissolved oxygen is between about 0% to about 10%, about 0% to about 20%, about 0% to about 30%, about 0% to about 40%, about 0% to about 50%, about 0% to about 60%, about 0% to about 70%, about 0% to about 80%, about 0% to about 90%, about 5% to about 10%, about 5% to about 20%, about 5% to about 30%, about 5% to about 40%, about 5% to about 50%, about 5% to about 60%, about 5% to about 70%, about 5% to about 80%, about 5% to about 90%, about 10% to about 20%, about 10% to about 30%, about 10% to about 40% or about 10% to about 50%. In some embodiments, the CO_2 level is between about 0.04% to about 0.1% CO_2, about 0.04% to about 1% CO_2, about 0.04% to about 5% CO_2, about 0.04% to about 10% CO_2, about 0.04% to about 20% CO_2, about 0.04% to about 30% CO_2, about 0.04% to about 40% CO_2, about 0.04% to about 50% CO_2, about 0.04% to about 60% CO_2, about 0.04% to about 70% CO_2, about 0.1% to about 5% CO_2, about 0.1% to about 10% CO_2, about 0.1% to about 20% CO_2, about 0.1% to about 30% CO_2, about 0.1% to about 40% CO_2, about 0.1% to about 50% CO_2, about 1% to about 5% CO_2, about 1% to about 10% CO_2, about 1% to about 20% CO_2, about 1% to about 30% CO_2, about 1% to about 40% CO_2, about 1% to about 50% CO_2, about 5% to about 10% CO_2, about 10% to about 20% CO_2, about 10% to about 30% CO_2, about 10% to about 40% CO_2, about 10% to about 50% CO_2, about 10% to about 60% CO_2, about 10% to about 70% CO_2, about 10% to about 80% CO_2, about 50% to about 60% CO_2, about 50% to about 70% CO_2, or about 50% to about 80% CO_2. genetically modified host cells disclosed herein disclosed herein can be grown under aerobic, anoxic, microaerobic, or anaerobic conditions based on the requirements of the cells.

[00713] Standard culture conditions and modes of fermentation, such as batch, fed-batch, or continuous fermentation that can be used are described in International Publication No. WO 2009/076676, U.S. Patent Application No. 12/335,071 (U.S. Publ. No. 2009/0203102), WO 2010/003007, US Publ. No. 2010/0048964, WO 2009/132220, US Publ. No. 2010/0003716, the contents of each of which are incorporated by reference herein in their entireties. Batch and Fed- Batch fermentations are common and well known in the

2018256863 11 Apr 2024

art and examples can be found in Brock, Biotechnology: A Textbook of Industrial Microbiology, Second Edition (1989) Sinauer Associates, Inc.

Production and Recovery of Produced Cannabinoids, Cannabinoid Precursors, Cannabinoid Derivatives or Cannabinoid Precursor Derivatives

[00714] In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative by genetically modified host cells of the disclosure in a recoverable amount of from about 1 mg/L culture medium to about 1 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 1 mg/L culture medium to about 500 mg/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 1 mg/L culture medium to about 100 mg/L culture medium. For example, in some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 1 mg/L culture medium to about 5 mg/L culture medium, from about 5 mg/L culture medium to about 10 mg/L culture medium, from about 10 mg/L culture medium to about 25 mg/L culture medium, from about 25 mg/L culture medium to about 50 mg/L culture medium, from about 50 mg/L culture medium to about 75 mg/L culture medium, or from about 75 mg/L culture medium to about 100 mg/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 100 mg/L culture medium to about 150 mg/L culture medium, from about 150 mg/L culture medium to about 200 mg/L culture medium, from about 200 mg/L culture medium to about 250 mg/L culture medium, from about 250 mg/L culture medium to about 500 mg/L culture medium, from about 500 mg/L culture medium to about 750 mg/L culture medium, or from about 750 mg/L culture medium to about 1 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about from about 50 mg/L culture medium to about 100 mg/L culture medium, 50 mg/L culture medium to about 150

mg/L culture medium, from about 50 mg/L culture medium to about 200 mg/L culture medium, from about 50 mg/L culture medium to about 250 mg/L culture medium, from about 50 mg/L culture medium to about 500 mg/L culture medium, or from about 50 mg/L culture medium to about 750 mg/L culture medium.

[00715] In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 100 mg/L culture medium to about 500 mg/L culture medium, or more than 500 mg/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 500 mg/L culture medium to about 1 g/L culture medium, or more than 1 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 1 g/L culture medium to about 10 g/L culture medium, or more than 10 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 10 g/L culture medium to about 100 g/L culture medium, or more than 100 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 1 g/L culture medium to about 20 g/L culture medium, or more than 20 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 1 g/L culture medium to about 30 g/L culture medium, or more than 30 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 1 g/L culture medium to about 40 g/L culture medium, or more than 40 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 1 g/L culture medium to about 50 g/L culture medium, or more than 50 g/L culture medium. In some embodiments, a method of the present disclosure provides for

2018256863 11 Apr 2024

production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 1 g/L culture medium to about 60 g/L culture medium, or more than 60 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 1 g/L culture medium to about 70 g/L culture medium, or more than 70 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 1 g/L culture medium to about 80 g/L culture medium, or more than 80 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 1 g/L culture medium to about 90 g/L culture medium, or more than 90 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 10 g/L culture medium to about 20 g/L culture medium, or more than 20 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 10 g/L culture medium to about 30 g/L culture medium, or more than 30 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 10 g/L culture medium to about 40 g/L culture medium, or more than 40 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 10 g/L culture medium to about 50 g/L culture medium, or more than 50 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 10 g/L culture medium to about 60 g/L culture medium, or more than 60 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 10 g/L culture

medium to about 70 g/L culture medium, or more than 70 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 10 g/L culture medium to about 80 g/L culture medium, or more than 80 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 10 g/L culture medium to about 90 g/L culture medium, or more than 90 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 50 g/L culture medium to about 100 g/L culture medium, or more than 100 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 50 g/L culture medium to about 60 g/L culture medium, or more than 60 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 50 g/L culture medium to about 70 g/L culture medium, or more than 70 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 50 g/L culture medium to about 80 g/L culture medium, or more than 80 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 50 g/L culture medium to about 90 g/L culture medium, or more than 90 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 20 g/L culture medium to about 100 g/L culture medium, or more than 100 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 20 g/L culture medium to about 30 g/L culture medium, or more than 30 g/L culture medium. In some embodiments, a method of the present disclosure provides for

production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 20 g/L culture medium to about 40 g/L culture medium, or more than 40 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 20 g/L culture medium to about 50 g/L culture medium, or more than 50 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 20 g/L culture medium to about 60 g/L culture medium, or more than 60 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 20 g/L culture medium to about 70 g/L culture medium, or more than 70 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 20 g/L culture medium to about 80 g/L culture medium, or more than 80 g/L culture medium. In some embodiments, a method of the present disclosure provides for production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in a recoverable amount of from about 20 g/L culture medium to about 90 g/L culture medium, or more than 90 g/L culture medium.

[00716] In some embodiments, the genetically modified host cell disclosed herein is cultured in a liquid medium comprising a precursor acid to generate acyl-CoA compounds or acyl-CoA compound derivatives. Suitable precursor acids may include, but are not limited to, carboxylic acids.

[00717] In some embodiments, a method of producing a cannabinoid, a cannabinoid derivative, cannabinoid precursor, cannabinoid precursor derivative, or cannabinoid precursor derivative may involve culturing a genetically modified yeast cell of the present disclosure under conditions that favor production of the cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative; where the cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative is produced by the genetically modified yeast cell and is present in the culture medium (e.g., a liquid culture medium) in which the genetically modified yeast cell is cultured. In some embodiments, the culture medium in which the genetically modified yeast cell is cultured

comprises a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in an amount of from 1 ng/L to 1 g/L (e.g., from 1 ng/L to 50 ng/L, from 50 ng/L to 100 ng/L, from 100 ng/L to 500 ng/L, from 500 ng/L to 1 μg/L, from 1 μg/L to 50 μg/L, from 50 μg/L to 100 μg/L, from 100 μg/L to 500 μg/L, from 500 μg/L to 1 mg/L, from 1 mg/L to 50 mg/L, from 50 mg/L to 100 mg/L, from 100 mg/L to 500 mg/L, or from 500 mg/L to 1 g/L). In some embodiments, the culture medium in which the genetically modified yeast cell is cultured comprises a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative in an amount more than 1 g/L.

[00718] In some embodiments, a method of producing a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative may involve culturing a genetically modified yeast cell of the present disclosure under conditions that favor fermentation of a sugar, and under conditions that favor production of the cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative; wherein the cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative is produced by the genetically modified yeast cell and is present in alcohol produced by the genetically modified yeast cell. The present disclosure provides an alcoholic beverage produced by the genetically modified yeast cell, where the alcoholic beverage comprises the cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative produced by the genetically modified yeast cell. Alcoholic beverages may include beer, wine, and distilled alcoholic beverages. In some embodiments, an alcoholic beverage of the present disclosure comprises a cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative in an amount of from 1 ng/L to 1 g/L (e.g., from 1 ng/L to 50 ng/L, from 50 ng/L to 100 ng/L, from 100 ng/L to 500 ng/L, from 500 ng/L to 1 μg/L, from 1 μg/L to 50 μg/L, from 50 μg/L to 100 μg/L, from 100 μg/L to 500 μg/L, from 500 μg/L to 1 mg/L, from 1 mg/L to 50 mg/L, from 50 mg/L to 100 mg/L, from 100 mg/L to 500 mg/L, or from 500 mg/L to 1 g/L). In some embodiments, an alcoholic beverage of the present disclosure comprises a cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative in an amount more than 1 g/L

[00719] In some embodiments, a method of the present disclosure provides for increased production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative. In certain such embodiments, culturing of the genetically modified host cell disclosed herein in a suitable medium provides for synthesis of

the cannabinoid, the cannabinoid derivative, the cannabinoid precursor, or the cannabinoid precursor derivative in an increased amount compared to a non-genetically modified host cell cultured under similar conditions. The production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative by the genetically modified host cells disclosed herein may be increased by about 5% to about 1,000,000 folds compared to a non-genetically modified host cell cultured under similar conditions. The production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative by the genetically modified host cells disclosed herein may be increased by about 10% to about 1,000,000 folds (e.g., about 50% to about 1,000,000 folds, about 1 to about 500,000 folds, about 1 to about 50,000 folds, about 1 to about 5,000 folds, about 1 to about 1,000 folds, about 1 to about 500 folds, about 1 to about 100 folds, about 1 to about 50 folds, about 5 to about 100,000 folds, about 5 to about 10,000 folds, about 5 to about 1,000 folds, about 5 to about 500 folds, about 5 to about 100 folds, about 10 to about 50,000 folds, about 50 to about 10,000 folds, about 100 to about 5,000 folds, about 200 to about 1,000 folds, about 50 to about 500 folds, or about 50 to about 200 folds) compared to the production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative by non-genetically modified host cells cultured under similar conditions. The production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative by genetically modified host cells disclosed herein may also be increased by at least about any of 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 1 fold, 2 folds, 5 folds, 10 folds, 20 folds, 50 folds, 100 folds, 200 folds, 500 folds, 1000 folds, 2000 folds, 5000 folds, 10,000 folds, 20,000 folds, 50,000 folds, 100,000 folds, 200,000 folds, 500,000 folds, or 1,000,000 folds compared to the production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative by non-genetically modified host cells cultured under similar conditions.

[00720] In some embodiments, the production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative by genetically modified host cells of the disclosure may also be increased by at least about any of 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% compared to the production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative by non-genetically modified host cells cultured under similar conditions. In some embodiments, the production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative by genetically modified host cells disclosed

herein may also be increased by at least about any of 1-20%, 2-20%, 5-20%, 10-20%, 15-20%, 1-15%, 1-10%, 2-15%, 2-10%, 5-15%, 10-15%, 1-50%, 10-50%, 20-50%, 30-50%, 40-50%, 50-100%, 50-60%, 50-70%, 50-80%, or 50-90% compared to the production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative by non-genetically modified host cells cultured under similar conditions.

[00721] In some embodiments, production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative by genetically modified host cells of the disclosure is determined by LC-MS analysis. In certain such embodiments, each cannabinoid, cannabinoid precursor, cannabinoid derivative, or cannabinoid precursor derivative is identified by retention time, determined from an authentic standard, and multiple reaction monitoring (MRM) transition.

[00722] In some embodiments, the genetically modified host cell of the disclosure is yeast cell. In certain such embodiments, the genetically modified host cell disclosed herein is cultured in a bioreactor. In some embodiments, the genetically modified host cell is cultured in a suitable medium supplemented with hexanoic acid, a carboxylic acid other than hexanoic acid, olivetolic acid, or an olivetolic acid derivative.

[00723] In some embodiments, the cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative is recovered from a cell lysate, e.g., by lysing the genetically modified host cell disclosed herein and recovering the cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative from the lysate. In other cases, the cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative is recovered from the culture medium in which the genetically modified host cell disclosed herein is cultured. In other cases, the cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative is recovered from both the cell lysate and the culture medium.

[00724] In some embodiments, the recovered cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative is then purified. In some embodiments, whole-cell broth from cultures comprising genetically modified host cells of the disclosure may be extracted with a suitable organic solvent to afford cannabinoids, cannabinoid precursors, cannabinoid derivatives, or cannabinoid precursor derivatives. Suitable organic solvents include, but are not limited to, hexane, heptane, ethyl acetate, petroleum ether, and di-ethyl ether, chloroform, and ethyl acetate. In some embodiments, the suitable organic solvent comprises hexane. In some embodiments, the suitable organic solvent may be added to the whole-cell broth from fermentations comprising genetically

modified host cells of the disclosure at a 10:1 ratio (10 parts whole-cell broth – 1 part organic solvent) and stirred for 30 minutes. In certain such embodiments, the organic fraction may be separated and extracted twice with an equal volume of acidic water (pH 2.5). The organic layer may then be separated and dried in a concentrator (rotary evaporator or thin film evaporator under reduced pressure) to obtain crude cannabinoid, cannabinoid precursor, cannabinoid derivative, or cannabinoid precursor derivative crystals. In certain such embodiments, the crude crystals may be heated to 105 °C for 15 minutes followed by 145 °C for 55 minutes to decarboxylate a crude cannabinoid or cannabinoid derivative. In certain such embodiments, the crude crystalline product may be re-dissolved and recrystallized in a suitable solvent (e.g., n-pentane) and filtered to remove any insoluble material. In certain such embodiments, the solvent may then be removed e.g. by rotary evaporation, to produce pure crystalline product.

[00725] In some embodiments, the cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative is pure, e.g., at least about 40% pure, at least about 50% pure, at least about 60% pure, at least about 70% pure, at least about 80% pure, at least about 90% pure, at least about 95% pure, at least about 98%, or more than 98% pure, where "pure" in the context of a cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative may refer to a cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative that is free from other cannabinoids, cannabinoid derivatives, cannabinoid precursors, cannabinoid precursor derivatives macromolecules, contaminants, etc.

Cell-Free Methods of Generating Cannabinoids, Cannabinoid Precursors, Olivetolic Acid Derivatives, Olivetolic Acid, Cannabinoid Derivatives, or Cannabinoid Precursor Derivatives

[00726] The methods of the disclosure may involve cell-free production of cannabinoids, cannabinoid precursors, cannabinoid precursor derivatives, or cannabinoid derivatives using one or more polypeptides disclosed herein expressed or overexpressed by a genetically modified host cell of the disclosure. In some embodiments, one or more polypeptides disclosed herein expressed or overexpressed by a genetically modified host cell of the disclosure are used in a cell-free system for the production of cannabinoids, cannabinoid precursors, cannabinoid precursor derivatives, or cannabinoid derivatives. In certain such embodiments, appropriate starting materials for use in producing cannabinoids, cannabinoid precursors, cannabinoid precursor derivatives, or cannabinoid derivatives may

be mixed together with one or more polypeptides disclosed herein expressed or overexpressed by a genetically modified host cell of the disclosure in a suitable reaction vessel to effect the reaction. The one or more polypeptides disclosed herein expressed or overexpressed by a genetically modified host cell of the disclosure may be used in combination to effect a complete synthesis of a cannabinoid, cannabinoid precursor, cannabinoid precursor derivative, or cannabinoid derivative from the appropriate starting materials. In some embodiments, the cannabinoid, cannabinoid precursor, the cannabinoid precursor derivative, or cannabinoid derivative is recovered from a cell-free reaction mixture comprising one or more of the polypeptides disclosed herein.

[00727] In some embodiments, the recovered cannabinoids, cannabinoid precursors, cannabinoid precursor derivatives, or cannabinoid derivatives are then purified. In certain such embodiments, a cell-free reaction mixture comprising one or more of the polypeptides disclosed herein may be extracted with a suitable organic solvent to afford cannabinoids, cannabinoid precursors, cannabinoid derivatives, or cannabinoid precursor derivatives. Suitable organic solvents include, but are not limited to, hexane, heptane, ethyl acetate, petroleum ether, and di-ethyl ether, chloroform, and ethyl acetate. In some embodiments, the suitable organic solvent comprises hexane. In some embodiments, the suitable organic solvent may be added to the cell-free reaction mixture comprising one or more of the polypeptides disclosed herein at a 10:1 ratio (10 parts reaction mixture – 1 part organic solvent) and stirred for 30 minutes. In certain such embodiments, the organic fraction may be separated and extracted twice with an equal volume of acidic water (pH 2.5). The organic layer may then be separated and dried in a concentrator (rotary evaporator or thin film evaporator under reduced pressure) to obtain crude cannabinoid, cannabinoid precursor, cannabinoid derivative, or cannabinoid precursor derivative crystals. In certain such embodiments, the crude crystals may be heated to 105 °C for 15 minutes followed by 145 °C for 55 minutes to decarboxylate a crude cannabinoid or cannabinoid derivative. In certain such embodiments, the crude crystalline product may be re-dissolved and recrystallized in a suitable solvent (e.g., n-pentane) and filtered to remove any insoluble material. In certain such embodiments, the solvent may then be removed e.g. by rotary evaporation, to produce pure crystalline product.

[00728] In some embodiments, a prenyl group acceptor molecule, a prenyl group donor molecule and a GOT polypeptide may be mixed together in a suitable reaction vessel to effect the reaction. In certain such embodiments, the GOT polypeptide can catalyze production of cannabigerolic acid from GPP and olivetolic acid in an amount at least ten

times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82. In some embodiments, the GOT polypeptide comprises an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110. In some embodiments, the prenyl group acceptor molecule is olivetolic acid or a derivative thereof. In some embodiments, the prenyl group donor molecule is GPP or a derivative thereof. In some embodiments, the reaction produces cannabigerolic acid or a derivative thereof.

[00729] In some embodiments, cell-free production of a cannabinoid, a cannabinoid derivative, a cannabinoid precursor, or a cannabinoid precursor derivative by one or more polypeptides disclosed herein expressed or overexpressed by a genetically modified host cell of the disclosure is determined by LC-MS analysis. In certain such embodiments, each cannabinoid, cannabinoid precursor, cannabinoid derivative, or cannabinoid precursor derivative is identified by retention time, determined from an authentic standard, and multiple reaction monitoring (MRM) transition.

EXAMPLES OF NON-LIMITING EMBODIMENTS OF THE DISCLOSURE

[00730] Embodiments, of the present subject matter disclosed herein may be beneficial alone or in combination, with one or more other embodiments. Without limiting the foregoing description, certain non-limiting embodiments of the disclosure, numbered I-1 to I-54, II-1 to II-55, and III-1 to III-81 are provided below. As will be apparent to those of skill in the art upon reading this disclosure, each of the individually numbered embodiments may be used or combined with any of the preceding or following individually numbered embodiments. This is intended to provide support for all such combinations of embodiments and is not limited to combinations of embodiments explicitly provided below. Some embodiments of the disclosure are of Embodiment I:

[00731] Embodiment I-1. A genetically modified host cell that produces a cannabinoid compound or a cannabinoid precursor, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode one or more polypeptides that generate hexanoyl-CoA or derivatives of hexanoyl-CoA; b) one or more heterologous nucleic acids that encode one or more polypeptides that generate geranyl pyrophosphate; c) one or more heterologous nucleic acids that encode one or more polypeptides that generate malonyl-CoA; d) one or more heterologous nucleic acids that

2018256863 11 Apr 2024

encode a fusion TKS/OAC polypeptide that condenses hexanoyl-CoA or its derivatives and malonyl-CoA to generate olivetolic acid or derivatives of olivetolic acid; e) one or more heterologous nucleic acids that encode a truncated geranyl pyrophosphate:olivetolic acid geranyltransferase (GOT) polypeptide or an NphB polypeptide; and f) one or more heterologous nucleic acids that encode a cannabinoid synthase polypeptide, wherein culturing of the genetically modified host cell in a suitable medium provides for synthesis of the cannabinoid compound or the cannabinoid precursor in a recoverable amount.

[00732] Embodiment I-2. The genetically modified host cell of Embodiment I-1, wherein the host cell is a eukaryotic cell.

[00733] Embodiment I-3. The genetically modified host cell of Embodiment I-2, wherein the host cell is a yeast cell.

[00734] Embodiment I-4. The genetically modified host cell of Embodiment I-3, wherein the host cell is *Saccharomyces cerevisiae*.

[00735] Embodiment I-5. The genetically modified host cell of Embodiment I-4, wherein the host cell is a protease-deficient strain of *Saccharomyces cerevisiae*.

[00736] Embodiment I-6. The genetically modified host cell of Embodiment I-2, wherein the host cell is a plant cell.

[00737] Embodiment I-7. The genetically modified host cell of Embodiment I-1, wherein the host cell is a prokaryotic cell.

[00738] Embodiment I-8. The genetically modified host cell of Embodiment I-1, wherein the one or more polypeptides that generate hexanoyl-CoA or a hexanoyl-CoA derivative is hexanoyl-CoA synthetase (HCS) polypeptide, and wherein the medium comprises hexanoate.

[00739] Embodiment I-9. The genetically modified host cell of Embodiment I-6, wherein the HCS polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, or SEQ ID NO:4.

[00740] Embodiment I-10. The genetically modified host cell of Embodiment I-1, wherein the one or more polypeptides that generate geranyl pyrophosphate comprise geranyl pyrophosphate synthetase (GPPS) polypeptide.

[00741] Embodiment I-11. The genetically modified host cell of Embodiment I-11, wherein the GPPS polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at

2018256863 11 Apr 2024

least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to one of the amino acid sequences set forth in SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, or SEQ ID NO:8.

[00742] Embodiment I-12. The genetically modified host cell of Embodiment I-10, wherein the GPPS polypeptide is a dominant negative variant that reduces the ability of an endogenous GPPS polypeptide to function as a farnesyl pyrophosphate synthetase (FPPS) polypeptide.

[00743] Embodiment I-13. The genetically modified host cell of Embodiment I-10, wherein the GPPS polypeptide comprises a K197G amino acid substitution.

[00744] Embodiment I-14. The genetically modified host cell of Embodiment I-10, wherein the GPPS polypeptide is a heterodimeric protein comprising a GPPS large subunit polypeptide and a GPPS small subunit polypeptide, or a homodimeric or monomeric GPPS polypeptide.

[00745] Embodiment I-15. The genetically modified host cell of Embodiment I-1, wherein the one or more polypeptides that generate malonyl-CoA comprises acetyl-CoA carboxylase-1 (ACC1) polypeptide.

[00746] Embodiment I-16. The genetically modified host cell of Embodiment I-15, wherein the ACC1 polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:9.

[00747] Embodiment I-17. The genetically modified host cell of Embodiment I-1, wherein the one or more polypeptides that generate hexanoyl-CoA comprise an MCT1 polypeptide, a PaaH1 polypeptide, a Crt polypeptide, a Ter polypeptide, and a BktB polypeptide.

[00748] Embodiment I-18. The genetically modified host cell of Embodiment I-1, wherein the one or more polypeptides that generate hexanoyl-CoA comprise a MCT1 polypeptide, a PhaB polypeptide, a PhaJ polypeptide, a Ter polypeptide, and a BktB polypeptide.

[00749] Embodiment I-19. The genetically modified host cell of Embodiment I-17, wherein: i) the PaaH1 polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:18 or SEQ ID NO:46; ii) the Crt polypeptide

comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:19 or SEQ ID NO:48; iii) the Ter polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:20 or SEQ ID NO:50; and iv) the BktB polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:21 or SEQ ID NO:44.

[00750] Embodiment I-20. The genetically modified host cell of Embodiment I-18, wherein: i) the PhaB polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:94; ii) the PhaJ polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:96; iii) the Ter polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:20 or SEQ ID NO:50; and iv) the BktB polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:21 or SEQ ID NO:44.

[00751] Embodiment I-21. The genetically modified host cell of Embodiment I-1, wherein the host cell is genetically modified with a heterologous nucleic acid encoding one or more polypeptides that modulate NADH redox balance.

[00752] Embodiment I-22. The genetically modified host cell of Embodiment I-1, wherein the host cell is genetically modified with one or more of: i) one or more heterologous nucleic acids encoding a HMG-CoA synthase polypeptide; ii) one or more heterologous nucleic acids encoding a 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMGR) polypeptide; iii) one or more heterologous nucleic acids encoding an MK polypeptide; and

iv) one or more heterologous nucleic acids encoding an isopentenyl diphosphate isomerase (IDI) polypeptide.

[00753] Embodiment I-23. The genetically modified host cell of Embodiment I-22, wherein the host cell is genetically modified to overexpress a heterologous IDI polypeptide.

[00754] Embodiment I-24. The genetically modified host cell of Embodiment I-23, wherein the IDI polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:58.

[00755] Embodiment I-25. The genetically modified host cell of Embodiment I-22, wherein the host cell is genetically modified to overexpress a truncated HMGR (tHMGR) polypeptide.

[00756] Embodiment I-26. The genetically modified host cell of Embodiment I-25, wherein the tHMGR polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:17.

[00757] Embodiment I-27. The genetically modified host cell of Embodiment I-22, wherein the HMGR polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:22.

[00758] Embodiment I-28. The genetically modified host cell of Embodiment I-22, wherein the HMGS polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the MvaS polypeptide amino acid sequence set forth in SEQ ID NO:23, or wherein the HMGS polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the ERG13 polypeptide amino acid sequence set forth in SEQ ID NO:24.

[00759] Embodiment I-29. The genetically modified host cell of Embodiment I-22, wherein the MK polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least

2018256863 11 Apr 2024

95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the ERG12 polypeptide amino acid sequence set forth in SEQ ID NO:64.

[00760] Embodiment I-30. The genetically modified host cell of Embodiment I-1, wherein the host cell is genetically modified with a heterologous nucleic acid encoding one or more polypeptides that condense two molecules of acetyl-CoA to generate acetoacetyl-CoA.

[00761] Embodiment I-31. The genetically modified host cell of Embodiment I-30, wherein the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide.

[00762] Embodiment I-32. The genetically modified host cell of Embodiment I-31, wherein the acetoacetyl-CoA thiolase polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:25.

[00763] Embodiment I-33. The genetically modified host cell of Embodiment I-1, wherein the host cell is genetically modified with a heterologous nucleic acid encoding one or more polypeptides that condense one molecule of acetyl-CoA and one molecule of malonyl-CoA to generate acetoacetyl-CoA.

[00764] Embodiment I-34. The genetically modified host cell of any one of Embodiments I-1 to I-33, wherein at least one of the one or more heterologous nucleic acids is integrated into the chromosome of the host cell.

[00765] Embodiment I-35. The genetically modified host cell of any one of Embodiments I-1 to I-33, wherein at least one of the one or more heterologous nucleic acids is maintained extrachromosomally.

[00766] Embodiment I-36. The genetically modified host cell of any one of Embodiments I-1 to I-33, wherein two or more of the one or more heterologous nucleic acids are present in a single expression vector.

[00767] Embodiment I-37. The genetically modified host cell of any one of Embodiments I-1 to I-36, wherein the cannabinoid compound is cannabigerolic acid.

[00768] Embodiment I-38. The genetically modified host cell of Embodiment I-1, wherein the genetically modified host cell is genetically modified with a heterologous nucleic acid encoding a Δ^9-THCA synthase polypeptide.

[00769] Embodiment I-39. The genetically modified host cell of Embodiment I-38, wherein the THCA synthase polypeptide comprises an amino acid having at least 50%

2018256863 11 Apr 2024

(at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to SEQ ID NO:14 or SEQ ID NO:15.

[00770] Embodiment I-40. The genetically modified host cell of Embodiment I-1, wherein the genetically modified host cell is genetically modified with a heterologous nucleic acid encoding a CBDA synthase polypeptide.

[00771] Embodiment I-41. The genetically modified host cell of Embodiment I-40, wherein the CBDA synthase polypeptide comprises an amino acid having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to SEQ ID NO:88 or SEQ ID NO:16.

[00772] Embodiment I-42. The genetically modified host cell of any one of Embodiments I-1 to I-41, wherein at least one of the heterologous nucleic acids is operably linked to an inducible promoter.

[00773] Embodiment I-43. The genetically modified host cell of any one of Embodiments I-1 to I-41, wherein at least one of the heterologous nucleic acids is operably linked to a constitutive promoter.

[00774] Embodiment I-44. The genetically modified host cell of Embodiment I-1, wherein the cannabinoid compound is cannabichromenic acid, cannabigerolic acid, Δ^9-tetrahydrocannabinolic acid, cannabidiolic acid, Δ^9-tetrahydrocannabinol, cannabidiol, or cannabichromene.

[00775] Embodiment I-45. A method of synthesizing a cannabinoid compound or cannabinoid precursor in a host cell, the method comprising: a) culturing a host cell of any one of Embodiments I-1 to I-40 in a suitable medium; and b) recovering the produced cannabinoid compound or cannabinoid precursor.

[00776] Embodiment I-46. The method of Embodiment I-45, wherein the medium comprises a fermentable sugar.

[00777] Embodiment I-47. The method of Embodiment I-45, wherein the medium comprises a pretreated cellulosic feedstock.

[00778] Embodiment I-48. A genetically modified host cell that produces an olivetolic acid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode an acyl-activating enzyme (AAE) polypeptide; and b) one or more heterologous nucleic acids that encode a TKS/OAC fusion polypeptide.

2018256863 11 Apr 2024

[00779] Embodiment I-49. The genetically modified host cell of Embodiment I-48, wherein the AAE polypeptide comprises an amino acid sequence having at least 50% (at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity to an AAE amino acid sequence set forth in SEQ ID NO:90 or SEQ ID NO:91.

[00780] Embodiment I-50. A genetically modified host cell that produces a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode an acyl-activating enzyme (AAE) polypeptide; b) one or more heterologous nucleic acids that encode a TKS/OAC fusion polypeptide; and c) one or more heterologous nucleic acids that encode a GOT polypeptide or an NphB polypeptide.

[00781] Embodiment I-51. The genetically modified host cell of Embodiment I-50, wherein the genetically modified host cell is genetically modified with a heterologous nucleic acid encoding a THCAS polypeptide.

[00782] Embodiment I-52. The genetically modified host cell of Embodiment I-50, wherein the genetically modified host cell is genetically modified with a heterologous nucleic acid encoding a CBDAS polypeptide.

[00783] Embodiment I-53. A method of producing an olivetolic acid derivative, the method comprising culturing the genetically modified host cell of Embodiment I-48 or Embodiment I-49 in a culture medium comprising a carboxylic acid.

[00784] Embodiment I-54. A method of producing a cannabinoid derivative, the method comprising culturing the genetically modified host cell of any one of Embodiments I-50 to I-52 in a culture medium comprising a carboxylic acid.

[00785] Some embodiments of the disclosure are of Embodiment II:

[00786] Embodiment II-1. A genetically modified host cell that produces a cannabinoid compound, a cannabinoid derivative, or a cannabinoid precursor, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode one or more polypeptides that generate hexanoyl-CoA or derivatives of hexanoyl-CoA; b) one or more heterologous nucleic acids that encode one or more polypeptides that generate geranyl pyrophosphate; c) one or more heterologous nucleic acids that encode one or more polypeptides that generate malonyl-CoA; d) one or more heterologous nucleic acids that encode a TKS polypeptide and an OAC polypeptide, or a fusion TKS and OAC polypeptide, that converts hexanoyl-CoA or its derivatives and malonyl-CoA to olivetolic acid or derivatives of olivetolic acid; e) one or more heterologous

2018256863 11 Apr 2024

nucleic acids that encode a geranyl pyrophosphate:olivetolic acid geranyltransferase (GOT) polypeptide or an NphB polypeptide; and f) one or more heterologous nucleic acids that encode a cannabinoid synthase polypeptide, wherein culturing of the genetically modified host cell in a suitable medium provides for synthesis of the cannabinoid compound, cannabinoid derivative, or the cannabinoid precursor in a recoverable amount.

[00787] Embodiment II-2. The genetically modified host cell of Embodiment II-1, wherein the host cell is a eukaryotic cell.

[00788] Embodiment II-3. The genetically modified host cell of Embodiment II-2, wherein the host cell is a yeast cell.

[00789] Embodiment II-4. The genetically modified host cell of Embodiment II-3, wherein the host cell is *Saccharomyces cerevisiae*.

[00790] Embodiment II-5. The genetically modified host cell of Embodiment II-4, wherein the host cell is a protease-deficient strain of *Saccharomyces cerevisiae*.

[00791] Embodiment II-6. The genetically modified host cell of Embodiment II-2, wherein the host cell is a plant cell.

[00792] Embodiment II-7. The genetically modified host cell of Embodiment II-1, wherein the host cell is a prokaryotic cell.

[00793] Embodiment II-8. The genetically modified host cell of Embodiment II-1, wherein the one or more polypeptides that generate hexanoyl-CoA or a hexanoyl-CoA derivative is a hexanoyl-CoA synthetase (HCS) polypeptide, and wherein the medium comprises hexanoate.

[00794] Embodiment II-9. The genetically modified host cell of Embodiment II-8, wherein the HCS polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, or SEQ ID NO:4.

[00795] Embodiment II-10. The genetically modified host cell of Embodiment II-1, wherein the one or more polypeptides that generate geranyl pyrophosphate comprise a geranyl pyrophosphate synthetase (GPPS) polypeptide.

[00796] Embodiment II-11. The genetically modified host cell of Embodiment II-11, wherein the GPPS polypeptide comprises an amino acid having at least 50% amino acid sequence identity to one of the amino acid sequences set forth in SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, or SEQ ID NO:8.

[00797] Embodiment II-12. The genetically modified host cell of Embodiment II-10, wherein the GPPS polypeptide is a dominant negative variant that reduces the ability of

2018256863 11 Apr 2024

an endogenous GPPS polypeptide to function as a farnesyl pyrophosphate synthetase (FPPS) polypeptide.

[00798] Embodiment II-13. The genetically modified host cell of Embodiment II-10, wherein the GPPS polypeptide comprises a K197G amino acid substitution.

[00799] Embodiment II-14. The genetically modified host cell of Embodiment II-10, wherein the GPPS polypeptide is a heterodimeric protein comprising a GPPS large subunit polypeptide and a GPPS small subunit polypeptide, or a homodimeric or monomeric GPPS polypeptide.

[00800] Embodiment II-15. The genetically modified host cell of Embodiment II-1, wherein the one or more polypeptides that generate malonyl-CoA comprises an acetyl-CoA carboxylase-1 (ACC1) polypeptide.

[00801] Embodiment II-16. The genetically modified host cell of Embodiment II-15, wherein the ACC1 polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:9.

[00802] Embodiment II-17. The genetically modified host cell of Embodiment II-1, wherein the one or more polypeptides that generate hexanoyl-CoA comprise a MCT1 polypeptide, a PaaH1 polypeptide, a Crt polypeptide, a Ter polypeptide, and a BktB polypeptide.

[00803] Embodiment II-18. The genetically modified host cell of Embodiment II-1, wherein the one or more polypeptides that generate hexanoyl-CoA comprise a MCT1 polypeptide, a PhaB polypeptide, a PhaJ polypeptide, a Ter polypeptide, and a BktB polypeptide.

[00804] Embodiment II-19 The genetically modified host cell of Embodiment II-17, wherein: i) the PaaH1 polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:18 or SEQ ID NO:46; ii) the Crt polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:19 or SEQ ID NO:48; iii) the Ter polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:20 or SEQ ID NO:50; and iv) the BktB polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:21 or SEQ ID NO:44.

[00805] Embodiment II-20. The genetically modified host cell of Embodiment II-18, wherein: i) the PhaB polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:94; ii) the PhaJ

polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:96; iii) the Ter polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:20 or SEQ ID NO:50; and iv) the BktB polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:21 or SEQ ID NO:44.

[00806] Embodiment II-21. The genetically modified host cell of Embodiment II-1, wherein the host cell is genetically modified with a heterologous nucleic acid encoding one or more polypeptides that modulate NADH redox balance.

[00807] Embodiment II-22. The genetically modified host cell of Embodiment II-1, wherein the host cell is genetically modified with one or more of: i) one or more heterologous nucleic acids that encode an HMG-CoA synthase polypeptide; ii) one or more heterologous nucleic acids that encode a 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMGR) polypeptide; iii) one or more heterologous nucleic acids that encode an MK polypeptide; and iv) one or more heterologous nucleic acids that encode an isopentenyl diphosphate isomerase (IDI) polypeptide.

[00808] Embodiment II-23. The genetically modified host cell of Embodiment II-22, wherein the host cell is genetically modified to overexpress a heterologous IDI polypeptide.

[00809] Embodiment II-24. The genetically modified host cell of Embodiment II-23, wherein the IDI polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:58.

[00810] Embodiment II-25. The genetically modified host cell of Embodiment II-22, wherein the host cell is genetically modified to overexpress a truncated HMGR (tHMGR) polypeptide.

[00811] Embodiment II-26. The genetically modified host cell of Embodiment II-25, wherein the tHMGR polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:17.

[00812] Embodiment II-27. The genetically modified host cell of Embodiment II-22, wherein the HMGR polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:22.

[00813] Embodiment II-28. The genetically modified host cell of Embodiment II-22, wherein the HMGS polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the MvaS polypeptide amino acid sequence set forth in SEQ ID NO:23,

2018256863 11 Apr 2024

or wherein the HMGS polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the ERG13 polypeptide amino acid sequence set forth in SEQ ID NO:24.

[00814] Embodiment II-29. The genetically modified host cell of Embodiment II-22, wherein the MK polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the ERG12 polypeptide amino acid sequence set forth in SEQ ID NO:64.

[00815] Embodiment II-30. The genetically modified host cell of Embodiment II-1, wherein the host cell is genetically modified with a heterologous nucleic acid encoding one or more polypeptides that condense two molecules of acetyl-CoA to generate acetoacetyl-CoA.

[00816] Embodiment II-31. The genetically modified host cell of Embodiment II-30, wherein the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide.

[00817] Embodiment II-32. The genetically modified host cell of Embodiment II-31, wherein the acetoacetyl-CoA thiolase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25.

[00818] Embodiment II-33. The genetically modified host cell of Embodiment II-1, wherein the host cell is genetically modified with a heterologous nucleic acid encoding one or more polypeptides that condense one molecule of acetyl-CoA and one molecule of malonyl-CoA to generate acetoacetyl-CoA.

[00819] Embodiment II-34. The genetically modified host cell of any one of Embodiments II-1 to II-33, wherein at least one of the one or more heterologous nucleic acids is integrated into the chromosome of the host cell.

[00820] Embodiment II-35. The genetically modified host cell of any one of Embodiments II-1 to II-33, wherein at least one of the one or more heterologous nucleic acids is maintained extrachromosomally.

[00821] Embodiment II-36. The genetically modified host cell of any one of Embodiments II-1 to II-33, wherein two or more of the one or more heterologous nucleic acids are present in a single expression vector.

[00822] Embodiment II-37. The genetically modified host cell of any one of Embodiments II-1 to II-36, wherein the cannabinoid compound is cannabigerolic acid.

[00823] Embodiment II-38. The genetically modified host cell of Embodiment II-1, wherein the genetically modified host cell is genetically modified with a heterologous nucleic acid encoding a Δ^9-THCA synthase polypeptide.

[00824] Embodiment II-39. The genetically modified host cell of Embodiment II-38, wherein the THCA synthase polypeptide comprises an amino acid having at least 50% amino acid sequence identity to SEQ ID NO:14 or SEQ ID NO:15.

[00825] Embodiment II-40. The genetically modified host cell of Embodiment II-1, wherein the genetically modified host cell is genetically modified with a heterologous nucleic acid encoding a CBDA synthase polypeptide.

[00826] Embodiment II-41. The genetically modified host cell of Embodiment II-40, wherein the CBDA synthase polypeptide comprises an amino acid having at least 50% amino acid sequence identity to SEQ ID NO:88 or SEQ ID NO:16.

[00827] Embodiment II-42. The genetically modified host cell of any one of Embodiments II-1 to II-41, wherein at least one of the heterologous nucleic acids is operably linked to an inducible promoter.

[00828] Embodiment II-43. The genetically modified host cell of any one of Embodiments II-1 to II-41, wherein at least one of the heterologous nucleic acids is operably linked to a constitutive promoter.

[00829] Embodiment II-44. The genetically modified host cell of Embodiment II-1, wherein the cannabinoid compound is cannabichromenic acid, cannabigerolic acid, Δ^9-tetrahydrocannabinolic acid, cannabidiolic acid, Δ^9-tetrahydrocannabinol, cannabidiol, or cannabichromene.

[00830] Embodiment II-45. A method of synthesizing a cannabinoid compound or cannabinoid precursor in a host cell, the method comprising: a) culturing a host cell of any one of Embodiments II-1 to II-40 in a suitable medium; and b) recovering the produced cannabinoid compound or cannabinoid precursor.

[00831] Embodiment II-46. The method of Embodiment II-45, wherein the medium comprises a fermentable sugar.

[00832] Embodiment II-47. The method of Embodiment II-45, wherein the medium comprises a pretreated cellulosic feedstock.

[00833] Embodiment II-48. A genetically modified host cell that produces an olivetolic acid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode an acyl-activating enzyme

2018256863 11 Apr 2024

(AAE) polypeptide; and b) one or more heterologous nucleic acids that encode a TKS/OAC fusion polypeptide.

[00834] Embodiment II-49. The genetically modified host cell of Embodiment II-48, wherein the AAE polypeptide comprises an amino acid sequence having at least 50% amino acid sequence identity to an AAE amino acid sequence set forth in SEQ ID NO:90 or SEQ ID NO:91.

[00835] Embodiment II-50. A genetically modified host cell that produces a cannabinoid derivative, wherein the genetically modified host cell is genetically modified with: a) one or more heterologous nucleic acids that encode an acyl-activating enzyme (AAE) polypeptide; b) one or more heterologous nucleic acids that encode a TKS/OAC fusion polypeptide; and c) one or more heterologous nucleic acids that encode a GOT polypeptide or an NphB polypeptide.

[00836] Embodiment II-51. The genetically modified host cell of Embodiment II-50, wherein the genetically modified host cell is genetically modified with a heterologous nucleic acid encoding a THCAS polypeptide.

[00837] Embodiment II-52. The genetically modified host cell of Embodiment II-50, wherein the genetically modified host cell is genetically modified with a heterologous nucleic acid encoding a CBDAS polypeptide.

[00838] Embodiment II-53. A method of producing an olivetolic acid derivative, the method comprising culturing the genetically modified host cell of Embodiment II-48 or II-49 in a culture medium comprising a carboxylic acid.

[00839] Embodiment II-54. A method of producing a cannabinoid derivative, the method comprising culturing the genetically modified host cell of any one of Embodiments II-50 to II-52 in a culture medium comprising a carboxylic acid.

[00840] Embodiment II-55. The genetically modified host cell of Embodiment II-1, wherein the geranyl pyrophosphate olivetolic acid geranyltransferase (GOT) polypeptide or the NphB polypeptide comprises an amino acid having at least 50% amino acid sequence identity to the amino acid sequence of a GOT polypeptide, including CsPT4t polypeptide, or NphB polypeptide sequence, respectively, disclosed herein.

[00841] Some embodiments of the disclosure are of Embodiment III:

[00842] Embodiment III-1. A genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, the genetically modified host cell comprising one or more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide, wherein said geranyl pyrophosphate:olivetolic acid

geranyltransferase polypeptide can catalyze production of cannabigerolic acid from geranyl pyrophosphate and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[00843] Embodiment III-2. A genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, the genetically modified host cell comprising one or more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110.

[00844] Embodiment III-3. A genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, the genetically modified host cell comprising one or more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100.

[00845] Embodiment III-4. The genetically modified host cell of any one of Embodiments III-1 to III-3, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a tetraketide synthase (TKS) polypeptide and one or more heterologous nucleic acids encoding an olivetolic acid cyclase (OAC) polypeptide, or one or more heterologous nucleic acids encoding a fusion TKS and OAC polypeptide.

[00846] Embodiment III-5. The genetically modified host cell of Embodiment III-4, wherein the TKS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:11 or SEQ ID NO:76.

[00847] Embodiment III-6. The genetically modified host cell of Embodiment III-4 or III-5, wherein the OAC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:10 or SEQ ID NO:78.

[00848] Embodiment III-7. The genetically modified host cell of any one of Embodiments III-1 to III-6, wherein the genetically modified host cell further comprises one or more of the following: a) one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; b) one or more heterologous nucleic acids encoding a polypeptide that generates geranyl pyrophosphate; or c) one or more heterologous nucleic acids encoding a polypeptide that generates malonyl-CoA.

[00849] Embodiment III-8. The genetically modified host cell of Embodiment III-7, wherein the genetically modified host cell further comprises one or more heterologous

2018256863 11 Apr 2024

nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative, wherein the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is an acyl-activating enzyme (AAE) polypeptide.

[00850] Embodiment III-9. The genetically modified host cell of Embodiment III-8, wherein the AAE polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90.

[00851] Embodiment III-10. The genetically modified host cell of Embodiment III-8, wherein the AAE polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:92 or SEQ ID NO:149.

[00852] Embodiment III-11. The genetically modified host cell of Embodiment III-7, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative, wherein the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is a fatty acyl-CoA ligase polypeptide.

[00853] Embodiment III-12. The genetically modified host cell of Embodiment III-11, wherein the fatty acyl-CoA ligase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:145 or SEQ ID NO:147.

[00854] Embodiment III-13. The genetically modified host cell of Embodiment III-7, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative, wherein the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is a fatty acyl-CoA synthetase (FAA) polypeptide.

[00855] Embodiment III-14. The genetically modified host cell of Embodiment III-13, wherein the FAA polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200.

[00856] Embodiment III-15. The genetically modified host cell of any one of Embodiments III-7 to III-14, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that generates geranyl pyrophosphate, wherein the polypeptide that generates geranyl pyrophosphate is a geranyl pyrophosphate synthetase (GPPS) polypeptide.

[00857] Embodiment III-16. The genetically modified host cell of Embodiment III-15, wherein the GPPS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:60.

2018256863 11 Apr 2024

[00858] Embodiment III-17. The genetically modified host cell of any one of Embodiments III-7 to III-16, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that generates malonyl-CoA, wherein the polypeptide that generates malonyl-CoA is an acetyl-CoA carboxylase-1 (ACC1) polypeptide.

[00859] Embodiment III-18. The genetically modified host cell of Embodiment III-17, wherein the ACC1 polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207.

[00860] Embodiment III-19. The genetically modified host cell of any one of Embodiments III-1 to III-18, wherein the genetically modified host cell further comprises one or more of the following: a) one or more heterologous nucleic acids encoding a HMG-CoA synthase (HMGS) polypeptide; b) one or more heterologous nucleic acids encoding a 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMGR) polypeptide; c) one or more heterologous nucleic acids encoding a mevalonate kinase (MK) polypeptide; d) one or more heterologous nucleic acids encoding a phosphomevalonate kinase (PMK) polypeptide; e) one or more heterologous nucleic acids encoding a mevalonate pyrophosphate decarboxylase (MVD) polypeptide; or f) one or more heterologous nucleic acids encoding a isopentenyl diphosphate isomerase (IDI) polypeptide.

[00861] Embodiment III-20. The genetically modified host cell of Embodiment III-19, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an IDI polypeptide.

[00862] Embodiment III-21. The genetically modified host cell of Embodiment III-20, wherein the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58.

[00863] Embodiment III-22. The genetically modified host cell of any one of Embodiments III-19 to III-21, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an HMGR polypeptide.

[00864] Embodiment III-23. The genetically modified host cell of Embodiment III-22, wherein the HMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:22.

[00865] Embodiment III-24. The genetically modified host cell of any one of Embodiments III-19 to III-21, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an HMGR polypeptide, wherein the HMGR polypeptide is a truncated HMGR (tHMGR) polypeptide.

2018256863 11 Apr 2024

[00866] Embodiment III-25. The genetically modified host cell of Embodiment III-24, wherein the tHMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:17, SEQ ID NO:52, SEQ ID NO:113, or SEQ ID NO:208.

[00867] Embodiment III-26. The genetically modified host cell of any one of Embodiments III-19 to III-25, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an HMGS polypeptide.

[00868] Embodiment III-27. The genetically modified host cell of Embodiment III-26, wherein the HMGS polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:23, SEQ ID NO:24, or SEQ ID NO:115.

[00869] Embodiment III-28. The genetically modified host cell of any one of Embodiments III-19 to III-27, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an MK polypeptide.

[00870] Embodiment III-29. The genetically modified host cell of Embodiment III-28, wherein the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64.

[00871] Embodiment III-30. The genetically modified host cell of any one of Embodiments III-19 to III-29, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a PMK polypeptide.

[00872] Embodiment III-31. The genetically modified host cell of Embodiment III-30, wherein the PMK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:62 or SEQ ID NO:205.

[00873] Embodiment III-32. The genetically modified host cell of any one of Embodiments III-19 to III-31, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a MVD polypeptide.

[00874] Embodiment III-33. The genetically modified host cell of Embodiment III-32, wherein the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66.

[00875] Embodiment III-34. The genetically modified host cell of any one of Embodiments III-1 to III-33, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA.

[00876] Embodiment III-35. The genetically modified host cell of Embodiment III-34, wherein the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide.

2018256863 11 Apr 2024

[00877] Embodiment III-36. The genetically modified host cell of Embodiment III-35, wherein the acetoacetyl-CoA thiolase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25.

[00878] Embodiment III-37. The genetically modified host cell of any one of Embodiments III-1 to III-36, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a pyruvate decarboxylase (PDC) polypeptide.

[00879] Embodiment III-38. The genetically modified host cell of Embodiment III-37, wherein the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117.

[00880] Embodiment III-39. The genetically modified host cell of any one of Embodiments III-1 to III-38, wherein the genetically modified host cell is a eukaryotic cell.

[00881] Embodiment III-40. The genetically modified host cell of Embodiment III-39, wherein the eukaryotic cell is a yeast cell.

[00882] Embodiment III-41. The genetically modified host cell of Embodiment III-40, wherein the yeast cell is *Saccharomyces cerevisiae*.

[00883] Embodiment III-42. The genetically modified host cell of Embodiment III-41, wherein the *Saccharomyces cerevisiae* is a protease-deficient strain of *Saccharomyces cerevisiae*.

[00884] Embodiment III-43. The genetically modified host cell of any one of Embodiments III-1 to III-39, wherein the genetically modified host cell is a plant cell.

[00885] Embodiment III-44. The genetically modified host cell of any one of Embodiments III-1 to III-38, wherein the genetically modified host cell is a prokaryotic cell.

[00886] Embodiment III-45. The genetically modified host cell of any one of Embodiments III-1 to III-44, wherein at least one of the one or more heterologous nucleic acids is integrated into the chromosome of the genetically modified host cell.

[00887] Embodiment III-46. The genetically modified host cell of any one of Embodiments III-1 to III-44, wherein at least one of the one or more heterologous nucleic acids is maintained extrachromosomally.

[00888] Embodiment III-47. The genetically modified host cell of any one of Embodiments III-1 to III-44, wherein two or more of the one or more heterologous nucleic acids are present in a single expression vector.

2018256863 11 Apr 2024

[00889] Embodiment III-48. The genetically modified host cell of any one of Embodiments III-1 to III-44, wherein at least one of the heterologous nucleic acids is operably linked to an inducible promoter.

[00890] Embodiment III-49. The genetically modified host cell of any one of Embodiments III-1 to III-44, wherein at least one of the heterologous nucleic acids is operably linked to a constitutive promoter.

[00891] Embodiment III-50. The genetically modified host cell of any one of Embodiments III-1 to III-49, wherein culturing of the genetically modified host cell in a suitable medium provides for synthesis of the cannabinoid or the cannabinoid derivative in an increased amount compared to a non-genetically modified host cell cultured under similar conditions.

[00892] Embodiment III-51. The genetically modified host cell of any one of Embodiments III-1 to III-50, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a cannabinoid synthase polypeptide.

[00893] Embodiment III-52. The genetically modified host cell of Embodiment III-51, wherein the cannabinoid synthase polypeptide is a tetrahydrocannabinolic acid (THCA) synthase polypeptide.

[00894] Embodiment III-53. The genetically modified host cell of Embodiment III-52, wherein the THCA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:14, SEQ ID NO:86, SEQ ID NO:104, SEQ ID NO:153, or SEQ ID NO:155.

[00895] Embodiment III-54. The genetically modified host cell of Embodiment III-51, wherein the cannabinoid synthase polypeptide is a cannabidiolic acid (CBDA) synthase polypeptide.

[00896] Embodiment III-55. The genetically modified host cell of Embodiment III-54, wherein the CBDA synthase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

[00897] Embodiment III-56. The genetically modified host cell of any one of Embodiments III-1 to III-55, wherein the cannabinoid is cannabigerolic acid, cannabigerol, Δ^9-tetrahydrocannabinolic acid, Δ^9-tetrahydrocannabinol, Δ^8-tetrahydrocannabinolic acid, Δ^8-tetrahydrocannabinol, cannabidiolic acid, cannabidiol, cannabichromenic acid, cannabichromene, cannabinolic acid, cannabinol, cannabidivarinic acid, cannabidivarin, tetrahydrocannabivarinic acid, tetrahydrocannabivarin, cannabichromevarinic acid,

2018256863 11 Apr 2024

cannabichromevarin, cannabigerovarinic acid, cannabigerovarin, cannabicyclolic acid, cannabicyclol, cannabielsoinic acid, cannabielsoin, cannabicitranic acid, or cannabicitran.

[00898] Embodiment III-57. The genetically modified host cell of Embodiment III-56, wherein the cannabinoid is cannabigerolic acid.

[00899] Embodiment III-58. A method of producing a cannabinoid or a cannabinoid derivative in a genetically modified host cell, the method comprising: a) culturing the genetically modified host cell of any one of Embodiments III-1 to III-57 in a suitable medium; and b) recovering the produced cannabinoid or cannabinoid derivative.

[00900] Embodiment III-59. A method of producing a cannabinoid or a cannabinoid derivative, the method comprising: a) culturing the genetically modified host cell of any one of Embodiments III-1 to III-57 in a suitable medium comprising a carboxylic acid; b) recovering the produced cannabinoid or cannabinoid derivative.

[00901] Embodiment III-60. A method of producing a cannabinoid or a cannabinoid derivative, the method comprising: a) culturing the genetically modified host cell of any one of Embodiments III-1 to III-57 in a suitable medium comprising olivetolic acid or an olivetolic acid derivative; b) recovering the produced cannabinoid or cannabinoid derivative.

[00902] Embodiment III-61. A method of producing a cannabinoid or a cannabinoid derivative in a genetically modified host cell, the method comprising: a) culturing a genetically modified host cell comprising one or more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide, wherein said geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide can catalyze production of cannabigerolic acid from geranyl pyrophosphate and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, in a suitable medium; and b) recovering the produced cannabinoid or cannabinoid derivative.

[00903] Embodiment III-62. A method of producing a cannabinoid or a cannabinoid derivative in a genetically modified host cell, the method comprising: a) culturing a genetically modified host cell comprising one or more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110 in a suitable medium; and b) recovering the produced cannabinoid or cannabinoid derivative.

[00904] Embodiment III-63. A method of producing a cannabinoid or a cannabinoid derivative in a genetically modified host cell, the method comprising: a) culturing a genetically modified host cell comprising one or more heterologous nucleic acids encoding a

2018256863 11 Apr 2024

geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100 in a suitable medium; and b) recovering the produced cannabinoid or cannabinoid derivative.

[00905] Embodiment III-64. The method of any one of Embodiments III-58 to III-63, wherein the suitable medium comprises a fermentable sugar.

[00906] Embodiment III-65. The method of any one of Embodiments III-58 to III-63, wherein the suitable medium comprises a pretreated cellulosic feedstock.

[00907] Embodiment III-66. The method of any one of Embodiments III-58 to III-63, wherein the suitable medium comprises a non-fermentable carbon source.

[00908] Embodiment III-67. The method of Embodiment III-66, wherein the non-fermentable carbon source comprises ethanol.

[00909] Embodiment III-68. The method of any one of Embodiments III-58 to III-67, wherein the cannabinoid is cannabigerolic acid, cannabigerol, Δ^9-tetrahydrocannabinolic acid, Δ^9-tetrahydrocannabinol, Δ^8-tetrahydrocannabinolic acid, Δ^8-tetrahydrocannabinol, cannabidiolic acid, cannabidiol, cannabichromenic acid, cannabichromene, cannabinolic acid, cannabinol, cannabidivarinic acid, cannabidivarin, tetrahydrocannabivarinic acid, tetrahydrocannabivarin, cannabichromevarinic acid, cannabichromevarin, cannabigerovarinic acid, cannabigerovarin, cannabicyclolic acid, cannabicyclol, cannabielsoinic acid, cannabielsoin, cannabicitranic acid, or cannabicitran.

[00910] Embodiment III-69. An isolated or purified geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide, wherein said geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide can catalyze production of cannabigerolic acid from geranyl pyrophosphate and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[00911] Embodiment III-70. An isolated or purified polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110.

[00912] Embodiment III-71. An isolated or purified polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100.

[00913] Embodiment III-72. An isolated or purified nucleic acid encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide, wherein said geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide can catalyze production of cannabigerolic acid from geranyl pyrophosphate and olivetolic acid in an amount at least ten

2018256863 11 Apr 2024

times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[00914] Embodiment III-73. An isolated or purified nucleic acid encoding a polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110.

[00915] Embodiment III-74. An isolated or purified nucleic acid encoding a polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100.

[00916] Embodiment III-75. A vector comprising a nucleic acid encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide, wherein said geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide can catalyze production of cannabigerolic acid from geranyl pyrophosphate and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

[00917] Embodiment III-76. A vector comprising a nucleic acid encoding a polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110.

[00918] Embodiment III-77. A vector comprising a nucleic acid encoding a polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100.

[00919] Embodiment III-78. A method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing one or more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide, wherein said geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide can catalyze production of cannabigerolic acid from geranyl pyrophosphate and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82, into a host cell.

[00920] Embodiment III-79. A method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing one or more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110 into a host cell.

[00921] Embodiment III-80. A method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing one or

2018256863 11 Apr 2024

more heterologous nucleic acids encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100 into a host cell.

[00922] Embodiment III-81. A method of making a genetically modified host cell for producing a cannabinoid or a cannabinoid derivative, comprising introducing the vector of any one of Embodiments III-75 to III-77 into a host cell.

[00923] Provided in Table 1 are amino acid and nucleotide sequences disclosed herein. Where a genus and/or species is noted, the sequence should not be construed to be limited only to the specified genus and/or species, but also includes other genera and/or species expressing said sequence.

Table 1: Amino acid and nucleotide sequences of the disclosure

SEQ ID	SEQUENCE
SEQ ID NO:1 hexanoyl-CoA synthetase (HCS) GenBank AFD33359 *Cannabis sativa*	MALELPHLLPYKLVKGQTLVAQAARAELASSSSSSVILKSNFINNNYIN YCNNNNNNERRLVVRRDWETMASSPSHSRNNNDIRTINHLRHVDSMA TMPSGAGKIPRLNAVILGEALATEENDLVFPTDEFSQQAHVPSPQKYLE MYKRSIEDPAGFWSEIASQFYWKQKWDDSVYSENLDVSKGRVNIEWF KGGITNICYNCLDKNVEAGLGDKIALYWEGNDTGFDDSLTYSQLLHK VCQLANYLKDMGVQKGDAVVIYLPMLLELPITMLACARIGAVHSVVF AGFSAESLSQRIIDCKPKVVITCNAVKRGPKIIHLKDIVDAALVESAKTG VPIDTCLVYENQLAMKRDITKWQDGRDIWWQDVIPKYPTECAVEWV DAEDPLFLLYTSGSTGKPKGVLHTTGGYMVYTATTFKYAFDYKPSDV YWCTADCGWITGHSYVTYGPLLNGASCIVFEGAPNYPDSGRCWDIVD KYKVTIFYTAPTLVRSLMRDGDEYVTRYSRKSLRILGSVGEPINPSAWR WFYNVVGDSRCPISDTWWQTETGGFMITPLPGAWPQKPGSATFPFFGV KPVIVDEKGVEIEGECSGYLCVKGSWPGAFRTLYGDYERYETTYFKPF TGYYFTGDGCSRDKDGYHWLTGRVDDVINVSGHRIGTAEVESALVSH PKCAEAAVVGIEHEVKGQAIYAFVTLVEGEPYSEELRKSLILSVRKQIG AFAAPERIHWAPGLPKTRSGKIMRRILRKIASGQLDELGDTSTLADPNV VEQLISLSNC
SEQ ID NO:2 RevS BAK64635.1| putative CoA ligase [Streptomyces sp. SN-593]	MELALPAELAPTLPEALRLRSEQQPDTVAYVFLRDGETPEETLTYGRL DRAARARAAALEAAGLAGGTAVLLYPSGLEFVAALLGCMYAGTAGA PVQVPTRRRGMERARRIADDAGAKTILTTTAVKREVEEHFADLLTGLT VIDTESLPDVPDDAPAVRLPGPDDVALLQYTSGSTGDPKGVEVTHANF RANVAETVELWPVRSDGTVVNWLPLFHDMGLMFGVVMPLFTGVPAY LMAPQSFIRRPARWLEAISRFRGTHAAAPSFAYELCVRSVADTGLPAG LDLSSWRVAVNGAEPVRWTAVADFTEAYAPAGFRPQAMCPGYGLAE NTLKLSGSPEDRPPTLLRADAAALQDGRVVPLTGPGTDGVRLVGSGVT VPSSRVAVVDPGTGTEQPAGRVGEIWINGPCVARGYHGRPAESAESFG ARIAGQEARGTWLRTGDLGFLHDGEVFVAGRLKDVVIHQGRNFYPQD IELSAEVSDRALHPNCAAAFALDDGRTERLVLLVEADGRALRNGGAD ALRARVHDAVWDRQRLRIDEIVLLRRGALPKTSSGKVQRRLARSRYL DGEFGPAPAREA
SEQ ID NO:3 hexanoyl-CoA	MVIQGKRLAASSIQLLASSLDAKKLCYEYDERQAPGVTQITEEAPTEQP PLSTPPSLPQTPNISPISASKIVIDDVALSRVQIVQALVARKLKTAIAQLP

SEQ ID	SEQUENCE
synthase (AflA); GenBank AAL99898 *Aspergillus* sp.	TSKSIKELSGGRSSLQNELVGDIHNEFSSIPDAPEQILLRDFGDANPTVQ LGKTSSAAVAKLISSKMPSDFNANAIRAHLANKWGLGPLRQTAVLLY AIASEPPSRLASSSAAEEYWDNVSSMYAESCGITLRPRQDTMNEDAMA SSAIDPAVVAEFSKGHRRLGVQQFQALAEYLQIDLSGSQASQSDALVA ELQQKVDLWTAEMTPEFLAGISPMLDVKKSRRYGSWWNMARQDVLA FYRRPSYSEFVDDALAFKVFLNRLCNRADEALLNMVRSLSCDAYFKQ GSLPGYHAASRLLEQAITSTVADCPKARLILPAVGPHTTITKDGTIEYAE APRQGVSGPTAYIQSLRQGASFIGLKSADVDTQSNLTDALLDAMCLAL HNGISFVGKTFLVTGAGQGSIGAGVVRLLLEGGARVLVTTSREPATTS RYFQQMYDNHGAKFSELRVVPCNLASAQDCEGLIRHVYDPRGLNWD LDAILPFAAASDYSTEMHDIRGQSELGHRLMLVNVFRVLGHIVHCKRD AGVDCHPTQVLLPLSPNHGIFGGDGMYPESKLALESLFHRIRSESWSDQ LSICGVRIGWTRSTGLMTAHDIIAETVEEHGIRTFSVAEMALNIAMLLT PDFVAHCEDGPLDADFTGSLGTLGSIPGFLAQLHQKVQLAAEVIRAVQ AEDEHERFLSPGTKPTLQAPVAPMHPRSSLRVGYPRLPDYEQEIRPLSP RLERLQDPANAVVVVGYSELGPWGSARLRWEIESQGQWTSAGYVEL AWLMNLIRHVNDESYVGWVDTQTGKPVRDGEIQALYGDHIDNHTGIR PIQSTSYNPERMEVLQEVAVEEDLPEFEVSQLTADAMRLRHGANVSIR PSGNPDACHVKLKRGAVILVPKTVPFVWGSCAGELPKGWTPAKYGIPE NLIHQVDPVTLYTICCVAEAFYSAGITHPLEVFRHIHLSELGNFIGSSMG GPTKTRQLYRDVYFDHEIPSDVLQDTYLNTPAAWVNMLLLGCTGPIKT PVGACATGVESIDSGYESIMAGKTKMCLVGGYDDLQEEASYGFAQLK ATVNVEEEIACGRQPSEMSRPMAESRAGFVEAHGCGVQLLCRGDIALQ MGLPIYAVIASSAMAADKIGSSVPAPGQGILSFSRERARSSMISVTSRPS SRSSTSSEVSDKSSLTSITSISNPAPRAQRARSTTDMAPLRAALATWGLT IDDLDVASLHGTSTRGNDLNEPEVIETQMRHLGRTPGRPLWAICQKSV TGHPKAPAAAWMLNGCLQVLDSGLVPGNRNLDTLDEALRSASHLCFP TRTVQLREVKAFLLTSFGFGQKGGQVVGVAPKYFFATLPRPEVEGYYR KVRVRTEAGDRAYAAAVMSQAVVKIQTQNPYDEPDAPRIFLDPLARIS QDPSTGQYRFRSDATPALDDDALPPPGEPTELVKGISSAWIEEKVRPHM SPGGTVGVDLVPLASFDAYKNAIFVERNYTVRERDWAEKSADVRAAY ASRWCAKEAVFKCLQTHSQGAGAAMKEIEIEHGGNGAPKVKLRGAA QTAARQRGLEGVQLSISYGDDAVIAVALGLMSGAS
SEQ ID NO:4 hexanoyl-CoA synthase (AflB) AAS66003.1| fatty acid synthase beta subunit [Aspergillus sp.]	MGSVSREHESIPIQAAQRGAARICAAFGGQGSNNLDVLKGLLELYKRY GPDLDELLDVASNTLSQLASSPAAIDVHEPWGFDLRQWLTTPEVAPSK EILALPPRSFPLNTLLSLALYCATCRELELDPGQFRSLLHSSTGHSQGIL AAVAITQAESWPTFYDACRTVLQISFWIGLEAYLFTPSSAASDAMIQDC IEHGEGLLSSMLSVSGLSRSQVERVIEHVNKGLGECNRWVHLALVNSH EKFVLAGPPQSLWAVCLHVRRIRADNDLDQSRILFRNRKPIVDILFLPIS APFHTPYLDGVQDRVIEALSSASLALHSIKIPLYHTGTGSNLQELQPHQ LIPTLIRAITVDQLDWPLVCRGLNATHVLDFGPGQTCSLIQELTQGTGV SVIQLTTQSGPKPVGGHLAAVNWEAEFGLRLHANVHGAAKLHNRMT TLLGKPPVMVAGMTPTTVRWDFVAAVAQAGYHVELAGGGYHAERQ FEAEIRRLATAIPADHGITCNLLYAKPTTFSWQISVIKDLVRQGVPVEGI TIGAGIPSPEVVQECVQSIGLKHISFKPGSFEAIHQVIQIARTHPNFLIGLQ WTAGRGGGHHSWEDFHGPILATYAQIRSCPNILLVVGSGFGGGPDTFP YLTGQWAQAFGYPCMPFDGVLLGSRMMVAREAHTSAQAKRLIIDAQ GVGDADWHKSFDEPTGGVVTVNSEFGQPIHVLATRGVMLWKELDNR VFSIKDTSKRLEYLRNHRQEIVSRLNADFARPWFAVDGHGQNVELED MTYLEVLRRLCDLTYVSHQKRWVDPSYRILLLDFVHLLRERFQCAIDN PGEYPLDIIVRVEESLKDKAYRTLYPEDVSLLMHLFSRRDIKPVPFIPRL DERFETWFKKDSLWQSEDVEAVIGQDVQRIFIIQGPMAVQYSISDDESV KDILHNICNHYVEALQADSRETSIGDVHSITQKPLSAFPGLKVTTNRVQ GLYKFEKVGAVPEMDVLFEHIVGLSKSWARTCLMSKSVFRDGSRLHN

2018256863 11 Apr 2024

SEQ ID	SEQUENCE
	PIRAALQLQRGDTIEVLLTADSEIRKIRLISPTGDGGSTSKVVLEIVSNDG QRVFATLAPNIPLSPEPSVVFCFKVDQKPNEWTLEEDASGRAERIKALY MSLWNLGFPNKASVLGLNSQFTGEELMITTDKIRDFERVLRQTSPLQL QSWNPQGCVPIDYCVVIAWSALTKPLMVSSLKCDLLDLLHSAISFHYA PSVKPLRVGDIVKTSSRILAVSVRPRGTMLTVSADIQRQGQHVVTVKS DFFLGGPVLACETPFELTEEPEMVVHVDSEVRRAILHSRKWLMREDRA LDLLGRQLLFRLKSEKLFRPDGQLALLQVTGSVFSYSPDGSTTAFGRV YFESESCTGNVVMDFLHRYGAPRAQLLELQHPGWTGTSTVAVRGPRR SQSYARVSLDHNPIHVCPAFARYAGLSGPIVHGMETSAMMRRIAEWAI GDADRSRFRSWHITLQAPVHPNDPLRVELQHKAMEDGEMVLKVQAF NERTEERVAEADAHVEQETTAYVFCGQGSQRQGMGMDLYVNCPEAK ALWARADKHLWEKYGFSILHIVQNNPPALTVHFGSQRGRRIRANYLR MMGQPPIDGRHPPILKGLTRNSTSYTFSYSQGLLMSTQFAQPALALME MAQFEWLKAQGVVQKGARFAGHSLGEYAALGACASFLSFEDLISLIFY RGLKMQNALPRDANGHTDYGMLAADPSRIGKGFEEASLKCLVHIIQQ ETGWFVEVVNYNINSQQYVCAGHFRALWMLGKICDDLSCHPQPETVE GQELRAMVWKHVPTVEQVPREDRMERGRATIPLPGIDIPYHSTMLRGE IEPYREYLSERIKVGDVKPCELVGRWIPNVVGQPFSVDKSYVQLVHGIT GSPRLHSLLQQMA
SEQ ID NO:5 AAF08793.1|AF182828_1 geranyl diphosphate synthase large subunit [Mentha x piperita]	MSALVNPVAKWPQTIGVKDVHGGRRRRSRSTLFQSHPLRTEMPFSLYF SSPLKAPATFSVSAVYTKEGSEIRDKDPAPSTSPAFDFDGYMLRKAKSV NKALEAAVQMKEPLKIHESMRYSLLAGGKRVRPMLCIAACELVGGDE STAMPAACAVEMIHTMSLMHDDLPCMDNDDLRRGKPTNHMAFGESV AVLAGDALLSFAFEHVAAATKGAPPERIVRVLGELAVSIGSEGLVAGQ VVDVCSEGMAEVGLDHLEFIHHHKTAALLQGSVVLGAILGGGKEEEV AKLRKFANCIGLLFQVVDDILDVTKSSKELGKTAGKDLVADKTTYPKL IGVEKSKEFADRLNREAQEQLLHFHPHRAAPLIALANYIAYRDN
SEQ ID NO:6 AAF08792.1|AF182827_1 geranyl diphosphate synthase small subunit [Mentha x piperita]	MAINLSHINSKTCFPLKTRSDLSRSSSARCMPTAAAAAFPTIATAAQSQ PYWAAIEADIERYLKKSITIRPPETVFGPMHHLTFAAPATAASTLCLAA CELVGGDRSQAMAAAAAIHLVHAAAYVHEHLPLTDGSRPVSKPAIQH KYGPNVELLTGDGIVPFGFELLAGSVDPARTDDPDRILRVIIEISRAGGP EGMISGLHREEEIVDGNTSLDFIEYVCKKKYGEMHACGAACGAILGGA AEEEIQKLRNFGLYQGTLRGMMEMKNSHQLIDENIIGKLKELALEELG GFHGKNAELMSSLVAEPSLYAA
SEQ ID NO:7 Erg20: farnesylpyro phosphate synthase (Saccharomyces sp.)	MASEKEIRRERFLNVFPKLVEELNASLLAYGMPKEACDWYAHSLNYN TPGGKLNRGLSVVDTYAILSNKTVEQLGQEEYEKVAILGWCIELLQAY FLVADDMMDKSITRRGQPCWYKVPEVGEIAINDAFMLEAAIYKLLKS HFRNEKYYIDITELFHEVTFQTELGQLMDLITAPEDKVDLSKFSLKKHS FIVTF**K**TAYYSFYLPVALAMYVAGITDEKDLKQARDVLIPLGEYFQIQD DYLDCFGTPEQIGKIGTDIQDNKCSWVINKALELASAEQRKTLDENYG KKDSVAEAKCKKIFNDLKIEQLYHEYEESIAKDLKAKISQVDESRGFKA DVLTAFLNKVYKRSK
SEQ ID NO:8 Mutated Erg20: farnesylpyro phosphate synthase (K197G)	MASEKEIRRERFLNVFPKLVEELNASLLAYGMPKEACDWYAHSLNYN TPGGKLNRGLSVVDTYAILSNKTVEQLGQEEYEKVAILGWCIELLQAY FLVADDMMDKSITRRGQPCWYKVPEVGEIAINDAFMLEAAIYKLLKS HFRNEKYYIDITELFHEVTFQTELGQLMDLITAPEDKVDLSKFSLKKHS FIVTF**G**TAYYSFYLPVALAMYVAGITDEKDLKQARDVLIPLGEYFQIQD DYLDCFGTPEQIGKIGTDIQDNKCSWVINKALELASAEQRKTLDENYG KKDSVAEAKCKKIFNDLKIEQLYHEYEESIAKDLKAKISQVDESRGFKA DVLTAFLNKVYKRSK

2018256863 11 Apr 2024

SEQ ID	SEQUENCE
SEQ ID NO:9 GenBank EXX73400 acetyl-CoA carboxylase (ACC1) Rhizophagus irregularis DAOM 197198w	MRAQAHLGGGLKRIETQHQKGKLTARERAELLLDPGSFNEYDTFVEH QCTDFGMDKNKIIGDGVVTGHGTINGRRVFTFSQDFTAFGGSLSKMHA QKICKIMDKAMLVGAPVIGLNDSGGARIQEGVDSLAGYADIFQRNVLS SGVVPQLSLIMGPCAGGAVYSPALTDFTFMVRDTSYLFVTGPEVVKAV CNEDVTQEELGGANTHTVISGVAHAAFENDIEAIQRIRDFMDFLPLSNR EQAPTRYSDDPIDREDPSLNHIIPVDSTKAYDMREIITRLIDDGHFFEIMP DYAKNIVVGFARMGGKTVSIVGNQPLVSSGVLDINSSVKAARFVRFCD AFNIPIITLVDVPGFLPGTAQEHNGIIRHGAKLLYAYAEATVPKITIITRK AYGGAYDVMSSKHLRGDMNYSWPTGEIAVMGAKGAVEIIFRHVEDR TQSEHEYIDKFANPIPAAQRGYIDDIILPAATRKRIIEDLFVLSHKQLPLI YKKHDNCPL
SEQ ID NO:10 Olivetolic acid cyclase (OAC) GenBank AFN42527 *Cannabis sativa*	MAVKHLIVLKFKDEITEAQKEEFFKTYVNLVNIIPAMKDVYWGKDVT QKNKEEGYTHIVEVTFESVETIQDYIIHPAHVGFGDVYRSFWEKLLIFD YTPRK
SEQ ID NO:11 Tetraketide synthase (TKS) GenBank B1Q2B6 Cannabis sativa	MNHLRAEGPASVLAIGTANPENILLQDEFPDYYFRVTKSEHMTQLKEK FRKICDKSMIRKRNCFLNEEHLKQNPRLVEHEMQTLDARQDMLVVEV PKLGKDACAKAIKEWGQPKSKITHLIFTSASTTDMPGADYHCAKLLGL SPSVKRVMMYQLGCYGGGTVLRIAKDIAENNKGARVLAVCCDIMACL FRGPSESDLELLVGQAIFGDGAAAVIVGAEPDESVGERPIFELVSTGQTI LPNSEGTIGGHIREAGLIFDLHKDVPMLISNNIEKCLIEAFTPIGISDWNSI FWITHPGGKAILDKVEEKLHLKSDKFVDSRHVLSEHGNMSSSTVLFVM DELRKRSLEEGKSTTGDGFEWGVLFGFGPGLTVERVVVRSVPIKY
SEQ ID NO:12 Truncated geranyl pyrophosphate olivetolic acid geranyltransferase (GOT)	MNSIRAATTNQTEPPESDNHSVATKILNFGKACWKLQRPYTIIAFTSCA CGLFGKELLHNTNLISWSLMFKAFFFLVAILCIASFTTTINQIYDLHIDRI NKPDLPLASGEISVNTAWIMSIIVALFGLIITIKMKGGPLYIFGYCFGIFG GIVYSVPPFRWKQNPSTAFLLNFLAHIITNFTFYYASRAALGLPFELRPS FTFLLAFMKSMGSALALIKDASDVEGDTKFGISTLASKYGSRNLTLFCS GIVLLSYVAAILAGIIWPQAFNSNVMLLSHAILAFWLILQTRDFALTNY DPEAGRRFYEFMWKLYYAEYLVYVFIGS
SEQ ID NO:13 Engineered geranyl pyrophosphate olivetolic acid geranyltransferase (GOT)	MKDQRGNSIRASAQIEDRPPESGNLSALTNVKDFVSVCWEYVRPYTAK GVIICSSCLFGRELLENPNLFSWPLIFKAFFFLVAILCIASFTTTINQIYDL HIDRINKPDLPLASGEISVNTAWIMSIIVALFGLIITIKMKGGPLYIFGYCF GIFGGIVYSVPPFRWKQNPSTAFLLNFLAHIITNFTFYYASRAALGLPFE LRPSFTFLLAFMKSMGSALALIKDASDVEGDTKFGISTLASKYGSRNLT LFCSGIVLLSYVAAILAGIIWPQAFNSNVMLLSHAILAFWLILQTRDFAL TNYDPEAGRRFYEFMWKLYYAEYLVYVFIGS
SEQ ID NO:14 Mutant tetrahydro cannabinolic acid synthase (THCAS)	MNCSAFSFWFVCKIIFFFLSFHIQISIANPRENFLKCFSKHIPNNVANPKL VYTQHDQLYMSILNSTIQNLRFISDTTPKPLVIVTPSNNSHIQATILCSKK VGLQIRTRSGGHDAEGMSYISQVPFVVVDLRNMHSIKIDVHSQTAWVE AGATLGEVYYWINEKNENLSFPGGYCPTVGVGGHFSGGGYGALMRN YGLAADNIIDAHLVNVDGKVLDRKSMGEDLFWAIRGGGGENFGIIAA WKIKLVAVPSKSTIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLVLM THFITKNITDNHGKNKTTVHGYFSSIFHGGVDSLVDLMNKSFPELGIKK TDCKEFSWIDTTIFYSGVVNFNTANFKKEILLDRSAGKKTAFSIKLDYV KKPIPETAMVKILEKLYEEDVGAGMYVLYPYGGIMEEISESAIPFPHRA GIMYELWYTASWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLN

2018256863 11 Apr 2024

SEQ ID	SEQUENCE
	YRDLDLGKTNHASPNNYTQARIWGEKYFGKNFNRLVKVKTKVDPNN FFRNEQSIPPLPPHHHGS
SEQ ID NO:15 Truncated tetrahydro cannabinolic acid synthase (THCAS)	MNPRENFLKCFSKHIPNNVANPKLVYTQHDQLYMSILNSTIQNLRFISD TTPKPLVIVTPSNNSHIQATILCSKKVGLQIRTRSGGHDAEGMSYISQVP FVVVDLRNMHSIKIDVHSQTAWVEAGATLGEVYYWINEKNENLSFPG GYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDR KSMGEDLFWAIRGGGGENFGIIAAWKIKLVAVPSKSTIFSVKKNMEIH GLVKLFNKWQNIAYKYDKDLVLMTHFITKNITDNHGKNKTTVHGYFS SIFHGGVDSLVDLMNKSFPELGIKKTDCKEFSWIDTTIFYSGVVNFNTA NFKKEILLDRSAGKKTAFSIKLDYVKKPIPETAMVKILEKLYEEDVGAG MYVLYPYGGIMEEISESAIPFPHRAGIMYELWYTASWEKQEDNEKHIN WVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNHASPNNYTQARIWG EKYFGKNFNRLVKVKTKVDPNNFFRNEQSIPPLPPHHHGS
SEQ ID NO:16 Truncated cannabidiolic acid synthase (CBDAS)	MNPRENFLKCFSQYIPNNATNLKLVYTQNNPLYMSVLNSTIHNLRFTS DTTPKPLVIVTPSHVSHIQGTILCSKKVGLQIRTRSGGHDSEGMSYISQV PFVIVDLRNMRSIKIDVHSQTAWVEAGATLGEVYYWVNEKNENLSLA AGYCPTVCAGGHFGGGGYGPLMRNYGLAADNIIDAHLVNVHGKVLD RKSMGEDLFWALRGGGAESFGIIVAWKIRLVAVPKSTMFSVKKIMEIH ELVKLVNKWQNIAYKYDKDLLLMTHFITRNITDNQGKNKTAIHTYFSS VFLGGVDSLVDLMNKSFPELGIKKTDCRQLSWIDTIIFYSGVVNYDTD NFNKEILLDRSAGQNGAFKIKLDYVKKPIPESVFVQILEKLYEEDIGAG MYALYPYGGIMDEISESAIPFPHRAGILYELWYICSWEKQEDNEKHLN WIRNIYNFMTPYVSKNPRLAYLNYRDLDIGINDPKNPNNYTQARIWGE KYFGKNFDRLVKVKTLVDPNNFFRNEQSIPPLPRHRHGS
SEQ ID NO:17 Truncated 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMGR)	MVLTNKTVISGSKVKSLSSAQSSSSGPSSSSEEDDSRDIESLDKKIRPLEE LEALLSSGNTKQLKNKEVAALVIHGKLPLYALEKKLGDTTRAVAVRR KALSILAEAPVLASDRLPYKNYDYDRVFGACCENVIGYMPLPVGVIGP LVIDGTSYHIPMATTEGCLVASAMRGCKAINAGGGATTVLTKDGMTR GPVVRFPTLKRSGACKIWLDSEEGQNAIKKAFNSTSRFARLQHIQTCLA GDLLFMRFRTTTGDAMGMNMISKGVEYSLKQMVEEYGWEDMEVVS VSGNYCTDKKPAAINWIEGRGKSVVAEATIPGDVVRKVLKSDVSALV ELNIAKNLVGSAMAGSVGGFNAHAANLVTAVFLALGQDPAQNVESSN CITLMKEVDGDLRISVSMPSIEVGTIGGGTVLEPQGAMLDLLGVRGPH ATAPGTNARQLARIVACAVLAGELSLCAALAAGHLVQSHMTHNRKPA EPTK PNNLDATDINRLKDGSVTCIKS
SEQ ID NO:18 PaaH1: 3-hydroxyacyl-CoA dehydrogenase (Ralstonia sp.)	MSIRTVGIVGAGTMGNGIAQACAVVGLNVVMVDISDAAVQKGVATV ASSLDRLIKKEKLTEADKASALARIKGSTSYDDLKATDIVIEAATENYD LKVKILKQIDGIVGENVIIASNTSSISITKLAAVTSRADRFIGMHFFNPVP VMALVELIRGLQTSDTTHAAVEALSKQLGKYPITVKNSPGFVVNRILCP MINEAFCVLGEGLASPEEIDEGMKLGCNHPIGPLALADMIGLDTMLAV MEVLYTEFADPKYRPAMLMREMVAAGYLGRKTGRGVYVYSK
SEQ ID NO:19 Crt: crotonase (Clostridium sp.)	MELNNVILEKEGKVAVVTINRPKALNALNSDTLKEMDYVIGEIENDSE VLAVILTGAGEKSFVAGADISEMKEMNTIEGRKFGILGNKVFRRLELLE KPVIAAVNGFALGGGCEIAMSCDIRIASSNARFGQPEVGLGITPGFGGT QRLSRLVGMGMAKQLIFTAQNIKADEALRIGLVNKVVEPSELMNTAK EIANKIVSNAPVAVKLSKQAINRGMQCDIDTALAFESEAFGECFSTEDQ KDAMTAFIEKRKIEGFKNR
SEQ ID NO:20 Ter: trans-2-enoyl-CoA reductase	MIVKPMVRNNICLNAHPQGCKKGVEDQIEYTKKRITAEVKAGAKAPK NVLVLGCSNGYGLASRITAAFGYGAATIGVSFEKAGSETKYGTPGWY NNLAFDEAAKREGLYSVTIDGDAFSDEIKAQVIEEAKKKGIKFDLIVYS

2018256863 11 Apr 2024

SEQ ID	SEQUENCE
(Treponema sp.)	LASPVRTDPDTGIMHKSVLKPFGKTFTGKTVDPFTGELKEISAEPANDE EAAATVKVMGGEDWERWIKQLSKEGLLEEGCITLAYSYIGPEATQAL YRKGTIGKAKEHLEATAHRLNKENPSIRAFVSVNKGLVTRASAVIPVIP LYLASLFKVMKEKGNHEGCIEQITRLYAERLYRKDGTIPVDEENRIRID DWELEEDVQKAVSALMEKVTGENAESLTDLAGYRHDFLASNGFDVE GINYEAEVERFDRI
SEQ ID NO:21 BktB: beta-ketothiolase (Ralstonia sp.)	MTREVVVVSGVRTAIGTFGGSLKDVAPAELGALVVREALARAQVSGD DVGHVVFGNVIQTEPRDMYLGRVAAVNGGVTINAPALTVNRLCGSGL QAIVSAAQTILLGDTDVAIGGGAESMSRAPYLAPAARWGARMGDAGL VDMMLGALHDPFHRIHMGVTAENVAKEYDISRAQQDEAALESHRRAS AAIKAGYFKDQIVPVVSKGRKGDVTFDTDEHVRHDATIDDMTKLRPV FVKENGTVTAGNASGLNDAAAAVVMMERAEAERRGLKPLARLVSYG HAGVDPKAMGIGPVPATKIALERAGLQVSDLDVIEANEAFAAQACAV TKALGLDPAKVNPNGSGISLGHPIGATGALITVKALHELNRVQGRYAL VTMCIGGGQGIAAIFERI
SEQ ID NO:22 MvaE: acetyl-CoA acetyltransferase/HMG-CoA reductase (Enterococcus sp.)	MKEVVMIDAARTPIGKYRGSLSPFTAVELGTLVTKGLLDKTKLKKDKI DQVIFGNVLQAGNGQNVARQIALNSGLPVDVPAMTINEVCGSGMKAV ILARQLIQLGEAELVIAGGTESMSQAPMLKPYQSETNEYGEPISSMVND GLTDAFSNAHMGLTAEKVATQFSVSREEQDRYALSSQLKAAHAVEAG VFSEEIIPVKISDEDVLSEDEAVRGNSTLEKLGTLRTVFSEEGTVTAGNA SPLNDGASVVILASKEYAENNNLPYLATIKEVAEVGIDPSIMGIAPIKAI QKLTDRSGMNLSTIDLFEINEAFAASSIVVSQELQLDEEKVNIYGGAIAL GHPIGASGARILTTLAYGLLREQKRYGIASLCIGGGLGLAVLLEANMEQ THKDVQKKKFYQLTPSERRSQLIEKNVLTQETALIFQEQTLSEELSDHM IENQVSEVEIPMGIAQNFQINGKKKWIPMATEEPSVIAAASNGAKICGNI CAETPQRLMRGQIVLSGKSEYQAVINAVNHRKEELILCANESYPSIVKR GGGVQDISTREFMGSFHAYLSIDFLVDVKDAMGANMINSILESVANKL REWFPEEEILFSILSNFATESLASACCEIPFERLGRNKEIGEQIAKKIQQA GEYAKLDPYRAATHNKGIMNGIEAVVAATGNDTRAVSASIHAYAARN GLYQGLTDWQIKGDKLVGKLTVPLAVATVGGASNILPKAKASLAMLD IDSAKELAQVIAAVGLAQNLAALRALVTEGIQKGHMGLQARSLAISIG AIGEEIEQVAKKLREAEKMNQQTAIQILEKIREK
SEQ ID NO:23 MvaS: HMG-CoA synthase (Lactobacillus plantarum)	MKIGIDKLHFATSHLYVDMAELATARQAEPDKYLIGIGQSKMAVIPPS QDVVTLAANAAAPMLTATDIAAIDLLVVGTESGIDNSKASAIYVAKLL GLSQRVRTIEMKEACYAATAGVQLAQDHVRVHPDKKALVIGSDVAR YGLNTPGEPTQGGGAVAMLISADPKVLVLGTESSLLSEDVMDFWRPL YHTEALVDGKYSSNIYIDYFQDVFKNYLQTTQTSPDTLTALVFHLPYT KMGLKALRSVLPLVDAEKQAQWLAHFEHARQLNRQVGNLYTGSLYL SLLSQLLTDPQLQPGNRLGLFSYGSGAEGEFYTGVIQPDYQTGLDHGLP QRLARRRRVSVAEYEALFSHQLQWRADDQSVSYADDPHRFVLTGQK NEQRQYLDQQV
SEQ ID NO:24 Erg13: HMG-CoA synthase (Saccharomyces cerevisiae)	MKLSTKLCWCGIKGRLRPQKQQQLHNTNLQMTELKKQKTAEQKTRP QNVGIKGIQIYIPTQCVNQSELEKFDGVSQGKYTIGLGQTNMSFVNDRE DIYSMSLTVLSKLIKSYNIDTNKIGRLEVGTETLIDKSKSVKSVLMQLFG ENTDVEGIDTLNACYGGTNALFNSLNWIESNAWDGRDAIVVCGDIAIY DKGAARPTGGAGTVAMWIGPDAPIVFDSVRASYMEHAYDFYKPDFTS EYPYVDGHFSLTCYVKALDQVYKSYSKKAISKGLVSDPAGSDALNVL KYFDYNVFHVPTCKLVTKSYGRLLYNDFRANPQLFPEVDAELATRDY DESLTDKNIEKTFVNVAKPFHKERVAQSLIVPTNTGNMYTASVYAAFA SLLNYVGSDDLQGKRVGLFSYGSGLAASLYSCKIVGDVQHIIKELDITN KLAKRITETPKDYEAAIELRENAHLKKNFKPQGSIEHLQSGVYYLTNID DKFRRSYDVKK

2018256863 11 Apr 2024

SEQ ID	SEQUENCE
SEQ ID NO:25 Erg10p: acetoacetyl CoA thiolase [Saccharomyces cerevisiae]	MSQNVYIVSTARTPIGSFQGSLSSKTAVELGAVALKGALAKVPELDAS KDFDEIIFGNVLSANLGQAPARQVALAAGLSNHIVASTVNKVCASAMK AIILGAQSIKCGNADVVVAGGCESMTNAPYYMPAARAGAKFGQTVLV DGVERDGLNDAYDGLAMGVHAEKCARDWDITREQQDNFAIESYQKS QKSQKEGKFDNEIVPVTIKGFRGKPDTQVTKDEEPARLHVEKLRSART VFQKENGTVTAANASPINDGAAAVILVSEKVLKEKNLKPLAIIKGWGE AAHQPADFTWAPSLAVPKALKHAGIEDINSVDYFEFNEAFSVVGLVNT KILKLDPSKVNVYGGAVALGHPLGCSGARVVVTLLSILQQEGGKIGVA AICNGGGGASSIVIEKI
SEQ ID NO:26 SCFA-TE: Short chain fatty acyl-CoA thioesterase From BMC Biochem. 2011 Aug 10;12:44. doi: 10.1186/1471-2091-12-44., e.g.: Bacteroides sp. (GenBank:CAH09236, Subfamily F)	MSDDKKIGSYKFIAEPFHVDFNGRLTMGVLGNHLLNCAGFHASERGF GIATLNEDNYTWVLSRLAIDLEEMPYQYEEFTVQTWVENVYRLFTDR NFAIIDKDGKKIGYARSVWAMINLNTRKPADLLTLHGGSIVDYVCDEP CPIEKPSRIKVATDQPCAKLTAKYSDIDINGHVNSIRYIEHILDLFPIDLY KSKRIQRFEMAYVAESYYGDELSFFEEEVSENEYHVEIKKNGSEVVCR AKVKFV
SEQ ID NO:27 SCFA-TE: Short chain fatty acyl-CoA thioesterase B. thetaiotaomicron (GenBank:AAO77182, Subfamily F)	MSEENKIGTYQFVAEPFHVDFNGRLTMGVLGNHLLNCAGFHASDRGF GIATLNEDNYTWVLSRLAIELDEMPYQYEKFSVQTWVENVYRLFTDR NFAVIDKDGKKIGYARSVWAMINLNTRKPADLLALHGGSIVDYICDEP CPIEKPSRIKVTSNQPVATLTAKYSDIDINGHVNSIRYIEHILDLFPIELYQ TKRIRRFEMAYVAESYFGDELSFFCDEVSENEFHVEVKKNGSEVVCRS KVIFE
SEQ ID NO:28 SCFA-TE: Short chain fatty acyl-CoA thioesterase Bryantella formatexigen (GenBank: EET61113, Subfamily H)	MIYMAYQYRSRIRYSEIGEDKKLTLPGLVNYFQDCSTFQSEALGIGLDT LGARQRAWLLASWKIVIDRLPRLGEEVVTETWPYGFKGFQGNRNFRM LDQEGHTLAAAASVWIYLNVESGHPCRIDGDVLEAYELEEELPLGPFS RKIPVPEESTERDSFLVMRSHLDTNHHVNNGQYILMAEEYLPEGFKVK QIRVEYRKAAVLHDTIVPFVCTEPQRCTVSLCGSDEKPFAVVEFSE
SEQ ID NO:29 SCFA-TE: Short chain fatty acyl-CoA thioesterase L. brevis (GenBank: ABJ63754, Subfamily J)	MAANEFSETHRVVYYEADDTGQLTLAMLINLFVLVSEDQNDALGLST AFVQSHGVGWVVTQYHLHIDELPRTGAQVTIKTRATAYNRYFAYREY WLLDDAGQVLAYGEGIWVTMSYATRKITTIPAEVMAPYHSEEQTRLP RLPRPDHFDEAVNQTLKPYTVRYFDIDGNGHVNNAHYFDWMLDVLP ATFLRAHHPTDVKIRFENEVQYGHQVTSELSQAAALTTQHMIKVGDLT AVKATIQWDNR
SEQ ID NO:30 SCFA-TE: Short chain fatty acyl-CoA thioesterase L. plantarum	MATLGANASLYSEQHRITYYECDRTGRATLTTLIDIAVLASEDQSDAL GLTTEMVQSHGVGWVVTQYAIDITRMPRQDEVVTIAVRGSAYNPYFA YREFWIRDADGQQLAYITSIWVMMSQTTRRIVKILPELVAPYQSEVVK RIPRLPRPISFEATDTTITKPYHVRFFDIDPNRHVNNAHYFDWLVDTLPA TFLLQHDLVHVDVRYENEVKYGQTVTAHANILPSEVADQVTTSHLIEV

SEQ ID	SEQUENCE
(GenBank:CAD63310, Subfamily J)	DDEKCCEVTIQWRTLPEPIQ
SEQ ID NO:31 SCFA-TE: Short chain fatty acyl-CoA thioesterase Streptococcus dysgalactiae (GenBank:BAH81730, Subfamily J)	MGLSYREDIKLPFELCDVKSDIKFPLLLDYCLTVSGRQSAQLGRSNDYL LEQYGLIWIVTDYEATIHRLPHFQETITIETKALSYNKFFCYRQFYIYDQ EGGLLVDILAYFALLNPDTRKVATIPEDLVAPFETDFVKKLHRVPKMP LLEQSIDRDYYVRYFDIDMNGHVNNSKYLDWMYDVLGCEFLKTHQPL KMTLKYVKEVSPGGQITSSYHLDQLTSYHQITSDGQLNAQAMIEWRAI KQTESEID
SEQ ID NO:32 DXS 1-deoxy-D-xylulose-5-phosphate synthase gene (dxs-AC# 16128405) Escherichia coli	MSFDIAKYPTLALVDSTQELRLLPKESLPKLCDELRRYLLDSVSRSSGH FASGLGTVELTVALHYVYNTPFDQLIWDVGHQAYPHKILTGRRDKIGT IRQKGGLHPFPWRGESEYDVLSVGHSSTSISAGIGIAVAAEKEGKNRRT VCVIGDGAITAGMAFEAMNHAGDIRPDMLVILNDNEMSISENVGALN NHLAQLLSGKLYSSLREGGKKVFSGVPPIKELLKRTEEHIKGMVVPGT LFEELGFNYIGPVDGHDVLGLITTLKNMRDLKGPQFLHIMTKKGRGYE PAEKDPITFHAVPKFDPSSGCLPKSSGGLPSYSKIFGDWLCETAAKDNK LMAITPAMREGSGMVEFSRKFPDRYFDVAIAEQHAVTFAAGLAIGGYK PIVAIYSTFLQRAYDQVLHDVAIQKLPVLFAIDRAGIVGADGQTHQGAF DLSYLRCIPEMVIMTPSDENECRQMLYTGYHYNDGPSAVRYPRGNAV GVELTPLEKLPIGKGIVKRRGEKLAILNFGTLMPEAAKVAESLNATLVD MRFVKPLDEALILEMAASHEALVTVEENAIMGGAGSGVNEVLMAHRK PVPVLNIGLPDFFIPQGTQEEMRAELGLDAAGMEAKIKAWLA
SEQ ID NO:33 DXR/IspC 1-deoxy-D-xylulose 5-phosphate reductoisomerase [Escherichia coli] AC# 16128166	MKQLTILGSTGSIGCSTLDVVRHNPEHFRVVALVAGKNVTRMVEQCL EFSPRYAVMDDEASAKLLKTMLQQQGSRTEVLSGQQAACDMAALED VDQVMAAIVGAAGLLPTLAAIRAGKTILLANKESLVTCGRLFMDAVK QSKAQLLPVDSEHNAIFQSLPQPIQHNLGYADLEQNGVVSILLTGSGGP FRETPLRDLATMTPDQACRHPNWSMGRKISVDSATMMNKGLEYIEAR WLFNASASQMEVLIHPQSVIHSMVRYQDGSVLAQLGEPDMRTPIAHT MAWPNRVNSGVKPLDFCKLSALTFAAPDYDRYPCLKLAMEAFEQGQ AATTALNAANEITVAAFLAQQIRFTDIAALNLSVLEKMDMREPQCVDD VLSVDANAREVARKEVMRLAS
SEQ ID NO:34 IspD 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase [Escherichia coli] AC# 190908496	MATTHLDVCAVVPAAGFGRRMQTECPKQYLSIGNQTILEHSVHALLA HPRVKRVVIAISPGDSRFAQLPLANHPRITVVDGGEERADSVLAGLKA AGDAQWVLVHDAARPCLHQDDLARLLALSETSRTGGILAAPVRDTM KRAEPGKNAIAHTVDRNGLWHALTPQFFPRELLHDCLTRALNEGATIT DEASALEYCGFHPQLVEGRADNIKVTRPEDLALAEFYLTRTIHQENT
SEQ ID NO:35 IspE 4-diphosphocytidyl-2-C-methylerythritol kinase [Escherichia coli] AC# 4062791	MRTQWPSPAKLNLFLYITGQRADGYHTLQTLFQFLDYGDTISIELRDD GDIRLLTPVEGVEHEDNLIVRAARLLMKTAADSGRLPTGSGANISIDKR LPMGGGLGGGSSNAATVLVALNHLWQCGLSMDELAEMGLTLGADVP VFVRGHAAFAEGVGEILTPVDPPEKWYLVAHPGVSIPTPVIFKDPELPR NTPKRSIETLLKCEFSNDCEVIARKRFREVDAVLSWLLEYAPSRLTGTG ACVFAEFDTESEARQVLEQAPEWLNGFVAKGANLSPLHRAML
SEQ ID NO:36 IspF 2C-methyl-D-erythritol 2,4-	MRIGHGFDVHAFGGEGPIIIGGVRIPYEKGLLAHSDGDVVLHALTDALL GAAALGDIGKLFPDTDPAFKGADSRELLREAWRRIQAKGYALGNVDV TIIAQAPRMLPHIPQMRVFIAEDLGCHMDDVNVKATTTEKLGFTGRGE

2018256863 11 Apr 2024

SEQ ID	SEQUENCE
cyclodiphosphate synthase [Escherichia coli F11] AC# 190908583	GIACEAVALLIKATK
SEQ ID NO:37 IspG 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase [Escherichia coli F11] CDU37657	MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQI KALERVGADIVRVSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALK VAEYGVDCLRINPGNIGNEERIRMVVDCARDKNIPIRIGVNAGSLEKDL QEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASDVFLAVESY RLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLA ADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLE DIITPMDVSIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLD NNDMIDQLEARIRAKASQLDEARRIDVQQVEK
SEQ ID NO:38 IspH 4-hydroxy-3-methylbut-2-enyl diphosphate reductase [Escherichia coli F11] AC# 190905591	MQILLANPRGFCAGVDRAISIVENALAIYGAPIYVRHEVVHNRYVVDS LRERGAIFIEQISEVPDGAILIFSAHGVSQAVRNEAKSRDLTVFDATCPL VTKVHMEVARASRRGEESILIGHAGHPEVEGTMGQYSNPEGGMYLVE SPDDVWKLTVKNEEKLSFMTQTTLSVDDTSDVIDALRKRFPKIVGPRK DDICYATTNRQEAVRALAEQAEVVLVVGSKNSSNSNRLAELAQRMGK RAFLIDDATDIQEEWVKEAKCVGVTAGASAPDILVQNVVARLQQLGG GEAIPLEGREENIVFEVPKELRVDIREVD
SEQ ID NO:39 IDI: isopentenyl diphosphate (IPP) isomerase (GenBank AKF73239) Escherichia coli	MQTEHVILLNAQGVPTGTLEKYAAHTADTRLHLAFSSWLFNAKGQLL VTRRALSKKAWPGVWTNSVCGHPQLGESNEDAVIRRCRYELGVEITPP ESIYPDFRYRATDPSGIVENEVCPVFAARTTSALQINDDEVMDYQWCD LADVLHGIDATPWAFSPWMVMQATNREARKRLSAFTQLK
SEQ ID NO:40 Mutated IspA* FPP synthase (S81F) for GPP production (ispA-AC# NP_414955)	MDFPQQLEACVKQANQALSRFIAPLPFQNTPVVETMQYGALLGGKRL RPFLVYATGHMFGVSTNTLDAPAAAVECIHAYSLIHDDLPAMDDDDL RRGLPTCHVKFGEANAILAGDALQTLAFSILSDADMPEVSDRDRISMIS ELASASGIAGMCGGQALDLDAEGKHVPLDALERIHRHKTGALIRAAV RLGALSAGDKGRRALPVLDKYAESIGLAFQVQDDILDVVGDTATLGK RQGADQQLGKSTYPALLGLEQARKKARDLIDDARQSLKQLAEQSLDT SALEALADYIIQRNK
SEQ ID NO:41 Malonyl CoA-acyl carrier protein transacylase (MCT1) Saccharomyces cerevisiae	atgaagctactaaccttcccaggtcaagggacctccatctccatttcgatattaaaagcgataataagaaacaaat caagagaattccaaacaatactgagtcagaacggcaaggaatcaaatgatctattgcagtacatcttccagaacc cttccagccccggaagcattgcagtctgctccaacctttctatcaattgtaccagatactctcgaatccttctgatcc tcaagatcaagcaccaaaaaatatgactaagatcgattcccccgacaagaaagacaatgaacaatgttacctttg ggtcactcgctaggcgagttaacatgtctgagtgttaattcactgttttcgttaaaggatcttttgatattgctaatttta gaaataagttaatggtaacatctactgaaaagtacttagtagcccacaatatcaacagatccaacaaatttgaaatg tgggcactctcttctccgagggccacagatttaccgcaagaagtgcaaaaactactaaattcccctaatttattatc atcttcacaaaataccatttctgtagcaaatgcaaattcagtaaagcaatgtgtagtcaccggtctggttgatgattta gagtccttaagaacagaattgaacttaaggttcccgcgtttaagaattacagaattaactaacccatacaacatccc cttccataatagcactgtgttgaggcccgttcaggaaccactctatgactacatttgggatatattaaagaaaaacg gaactcacacgttgatggagttgaaccatccaataatagctaacttagatggtaatatatcttactatattcatcatgc cctagatagattcgttaagtgttcaagcaggactgtgcaattcaccatgtgttatgataccataaactctggaaccc cagtggaaattgataagagtatttgctttggcccgggcaatgtgatttataaccttattcggagaaattgtccccaag tggacactatagaatacacctctttagcaactatagacgcttatcacaaggcggcagaggagaacaaagattga
SEQ ID NO:42 malonyl CoA-acyl carrier protein transacylase (MCT1)	MKLLTFPGQGTSISISILKAIIRNKSREFQTILSQNGKESNDLLQYIFQNPS SPGSIAVCSNLFYQLYQILSNPSDPQDQAPKNMTKIDSPDKKDNEQCYL LGHSLGELTCLSVNSLFSLKDLFDIANFRNKLMVTSTEKYLVAHNINRS NKFEMWALSSPRATDLPQEVQKLLNSPNLLSSSQNTISVANANSVKQC VVTGLVDDLESLRTELNLRFPRLRITELTNPYNIPFHNSTVLRPVQEPLY

2018256863 11 Apr 2024

SEQ ID	SEQUENCE
Saccharomyces cerevisiae	DYIWDILKKNGTHTLMELNHPIIANLDGNISYYIHHALDRFVKCSSRTV QFTMCYDTINSGTPVEIDKSICFGPGNVIYNLIRRNCPQVDTIEYTSLATI DAYHKAAEENKD*
SEQ ID NO:43 Artificial beta-ketothiolase (BktB) nucleotide sequence	atgactagagaagttgtcgtcgtttccggtgtccgtaccgctatcggtactttcggtggttccttaaaggatgttgct cctgctgaattgggtgctttagttgttagagaagctttggccagagcccaagtctccggtgacgacgttggtcacg tcgttttcggtaacgtcatccaaactgaaccacgtgacatgtacttgggtagagtcgccgctgttaacggtggtgtc accatcaacgctcctgccttaactgttaacagattatgtggttccggtttacaagctattgtctctgccgcccaaact atcttgttgggtgatactgacgttgctattggtggtggtgctgaatctatgtctagagctccatacttggctccagctg cccgttggggtgctagaatgggtgacgccggtttggtcgatatgatgttgggtgccttgcatgatcctttccacag aatccacatgggtgttaccgctgaaaacgttgctaaggaatacgatatctctagagctcaacaagatgaagccgc tttagaatctcacagacgtgcctccgccgctattaaggctggttacttcaaggaccaaattgttccagttgtctctaa gggtcgtaaaggtgatgttacctttgatactgacgaacacgttagacacgacgccactattgacgatatgactaaa ttaagaccagtctttgttaaggagaatggtaccgttactgctggtaacgcttctggtttgaacgatgccgccgctgc cgttgttatgatggaaagagctgaagccgaaagacgtggtttaaagccattggccagattagtctcctacggtca cgctggtgtcgacccaaaggctatgggtatcggtccagttcctgctactaagattgctttagaaagagctggtttg caagtttctgacttggacgtcatcgaagccaacgaagccttcgctgctcaagcttgtgctgtcaccaaggctttgg gtttggatccagctaaagttaaccctaatggttctggtatttccttgggtcacccaatcggtgctaccggtgctttaat cactgttaaagccttacacgaattgaacagagttcaaggtagatacgctttggtcactatgtgcatcggtggtggtc aaggtatcgctgctatcttcgaaagaatcggatcctaa
SEQ ID NO:44 Engineered beta-ketothiolase (BktB)	MTREVVVVSGVRTAIGTFGGSLKDVAPAELGALVVREALARAQVSGD DVGHVVFGNVIQTEPRDMYLGRVAAVNGGVTINAPALTVNRLCGSGL QAIVSAAQTILLGDTDVAIGGGAESMSRAPYLAPAARWGARMGDAGL VDMMLGALHDPFHRIHMGVTAENVAKEYDISRAQQDEAALESHRRAS AAIKAGYFKDQIVPVVSKGRKGDVTFDTDEHVRHDATIDDMTKLRPV FVKENGTVTAGNASGLNDAAAAVVMMERAEAERRGLKPLARLVSYG HAGVDPKAMGIGPVPATKIALERAGLQVSDLDVIEANEAFAAQACAV TKALGLDPAKVNPNGSGISLGHPIGATGALITVKALHELNRVQGRYAL VTMCIGGGQGIAAIFERIGS*
SEQ ID NO:45 Artificial PaaH1: 3-hydroxyacyl-CoA dehydrogenase nucleotide sequence	atgtccatcagaactgtcggtattgttggtgctggtactatgggtaacggtattgctcaagcctgtgctgtcgtcggt ttgaacgtcgtcatggtcgacatttctgacgctgctgttcaaaagggtgttgctactgtcgcttcctctttggacagat taattaagaaggaaaagttgaccgaagccgacaaggcctctgccttggccagaattaagggttccacttcttatg acgacttgaaagctaccgacattgttatcgaagctgctactgaaaactacgatttgaaagttaagatcttgaagcaa attgatggtatcgtcggtgagaacgtcattattgcttctaacacttcctccatttctatcactaaattagccgccgtca cctctagagccgacagatttatcggtatgcacttctttaatccagttccagtcatggctttggtcgaattaattagagg tttgcaaacctccgacaccacccacgccgccgttgaagctttgtctaagcaattgggtaagtacccaatcaccgtt aaaaattccccaggtttcgttgtcaaccgtattttgtgcccaatgatcaatgaagctttctgtgtcttgggtgagggtt tggcctccccagaagaaatcgatgaaggtatgaagttaggttgtaaccaccctattggtcctttagccttggccga catgatcggtttagacactatgttggccgttatggaagtcttgtacactgaattcgctgacccaaagtacagaccag ctatgttaatgagagaaatggttgctgccggttatttgggtagaaagactggtcgtggtgtttatgtctactctaaag ggatc
SEQ ID NO:46 Engineered PaaH1: 3-hydroxyacyl-CoA dehydrogenase	MSIRTVGIVGAGTMGNGIAQACAVVGLNVVMVDISDAAVQKGVATV ASSLDRLIKKEKLTEADKASALARIKGSTSYDDLKATDIVIEAATENYD LKVKILKQIDGIVGENVIIASNTSSISITKLAAVTSRADRFIGMHFFNPVP VMALVELIRGLQTSDTTHAAVEALSKQLGKYPITVKNSPGFVVNRILCP MINEAFCVLGEGLASPEEIDEGMKLGCNHPIGPLALADMIGLDTMLAV MEVLYTEFADPKYRPAMLMREMVAAGYLGRKTGRGVYVYSKGI
SEQ ID NO:47 Artificial crotonase (Crt) nucleotide sequence	atggaattgaacaacgttatttttggaaaaggaaggtaaggtcgctgtcgttactatcaacagaccaaaggctttaa acgctttgaactctgacaccttgaaagaaatggattatgttatcggtgaaatcgaaaatgactctgaagttttggcc gttatcttgactggtgctggtgaaaaatctttcgttgctggtgctgacatttctgaaatgaaggagatgaataccatt gaaggtagaaagttcggtatcttgggtaacaaggtttttagaagattggaattgttggaaaaaccagtcatcgctg ctgttaacggtttcgctttaggtggtggttgtgaaatcgctatgtcctgtgacattcgtatcgcctcctccaatgctag attcggtcaaccagaagttggtttaggtattactccaggtttcggtggtacccaaagattgtctagattggtcggtat

SEQ ID	SEQUENCE
	gggtatggctaagcaattaattttcactgctcaaaacattaaggctgatgaagccttacgtattggtttggtcaacaa ggtcgttgaaccatctgaattgatgaataccgctaaggaaattgctaacaaaattgtttctaatgccccagttgctgt caagttgtccaagcaagctattaacagaggtatgcaatgtgatattgacactgctttggctttcgaatccgaagcttt tggtgaatgtttttctaccgaagatcaaaaggatgctatgaccgctttcatcgagaagagaaagatcgaaggtttc aaaaacagaggatcctaa
SEQ ID NO:48 Engineered crotonase (Crt)	MELNNVILEKEGKVAVVTINRPKALNALNSDTLKEMDYVIGEIENDSE VLAVILTGAGEKSFVAGADISEMKEMNTIEGRKFGILGNKVFRRLELLE KPVIAAVNGFALGGGCEIAMSCDIRIASSNARFGQPEVGLGITPGFGGT QRLSRLVGMGMAKQLIFTAQNIKADEALRIGLVNKVVEPSELMNTAK EIANKIVSNAPVAVKLSKQAINRGMQCDIDTALAFESEAFGECFSTEDQ KDAMTAFIEKRKIEGFKNRGS*
SEQ ID NO:49 Artificial Ter: trans-2-enoyl-CoA reductase nucleotide sequence	atgattgtcaaaccaatggttcgtaacaacatttgtttaaatgcccacccacaaggttgtaagaagggtgttgaaga tcaaatcgaatacactaaaaagagaattaccgctgaagttaaagctggtgctaaggccccaaagaacgttttggtt ttgggttgttccaacggttacggtttggcctccagaattactgctgcttttggttacggtgccgctaccatcggtgtct ctttcgaaaaggccggttccgaaactaagtacggtactccaggttggtacaataacttggctttcgatgaagctgc taagagagaagaggtttgtattccgttactattgacggtgatgccttttctgacgaaatcaaagctcaagtcatcgaag aagccaaaaagaaaggtatcaagttcgatttgattgtctactctttagcctctcctgttagaactgatccagatactg gtattatgcacaaatccgttttgaagccattcggtaagaccttcactggtaaaactgtcgatcctttcactggtgaatt aaaggaaatctctgctgaacctgccaacgacgaagaagctgctgccactgttaaggttatgggtggtgaagact gggaaagatggatcaagcaattatctaaggaaggtttgttggaagaaggttgtatcaccttggcttactcttacatc ggtccagaagctacccaagctttgtacagaaagggtaccattggtaaggctaaagaacacttggaggctactgc tcatagattgaacaaggaaaatccatccatcagagcctttgtttccgtcaataaaggtttggtcactagagcctctg ccgtcattccagttatccctttatacttggcttctttgtttaaagtcatgaaggaaaagggtaaccatgaaggttgtat cgaacaaatcactcgtttgtacgctgaacgtttatacagaaaggacggtaccatccctgtcgatgaagaaaacag aatcagaatcgacgattgggaattggaagaagatgttcaaaaagccgtttccgccttgatggaaaaggtcaccg gtgaaaatgccgaatccttgactgacttagctggttacagacatgacttttagcttctaatggtttcgatgttgaagg tattaactatgaggctgaagtcgaaagatttgacagaatcggatcctaa
SEQ ID NO:50 Engineered Ter: trans-2-enoyl-CoA reductase	MIVKPMVRNNICLNAHPQGCKKGVEDQIEYTKKRITAEVKAGAKAPK NVLVLGCSNGYGLASRITAAFGYGAATIGVSFEKAGSETKYGTPGWY NNLAFDEAAKREGLYSVTIDGDAFSDEIKAQVIEEAKKKGIKFDLIVYS LASPVRTDPDTGIMHKSVLKPFGKTFTGKTVDPFTGELKEISAEPANDE EAAATVKVMGGEDWERWIKQLSKEGLLEEGCITLAYSYIGPEATQAL YRKGTIGKAKEHLEATAHRLNKENPSIRAFVSVNKGLVTRASAVIPVIP LYLASLFKVMKEKGNHEGCIEQITRLYAERLYRKDGTIPVDEENRIRID DWELEEDVQKAVSALMEKVTGENAESLTDLAGYRHDFLASNGFDVE GINYEAEVERFDRIGS*
SEQ ID NO:51 Truncated 3-hydroxy-3-methyl-glutaryl-CoA reductase (tHMG1)	atggcgcgtgaccaattggtgaaaactgaagtcaccaagaagtcttttactgctcctgtacaaaaggcttctacac cagttttaaccaataaaacagtcatttctggatcgaaagtcaaaagtttatcatctgcgcaatcgagctcatcagga ccttcatcatctagtgaggaagatgattcccgcgatattgaaagcttggataagaaaatacgtcctttagaagaatt agaagcattattaagtagtggaaatacaaaacaattgaagaacaaagaggtcgctgccttggttattcacggtaa gttacctttgtacgctttggagaaaaaattaggtgatactacgagagcggttgcggtacgtaggaaggctctttca attttggcagaagctcctgtattagcatctgatcgtttaccatataaaaattatgactacgaccgcgtatttggcgctt gttgtgaaaatgttataggttacatgcctttgcccgttggtgttataggccccttggttatcgatggtacatcttatcat ataccaatggcaactacagagggttgtttggtagcttctgccatgcgtggctgtaaggcaatcaatgctggcggt ggtgcaacaactgttttaactaaggatggtatgacaagaggcccagtagtccgtttcccaactttgaaaagatctg gtgcctgtaagatatggttagactcagaagagggacaaaacgcaattaaaaaagcttttaactctacatcaagattt gcacgtctgcaacatattcaaacttgtctagcaggagatttgtt
SEQ ID NO:52 Truncated 3-hydroxy-3-methyl-glutaryl-CoA reductase	MARDQLVKTEVTKKSFTAPVQKASTPVLTNKTVISGSKVKSLSSAQSS SSGPSSSSEEDDSRDIESLDKKIRPLEELEALLSSGNTKQLKNKEVAALV IHGKLPLYALEKKLGDTTRAVAVRRKALSILAEAPVLASDRLPYKNYD YDRVFGACCENVIGYMPLPVGVIGPLVIDGTSYHIPMATTEGCLVASA MRGCKAINAGGGATTVLTKDGMTRGPVVRFPTLKRSGACKIWLDSEE

2018256863 11 Apr 2024

SEQ ID	SEQUENCE
(tHMG1)	GQNAIKKAFNSTSRFARLQHIQTCLAGDL
SEQ ID NO:53 Artificial acetyl-CoA acetyltransferase/ HMG-CoA reductase (mvaE) nucleotide sequence	atgaagactgtcgttatcatagatgccttgagaacaccaatcggtaaatacaaaggttcattatcccaagtttccgc cgttgacttaggtactcatgttactacacaattgttgaagagacactccacaatcagtgaagaaatcgatcaagtca tattcggtaacgtattgcaagctggtaatggtcaaaacccagccagacaaatagctatcaattctggtttatcacat gaaattcctgctatgacagtaaacgaagtttgtggttcaggcatgaaagcagtcattttggccaagcaattgataca attaggtgaagcagaagttttaatcgccggtggtatagaaaacatgagtcaagctccaaaattgcaaagattcaat tacgaaactgaatcttacgatgcacctttctcttcgatgatgtatgatggtttgactgacgctttttctggtcaagcaat gggtttaacagctgaaaatgtcgcagaaaagtaccatgtaaccagagaagaacaagatcaattttccgttcacag tcaattaaaagctgcacaagcacaagccgaaggtattttcgccgacgaaatagctccattggaagtttctggtaca ttagtcgaaaaggatgaaggtattagacctaactccagtgttgaaaaattgggtactttgaagacagtattcaagg aagacggtacagttaccgctggtaatgcctctaccattaacgatggtgctagtgcattgattatagcttctcaagaa tatgccgaagctcatggtttgccatacttagctatcattagagatagtgtagaagttggtattgacccagcatacatg ggtatctctcctataaaagcaatccaaaagttgttagccagaaaccaattgaccactgaagaaattgatttgtacga aattaacgaagcatttgccgctacatcaatcgttgtccaaagagaattggcattgccagaagaaaaggttaacatc tatggtggtggtatctccttgggtcacgctataggtgcaaccggtgccagattgttgacttccttaagttaccaattg aaccaaaaggaaaagaaatacggtgttgcttctttatgcattggtggtggtttgggtttagcaatgttgttagaaaga ccacaacaaaagaaaaattctagattctaccaaatgtcccctgaagaaagattggcctcattgttaaatgaaggtc aaatttccgcagatactaagaaagaatttgaaaacaccgctttatcttcacaaatcgcaaaccatatgatcgaaaac caaatctctgaaacagaagttccaatgggtgtcggtttgcacttaactgtcgatgaaacagactatttggtaccaat ggctaccgaagaacctagtgttatcgcagccttatctaatggtgctaagatagcacaaggttttaagactgttaacc aacaaagattgatgagaggtcaaatcgtattctacgatgttgctgacccagaatcattaatcgataagttgcaagta agagaagccgaagtttttcaacaagctgaattgtcttacccttcaatagttaagagaggtggtggtttgagagattt gcaatacagaacttttgacgaatccttcgtcagtgtagatttcttagttgatgtcaaggacgccatgggtgctaatat tgttaacgcaatgttggaaggtgtcgccgaattgtttagagaatggttcgctgaacaaaagattttgttttctatcttgt caaactacgctacagaatctgtagttaccatgaaaactgcaattccagtttccagattgagtaagggttctaacggt agagaaatcgctgaaaagattgttttggcatcaagatatgcctccttagacccttacagagctgttactcataataa gggtataatgaacggtatcgaagctgtcgtattagcaaccggtaatgatactagagcagtatctgcctcatgtcac gcattcgccgttaaggaaggtagataccaaggtttgacatcatggaccttggatggtgaacaattaattggtgaaa tatccgttccattggctttagcaactgttggtggtgctacaaaagtcttgcctaagagtcaagctgcagccgatttgt tagccgtcactgacgctaaggaattgtctagagttgtcgctgcagtaggtttagctcaaaatttggccgctttaaga gcattggtttcagaaggtattcaaaaaggtcatatggctttgcaagcaagatccttagccatgacagttggtgctac cggtaaagaagtcgaagccgtagctcaacaattaaaaagacaaaagacaatgaaccaagacagagcaatggc tatattaaacgatttgagaaagcaataa
SEQ ID NO:54 MvaE: acetyl-CoA acetyltransferase/ HMG-CoA reductase (Enterococcus sp.)	MKTVVIIDALRTPIGKYKGSLSQVSAVDLGTHVTTQLLKRHSTISEEID QVIFGNVLQAGNGQNPARQIAINSGLSHEIPAMTVNEVCGSGMKAVIL AKQLIQLGEAEVLIAGGIENMSQAPKLQRFNYETESYDAPFSSMMYDG LTDAFSGQAMGLTAENVAEKYHVTREEQDQFSVHSQLKAAQAQAEGI FADEIAPLEVSGTLVEKDEGIRPNSSVEKLGTLKTVFKEDGTVTAGNAS TINDGASALIIASQEYAEAHGLPYLAIIRDSVEVGIDPAYMGISPIKAIQK LLARNQLTTEEIDLYEINEAFAATSIVVQRELALPEEKVNIYGGGISLGH AIGATGARLLTSLSYQLNQKEKKYGVASLCIGGGLGLAMLLERPQQK KNSRFYQMSPEERLASLLNEGQISADTKKEFENTALSSQIANHMIENQI SETEVPMGVGLHLTVDETDYLVPMATEEPSVIAALSNGAKIAQGFKTV NQQRLMRGQIVFYDVADPESLIDKLQVREAEVFQQAELSYPSIVKRGG GLRDLQYRTFDESFVSVDFLVDVKDAMGANIVNAMLEGVAELFREWF AEQKILFSILSNYATESVVTMKTAIPVSRLSKGSNGREIAEKIVLASRYA SLDPYRAVTHNKGIMNGIEAVVLATGNDTRAVSASCHAFAVKEGRYQ GLTSWTLDGEQLIGEISVPLALATVGGATKVLPKSQAAADLLAVTDAK ELSRVVAAVGLAQNLAALRALVSEGIQKGHMALQARSLAMTVGATG KEVEAVAQQLKRQKTMNQDRAMAILNDLRKQ*
SEQ ID NO:55 MvaS: HMG-CoA synthase	atgacaattgggattgataaaattagtttttttgtgccccttattatattgatatgacggcactggctgaagccagaa atgtagaccctggaaaatttcatattggtattgggcaagaccaaatggcggtgaacccaatcagccaagatattgt gacatttgcagccaatgccgcagaagcgatcttgaccaaagaagataaagaggccattgatatggtgattgtcg

2018256863 11 Apr 2024

SEQ ID	SEQUENCE
(Enterococcus sp.)	ggactgagtccagtatcgatgagtcaaaagcggccgcagttgtcttacatcgtttaatggggattcaacctttcgct cgctctttcgaaatcaaggaagcttgttacggagcaacagcaggcttacagttagctaagaatcacgtagccttac atccagataaaaaagtcttggtcgtagcggcagatattgcaaaatatggcttaaattctggcggtgagcctacaca aggagctggggcggttgcaatgttagttgctagtgaaccgcgcattttggctttaaaagaggataatgtgatgctg acgcaagatatctatgacttttggcgtccaacaggccacccgtatcctatggtcgatggtcctttgtcaaacgaaa cctacatccaatcttttgcccaagtctgggatgaacataaaaaacgaaccggtcttgattttgcagattatgatgcttt agcgttccatattccttacacaaaaatgggcaaaaaagccttattagcaaaaatctccgaccaaactgaagcaga acaggaacgaattttagcccgttatgaagaaagtatcgtctatagtcgtcgcgtaggaaacttgtatacgggttca ctttatctgggactcatttccctttagaaaatgcaacgactttaaccgcaggcaatcaaattggtttattcagttatgg ttctggtgctgtcgctgaatttttcactggtgaattagtagctggttatcaaaatcatttacaaaaagaaactcatttag cactgctggataatcggacagaactttctatcgctgaatatgaagccatgtttgcagaaactttagacacagacatt gatcaaacgttagaagatgaattaaaatatagtatttctgctattaataataccgttcgttcttatcgaaactaa
SEQ ID NO:56 MvaS: HMG-CoA synthase (Enterococcus sp.)	MTIGIDKISFFVPPYYIDMTALAEARNVDPGKFHIGIGQDQMAVNPISQ DIVTFAANAAEAILTKEDKEAIDMVIVGTESSIDESKAAAVVLHRLMGI QPFARSFEIKEACYGATAGLQLAKNHVALHPDKKVLVVAADIAKYGL NSGGEPTQGAGAVAMLVASEPRILALKEDNVMLTQDIYDFWRPTGHP YPMVDGPLSNETYIQSFAQVWDEHKKRTGLDFADYDALAFHIPYTKM GKKALLAKISDQTEAEQERILARYEESIVYSRRVGNLYTGSLYLGLISLL ENATTLTAGNQIGLFSYGSGAVAEFFTGELVAGYQNHLQKETHLALLD NRTELSIAEYEAMFAETLDTDIDQTLEDELKYSISAINNTVRSYRN*
SEQ ID NO:57 Isopentenyl pyrophosphate isomerase (Sc_IDI1) Saccharomyces sp.	atgactgccgacaacaatagtatgccccatggtgcagtatctagttacgccaaattagtgcaaaaccaaacacct gaagacattttggaagagtttcctgaaattattccattacaacaaagacctaatacccgatctagtgagacgtcaaa tgacgaaagcggagaaacatgtttttctggtcatgatgaggagcaaattaagttaatgaatgaaaattgtattgtttt ggattgggacgataatgctattggtgccggtaccaagaaagtttgtcatttaatggaaaatattgaaaagggtttac tacatcgtgcattctccgtctttattttcaatgaacaaggtgaattacttttacaacaaagagccactgaaaaaataac tttccctgatctttggactaacacatgctgctctcatccactatgtattgatgacgaattaggtttgaagggtaagcta gacgataagattaagggcgctattactgcggcggtgagaaaactagatcatgaattaggtattccagaagatgaa actaagacaaggggtaagtttcacttttaaacagaatccattacatggcaccaagcaatgaaccatggggtgaa catgaaattgattacatcctattttataagatcaacgctaaagaaaacttgactgtcaacccaaacgtcaatgaagtt agagacttcaaatgggtttcaccaaatgatttgaaaactatgtttgctgacccaagttacaagtttacgccttggttta agattatttgcgagaattacttattcaactggtgggagcaattagatgacctttctgaagtggaaaatgacaggcaa attcatagaatgctataa
SEQ ID NO:58 Isopentenyl pyrophosphate isomerase (Sc_IDI1) Saccharomyces sp.	MTADNNSMPHGAVSSYAKLVQNQTPEDILEEFPEIIPLQQRPNTRSSET SNDESGETCFSGHDEEQIKLMNENCIVLDWDDNAIGAGTKKVCHLME NIEKGLLHRAFSVFIFNEQGELLLQQRATEKITFPDLWTNTCCSHPLCID DELGLKGKLDDKIKGAITAAVRKLDHELGIPEDETKTRGKFHFLNRIHY MAPSNEPWGEHEIDYILFYKINAKENLTVNPNVNEVRDFKWVSPNDLK TMFADPSYKFTPWFKIICENYLFNWWEQLDDLSEVENDRQIHRML
SEQ ID NO:59 Mutant farnesyl pyrophosphate synthase (Erg20mut, F96W, N127W)	atggcttcagaaaaagaaattaggagagagagattcttgaacgttttccctaaattagtagaggaattgaacgcat cgcttttggcttacggtatgcctaaggaagcatgtgactggtatgcccactcattgaactacaacactccaggcgg taagctaaatagaggtttgtccgttgtggacacgtatgctattctctccaacaagaccgttgaacaattggggcaa gaagaatacgaaaaggttgccattctaggttggtgcattgagttgttgcaggcttactggttggtcgccgatgatat gatggacaagtccattaccagaagaggccaaccatgttggtacaaggttcctgaagttggggaaattgccatctg ggacgcattcatgttagaggctgctatctacaagcttttgaaatctcacttcagaaacgaaaaatactacatagatat caccgaattgttccatgaggtcaccttccaaaccgaattgggccaattgatggacttaatcactgcacctgaagac aaagtcgacttgagtaagttctccctaaagaagcactccttcatagttactttcaagactgcttactattctttctactt gcctgtcgcattggccatgtacgttgccggtatcacggatgaaaaggatttgaaacaagccagagatgtcttgatt ccattgggtgaatacttccaaattcaagatgactacttagactgcttcggtaccccagaacagatcggtaagatcg gtacagatatccaagataacaaatgttcttgggtaatcaacaaggcattggaacttgcttccgcagaacaaagaa agactttagacgaaaattacggtaagaaggactcagtcgcagaagccaaatgcaaaaagattttcaatgacttga aaattgaacagctataccacgaatatgaagagtctattgccaaggatttgaaggccaaaatttctcaggtcgatga gtctcgtggcttcaaagctgatgtcttaactgcgttcttgaacaaagtttacaagagaagcaaatag

SEQ ID	SEQUENCE
SEQ ID NO:60 Mutant farnesyl pyrophosphate synthase (Erg20mut, F96W, N127W)	MASEKEIRRERFLNVFPKLVEELNASLLAYGMPKEACDWYAHSLNYN TPGGKLNRGLSVVDTYAILSNKTVEQLGQEEYEKVAILGWCIELLQAY WLVADDMMDKSITRRGQPCWYKVPEVGEIAIWDAFMLEAAIYKLLKS HFRNEKYYIDITELFHEVTFQTELGQLMDLITAPEDKVDLSKFSLKKHS FIVTFKTAYYSFYLPVALAMYVAGITDEKDLKQARDVLIPLGEYFQIQD DYLDCFGTPEQIGKIGTDIQDNKCSWVINKALELASAEQRKTLDENYG KKDSVAEAKCKKIFNDLKIEQLYHEYEESIAKDLKAKISQVDESRGFKA DVLTAFLNKVYKRSK*
SEQ ID NO:61 Phosphomevalonate kinase (Sc_ERG8) Saccharomyces sp.	atgtcagagttgagagccttcagtgccccagggaaagcgttactagctggtggatatttagttttagatacaaaata tgaagcatttgtagtcggattatcggcaagaatgcatgctgtagcccatccttacggttcattgcaagggtctgata agtttgaagtgcgtgtgaaaagtaaacaatttaaagatggggagtggctgtaccatataagtcctaaaagtggctt cattcctgtttcgataggcggatctaagaaccctttcattgaaaaagttatcgctaacgtatttagctactttaaacctа acatggacgactactgcaatagaaacttgttcgttattgatattttctctgatgatgcctaccattctcaggaggatag cgttaccgaacatcgtggcaacagaagattgagttttcattcgcacagaattgaagaagttcccaaaacagggct gggctcctcggcaggtttagtcacagttttaactacagctttggcctccttttttgtatcggacctggaaaataatgta gacaaatatagagaagttattcataatttagcacaagttgctcattgtcaagctcagggtaaaattggaagcgggtt tgatgtagcggcggcagcatatggatctatcagatatagaagattcccacccgcattaatctctaatttgccagata ttggaagtgctacttacggcagtaaactggcgcatttggttgatgaagaagactggaatattacgattaaaagtaa ccatttaccttcgggattaactttatggatgggcgatattaagaatggttcagaaacagtaaaactggtccagaag gtaaaaaattggtatgattcgcatatgccagaaagcttgaaaatatatacagaactcgatcatgcaaattctagattt atggatggactatctaaactagatcgcttacacgagactcatgacgattacagcgatcagatatttgagtctcttga gaggaatgactgtacctgtcaaaagtatcctgaaatcacagaagttagagatgcagttgccacaattagacgttcc tttagaaaaataactaaagaatctggtgccgatatcgaacctcccgtacaaactagcttattggatgattgccagac cttaaaaggagttcttacttgcttaatacctggtgctggtggttatgacgccattgcagtgattactaagcaagatgtt gatcttagggctcaaaccgctaatgacaaaagattttctaaggttcaatggctggatgtaactcaggctgactggg gtgttaggaaagaaaaagatccggaaacttatcttgataaataa
SEQ ID NO:62 Phosphomevalonate kinase (Sc_ERG8) Saccharomyces sp.	MSELRAFSAPGKALLAGGYLVLDTKYEAFVVGLSARMHAVAHPYGSL QGSDKFEVRVKSKQFKDGEWLYHISPKSGFIPVSIGGSKNPFIEKVIANV FSYFKPNMDDYCNRNLFVIDIFSDDAYHSQEDSVTEHRGNRRLSFHSH RIEEVPKTGLGSSAGLVTVLTTALASFFVSDLENNVDKYREVIHNLAQ VAHCQAQGKIGSGFDVAAAAYGSIRYRRFPPALISNLPDIGSATYGSKL AHLVDEEDWNITIKSNHLPSGLTLWMGDIKNGSETVKLVQKVKNWYD SHMPESLKIYTELDHANSRFMDGLSKLDRLHETHDDYSDQIFESLERN DCTCQKYPEITEVRDAVATIRRSFRKITKESGADIEPPVQTSLLDDCQTL KGVLTCLIPGAGGYDAIAVITKQDVDLRAQTANDKRFSKVQWLDVTQ ADWGVRKEKDPETYLDK*
SEQ ID NO:63 ERG12 – mevalonate kinase (Saccharomyces sp.)	atgtcattaccgttcttaacttctgcaccgggaaaggttattatttttggtgaacactctgctgtgtacaacaagcctg ccgtcgctgctagtgtgtctgcgttgagaacctacctgctaataagcgagtcatctgcaccagatactattgaattg gacttcccggacattagctttaatcataagtggtccatcaatgatttcaatgccatcaccgaggatcaagtaaactc ccaaaaattggccaaggctcaacaagccaccgatggcttgtctcaggaactcgttagtcttttggatccgttgttag ctcaactatccgaatccttccactaccatgcagcgttttgtttcctgtatatgtttgtttgcctatgcccccatgccaag aatattaagttttctttaaagtctactttacccatcggtgctgggttgggctcaagcgcctctatttctgtatcactggc cttagctatggcctacttgggggggttaataggatctaatgacttggaaaagctgtcagaaaacgataagcatata gtgaatcaatgggccttcataggtgaaaagtgtattcacggtacccccttcaggaatagataacgctgtggccactt atggtaatgccctgctatttgaaaaagactcacataatggaacaataaacacaaacaattttaagttcttagatgattt cccagccattccaatgatcctaacctatactagaattccaaggtctacaaaagatcttgttgctcgcgttcgtgtgtt ggtcaccgagaaatttcctgaagttatgaagccaattctagatgccatgggtgaatgtgccctacaaggcttagag atcatgactaagttaagtaaatgtaaaggcaccgatgacgaggctgtagaaactaataatgaactgtatgaacaa ctattggaattgataagaataaatcatggactgcttgtctcaatcggtgtttctcatcctggattagaacttattaaaaa tctgagcgatgatttgagaattggctccacaaaacttaccggtgctggtggcggcggttgctctttgactttgttac gaagagacattactcaagagcaaattgacagcttcaaaaagaaattgcaagatgattttagttacgagacatttga aacagacttgggtgggactggctgctgtttgttaagcgcaaaaaatttgaataaagatcttaaaatcaaatccctag tattccaattatttgaaaataaaactaccacaaagcaacaaattgacgatctattattgccaggaaacacgaatttac

SEQ ID	SEQUENCE
	catggacttcataa
SEQ ID NO:64 ERG12 – mevalonate kinase (Saccharomyces sp.)	MSLPFLTSAPGKVIIFGEHSAVYNKPAVAASVSALRTYLLISESSAPDTI ELDFPDISFNHKWSINDFNAITEDQVNSQKLAKAQQATDGLSQELVSLL DPLLAQLSESFHYHAAFCFLYMFVCLCPHAKNIKFSLKSTLPIGAGLGS SASISVSLALAMAYLGGLIGSNDLEKLSENDKHIVNQWAFIGEKCIHGT PSGIDNAVATYGNALLFEKDSHNGTINTNNFKFLDDFPAIPMILTYTRIP RSTKDLVARVRVLVTEKFPEVMKPILDAMGECALQGLEIMTKLSKCK GTDDEAVETNNELYEQLLELIRINHGLLVSIGVSHPGLELIKNLSDDLRI GSTKLTGAGGGGCSLTLLRRDITQEQIDSFKKKLQDDFSYETFETDLGG TGCCLLSAKNLNKDLKIKSLVFQLFENKTTTKQQIDDLLLPGNTNLPW TS*
SEQ ID NO:65 Mevalonate pyrophosphate decarboxylase (Sc_ERG19) Saccharomyces sp.	atgaccgtttacacagcatccgttaccgcacccgtcaacatcgcaacccttaagtattgggggaaaagggacac gaagttgaatctgcccaccaattcgtccatatcagtgactttatcgcaagatgacctcagaacgttgacctctgcg gctactgcacctgagtttgaacgcgacactttgtggttaaatggagaaccacacagcatcgacaatgaaagaact caaaattgtctgcgcgacctacgccaattaagaaaggaaatggaatcgaaggacgcctcattgcccacattatct caatggaaactccacattgtctccgaaaataactttcctacagcagctggtttagcttcctccgctgctggctttgct gcattggtctctgcaattgctaagttataccaattaccacagtcaacttcagaaatatctagaatagcaagaaaggg gtctggttcagcttgtagatcgttgtttggcggatacgtggcctgggaaatgggaaaagctgaagatggtcatgat tccatggcagtacaaatcgcagacagctctgactggcctcagatgaaagcttgtgtcctagttgtcagcgatatta aaaaggatgtgagttccactcagggtatgcaattgaccgtggcaacctccgaactatttaaagaaagaattgaac atgtcgtaccaaagagatttgaagtcatgcgtaaagccattgttgaaaaagatttcgccacctttgcaaaggaaac aatgatggattccaactctttccatgccacatgtttggactctttccctccaatattctacatgaatgacacttccaag cgtatcatcagttggtgccacaccattaatcagttttacggagaaacaatcgttgcatacacgtttgatgcaggtcc aaatgctgtgttgtactacttagctgaaaatgagtcgaaactctttgcatttatctataaattgtttggctctgttcctgg atgggacaagaaatttactactgagcagcttgaggctttcaaccatcaatttgaatcatctaactttactgcacgtga attggatcttgagttgcaaaaggatgttgccagagtgattttaactcaagtcggttcaggcccacaagaaacaaac gaatctttgattgacgcaaagactggtctaccaaaggaataa
SEQ ID NO:66 Mevalonate pyrophosphate decarboxylase (Sc_ERG19) Saccharomyces sp.	MTVYTASVTAPVNIATLKYWGKRDTKLNLPTNSSISVTLSQDDLRTLT SAATAPEFERDTLWLNGEPHSIDNERTQNCLRDLRQLRKEMESKDASL PTLSQWKLHIVSENNFPTAAGLASSAAGFAALVSAIAKLYQLPQSTSEI SRIARKGSGSACRSLFGGYVAWEMGKAEDGHDSMAVQIADSSDWPQ MKACVLVVSDIKKDVSSTQGMQLTVATSELFKERIEHVVPKRFEVMR KAIVEKDFATFAKETMMDSNSFHATCLDSFPPIFYMNDTSKRIISWCHT INQFYGETIVAYTFDAGPNAVLYYLAENESKLFAFIYKLFGSVPGWDK KFTTEQLEAFNHQFESSNFTARELDLELQKDVARVILTQVGSGPQETNE SLIDAKTGLPKE*
SEQ ID NO:67 Engineered phosphomevalonate kinase/mevalonate kinase (Erg8-T2A-Erg12)	atgtcagagttgagagccttcagtgccccagggaaagcgttactagctggtggatatttagttttagatacaaaata tgaagcatttgtagtcggattatcggcaagaatgcatgctgtagcccatccttacggttcattgcaagggtctgata agtttgaagtgcgtgtgaaaagtaaacaatttaaagatggggagtggctgtaccatataagtcctaaaagtggctt cattcctgtttcgataggcggatctaagaacccttttcattgaaaaagttatcgctaacgtatttagctactttaaacct acatggacgactactgcaatagaaacttgttcgttattgatattttctctgatgatgcctaccattctcaggaggatag cgttaccgaacatcgtggcaacagaagattgagttttcattcgcacagaattgaagaagttcccaaaacagggct gggctcctcggcaggtttagtcacagttttaactacagctttggcctccttttttgtatcggacctggaaaataatgta gacaaatatagagaagttattcataatttagcacaagttgctcattgtcaagctcagggtaaaattggaagcgggtt tgatgtagcggcggcagcatatggatctatcagatatagaagattcccacccgcattaatctctaatttgccagata ttggaagtgctacttacggcagtaaactggcgcatttggttgatgaagaagactggaatattacgattaaaagtaa ccatttaccttcgggattaactttatggatgggcgatattaagaatggttcagaaacagtaaaactggtccagaag gtaaaaaattggtatgattcgcatatgccagaaagcttgaaaatatatacagaactcgatcatgcaaattctagattt atggatggactatctaaactagatcgcttacacgagactcatgacgattacagcgatcagatatttgagtctcttga gaggaatgactgtacctgtcaaaagtatcctgaaatcacagaagttagagatgcagttgccacaattagacgttcc tttagaaaaataactaaagaatctggtgccgatatcgaacctcccgtacaaactagcttattggatgattgccagac cttaaaaggagttcttacttgcttaatacctggtgctggtggttatgacgccattgcagtgattactaagcaagatgtt gatcttagggctcaaaccgctaatgacaaaagattttctaaggttcaatggctggatgtaactcaggctgactggg

SEQ ID	SEQUENCE
	gtgttaggaaagaaaaagatccggaaacttatcttgataaaaagcttgagggcagaggaagtcttctaacatgcg gtgacgtggaggagaatcccggccctgctagcatgtcattaccgttcttaacttctgcaccgggaaaggttattatt tttggtgaacactctgctgtgtacaacaagcctgccgtcgctgctagtgtgtctgcgttgagaacctacctgctaat aagcgagtcatctgcaccagatactattgaattggacttcccggacattagctttaatcataagtggtccatcaatg atttcaatgccatcaccgaggatcaagtaaactcccaaaaattggccaaggctcaacaagccaccgatggcttgt ctcaggaactcgttagtcttttggatccgttgttagctcaactatccgaatccttccactaccatgcagcgttttgtttc ctgtatatgtttgtttgcctatgcccccatgccaagaatattaagttttctttaaagtctactttacccatcggtgctggg ttgggctcaagcgcctctatttctgtatcactggccttagctatggcctacttgggggggttaataggatctaatgac ttggaaaagctgtcagaaaacgataagcatatagtgaatcaatgggccttcataggtgaaaagtgtattcacggta ccccttcaggaatagataacgctgtggccacttatggtaatgccctgctatttgaaaaagactcacataatggaac aataaacacaaacaattttaagttcttagatgatttcccagccattccaatgatcctaacctatactagaattccaagg tctacaaaagatcttgttgctcgcgttcgtgtgttggtcaccgagaaatttcctgaagttatgaagccaattctagat gccatgggtgaatgtgccctacaaggcttagagatcatgactaagttaagtaaatgtaaaggcaccgatgacga ggctgtagaaactaataatgaactgtatgaacaactattggaattgataagaataaatcatggactgcttgtctcaat cggtgtttctcatcctggattagaacttattaaaaatctgagcgatgatttgagaattggctccacaaaacttaccgg tgctggtggcggcggttgctctttgactttgttacgaagagacattactcaagagcaaattgacagcttcaaaaag aaattgcaagatgattttagttacgagacatttgaaacagacttgggtgggactggctgctgtttgttaagcgcaaa aaatttgaataaagatcttaaaatcaaatccctagtattccaattatttgaaaataaaactaccacaaagcaacaaatt gacgatctattattgccaggaaacacgaatttaccatggacttcataa
SEQ ID NO:68 Engineered phosphomevalonate kinase/mevalonate kinase (Erg8-T2A-Erg12)	MSELRAFSAPGKALLAGGYLVLDTKYEAFVVGLSARMHAVAHPYGSL QGSDKFEVRVKSKQFKDGEWLYHISPKSGFIPVSIGGSKNPFIEKVIANV FSYFKPNMDDYCNRNLFVIDIFSDDAYHSQEDSVTEHRGNRRLSFHSH RIEEVPKTGLGSSAGLVTVLTTALASFFVSDLENNVDKYREVIHNLAQ VAHCQAQGKIGSGFDVAAAAYGSIRYRRFPPALISNLPDIGSATYGSKL AHLVDEEDWNITIKSNHLPSGLTLWMGDIKNGSETVKLVQKVKNWYD SHMPESLKIYTELDHANSRFMDGLSKLDRLHETHDDYSDQIFESLERN DCTCQKYPEITEVRDAVATIRRSFRKITKESGADIEPPVQTSLLDDCQTL KGVLTCLIPGAGGYDAIAVITKQDVDLRAQTANDKRFSKVQWLDVTQ ADWGVRKEKDPETYLDKKLEGRGSLLTCGDVEENPGPASMSLPFLTS APGKVIIFGEHSAVYNKPAVAASVSALRTYLLISESSAPDTIELDFPDISF NHKWSINDFNAITEDQVNSQKLAKAQQATDGLSQELVSLLDPLLAQLS ESFHYHAAFCFLYMFVCLCPHAKNIKFSLKSTLPIGAGLGSSASISVSLA LAMAYLGGLIGSNDLEKLSENDKHIVNQWAFIGEKCIHGTPSGIDNAV ATYGNALLFEKDSHNGTINTNNFKFLDDFPAIPMILTYTRIPRSTKDLV ARVRVLVTEKFPEVMKPILDAMGECALQGLEIMTKLSKCKGTDDEAV ETNNELYEQLLELIRINHGLLVSIGVSHPGLELIKNLSDDLRIGSTKLTG AGGGGCSLTLLRRDITQEQIDSFKKKLQDDFSYETFETDLGGTGCCLLS AKNLNKDLKIKSLVFQLFENKTTTKQQIDDLLLPGNTNLPWTS*
SEQ ID NO:69 Artificial neryl pyrophosphate (NPP) synthase (NPPS) nucleotide sequence	atgtgctcacttaatttgcaaacggaaaagctatgctatgaagacaatgacaatgacttggacgaggaactgatg ccgaagcacatagcgctaatcatggatggtaatagacgttgggcaaaagacaagggcttagaagtgtacgaag ggcacaaacatataatcccgaaactaaaagaaatatgtgacatatcctccaagttggggattcagatcatcacag cgttcgcgttctccacagagaactggaagagatccaaggaggaagtcgatttcctattgcagatgtttgaagaaat ctatgacgaatttagccgttctggggtgagagtgagtatcatcggatgcaaaagcgatttgccgatgacccttcaa aaatgtatcgcattgacagaggaaacgacgaaaggcaataagggattacacctggtcatagcacttaactacgg tgggtattacgatatcctacaagcaacgaagtccattgtaaacaaggctatgaatggtttattggacgttgaagaca tcaataaaaatctgttcgaccaagaattagaaagcaaatgccctaaccctgacttgctgatcagaactgggggag aacagagggtctctaattttcttctatggcaattggcttatactgagttctattttaccaatactttattccctgactttgg tgaagaggacctgaaagaagccatcatgaattttcaacagagacaccgtagattcggaggacatacttattga
SEQ ID NO:70 Neryl pyrophosphate (NPP) synthase (NPPS)	MCSLNLQTEKLCYEDNDNDLDEELMPKHIALIMDGNRRWAKDKGLE VYEGHKHIIPKLKEICDISSKLGIQIITAFAFSTENWKRSKEEVDFLLQMF EEIYDEFSRSGVRVSIIGCKSDLPMTLQKCIALTEETTKGNKGLHLVIAL NYGGYYDILQATKSIVNKAMNGLLDVEDINKNLFDQELESKCPNPDLL IRTGGEQRVSNFLLWQLAYTEFYFTNTLFPDFGEEDLKEAIMNFQQRH

SEQ ID	SEQUENCE
Solanum sp.	RRFGGHTY*
SEQ ID NO:71 Artificial geranylgeranyl pyrophosphate synthase large subunit (GPPSlsu) nucleotide sequence	atgagcaccgtgaatctgacctgggtgcagacgtgctctatgttcaaccagggcgggcgttcccgttcattgtca accttcaacttaaatctgtaccatccattgaagaaaacgcctttctctatccagacacctaagcagaaaaggccaa cttccccttctcatctatcagtgccgtattaacggagcaggaagcagtaaaggagggtgacgaggaaaaaagc atatttaacttcaaatcttatatggttcagaaagctaatagcgtgaatcaggcactagattctgcggtgttattgagag accccattatgatacatgaatctatgcgttactctttgcttgcgggcggcaagcgtgtcagaccgatgttatgcttaa gtgcgtgcgagttagtaggaggtaaagagtctgtagcaatgcccgcagcatgtgctgtagaaatgatacacaca atgtcactgattcacgatgatcttccttgcatggataacgacgatcttcgtagaggtaagccaaccaaccacaagg tattcggggaagacgtggcagttttagcaggagacgcgctactagcgttcgcgtttgaacacatggcagttagca cagtaggagttccagcagcaaaaatagttagggctataggagagttagcaaagtccatcggtagcgagggcctt gttgccggacaggtagttgatatcgatagtgaagggttggctaacgtgggactagaacaactggagttcatccac ctacacaagacaggggcactgcttgaagcgagtgttgtacttggggctattctgggggggaggaacagatgagg aggtagaaaaactacgtagttttgccaggtgtataggactactatttcaagttgtagatgatatccttgacgtcacga agagtagtcaagagttaggaaaaacagcagggaaagatctagttgccgataaagtaacctaccccaggctaat gggtatcgataaatctcgtgagttcgccgaacaattaaatactgaggctaagcaacatttaagcgggtttgatccta ttaaggctgcgccgctgattgctctagcaaactatattgcatatagacagaactga
SEQ ID NO:72 Geranylgeranyl pyrophosphate synthase large subunit (GPPSlsu) Cannabis sativa	MSTVNLTWVQTCSMFNQGGRSRSLSTFNLNLYHPLKKTPFSIQTPKQK RPTSPFSSISAVLTEQEAVKEGDEEKSIFNFKSYMVQKANSVNQALDSA VLLRDPIMIHESMRYSLLAGGKRVRPMLCLSACELVGGKESVAMPAA CAVEMIHTMSLIHDDLPCMDNDDLRRGKPTNHKVFGEDVAVLAGDA LLAFAFEHMAVSTVGVPAAKIVRAIGELAKSIGSEGLVAGQVVDIDSE GLANVGLEQLEFIHLHKTGALLEASVVLGAILGGGTDEEVEKLRSFAR CIGLLFQVVDDILDVTKSSQELGKTAGKDLVADKVTYPRLMGIDKSRE FAEQLNTEAKQHLSGFDPIKAAPLIALANYIAYRQN*
SEQ ID NO:73 Artificial geranylgeranyl pyrophosphate synthase small subunit (GPPSssu) nucleotide sequence	atggctgtttacaacctttcaatcaactgttctcccagattcgtccatcatgtatacgtgccccattttacatgtaaatc aaataagagcctgagccatgtccccatgagaatcacgatgtcaaagcagcatcatcactcatactttgcctctaca acggcagatgtcgatgcccatctaaaacaatcaatcacaattaaaccccgttgtctgtccacgaagccatgtata actttatcttcagtacgccaccgaatttggcgccatcattatgtgtcgcagcatgtgaattggttgggggtcaccag gacaggcgatggcagcggccagcgcattaagggtaatacatgctagcatcgttacccacgatcaccttccgtt aacgggaaggccaaaccccacctcaccggaggccgctacgcacaattcctataatccaaacatacagttgttatt acctgacgccattacacccttcgggtttgagctattagcgtccagtgatgatcttacacacaacaagagtgagaga gttcttagggtgatcgttgaatttacgaggactttcggttccagaggcactatagacgcccaataccacgaaaagt tggctagtaggtttgatgtggatagccatgaggcaaagaccgtaggatgggggcattacccatcattgaaaaag gagggagccatgcacgcatgtgctgctgcctgcggagcaatattgggtgaggctcatgaagaagaagtggaa aaattgcgtacattcgggctgtatgtcggcatgatccaaggttatgcgaacagattcatcatgagcagtacagag gagaaaaaagaggctgacaggataattgaggagcttaccaatttagcgcgtcaggagctgaaatacttcgatgg aaggaacctagaaccgttttcaacattcttgttccgtttgtag
SEQ ID NO:74 Geranylgeranyl pyrophosphate synthase small subunit (GPPSssu) Cannabis sativa	MAVYNLSINCSPRFVHHVYVPHFTCKSNKSLSHVPMRITMSKQHHHSY FASTTADVDAHLKQSITIKPPLSVHEAMYNFIFSTPPNLAPSLCVAACEL VGGHQGQAMAAASALRVIHASIVTHDHLPLTGRPNPTSPEAATHNSYN PNIQLLLPDAITPFGFELLASSDDLTHNKSERVLRVIVEFTRTFGSRGTID AQYHEKLASRFDVDSHEAKTVGWGHYPSLKKEGAMHACAAACGAIL GEAHEEEVEKLRTFGLYVGMIQGYANRFIMSSTEEKKEADRIIEELTNL ARQELKYFDGRNLEPFSTFLFRL*
SEQ ID NO:75 Artificial tetraketide synthase (TKS) nucleotide sequence	atgaatcatttaagagctgaaggtccagcctccgttttggccatcggtaccgctaaccctgaaaacattttgttgca agacgaattcccagactactacttcagagtcactaagtccgaacacatgacccaattgaaggagaagttcagaa agatttgtgacaagtccatgattagaaagagaaactgtttcttgaacgaagaacacttgaagcaaaacccaagatt ggttgaacatgaaatgcaaactttggacgctagacaagacatgttggttgttgaagtccctaagttgggtaaggat gcctgtgctaaggccattaaagaatggggtcaacctaagtccaagattacccacttgattttcacctctgcctccac cactgacatgcctggtgctgattaccactgcgctaagttattgggtttgtctccatccgttaagagagttatgatgta ccaattgggttgctacggtggtggtactgttttaagaattgctaaggatattgctgaaaacaacaagggtgccaga gtcttagctgtctgctgtgacattatggcttgtttattcagaggtccatctgaatccgacttggaattgttggttggtca

SEQ ID	SEQUENCE
	agctatcttcggtgacggtgctgctgccgttattgttggtgctgaaccagacgaatccgttggtgaaagaccaattt ttgaattggtttccaccggtcaaactattttgccaaattccgaaggtaccatcggtggtcatatcagagaagccggt ttgatcttcgacttacataaggatgtcccaatgttgatctctaacaacattgaaaagtgtttgatcgaagcttttaccc caattggtatttctgactggaactctatcttctggattacccatcctggtggtaaggctattttggataaggtcgagga aaaattgcacttgaagtctgacaagttcgttgactctagacacgtcttgtccgaacatggtaatatgtcctcttccac cgttttattcgttatggatgagttgagaaagagatccttagaagaaggtaagtccaccaccggtgatggttttgagt ggggtgttttgttcggtttcggtccaggtttgaccgtcgaaagagttgttgttagatctgtcccaattaagtacggat cc
SEQ ID NO:76 Artificial tetraketide synthase (TKS)	MNHLRAEGPASVLAIGTANPENILLQDEFPDYYFRVTKSEHMTQLKEK FRKICDKSMIRKRNCFLNEEHLKQNPRLVEHEMQTLDARQDMLVVEV PKLGKDACAKAIKEWGQPKSKITHLIFTSASTTDMPGADYHCAKLLGL SPSVKRVMMYQLGCYGGGTVLRIAKDIAENNKGARVLAVCCDIMACL FRGPSESDLELLVGQAIFGDGAAAVIVGAEPDESVGERPIFELVSTGQTI LPNSEGTIGGHIREAGLIFDLHKDVPMLISNNIEKCLIEAFTPIGISDWNSI FWITHPGGKAILDKVEEKLHLKSDKFVDSRHVLSEHGNMSSSTVLFVM DELRKRSLEEGKSTTGDGFEWGVLFGFGPGLTVERVVVRSVPIKYGS
SEQ ID NO:77 Artificial olivetolic acid cyclase (OAC) nucleotide sequence	atggccgtcaagcacttgatcgttttgaagttcaaggatgaaatcactgaagctcaaaaggaagaattcttcaaaa cctacgtcaacttagtcaatattattccagccatgaaggacgtctattggggtaaggacgttactcaaaagaataa ggaggaaggttatactcatatcgttgaggtcactttcgaatctgttgagactattcaagactacatcatccacccag cccacgttggtttcggtgatgtttatcgttccttctgggaaaaattgttgatcttcgactacacccctagaaagggat cc
SEQ ID NO:78 Artificial olivetolic acid cyclase (OAC)	MAVKHLIVLKFKDEITEAQKEEFFKTYVNLVNIIPAMKDVYWGKDVT QKNKEEGYTHIVEVTFESVETIQDYIIHPAHVGFGDVYRSFWEKLLIFD YTPRKGS
SEQ ID NO:79 Fusion tetraketide synthase-olivetolic acid cyclase (TKS-OAC)	atgaatcatttaagagctgaaggtccagcctccgttttggccatcggtaccgctaaccctgaaaacattttgttgca agacgaattcccagactactacttcagagtcactaagtccgaacacatgacccaattgaaggagaagttcagaa agatttgtgacaagtccatgattagaaagagaaactgtttcttgaacgaagaacacttgaagcaaaacccaagatt ggttgaacatgaaatgcaaactttggacgctagacaagacatgttggttgttgaagtccctaagttgggtaaggat gcctgtgctaaggccattaaagaatggggtcaacctaagtccaagattacccacttgattttcacctctgcctccac cactgacatgcctggtgctgattaccactgcgctaagttattgggtttgtctccatccgttaagagagttatgatgta ccaattgggttgctacggtggtggtactgttttaagaattgctaaggatattgctgaaaacaacaagggtgccaga gtcttagctgtctgctgtgacattatggcttgtttattcagaggtccatctgaatccgacttggaattgttggttggtca agctatcttcggtgacggtgctgctgccgttattgttggtgctgaaccagacgaatccgttggtgaaagaccaattt ttgaattggtttccaccggtcaaactattttgccaaattccgaaggtaccatcggtggtcatatcagagaagccggt ttgatcttcgacttacataaggatgtcccaatgttgatctctaacaacattgaaaagtgtttgatcgaagcttttaccc caattggtatttctgactggaactctatcttctggattacccatcctggtggtaaggctattttggataaggtcgagga aaaattgcacttgaagtctgacaagttcgttgactctagacacgtcttgtccgaacatggtaatatgtcctcttccac cgttttattcgttatggatgagttgagaaagagatccttagaagaaggtaagtccaccaccggtgatggttttgagt ggggtgttttgttcggtttcggtccaggtttgaccgtcgaaagagttgttgttagatctgtcccaattaagtacgcag ccacaagcggttctacgggctccacgggctctaccggcagtgggaggagcactgggtcaacgggatcaacag gtagtggaagatcacacatggttgccgtcaagcacttgatcgttttgaagttcaaggatgaaatcactgaagctca aaaggaagaattcttcaaaacctacgtcaacttagtcaatattattccagccatgaaggacgtctattggggtaag gacgttactcaaaagaataaggaggaaggttatactcatatcgttgaggtcactttcgaatctgttgagactattca agactacatcatccacccagcccacgttggtttcggtgatgtttatcgttccttctgggaaaaattgttgatcttcgac tacacccctagaaagggtaactcgagagcttttgattaa
SEQ ID NO:80 Fusion tetraketide synthase-olivetolic acid cyclase (TKS-OAC)	MNHLRAEGPASVLAIGTANPENILLQDEFPDYYFRVTKSEHMTQLKEK FRKICDKSMIRKRNCFLNEEHLKQNPRLVEHEMQTLDARQDMLVVEV PKLGKDACAKAIKEWGQPKSKITHLIFTSASTTDMPGADYHCAKLLGL SPSVKRVMMYQLGCYGGGTVLRIAKDIAENNKGARVLAVCCDIMACL FRGPSESDLELLVGQAIFGDGAAAVIVGAEPDESVGERPIFELVSTGQTI LPNSEGTIGGHIREAGLIFDLHKDVPMLISNNIEKCLIEAFTPIGISDWNSI FWITHPGGKAILDKVEEKLHLKSDKFVDSRHVLSEHGNMSSSTVLFVM

SEQ ID	SEQUENCE
	DELRKRSLEEGKSTTGDGFEWGVLFGFGPGLTVERVVVRSVPIKYAAT SGSTGSTGSTGSGRSTGSTGSTGSGRSHMVAVKHLIVLKFKDEITEAQK EEFFKTYVNLVNIIPAMKDVYWGKDVTQKNKEEGYTHIVEVTFESVET IQDYIIHPAHVGFGDVYRSFWEKLLIFDYTPRKGNSRAFD*
SEQ ID NO:81 Artificial geranyl pyrophosphate olivetolic acid geranyltransferase (GOT) nucleotide sequence	atgggattgtccagcgtgtgcaccttctcattccaaaccaactaccatacacttctcaatccgcacaataataaccc gaaaaccagcttattatgttatagacacccgaagacgcctattaagtacagttataacaactttcctagcaagcatt gctctactaaaagttttcatctgcaaaacaagtgctctgagtccttgagtatagcaaagaatagcattagagctgca acgacaaatcaaaccgagccgccggagtctgataaccatagtgtggcgaccaagatactaaattttggcaaagc gtgttggaagctacaacgaccttatactattatcgcgtttacgagttgtgcatgtgggctgttcgggaaagagctctt gcacaatacaaacttaatcagttggagtttgatgttcaaagcatttttttttctcgtcgctatcttatgtatcgcgtcattt accacgaccataaatcaaatatacgatctgcatatcgatcgtatcaataagcccgacctcccactggcctcaggt gaaatttccgttaacacggcgtggattatgagtataatcgtagcactatttggacttattataaccatcaaaatgaag ggcggtcctctatacatttttggatattgttttgggatttttggaggtatagtctattccgtccccccattcagatggaa acaaaacccgtccaccgctttccttttaaatttcttggcacatatcatcacaaacttcacgttttactatgccagccga gccgcactgggactcccgttcgagttgcgtccgtcattcaccttccttttagcttttatgaaatctatgggaagcgct ttagctttaattaaggacgcgagcgacgtggaaggggacacgaaattcggtataagcacgctggcttcaaaatat ggaagtcgtaatctcactctattttgttctgggattgtactcctaagttacgtagctgcgatactcgcaggcattatat ggccacaagctttcaactccaacgtaatgttgctatcacatgcaatcttggccttctggctcatccttcaaactagag attttgcactaacgaactacgatccagaagcgggtcgtcgattttacgaatttatgtggaaactgtactatgctgagt acctcgtctatgtgttcata
SEQ ID NO:82 geranyl pyrophosphate olivetolic acid geranyltransferase (GOT) (CsPT1) *Cannabis sativa* 395 aa WO 2011/017798	MGLSSVCTFSFQTNYHTLLNPHNNNPKTSLLCYRHPKTPIKYSYNNFPS KHCSTKSFHLQNKCSESLSIAKNSIRAATTNQTEPPESDNHSVATKILNF GKACWKLQRPYTIIAFTSCACGLFGKELLHNTNLISWSLMFKAFFFLVA ILCIASFTTTINQIYDLHIDRINKPDLPLASGEISVNTAWIMSIIVALFGLII TIKMKGGPLYIFGYCFGIFGGIVYSVPPFRWKQNPSTAFLLNFLAHIITN FTFYYASRAALGLPFELRPSFTFLLAFMKSMGSALALIKDASDVEGDTK FGISTLASKYGSRNLTLFCSGIVLLSYVAAILAGIIWPQAFNSNVMLLSH AILAFWLILQTRDFALTNYDPEAGRRFYEFMWKLYYAEYLVYVFI
SEQ ID NO:83 Artificial aromatic prenyltransferase (NphB-ScCO) nucleotide sequence	atgtctgaggcggcagacgtagagagagtatacgctgctatggaggaagcggctggattattgggggtggctt gtgccagagacaagatatatccgttactgtctactttccaggacactcttgtagaaggagggagtgtggtggtgttt agtatggcatcaggccgtcattcaacagagctagatttcagtatatctgtgccaacaagtcacggtgatccatacg caaccgtagtcgagaagggtcttttcccggcaacagggcatcctgtagatgatttgcttgccgacacacagaag cacctgcccgtctccatgttcgcaatcgatggtgaggtgaccggaggatttaaaaagacttacgctttcttcccga ctgacaatatgccaggagttgccgagttgagtgcaataccatccatgccgccagcagtcgcggagaacgccga attgttcgcccgttacggcttggacaaagtccaaatgactagtatggactataaaaagaggcaggtgaatctatatt tcagcgaactttctgcccaaaccttggaggcggagagcgttttagcccttgttagggagttagggctacacgtcc cgaatgagttgggtttgaaattttgtaagcgtagcttttcagtatatccgacgctgaactgggaaactggaaagatt gacaggctatgctttgcagtgatttctaatgaccctacgcttgtaccttcctcagacgagggcgacatcgagaaatt ccacaactatgccacaaaagctccgtatgcctacgtcggcgaaaaacgtactctagtatacggtttgactctgagt cccaaggaagagtattacaagctaggagcgtactatcatatcactgatgtgcaacgtggcttgctgaaagccttc gactccttagaggac
SEQ ID NO:84 Aromatic prenyltransferase NphB-ScCO (Streptomyces sp.)	MSEAADVERVYAAMEEAAGLLGVACARDKIYPLLSTFQDTLVEGGSV VVFSMASGRHSTELDFSISVPTSHGDPYATVVEKGLFPATGHPVDDLL ADTQKHLPVSMFAIDGEVTGGFKKTYAFFPTDNMPGVAELSAIPSMPP AVAENAELFARYGLDKVQMTSMDYKKRQVNLYFSELSAQTLEAESVL ALVRELGLHVPNELGLKFCKRSFSVYPTLNWETGKIDRLCFAVISNDPT LVPSSDEGDIEKFHNYATKAPYAYVGEKRTLVYGLTLSPKEEYYKLGA YYHITDVQRGLLKAFDSLED
SEQ ID NO:85	atgaactgttccgcgtttagtttctggttcgtgtgcaagatcatcttctttttttctaagcttcaacattcaaatcagcatc gcgaatcctcaggagaacttcctgaagtgtttctcagaatacataccaaataatcccgccaatcctaaatttatatat

SEQ ID	SEQUENCE
Artificial Tetrahydro cannabinolic acid synthase (THCAS) nucleotide sequence	acccaacatgatcagctatacatgagtgtattgaactctacgattcagaatctaagattcacatctgatacaacgcc gaaacctctagtaatcgtgacaccgtctaatgtctcccatattcaagcttctatcttgtgctcaaagaaagtcggtctt caaataaggacacgttctggcgggcatgacgccgagggcatgtcatatatcagccaagtaccatttgtagtcgtg gatttaagaaacatgcattctataaaaatcgacgttcactcccaaacggcatgggtggaagctggagcgacactg gggggaggtgtactactggatcaatgaaaagaacgaaaatttttccttccccggaggatattgtccgacagttggg gtgggggccacttctctggcggcgggtacggcgctctgatgcgtaattatggactggccgcagataacataat cgacgcgcatttggtgaacgttgacgggaaggttttggataggaagtctatgggagaggacctattctgggcaat tagaggcggaggaggagagaattttggtattattgctgcatggaagattaaattggttgcggtgccgagtaaaag taccatcttttccgtcaagaaaaacatggagattcacggactagttaagctgtttaataaatggcaaaacatcgcct ataagtacgacaaagatttggttctgatgacgcatttcataactaagaatataactgataatcacggcaagaataag accactgtgcacggttattttagttcaatattccatggcggcgttgactcccttgtcgatttgatgaataagagcttcc ctgaattgggtatcaagaagacagactgcaaagaattctcctggattgatacgactatcttctattcaggggtcgtg aatttcaacactgcgaatttcaaaaaggagatattgttagaccgttccgcgggaaaaaaaactgcgttttctattaa actagattatgtgaaaaaaccgattcctgagacagccatggttaagattcttgaaaaattgtatgaagaggatgtcg ggtcggtatgtacgtcctttacccttacggaggaatcatggaagaaatatccgaatctgcaattcctttcccgcat cgtgccggtattatgtatgagctatggtacaccgctagctgggagaagcaggaagataacgagaagcatatcaa ttgggtgaggtctgtgtataattttacaacaccatacgtcagtcaaaaccctagattggcctatcttaactatcgtgat ctggacttgggaaaaacaaatccagaatccccaaataactacactcaagcccgtatatggggcgagaagtactt cggcaaaaatttcaatagactggtcaaagttaagacgaaagcagaccctaataatttcttccgtaacgaacaatca attcccccgcttccgccacaccatcac
SEQ ID NO:86 Tetrahydro cannabinolic acid synthase (THCAS) *Cannabis sativa*	MNCSAFSFWFVCKIIFFFLSFNIQISIANPQENFLKCFSEYIPNNPANPKFI YTQHDQLYMSVLNSTIQNLRFTSDTTPKPLVIVTPSNVSHIQASILCSKK VGLQIRTRSGGHDAEGMSYISQVPFVVVDLRNMHSIKIDVHSQTAWVE AGATLGEVYYWINEKNENFSFPGGYCPTVGVGGHFSGGGYGALMRN YGLAADNIIDAHLVNVDGKVLDRKSMGEDLFWAIRGGGGENFGIIAA WKIKLVAVPSKSTIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLVLM THFITKNITDNHGKNKTTVHGYFSSIFHGGVDSLVDLMNKSFPELGIKK TDCKEFSWIDTTIFYSGVVNFNTANFKKEILLDRSAGKKTAFSIKLDYV KKPIPETAMVKILEKLYEEDVGVGMYVLYPYGGIMEEISESAIPFPHRA GIMYELWYTASWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLN YRDLDLGKTNPESPNNYTQARIWGEKYFGKNFNRLVKVKTKADPNNF FRNEQSIPPLPPHHH
SEQ ID NO:87 Artificial Cannabidiolic acid synthase (CBDAS) nucleotide sequence	atgaaatgttctactttcagtttttggttcgtgtgtaagatcatctttttctttttcagcttcaatatacagacaagtatcgc caatccaagagaaaatttcttaaaatgtttttcacagtacatccctaataacgccactaacctgaaattagtgtacac ccaaaataatcctctttatatgtctgttttaaactccacgatccataatttaaggtttacatcagatacgacaccaaag cccttggtaatcgtgactcccagccacgtgagccacatacaggggaccatcctgtgctctaagaaagtaggcttg cagatcaggacaagatccggtggacacgacagtgagggaatgtcctatatttcacaagtccccttcgttatagta gatctgaggaacatgaggtccattaagattgatgtgcactcacaaacggcttgggttgaagctggagccacattg ggagaggtttattactgggtgaatgagaagaacgagaacctttcattagcagcgggatattgtcccacggtgtgc gcaggtgggcatttcgggggaggagggtacggccctttgatgagaaattacgggctagcggcagacaacatc atcgacgcccatctggtgaacgtgcatggaaaagtactggacagaaagtcaatgggcgaggacctgttttgggc tttgagagggggcggtgcagagtcatttggcatcatagttgcatggaaaatcagacttgttgccgtcccaaagtcc acaatgttctctgttaagaaaatcatggagatacacgaattggtgaaattagtgaataaatggcaaaacatagcgt acaagtacgacaaagacttactgctgatgacacactttatcacccgtaatattacagataatcagggtaagaacaa aaccgcgatccatacatatttttcatccgtttttctaggcggtgtcgattcattagtagatctgatgaacaaatctttcc ccgaacttggtatcaaaaagactgattgcagacagttatcatggattgatacaataattttctattctggtgtcgtaaa ttacgataccgataattttaataaggaaatactattagatcgttccgctgggcagaatggtgcattcaagataaaact tgattatgtcaaaaagcccattccagagagtgtctttgtgcagatccttgagaagttgtatgaagaagacattggtg cagggatgtacgcgctatatccgtacgggggtattatggacgagatttctgagagcgccataccattcccacaca gagcaggaattttatacgagttatggtatatctgctcatgggaaaaacaggaagacaacgagaagcacttaaact ggatacgtaatatctataatttttatgaccccatacgtatcaaaaaatccgcgtcttgcgtaccttaactacagggacc tggacataggtataaacgacccaaaaaatcccaataattacacccaagctagaatctgggggagaagtatttcg gtaagaactttgaccgtttggtaaaagtcaaaactctggtcgatccgaacaatttcttccgtaacgagcaatccata

2018256863 11 Apr 2024

SEQ ID	SEQUENCE
	cctccgctaccgagacatagacat
SEQ ID NO:88 GenBank A6P6V9 Cannabidiolic acid synthase (CBDAS) Cannabis sativa	MKCSTFSFWFVCKIIFFFFSFNIQTSIANPRENFLKCFSQYIPNNATNLKL VYTQNNPLYMSVLNSTIHNLRFTSDTTPKPLVIVTPSHVSHIQGTILCSK KVGLQIRTRSGGHDSEGMSYISQVPFVIVDLRNMRSIKIDVHSQTAWV EAGATLGEVYYWVNEKNENLSLAAGYCPTVCAGGHFGGGGYGPLMR NYGLAADNIIDAHLVNVHGKVLDRKSMGEDLFWALRGGGAESFGIIV AWKIRLVAVPKSTMFSVKKIMEIHELVKLVNKWQNIAYKYDKDLLLM THFITRNITDNQGKNKTAIHTYFSSVFLGGVDSLVDLMNKSFPELGIKK TDCRQLSWIDTIIFYSGVVNYDTDNFNKEILLDRSAGQNGAFKIKLDYV KKPIPESVFVQILEKLYEEDIGAGMYALYPYGGIMDEISESAIPFPHRAGI LYELWYICSWEKQEDNEKHLNWIRNIYNFMTPYVSKNPRLAYLNYRD LDIGINDPKNPNNYTQARIWGEKYFGKNFDRLVKVKTLVDPNNFFRNE QSIPPLPRHRH
SEQ ID NO:89 Artificial acyl-activating enzyme (CsAAE1) nucleotide sequence	atgggaaaaaactacaaaagtctggactccgtcgtcgcgtcagacttcattgccctaggcataacatcagaggta gcggaaaccttacacggcagactagccgagattgtttgtaactacggggcggctactccccagacttggatcaa tatagccaatcacatattaagccccgatttgccgttttcccttcaccaaatgttgttctacggctgctataaggacttt ggaccagcgccccccgcgtggattcctgatccggagaaagttaaatccacgaatcttggggcattactagaaaa acgtggcaaagaattcctaggagttaaatataaggaccccatatcttccttttcacactttcaagaattttcagttaga aacccagaggtttactggaggacagtattaatggatgagatgaagataagctttagtaaggatccggagtgtattc tgcgtagagatgacattaacaatcctggcggaagtgaatggctgcctggtgggtacctgaatagtgctaagaact gtttaaacgtcaactctaataaaaaattgaatgatacaatgattgtatggagagacgaagggaacgatgacctacc attgaacaagctgactctagatcagctacgtaaacgtgtatggttggtcgggtacgcgctggaggagatgggatt agaaaaaggatgcgcaattgctatcgacatgcctatgcatgtggacgcggtagtcatttacttggccattgtccta gcgggttacgtcgtcgtttcaattgcagacagcttttctgcacccgaaatcagtacccgtctgcgtttgtctaaagct aaggcaatatttacccaagaccatataattagaggcaagaagcgtataccgttgtacagtagggttgtagaggca aagtcacccatggctattgtgataccatgctctggctctaatataggagcggagcttagagatggtgacatctcct gggattactttcttgaacgtgctaaggagtttaaaaactgtgaatttactgcaagagagcagcccgtggatgcata cacaaacatattgttctccagcggtactacgggagaacctaaagcaataccttggacacaagctacacccttaa agcggccgctgacggatggtcccacctggatatcaggaagggtgacgtcatagtttggccgactaacctggga tggatgatgggcccttggctggtttacgctagccttctgaatggggccagcattgcattgtacaatggctcaccgc ttgtatcaggcttcgcgaagttcgtacaggacgccaaggtaacaatgctaggcgtagttccgtccatagttaggtc ttggaagagcacgaactgcgttagtggctacgattggagcactattcgttgtttcagctcttctggcgaggccagc aacgttgatgaatatttgtggttgatggggagagcgaactacaaacctgttattgagatgtgcggcggaactgag attgggggagcattctccgccggttcttttctacaagcccaaagtttatcctcttttagcagccagtgcatgggctgt acactatacattctggacaaaaatggttatccgatgccgaaaaacaagcccggcatcggagaactggccctagg acccgtgatgttcggcgctagtaagacgttgttgaatgggaatcaccacgacgtttatttttaagggaatgccaactt tgaatggcgaagtacttcgtagacacggagacatctttgagttgacttcaaacggttactaccacgctcatggacg tgccgatgatacgatgaacattgggggaattaaaatttcatccatagaaatagaacgtgtgtgtaacgaagtcgat gatcgtgtattcgagactacagcgatcggtgtcccaccgttgggtgggggaccagaacaattggtaatctttttgt tctgaaagactccaacgatacgaccatcgacctaaatcagctgaggctatcctttaatctgggcttgcagaaaaa gctaaatcctttattcaaagtcactagagttgttcctttatcttcattaccaagaactgcaacaaataaaataatgcgta gagttctaaggcagcagtttagtcatttcgaa
SEQ ID NO:90 Acyl-activating enzyme (CsAAE1) Cannabis sativa	MGKNYKSLDSVVASDFIALGITSEVAETLHGRLAEIVCNYGAATPQTW INIANHILSPDLPFSLHQMLFYGCYKDFGPAPPAWIPDPEKVKSTNLGA LLEKRGKEFLGVKYKDPISSFSHFQEFSVRNPEVYWRTVLMDEMKISFS KDPECILRRDDINNPGGSEWLPGGYLNSAKNCLNVNSNKKLNDTMIV WRDEGNDDLPLNKLTLDQLRKRVWLVGYALEEMGLEKGCAIAIDMP MHVDAVVIYLAIVLAGYVVVSIADSFSAPEISTRLRLSKAKAIFTQDHII RGKKRIPLYSRVVEAKSPMAIVIPCSGSNIGAELRDGDISWDYFLERAK EFKNCEFTAREQPVDAYTNILFSSGTTGEPKAIPWTQATPLKAAADGW SHLDIRKGDVIVWPTNLGWMMGPWLVYASLLNGASIALYNGSPLVSG FAKFVQDAKVTMLGVVPSIVRSWKSTNCVSGYDWSTIRCFSSSGEASN VDEYLWLMGRANYKPVIEMCGGTEIGGAFSAGSFLQAQSLSSFSSQCM

SEQ ID	SEQUENCE
	GCTLYILDKNGYPMPKNKPGIGELALGPVMFGASKTLLNGNHHDVYF KGMPTLNGEVLRRHGDIFELTSNGYYHAHGRADDTMNIGGIKISSIEIE RVCNEVDDRVFETTAIGVPPLGGGPEQLVIFFVLKDSNDTTIDLNQLRL SFNLGLQKKLNPLFKVTRVVPLSSLPRTATNKIMRRVLRQQFSHFE
SEQ ID NO:91 Artificial acyl-activating enzyme (CsAAE3) nucleotide sequence	atggaaaagtctggttatggtagagatggtatctacaggtctttaagaccaccattgcatttgccaaacaacaaca acttgtccatggtcagtttcttgttcagaaactcttcttcctacccacaaaaaccagccttgattgactctgaaactaa tcaaatcttgtccttctcccacttcaaatccaccgttattaaggtttctcacggtttcttgaacttgggtatcaagaaga atgactggttgatctacgctccaaactctattcatttcccagtttgctttttgggtattattgcttctggtgctattgctact acttccaacccattatacaccgtcagtgaattgtctaagcaagtcaaggattctaacccaaagttgattatcaccgtt ccacaattattggaaaaggtcaagggtttcaacttgccaaccattttgattggtccagactcagaacaagaatcctc ttcagataaggttatgaccttcaacgatttggttaacttgggtggttcttctggttctgaatttccaatcgttgatgactt caagcaatctgatactgctgctttgttgtactcttctggtactactggtatgtctaaaggttggttgactcacaagaac tttatcgcctcttctttgatggttaccatggaacaagacttggttggtgaaatggataacgttttcttgtgcttcttgcca atgttccatgttttcggtttggccattattacctacgctcaattgcaaagaggtaacactgttatttccgccagattcga tttggaaaagatgttgaaggacgtcgaaaagtatgttactcatttgtggtggcctccagttattttggctttgtctaaa aactccatggttaagttcaacttgtcatccatcaagtacattggttcaggtgctgctccattgggtaaggatttgatg gaagaatgttctaaatggccatacggtatagttgctcaaggttacggtatgactgaaacttgtggtatcgtttctatg gaagatatcagaggtggtaagagaaattctggttcagctggtatgttggcttcaggtgttgaagctcaaatagtttc tgttgataccttgaaaccattgccaccaaatcaattgggtgaaatttgggttaagggtccaaatatgatgcaaggtt acttcaacaatccacaagctaccaagttgaccattgataagaaaggttgggttcatactggtgacttgggttacttt gatgaagatggtcacttgtactgggacagaatcaaagaattgattaagtacaagggttttcaagtcgctccagctg aattggaaggtttgttggtttctcatccagaaatattggatgcctggattccatttccagatgctgaagctggtgaag ttccagttgcttattggagatcaccaaactcttcattgactgaaaacgacgtcaagaagttcattgctggtcaagttg cttctttcaagagattgagaaaggtcaccttcatcaactctgttccaaaatctgcttccggtaagatcttgagaagag aattgatccaaaaggtcagatccaatatg
SEQ ID NO:92 Acyl-activating enzyme (CsAAE3) Cannabis sativa	MEKSGYGRDGIYRSLRPPLHLPNNNNLSMVSFLFRNSSSYPQKPALIDS ETNQILSFSHFKSTVIKVSHGFLNLGIKKNDWLIYAPNSIHFPVCFLGIIA SGAIATTSNPLYTVSELSKQVKDSNPKLIITVPQLLEKVKGFNLPTILIGP DSEQESSSDKVMTFNDLVNLGGSSGSEFPIVDDFKQSDTAALLYSSGTT GMSKGWLTHKNFIASSLMVTMEQDLVGEMDNVFLCFLPMFHVFGLAI ITYAQLQRGNTVISARFDLEKMLKDVEKYVTHLWVPPVILALSKNSM VKFNLSSIKYIGSGAAPLGKDLMEECSKWPYGIVAQGYGMTETCGIVS MEDIRGGKRNSGSAGMLASGVEAQIVSVDTLKPLPPNQLGEIWVKGPN MMQGYFNNPQATKLTIDKKGWVHTGDLGYFDEDGHLYWDRIKELIK YKGFQVAPAELEGLLVSHPEILDAWIPFPDAEAGEVPVAYWRSPNSSL TENDVKKFIAGQVASFKRLRKVTFINSVPKSASGKILRRELIQKVRSNM
SEQ ID NO:93 Artificial acetoacetyl-CoA reductase (PhaB) nucleotide sequence	atgacgcagagaatcgcctatgtaacgggtgggatgggtgggataggaaccgccatatgtcagagactagcaa aggacggattcagggttgtagccggttgcggtcctaatagtccaagaagagagaaatggttggaacagcaaaa agctctaggatttgattttatagcatcagaagggaatgttgctgactgggattctacaaagacggcatttgacaaag tgaaatctgaagtcggcgaggtcgatgtcctaattaacaacgccggcatcaccagagatgtggttttcaggaaga tgactagggctgactgggacgccgtgatagacacaaatttgacgagcttgttcaacgtcacaaagcaagtaattg acggcatggcagatcgtgggtggggaaggatagtcaatatctccagcgtcaacggtcagaaaggccagttcgg acagactaactactccacagcgaaggctggcttacacggattcacgatggccttggcccaagaggtggctacta aaggggtgactgtgaacacagtgtcaccaggatacatcgcgacggatatggtcaaagctattagacaagatgtc ctggacaagattgttgccactattcccgtaaagaggcttgggttaccagaagagatagcttcaatttgcgcttggct atctagtgaggaatcagggttcagcactggggcggacttttcattaaacggtggattacacatgggaggatcc
SEQ ID NO:94 Mutant acetoacetyl-CoA reductase (PhaB)	MTQRIAYVTGGMGGIGTAICQRLAKDGFRVVAGCGPNSPRREKWLEQ QKALGFDFIASEGNVADWDSTKTAFDKVKSEVGEVDVLINNAGITRD VVFRKMTRADWDAVIDTNLTSLFNVTKQVIDGMADRGWGRIVNISSV NGQKGQFGQTNYSTAKAGLHGFTMALAQEVATKGVTVNTVSPGYIA TDMVKAIRQDVLDKIVATIPVKRLGLPEEIASICAWLSSEESGFSTGADF SLNGGLHMGGS

SEQ ID	SEQUENCE
SEQ ID NO:95 Artificial (R)-specific enoyl-CoA hydratase (PhaJ)	atgtctgcccagagtctggaagtcggtcaaaaagcaagactgtcaaaaagatttggggcggcagaggtagcgg cgttcgcggcgctgtctgaggattttaatccactgcacttagatcctgcgttcgccgcgacaacagcattcgagag gcccatcgtgcacggcatgctacttgcctctttgttctcaggtctactgggtcaacagttacctgggaaaggaagc atctatctgggacagtcattgtcttttaagctgcccgtcttcgtcggcgatgaggtgacagcagaagtagaagtca cagcattgagggaagacaagcctattgcgacccttactactcgtatttttactcagggcggagccttagcagtgac aggagaagctgtagtaaaactaccaggatcc
SEQ ID NO:96 Mutant (R)-specific enoyl-CoA hydratase (PhaJ)	MSAQSLEVGQKARLSKRFGAAEVAAFAALSEDFNPLHLDPAFAATTA FERPIVHGMLLASLFSGLLGQQLPGKGSIYLGQSLSFKLPVFVGDEVTA EVEVTALREDKPIATLTTRIFTQGGALAVTGEAVVKLPGS
SEQ ID NO:97 Mutated acetyl-CoA carboxylase (ACC1) (S659A, S1157A)	MSEESLFESSPQKMEYEITNYSERHTELPGHFIGLNTVDKLEESPLRDFV KSHGGHTVISKILIANNGIAAVKEIRSVRKWAYETFGDDRTVQFVAMA TPEDLEANAEYIRMADQYIEVPGGTNNNNYANVDLIVDIAERADVDA VWAGWGHASENPLLPEKLSQSKRKVIFIGPPGNAMRSLGDKISSTIVAQ SAKVPCIPWSGTGVDTVHVDEKTGLVSVDDDIYQKGCCTSPEDGLQK AKRIGFPVMIKASEGGGGKGIRQVEREEDFIALYHQAANEIPGSPIFIMK LAGRARHLEVQLLADQYGTNISLFGRDCSVQRRHQKIIEEAPVTIAKAE TFHEMEKAAVRLGKLVGYVSAGTVEYLYSHDDGKFYFLELNPRLQVE HPTTEMVSGVNLPAAQLQIAMGIPMHRISDIRTLYGMNPHSASEIDFEF KTQDATKKQRRPIPKGHCTACRITSEDPNDGFKPSGGTLHELNFRSSSN VWGYFSVGNNGNIHSFSDSQFGHIFAFGENRQASRKHMVVALKELSIR GDFRTTVEYLIKLLETEDFEDNTITTGWLDDLITHKMTAEKPDPTLAVI CGAATKAFLASEEARHKYIESLQKGQVLSKDLLQTMFPVDFIHEGKRY KFTVAKSGNDRYTLFINGSKCDIILRQLADGGLLIAIGGKSHTIYWKEE VAATRLSVDSMTTLLEVENDPTQLRTPSPGKLVKFLVENGEHIIKGQPY AEIEVMKMQMPLVSQENGIVQLLKQPGSTIVAGDIMAIMTLDDPSKVK HALPFEGMLPDFGSPVIEGTKPAYKFKSLVSTLENILKGYDNQVIMNAS LQQLIEVLRNPKLPYSEWKLHISALHSRLPAKLDEQMEELVARSLRRG AVFPARQLSKLIDMAVKNPEYNPDKLLGAVVEPLADIAHKYSNGLEA HEHSIFVHFLEEYYEVEKLFNGPNVREENIILKLRDENPKDLDKVALTV LSHSKVSAKNNLILAILKHYQPLCKLSSKVSAIFSTPLQHIVELESKATA KVALQAREILIQGALPSVKERTEQIEHILKSSVVKVAYGSSNPKRSEPDL NILKDLIDSNYVVFDVLLQFLTHQDPVVTAAAAQVYIRRAYRAYTIGDI RVHEGVTVPIVEWKFQLPSAAFSTFPTVKSKMGMNRAVSVADLSYVA NSQSSPLREGILMAVDHLDDVDEILSQSLEVIPRHQSSSNGPAPDRSGSS ASLSNVANVCVASTEGFESEEEILVRLREILDLNKQELINASIRRITFMF GFKDGSYPKYYTFNGPNYNENETIRHIEPALAFQLELGRLSNFNIKPIFT DNRNIHVYEAVSKTSPLDKRFFTRGIIRTGHIRDDISIQEYLTSEANRLM SDILDNLEVTDTSNSDLNHIFINFIAVFDISPEDVEAAFGGFLERFGKRLL RLRVSSAEIRIIIKDPQTGAPVPLRALINNVSGYVIKTEMYTEVKNAKGE WVFKSLGKPGSMHLRPIATPYPVKEWLQPKRYKAHLMGTTYVYDFPE LFRQASSSQWKNFSADVKLTDDFFISNELIEDENGELTEVEREPGANAI GMVAFKITVKTPEYPRGRQFVVVANDITFKIGSFGPQEDEFFNKVTEYA RKRGIPRIYLAANSGARIGMAEEIVPLFQVAWNDAANPDKGFQYLYLT SEGMETLKKFDKENSVLTERTVINGEERFVIKTIIGSEDGLGVECLRGSG LIAGATSRAYHDIFTITLVTCRSVGIGAYLVRLGQRAIQVEGQPIILTGA PAINKMLGREVYTSNLQLGGTQIMYNNGVSHLTAVDDLAGVEKIVEW MSYVPAKRNMPVPILETKDTWDRPVDFTPTNDETYDVRWMIEGRETE SGFEYGLFDKGSFFETLSGWAKGVVVGRARLGGIPLGVIGVETRTVEN LIPADPANPNSAETLIQEPGQVWHPNSAFKTAQAINDFNNGEQLPMMIL ANWRGFSGGQRDMFNEVLKYGSFIVDALVDYKQPIIIYIPPTGELRGGS WVVVDPTINADQMEMYADVNARAGVLEPQGMVGIKFRREKLLDTM NRLDDKYRELRSQLSNKSLAPEVHQQISKQLADRERELLPIYGQISLQF

SEQ ID	SEQUENCE
	ADLHDRSSRMVAKGVISKELEWTEARRFFFWRLRRRLNEEYLIKRLSH QVGEASRLEKIARIRSWYPASVDHEDDRQVATWIEENYKTLDDKLKG LKLESFAQDLAKKIRSDHDNAIDGLSEVIKMLSTDDKEKLLKTLK*
SEQ ID NO:98 Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsGOTt75	MAATTNQTEPPESDNHSVATKILNFGKACWKLQRPYTIIAFTSCACGLF GKELLHNTNLISWSLMFKAFFFLVAILCIASFTTTINQIYDLHIDRINKPD LPLASGEISVNTAWIMSIIVALFGLIITIKMKGGPLYIFGYCFGIFGGIVYS VPPFRWKQNPSTAFLLNFLAHIITNFTFYYASRAALGLPFELRPSFTFLL AFMKSMGSALALIKDASDVEGDTKFGISTLASKYGSRNLTLFCSGIVLL SYVAAILAGIIWPQAFNSNVMLLSHAILAFWLILQTRDFALTNYDPEAG RRFYEFMWKLYYAEYLVYVFI*
SEQ ID NO:99 Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsGOTt33	MSHPKTPIKYSYNNFPSKHCSTKSFHLQNKCSESLSIAKNSIRAATTNQT EPPESDNHSVATKILNFGKACWKLQRPYTIIAFTSCACGLFGKELLHNT NLISWSLMFKAFFFLVAILCIASFTTTINQIYDLHIDRINKPDLPLASGEIS VNTAWIMSIIVALFGLIITIKMKGGPLYIFGYCFGIFGGIVYSVPPFRWK QNPSTAFLLNFLAHIITNFTFYYASRAALGLPFELRPSFTFLLAFMKSMG SALALIKDASDVEGDTKFGISTLASKYGSRNLTLFCSGIVLLSYVAAILA GIIWPQAFNSNVMLLSHAILAFWLILQTRDFALTNYDPEAGRRFYEFM WKLYYAEYLVYVFI*
SEQ ID NO:100 Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4t	MSAGSDQIEGSPHHESDNSIATKILNFGHTCWKLQRPYVVKGMISIACG LFGRELFNNRHLFSWGLMWKAFFALVPILSFNFFAAIMNQIYDVDIDRI NKPDLPLVSGEMSIETAWILSIIVALTGLIVTIKLKSAPLFVFIYIFGIFAG FAYSVPPIRWKQYPFTNFLITISSHVGLAFTSYSATTSALGLPFVWRPAF SFIIAFMTVMGMTIAFAKDISDIEGDAKYGVSTVATKLGARNMTFVVS GVLLLNYLVSISIGIIWPQVFKSNIMILSHAILAFCLIFQTRELALANYAS APSRQFFEFIWLLYYAEYFVYVFI*
SEQ ID NO:101 Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT7t	MSTDTANQTEPPESNTKYSVVTKILSFGHTCWKLQRPYTFIGVISCACG LFGRELFHNTNLLSWSLMLKAFSSLMVILSVNLCTNIINQITDLDIDRIN KPDLPLASGEMSIETAWIMSIIVALTGLILTIKLNCGPLFISLYCVSILVG ALYSVPPFRWKQNPNTAFSSYFMGLVIVNFTCYYASRAAFGLPFEMSP PFTFILAFVKSMGSALFLCKDVSDIEGDSKHGISTLATRYGAKNITFLCS GIVLLTYVSAILAAIIWPQAFKSNVMLLSHATLAFWLIFQTREFALTNY NPEAGRKFYEFMWKLHYAEYLVYVFI*
SEQ ID NO:102 Truncated geranyl pyrophosphate olivetolic acid geranyltransferase HlPT1Lt	MDRPPESGNLSALTNVKDFVSVCWEYVRPYTAKGVIICSSCLFGRELL ENPNLFSWPLIFRALLGMLAILGSCFYTAGINQIFDMDIDRINKPDLPLV SGRISVESAWLLTLSPAIIGFILILKLNSGPLLTSLYCLAILSGTIYSVPPFR WKKNPITAFLCILMIHAGLNFSVYYASRAALGLAFVWSPSFSFITAFITF MTLTLASSKDLSDINGDRKFGVETFATKLGAKNITLLGTGLLLLNYVA AISTAIIWPKAFKSNIMLLSHAILAFSLFFQARELDRTNYTPEACKSFYEF IWILFSAEYVVYLFI*
SEQ ID NO:103 Truncated geranyl pyrophosphate olivetolic acid geranyltransferase HlPT2t	MGHLPRPNSLTAWSHQSEFPSTIVTKGSNFGHASWKFVRPIPFVAVSIIC TSLFGAELLKNPNLFSWQLMFDAFQGLVVILLYHIYINGLNQIYDLESD RINKPDLPLAAEEMSVKSAWFLTIFSAVASLLLMIKLKCGLFLTCMYCC YLVIGAMYSVPPFRWKMNTFTSTLWNFSEIGIGINFLINYASRATLGLPF QWRPPFTFIIGFVSTLSIILSILKDVPDVEGDKKVGMSTLPVIFGARTIVL VGSGFFLLNYVAAIGVAIMWPQAFKGYIMIPAHAIFASALIFKTWLLDK ANYAKEASDSYYHFLWFLMIAEYILYPFIST*
SEQ ID NO:104 Truncated tetrahydrocannabinolic acid synthase THCASt28	MNFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIQNLRFTSDTTPK PLVIVTPSNVSHIQASILCSKKVGLQIRTRSGGHDAEGMSYISQVPFVVV DLRNMHSIKIDVHSQTAWVEAGATLGEVYYWINEKNENFSFPGGYCP TVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMG EDLFWAIRGGGGENFGIIAAWKIKLVAVPSKSTIFSVKKNMEIHGLVKL

SEQ ID	SEQUENCE
	FNKWQNIAYKYDKDLVLMTHFITKNITDNHGKNKTTVHGYFSSIFHGG VDSLVDLMNKSFPELGIKKTDCKEFSWIDTTIFYSGVVNFNTANFKKEI LLDRSAGKKTAFSIKLDYVKKPIPETAMVKILEKLYEEDVGVGMYVLY PYGGIMEEISESAIPFPHRAGIMYELWYTASWEKQEDNEKHINWVRSV YNFTTPYVSQNPRLAYLNYRDLDLGKTNPESPNNYTQARIWGEKYFG KNFNRLVKVKTKADPNNFFRNEQSIPPLPPHHH*
SEQ ID NO:105 Truncated cannabidiolic acid synthase CBDASt28*	MNPRENFLKCFSQYIPNNATNLKLVYTQNNPLYMSVLNSTIHNLRFTS DTTPKPLVIVTPSHVSHIQGTILCSKKVGLQIRTRSGGHDSEGMSYISQV PFVIVDLRNMRSIKIDVHSQTAWVEAGATLGEVYYWVNEKNENLSLA AGYCPTVCAGGHFGGGGYGPLMRNYGLAADNIIDAHLVNVHGKVLD RKSMGEDLFWALRGGGAESFGIIVAWKIRLVAVPKSTMFSVKKIMEIH ELVKLVNKWQNIAYKYDKDLLLMTHFITRNITDNQGKNKTAIHTYFSS VFLGGVDSLVDLMNKSFPELGIKKTDCRQLSWIDTIIFYSGVVNYDTD NFNKEILLDRSAGQNGAFKIKLDYVKKPIPESVFVQILEKLYEEDIGAG MYALYPYGGIMDEISESAIPFPHRAGILYELWYICSWEKQEDNEKHLN WIRNIYNFMTPYVSKNPRLAYLNYRDLDIGINDPKNPNNYTQARIWGE KYFGKNFDRLVKVKTLVDPNNFFRNEQSIPPLPRHRH*
SEQ ID NO:106 Mutated fatty acid synthase (FAS1, I306A, R1834K)	MDAYSTRPLTLSHGSLEHVLLVPTASFFIASQLQEQFNKILPEPTEGFAA DDEPTTPAELVGKFLGYVSSLVEPSKVGQFDQVLNLCLTEFENCYLEG NDIHALAAKLLQENDTTLVKTKELIKNYITARIMAKRPFDKKSNSALFR AVGEGNAQLVAIFGGQGNTDDYFEELRDLYQTYHVLVGDLIKFSAETL SELIRTTLDAEKVFTQGLNILEWLENPSNTPDKDYLLSIPISCPLIGVIQL AHYVVTAKLLGFTPGELRSYLKGATGHSQGLVTAVAIAETDSWESFFV SVRKAITVLFFGGVRCYEAYPNTSLPPSILEDSLENNEGVPSPMLSISNL TQEQVQDYVNKTNSHLPAGKQVEISLVNGAKNLVVSGPPQSLYGLNL TLRKAKAPSGLDQSRIPFSERKLKFSNRFLPVASPFHSHLLVPASDLINK DLVKNNVSFNAKDIQIPVYDTFDGSDLRVLSGSISERIVDCIIRLPVKWE TTTQFKATHILDFGPGGASGLGVLTHRNKDGTGVRVIVAGTLDINPDD DYGFKQEIFDVTSNGLKKNPNWLEEYHPKLIKNKSGKIFVETKFSKLIG RPPLLVPGMTPCTVSPDFVAATTNAGYTIELAGGGYFSAAGMTAAIDS VVSQIEKGSTFGINLIYVNPFMLQWGIPLIKELRSKGYPIQFLTIGAGVPS LEVASEYIETLGLKYLGLKPGSIDAISQVINIAKAHPNFPIALQWTGGRG GGHHSFEDAHTPMLQMYSKIRRHPNIMLIFGSGFGSADDTYPYLTGEW STKFDYPPMPFDGFLFGSRVMIAKEVKTSPDAKKCIAACTGVPDDKWE QTYKKPTGGIVTVRSEMGEPIHKIATRGVMLWKEFDETIFNLPKNKLV PTLEAKRDYIISRLNADFQKPWFATVNGQARDLATMTYEEVAKRLVE LMFIRSTNSWFDVTWRTFTGDFLRRVEERFTKSKTLSLIQSYSLLDKPD EAIEKVFNAYPAAREQFLNAQDIDHFLSMCQNPMQKPVPFVPVLDRRF EIFFKKDSLWQSEHLEAVVDQDVQRTCILHGPVAAQFTKVIDEPIKSIM DGIHDGHIKKLLHQYYGDDESKIPAVEYFGGESPVDVQSQVDSSSVSE DSAVFKATSSTDEESWFKALAGSEINWRHASFLCSFITQDKMFVSNPIR KVFKPSQGMVVEISNGNTSSKTVVTLSEPVQGELKPTVILKLLKENIIQ MEMIENRTMDGKPVSLPLLYNFNPDNGFAPISEVMEDRNQRIKEMYW KLWIDEPFNLDFDPRDVIKGKDFEITAKEVYDFTHAVGNNCEDFVSRP DRTMLAPMDFAIVVGWRAIIKAIFPNTVDGDLLKLVHLSNGYKMIPGA KPLQVGDVVSTTAVIESVVNQPTGKIVDVVGTLSRNGKPVMEVTSSFF YRGNYTDFENTFQKTVEPVYQMHIKTSKDIAVLRSKEWFQLDDEDFD LLNKTLTFETETEVTFKNANIFSSVKCFGPIKVELPTKETVEIGIVDYEA GASHGNPVVDFLKRNGSTLEQKVNLENPIPIAVLDSYTPSTNEPYARVS GDLNPIHVSRHFASYANLPGTITHGMFSSASVRALIENWAADSVSSRVR GYTCQFVDMVLPNTALKTSIQHVGMINGRKLIKFETRNEDDVVVLTGE AEIEQPVTTFVFTGQGSQEQGMGMDLYKTSKAAQDVWNRADNHFKD TYGFSILDIVINNPVNLTIHFGGEKGKRIRENYSAMIFETIVDGKLKTEKI

2018256863 11 Apr 2024

SEQ ID	SEQUENCE
	FKEINEHSTSYTFRSEKGLLSATQFTQPALTLMEKAAFEDLKSKGLIPA DATFAGHSLGEYAALASLADVMSIESLVEVVFYFGMTMQVAVPRDEL GRSNYGMIAINPGRVAASFSQEALQYVVERVGKRTGWLVEIVNYNVE NQQYVAAGDLRALDTVTNVLNFIKLQKIDIIELQKSLSLEEVEGHLFEII DEASKKSAVKPRPLKLERGFACIPLVGISVPFHSTYLMNGVKPFKSFLK KNIIKENVKVARLAGKYIPNLTAKPFQVTKEYFQDVYDLTGSEPIKEIID NWEKYEQS*
SEQ ID NO:107 Mutated fatty acid synthase (FAS2, G1250S)	MKPEVEQELAHILLTELLAYQFASPVRWIETQDVFLKDFNTERVVEIGP SPTLAGMAQRTLKNKYESYDAALSLHREILCYSKDAKEIYYTPDPSEL AAKEEPAKEEAPAPTPAASAPAPAAAAPAPVAAAAPAAAAAEIADEPV KASLLLHVLVAHKLKKSLDSIPMSKTIKDLVGGKSTVQNEILGDLGKE FGTTPEKPEETPLEELAETFQDTFSGALGKQSSSLLSRLISSKMPGGFTIT VARKYLQTRWGLPSGRQDGVLLVALSNEPAARLGSEADAKAFLDSM AQKYASIVGVDLSSAASASGAAGAGAAAGAAMIDAGALEEITKDHKV LARQQLQVLARYLKMDLDNGERKFLKEKDTVAELQAQLDYLNAELG EFFVNGVATSFSRKKARTFDSSWNWAKQSLLSLYFEIIHGVLKNVDRE VVSEAINIMNRSNDALIKFMEYHISNTDETKGENYQLVKTLGEQLIENC KQVLDVDPVYKDVAKPTGPKTAIDKNGNITYSEEPREKVRKLSQYVQ EMALGGPITKESQPTIEEDLTRVYKAISAQADKQDISSSTRVEFEKLYSD LMKFLESSKEIDPSQTTQLAGMDVEDALDKDSTKEVASLPNKSTISKTV SSTIPRETIPFLHLRKKTPAGDWKYDRQLSSLFLDGLEKAAFNGVTFKD KYVLITGAGKGSIGAEVLQGLLQGGAKVVVTTSRFSKQVTDYYQSIYA KYGAKGSTLIVVPFNQGSKQDVEALIEFIYDTEKNGGLGWDLDAIIPFA AIPEQGIELEHIDSKSEFAHRIMLTNILRMMGCVKKQKSARGIETRPAQ VILPMSPNHGTFGGDGMYSESKLSLETLFNRWHSESWANQLTVCGAII GWTRGTGLMSANNIIAEGIEKMGVRTFSQKEMAFNLLGLLTPEVVELC QKSPVMADLNGGLQFVPELKEFTAKLRKELVETSEVRKAVSIETALEH KVVNGNSADAAYAQVEIQPRANIQLDFPELKPYKQVKQIAPAELEGLL DLERVIVVTGFAEVGPWGSARTRWEMEAFGEFSLEGCVEMAWIMGFI SYHNGNLKGRPYTGWVDSKTKEPVDDKDVKAKYETSILEHSGIRLIEP ELFNGYNPEKKEMIQEVIVEEDLEPFEASKETAEQFKHQHGDKVDIFEI PETGEYSVKLLKGATLYIPKALRFDRLVAGQIPTGWNAKTYGISDDIIS QVDPITLFVLVSVVEAFIASGITDPYEMYKYVHVSEVGNCSGSSMGGV SALRGMFKDRFKDEPVQNDILQESFINTMSAWVNMLLISSSGPIKTPVG ACATSVESVDIGVETILSGKARICIVGGYDDFQEEGSFEFGNMKATSNT LEEFEHGRTPAEMSRPATTTRNGFMEAQGAGIQIIMQADLALKMGVPI YGIVAMAATATDKIGRSVPAPGKGILTTAREHHSSVKYASPNLNMKYR KRQLVTREAQIKDWVENELEALKLEAEEIPSEDQNEFLLERTREIHNEA ESQLRAAQQQWGNDFYKRDPRIAPLRGALATYGLTIDDLGVASFHGTS TKANDKNESATINEMMKHLGRSEGNPVIGVFQKFLTGHPKGAAGAW MMNGALQILNSGIIPGNRNADNVDKILEQFEYVLYPSKTLKTDGVRAV SITSFGFGQKGGQAIVVHPDYLYGAITEDRYNEYVAKVSAREKSAYKF FHNGMIYNKLFVSKEHAPYTDELEEDVYLDPLARVSKDKKSGSLTFNS KNIQSKDSYINANTIETAKMIENMTKEKVSNGGVGVDVELITSINVEND TFIERNFTPQEIEYCSAQPSVQSSFAGTWSAKEAVFKSLGVKSLGGGAA LKDIEIVRVNKNAPAVELHGNAKKAAEEAGVTDVKVSISHDDLQAVA VAVSTKK*
SEQ ID NO:108 MBPtag	MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKF PQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWD AVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGK SALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKA GLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDT SKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLL

2018256863 11 Apr 2024

SEQ ID	SEQUENCE
	TDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNI PQMSAFWYAVRTAVINAASGRQTVDEALKDAQTRITK
SEQ ID NO:109 ProA tag	MIFDGTTMSIAIGLLSTLGIGAEA
SEQ ID NO:110 Geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4 Cannibis sativa	MGLSLVCTFSFQTNYHTLLNPHNKNPKNSLLSYQHPKTPIIKSSYDNFP SKYCLTKNFHLLGLNSHNRISSQSRSIRAGSDQIEGSPHHESDNSIATKIL NFGHTCWKLQRPYVVKGMISIACGLFGRELFNNRHLFSWGLMWKAFF ALVPILSFNFFAAIMNQIYDVDIDRINKPDLPLVSGEMSIETAWILSIIVA LTGLIVTIKLKSAPLFVFIYIFGIFAGFAYSVPPIRWKQYPFTNFLITISSH VGLAFTSYSATTSALGLPFVWRPAFSFIIAFMTVMGMTIAFAKDISDIEG DAKYGVSTVATKLGARNMTFVVSGVLLLNYLVSISIGIIWPQVFKSNIM ILSHAILAFCLIFQTRELALANYASAPSRQFFEFIWLLYYAEYFVYVFI*
SEQ ID NO:111 Artificial geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4 nucleotide sequence	atgggtttatctttggtctgcaccttctcctttcaaactaactaccacactttattgaatccacataataagaatcctaa gaactctttattgtcctaccaacacccaaagactcctattatcaagtcctcttacgataacttcccatctaagtactgtt tgactaagaatttccatttgttgggtttgaattctcacaacagaatttcctcccaatcccgttctattagagccggttct gatcaaatcgaaggttcccctcatcatgagtccgataactccattgctactaaaattttaaatttcggtcatacttgttg gaagttgcaacgtccttacgttgtcaagggtatgatctctattgcttgtggtttgttcggtagagaattgtttaacaac agacacttgttctcttggggtttgatgtggaaagctttcttcgctttggtcccaattttgtctttcaatttcttcgccgcca tcatgaaccaaatctacgatgttgatatcgaccgtatcaacaagccagacttacctttagtttccggtgaaatgtcca ttgaaactgcttggatcttgtctatcattgttgccttgactggtttaattgttactattaagttgaagtccgctccattgttt gtcttcatctacatcttcggtatcttcgctggtttcgcttactccgtcccacctattagatggaaacaatatccttttacc aatttcttgatcactatttcctctcatgttggtttggctttcacttcttactctgccaccacttctgctttaggtttgcctttc gtttggcgtcctgccttctctttcattattgctttcatgactgtcatgggtatgactattgcctttgctaaagacatttctg atatcgaaggtgatgctaagtacggtgtctctaccgttgctaccaagttaggtgctagaaatatgacttttgttgtttc tggtgtcttattgttgaactacttggtttctatctctattggtatcatttggccacaagttttcaagtctaacattatgatct tgtctcatgctattttggctttctgtttgatctttcaaactcgtgaattagccttagccaattatgcctctgccccatccc gtcaatttttcgaattcatctggttgttatactatgccgaatacttcgtttacgtcttcatttaa
SEQ ID NO:112 Truncated acyl activating enzyme AAE (CsAAE3 truncation)	MEKSGYGRDGIYRSLRPPLHLPNNNNLSMVSFLFRNSSSYPQKPALIDS ETNQILSFSHFKSTVIKVSHGFLNLGIKKNDWLIYAPNSIHFPVCFLGIIA SGAIATTSNPLYTVSELSKQVKDSNPKLIITVPQLLEKVKGFNLPTILIGP DSEQESSSDKVMTFNDLVNLGGSSGSEFPIVDDFKQSDTAALLYSSGTT GMSKGWLTHKNFIASSLMVTMEQDLVGEMDNVFLCFLPMFHVFGLAI ITYAQLQRGNTVISARFDLEKMLKDVEKYVTHLWVPPVILALSKNSM VKFNLSSIKYIGSGAAPLGKDLMEECSKWPYGIVAQGYGMTETCGIVS MEDIRGGKRNSGSAGMLASGVEAQIVSVDTLKPLPPNQLGEIWVKGPN MMQGYFNNPQATKLTIDKKGWVHTGDLGYFDEDGHLYWDRIKELIK YKGFQVAPAELEGLLVSHPEILDAWIPFPDAEAGEVPVAYWRSPNSSL TENDVKKFIAGQVASFKRLRKVTFINSVPKSASGKILR
SEQ ID NO:113 Truncated 3-hydroxy-3-methyl-glutaryl-CoA reductase (tHMG1)	MSDQLVKTEVTKKSFTAPVQKASTPVLTNKTVISGSKVKSLSSAQSSSS GPSSSSEEDDSRDIESLDKKIRPLEELEALLSSGNTKQLKNKEVAALVIH GKLPLYALEKKLGDTTRAVAVRRKALSILAEAPVLASDRLPYKNYDY DRVFGACCENVIGYMPLPVGVIGPLVIDGTSYHIPMATTEGCLVASAM RGCKAINAGGGATTVLTKDGMTRGPVVRFPTLKRSGACKIWLDSEEG QNAIKKAFNSTSRFARLQHIQTCLAGDLLFMRFRTTTGDAMGMNMISK GVEYSLKQMVEEYGWEDMEVVSVSGNYCTDKKPAAINWIEGRGKSV VAEATIPGDVVRKVLKSDVSALVELNIAKNLVGSAMAGSVGGFNAHA ANLVTAVFLALGQDPAQNVESSNCITLMKEVDGDLRISVSMPSIEVGTI GGGTVLEPQGAMLDLLGVRGPHATAPGTNARQLARIVACAVLAGELS LCAALAAGHLVQSHMTHNRKPAEPTKPNNLDATDINRLKDGSVTCIKS *
SEQ ID NO:114	atgtccgatcaattagtcaagaccgaagtcaccaagaagtccttcaccgccccagttcaaaaagcctccactcca

SEQ ID	SEQUENCE
Truncated 3-hydroxy-3-methyl-glutaryl-CoA reductase (tHMG1)	gtcttaactaacaagaccgttatttctggttccaaggttaagtctttatcctccgctcaatcctcctcttccggtccatc ttcttcttctgaagaagatgattcccgtgacattgagtccttggataagaaaatcagacctttggaagaattagaag ctttgttatcctctggtaatactaagcaattgaaaaacaaggaagttgctgctttggttattcacggtaaattacctttg tacgctttagagaagaagttaggtgacactacccgtgctgtcgccgttagaagaaaagctttgtctattttagctga ggcccctgttttggcttctgacagattaccatacaagaattacgattacgatagagttttcggtgcctgttgtgagaa cgttatcggttatatgccattaccagtcggtgttatcggtccattggttattgacggtacctcttaccacatcccaatg gctactactgaaggttgtttagtcgcctccgccatgagaggttgtaaggctatcaatgctggtggtggtgctactac cgtcttgactaaggatggtatgactagaggtccagttgtccgttttccaactttgaaaagatctggtgcttgtaagatt tggttggattctgaagaaggtcaaaatgccattaagaaggctttcaattccacctctagatttgccagattacaacat attcaaacctgtttagccggtgatttgttgttcatgagattcagaactactactggtgatgctatgggtatgaacatga tctctaagggtgtcgaatattctttaaaacaaatggttgaagagtatggttgggaagacatggaggtcgtctctgtct ctggtaactactgtactgataagaaaccagctgctatcaactggatcgaaggtcgtggtaagtctgttgttgccga agctactattccaggtgatgttgttagaaaggttttaaaatccgatgtctctgccttggttgagttgaacattgctaaa aacttggttggttctgctatggctggttctgtcggtggttttaatgcccatgccgccaacttagtcaccgccgttttctt agctttgggtcaagatccagctcaaaatgtcgaatcctccaactgtatcactttgatgaaagaggtcgacggtgac ttgcgtatctctgtttccatgccatctatcgaagttggtactatcggtggtggtactgttttggagccacaaggtgcta tgttggacttattgggtgttagaggtccacacgccactgctcctggtaccaacgccagacaattagctagaatcgt tgcctgtgccgtcttagctggtgagttgtctttatgtgctgccttagctgctggtcacttggtccaatcccacatgact cataacagaaagccagctgaacctaccaagcctaacaacttggatgccaccgatattaatcgtttaaaagatggtt ctgtcacttgcattaagtcctaa
SEQ ID NO:115 HMG-CoA synthase (Sc_ERG13) Saccharomyces sp.	MTELKKQKTAEQKTRPQNVGIKGIQIYIPTQCVNQSELEKFDGVSQGK YTIGLGQTNMSFVNDREDIYSMSLTVLSKLIKSYNIDTNKIGRLEVGTE TLIDKSKSVKSVLMQLFGENTDVEGIDTLNACYGGTNALFNSLNWIES NAWDGRDAIVVCGDIAIYDKGAARPTGGAGTVAMWIGPDAPIVFDSV RASYMEHAYDFYKPDFTSEYPYVDGHFSLTCYVKALDQVYKSYSKKA ISKGLVSDPAGSDALNVLKYFDYNVFHVPTCKLVTKSYGRLLYNDFRA NPQLFPEVDAELATRDYDESLTDKNIEKTFVNVAKPFHKERVAQSLIVP TNTGNMYTASVYAAFASLLNYVGSDDLQGKRVGLFSYGSGLAASLYS CKIVGDVQHIIKELDITNKLAKRITETPKDYEAAIELRENAHLKKNFKP QGSIEHLQSGVYYLTNIDDKFRRSYDVKK*
SEQ ID NO:116 Artificial HMG-CoA synthase (Sc_ERG13) nucleotide sequence	atgaccgaattgaagaagcaaaagactgctgaacaaaaaacccgtccacaaaatgttggtatcaagggtattca aatctacattccaactcaatgcgtcaaccaatctgaattggaaaaatttgatggtgtttctcaaggtaaatacactatt ggtttgggtcaaactaatatgtccttcgttaacgacagagaagatatttactccatgtccttgaccgtcttgtccaaat tgattaagtcttataatattgacaccaacaagatcggtagattggaagttggtactgaaactttgattgataagtctaa gtctgttaagtctgttttaatgcaattgttcggtgaaaatactgacgttgaaggtattgacactttgaacgcttgttacg gtggtactaatgctttatttaactctttgaactggattgaatccaacgcttgggacggtagagatgccattgtcgtttg tggtgacattgctatctatgacaagggtgccgctcgtccaactggtggtgctggtaccgttgctatgtggatcggt cctgacgccccaatcgtttttgattctgttcgtgcttcttacatggagcatgcttacgacttttataagccagactttac ctccgaatatccatacgtcgatggtcatttctctttgacctgctacgttaaagccttagatcaagtctacaagtcttact ccaagaaggccatttccaagggtttagtctccgatccagctggttccgatgctttaaacgttttaaagtacttcgatt acaacgtttttcatgtccctacttgtaaattggttaccaaatcttacggtagattattgtacaacgatttcagagctaat ccacaattatttccagaagtcgatgctgagttggctactagagattacgacgaatccttgaccgacaaaaatattga aaagacttttgttaacgttgctaagccatttcacaaagagagagttgcccaatctttgattgtcccaactaatactggt aatatgtatactgcttctgtttacgctgcctttgcttctttgttgaactatgtcggttctgacgacttacaaggtaagcgt gtcggtttgttctcctacggttccggtttggctgcctctttgtattcttgtaagattgtcggtgatgttcaacacatcatc aaggaattggatatcaccaataaattggccaagagaatcactgaaactcctaaagactatgaagctgctatcgaat tgagagaaaatgctcatttaaagaaaaactttaaaccacaaggttctattgaacacttgcaatccggtgtttactact taactaacatcgatgacaagttccgtagatcctacgacgtcaagaagtaa
SEQ ID NO:117 Pyruvate decarboxylase (Zm_PDC)	MSYTVGTYLAERLVQIGLKHHFAVAGDYNLVLLDNLLLNKNMEQVY CCNELNCGFSAEGYARAKGAAAAVVTYSVGALSAFDAIGGAYAENLP VILISGAPNNNDHAAGHVLHHALGKTDYHYQLEMAKNITAAAEAIYT PEEAPAKIDHVIKTALREKKPVYLEIACNIASMPCAAPGPASALFNDEA SDEASLNAAVEETLKFIANRDKVAVLVGSKLRAAGAEEAAVKFADAL

2018256863 11 Apr 2024

SEQ ID	SEQUENCE
Zymomonas mobilis	GGAVATMAAAKSFFPEENPHYIGTSWGEVSYPGVEKTMKEADAVIAL APVFNDYSTTGWTDIPDPKKLVLAEPRSVVVNGIRFPSVHLKDYLTRL AQKVSKKTGALDFFKSLNAGELKKAAPADPSAPLVNAEIARQVEALLT PNTTVIAETGDSWFNAQRMKLPNGARVEYEMQWGHIGWSVPAAFGY AVGAPERRNILMVGDGSFQLTAQEVAQMVRLKLPVIIFLINNYGYTIEV MIHDGPYNNIKNWDYAGLMEVFNGNGGYDSGAGKGLKAKTGGELAE AIKVALANTDGPTLIECFIGREDCTEELVKWGKRVAAANSRKPVNKLL *
SEQ ID NO:118 Artificial pyruvate decarboxylase (Zm_PDC) nucleotide sequence	atgtcctacaccgttggtacctacttagctgagcgtttggtccaaatcggtttgaagcaccatttcgccgttgctggt gattacaacttggtcttgttagataatttattattgaacaagaacatggaacaagtctactgctgtaatgaattgaact gtggtttctctgctgaaggttatgctagagctaaaggtgccgctgccgctgttgtcacttactctgttggtgctttgtc tgccttcgacgctattggtggtgcttacgccgagaatttacctgttatttttaatttctggtgcccctaacaataacgat catgctgctggtcatgttttacaccacgctttgggtaaaactgactaccattatcaattagagatggccaaaaacat caccgccgctgccgaggccatttacactccagaagaagccccagccaaaattgatcacgtcatcaaaaccgcc ttgagagagaaaaaacctgtttacttggaaatcgcctgtaatatcgcctctatgccttgcgccgctcctggtcctgc ttccgccttattcaacgatgaggcttctgatgaagcttccttaaacgctgctgttgaggagactttaaagttcatcgct aatagagataaggtcgctgttttagtcggttctaagttgcgtgctgccggtgccgaggaagctgctgttaaattcgc cgatgctttaggtggtgctgtcgccaccatggccgccgccaaatccttttccctgaagaaaacccacactacatc ggtacttcttggggtgaagtctcttacccaggtgtcgaaaagactatgaaggaagccgatgccgtcatcgccttg gccccagtttttaatgattattccaccactggttggactgatatcccagatcctaaaaagttagttttagccgagccta gatccgttgttgttaacggtattagattcccttccgttcacttgaaggattacttaactagattggctcaaaaggtttcc aagaagaccggtgctttggacttttcaaatctttgaacgccggtgagttaaagaaggccgcccctgctgacccat ctgctccattggttaacgctgagattgctagacaagtcgaagctttattgaccccaaacactaccgttatcgccgaa actggtgactcttggtttaatgctcaaagaatgaagttaccaaatggtgccagagttgagtacgaaatgcaatggg gtcatatcggttggtctgtcccagctgcttttggttatgctgttggtgcccctgagagaagaaacatcttgatggttg gtgacggttccttccaattgactgctcaagaagtcgctcaaatggttagattaaaattaccagtcatcatcttcttgat caataactacggttacactatcgaagtcatgattcacgatggtccttacaataatattaagaactgggactatgctg gtttgatggaagtctttaatggtaacggtggttacgattccggtgctggtaagggtttaaaggctaagactggtggt gaattagctgaagccattaaggttgccttggctaacaccgacggtcctactttaatcgaatgtttcattggtagaga ggattgtaccgaagagttagttaagtggggtaagagagttgccgctgctaattcccgtaagcctgtcaataaattg ttataa
SEQ ID NO:119 Truncated 3-hydroxy-3-methyl-glutaryl-CoA reductase (tHMG1)	atgcaattggtgaagactgaagtcaccaagaagtcttttactgctcctgtacaaaaggcttctacaccagttttaac caataaaacagtcatttctggatcgaaagtcaaaagtttatcatctgcgcaatcgagctcatcaggaccttcatcat ctagtgaggaagatgattcccgcgatattgaaagcttggataagaaaatacgtcctttagaagaattagaagcatt attaagtagtggaaatacaaaacaattgaagaacaaagaggtcgctgccttggttattcacggtaagttacctttgt acgctttggagaaaaaattaggtgatactacgagagcggttgcggtacgtaggaaggctctttcaattttggcaga agctcctgtattagcatctgatcgtttaccatataaaaattatgactacgaccgcgtatttggcgcttgttgtgaaaat gttataggttacatgcctttgcccgttggtgttataggcccttggttatcgatggtacatcttatcatataccaatggc aactacagagggttgtttggtagcttctgccatgcgtggctgtaaggcaatcaatgctggcggtggtgcaacaac tgttttaactaaggatggtatgacaagaggcccagtagtccgtttcccaactttgaaaagatctggtgcctgtaaga tatggttagactcagaagagggacaaaacgcaattaaaaaagcttttaactctacatcaagatttgcacgtctgca acatattcaaacttgtctagcaggagatttactcttcatgagatttagaacaactactggtgacgcaatgggtatgaa tatgatttctaagggtgtcgaatactcattaaagcaaatggtagaagagtatggctgggaagatatggaggttgtct ccgtttctggtaactactgtaccgacaaaaaaccagctgccatcaactggatcgaaggtcgtggtaagagtgtcg tcgcagaagctactattcctggtgatgttgtcagaaaagtgttaaaaagtgatgtttccgcattggttgagttgaaca ttgctaagaatttggttggatctgcaatggctgggtctgttggtggatttaacgcacatgcagctaatttagtgacag ctgttttcttggcattaggacaagatcctgcacaaaatgtcgaaagttccaactgtataacattgatgaaagaagtg gacggtgatttgagaatttccgtatccatgccatccatcgaagtaggtaccatcggtggtggtactgttctagaacc acaaggtgccatgttggacttattaggtgtaagaggcccacatgctaccgctcctggtaccaacgcacgtcaatt agcaagaatagttgcctgtgccgtcttggcaggtgaattatccttatgtgctgccctagcagccggccatttggttc aaagtcatatgacccacaacaggaaacctgctgaaccaacaaaacctaacaatttggacgccactgatataaatc gtttgaaagatgggtccgtcacctgcattaaatcctaa
SEQ ID NO:120	atgactgaactaaaaaaacaaaagaccgctgaacaaaaaaccagacctcaaaatgtcggtattaaaggtatcca

2018256863 11 Apr 2024

SEQ ID	SEQUENCE
HMG-CoA synthase (Sc_ERG13) Saccharomyces sp.	aatttacatcccaactcaatgtgtcaaccaatctgagctagagaaatttgatggcgtttctcaaggtaaatacacaat tggtctgggccaaaccaacatgtcttttgtcaatgacagagaagatatctactcgatgtccctaactgttttgtctaa gttgatcaagagttacaacatcgacaccaacaaaattggtagattagaagtcggtactgaaactctgattgacaag tccaagtctgtcaagtctgtcttgatgcaattgtttggtgaaaacactgacgtcgaaggtattgacacgcttaatgcc tgttacggtggtaccaacgcgttgttcaactctttgaactggattgaatctaacgcatgggatggtagagacgccat tgtagtttgcggtgatattgccatctacgataagggtgccgcaagaccaaccggtggtgccggtactgttgctatg tggatcggtcctgatgctccaattgtatttgactctgtaagagcttcttacatggaacacgcctacgatttttacaagc cagatttcaccagcgaatatccttacgtcgatggtcatttttcattaacttgttacgtcaaggctcttgatcaagtttac aagagttattccaagaaggctatttctaaagggttggttagcgatcccgctggttcggatgctttgaacgttttgaaa tatttcgactacaacgttttccatgttccaacctgtaaattggtcacaaaatcatacggtagattactatataacgattt cagagccaatcctcaattgttcccagaagttgacgccgaattagctactcgcgattatgacgaatctttaaccgata agaacattgaaaaaactttgttaatgttgctaagccattccacaaagagagagttgcccaatctttgattgttccaa caaacacaggtaacatgtacaccgcatctgtttatgccgcctttgcatctctattaaactatgttggatctgacgactt acaaggcaagcgtgttggtttattttcttacggttccggtttagctgcatctctatattcttgcaaaattgttggtgacgt ccaacatattatcaaggaattagatattactaacaaattagccaagagaatcaccgaaactccaaaggattacgaa gctgccatcgaattgagagaaaatgcccatttgaagaagaacttcaaacctcaaggttccattgagcatttgcaaa gtggtgtttactacttgaccaacatcgatgacaaatttagaagatcttacgatgttaaaaaataat
SEQ ID NO:121 Geranylgeranyl pyrophosphate synthase (Cr_GPPS) Catharanthus sp.	MLFSRGLYRIARTSLNRSRLLYPLQSQSPELLQSFQFRSPIGSSQKVSGF RVIYSCVSSALANVGQQVQRQSNSVAEEPLDPFSLVADELSILANRLRS MVVAEVPKLASAAEYFFKLGVEGKRFRPTVLLLMATAIDAPISRTPPD TSLDTLSTELRLRQQSIAEITEMIHVASLLHDDVLDDAETRRGIGSLNFV MGNKLAVLAGDFLLSRACVALASLKNTEVVSLLATVVEHLVTGETMQ MTTTSDQRCSMEYYMQKTYYKTASLISNSCKAIALLAGQTSEVAMLA YEYGKNLGLAFQLIDDVLDFTGTSASLGKGSLSDIRHGIVTAPILFAIEE FPELRAVVDEGFENPYNVDLALHYLGKSRGIQRTRELAIKHANLASDA IDSLPVTDDEHVLRSRRALVELTQRVITRRK*
SEQ ID NO:122 Artificial geranylgeranyl pyrophosphate synthase (Cr_GPPS) nucleotide sequence	atgttattctctcgtggtttatacagaatcgccagaacttctttgaacagatcccgtttgttgtaccctttacaatctcaa tctcctgaattgttacaatccttccaattcagatctccaatcggttcctctcaaaaggtttccggtttcagagttatcta ctcctgcgtttcctctgctttagctaacgttggtcaacaagtccaaagacaatctaattccgttgctgaagaacctttg gacccattctccttggttgccgatgaattatccattttagctaacagattgcgttctatggtcgtcgctgaagttccaa agttagcctccgccgccgaatatttcttcaagttgggtgtcgagggtaaaagattcagaccaactgttttgttgttaa tggccaccgccattgatgccccaatctctagaaccccacctgacacctccttagatactttatccaccgaattgcgt ttgagacaacaatctatcgccgaaattactgaaatgattcatgtcgcttccttgttgcacgatgatgttttggatgatg ctgaaactagaagaggtattggttctttaaatttgtcatgggtaacaaattggctgttttggccggtgacttcttattat ctagagcttgtgttgccttagcttctttgaaaaacactgaagtcgtctccttgttagccactgtcgttgaacacttagtt actggtgagactatgcaaatgactaccacctccgatcaaagatgttctatggaatactacatgcaaaagacctatt acaagactgcctctttgatttctaactcctgtaaagccattgccttgttagctggtcaaacttctgaagttgccatgttg gcttacgaatacggtaaaaacttgggtttggctttccaattgattgatgatgttttggatttcactggtacttctgcttcc ttaggtaaaggttctttgtctgatattcgtcacggtatcgttaccgccccaatcttgttcgctattgaagaattcccag agttaagagctgttgttgacgaaggtttcgaaaaccccttacaatgttgacttagccttgcactacttgggtaaatcta gaggtattcaacgtaccagagaattagccattaaacatgctaacttagcctctgacgccattgactctttaccagtc actgatgatgagcacgtcttacgttccagacgtgccttagttgaattgactcaaagagttattactagaagaaagta a
SEQ ID NO:123 Geranylgeranyl pyrophosphate synthase (Mi_GPPS1) Mangifera indica	MLFSYGLSRISINPRASLLTCRWLLSHLTGSLSPSTSSHTISDSVHKVWG CREAYTWSVPALHGFRHQIHHQSSSLIEDQLDPFSLVADELSLVANRLR SMVVTEVPKLASAAEYFFKMGVEGKRFRPAVLLLMATALNVHVLEPL PEGAGDALMTELRTRQQCIAEITEMIHVASLLHDDVLDDADTRRGIGS LNLVMGNKLAVLAGDFLLSRACVALASLKNTEVVSLLATVVEHLVTG ETMQMTTSSDQRCSMEYYMQKTYYKTASLISNSCKAIALLAGQSAEV AMLAFEFGKNLGLAYQLIDDVLDFTGTSASLGKGSLSDIRHGIVTAPIL FAMEEFPQLRAVIDQGFENPSNVDVALEYLGKSRGIQRTRELATNHAN LAAAAIDALPKTDNEEVRKSRRALLDLTQRVITRNK*

SEQ ID	SEQUENCE
SEQ ID NO:124 Artificial geranylgeranyl pyrophosphate synthase (Mi_GPPS1) nucleotide sequence	atgttattctcttatggtttatctcgtatttctattaaccctcgtgcctctttattgacttgtagatggttattatcccatttga ctggttctttatctccttccacttcttcccatactatttctgactccgtccataaagtctggggttgcagagaagcctat acttggtctgtcccagctttacatggttttagacatcaaatccaccatcaatcctcttccttgattgaagatcaattaga cccattctccttggtcgccgatgagttgtccttggttgctaaccgtttaagatctatggttgtcactgaagtccctaaa ttagcctctgccgccgaatacttttttcaagatgggtgtcgaaggtaagcgtttcagaccagctgtcttgttgttaatg gccactgccttaaacgttcatgttttggaacctttgcctgaaggtgctggtgacgctttaatgaccgagttgagaac ccgtcaacaatgcattgctgaaatcactgagatgatccacgtcgcctctttattgcatgacgatgttttagacgacg ctgatactagaagaggtattggttctttgaacttggttatgggtaacaaattggccgttttggccggtgatttcttgtta tcccgtgcttgcgttgctttagcttctttgaagaacactgaagttgtttctttgttggccaccgtcgttgaacacttagtt actggtgagactatgcaaatgaccacctcttctgaccaaagatgttccatggaatattacatgcaaaaaacttatta caaaaccgcctccttgatttctaactcctgtaaagccatcgccttattagctggtcaatctgctgaagttgccatgtta gccttcgagtttggtaagaacttgggtttagcttaccaattgatcgatgatgtcttggattttaccggtacctctgcttc tttgggtaagggttccttgtccgacattagacacggtattgttaccgccccaatcttattcgctatggaagagtttcc acaattgagagctgttatcgaccaaggtttcgagaacccatctaacgttgacgtcgccttagagtatttaggtaaat ctagaggtatccaacgtacccgtgaattagctactaaccatgctaacttagccgccgccgccatcgatgccttgc ctaaaaccgataatgaagaagtccgtaagtccagacgtgctttattagatttgactcaaagagtcatcaccagaaa caaatag
SEQ ID NO:125 Geranylgeranyl pyrophosphate synthase (Mi_GPPS2) Mangifera indica	MPFVVPRRNRSLSVSAVLTKEETLREEEEDPKPVFDFKSYMLQKGNSV NQALDAVVSIREPKKIHEAMRYSLLAGGKRVRPVLCIAACELVGGNES MAMPAACAVEMIHTMSLIHDDLPCMDNDDLRRGKPTNHKVFGEDVA VLAGDALLAFSFENMAVSTVGVLPSRVVKAVGELAKSIGIEGLVAGQV VDINSEGLKEVGLDHLEFIHQHKTAALLEGSVVLGAILGGGSDDEVEK LRTFARCIGLLFQVVDDILDVTKSSRELGKTAGKDLVADKVTYPKLLGI EKSRELADKLNKDAQQQLSGFDQEKAAPLIALSNYIAYRQN*
SEQ ID NO:126 Artificial geranylgeranyl pyrophosphate synthase (Mi_GPPS2) nucleotide sequence	atgccattcgttgttcctagaagaaaccgttctttgtccgtttccgccgttttgaccaaggaagaaactttaagagag gaagaagaagatccaaagccagttttcgacttcaaatcttacatgttacaaaagggtaattctgttaatcaagctttg gatgctgtcgtttccattagagaacctaagaaaatccatgaggctatgcgttactctttgttggctggtggtaagag agttcgtcctgttttgtgtattgccgcctgtgaattggtcggtggtaacgaatctatggctatgccagccgcctgtgc tgtcgaaatgatccacactatgtccttgattcacgatgatttgccatgtatggataatgacgatttgcgtcgtggtaa acctaccaaccataaagttttcggtgaagacgtcgccgttttggctggtgacgctttattagctttttccttcgagaac atggccgtttccactgttggtgtcttaccatccagagttgtcaaggctgttggtgaattggccaagtctatcggtatt gaaggtttggttgccggtcaagtcgtcgatattaattctgagggtttaaaagaggtcggtttagatcacttagaattt atccatcaacacaaaaccgctgctttgttggagggttctgttgttttgggtgctattttaggtggtggttctgatgatg aagtcgaaaagttgcgtaccttgctagatgtatcggtttgttgtttcaagttgttgacgatattttggatgtcactaag tcttccagagaattgggtaagactgccggtaaagatttggttgctgataaagttacttatccaaagttgttaggtattg aaaagtctcgtgaattggccgataagttaaacaaggatgctcaacaacaattatccggttttgatcaagagaaggc tgccccttaatcgctttgtccaattacatcgcctacagacaaaactag
SEQ ID NO:127 Truncated geranylgeranyl pyrophosphate synthase Cs2_GPPS_NTrunc	MVIAEVPKLASAAEYFFKMGVEGKRFRPTVLLLMATALNVRVPEPLH DGVEDASATELRTRQQCIAEITEMIHVASLLHDDVLDDADTRRGIGSL NFVMGNKLAVLAGDFLLSRACVALASLKNTEVVTLLATVVEHLVTGE TMQMTTSSDQRCSMDYYMQKTYYKTASLISNSCKAIALLAGQTAEVA ILAFDYGKNLGLAYQLIDDVLDFTGTSASLGKGSLSDIRHGIITAPILFA MEEFPQLRTVVEQGFEDSSNVDIALEYLGKSRGIQKTRELAVKHANLA AAAIDSLPENNDEDVTKSRRALLDLTHRVITRNK*
SEQ ID NO:128 Truncated geranylgeranyl pyrophosphate synthase Cs2_GPPS_NTrunc	atggtcattgctgaagttcctaaattagcctctgccgccgaatacttcttcaagatgggtgtcgagggtaagagatt tcgtcctaccgttttgttgttaatggccaccgccttaaacgtcagagtccctgaaccattacatgatggtgttgaaga tgcctctgccaccgagttgagaactagacaacaatgtattgctgaaatcaccgagatgattcacgttgcctctttgtt gcacgatgatgttttggatgatgctgatacccgtcgtggtatcggttctttgaactttgtcatgggtaacaagttggct gtcttggctggtgatttcttattgtctcgtgcctgcgttgccttagcctctttaaaaaatacgaagttgttacttttattg gccactgttgttgagcacttggttactggtgaaactatgcaaatgaccacctcttccgaccaacgttgttccatgga ctattacatgcaaaagacctactacaagaccgcttctttgatttccaattcttgtaaagccattgccttattagctggtc aaactgctgaagttgccatcttggccttcgactacggtaaaaacttgggtttagcttaccaattgattgatgacgtttt agattttactggtacttctgcttctttgggtaaaggttctttatccgatattcgtcatggtatcattaccgctccaatctta

2018256863 11 Apr 2024

SEQ ID	SEQUENCE
	ttcgctatggaagaatttcctcaattgcgtactgtcgttgaacaaggtttcgaagactcctccaacgttgacattgcc ttagaatacttgggtaagtctcgtggtattcaaaagacccgtgaattagccgttaaacatgccaacttagccgccg ccgccatcgattccttgcctgaaaacaacgatgaggatgtcaccaagtcccgtcgtgctttgttagatttaactcac agagttattacccgtaacaagtaa
SEQ ID NO:129 Geranylgeranyl pyrophosphate synthase (Qr_GPPS) Quercus sp.	MLFSRISRIRRPGSNGFRWFLSHKTHLQFLNPPAYSYSSTHKVLGCREIF SWGLPALHGFRHNIHHQSSSIVEEQNDPFSLVADELSMVANRLRSMVV TEVPKLASAAEYFFKMGVEGKRFRPTVLLLMATAMNISILEPSLRGPG DALTTELRARQQRIAEITEMIHVASLLHDDVLDDADTRRGIGSLNFVM GNKLAVLAGDFLLSRACVALASLKNTEVVSLLAKVVEHLVTGETMQ MTTTCEQRCSMEYYMQKTYYKTASLISNSCKAIALLGGQTSEVAMLA YEYGKNLGLAYQLIDDVLDFTGTSASLGKGSLSDIRHGIITAPILFAMEE FPQLREVVDRGFDDPANVDVALDYLGKSRGIQRARELAKKHANIAAE AIDSLPESNDEDVRKSRRALLDLTERVITRTK*
SEQ ID NO:130 Artificial geranylgeranyl pyrophosphate synthase (Qr_GPPS) nucleotide sequence	atgttgttctctcgtatttctcgtatccgtagaccaggttctaatggtttcagatggttcttgtcccataagactcattta caattcttgaaccctccagcttattcctactcttccactcataaggtcttgggttgtagagaaatttttttcctggggttta cctgccttacatggtttcagacacaacattcaccaccaatcttcctctattgttgaagaacaaaatgacccttctctttt ggtcgctgatgagttgtccatggttgctaacagattgcgttctatggttgttactgaagttcctaaattagcctccgcc gctgaatacttttttaaaatgggtgttgaaggtaagagattcagaccaactgttttattgttgatggctaccgccatga acatttccatcttagaaccatctttgagaggtccaggtgacgctttgaccactgaattgagagccagacaacaaag aattgctgaaattaccgagatgatccacgttgcttccttgttgcacgatgacgttttggatgacgctgatactagaag aggtattggttccttaaactttgtcatgggtaataaattagctgttttggctggtgatttttttgttatctcgtgcctgtgttg ctttagcttctttgaagaacaccgaagttgtctccttgttagccaaggtcgtcgaacacttggttactggtgaaactat gcaaatgaccactacttgtgaacaaagatgttccatggaatactacatgcaaaagacttactataagaccgcttctt taatttccaactcctgtaaagccattgctttattaggtggtcaaacttctgaggtcgctatgttagcctacgaatatgg taaaaacttgggtttagcttaccaattgattgatgatgtcttggatttcactggtacttctgcttccttgggtaagggttc cttgtctgatattagacatggtatcattactgctccaattttgtttgctatggaagaattcccacaattacgtgaagttgt cgatagaggtttcgacgatcctgccaacgtcgatgttgccttggactacttgggtaagtctagaggtatccaaaga gccagagagttagctaaaaaacacgctaacattgctgccgaagccatcgactctttgccagaatccaacgacga ggacgtcagaaagtcccgtcgtgctttgttggacttgaccgaaagagtcattactcgtactaagtaa
SEQ ID NO:131 Truncated geranylgeranyl pyrophosphate synthase Pa_GPPS_Ntrunc	MYTRCILRDKYSRFNLRRKFFTSAKSINALNGLPDSGNPRGESNGISQF EIQQVFRCKEYIWIDRHKFHDVGFQAHHKGSITDEEQVDPFSLVADELS ILANRLRSMILTEIPKLGTAAEYFFKLGVEGKRFRPMVLLLMASSLTIGI PEVAADCLRKGLDEEQRLRQQRIAEITEMIHVASLLHDDVLDDADTRR GVGSLNFVMGNKLAVLAGDFLLSRASVALASLKNTEVVELLSKVLEH LVTGEIMQMTNTNEQRCSMEYYMQKTFYKTASLMANSCKAIALIAGQ PAEVCMLAYDYGRNLGLAYQLLDDVLDFTGTTASLGKGSLSDIRQGIV TAPILFALEEFPQLHDVINRKFKKPGDIDLALEFLGKSDGIRKAKQLAA QHAGLAAFSVESFPPSESEYVKLCRKALIDLSEKVITRTR*
SEQ ID NO:132 Truncated geranylgeranyl pyrophosphate synthase Pa_GPPS_N_trunc	atgtatacccgttgcattttaagagacaagtattctcgtttcaacttgagacgtaaattcttcacttccgctaaatccat caatgccttgaatggtttacctgactctggtaaccctagaggtgaatctaacggtatctcccaattcgaaattcaac aagttttccgttgtaaagaatacatttggatcgatcgtcacaagttccacgatgttggttttcaagctcatcacaagg gttccatcactgacgaggaacaagttgacccttttttctttagtcgctgatgaattgtccatcttagctaatcgtttaaga tccatgatcttaaccgagattccaaagttaggtaccgctgccgaatacttttttcaagttgggtgtcgaaggtaagag atttagaccaatggttttgttgttgatggcctcctctttaactattggtatccctgaagttgccgctgattgtttgcgtaa gggtttggacgaagaacaaagattacgtcaacaacgtatcgctgaaattactgaaatgattcatgtcgcctctttgt tgcacgatgatgttttggatgacgccgatactagacgtggtgttggttccttgaactttgttatgggtaacaagttgg ctgttttagccggtgatttcttgttatctagagcttctgttgccttagcttctttaaagaacactgaggttgttgagttatt gtctaaggttttggagcacttagtcactggtgagatcatgcaaatgactaacactaatgaacaaagatgttctatgg aatattacatgcaaaagactttctacaagaccgcctctttgatggctaattcttgtaaagccattgccttgatcgctgg tcaacctgccgaagtctgcatgttggcctacgactacggtagaaacttgggtttagcttatcaattattggatgacgt tttggatttcactggtaccactgcttctttaggtaagggttccttatccgacatcagacaaggtattgttactgcccct attttattcgctttggaagaattccctcaattacacgacgtcatcaaccgtaagttcaaaaaaccaggtgacatcgat ttggccttggaatttttgggtaagtctgatggtatccgtaaagccaaacaattggctgctcaacatgctggtttagct

SEQ ID	SEQUENCE
	gccttttctgtcgaatcctttccaccatctgaatccgaatacgttaagttatgtagaaaggccttgatcgatttgtctga aaaggtcattactcgtaccagataa
SEQ ID NO:133 Geranylgeranyl pyrophosphate synthase Ag_GPPS Abies grandis	MAYSAMATMGYNGMAASCHTLHPTSPLKPFHGASTSLEAFNGEHMG LLRGYSKRKLSSYKNPASRSSNATVAQLLNPPQKGKKAVEFDFNKYM DSKAMTVNEALNKAIPLRYPQKIYESMRYSLLAGGKRVRPVLCIAACE LVGGTEELAIPTACAIEMIHTMSLMHDDLPCIDNDDLRRGKPTNHKIFG EDTAVTAGNALHSYAFEHIAVSTSKTVGADRILRMVSELGRATGSEGV MGGQMVDIASEGDPSIDLQTLEWIHIHKTAMLLECSVVCGAIIGGASEI VIERARRYARCVGLLFQVVDDILDVTKSSDELGKTAGKDLISDKATYP KLMGLEKAKEFSDELLNRAKGELSCFDPVKAAPLLGLADYVAFRQN*
SEQ ID NO:134 Artificial geranylgeranyl pyrophosphate synthase (Ag_GPPS) nucleotide sequence	atggcttattctgctatggctactatgggttacaacggtatggctgcttcttgtcacactttacacccaacttctccatt gaaaccttttcacggtgcttctacttccttggaagccttcaatggtgaacacatgggtttgttaagaggttattctaag cgtaagttgtcctcttacaaaaatccagcttctcgttcctccaatgctaccgtcgctcaattattgaacccaccacaa aagggtaagaaggctgttgaatttgacttcaataagtatatggattctaaggctatgaccgtcaacgaggctttgaa taaagccatcccattgcgttacccacaaaagatctacgaatctatgagatattctttgttagctggtggtaagagagt ccgtccagttttgtgtatcgccgcttgtgaattagtcggtggtactgaggagttagctattccaaccgcctgtgccat cgaaatgatccacaccatgtctttgatgcacgatgatttgccatgtatcgacaacgatgacttgagacgtggtaaa cctaccaatcataagattttcggtgaagatactgctgttactgccggtaacgctttacactcttacgccttcgaacat attgctgtttctacttccaagactgttggtgctgatagaattttgagaatggtttctgaattaggtcgtgctactggttc cgaaggtgttatgggtggtcaaatggtcgatattgcttctgaaggtgacccttccattgatttgcaaactttagaatg gatccacatccacaagactgctatgttattagaatgttctgttgtctgtggtgccatcatcggtggtgcttctgaaatt gttattgagagagccagacgttatgctcgttgtgtcggtttattgtttcaagttgttgacgacattttagatgttaccaa atcttctgacgaattgggtaaaactgctggtaaagatttaatctccgataaagccacctaccctaagttgatgggttt ggagaaggccaaagagttttccgatgaattattaaacagagctaaaggtgaattgtcttgcttcgatccagttaag gctgccccattgttaggtttggctgactacgttgccttcagacaaaactaa
SEQ ID NO:135 Truncated geranylgeranyl pyrophosphate synthase Pb_GPPS_NTrunc	MAAIFPSIPSNFKPPQISQTLTRRRRPNRTLCTATSDQSYLSASSADIYSH LLRSLPATIHPSVKAPIHSLLSSPIPPTIAPPLCLAATELVGGNPNSAINAA CAIHLIHAVTHTRTAPPLAEFSPGVLLMTGDGLLVLAYEMLARSPAVD ADTSVRVLKEVARTAAAVAAAYEGGREGELAAGAAACGVILGGGNE EEVERGRRVGMFAGKMELVEAEVELRLGFEDAKAGAVRRLLEEMRF TQSFVNVRNPFYGK*
SEQ ID NO:136 Truncated geranylgeranyl pyrophosphate synthase Pb_GPPS_NTrunc	atggctgctatctttccatccattccatccaacttcaaaccacctcaaatctctcaaactttgaccagacgtagaaga ccaaaccgtactttatgtactgccacctctgaccaatcttacttgtccgcttcttctgccgacatttattctcatttgtta agatctttaccagctactattcatccatctgttaaagccccaatccattctttattgtcctctccaattcctccaaccatc gctccacctttgtgtttagctgctaccgaattggttggtggtaaccctaactctgccattaacgccgcctgtgccatt catttgattcatgctgttactcatactagaaccgctccaccattagctgaattttctcctggtgttttgttgatgactggt gatggtttattagttttggcttacgagatgttggccagatccccagctgttgatgccgatacttctgtccgtgttttgaa ggaagtcgctagaaccgccgccgccgtcgccgctgcttatgaaggtggtagagaagggtgaattagctgccggt gccgctgcttgtggtgtcattttgggtggtggtaacgaagaagaggtcgaaagaggtcgtagagtcggtatgttc gctggtaaaatggaattagttgaagctgaagtcgaattgagattgggtttcgaagatgctaaagccggtgccgtta gaagattgttggaagaaatgcgtttcacccaatcttttgtcaacgttagaaaccctttttatggtaagtaa
SEQ ID NO:137 Geranylgeranyl pyrophosphate synthase (Ai_GPPS) Azadirachta indica	MLFSRGLSRISRIPRNSLIGCRWLVSYRPDTILSGSSHSVGDSTQKVLGC REAYLWSLPALHGIRHQIHQQSSSLIEEELDPFSLVADELSLVANRLRS MVVAEVPKLASAAEYFFKMGVEGKRFRPTVLLLMASALNVQVPQPLS DGVGDALTTELRTRQQCIAEITEMIHVASLLHDDVLDDADTRRGIGSL NFVMGNKLAVLAGDFLLSRACVALASLKNTEVVSLLATVVEHLVTGE TMQMTTTAEQRRSMDYYMQKTYYKTASLISNSCKAIALLAGQTTEVA MLAFDYGKNLGLAFQLIDDVLDFTGTSASLGKGSLSDIRHGIVTAPILF AMEEFPELRKVVDKGFDDPSNVDIALEYLGKSRGIQRTRELAQKHANL ATVALDSLPESNDDDVKKSRRALLDLAQRVITRNK*
SEQ ID NO:138	atgttgttttccagaggtttatctcgtatttccagaatcccacgtaactctttgatcggttgtagatggttagtttcttacc

2018256863 11 Apr 2024

SEQ ID	SEQUENCE
Artificial geranylgeranyl pyrophosphate synthase (Ai_GPPS) nucleotide sequence	gtcctgataccatttttatctggttcctctcactccgttggtgactctactcaaaaggttttaggttgtcgtgaagcttact tgtggtctttaccagccttgcacggtattagacaccaaattcatcaacaatcctcttctttgattgaagaagaattgga tccattctctttagttgctgatgaattgtctttagtcgctaaccgtttgagatccatggtcgtcgctgaagtcccaaaat tagcctccgccgccgagtacttcttcaagatgggtgttgagggtaagagattccgtccaactgtcttattgttgatg gcctccgccttaaacgttcaagtcccacaacctttgtctgacggtgttggtgatgctttgactaccgagttgagaac tagacaacaatgcattgctgagattactgaaatgatccatgttgcttctttgttgcatgacgacgttttggatgatgct gacactagacgtggtatcggttctttgaacttcgttatgggtaacaagttggctgtcttggctggtgatttcttgttgtc cagagcctgtgttgctttagcttccttgaagaatactgaggttgtctctttgttggccaccgttgttgaacacttggtc accggtgaaactatgcaaatgactactactgctgaacaaagacgttccatggattattacatgcaaaagacttact ataagaccgcctctttgatttccaactcttgtaaagccattgccttgttagctggtcaaactaccgaagttgctatgtt ggctttcgattacggtaagaatttgggtttagcttttcaattgatcgatgacgtcttggattttactggtacctctgcttc tttaggtaaaggttccttgtctgatattagacacggtatcgttaccgctccaattttattcgctatggaagaattccca gaattaagaaaggttgttgataagggttttgacgacccttccaacgttgacattgctttggagtatttgggtaagtct agaggtattcaaagaaccagagaattggctcaaaaacatgccaatttggccaccgtcgccttggattctttaccag aatccaacgacgacgatgttaagaagtctcgtagagctttattggacttggctcaaagagttattactagaaacaa gtaa
SEQ ID NO:139 Truncated geranylgeranyl pyrophosphate synthase Es_GPPS_NTrunc	MRRSGSATAAAAATLARHANACCRARSPALGLLPGAAASSSTHRAAL SSNSGHGGDGSGHYDAAMRRRESCASRSRHRWSGQEAAAASATTTT ARRAPGGVAGASGQGSAAGSVRALSSSFLADAVRETATNHCIDRVVN GGLDGSVPVDKDTPTVEVQDFVYDIDFAQRPSGASQSLADGPDPFELV SAELAGLSDGIKSLIGTEHAVLNAAAKYFFELDGGKKIRPTMVILMSQA CNSNSQQVRPDVQPGTELVNPLQLRLAEITEMIHAASLFHDDVIDEADT RRGVPSVNKVFGNKLAILAGDFLLARSSMSLARLRSLESVELMSAAIE HLVKGEVLQMRPTEDGGGAFEYYVRKNYYKTGSLMANSCKASAVLG QHDLEVQEVAFEYGKRVGLAFQLVDDILDFEGNTFTLGKPALNDLRQ GLATAPVLLAAEQQPGLAKLISRKFRGPGDVDEALELVHRSDGIARAK EVAVVQAEKAMSAILTLHDSPAQNALVQLAHKIVNRNH*
SEQ ID NO:140 Truncated geranylgeranyl pyrophosphate synthase Es_GPPS_NTrunc	atgcgtagatccggttccgctaccgccgctgccgctgccaccttagccagacacgccaacgcctgttgtagagc ccgttccccagctttaggtttgttgcctggtgccgccgcttcttcctctactcacagagccgccttgtcttctaattct ggtcatggtggtgatggttccggtcattacgacgctgctatgagaagaagagaatcttgcgcttccagatctcgtc acagatggtccggtcaagaagctgccgccgcctccgccactaccaccaccgctcgtcgtgctccaggtggtgt cgccggtgcttctggtcaaggttctgctgccggttctgttagagccttatcctcttcttttttagccgatgccgttcgtg aaaccgctactaaccactgtatcgaccgtgttgtcaacggtggtttggacggttctgtcccagtcgataaagatac cccaactgtcgaagttcaagactttgtttatgatattgactttgctcaacgtccatccggtgcctctcaatctttagctg acggtccagatccattcgagttagtttccgctgagttggccggtttgtctgatggtattaagtctttgattggtaccga acatgctgtcttgaacgccgccgccaaatatttcttcgaattagatggtggtaaaaagatcagacctactatggttat cttaatgtcccaagcttgtaactctaattcccaacaagttcgtcctgacgttcaaccaggtactgaattagtcaatcct ttgcaattaagattggctgaaatcaccgagatgattcatgctgcttctttattccacgacgatgttattgatgaggctg atactagacgtggtgtcccttctgttaataaagttttcggtaacaaattagccatcttggccggtgacttcttattggct agatcctctatgtccttggcccgtttaagatccttggagtccgtcgaattgatgtccgccgctatcgaacacttggtc aaaggtgaagttttacaaatgcgtccaactgaggacggtggtggtgctttcgagtactacgtcagaaaaaattact acaagactggttctttgatggctaactcctgtaaggcctccgccgttttaggtcaacacgacttagaagtccaaga ggtcgcttttgaatacggtaagagagtcggtttggctttccaattggttgacgatattttagattttgaaggtaatactt tcactttgggtaagccagctttaaacgacttgagacaaggtttagccactgcccctgtcttgttagctgctgaacaa caacctggtttagctaaattgatctccagaaagtttagaggtcctggtgatgtcgatgaagctttggaattggtcca cagatccgacggtattgctagagctaaggaggttgctgttgtccaagccgaaaaagctatgtctgccattttgacc ttgcatgactccccagctcaaaatgctttggttcaattggctcacaaaatcgtcaatcgtaaccattag
SEQ ID NO:141 Geranylgeranyl pyrophosphate synthase (Si_GPPS) Solanum sp.	MIFSKGLAQISRNRFSRCRWLFSLRPIPQLHQSNHIHDPPKVLGCRVIHS WVSNALSGIGQQIHQQSTAVAEEQVDPFSLVADELSLLTNRLRSMVVA EVPKLASAAEYFFKLGVEGKRFRPTVLLLMATALNVQIPRSAPQVDVD SFSGDLRTRQQCIAEITEMIHVASLLHDDVLDDADTRRGIGSLNFVMG NKLAVLAGDFLLSRACVALASLKNTEVVCLLATVVEHLVTGETMQMT TSSDERCSMEYYMQKTYYKTASLISNSCKAIALLAGHSAEVSVLAFDY

SEQ ID	SEQUENCE
	GKNLGLAFQLIDDVLDFTGTSATLGKGSLSDIRHGIVTAPILYAMEEFP QLRTLVDRGFDDPVNVEIALDYLGKSRGIQRTRELARKHASLASAAID SLPESDDEEVQRSRRALVELTHRVITRTK*
SEQ ID NO:142 Artificial geranylgeranyl pyrophosphate synthase (Si_GPPS) nucleotide sequence	atgatcttttccaagggtttagctcaaatctctcgtaatagattctctcgttgcagatggttattctctttgcgtccaattc ctcaattacaccaatccaatcacatccacgacccaccaaaagttttgggttgtcgtgtcattcactcttgggtttctaa tgccttgtctggtatcggtcaacaaatccatcaacaatctactgccgttgccgaggaacaagtcgaccctttttcttt ggttgctgatgagttatccttgttaaccaacagattgagatccatggttgtcgctgaagtccctaagttagcctccgc cgctgagtatttctttaagttaggtgtcgaaggtaaacgtttccgtccaactgtcttgttgttgatggccactgcctta aacgtccaaattcctcgttctgctccacaagttgacgttgactcttttctggtgacttgagaactagacaacaatgta tcgctgaaattactgaaatgattcacgtcgcctctttgttgcatgatgacgtcttagatgatgctgatactagaagag gtattggttccttaaattttgttatgggtaataagttggctgttttggctggtgatttcttgttatccagagcctgcgtcg ccttagcctccttgaagaacaccgaagttgtctgtttattggccaccgttgtcgaacatttggttaccggtgaaacta tgcaaatgactacctcctccgatgaaagatgttccatggaatactacatgcaaaagacctactataagactgcctct ttgatttctaactcttgtaaagccattgccttgttagccggtcactctgctgaagtttctgtcttggccttcgattacggt aagaacttaggtttggcttttcaattgatcgacgatgttttggacttcaccggtacctctgctactttgggtaaaggtt ccttgtccgatatcagacatggtatcgttactgctcctattttgtatgctatggaagaattccctcaattacgtactttg gttgacagaggtttcgatgatccagttaatgttgagatcgctttggattacttgggtaaatcccgtggtattcaaaga actagagaattagccagaaagcatgcctctttagcctctgccgccatcgattccttgcctgaatccgacgatgagg aagttcaaagatctcgtagagctttggtcgaattgacccatagagtcattactcgtactaagtaa
SEQ ID NO:143 Geranylgeranyl pyrophosphate synthase Hb_GPPS Hevea brasiliensis	MQFLRGLSPISRSGLRLFLSRQLYPFPVANSSQLLGDSTQKVFNRRETY SWSLVDSHGFKQQIHHQSSFLSEEPLDPFSLVADELSLVANRLRAMLVS EVPKLASAAEYFFKMGVEGKRLRPTVLLLMATALNVHIHEPMPNGVG DTLGAELRTRQQCIAEITEMIHVASLLHDDVLDDADTRRGIGSLNFVM GNKVAVLAGDFLLSRACVALASLKNTEVVSLLATVVEHLVTGETMQ MTSTSEQRCSMDHYMQKTYYKTASLISDSCKAIALLAGQTTEVAMLA FEYGKSLGLAFQLIDDVLDFTGTSASLGKGSLSDIRHVIRLSLI*
SEQ ID NO:144 Artificial geranylgeranyl pyrophosphate synthase (Hb_GPPS) nucleotide sequence	atgcaatttttgagaggtttgtcccctatttccagatccggtttgcgtttattcttatctcgtcaattatatccattcccag tcgccaactcctcccaattattaggtgactctactcaaaaggtttttaacagacgtgagacttactcttggtctttggt cgactctcacggttttaagcaacaaattcatcaccaatcctctttttttgtctgaagaaccattggatccattctctttgg ttgctgatgaattatccttggtcgctaacagattgcgtgctatgttggtttctgaagtcccaaaattagcctccgccg ctgaatatttttttcaagatgggtgttgaaggtaagagattgcgtccaaccgtcttgttattaatggccactgctttaaa cgttcatatccatgaacctatgcctaacggtgttggtgacactttgggtgccgaattgagaactagacaacaatgc atcgctgaaatcaccgaaatgatccatgttgcttctttattacatgacgacgttttagacgatgccgataccagaag aggtattggttctttgaacttcgttatgggtaacaaggttgctgttttggccggtgacttttttgttgtccagagcttgtgt tgccttagcttctttgaagaataccgaagtcgtttctttattggccaccgtcgtcgaacacttggttactggtgagact atgcaaatgacctccacttctgagcaacgttgttccatggatcattatatgcaaaagacttactataagaccgcttcc ttaatctctgattcctgtaaagccatcgccttgttagctggtcaaactaccgaggtcgccatgttggccttcgaatat ggtaagtctttgggtttagcttttcaattaatcgacgatgttttagacttcaccggtacttctgcttccttgggtaaggg ttctttgtccgacattagacacgttattagattatccttaatttaa
SEQ ID NO:145 Mutant medium-chain fatty acid CoA ligase Ec_FADK_v1	MHPTGPHLGPDVLFRESNMKVTLTFNEQRRAAYRQQGLWGDASLAD YWQQTARAMPDKIAVVDNHGASYTYSALDHAASCLANWMLAKGIES GDRIAFQLPGWCEFTVIYLACLKIGAVSVPLLPSWREAELVWVLNKCQ AKMFFAPTLFKQTRPVDLILPLQNQLPQLQQIVGVDKLAPATSSLSLSQI IADNTSLTTAITTHGDELAAVLFTSGTEGLPKGVMLTHNNILASERAYC ARLNLTWQDVFMMPAPLGHATGFLHGVTAPFLIGARSVLLDIFTPDAC LALLEQQRCTCMLGATPFVYDLLNVLEKQPADLSALRFFLCGGTTIPK KVARECQQRGIKLLSVYGSTESSPHAVVNLDDPLSRFMHTDGYAAAG VEIKVVDDARKTLPPGCEGEEASRGPNVFMGYFDEPELTARALDEEG WYYSGDLCRMDEAGYIKITGRKKDIIVRGGENISSREVEDILLQHPKIH DACVVAMSDERLGERSCAYVVLKAPHHSLSLEEVVAFFSRKRVAKYK YPEHIVVIEKLPRTTSGKIQKFLLRKDIMRRLTQDVCEEIE*
SEQ ID NO:146	atgcatccaactggtccacacttaggtcctgatgtcttatttagagaatctaatatgaaagtcactttgacctttaatg

SEQ ID	SEQUENCE
Artificial medium-chain fatty acid CoA ligase Ec_FADK_v1 nucleotide sequence	aacaaagacgtgccgcttacagacaacaaggtttgtggggtgacgcttctttggctgactactggcaacaaactg ctagagctatgccagacaagatcgccgttgtcgataaccacggtgcttcttatacctactctgctttggatcatgcc gcttcttgtttggctaattggatgttggctaagggtatcgaatctggtgatcgtattgcttttcaattgccaggttggtg tgaatttaccgttatctacttggcttgtttgaagattggtgctgtttctgtcccattgttgccatcttggagagaagccg aattggtttgggttttgaacaaatgccaagctaagatgttctttgctccaaccttgttcaagcaaactagaccagttg acttgattttaccttttacaaaatcaattaccacaattgcaacaaatcgttggtgttgacaagttagctccagccacctc ctctttgtccttgtcccaaattatcgctgataatacttctttaaccaccgctatcactactcacggtgatgagttggctg ctgttttgttcacttccggtactgagggtttgccaaagggtgttatgttgacccacaataacattttggcttccgaaag agcttattgtgctcgtttgaacttgacctggcaagatgttttcatgatgccagctccattgggtcatgctactggtttct tgcacggtgttactgccccattcttgattggtgctagatctgtcttgttggatatctttaccccagacgcttgcttagct ttattggaacaacaaagatgtacctgtatgttaggtgctactccatttgtttacgatttattgaacgttttggaaaaaca accagctgatttgtctgccttgagattctttttgtgtggtggtactactattccaaagaaagttgctagagaatgccaa caaagaggtatcaagttgttgtccgtctatggttccactgaatcttctcctcatgctgttgtcaatttagatgacccatt gtctagattcatgcacaccgatggttacgccgctgctggtgttgagattaaggttgtcgacgatgctagaaagacc ttacctccaggttgtgaaggtgaagaagcctctagaggtccaaatgtctttatgggttacttcgacgagccagaatt gactgctagagctttagatgaggaaggttggtattactctggtgatttgtgtagaatggatgaagctggttacattaa aatcactggtagaaagaaggacattattgttagaggtggtgaaaatatctcctccagagaagttgaagatattttatt gcaacacccaaagattcatgatgcttgtgttgttgctatgtccgatgagagattaggtgaaagatcttgtgcttacgt tgttttgaaggctccacatcactctttgtctttagaagaagtcgttgctttcttctctagaaagagagtcgccaagtac aagtacccagaacacattgttgttatcgaaaaattgcctagaactacttctggtaaaattcaaaaattcttgttgaga aaggatatcatgagacgtttgacccaagatgtctgtgaagaaattgaataa
SEQ ID NO:147 Medium-chain fatty acid CoA ligase Ec_FADK_v2 Escherichia coli	MHPTGPHLGPDVLFRESNMKVTLTFNEQRRAAYRQQGLWGDASLAD YWQQTARAMPDKIAVVDNHGASYTYSALDHAASCLANWMLAKGIES GDRIAFQLPGWCEFTVIYLACLKIGAVSVPLLPSWREAELVWVLNKCQ AKMFFAPTLFKQTRPVDLILPLQNQLPQLQQIVGVDKLAPATSSLSLSQI IADNTSLTTAITTHGDELAAVLFTSGTEGLPKGVMLTHNNILASERAYC ARLNLTWQDVFMMPAPLGHATGFLHGVTAPFLIGARSVLLDIFTPDAC LALLEQQRCTCMLGATPFVYDLLNVLEKQPADLSALRFFLCGGTTIPK KVARECQQRGIKLLSVYGSTESSPHAVVNLDDPLSRFMHTDGYAAAG VEIKVVDDARKTLPPGCEGEEASRGPNVFMGYFDEPELTARALDEEG WYYSGDLCRMDEAGYIKITGRKKDIIVRGGENISSREVEDILLQHPKIH DACVVAMSDERLGERSCAYVVLKAPHHSLSLEEVVAFFSRKRVAKYK YPEHIVVIEKLPRTTSGKIQKFLLRKDIMRRLTQDVCEEIE*
SEQ ID NO:148 Artificial medium-chain fatty acid CoA ligase Ec_FADK_v2 nucleotide sequence	atgcatccaactggtcctcacttaggtccagatgtcttattcagagaatctaacatgaaagtcactttaacttttaacg aacaacgtagagctgcttatagacaacaaggtttgtggggtgatgcttccttggctgactactggcaacaaactgc tagagccatgccagataaaattgccgttgttgacaatcacggtgcttcttacacttattctgccttagatcacgctgc ttcctgtttagctaactggatgttagctaagggtattgaatccggtgatagaattgctttccaattgccaggttggtgc gaatttactgtcatttatttagcttgtttaaagattggtgccgtctccgtccctttgttgccatcctggagagaggccg agttggtttgggttttaaacaagtgtcaagctaaaatgttctttgctcctaccttgttcaagcaaaccagaccagttga cttaattttgccattacaaaaccaattaccacaattgcaacaaatcgtcggtgttgataaattagctccagccacttct tctttgtccttatcccaaattattgctgataacacttctttaactactgctattactactcacggtgatgaattggccgct gttttgttcacttccggtactgaaggtttgcctaaaggtgtcatgttgactcacaacaacattttggcctctgaaaga gcttactgtgcccgtttaaatttgacctggcaagatgtcttcatgatgcctgctccattgggtcacgctaccggtttct tacacggtgtcactgccccattcttgatcggtgctcgttctgttttattggatatctttactccagatgcttgcttggcttt attggaacaacaaagatgtacctgcatgttaggtgctactcctttcgtctatgatttattgaacgtcttagaaaaacaa ccagctgatttatccgctttaagattctttttgtgtggtggtactactatcccaaaaaaggtcgccagagaatgtcaa caaagaggtattaaattattgtccgtttatggttccactgaatcttcccctcatgctgttgtcaatttagacgacccttt gtccagattcatgcacactgatggttacgccgctgctggtgtcgaaatcaaggttgttgatgacgctagaaaaact ttaccacctggttgcgaaggtgaagaggcttccagaggtccaaacgtctttatgggttactttgatgaaccagaatt gactgccagagctttggatgaggaaggttggtattattctggtgatttgtgtagaatggatgaagccggttacatca agatcaccggtagaaagaaagacatcatcgttagaggtggtgaaaacatttcttctagagaagttgaagacatttt gttgcaacacccaaagatccacgacgcttgtgtcgtcgccatgtctgacgaaagattgggtgaacgttcttgtgct tacgtcgtcttgaaagccccacaccactctttgtctttggaagaagtcgttgctttttctctcgtaagcgtgttgcca

2018256863 11 Apr 2024

SEQ ID	SEQUENCE
	agtacaagtacccagagcacatcgttgttattgaaaaattgcctcgtactacttccggtaagattcaaaagttcttatt acgtaaggacatcatgagaagattgactcaagacgtctgcgaagaaattgaataa
SEQ ID NO:149 Truncated acyl activating enzyme (Cs_AAE3_Ctrunc)	MEKSGYGRDGIYRSLRPPLHLPNNNNLSMVSFLFRNSSSYPQKPALIDS ETNQILSFSHFKSTVIKVSHGFLNLGIKKNDWLIYAPNSIHFPVCFLGIIA SGAIATTSNPLYTVSELSKQVKDSNPKLIITVPQLLEKVKGFNLPTILIGP DSEQESSSDKVMTFNDLVNLGGSSGSEFPIVDDFKQSDTAALLYSSGTT GMSKGWLTHKNFIASSLMVTMEQDLVGEMDNVFLCFLPMFHVFGLAI ITYAQLQRGNTVISARFDLEKMLKDVEKYVTHLWWPPVILALSKNSM VKFNLSSIKYIGSGAAPLGKDLMEECSKWPYGIVAQGYGMTETCGIVS MEDIRGGKRNSGSAGMLASGVEAQIVSVDTLKPLPPNQLGEIWVKGPN MMQGYFNNPQATKLTIDKKGWVHTGDLGYFDEDGHLYWDRIKELIK YKGFQVAPAELEGLLVSHPEILDAWIPFPDAEAGEVPVAYWRSPNSSL TENDVKKFIAGQVASFKRLRKVTFINSVPKSASGKIL
SEQ ID NO:150 Truncated acyl activating enzyme (Cs_AAE3_Ctrunc)	atggaaaaatctggttatggtagagacggtatctacagatccttgcgtcctccattacacttgccaaacaataataa cttatctatggtttccttttgttccgtaactcttcctcttacccacaaaaacctgctttgattgactccgaaaccaatca aatcttgtccttttcccacttcaaatctactgtcattaaagtctctcacggtttcttgaacttaggtattaagaagaacg actggttgatctacgctcctaattccatccactttccagtttgtttcttgggtatcattgcttctggtgccattgctacca cttctaacccttttatacactgtttctgagttatctaagcaagttaaagattctaacccaaaattgattatcactgtccca caattattagaaaaggtcaagggtttcaatttaccaaccattttaatcggtccagactccgaacaagagtcttcttcc gataaagttatgacttttaacgacttagttaacttgggtggttcttctggttctgagttcccaatcgtcgatgatttcaa gcaatctgacaccgccgctttattgtattcctctggtactactggtatgtctaagggttggttgactcacaaaaacttt atcgcttcctctttgatggttaccatggaacaagacttggttggtgaaatggataacgtcttcttgtgtttttaccaat gttccatgttttcggtttagctatcattacttacgctcaattacaaagaggtaacactgtcatctctgctcgttttgactt agaaaagatgttgaaagacgttgaaaagtacgttactcacttgtggtggcctcctgttattttagctttgtctaagaat tctatggttaaattcaacttgtcctctatcaagtacattggttctggtgccgctccattaggtaaggacttgatggaag aatgttctaaatggccttacggtatcgtcgctcaaggttacggtatgactgaaacttgtggtatcgtttctatggaag acatcagaggtggtaagcgtaactccggttctgctggtatgttggcttccggtgttgaagcccaaattgtttctgtc gatactttgaaaccttgccacctaaccaattaggtgaaatttgggttaaaggtcctaacatgatgcaaggttacttc aataaccctcaagctactaagttaactattgataagaagggttgggttcatactggtgatttgggttacttcgatgaa gatggtcatttgtactgggatagaatcaaagaattaattaagtataaaggtttccaagttgccccagctgaattgga aggtttgttggtttctcatcctgaaattttagatgcttggattcctttcccagacgctgaagccggtgaagttccagtt gcttactggagatcccctaactcttccttgactgaaaacgacgtcaagaagttcatcgctggtcaagttgcttccttt aagagattaagaaaagtcaccttcatcaactccgttccaaagtctgcttccggtaagattttg
SEQ ID NO:151 Truncated cannabidiolic acid synthase Cs_CBDASt28	MSNPRENFLKCFSQYIPNNATNLKLVYTQNNPLYMSVLNSTIHNLRFTS DTTPKPLVIVTPSHVSHIQGTILCSKKVGLQIRTRSGGHDSEGMSYISQV PFVIVDLRNMRSIKIDVHSQTAWVEAGATLGEVYYWVNEKNENLSLA AGYCPTVCAGGHFGGGGYGPLMRNYGLAADNIIDAHLVNVHGKVLD RKSMGEDLFWALRGGGAESFGIIVAWKIRLVAVPKSTMFSVKKIMEIH ELVKLVNKWQNIAYKYDKDLLLMTHFITRNITDNQGKNKTAIHTYFSS VFLGGVDSLVDLMNKSFPELGIKKTDCRQLSWIDTIIFYSGVVNYDTD NFNKEILLDRSAGQNGAFKIKLDYVKKPIPESVFVQILEKLYEEDIGAG MYALYPYGGIMDEISESAIPFPHRAGILYELWYICSWEKQEDNEKHLN WIRNIYNFMTPYVSKNPRLAYLNYRDLDIGINDPKNPNNYTQARIWGE KYFGKNFDRLVKVKTLVDPNNFFRNEQSIPPLPRHRH*
SEQ ID NO:152 Truncated cannabidiolic acid synthase Cs_CBDASt28	atgtctaatccaagagagaatttcttaaagtgtttttctcaatacatcccaaacaatgctactaacttaaagttggttta cactcaaaataacccattgtacatgtctgtcttgaactctaccattcacaatttgcgttttacttctgacaccaccccta agccattagttattgttaccccatcccacgtctctcacatccaaggtactattttgtgttctaaaaaggttggtttgcaa attagaactagatctggtggtcacgactccgagggtatgtcttacatctctcaagttccattcgttattgtcgacttgc gtaacatgcgttccatcaaaatcgatgttcactcccaaactgcttgggtcgaagccggtgccactttaggtgaggt ttattactgggtcaatgagaagaatgagaatttgtccttggctgctggttattgtccaaccgtctgtgctggtggtcat tttggtggtggtggttacggtccattaatgagaaactatggtttggctgccgataacattatcgacgctcacttggtt aatgtccacggtaaggtcttagatagaaaatccatgggtgaggacttgttctgggctttgagaggtggtggtgctg agtcctttggtatcatcgttgcttggaaaattcgtttagttgctgtcccaaaatctactatgttttctgttaagaagatca

SEQ ID	SEQUENCE
	tggaaattcacgagttggttaagttggttaataagtggcaaaatattgcctacaagtatgacaaagacttgttattga tgactcacttcatcactagaaacatcaccgataaccaaggtaaaaataaaactgctatccatacctacttctcctcc gttttcttgggtggtgtcgactccttagttgatttgatgaacaaatcttttcctgaattaggtatcaagaagactgattgt cgtcaattgtcctggattgataccattatcttttactctggtgtcgtcaattacgacaccgataatttcaataaggaaat tttattggacagatctgccggtcaaaacggtgctttcaagatcaagttggactacgttaaaaaaccaatcccagaat ccgtctttgtccaaattttggagaagttatacgaggaagacatcggtgctggtatgtatgccttatatccatacggtg gtattatggatgaaatttccgaatctgctatcccatttccacatcgtgctggtattttgtatgaattatggtacatttgttc ctgggaaaagcaagaagataacgagaagcacttgaattggatcagaaatatctacaatttcatgactccttacgttt ctaagaatcctcgtttggcttacttgaactacagagatttggacatcggtattaatgacccaaagaacccaaataac tatactcaagctagaatttggggtgaaaagtacttcggtaaaaactttgacagattggttaaggttaagactttagtt gatccaaataacttcttcagaaatgaacaatccatcccaccattgcctagacacagacactaa
SEQ ID NO:153 Truncated tetrahydro cannabinolic acid synthase Cs_THCASt28	MSNPRENFLKCFSKHIPNNVANPKLVYTQHDQLYMSILNSTIQNLRFIS DTTPKPLVIVTPSNNSHIQATILCSKKVGLQIRTRSGGHDAEGMSYISQV PFVVVDLRNMHSIKIDVHSQTAWVEAGATLGEVYYWINEKNENLSFP GGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLD RKSMGEDLFWAIRGGGGENFGIIAAWKIKLVAVPSKSTIFSVKKNMEIH GLVKLFNKWQNIAYKYDKDLVLMTHFITKNITDNHGKNKTTVHGYFS SIFHGGVDSLVDLMNKSFPELGIKKTDCKEFSWIDTTIFYSGVVNFNTA NFKKEILLDRSAGKKTAFSIKLDYVKKPIPETAMVKILEKLYEEDVGAG MYVLYPYGGIMEEISESAIPFPHRAGIMYELWYTASWEKQEDNEKHIN WVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNHASPNNYTQARIWG EKYFGKNFNRLVKVKTKVDPNNFFRNEQSIPPLPPHHH*
SEQ ID NO:154 Truncated tetrahydrocannabinolic acid synthase Cs_THCASt28	atgtctaaccctcgtgagaacttcttgaaatgtttctccaaacatatcccaaacaatgtcgctaaccctaagttagttt acactcaacatgatcaattatatatgtctatcttgaactctaccatccaaaacttgagattcatctccgataccacccc aaaaccattggttattgttaccccatccaacaattctcatattcaagctaccattttgtgctccaaaaaggtcggtttg caaatccgtactagatctggtggtcacgatgctgaaggtatgtcttacatttcccaagtcccattcgttgttgtcgatt taagaaatatgcactctatcaaaatcgacgttcactctcaaactgcttgggttgaagccggtgccactttaggtgag gtttactactggattaacgaaaagaatgaaaacttatcctttccaggtggttactgtccaactgttggtgttggtggtc acttctctggtggtggttatggtgccttgatgagaaactacggtttagctgctgataatattatcgacgctcacttggt taatgtcgacggtaaggttttggacagaaaatccatgggtgaagatttattctgggccattagaggtggtggtggt gaaaacttcggtatcattgctgcttggaaaattaaattggtcgctgtcccatccaagtctactattttctccgtcaaga aaaacatggaaattcatggtttggttaaattattcaacaagtggcaaaacattgcttacaaatacgacaaagactta gttttgatgacccacttcattactaaaaacattaccgacaaccatggtaaaaataaaactactgttcacggttacttct cttccatttttcatggtggtgtcgactccttggtcgatttaatgaacaaatctttccctgagttgggtatcaagaagac cgactgtaaagaattctcttggatcgacactactattttctactctggtgtcgttaacttcaacaccgctaatttcaaga aggaaattttattagatagatccgctggtaaaaagaccgctttctctatcaaattagactacgttaaaaaaccaatcc cagaaaccgctatggtcaaaatcttggaaaaattatatgaagaagacgttggtgccggtatgtacgtcttatatcca tatggtggtattatggaagagatctctgaatccgctatccctttccacacagagccggtattatgtacgaattatgg tacactgcttcctgggagaaacaagaagataatgaaaagcacattaactgggttagatctgtttacaacttcactac tccatacgtctctcaaaacccaagattagcctacttaaactaccgtgatttggatttaggtaaaactaatcacgcttc cccaaacaactacacccaagctagaatttggggtgagaagtactttggtaagaacttcaaccgtttagtcaaggtc aagactaaagttgatccaaacaatttttttcagaaacgaacaatctatcccacctttaccaccacaccaccattag
SEQ ID NO:155 GenBank AB057805.1 Tetrahydro cannabinolic acid synthase (THCAS, Cs_THCAS_full) *Cannabis sativa*	MNCSAFSFWFVCKIIFFFLSFHIQISIANPRENFLKCFSKHIPNNVANPKL VYTQHDQLYMSILNSTIQNLRFISDTTPKPLVIVTPSNNSHIQATILCSKK VGLQIRTRSGGHDAEGMSYISQVPFVVVDLRNMHSIKIDVHSQTAWVE AGATLGEVYYWINEKNENLSFPGGYCPTVGVGGHFSGGGYGALMRN YGLAADNIIDAHLVNVDGKVLDRKSMGEDLFWAIRGGGGENFGIIAA WKIKLVAVPSKSTIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLVLM THFITKNITDNHGKNKTTVHGYFSSIFHGGVDSLVDLMNKSFPELGIKK TDCKEFSWIDTTIFYSGVVNFNTANFKKEILLDRSAGKKTAFSIKLDYV KKPIPETAMVKILEKLYEEDVGAGMYVLYPYGGIMEEISESAIPFPHRA GIMYELWYTASWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLN YRDLDLGKTNHASPNNYTQARIWGEKYFGKNFNRLVKVKTKVDPNN

2018256863 11 Apr 2024

SEQ ID	SEQUENCE
	FFRNEQSIPPLPPHHH*
SEQ ID NO:156 Artificial Tetrahydrocannabinolic acid synthase Cs_THCAS_full nucleotide sequence	atgaattgttctgctttctctttctggttcgtttgtaagatcatctttttcttcttatctttccatattcaaatctctatcgctaa ccctcgtgagaacttcttgaaatgtttctccaaacatatcccaaacaatgtcgctaaccctaagttagtttacactca acatgatcaattatatatgtctatcttgaactctaccatccaaaacttgagattcatctccgataccaccccaaaacc attggttattgttaccccatccaacaattctcatattcaagctaccattttgtgctccaaaaaggtcggtttgcaaatcc gtactagatctggtggtcacgatgctgaaggtatgtcttacatttcccaagtcccattcgttgttgtcgatttaagaaa tatgcactctatcaaaatcgacgttcactctcaaactgcttgggttgaagccggtgccactttaggtgaggtttacta ctggattaacgaaaagaatgaaaacttatcctttccaggtggttactgtccaactgttggtgttggtggtcacttctct ggtggtggttatggtgccttgatgagaaactacggtttagctgctgataatattatcgacgctcacttggttaatgtc gacggtaaggttttggacagaaaatccatgggtgaagatttattctgggccattagaggtggtggtggtgaaaact tcggtatcattgctgcttggaaaattaaattggtcgctgtcccatccaagtctactattttctccgtcaagaaaaacat ggaaattcatggtttggttaaattattcaacaagtggcaaaacattgcttacaaatacgacaaagacttagttttgat gacccacttcattactaaaaacattaccgacaaccatggtaaaaataaaactactgttcacggttacttctcttccatt tttcatggtggtgtcgactccttggtcgatttaatgaacaaatctttccctgagttgggtatcaagaagaccgactgt aaagaattctcttggatcgacactactattttctactctggtgtcgttaacttcaacaccgctaatttcaagaaggaaa ttttattagatagatccgctggtaaaaagaccgctttctctatcaaattagactacgttaaaaaaccaatcccagaaa ccgctatggtcaaaatcttggaaaaattatatgaagaagacgttggtgccggtatgtacgtcttatatccatatggtg gtattatggaagagatctctgaatccgctatccctttccacacagagccggtattatgtacgaattatggtacactg cttcctgggagaaacaagaagataatgaaaagcacattaactgggttagatctgtttacaacttcactactccatac gtctctcaaaacccaagattagcctacttaaactaccgtgatttggatttaggtaaaactaatcacgcttccccaaac aactacacccaagctagaatttggggtgagaagtactttggtaagaacttcaaccgtttagtcaaggtcaagacta aagttgatccaaacaatttttttcagaaacgaacaatctatcccacctttaccaccacaccaccattag
SEQ ID NO:157 Artificial Erg10p: acetoacetyl CoA thiolase nucleotide sequence	atgtcccaaaatgtttacattgtttctactgctagaactcctatcggttccttccaaggttccttatcttccaaaactgcc gtcgaattgggtgccgttgccttgaaaggtgctttagctaaagttccagagttagacgcttccaaagatttcgatga aattatcttcggtaacgttttatccgctaacttgggtcaagctccagccagacaagttgccttggctgccggtttgtc taatcacatcgttgcttctactgtcaacaaagtttgtgcctctgctatgaaagctatcattttaggtgcccaatctatta aatgtggtaatgctgacgttgttgtcgctggtggttgtgagtccatgaccaacgccccttactacatgccagccgc cagagccggtgccaaattcggtcaaactgttttggttgacggtgttgaaagagatggtttgaacgatgcctatgac ggtttggctatgggtgttcacgctgaaaagtgtgctagagactgggacattaccagagaacaacaagataatttc gctattgaatcttaccaaaagtcccaaaaatctcaaaaggaaggtaagtttgacaatgaaatcgttccagttactat caagggttttcgtggtaagcctgatactcaagtcaccaaggatgaagaaccagcccgtttacacgtcgaaaagtt gagatctgccagaaccgttttccaaaaagaaaacggtaccgttactgctgccaatgcttctccaatcaacgatggt gccgctgctgttattttagtctctgagaaggttttgaaggagaaaaatttgaagcctttagccatcattaagggttgg ggtgaagctgctcaccaaccagctgatttcacttgggccccttctttagctgtcccaaaggctttaaaacacgctg gtattgaagatatcaactctgttgactacttcgaattcaatgaagctttctctgtcgtcggtttggtcaataccaaaatc ttgaagttggatccttctaaggttaacgtttacggtggtgctgtcgccttaggtcacccttaggttgttctggtgcta gagttgttgtcaccttgttgtccattttacaacaagaaggtggtaagatcggtgttgctgctatctgtaacggtggtg gtggtgcttcttccattgtcatcgaaaagatctag
SEQ ID NO:158 Artificial mevalonate pyrophosphate decarboxylase (Sc_ERG19) nucleotide sequence	atgactgtctacactgcctccgttactgcccctgtcaacattgccaccttgaagtattggggtaaaagagatactaa attgaacttaccaactaactcctccatttctgtcactttgtctcaagatgatttgagaaccttgacttccgctgccacc gcccctgaatttgagagagatactttgtggttaaatggtgaacctcattctattgacaacgaaagaacccaaaact gtttacgtgacttgagacaattgcgtaaggaaatggaatctaaagacgcttctttacctaccttgtctcaatggaaat tgcatatcgtttctgaaaataacttccctactgctgccggtttggcttcctccgctgctggttttgctgctttagtttctg ccatcgccaaattatatcaattgccacaatccacttccgaaatctctagaatcgctagaaaaggttccggttctgctt gtagatccttgttcggtggttacgttgcttgggaaatgggtaaagctgaagacggtcatgattctatggccgttcaa attgccgactcctccgattggcctcaaatgaaagcttgtgtcttggttgtctccgatatcaaaaaggatgtctcttcta ctcaaggtatgcaattaactgttgccacttccgaattgttcaaagagcgtatcgaacacgttgttccaaagagatttg aagttatgagaaaagctatcgtcgaaaaggacttcgctacctttgccaaggagactatgatggattctaactccttc cacgctacttgtttggattcctttccacctattttctacatgaatgacacctccaaacgtattatctcttggtgtcacacc attaaccaattttatggtgaaactatcgtcgcttacactttcgatgccggtccaaacgctgtcttgtactatttggctga aaacgaatccaagttatttgcttttatctataagttgttcggttccgtccctggttgggacaagaaattcaccactgaa caattggaagctttcaaccaccaattcgaatcttccaatttcactgctagagaattagatttggaattacaaaaggat gtcgctagagtcatcttaactcaagttggttccggtccacaagaaactaacgaatctttgattgatgctaaaactggt

SEQ ID	SEQUENCE
	ttgcctaaagaataa
SEQ ID NO:159 Artificial isopentenyl pyrophosphate isomerase Sc_IDI1 nucleotide sequence	atgaccgctgacaacaactccatgccacatggtgctgtctcctcctacgctaaattagtccaaaaccaaacccctg aagacattttagaagagttccctgaaatcattccattgcaacaaagaccaaacactagatcctccgagacttctaa cgatgaatctggtgaaacttgtttttctggtcatgatgaagaacaaatcaagttgatgaacgagaattgtattgttttg gactgggatgacaacgctatcggtgctggtaccaaaaaggtctgtcacttgatggaaaacatcgaaaagggtttg ttgcatagagccttttccgtcttcatcttcaacgaacaaggtgagttattattgcaacaaagagccactgaaaaaatc acctttccagatttatggaccaacacctgttgctcccatccattgtgtattgatgatgaattgggtttgaaaggtaagt tggacgacaagattaaaggtgccatcaccgccgctgttcgtaagttagaccatgaattgggtatccctgaagacg aaactaagactagaggtaaattccatttcttgaatcgtattcactacatggctccttccaatgaaccatggggtgaa cacgaaatcgactacattttgttttacaaaattaatgctaaagaaaatttaaccgttaacccaaacgtcaacgaggtt agagatttcaagtgggtctctccaaacgatttgaagactatgttcgctgacccatcctacaagttcactccatggttt aagatcatctgtgaaaactatttgtttaactggtgggagcaattggacgacttatctgaagttgaaaatgatcgtcaa attcaccgtatgttgtaa
SEQ ID NO:160 Artificial phosphomevalonate kinase Sc_ERG8 nucleotide sequence	atgtccgagttaagagccttctccgctcctggtaaagccttattagctggtggttacttagtcttggatactaaatatg aagccttcgtcgtcggtttatctgccagaatgcatgccgtcgcccatccatacggttccttgcaaggttctgacaag tttgaggtccgtgtcaagtctaaacaattcaaagatggtgaatggttgtatcatatttctccaaaatccggtttcattcc agtttctatcggtggttctaagaacccattcatcgaaaaagtcatcgctaacgttttctcttacttcaagcctaatatg gatgattattgcaatagaaatttattcgttattgatatcttctccgatgacgcctatcattcccaagaagactctgttac cgagcatagaggtaacagaagattatctttccactctcacagaattgaagaagttccaaaaactggtttaggttctt ctgctggtttagtcaccgttttaaccactgccttggcttctttctttgtttccgacttagaaaataacgtcgacaagtat cgtgaagtcatccacaacttggcccaagttgctcattgtcaagctcaaggtaagattggttccggtttcgatgttgct gccgccgcctacggttccatcagatatagaagattccctccagctttgatttctaacttaccagatattggttctgcta cttatggttccaagttggctcacttggttgacgaagaagattggaacattaccatcaagtccaatcacttgccatctg gtttaactttgtggatgggtgatatcaagaacggttctgaaactgtcaaattggtccaaaaggtcaaaaattggtac gattcccatatgccagagtctttgaagatctatactgaattggaccacgctaactctcgtttcatggatggtttgtcta agttggacagattgcatgaaactcacgacgactactctgaccaaattttcgagtccttggaaagaaacgactgca cttgtcaaaagtatccagaaatcaccgaggttagagatgccgttgctactattagaagatccttcagaaagattacc aaggaatctggtgctgatattgagcctccagttcaaacttctttgttggatgattgccaaactttaaaaggtgttttaa cttgtttaattcctggtgctggtggttacgacgccatcgccgttatcaccaaacaagacgtcgacttaagagccca aactgccaacgacaaaagattctccaaggttcaatggttggacgtcactcaagctgattggggtgttagaaaaga aaaggacccagagacttacttggataaatag
SEQ ID NO:161 Mutant farnesyl pyrophosphate synthase (Erg20mut, F96W, N127W)	atggcttctgagaaggagattcgtcgtgagagattcttgaatgtttttcctaaattagtcgaggaattgaacgcttctt tgttggcttatggtatgcctaaggaagcttgtgattggtatgctcactccttgaattataatactccaggtggtaaatt gaaccgtggtttgtctgttgttgacacttacgctattttatctaacaagaccgtcgagcaattgggtcaagaagagta tgaaaaggtcgctattttaggttggtgtattgaattgttgcaagcttactggttggttgccgatgacatgatggacaa gtctattactcgtcgtggtcaaccttgctggtataaggtcccagaggttggtgaaattgctatctgggacgctttcat gttggaagctgctatctataaattgttgaaatcccacttcagaaacgagaaatactacattgacatcaccgagttgtt ccacgaagtcactttccaaactgagttaggtcaattaatggacttgatcaccgctccagaagacaaagttgacttgt ccaagttttccttgaaaaagcactctttcatcgttactttcaagactgcttattactctttctacttaccagttgccttggc tatgtacgtcgccggtatcactgacgaaaaggacttgaagcaagctcgtgacgttttgattccattaggtgaatattt ccaaatccaagatgactacttagactgttttggtacccctgaacaaatcggtaagatcggtactgatattcaagata acaagtgctcttgggttatcaacaaggctttagagttagcctccgccgaacaacgtaaaactttagatgaaaacta cggtaaaaaagactctgttgctgaggccaagtgtaagaagatttttaacgatttaaaaatcgaacaattgtatcacg aatatgaagagtccattgctaaggatttgaaggctaaaatttctcaagttgacgaatcccgtggtttcaaagctgac gttttgactgctttttaaacaaggtttacaagcgttccaaataa
SEQ ID NO:162 Artificial tetraketide synthase (TKS) nucleotide sequence	atgaaccatttaagagctgagggtccagcttccgtcttggctatcggtactgctaatccagagaacattttattacaa gatgagtttccagattactatttccgtgttactaagtccgagcatatgacccaattgaaagaaaagttccgtaaaatc tgtgataaatctatgattagaaaaagaaactgctttttaaacgaagaacacttgaagcaaaacccaagattagttga acacgagatgcaaaccttggacgctagacaagatatgttggttgtcgaggttcctaaattgggtaaagacgcctg tgctaaagctatcaaagagtggggtcaacctaagtccaagatcactcacttaatcttcacttccgcttccaccactg acatgcctggtgctgattaccactgtgccaagttgttgggtttgtctccttctgtcaagagagttatgatgtaccaatt aggttgttacggtggtggtactgtcttaagaattgctaaggacatcgctgaaaacaacaaaggtgctagagtttta gccgtttgttgtgacatcatggcttgtttatttcgtggtccatctgaatctgacttggagttgttggttggtcaagctatt

SEQ ID	SEQUENCE
	tttggtgatggtgccgctgccgtcatcgttggtgctgagccagatgaatccgttggtgaaagaccaattttcgaatt agtctctactggtcaaactattttgccaaactccgagggtactatcggtggtcatattcgtgaagccggtttaatcttt gatttgcacaaagacgttccaatgttgatctctaacaacatcgaaaagtgtttaattgaggcttttactccaattggta tctctgactggaactctatcttctggatcactcatccaggtggtaaggctatcttggacaaggttgaagaaaaatta catttaaagtccgataaattcgtcgattctcgtcatgttttgtctgaacacggtaacatgtcttcctccactgtcttgttt gttatggatgaattacgtaagagatctttggaggagggtaagtctactactggtgatggtttcgaatggggtgttttg ttcggtttcggtcctggtttgactgttgaacgtgttgttgttagatctgttccaattaagtactag
SEQ ID NO:163 Artificial olivetolic acid cyclase (OAC) nucleotide sequence	atggccgtcaaacacttgatcgtcttaaaattcaaggatgaaattactgaagctcaaaaagaagagttcttcaaaa cctatgtcaatttagtcaacattattcctgctatgaaggacgtttactggggtaaggatgtcacccaaaagaacaag gaagaaggttacactcacattgttgaagtcactttcgaatctgttgaaactatccaagattatattatccacccagct catgtcggttttggtgatgtttacagatctttttgggaaaaattgttgatctttgactatactccaagaaaataa
SEQ ID NO:164 Artificial acyl-activating enzyme Cs_AAE1_v1 nucleotide sequence	atgggtaagaattacaagtccttagactctgttgttgcttctgactttattgctttaggtattacttccgaagttgctgaa accttacacggtagattggctgaaattgtttgcaactacggtgctgctacccctcaaacttggattaacattgctaat catattttgtctccagatttgccattttctttacaccaaatgttgttctacggttgttacaaggatttcggtcctgctcctc cagcttggattcctgatccagaaaaagtcaaatctactaacttgggtgctttgttggaaaagagaggtaaggagttt ttgggtgttaagtacaaggacccaatttcttctttctctcacttccaagaattctctgttagaaaccctgaagtttactg gagaactgttttgatggatgagatgaagatttctttttctaaggacccagagtgtatcttaagaagagacgacattaa caatccaggtggttctgagtggttaccaggtggttacttgaactctgccaaaaattgcttgaacgttaactctaaca agaaattgaatgacactatgattgtctggagagatgagggtaacgatgatttgcctttgaataaattgactttggatc aattgagaaaaagagtctggttggttggttacgctttggaagaaatgggtttagaaaaaggttgtgctatcgccatc gatatgcctatgcacgttgatgctgttgttatttatttggctattgttttagctggttatgttgttgtttccatcgccgactc cttctctgctccagaaatctccaccagattgagattgtctaaagccaaagccattttcacccaagaccacatcatta gaggtaagaagcgtattccattgtattctcgtgttgttgaagctaaatctcctatggctatcgtcatcccatgctctgg ttctaacatcggtgctgaattaagagacggtgatatttcttgggactacttttagaaagagctaaagaattcaaaaa ctgcgagtttactgctagagaacaacctgtcgacgcttatactaatattttattctcttctggtactactggtgaaccta aggctattccatggacccaagctactcctttgaaagccgctgctgatggttggtcccatttagacatcagaaaagg tgatgtcatcgtctggccaactaacttaggttggatgatgggtccatggttagtctacgcttctttgttgaatggtgcc tctatcgccttatataataggttcccctttagtctctggttttgctaaattcgttcaagatgctaaggttaccatgttaggt gttgtcccttctatcgttagatcttggaaatctactaactgtgtttctggttacgactggtccactattcgttgtttctcttc ttctggtgaagcttccaatgtcgatgagtacttatggttaatgggtcgtgctaactacaagccagtcatcgaaatgt gcggtggtactgaaattggtggtgctttttccgctggttctttttacaagcccaatccttgtcttccttctcctctcaat gtatgggttgtacttatatatcttagataagaatggttaccctatgcctaaaaacaagccaggtattggtgaattagc tttgggtcctgttatgtttggtgcttctaaaaccttgttaaatggtaatcatcacgacgtttacttcaaaggtatgcctac tttgaacggtgaggttttgagacgtcatggtgatattttcgaattaacttccaacggttattatcacgctcacggtaga gctgatgatactatgaacattggtggtattaagatctcttccatcgaaattgagagagtttgtaacgaggttgacgat cgtgttttcgaaactactgctattggtgtccctcctttaggtggtggtccagaacaattggttatctttttcgtcttgaag gactccaacgacaccactatcgacttaaaccaattaagattgtctttcaacttgggtttgcaaaagaagttgaatcc attatttaaggttactcgtgtcgttccattgtcctccttgccaagaactgctaccaacaagattatgcgtagagtcttg agacaacaattctctcactttgagtaa
SEQ ID NO:165 Artificial acyl-activating enzyme Cs_AAE1_v2 nucleotide sequence	atgggtaagaactacaaatccttagattccgtcgtcgcttctgatttcatcgctttgggtattacttctgaagttgctga aaccttgcatggtagattggctgaaattgtctgtaactacggtgctgctaccccacaaacttggatcaacattgcta accacatcttatcccctgacttgccattctccttacaccaaatgttgttctacggttgttataaagatttcggtccagct cctcctgcttggattcctgacccagagaaggttaagtctactaatttaggtgctttgttagagaagagaggtaagga atttttaggtgttaagtataaagatccaatttcttccttctctcacttccaagaattttctgttagaaacccagaagtttac tggagaactgttttgatggatgaaatgaagatctctttttccaaggacccagagtgtattttgagacgtgatgacatc aacaatccaggtggttctgagtggttaccaggtggttacttgaactctgccaagaattgtttgaacgttaactctaac aaaaagttgaacgataccatgattgtttggagagacgaaggtaacgatgatttgccattgaataagttaaccttgg atcaattgagaaaaagagtctggttagtcggttacgctttggaagagatgggtttggaaaagggttgtgctatcgc catcgatatgccaatgcatgttgatgctgttgttatctatttggccattgttttggctggttacgttgttgtttccatcgct gactccttctctgctccagaaatttctactagattaagattgtctaaagccaaagccattttcactcaagaccatattat tagaggtaagaaaagaattccattgtattccagagttgttgaagctaaatccccaatggccatcgtcatcccatgct ctggttctaatattggtgccgaattgagagacggtgatatctcttgggactacttttggagcgtgctaaagaattta aaaactgcgaattcaccgccagagaacaaccagttgacgcctacactaacattttgttttcttctggtactactggt

SEQ ID	SEQUENCE
	gaacctaaggctattccatggactcaagctactccattgaaagccgccgccgatggttggtcccacttagatatta gaaagggtgatgtcatcgtctggcctactaacttgggttggatgatgggtccttggttggtttacgcttccttattga acggtgcctctatcgctttatataatggttccccttagtttctggttttgctaaattcgttcaagatgctaaggttactat gttgggtgtcgtcccatccattgtccgttcctggaagtctaccaattgtgtttctggttatgattggtctactattcgttg ttttcttcctctggtgaagcttctaatgtcgatgaatatttgtggttaatgggtagagctaactacaagccagttattg aaatgtgtggtggtactgaaattggtggtgctttctctgctggttccttttgcaagctcaatccttgtcttctttctcctc ccaatgtatgggttgcactttatacatcttggacaagaatggttaccctatgccaaagaataaaccaggtattggtg aattggctttgggtccagtcatgttcggtgcttctaagactttgttgaacggtaaccatcatgacgtctacttcaagg gtatgcctaccttgaacggtgaagttttaagacgtcacggtgacattttcgaattgacttccaacggttattatcatgc tcacggtagagctgacgacactatgaacatcggtggtattaagatctcttctatcgaaattgaaagagtttgcaacg aggttgatgatcgtgtcttcgaaaccactgctattggtgtccctcctttaggtggtggtcctgagcaattggttattttc tttgtcttaaaggattctaacgacaccactattgacttaaatcaattgagattgtccttcaatttgggtttgcaaaagaa gttgaacccattattcaaggttactcgtgtcgttcctttgtcctctttgccaagaaccgctaccaataaaattatgaga cgtgttttgcgtcaacaattctctcactttgaataa
SEQ ID NO:166 Artificial acyl-activating enzyme Cs_AAE3 nucleotide sequence	atggaaaaatctggttatggtagagacggtatctacagatccttgcgtcctccattacacttgccaaacaataataa cttatctatggtttccttttgttccgtaactcttcctcttacccacaaaaacctgctttgattgactccgaaaccaatca aatcttgtcctttcccacttcaaatctactgtcattaaagtctctcacggtttcttgaacttaggtattaagaagaacg actggttgatctacgctcctaattccatccactttccagtttgtttcttgggtatcattgcttctggtgccattgctacca cttctaacccttttatacactgtttctgagttatctaagcaagttaaagattctaacccaaaattgattatcactgtccca caattattagaaaaggtcaagggtttcaatttaccaaccattttaatcggtccagactccgaacaagagtcttcttcc gataaagttatgacttttaacgacttagttaacttgggtggttcttctggttctgagttcccaatcgtcgatgatttcaa gcaatctgacaccgccgctttattgtattcctctggtactactggtatgtctaagggttggttgactcacaaaaacttt atcgcttcctctttgatggttaccatggaacaagacttggttggtgaaatggataacgtcttcttgtgtttttttaccaat gttccatgttttcggtttagctatcattacttacgctcaattacaaagaggtaacactgtcatctctgctcgttttgactt agaaaagatgttgaaagacgttgaaaagtacgttactcacttgtggtggcctcctgttattttagctttgtctaagaat tctatggttaaattcaacttgtcctctatcaagtacattggttctggtgccgctccattaggtaaggacttgatggaag aatgttctaaatggccttacggtatcgtcgctcaaggttacggtatgactgaaacttgtggtatcgtttctatggaag acatcagaggtggtaagcgtaactccggttctgctggtatgttggcttccggtgttgaagcccaaattgtttctgtc gatactttgaaacctttgccacctaaccaattaggtgaaatttgggttaaaggtcctaacatgatgcaaggttacttc aataaccctcaagctactaagttaactattgataagaagggttgggttcatactggtgatttgggttacttcgatgaa gatggtcatttgtactgggatagaatcaaagaattaattaagtataaaggtttccaagttgccccagctgaattgga aggtttgttggtttctcatcctgaaattttagatgcttggattcctttcccagacgctgaagccggtgaagttccagtt gcttactggagatcccctaactcttccttgactgaaaacgacgtcaagaagttcatcgctggtcaagttgcttccttt aagagattaagaaaagtcaccttcatcaactccgttccaaagtctgcttccggtaagattttgagaagagaattaat ccaaaaggttcgttccaacatgtag
SEQ ID NO:167 Artificial cannabidiolic acid synthase (CBDAS) nucleotide sequence	atgaaatgttccacctttttctttctggtttgtttgtaagatcatcttcttcttcttctccttcaacatccaaacttccatcgct aatccaagagagaatttcttaaagtgtttttctcaatacatcccaaacaatgctactaacttaaagttggtttacactca aaataacccattgtacatgtctgtcttgaactctaccattcacaatttgcgttttacttctgacaccacccctaagccat tagttattgttaccccatcccacgtctctcacatccaaggtactattttgtgttctaaaaaggttggtttgcaaattaga actagatctggtggtcacgactccgagggtatgtcttacatctctcaagttccattcgttattgtcgacttgcgtaaca tgcgttccatcaaaatcgatgttcactcccaaactgcttgggtcgaagccggtgccacttaggtgaggtttattact gggtcaatgagaagaatgagaatttgtccttggctgctggttattgtccaaccgtctgtgctggtggtcattttggtg gtggtggttacggtccattaatgagaaactatggtttggctgccgataacattatcgacgctcacttggttaatgtcc acggtaaggtcttagatagaaaatccatgggtgaggacttgttctgggctttgagaggtggtggtgctgagtccttt ggtatcatcgttgcttggaaaattcgtttagttgctgtcccaaaatctactatgttttctgttaagaagatcatggaaatt cacgagttggttaagttggttaataagtggcaaaatattgcctacaagtatgacaaagacttgttattgatgactcac ttcatcactagaaacatcaccgataaccaaggtaaaaataaaactgctatccatacctacttctcctccgttttcttgg gtggtgtcgactccttagttgatttgatgaacaaatcttttcctgaattaggtatcaagaagactgattgtcgtcaattg tcctggattgataccattatctttactctggtgtcgtcaattacgacaccgataatttcaataaggaaatttttattggac agatctgccggtcaaaacggtgctttcaagatcaagttggactacgttaaaaaaccaatcccagaatccgtctttgt ccaaattttggagaagttatacgaggaagacatcggtgctggtatgtatgccttatatccatacggtggtattatgg atgaaatttccgaatctgctatcccatttccacatcgtgctggtattttgtatgaattatggtacatttgttcctgggaaa agcaagaagataacgagaagcacttgaattggatcagaaatatctacaatttcatgactccttacgtttctaagaat

SEQ ID	SEQUENCE
	cctcgtttggcttacttgaactacagagatttggacatcggtattaatgacccaaagaacccaaataactatactca agctagaatttggggtgaaaagtacttcggtaaaaactttgacagattggttaaggttaagactttagttgatccaaa taacttcttcagaaatgaacaatccatcccaccattgcctagacacagacactaa
SEQ ID NO:168 Medium chain fatty acyl-CoA synthetase Sc_FAA2 nucleotide sequence Saccharomyces sp.	atggccgctccagattatgcacttaccgatttaattgaatcggatcctcgtttcgaaagtttgaagacaagattagcc ggttacaccaaaggctctgatgaatatattgaagagctatactctcaattaccactgaccagctaccccaggtaca aaacatttttaaagaaacaggcggttgccatttcgaatccggataatgaagctggttttagctcgatttataggagtt ctctttcttctgaaaatctagtgagctgtgtggataaaaacttaagaactgcatacgatcacttcatgttttctgcaag gagatggcctcaacgtgactgtttaggttcaaggccaattgataaagccacaggcacctgggaggaaacattcc gtttcgagtcgtactccacggtatctaaaagatgtcataatatcggaagtggtatattgtctttggtaaacacgaaaa ggaaacgtcctttggaagccaatgattttgttgttgctatcttatcacacaacaaccctgaatggatcctaacagattt ggcctgtcaggcctattctctaactaacacggctttgtacgaaacattaggtccaaacacctccgagtacatattga atttaaccgaggccccattctgatttttgcaaaatcaaatatgtatcatgtattgaagatggtgcctgatatgaaattt gttaatactttggtttgtatggatgaattaactcatgacgagctccgtatgctaaatgaatcgttgctacccgttaagt gcaactctctcaatgaaaaaatcacatttttttcattggagcaggtagaacaagttggttgctttaacaaaattcctgc aattccacctaccccagattccttgtatactatttcgtttacttctggtactacaggtttacctaaaggtgtggaaatgt ctcacagaaacattgcgtctgggatagcatttgctttttctaccttcagaataccgccagataaaagaaaccaaca gttatatgatatgtgttttttgccattggctcatatttttgaaagaatggttattgcgtatgatctagccatcgggtttgga ataggcttcttacataaaccagacccaactgtattggtagaggatttgaagattttgaaaccttacgcggttgccct ggttcctagaatattaacacggtttgaagccggtataaaaaatgctttggataaatcgactgtccagaggaacgta gcaaatactatattggattctaaatcggccagatttaccgcaagaggtggtccagataaatcgattatgaattttcta gtttatcatcgcgtattgattgataaaatcagagactctttaggtttgtccaataactcgtttataattaccggatcagc tcccatatctaaagataccttactatttttaagaagcgccttggatattggtataagacagggctacggcttaactga aactttgctggtgtctgtttaagcgaaccgtttgaaaaagatgtcggatcttgtggtgccataggtatttctgcaga atgtagattgaagtctgttccagaaatgggttaccatgccgacaaggatttaaaaggtgaactgcaaattcgtggc ccacaggtttttgaaagatattttaaaaatccgaatgaaacttcaaaagccgttgaccaagatggttggttttccacg ggagatgttgcatttatcgatgcaaaaggtcgcatcagcgtcattgatcgagtcaagaacttttttcaagctagcaca tggtgaatatattgctccagagaaaatcgaaaatatttatttatcatcatgcccctatatcacgcaaatatttgtctttg gagatcctttgaagacatttttagttggcatcgttggtgttgatgttgatgcagcgcaaccgattttagctgcaaagc acccagaggtgaaaacgtggactaaggaagtgctagtagaaaacttaaatcgtaataaaaagctaaggaagga atttttaaacaaaattaataaatgcatcgatgggctacaaggatttgaaaaattgcacaacatcaaagtcggacttg agcctttgactctcgaggatgatgttgtgacgccaacttttaaaataaagcgtgccaaagcatcaaaattcttcaaa gatacattagaccaactatacgccgaaggttcactagtcaagacagaaaagctttag
SEQ ID NO:169 Medium chain fatty acyl-CoA synthetase Sc_FAA2 Saccharomyces sp.	MAAPDYALTDLIESDPRFESLKTRLAGYTKGSDEYIEELYSQLPLTSYP RYKTFLKKQAVAISNPDNEAGFSSIYRSSLSSENLVSCVDKNLRTAYDH FMFSARRWPQRDCLGSRPIDKATGTWEETFRFESYSTVSKRCHNIGSGI LSLVNTKRKRPLEANDFVVAILSHNNPEWILTDLACQAYSLTNTALYE TLGPNTSEYILNLTEAPILIFAKSNMYHVLKMVPDMKFVNTLVCMDEL THDELRMLNESLLPVKCNSLNEKITFFSLEQVEQVGCFNKIPAIPPTPDS LYTISFTSGTTGLPKGVEMSHRNIASGIAFAFSTFRIPPDKRNQQLYDMC FLPLAHIFERMVIAYDLAIGFGIGFLHKPDPTVLVEDLKILKPYAVALVP RILTRFEAGIKNALDKSTVQRNVANTILDSKSARFTARGGPDKSIMNFL VYHRVLIDKIRDSLGLSNNSFIITGSAPISKDTLLFLRSALDIGIRQGYGL TETFAGVCLSEPFEKDVGSCGAIGISAECRLKSVPEMGYHADKDLKGE LQIRGPQVFERYFKNPNETSKAVDQDGWFSTGDVAFIDAKGRISVIDR VKNFFKLAHGEYIAPEKIENIYLSSCPYITQIFVFGDPLKTFLVGIVGVD VDAAQPILAAKHPEVKTWTKEVLVENLNRNKKLRKEFLNKINKCIDGL QGFEKLHNIKVGLEPLTLEDDVVTPTFKIKRAKASKFFKDTLDQLYAE GSLVKTEKL*
SEQ ID NO:170 MBPtag	atgaaaatcgaagagggtaaattggtcatctggatcaatggtgacaaaggttacaacggtttggctgaagtcggt aaaaaattcgagaaagacactggtattaaggttaccgtcgaacacccagataagttggaagaaaaatttccacaa gttgccgctactggtgatggtccagacatcattttctgggcccacgacagatttggtggttatgctcaatctggtttg ttagccgagatcaccccagacaaagcctttcaagataaattatacccatttacctgggatgctgtccgttacaacg gtaagttgatcgcttacccaatcgccgttgaagctttgtctttaatctacaataaagacttattgccaaaccctccaaa

SEQ ID	SEQUENCE
	gacctgggaagaaattcctgccttggataaggaattaaaggctaaaggtaaatctgccttaatgttcaacttacaa gagccttactttacttggccattgattgctgctgatggtggttatgcttttaagtacgaaaatggtaaatacgacatta aagatgttggtgttgacaatgccggtgctaaagccggtttaactttcttagtcgacttgatcaagaacaagcacatg aatgctgacactgattattctatcgctgaagccgccttcaacaagggtgaaactgctatgactatcaatggtccttg ggcctggtctaatattgacacctccaaagtcaactacggtgttactgtcttaccaactttcaaaggtcaaccttccaa gccatttgtcggtgttttgtctgctggtattaacgctgcctctccaaacaaagaattggccaaggaatttttggaaaa ctacttgttgactgacgaaggtttagaggctgttaacaaagacaaaccattgggtgctgtcgccttgaaatcctac gaagaagaattagccaaggatccaagaatcgccgctaccatggaaaatgctcaaaaaggtgaaattatgccaa acattccacaaatgtccgcttttggtacgctgttagaactgctgttattaatgctgcttctggtagacaaactgtcga tgaagctttgaaggacgctcaaaccagaatcactaag
SEQ ID NO:171 GS12 Linker	ggaggtggaggaggtggttccggaggaggtggttct
SEQ ID NO:172 GS12 Linker	GGGGGGSGGGGS
SEQ ID NO:173 GB1 tag	atgtctgacacttacaagttgatcttgaacggtaagactttgaaaggtgaaactactaccgaagctgttgatgctgc cactgctgaaaaggtttttaagcaatacgccaatgataacggtgtcgacggtgaatggacttacgatgatgccact aagacttttaccgttactgaa
SEQ ID NO:174 GB1 tag	MSDTYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWT YDDATKTFTVTE
SEQ ID NO:175 MFalpha1_1-19	atgagatttccttcaatttttactgcagttttattcgcagcatcctccgcattagct
SEQ ID NO:176 MFalpha1_1-19	MRFPSIFTAVLFAASSALA
SEQ ID NO:177 MFalpha1_1-89	atgagatttccttcaatttttactgcagttttattcgcagcatcctccgcattagctgctccagtcaacactacaacag aagatgaaacggcacaaattccggctgaagctgtcatcggttacttagatttagaaggggatttcgatgttgctgtt ttgccattttccaacagcacaaataacgggttattgtttataaatactactattgccagcattgctgctaaagaagaa ggggtatctttggataaaagagaggctgaagct
SEQ ID NO:178 MFalpha1_1-89	MRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYLDLEGDFD VAVLPFSNSTNNGLLFINTTIASIAAKEEGVSLDKREAEA
SEQ ID NO:179 DasherGFP	atgaccgcactaacagaaggagctaaactattcgaaaaggagattccttacattacagaattagagggtgatgtc gaaggaatgaaattcattatcaagggcgagggtactggtgacgctactaccggtacgattaaagcaaagtacat ctgtacaacaggtgaccttcctgttccgtgggctactctggtgagcactttgtcttatggagttcaatgttttgctaaa taccttcgcacattaaagacttttcaaaagtgcaatgcctgagggctatactcaggagagaacaatatctttcga aggagatggtgtgtataagactagggctatggtcacgtatgaaagaggatccatctacaatagagtaactttaact ggtgaaaacttcaaaaaggacggtcacatccttagaaagaatgttgcctttcaatgcccaccatccatcttgtacat tttgccagacacagttaacaatggtatcagagttgagtttaaccaagcttatgacatagagggtgtcaccgaaaag ttggttacaaaatgttcacagatgaatcgtcccctggcaggatcagctgccgtccatatcccacgttaccatcatat cacttatcataccaagctgtccaaagatcgtgatgagagaagggatcacatgtgtttggttgaagtggtaaaggc cgtggatttggatacttaccaaggttga
SEQ ID NO:180 DasherGFP	MTALTEGAKLFEKEIPYITELEGDVEGMKFIIKGEGTGDATTGTIKAKYI CTTGDLPVPWATLVSTLSYGVQCFAKYPSHIKDFFKSAMPEGYTQERT ISFEGDGVYKTRAMVTYERGSIYNRVTLTGENFKKDGHILRKNVAFQC PPSILYILPDTVNNGIRVEFNQAYDIEGVTEKLVTKCSQMNRPLAGSAA VHIPRYHHITYHTKLSKDRDERRDHMCLVEVVKAVDLDTYQG*

SEQ ID	SEQUENCE
SEQ ID NO:181 ER1 tag	atattagagcaacctctgaaatttgtgcttactgcggccgtcgtgctcttgacgacgtcggttctttgttgtgtagtatt tacataa
SEQ ID NO:182 ER1 tag	ILEQPLKFVLTAAVVLLTTSVLCCVVFT*
SEQ ID NO:183 ER2 tag	tctacctctgaaaaccaaagtaaaggtagtggtacattggttgtcatattggccattttaatgctaggtgttgcttatta tttgttgaacgaataa
SEQ ID NO:184 ER2 tag	STSENQSKGSGTLVVILAILMLGVAYYLLNE*
SEQ ID NO:185 PM1 tag	tggtacaaggatctaaaaatgaagatgtgtctggctttagtaatcatcatattgcttgttgtaatcatcgtccccattg ctgttcactttagtcgataa
SEQ ID NO:186 PM1 tag	WYKDLKMKMCLALVIIILLVVIIVPIAVHFSR*
SEQ ID NO:187 VC1 tag	aatataaaagaaataatgtggtggcagaaggtcaaaaatattacgttattaactttcactattatactatttgtaagtg ctgctttcatgtttttctatctgtggtaa
SEQ ID NO:188 VC1 tag	NIKEIMWWQKVKNITLLTFTIILFVSAAFMFFYLW*
SEQ ID NO:189 PEX8 tag	tctaaattataa
SEQ ID NO:190 PEX8 tag	SKL*
SEQ ID NO:191 Long chain fatty acyl-CoA synthetase Sc_FAA1 Saccharomyces cerevisiae	atggttgctcaatataccgttccagttgggaaagccgccaatgagcatgaaactgctccaagaagaaattatcaat gccgcgagaagccgctcgtcagaccgcctaacacaaagtgttccactgtttatgagtttgttctagagtgctttca gaagaacaaaaattcaaatgctatgggttggagggatgttaaggaaattcatgaagaatccaaatcggttatgaa aaaagttgatggcaaggagacttcagtggaaaagaaatggatgtattatgaactatcgcattatcattataattcatt tgaccaattgaccgatatcatgcatgaaattggtcgtgggttggtgaaaataggattaaagcctaatgatgatgac aaattacatctttacgcagccacttctcacaagtggatgaagatgttcttaggagcgcagtctcaaggtattcctgtc gtcactgcctacgatactttgggagagaaagggctaattcattctttggtgcaaacggggtctaaggccatttttac cgataactctttattaccatccttgatcaaaccagtgcaagccgctcaagacgtaaaatacataattcatttcgattc catcagttctgaggacaggaggcaaagtggtaagatctatcaatctgctcatgatgccatcaacagaattaaaga agttagacctgatatcaagacctttagctttgacgacatcttgaagctaggtaaagaatcctgtaacgaaatcgatg ttcatccacctggcaaggatgatctttgttgcatcatgtatacgtctggttctacaggtgagccaaagggtgttgtctt gaaacattcaaatgttgtcgcaggtgttggtggtgcaagtttgaatgttttgaagtttgtgggcaataccgaccgtgt tatctgtttttgccactagctcatatttttgaattggttttcgaactattgtccttttattgggggggcctgcattggttatg ccaccgtaaaaactttaactagcagctctgtgagaaattgtcaaggtgatttgcaagaattcaagcccacaatcat ggttggtgtcgccgctgtttgggaaacagtgagaaaagggatcttaaaccaaattgataatttgcccttcctcacc aagaaaatcttctggaccgcgtataataccaagttgaacatgcaacgtctccacatccctggtggcggcgcctta ggaaacttggttttcaaaaaaatcagaactgccacaggtggccaattaagatatttgttaaacggtggttctccaat cagtcgggatgctcaggaattcatcacaaatttaatctgccctatgcttattggttacggtttaaccgagacatgcg ctagtaccaccatcttggatcctgctaattttgaactcggcgtcgctggtgacctaacaggttgtgttaccgtcaaa ctagttgatgttgaagaattaggttattttgctaaaaacaaccaaggtgaagtttggatcacaggtgccaatgtcac gcctgaatattataagaatgaggaagaaacttctcaagctttaacaagcgatggttggttcaagaccggtgacatc ggtgaatgggaagcaaatggccatttgaaaataattgacaggaagaaaaacttggtcaaaacaatgaacggtga

2018256863 11 Apr 2024

SEQ ID	SEQUENCE
	atatatcgcactcgagaaattagagtccgtttacagatctaacgaatatgttgctaacatttgtgtttatgccgaccaa tctaagactaagccagttggtattattgtaccaaatcatgctccattaacgaagcttgctaaaaagttgggaattatg gaacaaaaagacagttcaattaatatcgaaaattatttggaggatgcaaaattgattaaagctgtttattctgatctttt gaagacaggtaaagaccaaggtttggttggcattgaattactagcaggcatagtgttctttgacggcgaatggact ccacaaaacggttttgttacgtccgctcagaaattgaaaagaaaagacattttgaatgctgtcaaagataaagttga cgccgtttatagttcgtcttaa
SEQ ID NO:192 Long chain fatty acyl-CoA synthetase Sc_FAA1 Saccharomyces cerevisiae	MVAQYTVPVGKAANEHETAPRRNYQCREKPLVRPPNTKCSTVYEFVL ECFQKNKNSNAMGWRDVKEIHEESKSVMKKVDGKETSVEKKWMYY ELSHYHYNSFDQLTDIMHEIGRGLVKIGLKPNDDDKLHLYAATSHKW MKMFLGAQSQGIPVVTAYDTLGEKGLIHSLVQTGSKAIFTDNSLLPSLI KPVQAAQDVKYIIHFDSISSEDRRQSGKIYQSAHDAINRIKEVRPDIKTF SFDDILKLGKESCNEIDVHPPGKDDLCCIMYTSGSTGEPKGVVLKHSNV VAGVGGASLNVLKFVGNTDRVICFLPLAHIFELVFELLSFYWGACIGY ATVKTLTSSSVRNCQGDLQEFKPTIMVGVAAVWETVRKGILNQIDNLP FLTKKIFWTAYNTKLNMQRLHIPGGGALGNLVFKKIRTATGGQLRYLL NGGSPISRDAQEFITNLICPMLIGYGLTETCASTTILDPANFELGVAGDL TGCVTVKLVDVEELGYFAKNNQGEVWITGANVTPEYYKNEEETSQAL TSDGWFKTGDIGEWEANGHLKIIDRKKNLVKTMNGEYIALEKLESVY RSNEYVANICVYADQSKTKPVGIIVPNHAPLTKLAKKLGIMEQKDSSIN IENYLEDAKLIKAVYSDLLKTGKDQGLVGIELLAGIVFFDGEWTPQNG FVTSAQKLKRKDILNAVKDKVDAVYSSS*
SEQ ID NO:193 Truncated medium chain fatty acyl-CoA synthetase Sc_FAA2_Ctrunc	atggccgctccagattatgcacttaccgatttaattgaatcggatcctcgtttcgaaagtttgaagacaagattagcc ggttacaccaaaggctctgatgaatatattgaagagctatactctcaattaccactgaccagctaccccaggtaca aaacatttttaaagaaacaggcggttgccatttcgaatccggataatgaagctggttttagctcgatttataggagtt ctctttcttctgaaaatctagtgagctgtgtggataaaaacttaagaactgcatacgatcacttcatgttttctgcaag gagatggcctcaacgtgactgtttaggttcaaggccaattgataaagccacaggcacctgggaggaaacattcc gtttcgagtcgtactccacggtatctaaaagatgtcataatatcggaagtggtatattgtctttggtaaacacgaaaa ggaaacgtcctttggaagccaatgattttgttgttgctatcttatcacacaacaaccctgaatggatcctaacagattt ggcctgtcaggcctattctctaactaacacggctttgtacgaaacattaggtccaaacacctccgagtacatattga atttaaccgaggcccccattctgatttttgcaaaatcaaatatgtatcatgtattgaagatggtgcctgatatgaaattt gttaatactttggtttgtatggatgaattaactcatgacgagctccgtatgctaaatgaatcgttgctacccgttaagt gcaactctctcaatgaaaaaatcacatttttttcattggagcaggtagaacaagttggttgctttaacaaaattcctgc aattccacctaccccagattccttgtatactatttcgtttacttctggtactacaggtttacctaaaggtgtggaaatgt ctcacagaaacattgcgtctgggatagcatttgctttttctaccttcagaataccgccagataaaagaaaccaaca gttatatgatatgtgttttttgccattggctcatatttttgaaagaatggttattgcgtatgatctagccatcgggtttgga ataggcttcttacataaaccagacccaactgtattggtagaggatttgaagattttgaaaccttacgcggttgccct ggttcctagaatattaacacggtttgaagccggtataaaaaatgctttggataaatcgactgtccagaggaacgta gcaaatactatattggattctaaatcggccagatttaccgcaagaggtggtccagataaatcgattatgaattttcta gtttatcatcgcgtattgattgataaaatcagagactctttaggtttgtccaataactcgtttataattaccggatcagc tcccatatctaaagataccttactatttttaagaagcgccttggatattggtataagacagggctacggcttaactga aacttttgctggtgtctgtttaagcgaaccgtttgaaaaagatgtcggatcttgtggtgccataggtatttctgcaga atgtagattgaagtctgttccagaaatgggttaccatgccgacaaggatttaaaaggtgaactgcaaattcgtggc ccacaggtttttgaaagatattttaaaaatccgaatgaaacttcaaaagccgttgaccaagatggttggttttccacg ggagatgttgcatttatcgatgcaaaaggtcgcatcagcgtcattgatcgagtcaagaacttttcaagctagcaca tggtgaatatattgctccagagaaaatcgaaaatatttatttatcatcatgcccctatatcacgcaaatatttgtctttg gagatcctttgaagacatttttagttggcatcgttggtgttgatgttgatgcagcgcaaccgattttagctgcaaagc acccagaggtgaaaacgtggactaaggaagtgctagtagaaaacttaaatcgtaataaaaagctaaggaagga atttttaaacaaaattaataaatgcatcgatgggctacaaggatttgaaaaattgcacaacatcaaagtcggacttg agcctttgactctcgaggatgatgttgtgacgccaacttttaaaataaagcgtgccaaagcatcaaaattcttcaaa gatacattagaccaactatacgccgaaggttcactagtcaagacatag
SEQ ID NO:194 Truncated medium chain fatty acyl-CoA	MAAPDYALTDLIESDPRFESLKTRLAGYTKGSDEYIEELYSQLPLTSYP RYKTFLKKQAVAISNPDNEAGFSSIYRSSLSSENLVSCVDKNLRTAYDH FMFSARRWPQRDCLGSRPIDKATGTWEETFRFESYSTVSKRCHNIGSGI

2018256863 11 Apr 2024

SEQ ID	SEQUENCE
synthetase Sc_FAA2_Ctrunc	LSLVNTKRKRPLEANDFVVAILSHNNPEWILTDLACQAYSLTNTALYE TLGPNTSEYILNLTEAPILIFAKSNMYHVLKMVPDMKFVNTLVCMDEL THDELRMLNESLLPVKCNSLNEKITFFSLEQVEQVGCFNKIPAIPPTPDS LYTISFTSGTTGLPKGVEMSHRNIASGIAFAFSTFRIPPDKRNQQLYDMC FLPLAHIFERMVIAYDLAIGFGIGFLHKPDPTVLVEDLKILKPYAVALVP RILTRFEAGIKNALDKSTVQRNVANTILDSKSARFTARGGPDKSIMNFL VYHRVLIDKIRDSLGLSNNSFIITGSAPISKDTLLFLRSALDIGIRQGYGL TETFAGVCLSEPFEKDVGSCGAIGISAECRLKSVPEMGYHADKDLKGE LQIRGPQVFERYFKNPNETSKAVDQDGWFSTGDVAFIDAKGRISVIDR VKNFFKLAHGEYIAPEKIENIYLSSCPYITQIFVFGDPLKTFLVGIVGVD VDAAQPILAAKHPEVKTWTKEVLVENLNRNKKLRKEFLNKINKCIDGL QGFEKLHNIKVGLEPLTLEDDVVTPTFKIKRAKASKFFKDTLDQLYAE GSLVKT*
SEQ ID NO:195 Mutated medium chain fatty acyl-CoA synthetase Sc_FAA2_Cmut	atggccgctccagattatgcacttaccgatttaattgaatcggatcctcgtttcgaaagtttgaagacaagattagcc ggttacaccaaaggctctgatgaatatattgaagagctatactctcaattaccactgaccagctaccccaggtaca aaacatttttaaagaaacaggcggttgccatttcgaatccggataatgaagctggttttagctcgatttataggagtt ctctttcttctgaaaatctagtgagctgtgtggataaaaacttaagaactgcatacgatcacttcatgttttctgcaag gagatggcctcaacgtgactgtttaggttcaaggccaattgataaagccacaggcacctgggaggaaacattcc gtttcgagtcgtactccacggtatctaaaagatgtcataatatcggaagtggtatattgtctttggtaaacacgaaaa ggaaacgtcctttggaagccaatgattttgttgttgctatcttatcacacaacaaccctgaatggatcctaacagattt ggcctgtcaggcctattctctaactaacacggctttgtacgaaacattaggtccaaacacctccgagtacatattga atttaaccgaggccccattctgatttttgcaaaatcaaatatgtatcatgtattgaagatggtgcctgatatgaaattt gttaatactttggtttgtatggatgaattaactcatgacgagctccgtatgctaaatgaatcgttgctacccgttaagt gcaactctctcaatgaaaaaatcacatttttttcattggagcaggtagaacaagttggttgctttaacaaaattcctgc aattccacctaccccagattccttgtatactatttcgtttacttctggtactacaggtttacctaaaggtgtggaaatgt ctcacagaaacattgcgtctgggatagcatttgctttttctaccttcagaataccgccagataaaagaaaccaaca gttatatgatatgtgtttttgccattggctcatatttttgaaagaatggttattgcgtatgatctagccatcgggtttgga ataggcttcttacataaaccagacccaactgtattggtagaggatttgaagattttgaaaccttacgcggttgccct ggttcctagaatattaacacggtttgaagccggtataaaaaatgctttggataaatcgactgtccagaggaacgta gcaaatactatattggattctaaatcggccagatttaccgcaagaggtggtccagataaatcgattatgaattttcta gtttatcatcgcgtattgattgataaaatcagagactctttaggtttgtccaataactcgtttataattaccggatcagc tcccatatctaaagataccttactatttttaagaagcgccttggatattggtataagacagggctacggcttaactga aacttttgctggtgtctgtttaagcgaaccgtttgaaaaagatgtcggatcttgtggtgccataggtatttctgcaga atgtagattgaagtctgttccagaaatgggttaccatgccgacaaggatttaaaaggtgaactgcaaattcgtggc ccacaggtttttgaaagatattttaaaaatccgaatgaaacttcaaaagccgttgaccaagatggttggttttccacg ggagatgttgcatttatcgatgcaaaaggtcgcatcagcgtcattgatcgagtcaagaacttttcaagctagcaca tggtgaatatattgctccagagaaaatcgaaaatatttatttatcatcatgcccctatatcacgcaaatatttgtctttg gagatcctttgaagacatttttagttggcatcgttggtgttgatgttgatgcagcgcaaccgattttagctgcaaagc acccagaggtgaaaacgtggactaaggaagtgctagtagaaaacttaaatcgtaataaaaagctaaggaagga atttttaaacaaaattaataaatgcatcgatgggctacaaggatttgaaaaattgcacaacatcaaagtcggacttg agcctttgactctcgaggatgatgttgtgacgccaacttttaaaataaagcgtgccaaagcatcaaaattcttcaaa gatacattagaccaactatacgccgaaggttcactagtcaagacagaaaagcttaaatag
SEQ ID NO:196 Mutated medium chain fatty acyl-CoA synthetase Sc_FAA2_Cmut	MAAPDYALTDLIESDPRFESLKTRLAGYTKGSDEYIEELYSQLPLTSYP RYKTFLKKQAVAISNPDNEAGFSSIYRSSLSSENLVSCVDKNLRTAYDH FMFSARRWPQRDCLGSRPIDKATGTWEETFRFESYSTVSKRCHNIGSGI LSLVNTKRKRPLEANDFVVAILSHNNPEWILTDLACQAYSLTNTALYE TLGPNTSEYILNLTEAPILIFAKSNMYHVLKMVPDMKFVNTLVCMDEL THDELRMLNESLLPVKCNSLNEKITFFSLEQVEQVGCFNKIPAIPPTPDS LYTISFTSGTTGLPKGVEMSHRNIASGIAFAFSTFRIPPDKRNQQLYDMC FLPLAHIFERMVIAYDLAIGFGIGFLHKPDPTVLVEDLKILKPYAVALVP RILTRFEAGIKNALDKSTVQRNVANTILDSKSARFTARGGPDKSIMNFL VYHRVLIDKIRDSLGLSNNSFIITGSAPISKDTLLFLRSALDIGIRQGYGL TETFAGVCLSEPFEKDVGSCGAIGISAECRLKSVPEMGYHADKDLKGE

SEQ ID	SEQUENCE
	LQIRGPQVFERYFKNPNETSKAVDQDGWFSTGDVAFIDAKGRISVIDR VKNFFKLAHGEYIAPEKIENIYLSSCPYITQIFVFGDPLKTFLVGIVGVD VDAAQPILAAKHPEVKTWTKEVLVENLNRNKKLRKEFLNKINKCIDGL QGFEKLHNIKVGLEPLTLEDDVVTPTFKIKRAKASKFFKDTLDQLYAE GSLVKTEKLK*
SEQ ID NO:197 Long-chain fatty acyl-CoA synthetase Sc_FAA3 Saccharomyces cerevisiae	atgtccgaacaacactctgtcgcagtcggtaaagctgctaatgagcacgagactgcccctaggagaaatgttag agtcaagaagcggcccttaattagaccattgaactcgtcagcatctacgctgtatgaatttgccctagagtgtttca acaagggtggaaaacgagatggtatggcttggagagatgtcatcgagattcatgagacaaagaaaaccattgtg agaaaggtagacggcaaggataaatctatagaaaagacatggctgtattatgaaatgtcaccatataaaatgatg acctaccaggaactgatctgggtgatgcacgatatgggccgtgggctggcaaaaataggcatcaagcccaatg gagaacacaaattccacatcttcgcatctacttcccataaatggatgaagattttccttggttgcatatcccagggta tccccgtagtaaccgcgtatgatactttgggtgagagcggtttgattcactccatggttgaaaccgagtctgctgct attttcactgataatcaattattggctaaaatgatagtgcctttgcaatctgctaaagatatcaaatttcttatccataac gaacctatcgaccccaatgacagaagacaaaacggcaaactttacaaggctgctaaggatgccattaataagat cagagaagttaggccagacataaaaatttatagttttgaagaagttgtcaagataggtaaaaaaagtaaagatga ggtcaaacttcatccacctgagccaaaagatttggcttgtatcatgtacacctcgggctcgatcagtgcaccaaaa ggtgtagtattgactcattataatattgtttcgggtatcgctggtgtaggtcacaacgtctttggatggatcggctcta cagaccgtgttttgtcgttcttgccattggctcatatttttgaactggtctttgaattcgaagccttttactggaacggta ttcttgggtacggtagtgttaagactttgactaatacttcgactcgtaattgtaagggtgacctggttgagtttaagcc tactattatgatcggtgtggctgccgtttgggaaactgtgagaaaagctattttggaaaagatcagcgatttaactc ccgtactccaaaagatttttggtctgcctatagtatgaaagaaaagagtgtaccatgcaccgggtttttaagtcgta tggtcttcaagaaagtcagacaagccaccggtggtcatcttaagtatattatgaacggtgggtctgcgatcagtatt gatgctcagaaattcttttctatcgtcctgtgtcctatgattatcggttacggccttactgaaacagttgcgaatgcttg tgttttggagcctgatcatttcgaatatggtatagttggtgatcttgttggatcggtcactgccaaattggtggatgtta aggacctaggttattatgcaaaaaacaatcaaggtgaattgcttctaaagggtgcgccggtctgttctgaatattat aagaatccaatagaaacggcggtctctttcacttacgatggatggtttcgtactggtgatattgttgaatggactccc aagggacaacttaaaattattgatagaagaaagaatttggttaaaaccctaaatggtgaatatattgcattagaaaa gttagaatctgtttacaggtcaaactcctatgtgaaaaatatctgtgtttatgccgatgaaagtagggttaaaccggt gggtattgtggtacccaacccaggacccctatctaaatttgctgtcaaattgcgtattatgaaaaagggtgaagac atcgaaaactatatccatgacaaagcattacgaaatgctgttttcaaagagatgatcgcaacagccaaatctcaag gtttggttggtattgaactattatgtggtattgttttctttgatgaagaatggacacctgaaaatggctttgtcacatctg ctcaaaaattaaagagaagagaaatcttagccgctgttaaatcagaagtcgaaagggtttacaaagaaaattctta g
SEQ ID NO:198 Long-chain fatty acyl-CoA synthetase Sc_FAA3 Saccharomyces cerevisiae	MSEQHSVAVGKAANEHETAPRRNVRVKKRPLIRPLNSSASTLYEFALE CFNKGGKRDGMAWRDVIEIHETKKTIVRKVDGKDKSIEKTWLYYEMS PYKMMTYQELIWVMHDMGRGLAKIGIKPNGEHKFHIFASTSHKWMKI FLGCISQGIPVVTAYDTLGESGLIHSMVETESAAIFTDNQLLAKMIVPLQ SAKDIKFLIHNEPIDPNDRRQNGKLYKAAKDAINKIREVRPDIKIYSFEE VVKIGKKSKDEVKLHPPEPKDLACIMYTSGSISAPKGVVLTHYNIVSGI AGVGHNVFGWIGSTDRVLSFLPLAHIFELVFEFEAFYWNGILGYGSVK TLTNTSTRNCKGDLVEFKPTIMIGVAAVWETVRKAILEKISDLTPVLQK IFWSAYSMKEKSVPCTGFLSRMVFKKVRQATGGHLKYIMNGGSAISID AQKFFSIVLCPMIIGYGLTETVANACVLEPDHFEYGIVGDLVGSVTAKL VDVKDLGYYAKNNQGELLLKGAPVCSEYYKNPIETAVSFTYDGWFRT GDIVEWTPKGQLKIIDRRKNLVKTLNGEYIALEKLESVYRSNSYVKNIC VYADESRVKPVGIVVPNPGPLSKFAVKLRIMKKGEDIENYIHDKALRN AVFKEMIATAKSQGLVGIELLCGIVFFDEEWTPENGFVTSAQKLKRREI LAAVKSEVERVYKENS*
SEQ ID NO:199 Long-chain fatty acyl-CoA synthetase Sc_FAA4	atgaccgaacaatattccgttgcagttggcgaagccgacaatgagcatgaaaccgctccaagaagaaatatcag ggttaaagacaagcctttgattagacccataaactcctcagcatctacactgtacgaattcgccctggaatgttttac caaaggtggtaagagagacggtatggcatggagagatattatagatatacatgagacgaaaaaaaccatagtca agagggtggatggtaaggataagcccatcgaaaaaacatggttgtactacgaactgactccctacataaccatg acatacgaggagatgatctgcgtaatgcacgacattggacgtgggctgataaagattggtgttaaacctaacggt

SEQ ID	SEQUENCE
Saccharomyces cerevisiae	gagaacaagttccacatctttgcctctacatctcacaagtggatgaaaacttttcttggttgcatgtcacaaggtattc ctgtggtcaccgcgtacgacactttgggtgagagcggtttgattcactccatggtggaaacggattccgtcgccat tttcacggacaaccagctgttgtccaaattagcagttcctttgaaaaccgccaagaacgtaaaattcgtcattcaca acgaacccatcgatccaagtgacaaaagacaaaatggtaagctttacaaggctgccaaggatgctgttgacaaa atcaaggaagttagaccggacataaaaatctacagtttcgatgaaattattgagataggtaaaaaggccaaggac gaggttgaattgcatttccccaagcctgaagatccagcttgtatcatgtacacttctggttccactggtacaccaaa gggtgtggtattgacacattacaacattgtagctggtattggtggtgtgggccataacgttatcggatggattggcc caacagaccgtattatcgcattcttgccattggctcatatttttgaattaatctttgaattcgaagcgttctactggaat ggtatcctagggtacgccactgtcaagactttaaccccaacttctacacgtaattgccaaggtgacctgatggagt ttaaacctaccgtaatggtaggtgttgccgcagtttgggaaacagtgagaaaaggtatcctggccaagatcaacg aattgcccggttggtctcaaacgcttttctggactgtctatgctttgaaagagagaaatataccatgcagcggcttg ctgagtgggttgatcttcaagagaatcagagaagcaaccggtggaaacttaaggtttattctgaacggtgggtct gcaatcagcatagacgcccaaaaattcctctccaaccttctatgtcctatgctcattggatatgggctaactgaggg tgtggctaatgcctgtgtcctggagcctgaacattttgattacggtattgctggtgaccttgtcggaactattacagc taaattggtggatgtcgaagatttgggctattttgccaagaataaccaaggtgaattgctgttaaagggtgcaccca tctgttctgaatactataagaatcctgaagaaactgctgcggcctttaccgatgatggctggttccgtaccggtgat atcgctgaatggacccccaagggacaaattaagatcattgatagaaagaaaaatttggtcaagaccttaaacggt gagtacattgcattggaaaaattagaatccatttacagatcaaatccttacgtccaaaacatctgtgtctacgctgat gaaaacaaagttaagcctgtcggtattgtggtccctaacttaggacacttgtctaagctggctatcgaattaggtat aatggtaccaggtgaagatgtcgaaagctatatccatgaaaagaagctacaggatgccgtttgcaaagatatgct gtcaactgccaaatctcaaggcttgaatggtattgaattattatgtggcattgttttctttgaagaagaatggactcca gaaaacggtcttgttacatccgcccaaaaattaaagagaagagatattctagcggctgtcaagccagatgtggaa agagtttataaagaaaacacttaa
SEQ ID NO:200 Sc_FAA4 Saccharomyces cerevisiae	MTEQYSVAVGEADNEHETAPRRNIRVKDKPLIRPINSSASTLYEFALEC FTKGGKRDGMAWRDIIDIHETKKTIVKRVDGKDKPIEKTWLYYELTPY ITMTYEEMICVMHDIGRGLIKIGVKPNGENKFHIFASTSHKWMKTFLG CMSQGIPVVTAYDTLGESGLIHSMVETDSVAIFTDNQLLSKLAVPLKTA KNVKFVIHNEPIDPSDKRQNGKLYKAAKDAVDKIKEVRPDIKIYSFDEII EIGKKAKDEVELHFPKPEDPACIMYTSGSTGTPKGVVLTHYNIVAGIGG VGHNVIGWIGPTDRIIAFLPLAHIFELIFEFEAFYWNGILGYATVKTLTPT STRNCQGDLMEFKPTVMVGVAAVWETVRKGILAKINELPGWSQTLF WTVYALKERNIPCSGLLSGLIFKRIREATGGNLRFILNGGSAISIDAQKF LSNLLCPMLIGYGLTEGVANACVLEPEHFDYGIAGDLVGTITAKLVDV EDLGYFAKNNQGELLLKGAPICSEYYKNPEETAAAFTDDGWFRTGDIA EWTPKGQIKIIDRKKNLVKTLNGEYIALEKLESIYRSNPYVQNICVYAD ENKVKPVGIVVPNLGHLSKLAIELGIMVPGEDVESYIHEKKLQDAVCK DMLSTAKSQGLNGIELLCGIVFFEEEWTPENGLVTSAQKLKRRDILAA VKPDVERVYKENT*
SEQ ID NO:201 Mutated acetyl-CoA carboxylase (ACC1) (S659A, S1157A)	atgagcgaagaaagcttattcgagtcttctccacagaagatggagtacgaaattacaaactactcagaaagacat acagaacttccaggtcatttcattggcctcaatacagtagataaactagaggagtccccgttaagggactttgttaa gagtcacggtggtcacacggtcatatccaagatcctgatagcaaataatggtattgccgccgtgaaagaaattag atccgtcagaaaatgggcatacgagacgttcggcgatgacagaaccgtccaattcgtcgccatggccacccca gaagatctggaggccaacgcagaatatatccgtatggccgatcaatacattgaagtgccaggtggtactaataat aacaactacgctaacgtagacttgatcgtagacatcgccgaaagagcagacgtagacgccgtatgggctggct ggggtcacgcctccgagaatccactattgcctgaaaaattgtcccagtctaagaggaaagtcatctttattgggcc tccaggtaacgccatgaggtctttaggtgataaaatctcctctaccattgtcgctcaaagtgctaaagtcccatgtat tccatggtctggtaccggtgttgacaccgttcacgtggacgagaaaaccggtctggtctctgtcgacgatgacat ctatcaaaagggttgttgtacctctcctgaagatggtttacaaaaggccaagcgtattggttttcctgtcatgattaa ggcatccgaaggtggtggtggtaaaggtatcagacaagttgaacgtgaagaagatttcatcgctttataccacca ggcagccaacgaaattccaggctcccccattttcatcatgaagttggccggtagagcgcgtcacttggaagttca actgctagcagatcagtacggtacaaatatttccttgttcggtagagactgttccgttcagagacgtcatcaaaaaa ttatcgaagaagcaccagttacaattgccaaggctgaaacatttcacgagatggaaaaggctgccgtcagactg gggaaactagtcggttatgtctctgccggtaccgtggagtatctatattctcatgatgatggaaaattctacttttag

SEQ ID	SEQUENCE
	aattgaacccaagattacaagtcgagcatccaacaacggaaatggtctccggtgttaacttacctgcagctcaatt acaaatcgctatgggtatccctatgcatagaataagtgacattagaactttatatggtatgaatcctcattctgcctca gaaatcgatttcgaattcaaaactcaagatgccaccaagaaacaaagaagacctattccaaagggtcattgtacc gcttgtcgtatcacatcagaagatccaaacgatggattcaagccatcgggtggtactttgcatgaactaaacttcc gttcttcctctaatgtttggggttacttctccgtgggtaacaatggtaatattcactccttttcggactctcagttcggc catatttttgcttttggtgaaaatagacaagcttccaggaaacacatggttgttgccctgaaggaattgtccattagg ggtgatttcagaactactgtggaatacttgatcaaacttttggaaactgaagatttcgaggataacactattaccacc ggttggttggacgatttgattactcataaaatgaccgctgaaaagcctgatccaactcttgccgtcatttgcggtgc cgctacaaaggctttcttagcatctgaagaagcccgccacaagtatatcgaatccttacaaaagggacaagttcta tctaaagacctactgcaaactatgttccctgtagattttatccatgagggtaaaagatacaagttcaccgtagctaaa tccggtaatgaccgttacacattatttatcaatggttctaaatgtgatatcatactgcgtcaactatctgatggtggtct tttgattgccataggcggtaaatcgcataccatctattggaaagaagaagttgctgctacaagattatccgttgact ctatgactactttgttggaagttgaaaacgatccaacccagttgcgtactccatccccctggtaaattggttaaattctt ggtggaaaatggtgaacacattatcaagggccaaccatatgcagaaattgaagttatgaaaatgcaaatgccttt ggtttctcaagaaaatggtatcgtccagttattaaagcaacctggttctaccattgttgcaggtgatatcatggctatt atgactcttgacgatccatccaaggtcaagcacgctctaccatttgaaggtatgctgccagattttggttctccagtt atcgaaggaaccaaacctgcctataaattcaagtcattagtgtctactttggaaaacattttgaagggttatgacaa ccaagttattatgaacgcttccttgcaacaattgatagaggttttgagaaatccaaaactgccttactcagaatgga aactacacatctctgctttacattcaagattgcctgctaagctagatgaacaaatggaagagttagttgcacgttctttt gagacgtggtgctgttttcccagctagacaattaagtaaattgattgatatggccgtgaagaatcctgaatacaacc ccgacaaattgctgggcgccgtcgtggaaccattggcggatattgctcataagtactctaacgggttagaagccc atgaacattctatatttgtccatttcttggaagaatattacgaagttgaaaagttattcaatggtccaaatgttcgtgag gaaaatatcattctgaaattgcgtgatgaaaaccctaaagatctagataaagttgcgctaactgttttgtctcattcga aagtttcagcgaagaataacctgatcctagctatcttgaaacattatcaaccattgtgcaagttatcttctaaagtttct gccattttctctactcctctacaacatattgttgaactagaatctaaggctaccgctaaggtcgctctacaagcaaga gaaattttgattcaaggcgctttaccttcggtcaaggaaagaactgaacaaattgaacatatcttaaaatcctctgtt gtgaaggttgcctatggctcatccaatccaaagcgctctgaaccagatttgaatatcttgaaggacttgatcgattc taattacgttgtgttcgatgttttacttcaattcctaacccatcaagacccagttgtgactgctgcagctgctcaagtct atattcgtcgtgcttatcgtgcttacaccataggagatattagagttcacgaaggtgtcacagttccaattgttgaat ggaaattccaactaccttcagctgcgttctccacctttccaactgttaaatctaaaatgggtatgaacagggctgttt ctgtttcagatttgtcatatgttgcaaacagtcagtcatctccgttaagagaaggtattttgatggctgtggatcattta gatgatgttgatgaaattttgtcacaaagtttggaagttattcctcgtcaccaatcttcttctaacggacctgctcctga tcgttctggtagctccgcatcgttgagtaatgttgctaatgtttgtgttgcttctacagaaggtttcgaatctgaagag gaaattttggtaaggttgagagaaattttggatttgaataagcaggaattaatcaatgcttctatccgtcgtatcacat ttatgttcggttttaaagatgggtcttatccaaagtattatacttttaacggtccaaattataacgaaaatgaaacaatt cgtcacattgagccggctttggccttccaactggaattaggaagattgtccaacttcaacattaaaccaatttttcact gataatagaaacatccatgtctacgaagctgttagtaagacttctccattggataagagattctttacaagaggtatt attagaacgggtcatatccgtgatgacatttctattcaagaatatctgacttctgaagctaacagattgatgagtgat atattggataatttagaagtcaccgacacttcaaattctgatttgaatcatatcttcatcaacttcattgcggtgtttgat atctctccagaagatgtcgaagccgccttcggtggtttcttagaaagatttggtaagagattgttgagattgcgtgtt tcttctgccgaaattagaatcatcatcaaagatcctcaaacaggtgccccagtaccattgcgtgccttgatcaataa cgtttctggttatgttatcaaaacagaaatgtacaccgaagtcaagaacgcaaaaggtgaatgggtatttaagtctt tgggtaaacctggatccatgcatttaagacctattgctactccttaccctgttaaggaatggttgcaaccaaaacgtt ataaggcacacttgatgggtaccacatatgtctatgacttcccagaattattccgccaagcatcgtcatcccaatgg aaaaatttctctgcagatgttaagttaacagatgatttctttatttccaacgagttgattgaagatgaaaacggcgaat taactgaggtggaaagagaacctggtgccaacgctattggtatggttgcctttaagattactgtaaagactcctga atatccaagaggccgtcaatttgttgttgttgctaacgatatcacattcaagatcggttcctttggtccacaagaaga cgaattcttcaataaggttactgaatatgctagaaagcgtggtatcccaagaatttacttggctgcaaactcaggtg ccagaattggtatggctgaagagattgttccactatttcaagttgcatggaatgatgctgccaatccggacaaggg cttccaatacttatacttaacaagtgaaggtatggaaactttaaagaaatttgacaaagaaaattctgttctcactgaa cgtactgttataaacggtgaagaaagatttgtcatcaagacaattattggttctgaagatgggttaggtgtcgaatgt ctacgtggatctggtttaattgctggtgcaacgtcaagggcttaccacgatatcttcactatcaccttagtcacttgta gatccgtcggtatcggtgcttatttggttcgtttgggtcaaagagctattcaggtcgaaggccagccaattattttaa ctggtgctcctgcaatcaacaaaatgctgggtagagaagtttatacttctaacttacaattgggtggtactcaaatca

SEQ ID	SEQUENCE
	tgtataacaacggtgtttcacatttgactgctgttgacgatttagctggtgtagagaagattgttgaatggatgtcttat gttccagccaagcgtaatatgccagttcctatcttggaaactaaagacacatgggatagaccagttgatttcactcc aactaatgatgaaacttacgatgtaagatggatgattgaaggtcgtgagactgaaagtggatttgaatatggtttgtt tgataaagggtctttctttgaaactttgtcaggatgggccaaaggtgttgtcgttggtagagcccgtcttggtggtat tccactgggtgttattggtgttgaaacaagaactgtcgagaacttgattcctgctgatccagctaatccaaatagtg ctgaaacattaattcaagaacctggtcaagtttggcatccaaactccgccttcaagactgctcaagctatcaatgac tttaacaacggtgaacaattgccaatgatgattttggccaactggagaggtttctctggtggtcaacgtgatatgttc aacgaagtcttgaagtatggttcgtttattgttgacgcattggtggattacaaacaaccaattattatctatatcccac ctaccggtgaactaagaggtggttcatgggttgttgtcgatccaactatcaacgctgaccaaatggaaatgtatgc cgacgtcaacgctagagctggtgttttggaaccacaaggtatggttggtatcaagttccgtagagaaaaattgctg gacaccatgaacagattggatgacaagtacagagaattgagatctcaattatccaacaagagtttggctccagaa gtacatcagcaaatatccaagcaattagctgatcgtgagagagaactattgccaatttacggacaaatcagtcttc aatttgctgatttgcacgataggtcttcacgtatggtggccaagggtgttatttctaaggaactggaatggaccgag gcacgtcgtttcttcttctggagattgagaagaagattgaacgaagaatatttgattaaaaggttgagccatcaggt aggcgaagcatcaagattagaaaagatcgcaagaattagatcgtggtaccctgcttcagtggaccatgaagatg ataggcaagtcgcaacatggattgaagaaaactacaaaactttggacgataaactaaagggtttgaaattagagt cattcgctcaagacttagctaaaaagatcagaagcgaccatgacaatgctattgatggattatctgaagttatcaag atgttatctaccgatgataaagaaaaattgttgaagactttgaaataa
SEQ ID NO:202 Truncated geranylgeranyl pyrophosphate synthase Ag_GPPS_Ntrunc	atgtcttttgacttcaataagtatatggattctaaggctatgaccgtcaacgaggctttgaataaagccatcccattg cgttacccacaaaagatctacgaatctatgagatattctttgttagctggtggtaagagagtccgtccagttttgtgt atcgccgcttgtgaattagtcggtggtactgaggagttagctattccaaccgcctgtgccatcgaaatgatccaca ccatgtctttgatgcacgatgatttgccatgtatcgacaacgatgacttgagacgtggtaaacctaccaatcataag attttcggtgaagatactgctgttactgccggtaacgctttacactcttacgccttcgaacatattgctgtttctacttc caagactgttggtgctgatagaattttgagaatggtttctgaattaggtcgtgctactggttccgaaggtgttatggg tggtcaaatggtcgatattgcttctgaaggtgacccttccattgatttgcaaactttagaatggatccacatccacaa gactgctatgttattagaatgttctgttgtctgtggtgccatcatcggtggtgcttctgaaattgttattgagagagcc agacgttatgctcgttgtgtcggtttattgtttcaagttgttgacgacattttagatgttaccaaatcttctgacgaattg ggtaaaactgctggtaaagatttaatctccgataaagccacctaccctaagttgatgggtttggagaaggccaaa gagttttccgatgaattattaaacagagctaaaggtgaattgtcttgcttcgatccagttaaggctgccccattgtta ggtttggctgactacgttgccttcagacaaaactaa
SEQ ID NO:203 Truncated geranylgeranyl pyrophosphate synthase Ag_GPPS_Ntrunc	MSFDFNKYMDSKAMTVNEALNKAIPLRYPQKIYESMRYSLLAGGKRV RPVLCIAACELVGGTEELAIPTACAIEMIHTMSLMHDDLPCIDNDDLRR GKPTNHKIFGEDTAVTAGNALHSYAFEHIAVSTSKTVGADRILRMVSE LGRATGSEGVMGGQMVDIASEGDPSIDLQTLEWIHIHKTAMLLECSVV CGAIIGGASEIVIERARRYARCVGLLFQVVDDILDVTKSSDELGKTAGK DLISDKATYPKLMGLEKAKEFSDELLNRAKGELSCFDPVKAAPLLGLA DYVAFRQN*
SEQ ID NO:204 Phosphomevalonate kinase Sc_ERG8 Saccharomyces cerevisiae	ttatttatcaagataagtttccggatctttttctttcctaacaccccagtcagcctgagttacatccagccattgaacctt agaaaatcttttgtcatcagcggtttgagccctaagatcaacatcttgcttagcaatcactgcaatggcgtcataac caccagcaccaggtattaagcaagtaagaactccttttaaggtctggcaatcatccaataagctagtttgtacggg aggttcgatatcggcaccagattctttagttatttttctaaaggaacgtctaattgtggcaactgcatctctaacttctg tgatctcaggatactttgacaggtacagtcattcctctcaagagactcaaatatctgatcgctgtaatcgtcatgag tctcgtgtaagcgatctagtttagatagtccatccataaatctagaatttgcatgatcgagttctgtatatattttcaag ctttccggcatatgcgaatcataccaatttttttaccttctggaccagttttactgtttctgaaccattcttaatatcgccc atccataaagttaatcccgaaggtaaatggttacttttaatcgttatattccagtcttcttcattaaccaaatgcgccag tttactgccgtaagtagcacttccaatatctggcaaattagagattaatgcgggtgggaatcttctatatctgatagat ccatatgctgccgccgctacatcaaacccgcttccaattttaccctgagcttgacaatgagcaacttgtgataaatt atgaataacttctctatatttgtctacattattttccaggtccgatacaaaaaaggaggccaaagctgtagttaaaact gtgactaaacctgccgaggagcccagccctgttttgggaacttcttcaattctgtgcgaatgaaaactcaatcttct gttgccacgatgttcggtaacgctgtcctcctgagaatggtaggcatcatcagagaaaatatcaataacgaacaa gtttctattgcagtagtcgtccatgttaggcttaaagtagctaaatacgttagcgataacttttcaatgaaagggttct tagatccgcctatcgaaacaggaatgaagccagttttaggacttatatggtacagccactccccatctttaaattgtt

SEQ ID	SEQUENCE
	tacttttcacacgcacttcaaacttatcagactcttgcaatgaaccgtaaggatgggctacagcatgcattcttgcc gataatccgactacaaatgcttcatatttcggatctaaaactaaatatccaccagctagtaacgctttccctggggc actgaaggctctcaactctgacat
SEQ ID NO:205 Phosphomevalonate kinase Sc_ERG8 Saccharomyces cerevisiae	MSELRAFSAPGKALLAGGYLVLDPKYEAFVVGLSARMHAVAHPYGSL QESDKFEVRVKSKQFKDGEWLYHISPKTGFIPVSIGGSKNPFIEKVIANV FSYFKPNMDDYCNRNLFVIDIFSDDAYHSQEDSVTEHRGNRRLSFHSH RIEEVPKTGLGSSAGLVTVLTTALASFFVSDLENNVDKYREVIHNLSQV AHCQAQGKIGSGFDVAAAAYGSIRYRRFPPALISNLPDIGSATYGSKLA HLVNEEDWNITIKSNHLPSGLTLWMGDIKNGSETVKLVQKVKNWYDS HMPESLKIYTELDHANSRFMDGLSKLDRLHETHDDYSDQIFESLERND CTCQKYPEITEVRDAVATIRRSFRKITKESGADIEPPVQTSLLDDCQTLK GVLTCLIPGAGGYDAIAVIAKQDVDLRAQTADDKRFSKVQWLDVTQA DWGVRKEKDPETYLDK*
SEQ ID NO:206 Mevalonate kinase Erg12 Saccharomyces cerevisiae	atgtcattaccgttcttaacttctgcaccgggaaaggttattatttttggtgaacactctgctgtgtacaacaagcctg ccgtcgctgctagtgtgtctgcgttgagaacctacctgctaataagcgagtcatctgcaccagatactattgaattg gacttcccggacattagctttaatcataagtggtccatcaatgatttcaatgccatcaccgaggatcaagtaaactc ccaaaaattggccaaggctcaacaagccaccgatggcttgtctcaggaactcgttagtcttttggatccgttgttag ctcaactatccgaatccttccactaccatgcagcgttttgtttcctgtatatgtttgtttgcctatgcccccatgccaag aatattaagttttctttaaagtctactttacccatcggtgctgggttgggctcaagcgcctctatttctgtatcactggc cttagctatggcctacttgggggggttaataggatctaatgacttggaaaagctgtcagaaaacgataagcatata gtgaatcaatgggccttcataggtgaaaagtgtattcacggtacccccttcaggaatagataacgctgtggccactt atggtaatgccctgctatttgaaaaagactcacataatggaacaataaacacaaacaatttttaagttcttagatgattt cccagccattccaatgatcctaacctatactagaattccaaggtctacaaaagatcttgttgctcgcgttcgtgtgtt ggtcaccgagaaatttcctgaagttatgaagccaattctagatgccatgggtgaatgtgccctacaaggcttagag atcatgactaagttaagtaaatgtaaaggcaccgatgacgaggctgtagaaactaataatgaactgtatgaacaa ctattggaattgataagaataaatcatggactgcttgtctcaatcggtgtttctcatcctggattagaacttattaaaaa tctgagcgatgatttgagaattggctccacaaaacttaccggtgctggtggcggcggttgctctttgactttgttac gaagagacattactcaagagcaaattgacagtttcaaaaagaaattgcaagatgattttagttacgagacatttgaa acagacttgggtgggactggctgctgtttgttaagcgcaaaaaatttgaataaagatcttaaaatcaaatccctagt attccaattatttgaaaataaaactaccacaaagcaacaaattgacgatctattattgccaggaaacacgaatttac catggacttcataa
SEQ ID NO:207 Mutated acetyl-CoA carboxylase (ACC1) (S659A, S1157A)	MSEESLFESSPQKMEYEITNYSERHTELPGHFIGLNTVDKLEESPLRDFV KSHGGHTVISKILIANNGIAAVKEIRSVRKWAYETFGDDRTVQFVAMA TPEDLEANAEYIRMADQYIEVPGGTNNNNYANVDLIVDIAERADVDA VWAGWGHASENPLLPEKLSQSKRKVIFIGPPGNAMRSLGDKISSTIVAQ SAKVPCIPWSGTGVDTVHVDEKTGLVSVDDDIYQKGCCTSPEDGLQK AKRIGFPVMIKASEGGGGKGIRQVEREEDFIALYHQAANEIPGSPIFIMK LAGRARHLEVQLLADQYGTNISLFGRDCSVQRRHQKIIEEAPVTIAKAE TFHEMEKAAVRLGKLVGYVSAGTVEYLYSHDDGKFYFLELNPRLQVE HPTTEMVSGVNLPAAQLQIAMGIPMHRISDIRTLYGMNPHSASEIDFEF KTQDATKKQRRPIPKGHCTACRITSEDPNDGFKPSGGTLHELNFRSSSN VWGYFSVGNNGNIHSFSDSQFGHIFAFGENRQASRKHMVVALKELSIR GDFRTTVEYLIKLLETEDFEDNTITTGWLDDLITHKMTAEKPDPTLAVI CGAATKAFLASEEARHKYIESLQKGQVLSKDLLQTMFPVDFIHEGKRY KFTVAKSGNDRYTLFINGSKCDIILRQLSDGGLLIAIGGKSHTIYWKEEV AATRLSVDSMTTLLEVENDPTQLRTPSPGKLVKFLVENGEHIIKGQPYA EIEVMKMQMPLVSQENGIVQLLKQPGSTIVAGDIMAIMTLDDPSKVKH ALPFEGMLPDFGSPVIEGTKPAYKFKSLVSTLENILKGYDNQVIMNASL QQLIEVLRNPKLPYSEWKLHISALHSRLPAKLDEQMEELVARSLRRGA VFPARQLSKLIDMAVKNPEYNPDKLLGAVVEPLADIAHKYSNGLEAHE HSIFVHFLEEYYEVEKLFNGPNVREENIILKLRDENPKDLDKVALTVLS HSKVSAKNNLILAILKHYQPLCKLSSKVSAIFSTPLQHIVELESKATAKV ALQAREILIQGALPSVKERTEQIEHILKSSVVKVAYGSSNPKRSEPDLNI

SEQ ID	SEQUENCE
	LKDLIDSNYVVFDVLLQFLTHQDPVVTAAAAQVYIRRAYRAYTIGDIR VHEGVTVPIVEWKFQLPSAAFSTFPTVKSKMGMNRAVSVSDLSYVAN SQSSPLREGILMAVDHLDDVDEILSQSLEVIPRHQSSSNGPAPDRSGSSA SLSNVANVCVASTEGFESEEEILVRLREILDLNKQELINASIRRITFMFGF KDGSYPKYYTFNGPNYNENETIRHIEPALAFQLELGRLSNFNIKPIFTDN RNIHVYEAVSKTSPLDKRFFTRGIIRTGHIRDDISIQEYLTSEANRLMSDI LDNLEVTDTSNSDLNHIFINFIAVFDISPEDVEAAFGGFLERFGKRLLRL RVSSAEIRIIIKDPQTGAPVPLRALINNVSGYVIKTEMYTEVKNAKGEW VFKSLGKPGSMHLRPIATPYPVKEWLQPKRYKAHLMGTTYVYDFPEL FRQASSSQWKNFSADVKLTDDFFISNELIEDENGELTEVEREPGANAIG MVAFKITVKTPEYPRGRQFVVVANDITFKIGSFGPQEDEFFNKVTEYAR KRGIPRIYLAANSGARIGMAEEIVPLFQVAWNDAANPDKGFQYLYLTS EGMETLKKFDKENSVLTERTVINGEERFVIKTIIGSEDGLGVECLRGSG LIAGATSRAYHDIFTITLVTCRSVGIGAYLVRLGQRAIQVEGQPIILTGA PAINKMLGREVYTSNLQLGGTQIMYNNGVSHLTAVDDLAGVEKIVEW MSYVPAKRNMPVPILETKDTWDRPVDFTPTNDETYDVRWMIEGRETE SGFEYGLFDKGSFFETLSGWAKGVVVGRARLGGIPLGVIGVETRTVEN LIPADPANPNSAETLIQEPGQVWHPNSAFKTAQAINDFNNGEQLPMMIL ANWRGFSGGQRDMFNEVLKYGSFIVDALVDYKQPIIIYIPPTGELRGGS WVVVDPTINADQMEMYADVNARAGVLEPQGMVGIKFRREKLLDTM NRLDDKYRELRSQLSNKSLAPEVHQQISKQLADRERELLPIYGQISLQF ADLHDRSSRMVAKGVISKELEWTEARRFFFWRLRRRLNEEYLIKRLSH QVGEASRLEKIARIRSWYPASVDHEDDRQVATWIEENYKTLDDKLKG LKLESFAQDLAKKIRSDHDNAIDGLSEVIKMLSTDDKEKLLKTLK*
SEQ ID NO:208 Truncated 3-hydroxy-3-methyl-glutaryl-CoA reductase (Sc_tHMG1)	MQLVKTEVTKKSFTAPVQKASTPVLTNKTVISGSKVKSLSSAQSSSSGP SSSSEEDDSRDIESLDKKIRPLEELEALLSSGNTKQLKNKEVAALVIHGK LPLYALEKKLGDTTRAVAVRRKALSILAEAPVLASDRLPYKNYDYDR VFGACCENVIGYMPLPVGVIGPLVIDGTSYHIPMATTEGCLVASAMRG CKAINAGGGATTVLTKDGMTRGPVVRFPTLKRSGACKIWLDSEEGQN AIKKAFNSTSRFARLQHIQTCLAGDLLFMRFRTTTGDAMGMNMISKGV EYSLKQMVEEYGWEDMEVVSVSGNYCTDKKPAAINWIEGRGKSVVA EATIPGDVVRKVLKSDVSALVELNIAKNLVGSAMAGSVGGFNAHAAN LVTAVFLALGQDPAQNVESSNCITLMKEVDGDLRISVSMPSIEVGTIGG GTVLEPQGAMLDLLGVRGPHATAPGTNARQLARIVACAVLAGELSLC AALAAGHLVQSHMTHNRKPAEPTKPNNLDATDINRLKDGSVTCIKS*
SEQ ID NO:209 Erg10p: acetoacetyl CoA thiolase [Saccharomyces cerevisiae].	atgtctcagaacgtttacattgtatcgactgccagaaccccaattggttcattccagggttctctatcctccaagaca gcagtggaattgggtgctgttgctttaaaaggcgccttggctaaggttccagaattggatgcatccaaggattttga cgaaattatttttggtaacgttctttctgccaatttgggccaagctccggccagacaagttgctttggctgccggtttg agtaatcatatcgttgcaagcacagttaacaaggtctgtgcatccgctatgaaggcaatcattttgggtgctcaatc catcaaatgtggtaatgctgatgttgtcgtagctggtggttgtgaatctatgactaacgcaccatactacatgccag cagcccgtgcgggtgccaaatttggccaaactgttcttgttgatggtgtcgaaagagatgggttgaacgatgcgt acgatggtctagccatgggtgtacacgcagaaaagtgtgcccgtgattgggatattactagagaacaacaagac aattttgccatcgaatcctaccaaaaatctcaaaaatctcaaaaggaaggtaaattcgacaatgaaattgtacctgtt accattaagggatttagaggtaagcctgatactcaagtcacgaaggacgaggaacctgctagattacacgttgaa aaattgagatctgcaaggactgttttccaaaaagaaaacggtactgttactgccgctaacgcttctccaatcaacg atggtgctgcagccgtcatcttggtttccgaaaaagttttgaaggaaaagaatttgaagcctttggctattatcaaag gttggggtgaggccgctcatcaaccagctgattttacatgggctccatctcttgcagttccaaaggctttgaaacat gctggcatcgaagacatcaattctgttgattactttgaattcaatgaagccttttcggttgtcggtttggtgaacacta agattttgaagctagacccatctaaggttaatgtatatggtggtgctgttgctctaggtcacccattgggttgttctgg tgctagagtggttgttacactgctatccatcttacagcaagaaggaggtaagatcggtgttgccgccatttgtaatg gtggtggtggtgcttcctctattgtcattgaaaagatatga
SEQ ID NO:210 Artificial truncated	atgtcttacgttgtcaagggtatgatctctattgcttgtggtttgttcggtagagaattgtttaacaacagacacttgttc tcttggggtttgatgtggaaagctttcttcgctttggtcccaattttgtctttcaatttcttcgccgccatcatgaaccaa

SEQ ID	SEQUENCE
geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t112	atctacgatgttgatatcgaccgtatcaacaagccagacttacctttagtttccggtgaaatgtccattgaaactgctt ggatcttgtctatcattgttgccttgactggtttaattgttactattaagttgaagtccgctccattgtttgtcttcatctac atcttcggtatcttcgctggtttcgcttactccgtcccacctattagatggaaacaatatccttttaccaatttcttgatc actatttcctctcatgttggtttggctttcacttcttactctgccaccacttctgctttaggtttgcctttcgtttggcgtcc tgccttctctttcattattgctttcatgactgtcatgggtatgactattgcctttgctaaagacatttctgatatcgaaggt gatgctaagtacggtgtctctaccgttgctaccaagttaggtgctagaaatatgacttttgttgtttctggtgtcttatt gttgaactacttggtttctatctctattggtatcatttggccacaagttttcaagtctaacattatgatcttgtctcatgct attttggctttctgtttgatctttcaaactcgtgaattagccttagccaattatgcctctgccccatcccgtcaatttttcg aattcatctggttgttatactatgccgaatacttcgtttacgtcttcatttaa
SEQ ID NO:211 Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t112	MSYVVKGMISIACGLFGRELFNNRHLFSWGLMWKAFFALVPILSFNFF AAIMNQIYDVDIDRINKPDLPLVSGEMSIETAWILSIIVALTGLIVTIKLK SAPLFVFIYIFGIFAGFAYSVPPIRWKQYPFTNFLITISSHVGLAFTSYSAT TSALGLPFVWRPAFSFIIAFMTVMGMTIAFAKDISDIEGDAKYGVSTVA TKLGARNMTFVVSGVLLLNYLVSISIGIIWPQVFKSNIMILSHAILAFCLI FQTRELALANYASAPSRQFFEFIWLLYYAEYFVYVFI
SEQ ID NO:212 Artificial truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t131 nucleotide sequence	atgtctaacaacagacacttgttctcttggggtttgatgtggaaagctttcttcgctttggtcccaattttgtctttcaatt tcttcgccgccatcatgaaccaaatctacgatgttgatatcgaccgtatcaacaagccagacttacctttagtttccg gtgaaatgtccattgaaactgcttggatcttgtctatcattgttgccttgactggtttaattgttactattaagttgaagtc cgctccattgtttgtcttcatctacatcttcggtatcttcgctggtttcgcttactccgtcccacctattagatggaaaca atatccttttaccaatttcttgatcactatttcctctcatgttggtttggctttcacttcttactctgccaccacttctgcttta ggtttgcctttcgtttggcgtcctgccttctctttcattattgctttcatgactgtcatgggtatgactattgcctttgctaa agacatttctgatatcgaaggtgatgctaagtacggtgtctctaccgttgctaccaagttaggtgctagaaatatga ctttgttgtttctggtgtcttattgttgaactacttggtttctatctctattggtatcatttggccacaagttttcaagtcta acattatgatcttgtctcatgctattttggctttctgtttgatctttcaaactcgtgaattagccttagccaattatgcctct gccccatcccgtcaatttttcgaattcatctggttgttatactatgccgaatacttcgtttacgtcttcatttaa
SEQ ID NO:213 Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t131	MSNNRHLFSWGLMWKAFFALVPILSFNFFAAIMNQIYDVDIDRINKPD LPLVSGEMSIETAWILSIIVALTGLIVTIKLKSAPLFVFIYIFGIFAGFAYS VPPIRWKQYPFTNFLITISSHVGLAFTSYSATTSALGLPFVWRPAFSFIIA FMTVMGMTIAFAKDISDIEGDAKYGVSTVATKLGARNMTFVVSGVLL LNYLVSISIGIIWPQVFKSNIMILSHAILAFCLIFQTRELALANYASAPSR QFFEFIWLLYYAEYFVYVFI
SEQ ID NO:214 Artificial truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t142 nucleotide sequence	atgtcttggaaagctttcttcgctttggtcccaattttgtctttcaatttcttcgccgccatcatgaaccaaatctacgat gttgatatcgaccgtatcaacaagccagacttacctttagtttccggtgaaatgtccattgaaactgcttggatcttgt ctatcattgttgccttgactggtttaattgttactattaagttgaagtccgctccattgtttgtcttcatctacatcttcggt atcttcgctggtttcgcttactccgtcccacctattagatggaaacaatatccttttaccaatttcttgatcactatttcct ctcatgttggtttggctttcacttcttactctgccaccacttctgctttaggtttgccttcgtttggcgtcctgccttctct ttcattattgctttcatgactgtcatgggtatgactattgcctttgctaaagacatttctgatatcgaaggtgatgctaa gtacggtgtctctaccgttgctaccaagttaggtgctagaaatatgacttttgttgtttctggtgtcttattgttgaacta cttggtttctatctctattggtatcatttggccacaagttttcaagtctaacattatgatcttgtctcatgctattttggcttt ctgtttgatctttcaaactcgtgaattagccttagccaattatgcctctgccccatcccgtcaatttttcgaattcatctg gttgttatactatgccgaatacttcgtttacgtcttcatttaa
SEQ ID NO:215 Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t142	MSWKAFFALVPILSFNFFAAIMNQIYDVDIDRINKPDLPLVSGEMSIETA WILSIIVALTGLIVTIKLKSAPLFVFIYIFGIFAGFAYSVPPIRWKQYPFTN FLITISSHVGLAFTSYSATTSALGLPFVWRPAFSFIIAFMTVMGMTIAFA KDISDIEGDAKYGVSTVATKLGARNMTFVVSGVLLLNYLVSISIGIIWP QVFKSNIMILSHAILAFCLIFQTRELALANYASAPSRQFFEFIWLLYYAE YFVYVFI

SEQ ID	SEQUENCE
SEQ ID NO:216 Artificial truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t166 nucleotide sequence	Atgtctgatgttgatatcgaccgtatcaacaagccagacttacctttagtttccggtgaaatgtccattgaaactgct tggatcttgtctatcattgttgccttgactggtttaattgttactattaagttgaagtccgctccattgtttgtcttcatcta catcttcggtatcttcgctggtttcgcttactccgtcccacctattagatggaaacaatatccttttaccaatttcttgat cactatttcctctcatgttggtttggctttcacttcttactctgccaccacttctgctttaggtttgcctttcgtttggcgtc ctgccttctctttcattattgctttcatgactgtcatgggtatgactattgcctttgctaaagacatttctgatatcgaag gtgatgctaagtacggtgtctctaccgttgctaccaagttaggtgctagaaatatgacttttgttgtttctggtgtctta ttgttgaactacttggtttctatctctattggtatcatttggccacaagttttcaagtctaacattatgatcttgtctcatgc tattttggctttctgtttgatctttcaaactcgtgaattagccttagccaattatgcctctgccccatcccgtcaatttttc gaattcatctggttgttatactatgccgaatacttcgtttacgtcttcatttaa
SEQ ID NO:217 Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t166	MSDVDIDRINKPDLPLVSGEMSIETAWILSIIVALTGLIVTIKLKSAPLFV FIYIFGIFAGFAYSVPPIRWKQYPFTNFLITISSHVGLAFTSYSATTSALGL PFVWRPAFSFIIAFMTVMGMTIAFAKDISDIEGDAKYGVSTVATKLGAR NMTFVVSGVLLLNYLVSISIGIIWPQVFKSNIMILSHAILAFCLIFQTREL ALANYASAPSRQFFEFIWLLYYAEYFVYVFI
SEQ ID NO:218 Artificial truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t186 nucleotide sequence	atgtctattgaaactgcttggatcttgtctatcattgttgccttgactggtttaattgttactattaagttgaagtccgctc cattgtttgtcttcatctacatcttcggtatcttcgctggtttcgcttactccgtcccacctattagatggaaacaatatc cttttaccaatttcttgatcactatttcctctcatgttggtttggctttcacttcttactctgccaccacttctgctttaggttt gcctttcgtttggcgtcctgccttctctttcattattgctttcatgactgtcatgggtatgactattgcctttgctaaagac atttctgatatcgaaggtgatgctaagtacggtgtctctaccgttgctaccaagttaggtgctagaaatatgacttttg ttgtttctggtgtcttattgttgaactacttggtttctatctctattggtatcatttggccacaagttttcaagtctaacatta tgatcttgtctcatgctattttggctttctgtttgatctttcaaactcgtgaattagccttagccaattatgcctctgcccc atcccgtcaatttttcgaattcatctggttgttatactatgccgaatacttcgtttacgtcttcatttaa
SEQ ID NO:219 Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t186	MSIETAWILSIIVALTGLIVTIKLKSAPLFVFIYIFGIFAGFAYSVPPIRWK QYPFTNFLITISSHVGLAFTSYSATTSALGLPFVWRPAFSFIIAFMTVMG MTIAFAKDISDIEGDAKYGVSTVATKLGARNMTFVVSGVLLLNYLVSI SIGIIWPQVFKSNIMILSHAILAFCLIFQTRELALANYASAPSRQFFEFIWL LYYAEYFVYVFI
SEQ ID NO:220 Artificial geranyl pyrophosphate olivetolic acid geranyltransferase CsGOT (CsPT1) nucleotide sequence	atgggtttatcttccgtttgtacttttttctttccaaactaactaccacactttgttaaatccacacaacaacaaccctaaa acctccttgttatgttacagacacccaaagacccctattaaatactcctacaacaacttcccatccaaacactgctc cactaagtcctttcacttgcaaaacaagtgttctgaatccttgtccattgccaagaactctattcgtgccgctactact aaccaaactgagccacctgaatccgataaccactccgtcgccaccaagatcttgaattttggtaaagcttgctgga aattgcaaagaccatacactattattgctttcacttcctgtgcttgtggtttattcggtaaggaattattgcataacacc aacttgatttcttggtccttaatgttcaaagccttcttcttttagttgccattttatgtattgcttctttcactactactattaa tcaaatttacgatttgcacattgacagaatcaataagcctgacttgccattagcttccggtgaaatttctgttaacact gcttggatcatgtccatcattgtcgctttgttcggtttaattatcaccatcaaaatgaagggtggtcctttgtacatcttc ggttattgcttcggtattttcggtggtattgtctactctgtcccaccattcagatggaagcaaaacccatccactgcct ttttgttgaatttcttggctcacatcattaccaattttactttctactatgcctcccgtgctgctttaggtttgccttttgagt tacgtccatccttcacttttttattggcttttatgaagtccatgggttctgctttagccttaattaaggacgcctctgacg ttgaaggtgatactaagttcggtatctctactttagcctctaagtacggttctcgtaacttgaccttgttctgttctggta ttgtcttgttgtcttacgtcgccgctattttggccggtatcatctggccacaagctttcaactctaacgttatgttgttgt ctcatgctatcttagctttctggttgatcttacaaaccagagacttcgctttgactaactacgacccagaagccggtc gtagattctacgaattcatgtggaaattgtactacgccgagtacttggtctacgttttcatttag
SEQ ID NO:221	atgtctgctggctctgaccaaattgaaggttccccgcatcacgaatcagataatagtattgccacaaagatcttaaa

SEQ ID	SEQUENCE
Artificial truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4t nucleotide sequence	ctttgggcatacatgttggaaattacaaaggccctacgtcgtcaaaggaatgataagcatcgcttgcggtctgttc ggaagggaattatttaacaataggcatctattcagctgggggttaatgtggaaagctttcttcgcgttagtgccaat cctaagctttaacttttcgccgccatcatgaaccagatttatgatgttgatatcgacaggataaataagccagatctt ccattggtatccggtgaaatgtcaatagaaactgcatggatattatctattatcgttgcgctgaccggactgatagta acaatcaaattgaaatctgcacccctgtttgtttttatatatatatttggtattttcgctggattcgcttactcagtgccac ctatcaggtggaagcagtacccattcacgaattttctgatcacgatctctagccacgtcgggttagcgttcacatctt actctgcaaccacgagtgccttggggcttcctttcgtctggcgtccagcttttagttttatcattgcctttatgaccgta atgggaatgacgatcgcattcgcaaaggacatttctgacatagagggggatgcaaaatacggtgtctccactgtg gcgacaaaattaggagctaggaatatgactttcgtggtgtccggtgtattattactaaattatctggtatctataagta tcggcatcatatggccgcaagtgtttaaatccaacattatgatactgagtcatgctattttggctttttgtctgatttttc agacgcgtgagttggcgcttgcaaactatgcctctgcgcccagcaggcagtttttgaattcatatggttattgtact atgccgagtatttcgtctacgtatttatttaa
SEQ ID NO:222 Artificial truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsGOT_t75 (CsPT1_t75) nucleotide sequence	atgtctgccgctactactaaccaaactgagccacctgaatccgataaccactccgtcgccaccaagatcttgaatt ttggtaaagcttgctggaaattgcaaagaccatacactattattgctttcacttcctgtgcttgtggtttattcggtaag gaattattgcataacaccaacttgatttcttggtccttaatgttcaaagccttcttcttttttagttgccattttatgtattgct tctttcactactactattaatcaaatttacgatttgcacattgacagaatcaataagcctgacttgccattagcttccgg tgaaatttctgttaacactgcttggatcatgtccatcattgtcgctttgttcggtttaattatcaccatcaaaatgaagg gtggtcctttgtacatcttcggttattgcttcggtattttcggtggtattgtctactctgtcccaccattcagatggaagc aaaacccatccactgccttttttgttgaatttcttggctcacatcattaccaattttactttctactatgcctcccgtgctg ctttaggtttgccttttgagttacgtccatccttcacttttttattggcttttatgaagtccatgggttctgctttagccttaa ttaaggacgcctctgacgttgaaggtgatactaagttcggtatctctactttagcctctaagtacggttctcgtaactt gaccttgttctgttctggtattgtcttgttgtcttacgtcgccgctattttggccggtatcatctggccacaagctttcaa ctctaacgttatgttgttgtctcatgctatcttagctttctggttgatcttacaaaccagagacttcgctttgactaacta cgacccagaagccggtcgtagattctacgaattcatgtggaaattgtactacgccgagtacttggtctacgttttca tttag
SEQ ID NO:223 Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsGOT_t75 (CsPT1_t75)	MSAATTNQTEPPESDNHSVATKILNFGKACWKLQRPYTIIAFTSCACGL FGKELLHNTNLISWSLMFKAFFFLVAILCIASFTTTINQIYDLHIDRINKP DLPLASGEISVNTAWIMSIIVALFGLIITIKMKGGPLYIFGYCFGIFGGIV YSVPPFRWKQNPSTAFLLNFLAHIITNFTFYYASRAALGLPFELRPSFTF LLAFMKSMGSALALIKDASDVEGDTKFGISTLASKYGSRNLTLFCSGIV LLSYVAAILAGIIWPQAFNSNVMLLSHAILAFWLILQTRDFALTNYDPE AGRRFYEFMWKLYYAEYLVYVFI
SEQ ID NO:224 Artificial truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t76 (CsPT4t) nucleotide sequence	Atgtccgccggttctgatcaaatcgaaggttcccctcatcatgagtccgataactccattgctactaaaatttaaat ttcggtcatacttgttggaagttgcaacgtccttacgttgtcaagggtatgatctctattgcttgtggtttgttcggtag agaattgtttaacaacagacacttgttctcttggggtttgatgtggaaagctttcttcgctttggtcccaattttgtctttc aatttcttcgccgccatcatgaaccaaatctacgatgttgatatcgaccgtatcaacaagccagacttacctttagttt ccggtgaaatgtccattgaaactgcttggatcttgtctatcattgttgccttgactggtttaattgttactattaagttga agtccgctccattgtttgtcttcatctacatcttcggtatcttcgctggtttcgcttactccgtcccacctattagatgga aacaatatccttttaccaatttcttgatcactatttcctctcatgttggtttggctttcacttcttactctgccaccacttct gctttaggtttgcctttcgtttggcgtcctgccttctctttcattattgctttcatgactgtcatgggtatgactattgcctt tgctaaagacatttctgatatcgaaggtgatgctaagtacggtgtctctaccgttgctaccaagttaggtgctagaa atatgacttttgttgtttctggtgtcttattgttgaactacttggtttctatctctattggtatcatttggccacaagttttca agtctaacattatgatcttgtctcatgctattttggctttctgtttgatctttcaaactcgtgaattagccttagccaattat gcctctgccccatcccgtcaatttttcgaattcatctggttgttatactatgccgaatacttcgtttacgtcttcatttaa
SEQ ID NO:225 Geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4	atgggactctcattagtttgtaccttttcatttcaaactaattatcatactttattaaaccctcataataagaatcccaaa aactcattattatcttatcaacaccccaaaacaccaataattaaatcctcttatgataatttttccctctaaatattgcttaa ccaagaactttcatttacttggactcaattcacacaacagaataagctcacaatcaaggtccattagggcaggtag cgatcaaattgaaggttctcctcatcatgaatctgataattcaatagcaactaaaattttaaattttggacatacttgtt ggaaacttcaaagaccatatgtagtaaaagggatgatttcaatcgcttgtggtttgtttgggagagagttgttcaata acagacatttattcagttggggtttgatgtggaaggcattctttgctttggtgcctatattgtccttcaatttctttgcag caatcatgaatcaaatttacgatgtggacatcgacaggataaacaagcctgatctaccactagtttcaggggaaat

2018256863 11 Apr 2024

SEQ ID	SEQUENCE
Cannabis sativa	gtcaattgaaacagcttggattttgagcataattgtggcactaactgggttgatagtaactataaaattgaaatctgc accacttttttgttttcatttacatttttggtatatttgctgggtttgcctattctgttccaccaattagatggaagcaatatc cttttaccaatttttctaattaccatatcgagtcatgtgggcttagctttcacatcatattctgcaaccacatcagctcttg gtttaccatttgtgtggaggcctgcttttagtttcatcatagcattcatgacagttatgggtatgactattgcttttgcca aagatatttcagatattgaaggcgacgccaaatatggggtatcaactgttgcaaccaaattaggtgctaggaacat gacatttgttgtttctggagttcttcttctaaactacttggtttctatatctattgggataatttggcctcaggttttcaaga gtaacataatgatactttctcatgcaatcttagcattttgcttaatcttccagactcgtgagcttgctctagcaaattac gcctcggcgccaagcagacaattcttcgagtttatctggttgctatattatgctgaatactttgtatatgtatttatataa

EXAMPLES

[00924] The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the present disclosure, and are not intended to limit the scope of what the inventors regard as their disclosure nor are they intended to represent that the experiments below are all or the only experiments performed. Efforts have been made to ensure accuracy with respect to numbers used (e.g. amounts, temperature, etc.) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is weight average molecular weight, temperature is in degrees Celsius, and pressure is at or near atmospheric. Standard abbreviations may be used, e.g., bp, base pair(s); kb, kilobase(s); s or sec, second(s); min, minute(s); h or hr, hour(s); aa, amino acid(s); kb, kilobase(s); bp, base pair(s); nt, nucleotide(s); and the like.

General Methods of the Examples

Yeast transformation methods

[00925] Each DNA construct comprising one or more heterologous nucleic acids disclosed herein (e.g., constructs detailed in Table 11) was integrated into Saccharomyces cerevisiae (CEN.PK2) with standard molecular biology techniques in an optimized lithium acetate (LiAc) transformation. Briefly, cells were grown overnight in yeast extract peptone dextrose (YPD) media at 30 °C with shaking (200 rpm), diluted to an OD600 of 0.1 in 100 mL YPD, and grown to an OD600 of 0.6 - 0.8. For each transformation, 5 mL of culture was harvested by centrifugation, washed in 5 mL of sterile water, spun down again, resuspended in 1 mL of 100 mM LiAc, and transferred to a microcentrifuge tube. Cells were spun down (13,000 x g) for 30 seconds, the supernatant was removed, and the cells were resuspended in a transformation mix consisting of 240 µL 50% PEG, 36 µL 1M LiAc, 10 µL boiled salmon sperm DNA, and 74 µL of donor DNA. Following a heat shock at 42 °C for 40 minutes, cells were recovered overnight in YPD media before plating on selective media. DNA integration was confirmed by colony PCR with primers specific to the integrations.

2018256863 11 Apr 2024

Yeast culturing conditions

[00926] Yeast colonies verified to contain the expected DNA assembly comprising one or more heterologous nucleic acids disclosed herein, genetically modified host cells, were picked into 96-well microtiter plates containing 360 µL of YPD (10 g/L yeast extract, 20 g/L Bacto peptone, 20 g/L dextrose (glucose)) and sealed with a breathable film seal. Cells were cultured at 30 °C in a high capacity microtiter plate incubator shaking at 1000 rpm and 80% humidity for 3 days until the cultures reached carbon exhaustion. The growth-saturated cultures were subcultured into fresh plates containing YPGAL and either olivetolic acid or hexanoic acid, or an olivetolic acid derivative or a carboxylic acid other than hexanoic acid (10 g/L yeast extract, 20 g/L Bacto peptone, 20 g/L galactose, 1 g/L glucose and either 1 mM olivetolic acid or 2 mM hexanoic acid, or 1 mM of an olivetolic acid derivative or 2 mM of a carboxylic acid other than hexanoic acid), by taking 14.4 µL from the saturated cultures and diluting into 360 µL of fresh media and sealed with a breathable film seal. Genetically modified host cells in the production media were cultured at 30 °C in a high capacity microtiter plate shaker at 1000 rpm and 80% humidity for an additional 3 days prior to extraction and analysis. Upon completion, 100 µL of whole cell broth was diluted into 900 µL of methanol, sealed with a foil seal, and shaken at 1500 rpm for 60 seconds to extract the cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives. After shaking, the plate was centrifuged at 1000 x g for 60 seconds to remove any solids. After centrifugation, 12 µL of supernatant was transferred to a fresh assay plate containing 228 µL of methanol, sealed with a foil seal, shaken for 60 seconds at 900 rpm, and analyzed by LC-MS.

Analytical methods

[00927] Samples were analyzed by LC-MS mass spectrometer (Agilent 6470) using an Agilent Poroshell 120 Phenyl Hexyl 2.1 x 50 mm, 1.9 µm analytical column with the following gradient (Mobile Phase A: LC-MS grade water with 0.1% formic acid; Mobile Phase B: LC-MS grade acetonitrile with 0.1% formic acid):

Time (minutes)	**%B**
0	40
0.1	40
0.6	60
1	65
1.01	95
2.01	95
2.02	40

2018256863 11 Apr 2024

Time (minutes)	%B
2.5	40

[00928] The mass spectrometer was operated in negative ion multiple reaction monitoring mode. Each cannabinoid, cannabinoid precursor, cannabinoid derivative, or cannabinoid precursor derivative was identified by retention time, determined from an authentic standard, and multiple reaction monitoring (MRM) transition:

Compound Name	Q1 Mass (Da)	Q3 Mass (Da)
CBGA	359.2	341.1
CBGA	359.2	315.2
CBDA	357.2	339.1
CBDA	357.2	245.1
THCA	357.0	313.0

Recovery and purifications

[00929] Whole-cell broth from cultures comprising genetically modified host cells of the disclosure are extracted with a suitable organic solvent to afford cannabinoids, cannabinoid precursors, cannabinoid derivatives, or cannabinoid precursor derivatives. Suitable organic solvents include, but are not limited to, hexane, heptane, ethyl acetate, petroleum ether, and di-ethyl ether, chloroform, and ethyl acetate. The suitable organic solvent, such as hexane, is added to the whole-cell broth from fermentations comprising genetically modified host cells of the disclosure at a 10:1 ratio (10 parts whole-cell broth – 1 part organic solvent) and stirred for 30 minutes. The organic fraction is separated and extracted twice with an equal volume of acidic water (pH 2.5). The organic layer is then separated and dried in a concentrator (rotary evaporator or thin film evaporator under reduced pressure) to obtain crude cannabinoid, cannabinoid precursor, cannabinoid derivative, or cannabinoid precursor derivative crystals. The crude crystals may then be heated to 105 °C for 15 minutes followed by 145 °C for 55 minutes to decarboxylate a crude cannabinoid or cannabinoid derivative. The crude crystalline product is re-dissolved and recrystallized in a suitable solvent (e.g., n-pentane) and filtered through a 1 μm filter to remove any insoluble material. The solvent is then removed e.g. by rotary evaporation, to produce pure crystalline product.

In vitro enzyme assay and cell-free production of cannabinoids or cannabinoid derivatives

[00930] In some embodiments, genetically modified host cells, e.g., genetically modified yeast cells, verified to comprise one or more heterologous nucleic acids encoding a

2018256863 11 Apr 2024

GOT polypeptide that catalyzes production of cannabigerolic acid from geranyl pyrophosphate and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82 or a polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110, are cultured in 96-well microtiter plates containing 360 µL of YPD (10 g/L yeast extract, 20 g/L Bacto peptone, 20 g/L dextrose (glucose)) and sealed with a breathable film seal. Cells are then cultured at 30 °C in a high capacity microtiter plate incubator shaking at 1000 rpm and 80% humidity for 3 days until the cultures reach carbon exhaustion. The growth-saturated cultures are then subcultured into 200 mL of YPGAL media to an OD600 of 0.2 and incubated with shaking for 20 hours at 30 °C. Cells are then harvested by centrifugation at 3000 x g for 5 minutes at 4 °C. Harvested cells are then resuspended in 50 mL buffer (50 mM Tris-HCl, 1 mM EDTA, 0.1 M KCl, pH 7.4, 125 units Benzonase) and then lysed (Emulsiflex C3, Avestin, INC., 60 bar, 10 min). Cells debris is removed by centrifugation (10,000 × g, 10 min, 4 °C). Subsequently, the supernatant is then subjected to ultracentrifugation (150,000 × g, 1 h, 4 °C, Beckman Coulter L-90K, TI-70). The resulting membrane fractions of the GOT polypeptide that catalyzes production of cannabigerolic acid from geranyl pyrophosphate and olivetolic acid in an amount at least ten times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82 or the polypeptide comprising an amino acid sequence having at least 65% (e.g., at least 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, or 100%) sequence identity to SEQ ID NO:100 or SEQ ID NO:110 are then resuspended in 3.3 mL buffer (10 mM Tris-HCl, 10 mM $MgCl_2$, pH 8.0, 10% glycerol) and solubilized with a tissue grinder. Then, 0.02% (v/v) of the respective membrane preparations are then dissolved in reaction buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, pH 8.5) and substrate (500 µM olivetolic acid, 500 µM GPP) to a total volume of 50 µL and incubated for 1 hour at 30 °C. Assays are then extracted by adding two reaction volumes of ethyl acetate followed by vortexing and centrifugation. The organic layer is evaporated for 30 minutes, resuspended in acetonitrile/H_2O/formic acid (80:20:0.05%) and filtered with Ultrafree® -MC columns (0.22 µm pore size, PVDF membrane material). Cannabinoids or cannabinoid derivatives are then detected via LC-MS and/or recovered and purified.

2018256863 11 Apr 2024

Yeast cultivation in a bioreactor

[00931] Single yeast colonies comprising genetically modified host cells disclosed herein are grown in 15 mL of Verduyn medium (originally described by Verduyn et al, Yeast 8(7): 501-17) with 50 mM succinate (pH 5.0) and 2% glucose in a 125 mL flask at 30 °C, with shaking at 200 rpm to an OD600 between 4 to 9. Glycerol is then added to the culture to a concentration of 20% and 1 mL vials of the genetically modified host cell suspension are stored at −80 °C. One to two vials of genetically modified host cells are thawed and grown in Verduyn medium with 50 mM succinate (pH 5.0) and 4% sucrose for 24 hours, then sub-cultured to an OD600 reading of 0.1 in the same media. After 24 hours of growth at 30 °C with shaking, 65 mL of culture is used to inoculate a 1.3-liter fermenter (Eppendorf DASGIP Bioreactor) with 585 mL of Verduyn fermentation media containing 20 g/L galactose supplemented with hexanoic acid (2 mM), a carboxylic acid other than hexanoic acid (2 mM), olivetolic acid (1 mM), or an olivetolic acid derivative (1 mM). A poly-alpha-olefin may be added to the fermenter as an extractive agent. The fermenter is maintained at 30 °C and pH 5.0 with addition of NH_4OH. In an initial batch phase, the fermenter is aerated at 0.5 volume per volume per minute air (VVM) and agitation ramped to maintain 30% dissolved oxygen. After the initial sugar is consumed, the rise in dissolved oxygen triggers feeding of galactose + hexanoic acid (800 g galactose per liter + 9.28 g hexanoic acid per liter) at 10 g galactose per liter per hour in pulses of 10 g galactose per liter doses (alternatively, rather than feeding the genetically modified host cells disclosed herein hexanoic acid, olivetolic acid, an olivetolic acid derivative, or a carboxylic acid other than hexanoic acid is fed to the genetically modified host cells).

[00932] Between pulses, the feed rate is lowered to 5 g galactose per liter per hour. Upon a 10% rise in dissolved oxygen, the feed rate is resumed at 10 g L^{-1} $hour^{-1}$. As genetically modified host cell density increases, dissolved oxygen is allowed to reach 0%, and the pulse dose is increased to 50 g galacose per liter. Oxygen transfer rate is maintained at rates representative of full-scale conditions of 100 mM per liter per hour by adjusting agitation as volume increased. Feed rate is adjusted dynamically to meet demand using an algorithm that alternates between a high feed rate and low feed rate. During the low feed rate, genetically modified host cells should consume galactose and hexanoic acid, or, alternatively, olivetolic acid, an olivetolic acid derivative, or a carboxylic acid other than hexanoic acid, and any overflow metabolites accumulated during the high feed rate. A rise in dissolved oxygen triggers the high feed rate to resume. The length of time spent in the low feed rate reflects the extent to which genetically modified host cells are over- or under-fed in

2018256863 11 Apr 2024

the prior high feed rate pulse; this information is then monitored and used to tune the high feed rate up or down, keeping the low feed rate within a defined range.

[00933] Over time, the feed rate matches sugar and hexanoic acid, or, alternatively, olivetolic acid, an olivetolic acid derivative, or a carboxylic acid other than hexanoic acid, demand from genetically modified host cells. This algorithm ensures minimal net accumulation of fermentation products other than cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives; biomass; and CO_2. In some embodiments, the process continues for 5 to 14 days. In certain such embodiments, accumulated broth is removed daily and assayed for biomass and cannabinoid, cannabinoid derivative, cannabinoid precursor, or cannabinoid precursor derivative concentration. A concentrated solution of $NH_4H_2PO_4$, trace metals and vitamins are added periodically to maintain steady state concentrations.

Example 1 – Synthesis of Olivetolic Acid or Derivatives Thereof

[00934] The cannabinoid pathway is composed of four biosynthetic steps using the precursors hexanoyl-CoA, malonyl-CoA, and geranyl pyrophosphate (FIG. 1, Box 4). *Saccharomyces cerevisiae* has previously been engineered to produce high levels of malonyl-CoA. To increase geranyl pyrophosphate supply, the engineering strategy as outlined in FIG. 1, Box 3, was carried out. In addition, a heterologous nucleic acid encoding an ACC1 polypeptide was overexpressed to increase flux to malonyl-CoA.

[00935] FIG. 1: Diagram illustrating biosynthetic pathways for converting sugar or hexanoate to the cannabinoids Δ^9-THC and CBD

[00936] To date, engineering biosynthesis of the precursor hexanoyl-CoA in *Saccharomyces cerevisiae* has not been described. Strategies for hexanoyl-CoA biosynthesis in *Saccharomyces cerevisiae*, as outlined in FIG. 1, were conceived: Pathway 1a) hexanoyl-CoA synthetase polypeptide from *C. sativa* converts hexanoate to hexanoyl-CoA (FIG. 1, box 1a). A heterologous nucleic acid encoding a hexanoyl-CoA synthetase polypeptide from *C. sativa* was integrated into *S. cerevisiae*. The resulting cells were fed hexanoate to increase hexanoyl-CoA supply (FIG. 2).

[00937] FIG. 2: Intracellular olivetolic acid production using pathway 1a and TKS/OAC polypeptides. Yeast strains yWL004 expressing a heterologous nucleic acid encoding a TKS polypeptide and a heterologous nucleic acid encoding an OAC polypeptide and yWL009 expressing a heterologous nucleic acid encoding a TKS polypeptide, a heterologous nucleic acid encoding an OAC polypeptide, and a heterologous nucleic acid

2018256863 11 Apr 2024

encoding a HCS polypeptide were grown in YPG and either in the absence or presence of 1mM hexanoate for olivetolic acid production. Addition of hexanoate leads to six-fold increase in olivetolic acid production in strain yWL004, indicating an endogenous acyl-CoA ligase activity. Chromosomal integration of the heterologous nucleic acid encoding a hexanoyl-CoA synthetase polypeptide leading to strain yWL009 shows an additional two-fold increase in olivetolic acid production due to increased hexanoyl-CoA supply.

[00938] Biosynthesis of hexanoyl-CoA from fermentable sugars was increased by integrating pathway 1b, which comprises four enzymes encoding a reverse β-oxidation pathway that has been optimized in *E. coli* for production of hexanol (FIG. 1, box 1b) (Machado et al. 2012). LC-MS analysis confirmed that this pathway was also functional in *S. cerevisiae* and its activity is comparable to pathway 1a, since olivetolic acid yields were similar between the engineered strains yWL009 and yWL0013 (FIG. 3).

[00939] FIG. 3: Intracellular olivetolic acid production comparing pathway 1a and 1b. Yeast strain yWL009 expressing a heterologous nucleic acid encoding a TKS polypeptide, a heterologous nucleic acid encoding an OAC polypeptide, and a heterologous nucleic acid encoding a HCS polypeptide was grown in producing conditions (YPG with 1mM hexanoate) and compared to yWL013 expressing a heterologous nucleic acid encoding a TKS polypeptide, a heterologous nucleic acid encoding a OAC polypeptide, and the hexanoyl-CoA supply pathway 1b grown in non-producing (YPD) as well as producing (YPG) conditions. Chromosomal integration of the hexanoyl-CoA pathway yields similar levels in olivetolic acid production as strain yWL009 when grown in YPG with 1 mM hexanoate.

Table 2: List of strains used in this study

yWL004	Cen.PK2, ACC1::TKS-OAC, tHMGR::MvaE/S,
yWL009	Cen.PK2, ACC1::TKS-OAC, tHMGR::MvaE/S, URA3::HCS
yWL0013	CenPK2, ACC1::TKS-OAC, URA3::HexCoA

[00940] Codon optimized genes were synthesized and used in this study for the polypeptides listed in Table 3.

Table 3: List of polypeptides used in this study

Polypeptide	**Function**	**Original host**
BktB	β-ketothiolase	*Ralstonia eutropha*

2018256863 11 Apr 2024

Polypeptide	Function	Original host
PaaH1	3-Hydroxyacyl-CoA dehydrogenase	*R. eutropha*
Crt	Crotonase	*Clostridium acetobutylicum*
Ter	Trans-2-enoyl-CoA reductase	*Treponema denticola*
HCS	Hexanoyl-CoA synthetase	*Cannabis sativa*
ERG10	Acetyl-CoA acetyltransferase	*Saccharomyces cerevisiae*
ERG13	HMG-CoA synthase	*S. cerevisiae*
tHMG1	HMG-CoA reductase	*S. cerevisiae*
ERG12	Mevalonate kinase	*S. cerevisiae*
IDI1	Isopentenyl diphosphate:dimethylallyl diphosphate isomerase	*S. cerevisiae*
ERG20	Farnesylpyrophosphate synthetase	*S. cerevisiae*
MvaE	acetyl-CoA acetyltransferase/HMG-CoA reductase	*Escherichia coli*
MvaS	HMG-CoA synthase	*E. coli*
TKS	Tetraketide Synthase (Type III PKS)	*C. sativa*
OAC	Olivetolic acid cyclase	*C. sativa*
GOT	geranyl pyrophosphate:olivetolate geranyltransferase	*C. sativa*
Δ^9-THCAS	Δ^9-tetrahyrdocannabinoidic acid synthase	*C. sativa*
CBDAS	cannabidiolic acid synthase	*C. sativa*
DXS	1-deoxy-D-xylulose-5-phosphate synthase gene	*E. coli*
IspC	1-deoxy-D-xylulose 5-phosphate reductoisomerase	*E. coli*
IspD	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	*E. coli*
IspE	4-diphosphocytidyl-2-C-methylerythritol kinase	*E. coli*
IspF	2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	*E. coli*
IspG	4-hydroxy-3-methylbut-2-en-1-yl diphosphate	*E. coli*

2018256863 11 Apr 2024

Polypeptide	Function	Original host
	synthase	
IspH	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	*E. coli*
IDI	Isopentenyl diphosphate (IPP) isomerase	*E. coli*
IspA*	mutated FPP synthase (S81F) for GPP production	*E. coli*
AflA	Hexanoyl-CoA synthase, subunit A	*Aspergillus parasiticus*
AflB	Hexanoyl-CoA synthase, subunit B	*A. parasiticus*
SCFA-TE	Short chain fatty acyl-CoA Thioesterase	Various microbes

Example 2 – Synthesis of Olivetolic Acid or Derivatives Thereof or Cannabinoids or Derivatives Thereof

[00941] Multiple polypeptides in pathway 1b require NADH as a co-factor. In order to maximize flux through pathway 1b, other biosynthetic pathways that compete for NADH supply are modified (FIG. 1, Box 2). One target can be the ethanol pathway, mediated by various alcohol dehydrogenase polypeptides, but may also include other pathways that consume NADH, such as the glycerol biosynthesis pathway.

[00942] Another route conceived towards hexanoyl-CoA is described in pathway 1c: The alfatoxin biosynthetic gene cluster (iterative type I PKS) encodes a fatty acid synthase-based mechanism (FasA and FasB) for production of hexanoyl-CoA. In some embodiments, a heterologous nucleic acid encoding a thioesterase polypeptide and a heterologous nucleic acid encoding a CoA ligase polypeptide similar to a C6-tolerant thioesterase polypeptide (see BMC Biochem. 2011 Aug 10;12:44. doi: 10.1186/1471-2091-12-44) and a heterologous nucleic acid encoding a HCS polypeptide are expressed to facilitate release of hexanoyl-ACP and activate free hexanoate to its acyl-CoA compound. Additionally, various type II PKS biosynthetic pathways (e.g. benastatin, R1128) contain a FabH-like KSIII (e.g. BenQ, ZhuH), AT and ACP component, which are crucial for providing and selecting the rare hexanoate PKS starter unit. Lastly, the type I PKS pathway for reveromycin biosynthesis encodes the fatty acyl-CoA ligase RevS polypeptide and the FabH-like KASIII component RevR polypeptide, which are suggested to provide hexanoyl-CoA via fatty acid degradation as well as de novo fatty acid biosynthesis.

2018256863 11 Apr 2024

[00943] To avoid competitive consumption of hexanoyl-CoA via β-oxidation, the fatty acid degradation pathway is engineered to have lowered activity. Alternatively, yeast are grown in presence of oleic acid to avoid competition for fatty acids as energy source.

[00944] The pathway of four genes encoding the NADH pathway for hexanoyl-CoA production, including polypeptides PaaH1, Crt, Ter, and BktB, was constructed under the control of Gal1, Gal10, Gal7, and TEF2 promoters, respectively. FIG. 4. The whole cassette was inserted between the upstream and downstream homology region of ADE2 and was integrated into the genome of *S. cerevisiae* using CRISPR/Cas9 to generate yXL001 (using Construct 1/pXL044 as shown in FIG. 4). The pathway of four genes encoding the NADPH pathway (including PhaB, PhaJ, Ter, and BktB polypeptides) was introduced into to *S. cerevisiae* in the same way to generate yXL002 (using Construct 1/pXL072 as shown in FIG. 4). The MCT1 gene under the control of Gal1 promoter flanked by the 1622b homology region (Construct 2; FIG. 4) was introduced into the genome of yXL001 and yXL002 using CRISPR/Cas9 to generate yXL003 and yXL004 (FIG. 4).

[00945] A cassette encoding TKS and OAC genes under the control of Gal1 and Gal10 promoters flanked by ACC1 homology region (Construct 4; FIG. 5) was introduced into the genome of yXL003 and yXL004 using CRISPR/Cas9 to generate yXL005 and yXL006. A heterologous nucleic acid encoding a TKS-OAC fusion polypeptide under the control of a Gal1 promoter (Construct 5; FIG. 5) was introduced into yXL003 and yXL004 to generate yXL007 and yXL008. The resulting strains were inoculated into 10 mL YP medium supplemented with 2% dextrose. After an overnight culture at 30 °C and centrifugation at 3,000 × g for 5 mins, the pellet was resuspended into YP medium supplemented with 2% galactose. After two days expression, the culture supernatant was extracted with equal volume of ethyl acetate, and, after evaporation and filtration, the samples were analyzed by LC-MS, which showed the production of a significant amount of olivetolic acid (FIG. 9 and FIG. 10).

[00946] CsAAE (Construct 3; FIG. 4), TKS, and OAC genes (Construct 4; FIG. 5) were introduced into the genome of *S. cerevisiae* using CRISPR/Cas9 to generate yXL009, which can produce higher level of olivetolic acid in the presence of exogenously supplied hexanoate (FIG. 11).

[00947] In addition, by supplementing the growth medium with various aliphatic acids, from C_4-C_{10}, various olivetolic acid derivatives can be produced from yXL009 (FIG. 11 and FIG. 12). Some of the olivetolic acid derivatives can be further modified by biological or chemical means to covalently attach to other compounds. For example, click

2018256863 11 Apr 2024

chemistry can be performed on the olivetolic derivative containing alkyne functional group. The olivetolic derivative is dissolved in biology grade dimethyl sulfoxide (DMSO) and treated with a DMSO solution of crosslinker containing an azide group (1.0 equiv.), TBTA (DMSO: tBuOH 1:1), $CuSO_4$ $5H_2O$, sodium ascorbate and HEPES-KOH pH: 7.0 (final HEPES-KOH ≈250 mM). The reaction is placed on a water bath at 37 °C for 12 to 16 hours. Liquid chromatograph-mass spectrometry (LC-MS) analysis of the reaction mixture shows reaction completion after 16 hours to obtain the further modified olivetolic acid.

[00948] The GPPS large subunit (GPPSlsu) and small subunit (GPPSssu) genes from *Cannabis sativa* under the control of Gal1 and Gal10 promoters flanked by ADE1 homology region (Construct 10; FIG. 7) were introduced into yXL008 and yXL009 to generate yXL010 and yXL011. A cassette encoding a NphB polypeptide and a THCAS polypeptide under the control of Gal1 and Gal10 promoters flanked by 1014a homology region (Construct 12; FIG. 8) was introduced into the genome of yXL010 and yXL011 to generate yXL012 and yXL013 using CRISPR/Cas9. The resulting strains were inoculated into 10 mL YP medium supplemented with 2% dextrose. After an overnight culture at 30 °C and centrifugation at 3,000 × g for 5 mins, the pellet was resuspended into YP medium supplemented with 2% galactose. After two days expression, the culture supernatant was extracted with equal volume of ethyl acetate, and, after evaporation and filtration, the samples were analyzed by LC-MS, which showed that the overexpression of NphB in yXL010 resulted in the production of cannabigerolic acid (FIGS. 14 and 15). In the presence of a THCAS polypeptide, the cannabigerolic acid was transformed into THCA or into THC. With yXL013, C_4-C_{10} acids were added to the expression medium, resulting in the production of cannabigerolic acid derivatives, which were then modified by a THCAS polypeptide to produce THCA or THC derivatives. Those derivatives can then be further modified by chemical reactions (FIG. 13).

Example 3 – Synthesis of Cannabinoid Precursors, Cannabinoids, or Derivatives of the Foregoing

[00949] To recreate cannabinoid production in microorganisms, chassis *S. cerevisiae* strains were developed containing metabolic pathways for the production of (1) GPP through the mevalonate (Mva) pathway, (2), olivetolic acid or derivatives, (3) CBGA or derivatives, and (4) different cannabinoids or cannabinoid derivatives produced by cannabinoid synthase polypeptides.

2018256863 11 Apr 2024

Production of GPP

[00950] A GPP-overproducing strain, GTY23, was produced by overexpressing Mva pathway genes and introducing a repressible promoter on ERG9. A previously described ERG20 F96W-N127W mutant, ERG20mut, was added to provide a source of GPP precursor in the cell (FIG. 16). This strain was used to screen GOT polypeptide candidates.

Production of Olivetolic Acid or Derivatives Thereof

[00951] Olivetolic acid was produced from sugar by introducing genes CsTKS and CsOAC, and pathways to produce hexanoyl-CoA. Pathways for the production of hexanoate and hexanoyl-CoA are known in the art (e.g., Gajewski et al, "Engineering fungal de novo fatty acid synthesis for short chain fatty acid production," Nature Communications 2017). To produce olivetolic acid or its derivatives, rather than using hexanoyl-CoA pathways, a previously reported acyl-CoA ligase polypeptide, such as a CsAAE1 or CsAAE3 polypeptide, was introduced and exogenously fed cells hexanoate or a carboxylic acid other than hexanoate (FIGS. 17-19). These pathways allow for the production of non-naturally occurring cannabinoids.

Production of CBGA

[00952] The mother cannabinoid CBGA, or derivatives thereof, was produced by a GOT polypeptide. A *C. sativa* GOT polypeptide was identified in the 1990s, yet no report was identified describing reconstituting GOT polypeptide activity *in vivo*. Twenty-five polypeptide variants were screened for *in vivo* production of CBGA in strains containing GPP pathways and exogenously fed olivetolic acid. These genes were all chromosomally integrated driven by GAL1 promoters and screened for activity in yeast extract peptone galactose (YPG) media. GC-MS and LC-MS analysis demonstrated *in vivo* production of CBGA from a CsPT4t polypeptide (FIGS 26A-C). The gene sequence of the CsPT4t polypeptide is referred to as a GOT polypeptide (FIG. 20). yL444 was the strain used in the production of CBGA and expresses the following genotype: CEN.PK2-1D {1114a::GAL1p-CsPT4t-TDH1t; 308a::GAL1p-ERG20(F96W-N127W)-TDH1t; erg9::KanMX_CTR3p-ERG9; leu2-3,112::His3MX6_GAL1p-ERG19/GAL10p-ERG8; ura3-52::ura3/GAL1p-MvaS(A110G)/GAL10p-MvaE; his3_1::hphMX4_GAL1p-ERG12/GAL10p-IDI1; MATa} (FIGS. 6 and 20). LC-MS was carried out as follows (FIGS 26A-C):

Column info: 2015 Kinetex XB-C18 2.1x100 mm RES6 method 10.6 min

Method info:

0-5.6 mins, 45%-73%B, 0.2 mL/min

5.6-6.2 mins, 73%-97%B, 0.2 mL/min

2018256863 11 Apr 2024

6.2-11.3 mins, 97%B, 0.3 mL/min

11.3-12.7, 97-45%B, 0.3 mL/min

12.7-15.5, 45%B, 0.3 mL/min

A: H2O+0.05% TFA

Production of THCA and CBDA

[00953] Cannabinoid synthase genes have been identified from the *Cannabis* genome (including but not limited to THCA synthase (THCAS), CBDA synthase (CBDAS), JP450547, JP454863, JP471546, JP452622). To produce THCA and CBDA, the corresponding THCA synthase and CBDA synthase, respectively, were introduced into a strain producing CBGA containing a heterologous nucleic acid encoding a CsPT4t polypeptide. The synthases were introduced as N-terminal truncated polypeptides with polypeptide tags, e.g., ProA signal sequence (MIFDGTTMSIAIGLLSTLGIGAEA, from proteinase A with UniProt accession number F2QUG8) attached and the transcription of both synthases were under the control of GAL10 promoter. The final plasmid constructs were named as pESC-ProA-THCAS and pESC-ProA-CBDAS. Both plasmids were transformed individually into the above-mentioned strain, which has high CBGA production in the presence of olivetolic acid, to give strains yXL046 and yXL047 (FIGS. 21-25).

[00954] After confirming the transformation by PCR of THCAS or CBDAS, two colonies from each culture were inoculated into a defined medium (SC-Leu + 2% Dextrose) and were incubated at 30 °C with shaking at 800 RPM. After two-day growth, the cultures were back-diluted 1:50 into inducing medium (SC-Leu + 2% galactose + 1 mM olivetolic acid + $CuSO_4$) and incubated at 30 °C with shaking at 800 RPM for 4 days. After 4-day incubation, equal volume of ethyl acetate was added to the expression cultures and the mixtures were subjected to three rounds of bead beating. Then the mixtures were then spun down at 5000 RPM and the organic layers were sent for LC-MS analysis, which showed the production of THCA and CBGA from the corresponding cultures (FIGS. 27 and 28).

Example 4 – Generation of a Base Yeast Strain Capable of High Flux to CBGA with Olivetolic Acid Feeding

[00955] CBGA production strains were created from wild-type *Saccharomyces cerevisiae* strain (CEN.PK2) by expressing genes of the mevalonate pathway polypeptides and a GOT polypeptide under control of the GAL1 or GAL10 promoter. The S21 strain comprised the following chromosomally integrated mevalonate pathway genes from *S. cerevisiae*: ERG10, ERG13, truncated HMG1 (tHMGR), ERG12, ERG8, ERG19, and IDI1.

2018256863 11 Apr 2024

The S21 strain additionally comprised the chromosomally integrated pyruvate decarboxylase (PDC) from *Zymomonas mobilis* to increase flux from pyruvate towards acetyl-CoA.

[00956] To generate additionally strains, a mutant form of ERG20, ERG20mut, which preferentially generates GPP was added to the S21 strain with the following chromosomally integrated GOTs from *C. sativa*: CsPT1 (S164), a truncated CsPT1 (CsPT1_t75, S165), or CsPT4 (S29). Constructs used in S29, S164, and S165 are shown in Table 11.

[00957] Yeast colonies verified to contain the expected DNA assembly comprising one or more heterologous nucleic acids disclosed herein were picked into 96-well microtiter plates containing 360 µL of YPD (10 g/L yeast extract, 20 g/L Bacto peptone, 20 g/L dextrose (glucose)) and sealed with a breathable film seal. Cells were cultured at 30 °C in a high capacity microtiter plate incubator shaking at 1000 rpm and 80% humidity for 3 days until the cultures reached carbon exhaustion. The growth-saturated cultures were subcultured into fresh plates containing YPGAL and olivetolic acid (10 g/L yeast extract, 20 g/L Bacto peptone, 20 g/L galactose, 1 g/L glucose and 1 mM olivetolic acid) by taking 14.4 µL from the saturated cultures and diluting into 360 µL of fresh media and sealed with a breathable film seal. Genetically modified host cells in the production media were cultured at 30 °C in a high capacity microtiter plate shaker at 1000 rpm and 80% humidity for an additional 3 days prior to extraction and analysis. Upon completion, 100 µL of whole cell broth was diluted into 900 µL of methanol, sealed with a foil seal, and shaken at 1500 rpm for 60 seconds to extract the cannabinoids. After shaking, the plate was centrifuged at 1000 x g for 60 seconds to remove any solids. After centrifugation, 12 µL of supernatant was transferred to a fresh assay plate containing 228 µL of methanol, sealed with a foil seal, shaken for 60 seconds at 900 rpm, and analyzed by LC-MS.

[00958] Samples were analyzed by LC-MS mass spectrometer (Agilent 6470) using an Agilent Poroshell 120 Phenyl Hexyl 2.1 x 50 mm, 1.9 µm analytical column with the following gradient (Mobile Phase A: LC-MS grade water with 0.1% formic acid; Mobile Phase B: LC-MS grade acetonitrile with 0.1% formic acid):

Time (minutes)	**%B**
0	40
0.1	40
0.6	60
1	65
1.01	95
2.01	95
2.02	40
2.5	40

2018256863 11 Apr 2024

[00959] The mass spectrometer was operated in negative ion multiple reaction monitoring mode. Each cannabinoid was identified by retention time, determined from an authentic standard, and MRM transition (see FIGS. 77 and 78):

Compound Name	Q1 Mass (Da)	Q3 Mass (Da)
CBGA	359.2	341.1
CBGA	359.2	315.2

[00960] CsPT1 polypeptide and CsPT1_t75 polypeptide produced equivalent amounts of CBGA *in vivo* (1.3 mg/L CBGA). However, CsPT4 polypeptide produced 216 mg/L CBGA *in vivo* (see FIGS 79-82).

Example 5 – Determining the Minimal Catalytic Domain of CsPT4

[00961] To determine the minimal catalytic domain of CsPT4 polypeptide required for the conversion of GPP and olivetolic acid to CBGA, multiple N-terminal truncations of the CsPT4 polypeptide were generated (see Table 11) and expressed *in vivo* in the S21 strain with feeding of 1 mM olivetolic acid. Only full length CsPT4 polypeptide and CsPT4_t76 polypeptide (CsPT4t) displayed activity *in vivo* (Table 4).

Table 4: Screening of CsPT4 truncated polypeptides

CsPT4 construct	Strain	Peak intensity
CsPT4	S29	8901
CsPT4_t76	S147	6859
CsPT4_t112	S166	19
CsPT4_t131	S167	24
CsPT4_t142	S168	20
CsPT4_t166	S169	21
CsPT4_t186	S170	29

Example 6 – Generation of a Base Yeast Strain Capable of High Flux to CBGA with Hexanoic Acid (Caproic Acid) Feeding

[00962] To convert the high flux strain for the production of CBGA with olivetolic acid (S29) to a high flux stain for the production of CBGA with hexanoic acid, genes responsible for the production to olivetolic acid from fatty acids were expressed using the GAL1 or GAL10 promoter in S29. The strain comprised the following chromosomally integrated olivetolic acid pathway genes from *C. sativa*: three copies of TKS and three copies of OAC. Three different strains were generated with two copies of *C. sativa* AAE1

2018256863 11 Apr 2024

(S78), two copies of *C. sativa* AAE3 (S81), or two copies of *S. cerevisiae* FAA2 (S83) (see Table 11 for information on the strains). The strains were grown and tested as in Examples 4 and 5 but with 2 mM hexanoic acid added to the media instead of 1 mM olivetolic acid. Production of CBGA by the strains was observed (Table 5).

Table 5: Generation of CBGA

		Titer (mg/L)		
Feed compound	**Product**	**AAE1v1 (S78)**	**AAE3-Ctrunc (S81)**	**FAA2 (S83)**
Hexanoic acid	CBGA	38.5	32.1	35.1

Example 7 – Generation of a Base Yeast Strain Capable of High Flux to CBDA and THCA

[00963] To convert the high flux strain for the production of CBGA to a high flux strain for the production of CBDA or THCA, a heterologous nucleic acid encoding CBDA synthase polypeptide (S34) or THCA synthase polypeptide (S123) was added to Strain S29 (see Table 11 for information on the strains). The strains were tested as in Examples 4 and 5 with 1 mM olivetolic acid in the media. CBDA and THCA were produced by the strains, as shown in FIGS. 83 and 84.

Example 8 – Feeding of Cannabinoid Precursor Derivatives to Yeast to Produce Rare and Non-Naturally Occurring CBGA Derivatives

[00964] Strains from Example 6 (S78, S81, and S83) were grown as in Examples 4 and 5 but with 2 mM of a carboxylic acid (detailed in Table 6) added to the media and analyzed as in Example 4. Table 6 details the products produced by the strains (product peak intensity).

Table 6: CBGA Derivatives Produced

Feed compound	**Product (IUPAC name)**	**Transition 1**	**Transition 2**	**AAE1v1 (S78)**	**AAE3-Ctrunc (S81)**	**FAA2 (S83)**
2-methyl hexanoic acid	3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-6-(hexan-2-yl)-2,4-dihydroxybenzoic acid	373 --> 355	373 --> 329	1136	1255	1301

4-methyl hexanoic acid	3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-(3-methylpentyl)benzoic acid	373 --> 355	373 --> 329	82453	91493	82517
5-methyl hexanoic acid	3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-(4-methylpentyl)benzoic acid	373 --> 355	373 --> 329	76145	77270	77145
2-hexenoic acid	3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-[(1E)-pent-1-en-1-yl]benzoic acid	357 --> 339	357 --> 313	311	536	588
3-hexenoic acid	3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-[(2E)-pent-2-en-1-yl]benzoic acid	357 --> 339	357 --> 313	90422	104366	112440
5-hexenoic acid	3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-(pent-4-en-1-yl)benzoic acid	357 --> 339	357 --> 313	302499	325854	365798
butanoic acid	3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-propylbenzoic acid	331 --> 313	331 --> 287	92181	106229	103368
pentanoic acid	3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-butylbenzoic acid	345 --> 327	345 --> 301	224003	232206	236366
heptanoic acid	3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-hexylbenzoic acid	373 --> 355	373 --> 329	66544	67766	66570
octanoic acid	3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-heptylbenzoic acid	387 --> 369	387 --> 343	4225	3212	3603
5-chloro pentanoic acid	6-(4-chlorobutyl)-3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxybenzoic acid	379 -->361	379 --> 335	1023	947	902

2018256863 11 Apr 2024

5-(methyl sulfanyl) pentanoic acid	3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-[4-(methylsulfanyl) butyl]benzoic acid	391 --> 373	391 --> 347	18396	18704	19412

Example 9 – Feeding of Cannabinoid Precursor Derivatives to Yeast to Produce Rare and Non-Naturally Occurring CBDA Derivatives

[00965] Strains with (S34) or without a CBDA synthase polypeptide (S29) were tested as in Examples 4 and 5 with 1 mM of an olivetolic acid derivative (detailed in Table 7). Table 7 details the products produced by the strains.

Table 7: CBDA Derivatives Produced

Feed compound	**Product (IUPAC name)**	**Transition 1**	**Transition 2**	**CBGA derivative titer (mg/L) (S29)**	**CBDA derivative titer (mg/L) (S34)**
Orsellinic Acid	3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-methylbenzoic acid	303 --> 285	303 --> 259	1.86	1.05
Divarinic Acid	3-[(2E)-3,7-dimethylocta-2,6-dien-1-yl]-2,4-dihydroxy-6-propylbenzoic acid	331 --> 313	331 --> 287	29.54	3.06

Example 10 – Feeding of Cannabinoid Precursors to Yeast to Produce CBDA or CBGA

[00966] As there are numerous ways to produce cannabinoid precursors (e.g., GPP), a number of different genes were tested *in vivo* to optimize cannabinoid production. Different GPP synthase polypeptides, CBDA synthase polypeptides, TKS polypeptides, OAC polypeptides, medium and long chain fatty acyl-CoA synthetase polypeptides were tested in various combinations as described below (see Table 11 for information on the strains).

[00967] Strains were constructed with different GPP synthase polypeptides to identify the best producer of GPP for production of CBGA when fed 1 mM olivetolic acid. Strain S21 was transformed with heterologous nucleic acids encoding a CsPT4 polypeptide and a GPP synthase polypeptide. CBGA titer was measured as described in Example 4. CBGA titers, titer standard deviations (SD) and number of replicates tested are indicated in Table 8.

2018256863 11 Apr 2024

Table 8: Production of CBGA

Product	Strain	Titer (mg/L)	SD	n
CBGA	S29	215.6	12.2	8
CBGA	S114	6.8	0.7	3
CBGA	S116	15.5	2.0	4
CBGA	S108	8.5	1.7	4
CBGA	S112	9.9	1.6	4
CBGA	S104	10.2	1.6	3
CBGA	S115	9.2	1.9	4
CBGA	S118	5.1	NA	1

[00968] To optimize production of CBDA, strain S29 was transformed with a series of constructs with two copies of a CBDA synthase polypeptide encoding heterologous nucleic acid and grown as in Examples 4 and 5 with 1 mM olivetolic acid. CBDA was measured as described in Example 4. CBDA peak intensity, peak intensity standard deviations (SD) and number of replicates tested are indicated in Table 9.

Table 9: Production of CBDA

Product	Strain	Peak Area	SD	n
CBDA	S34	1651	329	4
CBDA	S35	831	72	4
CBDA	S37	505	26	4
CBDA	S38	658	31	4
CBDA	S39	1274	85	4
CBDA	S41	2129	462	4
CBDA	S42	72	4	4
CBDA	S43	419	481	4
CBDA	S44	758	68	4
CBDA	S45	1253	177	4
CBDA	S46	670	112	4
CBDA	S47	300	15	4

[00969] To optimize production of CBGA from hexanoic acid, different combinations of TKS polypeptide, OAC polypeptide, and medium and long chain fatty acyl-CoA synthetase polypeptide were tested *in vivo*. All strains were daughters or granddaughters of strain S29. All strains were tested as described in Example 4 with 2 mM hexanoic acid added to the production media. CBGA titers, titer standard deviations (SD) and number of replicates tested are indicated in Table 10.

Table 10: Production of CBGA

Product	Strain	Titer (mg/L)	SD	n
CBGA	S31	53.6	12.2	8

Product	Strain	Titer (mg/L)	SD	n
CBGA	S49	55.7	9.3	8
CBGA	S50	22.9	7.0	8
CBGA	S90	67.5	2.8	4
CBGA	S91	63.5	4.2	4
CBGA	S78	38.5	2.5	4
CBGA	S80	37.5	1.8	4
CBGA	S81	32.1	5.8	4
CBGA	S82	35.1	7.0	4
CBGA	S83	35.1	2.6	4
CBGA	S84	36.4	3.5	4
CBGA	S85	34.4	4.3	4
CBGA	S86	36.6	1.8	4
CBGA	S87	32.2	4.9	4
CBGA	S88	40.9	1.4	4
CBGA	S89	39.3	2.7	4
CBGA	S94	59.6	7.9	8
CBGA	S95	58.5	9.2	8
CBGA	S97	72.9	5.5	8

Table 11: Constructs and strains used in the Examples

Strain (Constructs)	Parent Strain*	Polypeptide SEQ ID NOs (Nucleotide SEQ ID NOs)
S21 (FIGS. 29A and 29B)		Sc_tHMG1: SEQ ID NO:208 (SEQ ID NO:119) Sc_ERG13: SEQ ID NO:115 (SEQ ID NO:120) Sc_ERG10: SEQ ID NO:25 (SEQ ID NO:209) Sc_MVD1 (Sc_ERG19): SEQ ID NO:66 (SEQ ID NO:65) Sc_IDI1: SEQ ID NO:58 (SEQ ID NO:57) Zm_PDC: SEQ ID NO:117 (SEQ ID NO:118) Sc_ERG8: SEQ ID NO:205 (SEQ ID NO:204) Sc_ERG12: SEQ ID NO:64 (SEQ ID NO:206)
S29 (FIG. 86)	S21	Cs_PT4: SEQ ID NO:110 (SEQ ID NO:111) Sc_ERG20_mut: SEQ ID NO:60 (SEQ ID NO:161)
S31 (FIGS. 30A, 30B, and 30C)	S29	Cs_OAC: SEQ ID NO:10 (SEQ ID NO:163) Cs_TKS: SEQ ID NO:11 (SEQ ID NO:162) Cs_AAE1_v1: SEQ ID NO:90 (SEQ ID NO:164) Sc_FAA2: SEQ ID NO:169 (SEQ ID NO:168)
S34 (FIG. 85)	S29	Cs_CBDAS_co1: SEQ ID NO:88 (SEQ ID NO:167)
S35 (FIG. 31)	S29	Cs_CBDAS_t28: SEQ ID NO:151 (SEQ ID NO:152)

2018256863 11 Apr 2024

Strain (Constructs)	Parent Strain*	Polypeptide SEQ ID NOs (Nucleotide SEQ ID NOs)
S37 (FIG. 32)	S29	MBP_co1: SEQ ID NO:108 (SEQ ID NO:170) Cs_CBDAS_t28: SEQ ID NO:151 (SEQ ID NO:152) GS12: SEQ ID NO:172 (SEQ ID NO:171)
S38 (FIG. 33)	S29	Cs_CBDAS_t28: SEQ ID NO:151 (SEQ ID NO:152) GB1: SEQ ID NO:174 (SEQ ID NO:173) GS12: SEQ ID NO:172 (SEQ ID NO:171)
S39 (FIG. 34)	S29	Cs_CBDAS_t28: SEQ ID NO:151 (SEQ ID NO:152) Sc_MFalpha1_1-19: SEQ ID NO:176 (SEQ ID NO:175)
S41 (FIG. 35)	S29	Cs_CBDAS_t28: SEQ ID NO:151 (SEQ ID NO:152) Sc_MFalpha1_1-89: SEQ ID NO:178 (SEQ ID NO:177)
S42 (FIG. 36)	S29	Cs_CBDAS_t28: SEQ ID NO:151 (SEQ ID NO:152) DasherGFP: SEQ ID NO:180 (SEQ ID NO:179) GS12: SEQ ID NO:172 (SEQ ID NO:171)
S43 (FIG. 37)	S29	Cs_CBDAS_t28: SEQ ID NO:151 (SEQ ID NO:152) GS12: SEQ ID NO:172 (SEQ ID NO:171) ER1_tag: SEQ ID NO:182 (SEQ ID NO:181)
S44 (FIG. 38)	S29	Cs_CBDAS_t28: SEQ ID NO:151 (SEQ ID NO:152) GS12: SEQ ID NO:172 (SEQ ID NO:171) ER2_tag: SEQ ID NO:184 (SEQ ID NO:183)
S45 (FIG. 39)	S29	Cs_CBDAS_t28: SEQ ID NO:151 (SEQ ID NO:152) GS12: SEQ ID NO:172 (SEQ ID NO:171) PM1_tag: SEQ ID NO:186 (SEQ ID NO:185)
S46 (FIG. 40)	S29	Cs_CBDAS_t28: SEQ ID NO:151 (SEQ ID NO:152) GS12: SEQ ID NO:172 (SEQ ID NO:171) VC1_tag: SEQ ID NO:188 (SEQ ID NO:187)
S47 (FIG. 41)	S29	Cs_CBDAS_t28: SEQ ID NO:151 (SEQ ID NO:152)

2018256863 11 Apr 2024

Strain (Constructs)	Parent Strain*	Polypeptide SEQ ID NOs (Nucleotide SEQ ID NOs)
		PEX8_tag: SEQ ID NO:190 (SEQ ID NO:189)
S49 (FIGS. 42A, 42B, and 42C)	S29	Cs_OAC: SEQ ID NO:10 (SEQ ID NO:163) Cs_TKS: SEQ ID NO:11 (SEQ ID NO:162) Cs_AAE1_v1: SEQ ID NO:90 (SEQ ID NO:164) Cs_AAE_v1: SEQ ID NO:90 (SEQ ID NO:164)
S50 (FIGS. 43A, 43B, and 43C)	S29	Cs_OAC: SEQ ID NO:10 (SEQ ID NO:163) Cs_TKS: SEQ ID NO:11 (SEQ ID NO:162) Sc_FAA2: SEQ ID NO:169 (SEQ ID NO:168)
S51 (FIGS. 44A, 44B, and 44C)	S29	Cs_OAC: SEQ ID NO:10 (SEQ ID NO:163) Cs_TKS: SEQ ID NO:11 (SEQ ID NO:162)
S78 (FIG. 45)	S51	Cs_AAE1_v1: SEQ ID NO:90 (SEQ ID NO:164) Cs_AAE_v1: SEQ ID NO:90 (SEQ ID NO:164) GB1: SEQ ID NO:174 (SEQ ID NO:173)
S80 (FIG. 46)	S51	Cs_AAE3: SEQ ID NO:92 (SEQ ID NO:166)
S81 (FIG. 47)	S51	Cs_AAE3_Ctrunc: SEQ ID NO:149 (SEQ ID NO:150)
S82 (FIG. 48)	S51	Sc_FAA1: SEQ ID NO:192 (SEQ ID NO:191) FAA1: SEQ ID NO:192 (SEQ ID NO:191)
S83 (FIG. 49)	S51	Sc_FAA2: SEQ ID NO:169 (SEQ ID NO:168)
S84 (FIG. 50)	S51	Sc_FAA2_Ctrunc: SEQ ID NO:194 (SEQ ID NO:193)
S85 (FIG. 51)	S51	Sc_FAA2_Cmut: SEQ ID NO:196 (SEQ ID NO:195) Sc_FAA2: SEQ ID NO:169 (SEQ ID NO:168)
S86 (FIG. 52)	S51	Sc_FAA3: SEQ ID NO:198 (SEQ ID NO:197)
S87 (FIG. 53)	S51	Sc_FAA4: SEQ ID NO:200 (SEQ ID NO:199)
S88 (FIG. 54)	S51	Cs_AAE1_v1: SEQ ID NO:90 (SEQ ID NO:164) Sc_ACC1_act: SEQ ID NO:207 (SEQ ID NO:201)
S89 (FIG. 55)	S51	Sc_FAA2: SEQ ID NO:169 (SEQ ID NO:168) Sc_ACC1_act: SEQ ID NO:207 (SEQ ID NO:201)
S90 (FIGS. 56A, 56B, and 56C)	S29	Cs_OAC: SEQ ID NO:10 (SEQ ID NO:163) Cs_TKS: SEQ ID NO:11 (SEQ ID NO:162) Cs_AAE1_v1: SEQ ID NO:90 (SEQ ID NO:164)

Strain (Constructs)	Parent Strain*	Polypeptide SEQ ID NOs (Nucleotide SEQ ID NOs)
		Cs_AAE_v1: SEQ ID NO:90 (SEQ ID NO:164)
S91 (FIGS. 57A, 57B, and 57C)	S29	Cs_OAC: SEQ ID NO:10 (SEQ ID NO:163) Cs_TKS: SEQ ID NO:11 (SEQ ID NO:162) Sc_FAA2: SEQ ID NO:169 (SEQ ID NO:168)
S94 (FIG. 58)	S31	Cs_PT4_full: SEQ ID NO:110 (SEQ ID NO:111)
S95 (FIG. 59)	S31	GB1: SEQ ID NO:174 (SEQ ID NO:173) Cs_OAC: SEQ ID NO:10 (SEQ ID NO:163)
S97 (FIG. 60)	S31	Cs_OAC: SEQ ID NO:10 (SEQ ID NO:163) Cs_TKS: SEQ ID NO:11 (SEQ ID NO:162) GS12: SEQ ID NO:172 (SEQ ID NO:171)
S104 (FIG. 61)	S21	Cs_PT4: SEQ ID NO:110 (SEQ ID NO:111) Ag_GPPS: SEQ ID NO:133 (SEQ ID NO:134) GB1: SEQ ID NO:174 (SEQ ID NO:173)
S108 (FIG. 62)	S21	Cs_PT4: SEQ ID NO:110 (SEQ ID NO:111) Hb_GPPS: SEQ ID NO:143 (SEQ ID NO:144) GB1: SEQ ID NO:174 (SEQ ID NO:173)
S112 (FIG. 63)	S21	Cs_PT4: SEQ ID NO:110 (SEQ ID NO:111) Cs_GPPS_NTrunc: SEQ ID NO:127 (SEQ ID NO:128)
S114 (FIG. 64)	S21	Cs_PT4: SEQ ID NO:110 (SEQ ID NO:111) Pa_GPPS_NTrunc: SEQ ID NO:131 (SEQ ID NO:132)
S115 (FIG. 65)	S21	Cs_PT4: SEQ ID NO:110 (SEQ ID NO:111) Ag_GPPS_NTrunc: SEQ ID NO:203 (SEQ ID NO:202)
S116 (FIG. 66)	S21	Cs_PT4: SEQ ID NO:110 (SEQ ID NO:111) Pb_GPPS_NTrunc: SEQ ID NO:135 (SEQ ID NO:136)
S118 (FIG. 67)	S21	Cs_PT4: SEQ ID NO:110 (SEQ ID NO:111) Es_GPPS_NTrunc: SEQ ID NO:139 (SEQ ID NO:140)
S123 (FIG. 68)	S29	Cs_THCAS_full: SEQ ID NO:155 (SEQ ID NO:156)
S147 (FIG. 69)	S21	Cs_PT4t: SEQ ID NO:100 (SEQ ID NO:224) Sc_ERG20_mut: SEQ ID NO:60 (SEQ ID NO:161)
S164 (FIG. 70)	S21	Cs_PT1: SEQ ID NO:82 (SEQ ID NO:220)

2018256863 11 Apr 2024

Strain (Constructs)	Parent Strain*	Polypeptide SEQ ID NOs (Nucleotide SEQ ID NOs)
		Sc_ERG20_mut: SEQ ID NO:60 (SEQ ID NO:161)
S165 (FIG. 71)	S21	CsPT1_t75: SEQ ID NO:223 (SEQ ID NO:222) Sc_ERG20_mut: SEQ ID NO:60 (SEQ ID NO:161)
S166 (FIG. 72)	S21	CsPT4_t112: SEQ ID NO:211 (SEQ ID NO:210) Sc_ERG20_mut: SEQ ID NO:60 (SEQ ID NO:161)
S167 (FIG. 73)	S21	CsPT4_t131: SEQ ID NO:213 (SEQ ID NO:212) Sc_ERG20_mut: SEQ ID NO:60 (SEQ ID NO:161)
S168 (FIG. 74)	S21	CsPT4_t142: SEQ ID NO:215 (SEQ ID NO:214) Sc_ERG20_mut: SEQ ID NO:60 (SEQ ID NO:161)
S169 (FIG. 75)	S21	CsPT4_t166: SEQ ID NO:217 (SEQ ID NO:216) Sc_ERG20_mut: SEQ ID NO:60 (SEQ ID NO:161)
S170 (FIG. 76)	S21	CsPT4_t186: SEQ ID NO:219 (SEQ ID NO:218) Sc_ERG20_mut: SEQ ID NO:60 (SEQ ID NO:161)

* If a strain has a parent strain, it is a child strain. All of the constructs present in the parent strain are also all present in the child strain.

[00970] While the present disclosure has been described with reference to the specific embodiments thereof, it should be understood by those skilled in the art that various changes may be made and equivalents may be substituted without departing from the true spirit and scope of the disclosure. In addition, many modifications may be made to adapt a particular situation, material, composition of matter, process, process step or steps, to the objective, spirit and scope of the present disclosure. All such modifications are intended to be within the scope of the claims appended hereto.

[00971] In the claims which follow and in the preceding description of the invention, except where the context requires otherwise due to express language or necessary implication, the word "comprise" or variations such as "comprises" or "comprising" is used in an inclusive sense, i.e. to specify the presence of the stated features but not to preclude the presence or addition of further features in various embodiments of the invention.

2018256863 11 Apr 2024

CLAIMS

What is claimed is:

1. A genetically modified yeast cell for producing a cannabinoid or a cannabinoid derivative, the genetically modified yeast cell comprising one or more heterologous nucleic acids integrated into a chromosome of the genetically modified yeast cell and encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:110 or SEQ ID NO:100.

2. The genetically modified yeast cell of claim 1, wherein the genetically modified yeast cell comprises one or more heterologous nucleic acids integrated into a chromosome of the genetically modified yeast cell and encoding a tetraketide synthase (TKS) polypeptide and one or more heterologous nucleic acids integrated into a chromosome of the genetically modified yeast cell and encoding an olivetolic acid (OAC) polypeptide.

3. The genetically modified yeast cell of claim 1 or 2, wherein the genetically modified yeast cell comprises one or more of the following:

a) one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative; or

b) one or more heterologous nucleic acids encoding a polypeptide that generates geranyl pyrophosphate.

4. The genetically modified yeast cell of any one of claims 1-3, wherein the genetically modified yeast cell comprises one or more of the following:

a) one or more heterologous nucleic acids encoding a HMG-CoA synthase (HMGS) polypeptide;

b) one or more heterologous nucleic acids encoding a 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMGR) polypeptide;

c) one or more heterologous nucleic acids encoding a mevalonate kinase (MK) polypeptide;

d) one or more heterologous nucleic acids encoding a phosphomevalonate kinase (PMK) polypeptide;

e) one or more heterologous nucleic acids encoding a mevalonate pyrophosphate decarboxylase (MVD) polypeptide; or

f) one or more heterologous nucleic acids encoding a isopentenyl diphosphate isomerase (IDI) polypeptide.

5. The genetically modified yeast cell of claim 4, wherein the genetically modified yeast cell comprises one or more heterologous nucleic acids encoding a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA and wherein the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide.

6. The genetically modified yeast cell of any one of claims 1-5, wherein the yeast cell is *Saccharomyces cerevisiae*.

7. The genetically modified yeast cell of any one of claims 1-6, wherein the genetically modified yeast cell comprises one or more heterologous nucleic acids encoding a cannabinoid synthase polypeptide, wherein the cannabinoid synthase polypeptide is a tetrahydrocannabinolic acid (THCA) synthase polypeptide or a cannabidiolic acid (CBDA) synthase polypeptide.

8. The genetically modified yeast cell of claim 7, wherein the THCA synthase polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO:155 and the CBDA synthase polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO:88 or SEQ ID NO:151.

9. The genetically modified yeast cell of any one of claims 1-8, wherein the genetically modified yeast cell comprises one or more of the following:

a) one or more heterologous nucleic acids encoding a polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative, wherein the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is an acyl-activating enzyme (AAE) polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:90, SEQ ID NO:92, or SEQ ID NO:149; a fatty acyl-CoA ligase polypeptide comprising an amino acid

sequence having at least 85% sequence identity to SEQ ID NO:145 or SEQ ID NO:147; or a fatty acyl-CoA synthetase (FAA) polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200;

b) one or more heterologous nucleic acids encoding a polypeptide that generates geranyl pyrophosphate, wherein the polypeptide that generates geranyl pyrophosphate is a geranyl pyrophosphate synthetase (GPPS) polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:60;

c) one or more heterologous nucleic acids encoding a HMG-CoA synthase (HMGS) polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:115;

d) one or more heterologous nucleic acids encoding a 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMGR) polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:208;

e) one or more heterologous nucleic acids encoding a mevalonate kinase (MK) polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:64;

f) one or more heterologous nucleic acids encoding a phosphomevalonate kinase (PMK) polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:205;

g) one or more heterologous nucleic acids encoding a mevalonate pyrophosphate decarboxylase (MVD) polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:66;

h) one or more heterologous nucleic acids encoding an isopentenyl diphosphate isomerase (IDI) polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:58;

i) one or more heterologous nucleic acids encoding a polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA, wherein the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:25; or

j) one or more heterologous nucleic acids integrated into a chromosome of the genetically modified yeast cell and encoding a tetraketide synthase (TKS)

polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:11 and one or more heterologous nucleic acids integrated into a chromosome of the genetically modified yeast cell and encoding an olivetolic acid cyclase (OAC) polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:10.

10. A method of producing a cannabinoid or a cannabinoid derivative, the method comprising:

a) culturing a genetically modified yeast cell comprising:

i) one or more heterologous nucleic acids integrated into a chromosome of the genetically modified yeast cell and encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:110 or SEQ ID NO:100; and

ii) one or more heterologous nucleic acids integrated into a chromosome of the genetically modified yeast cell and encoding a tetraketide synthase (TKS) polypeptide comprising the amino acid sequence of SEQ ID NO:11; and/or

iii) one or more heterologous nucleic acids integrated into a chromosome of the genetically modified yeast cell and encoding an olivetolic acid cyclase (OAC) polypeptide comprising the amino acid sequence of SEQ ID NO:10;

in a suitable medium containing a carboxylic acid; and

b) recovering the produced cannabinoid or cannabinoid derivative.

11. The method of claim 10, wherein the TKS polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO:11 and the OAC polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO:10.

12. The method of claim 10 or 11, wherein the carboxylic acid is hexanoic acid, butyric acid, valeric acid or derivative thereof comprising one or more functional or reactive groups selected from azido, halo, methyl, alkyl, alkynyl, alkenyl, methoxy, alkoxy, acetyl, amino, carboxyl, carbonyl, oxo, ester, hydroxyl, thio, cyano, aryl, heteroaryl, cycloalkyl, cycloalkenyl, cycloalkylalkenyl, cycloalkylalkynyl, cycloalkenylalkyl, cycloalkenylalkenyl, cycloalkenylalkynyl, heterocyclylalkenyl, heterocyclylalkynyl, heteroarylalkenyl, heteroarylalkynyl,

arylalkenyl, arylalkynyl, spirocyclyl, heterospirocyclyl, heterocyclyl, thioalkyl, sulfone, sulfonyl, sulfoxide, amido, alkylamino, dialkylamino, arylamino, alkylarylamino, diarylamino, N-oxide, imide, enamine, imine, oxime, hydrazone, nitrile, aralkyl, cycloalkylalkyl, haloalkyl, heterocyclylalkyl, heteroarylalkyl, nitro, thioxo, azide, amine, ester, cyano, thioester, thioether, sulfonyl halide, alcohol, thiol, succinimidyl ester, isothiocyanate, iodoacetamide, maleimide, hydrazine, alkynyl, and alkenyl, wherein, optionally, the carboxylic acid is hexanoic acid and wherein the cell culture produces the cannabinoid or cannabinoid derivative at a concentration of about 1 mg/L to about 500 mg/L.

13. The method of claim 10 or 11, wherein the cannabinoid is cannabigerolic acid, cannabigerolic acid monomethylether (CBGAM), cannabigerol, cannabigerol monomethylether, Δ^9-trans-tetrahydrocannabinol, Δ^9-tetrahydrocannabinolic acid, Δ^9-tetrahydrocannabinolic acid A, Δ^9-tetrahydrocannabinolic acid B, Δ^9-tetrahydrocannabinolic acid-C4, Δ^9-tetrahydrocannabinol-C4, Δ^9-tetrahydrocannabivarinic acid, Δ^9-tetrahydrocannabivarin, Δ^9-tetrahydrocannabiorcolic acid, Δ^9-tetrahydrocannabiorcol, Δ^7-cis-iso-tetrahydrocannabivarin, Δ^9-tetrahydrocannabinol, Δ^8-tetrahydrocannabinolic acid, Δ^8-trans-tetrahydrocannabinol, Δ^8-tetrahydrocannabinol, cannabidiolic acid, cannabidiol, cannabidiol monomethylether (CBDM), cannabidiol-C4 (CBD-C_4), cannabidiorcol (CBD-C_1), cannabitriol, cannabichromenic acid, cannabichromene, cannabinolic acid, cannabinol, cannabinol methylether, cannabinol-C4, cannabinol-C2, cannabidivarinic acid, cannabidivarin, tetrahydrocannabivarinic acid, tetrahydrocannabivarin, cannabichromevarinic acid, cannabichromevarin, cannabigerovarinic acid, cannabigerovarin, cannabicyclolic acid, cannabicyclol, cannabicyclovarin, cannabielsoic acid A, cannabielsoic acid B, cannabielsoinic acid, cannabielsoin, cannabicitranic acid, or cannabicitran.

14. A method of producing a cannabinoid or cannabinoid derivative, the method comprising:

a) culturing a genetically modified yeast cell comprising:

i) one or more heterologous nucleic acids integrated into a chromosome of the genetically modified yeast cell and encoding a geranyl pyrophosphate:olivetolic acid

geranyltransferase polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:110 or SEQ ID NO:100;

ii) one or more heterologous nucleic acids integrated into a chromosome of the genetically modified yeast cell and encoding a tetraketide synthase (TKS) polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:11;

iii) one or more heterologous nucleic acids integrated into a chromosome of the genetically modified yeast cell and encoding an olivetolic acid cyclase (OAC) polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:10;

iv) one or more heterologous nucleic acids encoding one or more polypeptides that generate geranyl pyrophosphate, wherein the polypeptide that generates geranyl pyrophosphate is a geranyl pyrophosphate synthetase (GPPS) polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:60;

v) one or more heterologous nucleic acids encoding one or more polypeptides that generate an acyl-CoA compound or an acyl-CoA compound derivative wherein the polypeptide that generates an acyl-CoA compound or an acyl-CoA compound derivative is an acyl-activating enzyme (AAE) polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:90, SEQ ID NO:92, or SEQ ID NO:149; a fatty acyl-CoA ligase polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:145 or SEQ ID NO:147; or a fatty acyl-CoA synthetase (FAA) polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:169, SEQ ID NO:192, SEQ ID NO:194, SEQ ID NO:196, SEQ ID NO:198, or SEQ ID NO:200; and

vi) one or more heterologous nucleic acids encoding one or more polypeptides that generate malonyl-CoA, wherein the polypeptide that generates malonyl-CoA is an acetyl-CoA carboxylase-1 (ACC1) polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:9, SEQ ID NO:97, or SEQ ID NO:207;

in a suitable medium; and

b) recovering the produced cannabinoid or cannabinoid derivative,

2018256863 11 Apr 2024

wherein, optionally, the genetically modified yeast cell comprises one or more heterologous nucleic acids encoding a THCA synthase polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:155 or a CBDA synthase polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:88 or SEQ ID NO:151; and/or

wherein at least one of the heterologous nucleic acids integrated into a chromosome of the genetically modified yeast cell is operably linked to an inducible promoter.

15. A method of producing a cannabinoid or a cannabinoid derivative, the method comprising use of a genetically modified yeast cell comprising a heterologous nucleic acid integrated into a chromosome of the genetically modified yeast cell and encoding a geranyl pyrophosphate:olivetolic acid geranyltransferase polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO:110 or SEQ ID NO: 100.

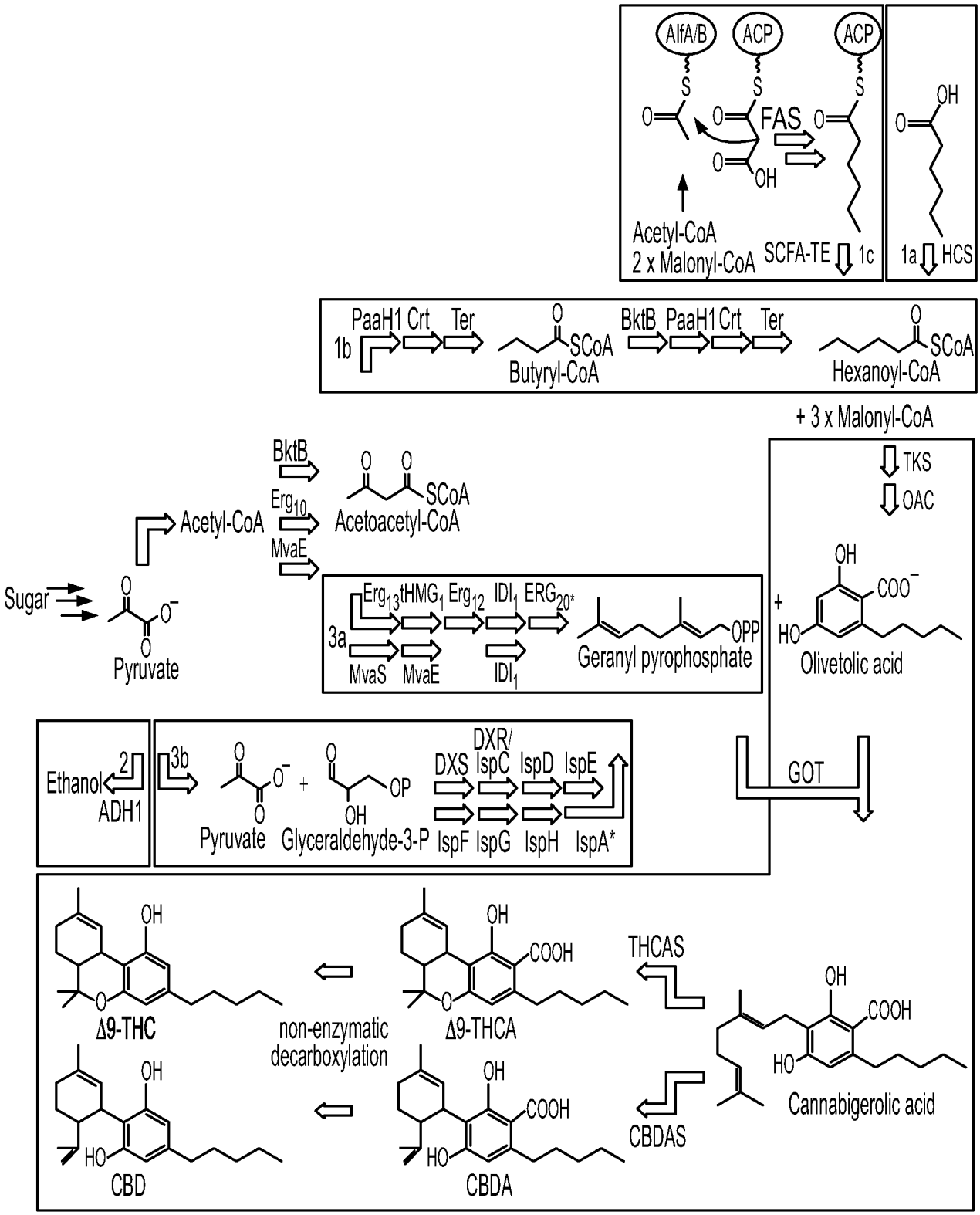

AlfA/B
ACP
ACP
FAS
Acetyl-CoA
2 x Malonyl-CoA
SCFA-TE
1c
1a
HCS
PaaH1 Crt Ter
1b
Butyryl-CoA
BktB PaaH1 Crt Ter
Hexanoyl-CoA
+ 3 x Malonyl-CoA
TKS
OAC
Olivetolic acid
BktB
Erg10
MvaE
Acetoacetyl-CoA
Acetyl-CoA
Sugar
Pyruvate
Erg13 tHMG1 Erg12 IDI1 ERG20*
3a
MvaS MvaE IDI1
Geranyl pyrophosphate
Ethanol
2
ADH1
3b
Pyruvate
Glyceraldehyde-3-P
DXS DXR/ IspC IspD IspE
IspF IspG IspH IspA*
GOT
Δ9-THC
non-enzymatic decarboxylation
Δ9-THCA
THCAS
Cannabigerolic acid
CBD
CBDA
CBDAS

FIG. 1

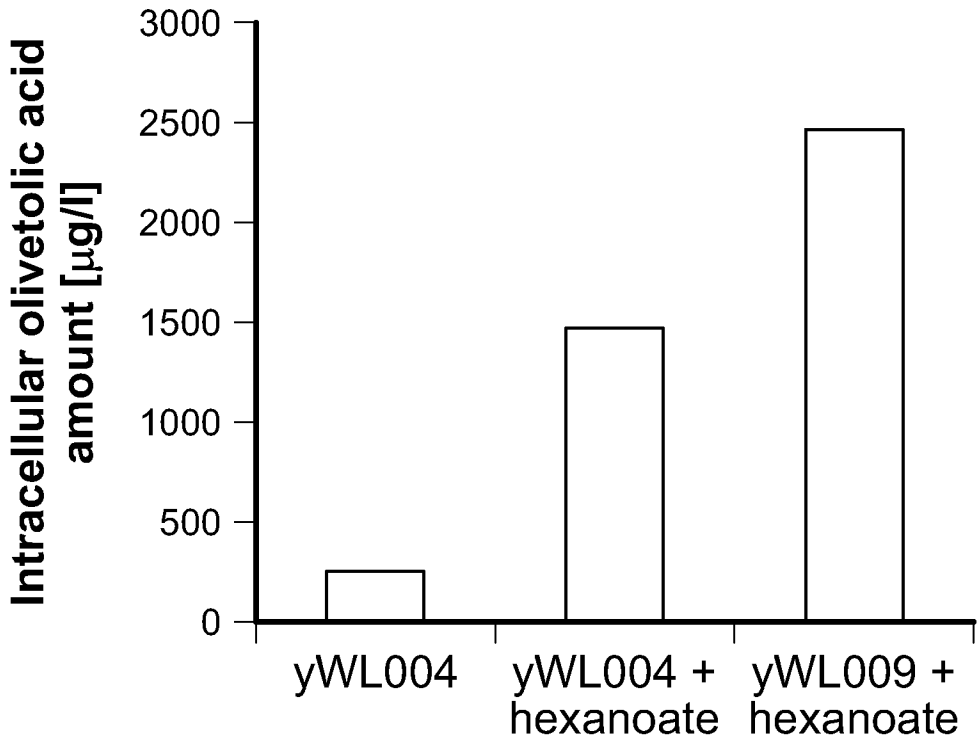

Intracellular olivetolic acid amount [μg/l]
3000
2500
2000
1500
1000
500
0
yWL004
yWL004 + hexanoate
yWL009 + hexanoate

FIG. 2

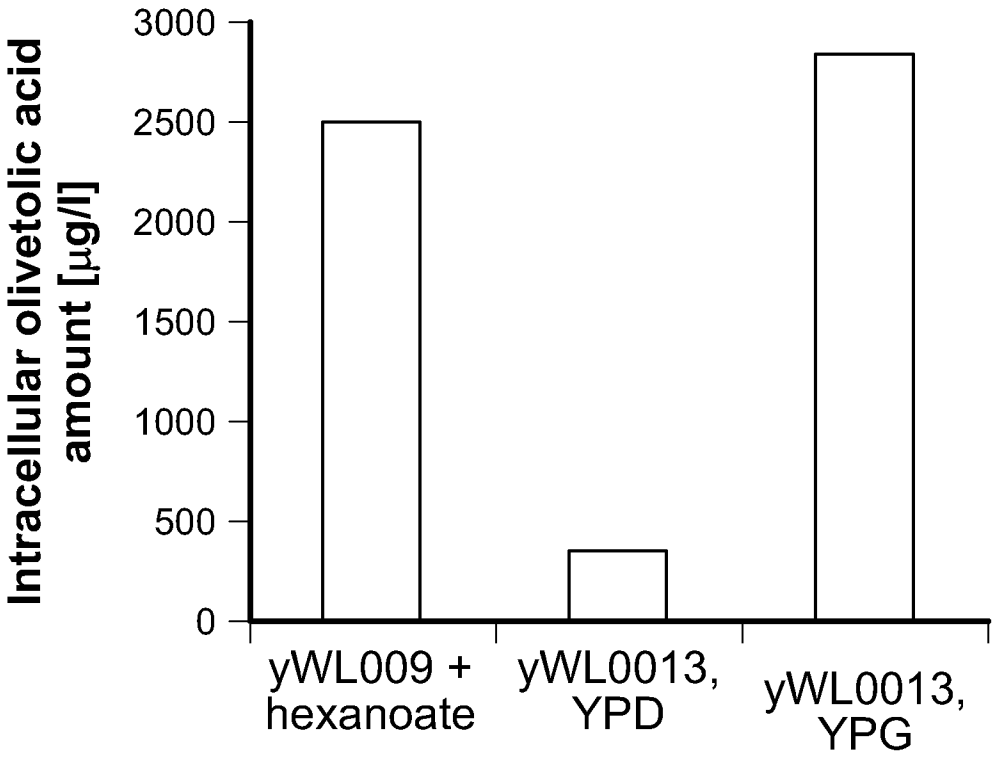

Intracellular olivetolic acid amount [μg/l]
3000
2500
2000
1500
1000
500
0
yWL009 + hexanoate
yWL0013, YPD
yWL0013, YPG

FIG. 3

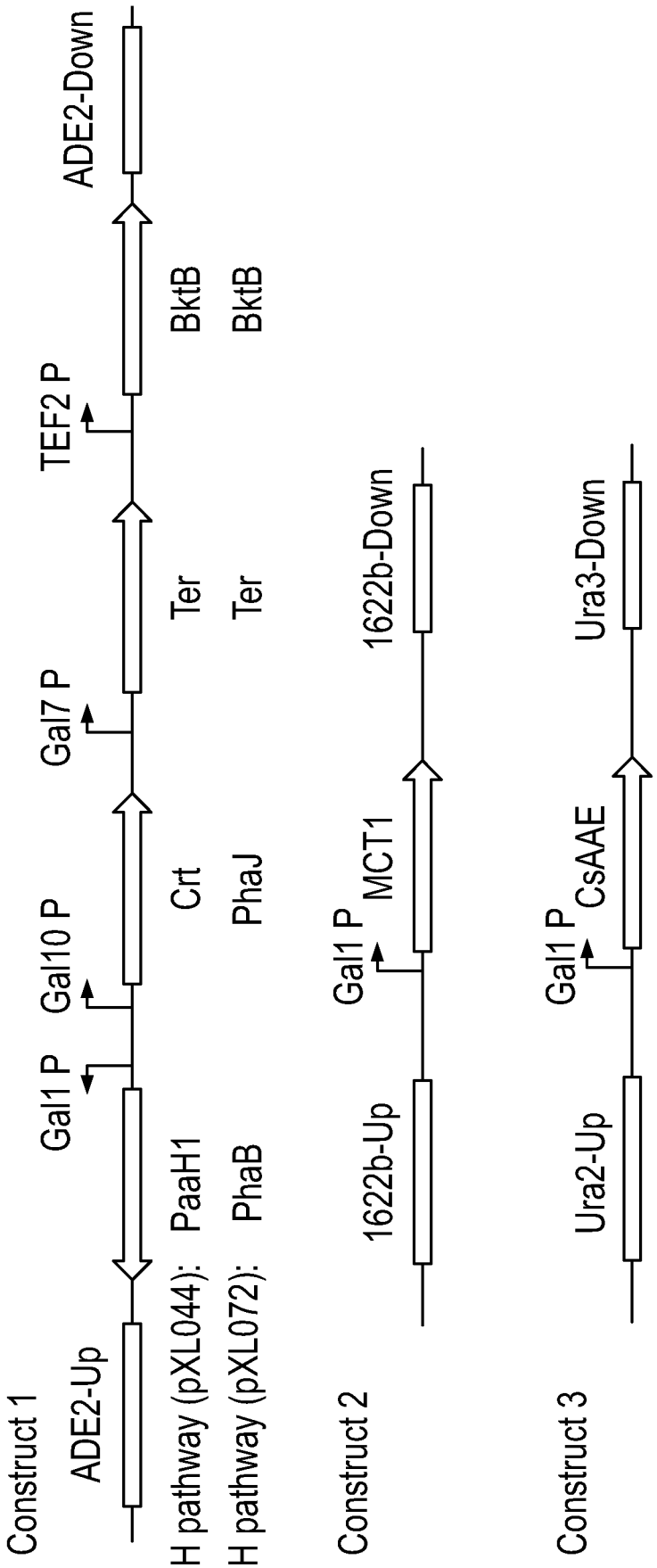

Construct 1
ADE2-Up
Gal1 P
Gal10 P
Gal7 P
TEF2 P
ADE2-Down
NADH pathway (pXL044): PaaH1 Crt Ter BktB
NADPH pathway (pXL072): PhaB PhaJ Ter BktB
Construct 2
1622b-Up
Gal1 P
MCT1
1622b-Down
Construct 3
Ura2-Up
Gal1 P
CsAAE
Ura3-Down

FIG. 4

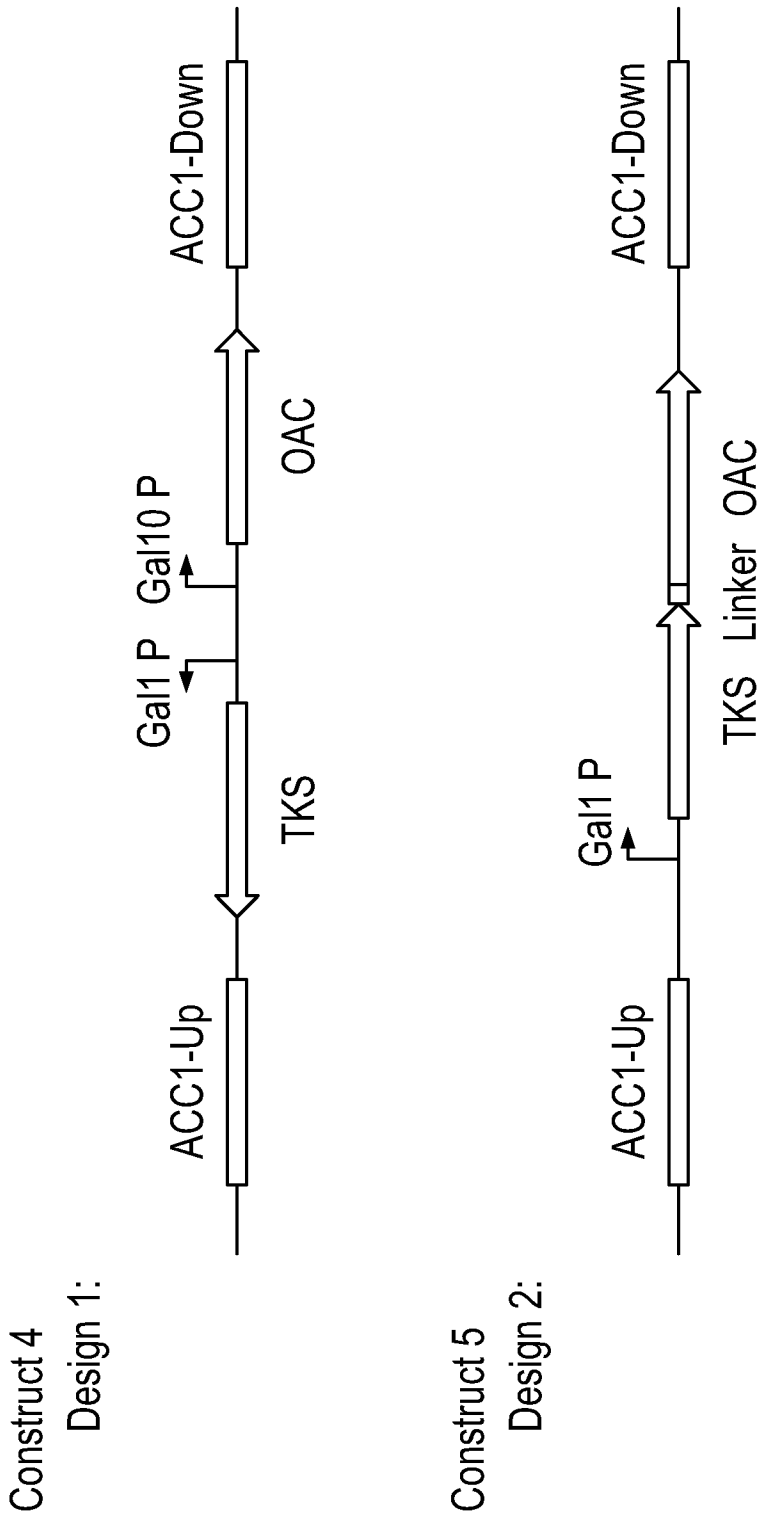

Construct 4
Design 1:
ACC1-Up
TKS
Gal1 P
Gal10 P
OAC
ACC1-Down
Construct 5
Design 2:
ACC1-Up
Gal1 P
TKS
Linker
OAC
ACC1-Down

FIG. 5

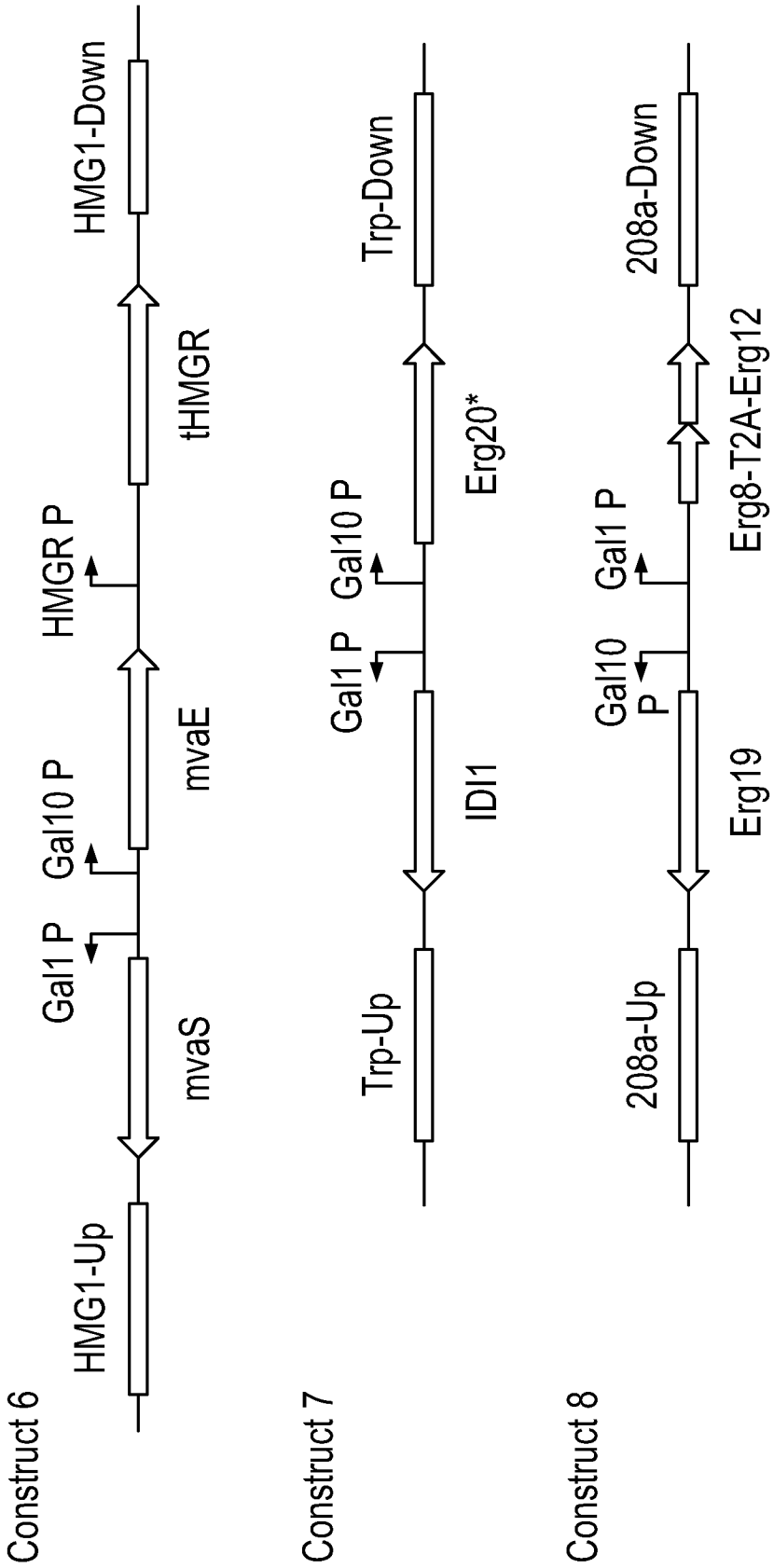

Construct 6
HMG1-Up
Gal1 P
Gal10 P
mvaS
mvaE
HMGR P
tHMGR
HMG1-Down
Construct 7
Trp-Up
Gal1 P
Gal10 P
IDI1
Erg20*
Trp-Down
Construct 8
208a-Up
Gal10 P
Gal1 P
Erg19
Erg8-T2A-Erg12
208a-Down

FIG. 6

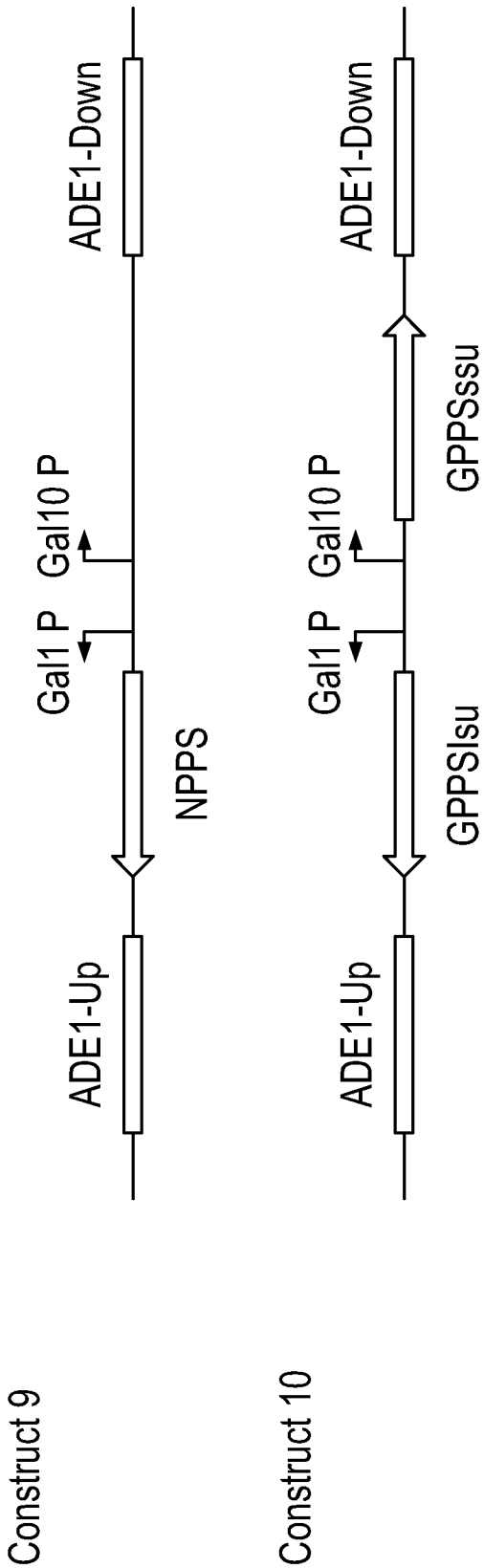

ADE1-Down
Gal10 P
Gal1 P
NPPS
ADE1-Up
Construct 9
ADE1-Down
GPPSssu
Gal10 P
Gal1 P
GPPSlsu
ADE1-Up
Construct 10

FIG. 7

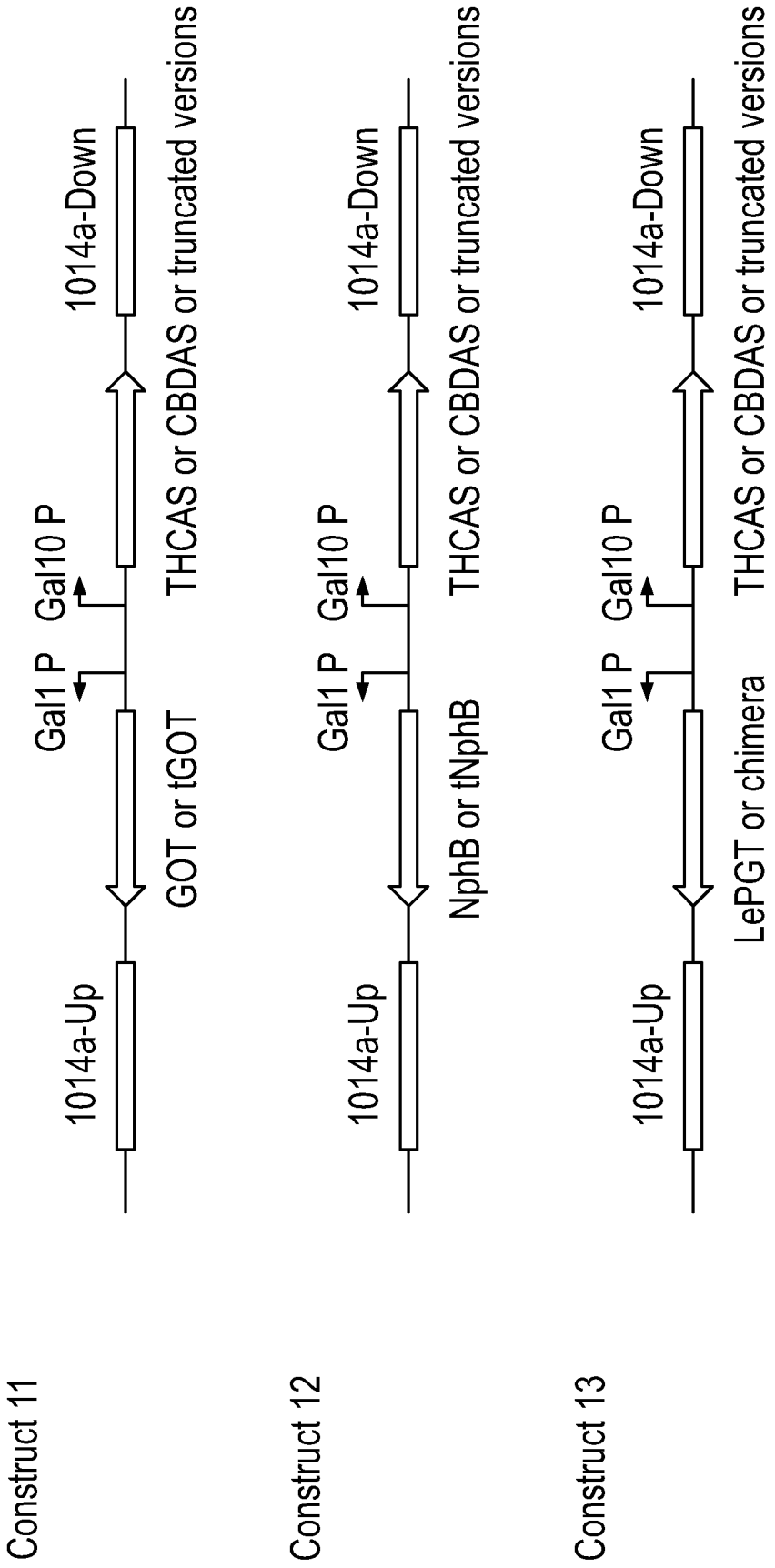

Construct 11
1014a-Up
Gal1 P
GOT or tGOT
Gal10 P
THCAS or CBDAS or truncated versions
1014a-Down
Construct 12
1014a-Up
Gal1 P
NphB or tNphB
Gal10 P
THCAS or CBDAS or truncated versions
1014a-Down
Construct 13
1014a-Up
Gal1 P
LePGT or chimera
Gal10 P
THCAS or CBDAS or truncated versions
1014a-Down

FIG. 8

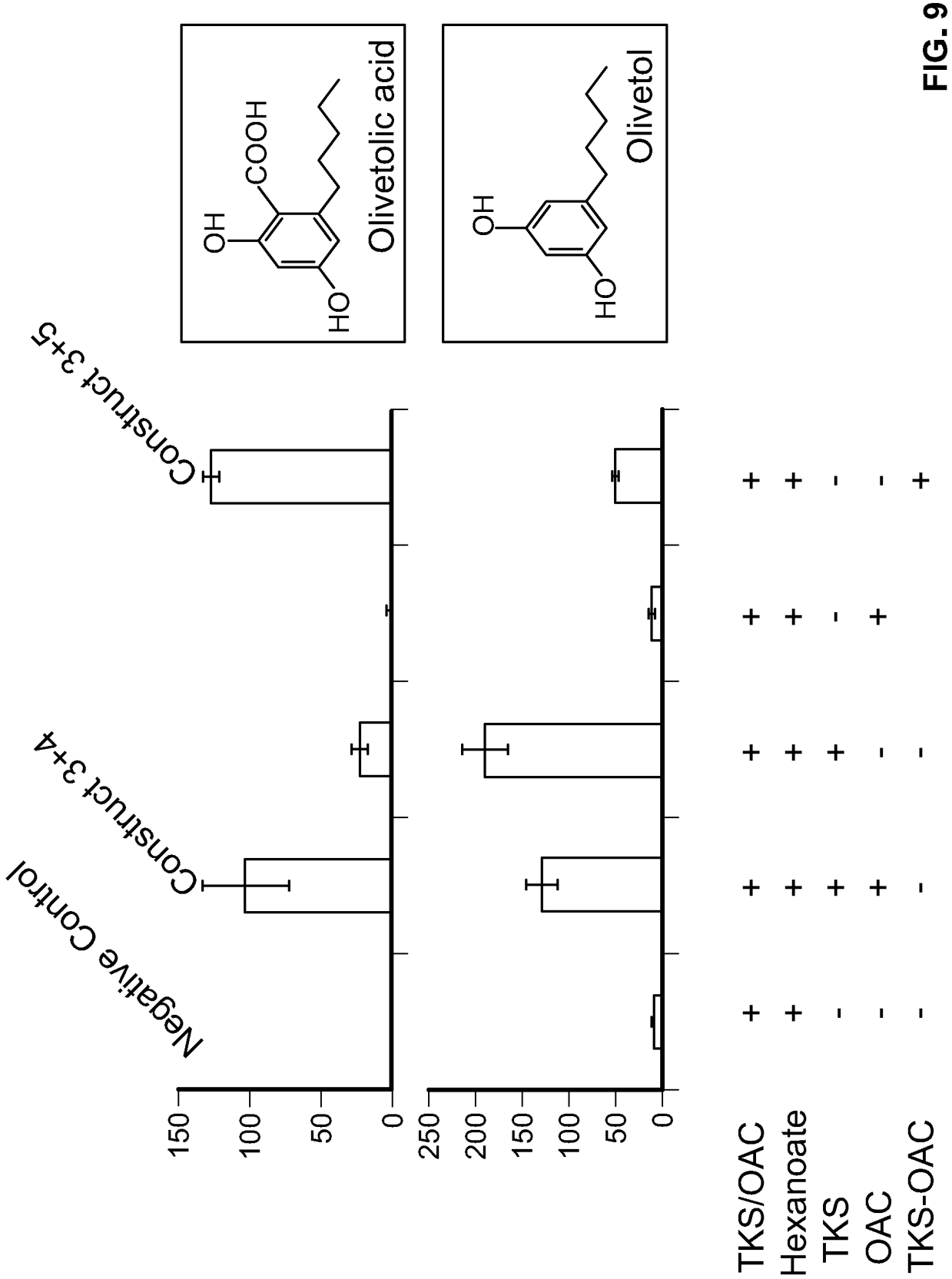

Olivetolic acid
COOH
OH
HO
Olivetol
OH
HO
Negative Control
Construct 3+4
Construct 3+5
150
100
50
0
250
200
150
100
50
0
TKS/OAC
Hexanoate
TKS
OAC
TKS-OAC
+ + - - +
+ + + - -
+ + - + -
+ + - - -

FIG. 9

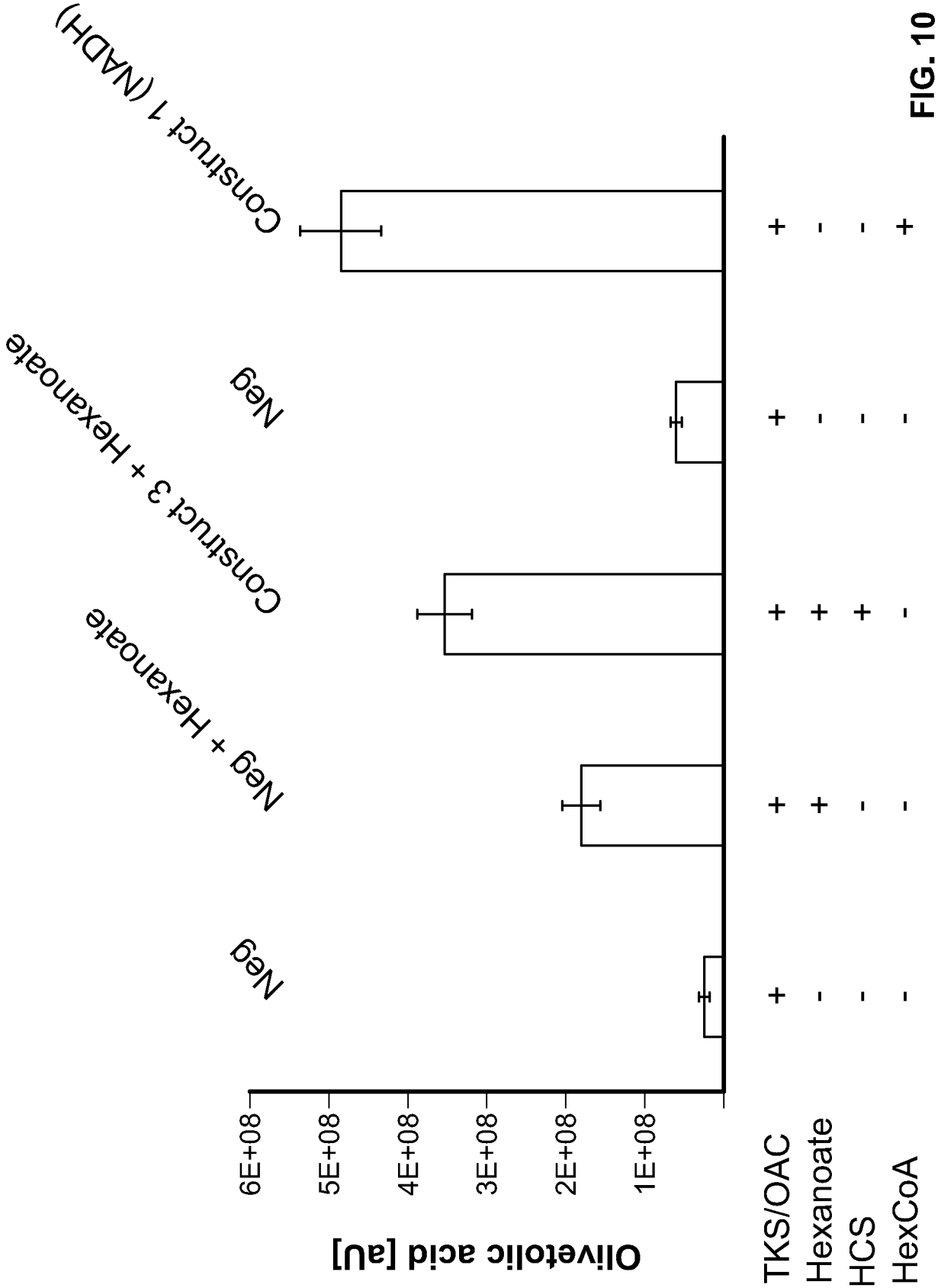

Oliveto lic acid [aU]
6E+08
5E+08
4E+08
3E+08
2E+08
1E+08
TKS/OAC
Hexanoate
HCS
HexCoA
+ - - -
+ + - -
+ + + -
+ - - -
Neg
Neg + Hexanoate
Construct 3 + Hexanoate
Construct 1 (NADH)

FIG. 10

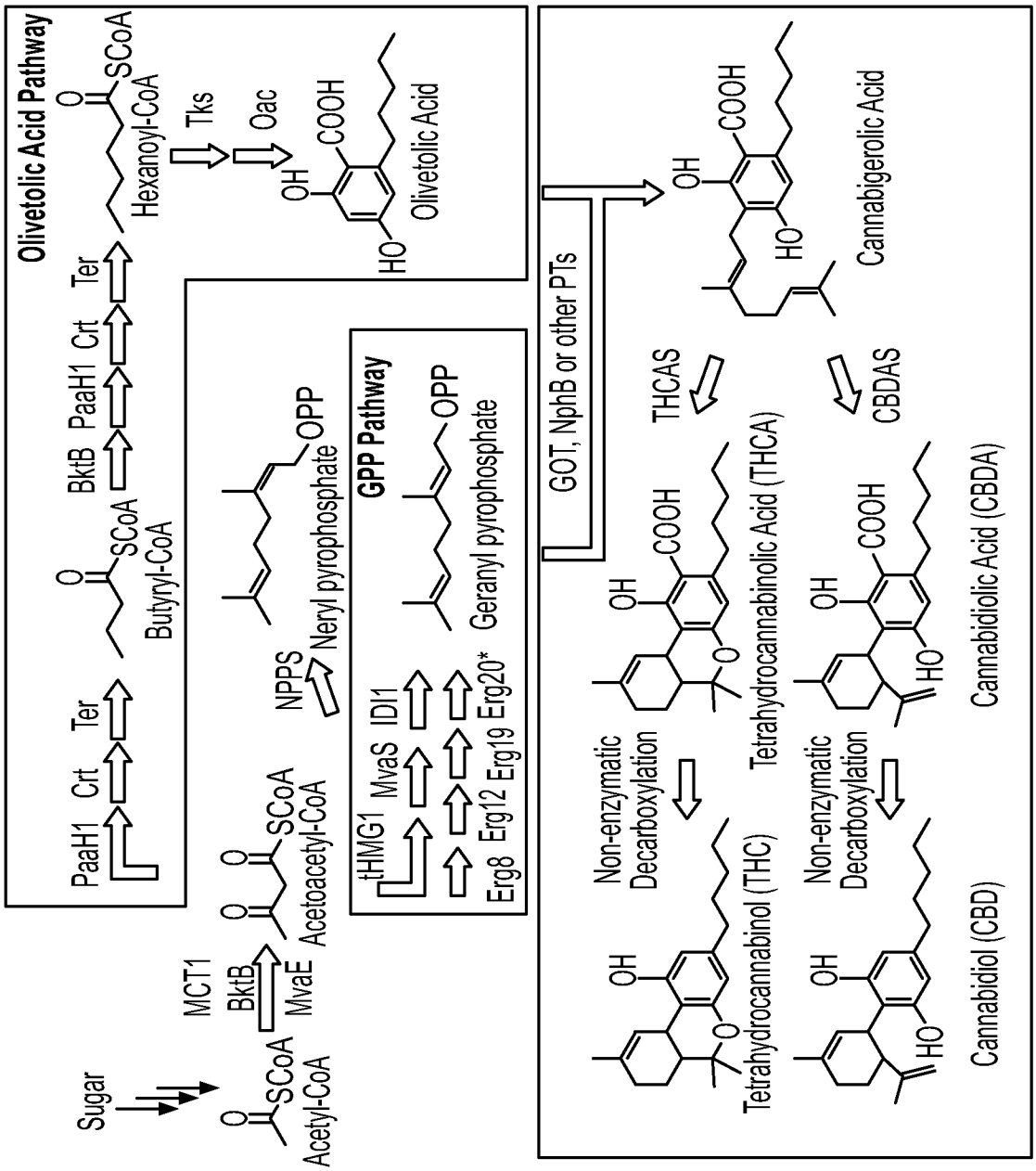

Olivetolic Acid Pathway
PaaH1
Crt
Ter
BktB
PaaH1
Crt
Ter
Butyryl-CoA
Hexanoyl-CoA
Tks
Oac
Olivetolic Acid
GPP Pathway
NPPS
Neryl pyrophosphate
Geranyl pyrophosphate
tHMG1
MvaS
IDI1
Erg8
Erg12
Erg19
Erg20*
Sugar
MCT1
Acetyl-CoA
BktB
MvaE
Acetoacetyl-CoA
GOT, NphB or other PTs
THCAS
CBDAS
Cannabigerolic Acid
Tetrahydrocannabinolic Acid (THCA)
Cannabidiolic Acid (CBDA)
Non-enzymatic Decarboxylation
Tetrahydrocannabinol (THC)
Cannabidiol (CBD)

FIG. 11

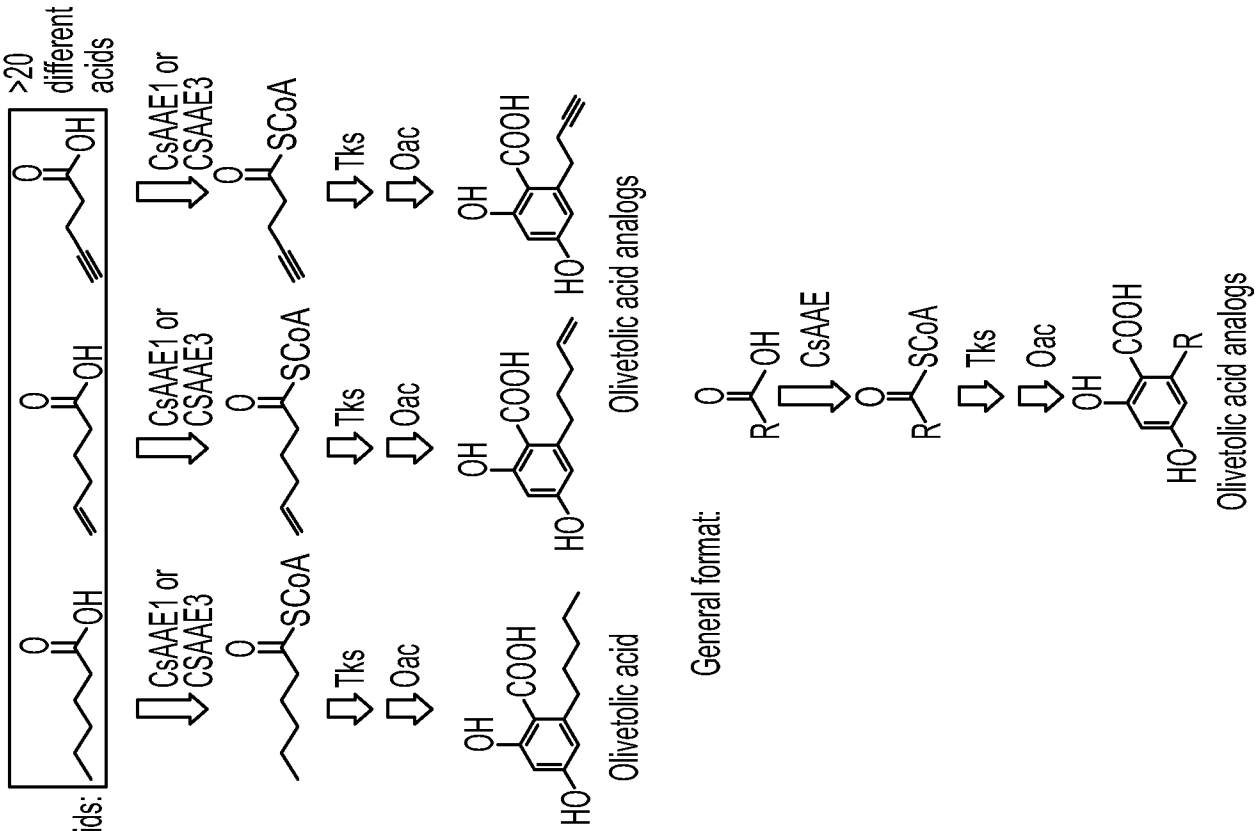

Feeding acids:
>20 different acids
CsAAE1 or CSAAE3
Tks
Oac
Olivetolic acid
Olivetolic acid analogs
General format:
CsAAE
Olivetolic acid analogs

FIG. 11 (Cont.)

12/101

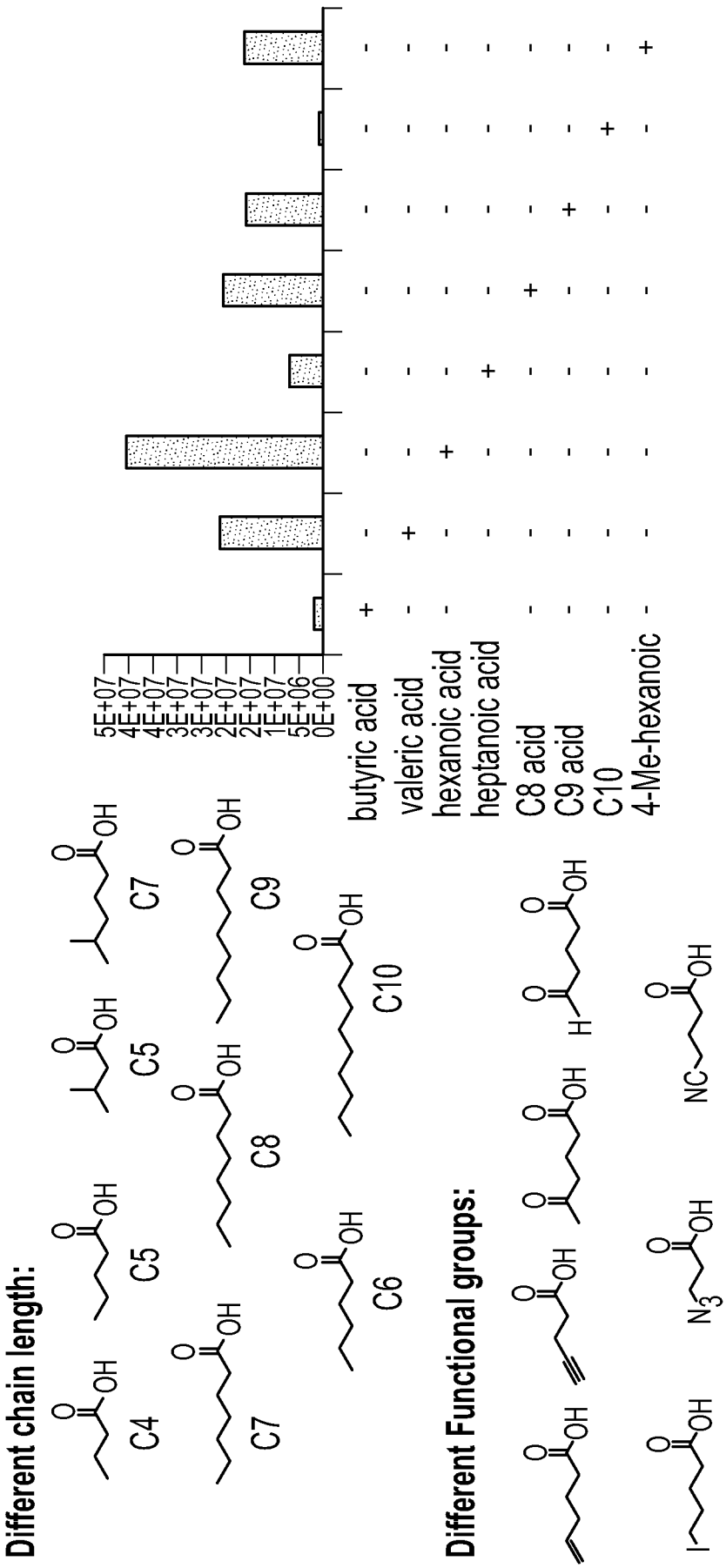

Different chain length:
C4
C5
C5
C7
C7
C8
C9
C6
C10
Different Functional groups:
N3
NC
5E+07
4E+07
4E+07
3E+07
3E+07
2E+07
2E+07
1E+07
5E+06
0E+00
butyric acid
valeric acid
hexanoic acid
heptanoic acid
C8 acid
C9 acid
C10
4-Me-hexanoic

FIG. 12

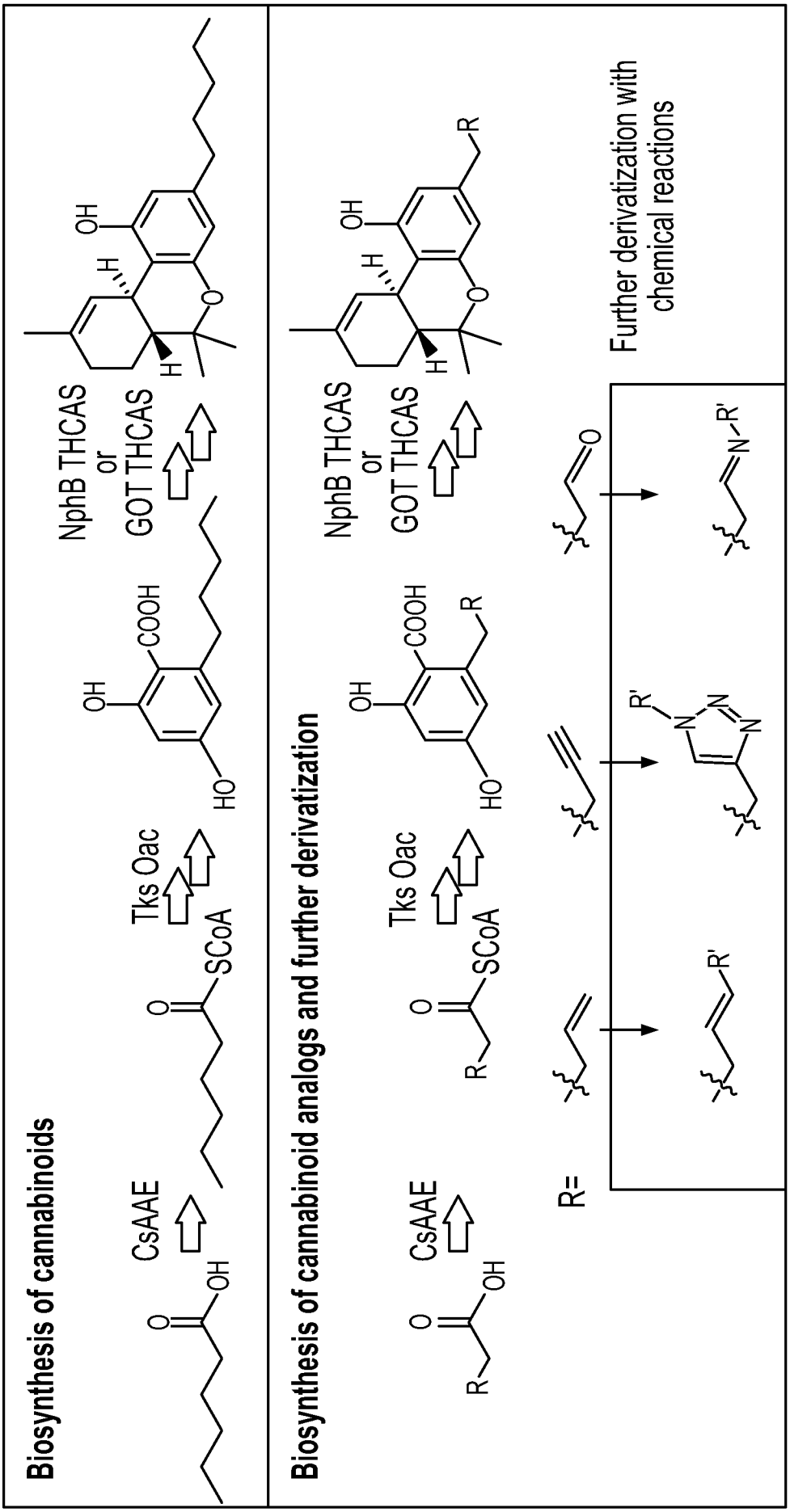

Biosynthesis of cannabinoids
CsAAE
Tks Oac
NphB THCAS
or
GOT THCAS
Biosynthesis of cannabinoid analogs and further derivatization
CsAAE
Tks Oac
NphB THCAS
or
GOT THCAS
Further derivatization with
chemical reactions
R=

FIG. 13

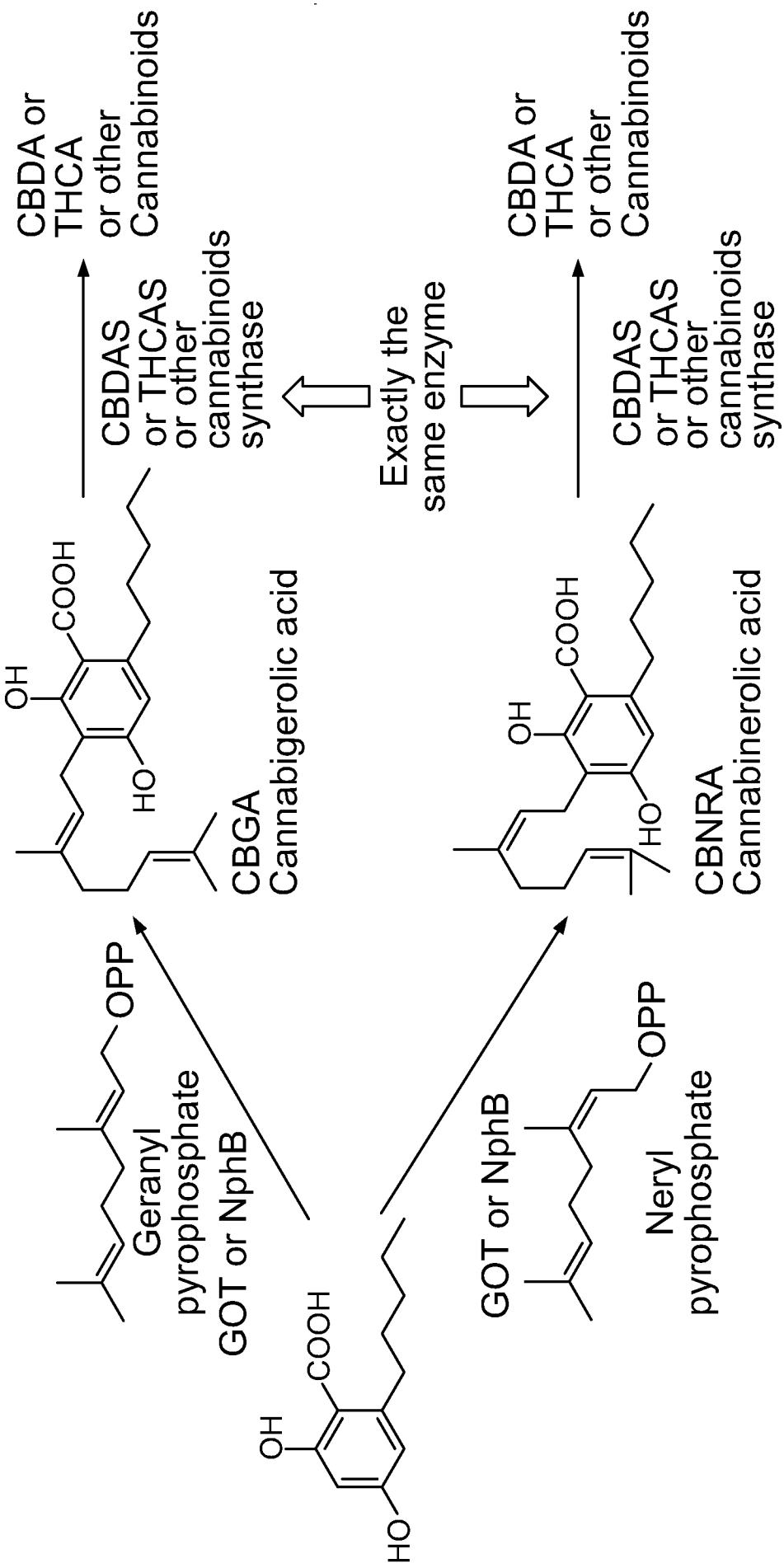

CBDA or
THCA
or other
Cannabinoids
CBDAS
or THCAS
or other
cannabinoids
synthase
Exactly the
same enzyme
CBDA or
THCA
or other
Cannabinoids
CBDAS
or THCAS
or other
cannabinoids
synthase
COOH
OH
HO
CBGA
Cannabigerolic acid
CBNRA
Cannabinerolic acid
OPP
Geranyl
pyrophosphate
GOT or NphB
GOT or NphB
Neryl
pyrophosphate

FIG. 14

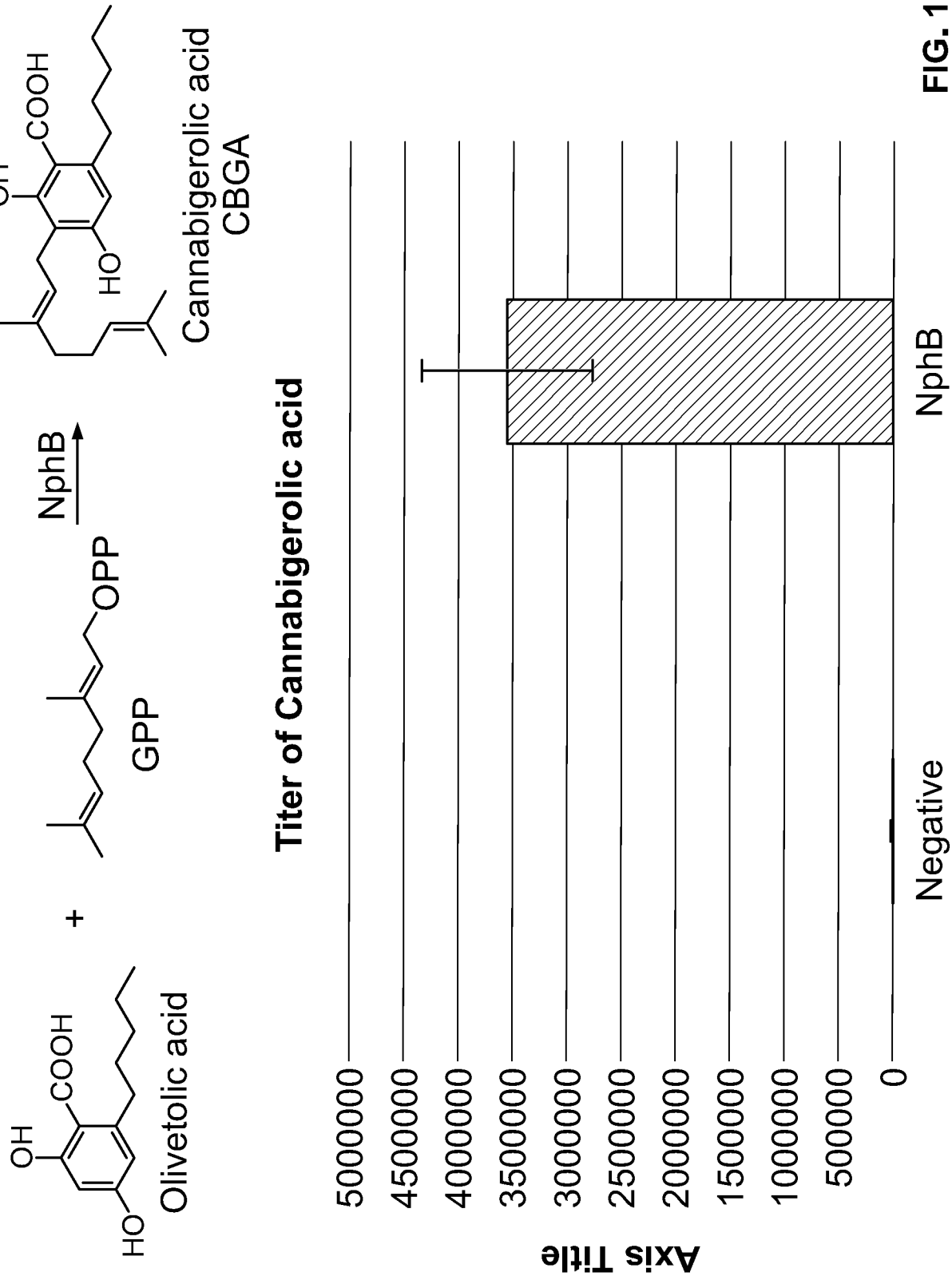

Olivetolic acid
+
GPP
OPP
NphB
Cannabigerolic acid
CBGA
OH
COOH
HO
Titer of Cannabigerolic acid
Axis Title
5000000
4500000
4000000
3500000
3000000
2500000
2000000
1500000
1000000
500000
0
Negative
NphB

FIG. 15

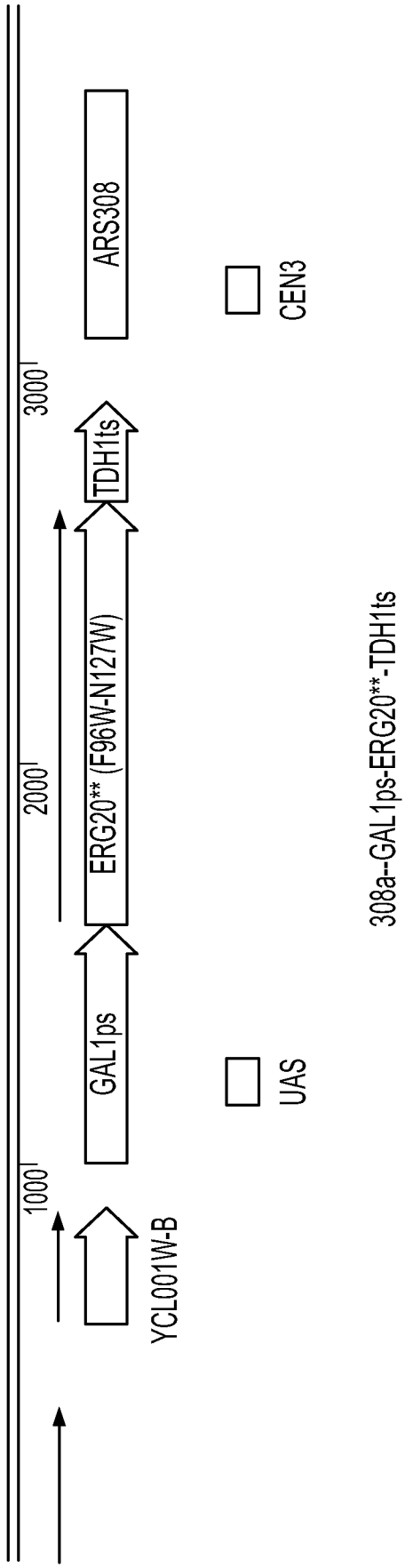

1000
2000
3000
YCL001W-B
GAL1ps
ERG20** (F96W-N127W)
TDH1ts
ARS308
UAS
CEN3
308a--GAL1ps-ERG20**-TDH1ts

FIG. 16

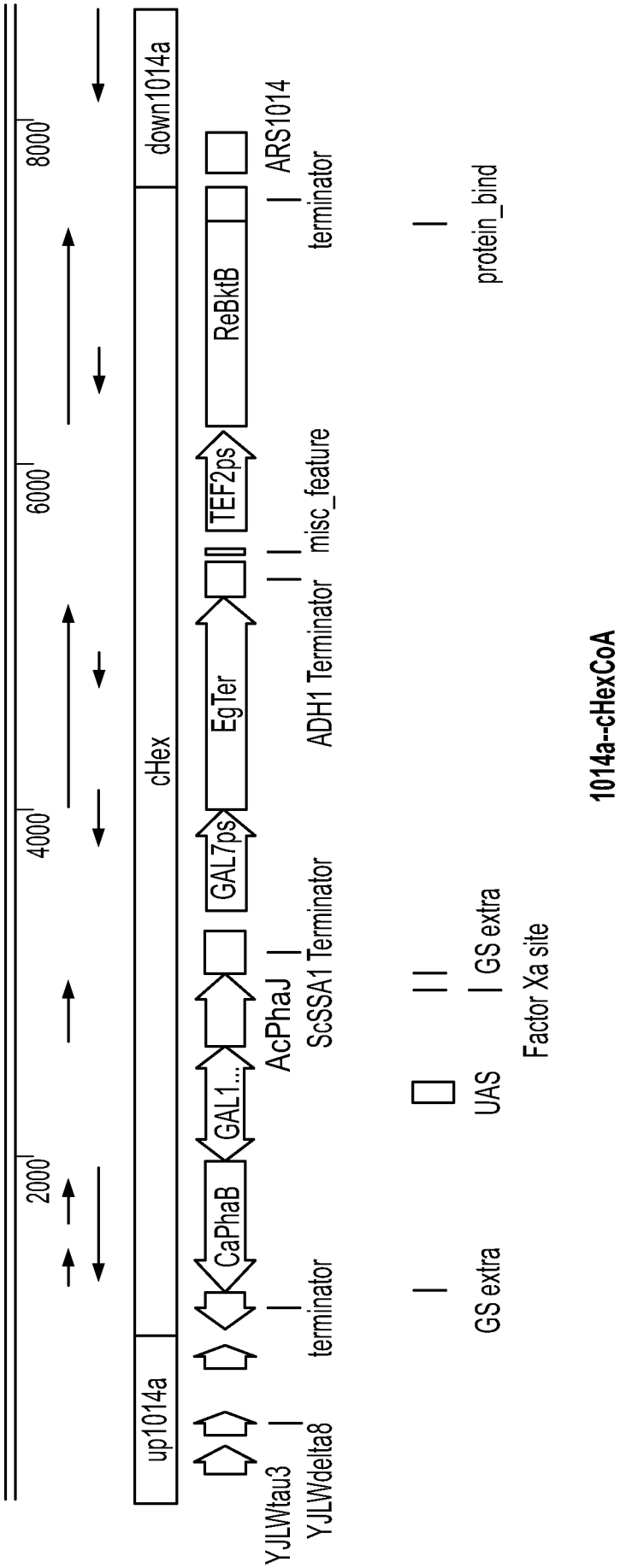

2000
4000
6000
8000
up1014a
cHex
down1014a
YJLWtau3
YJLWdelta8
terminator
CaPhaB
GAL1...
AcPhaJ
ScSSA1 Terminator
GAL7ps
EgTer
ADH1 Terminator
TEF2ps
misc_feature
ReBktB
terminator
ARS1014
GS extra
UAS
GS extra
Factor Xa site
protein_bind
1014a--cHexCoA

FIG. 17

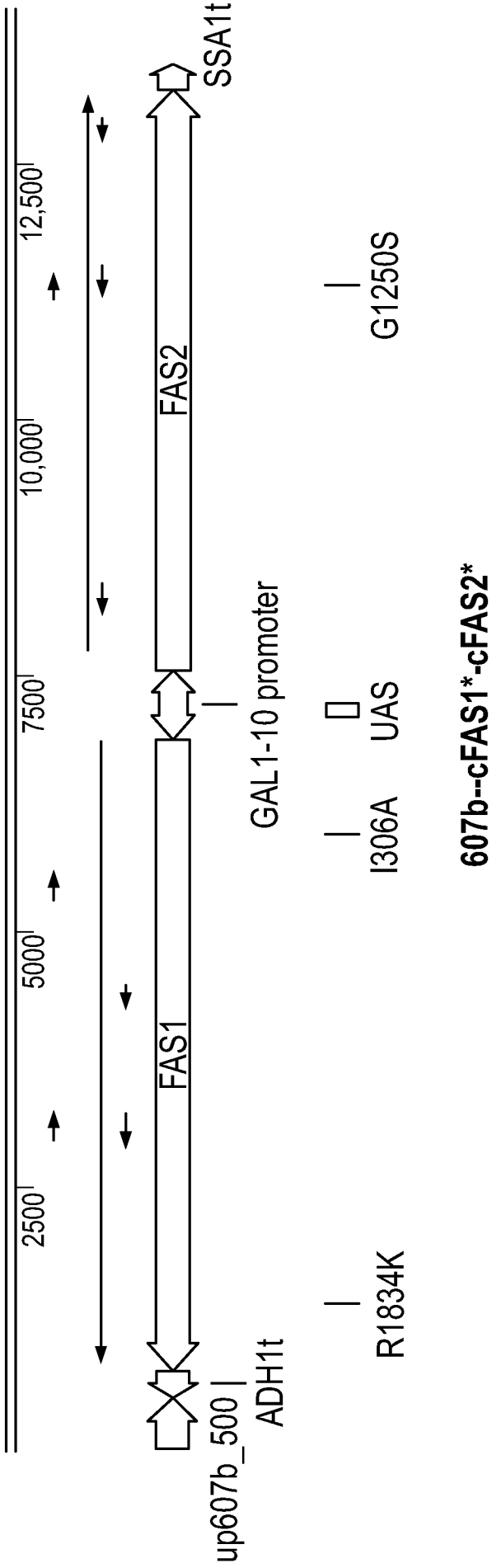
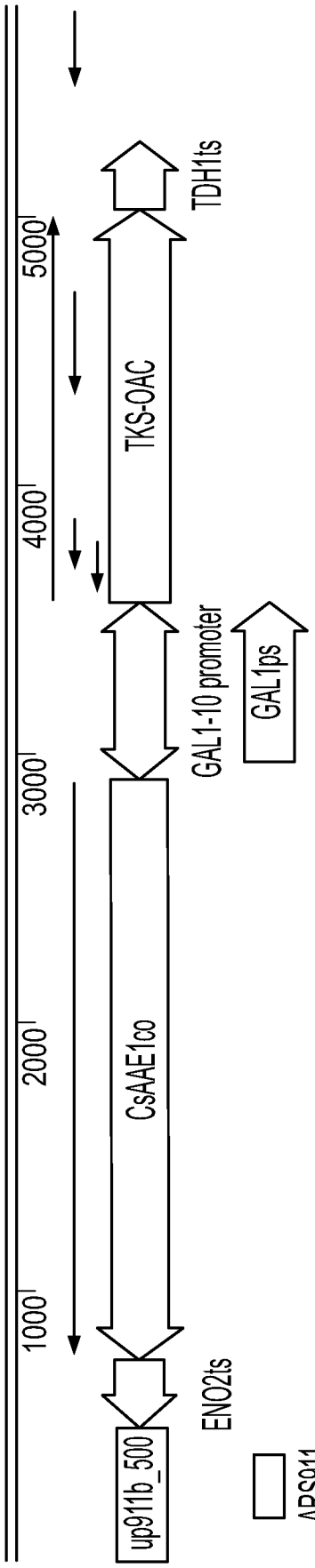
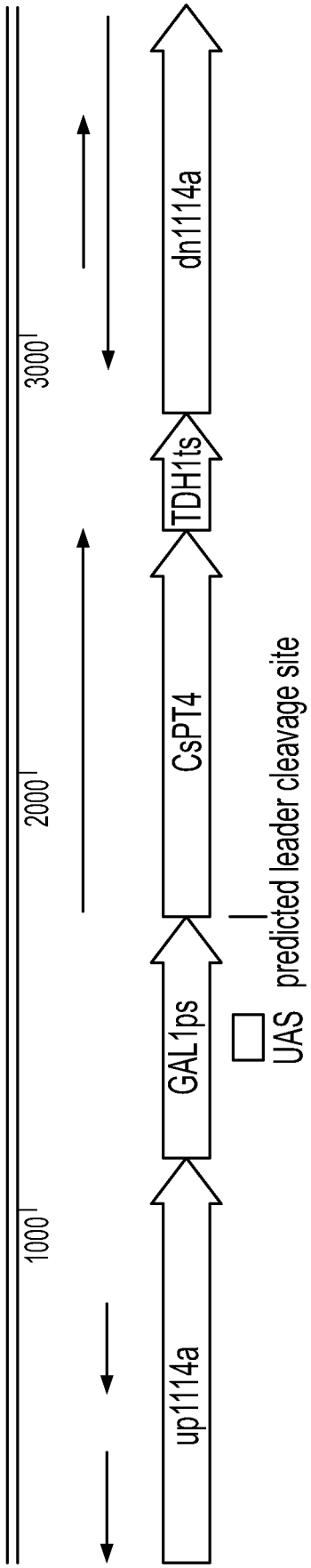

2500
5000
7500
10,000
12,500
FAS1
FAS2
SSA1t
ADH1t
up607b_500
GAL1-10 promoter
UAS
R1834K
I306A
G1250S
607b--cFAS1*-cFAS2*

FIG. 18



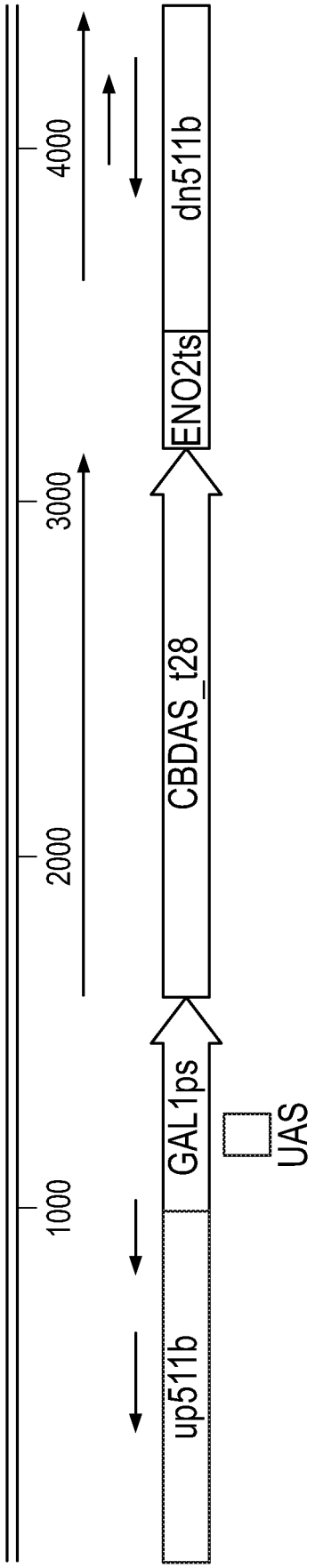
911b--cAAE1--cTKS-OAC

FIG. 19



1114a--tCsPT4

FIG. 20

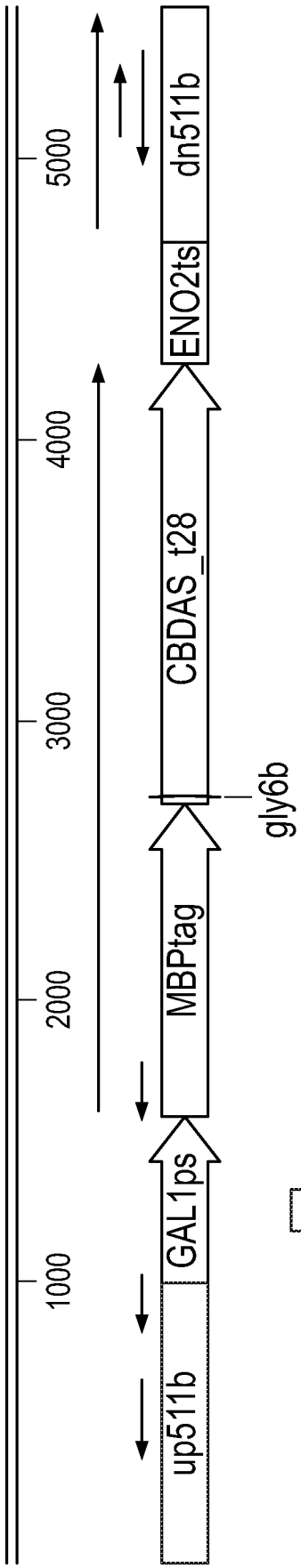


511b-GAL1ps-CBDASt28-ENO2ts

FIG. 21

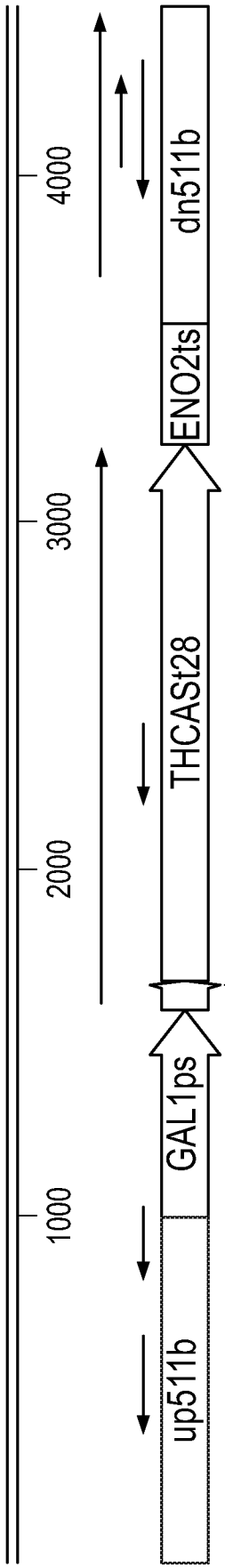
511b--GAL1ps-ProA-CBDAS-ENO2ts

FIG. 22



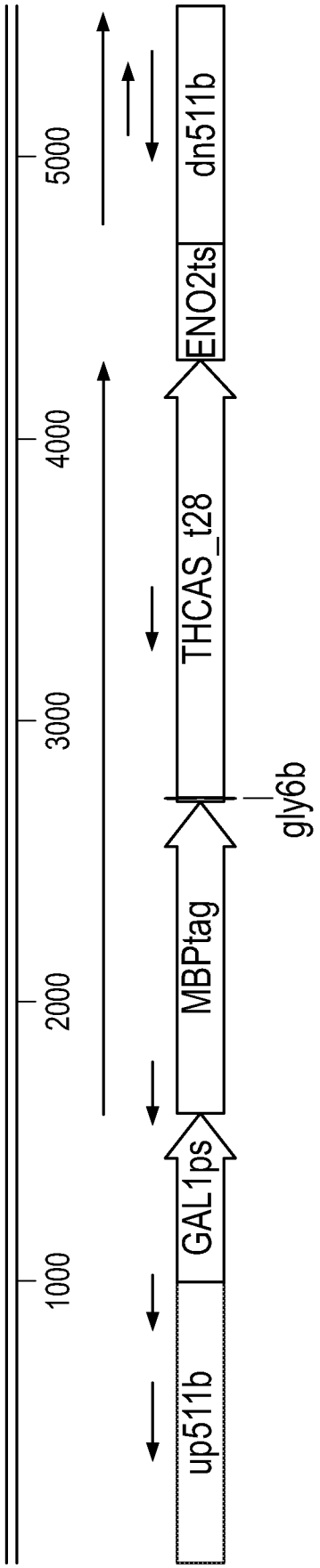
511b--GAL1ps-MBP-CBDAS-ENO2ts

FIG. 23



511b--GAL1ps-ProA-THCAS-ENO2ts

FIG. 24



511b--GAL1ps-MBP-THCAS-ENO2ts

FIG. 25

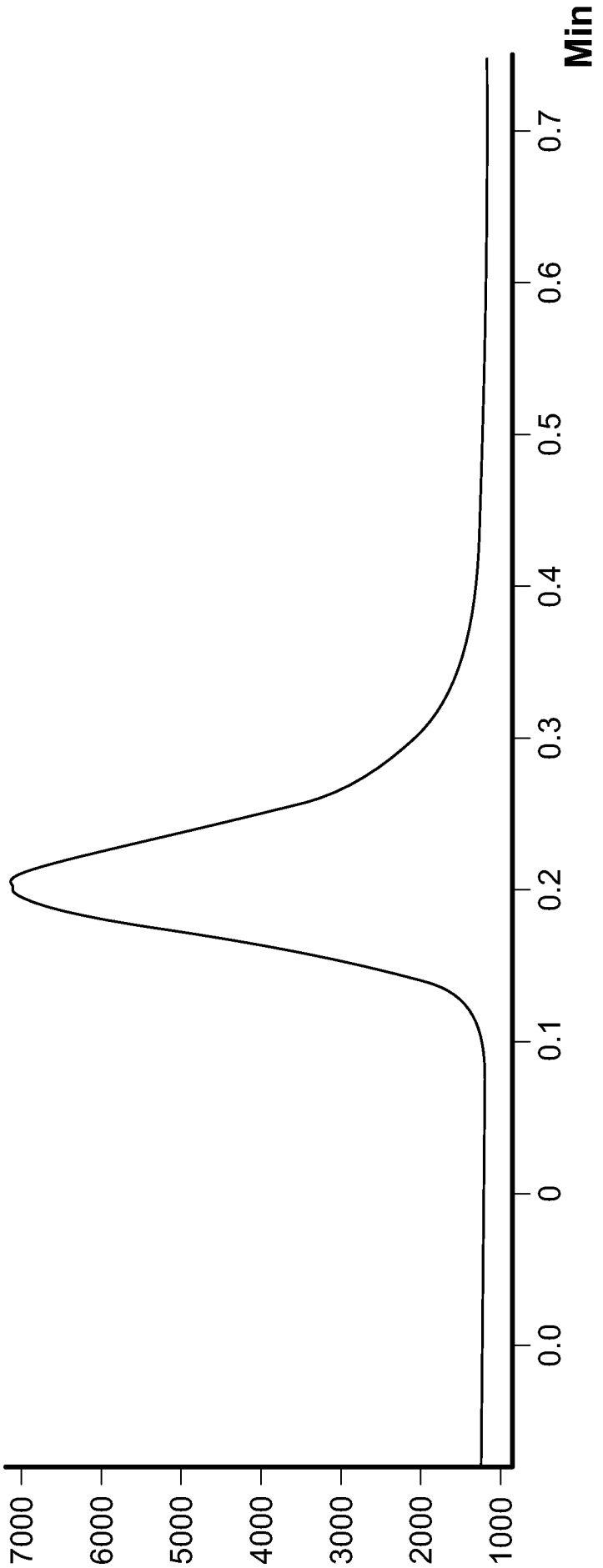


FIG. 26A

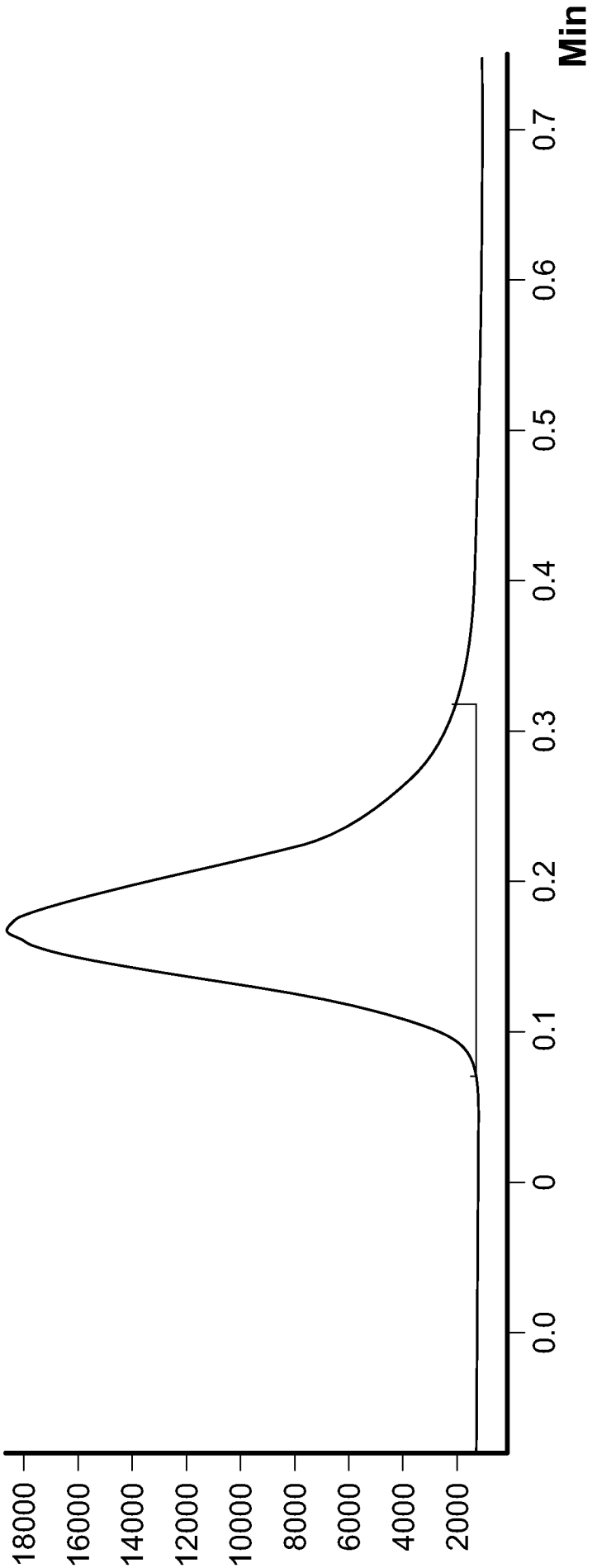

Min
0.0
0.1
0
0.2
0.3
0.4
0.5
0.6
0.7
2000
4000
6000
8000
10000
12000
14000
16000
18000

FIG. 26B

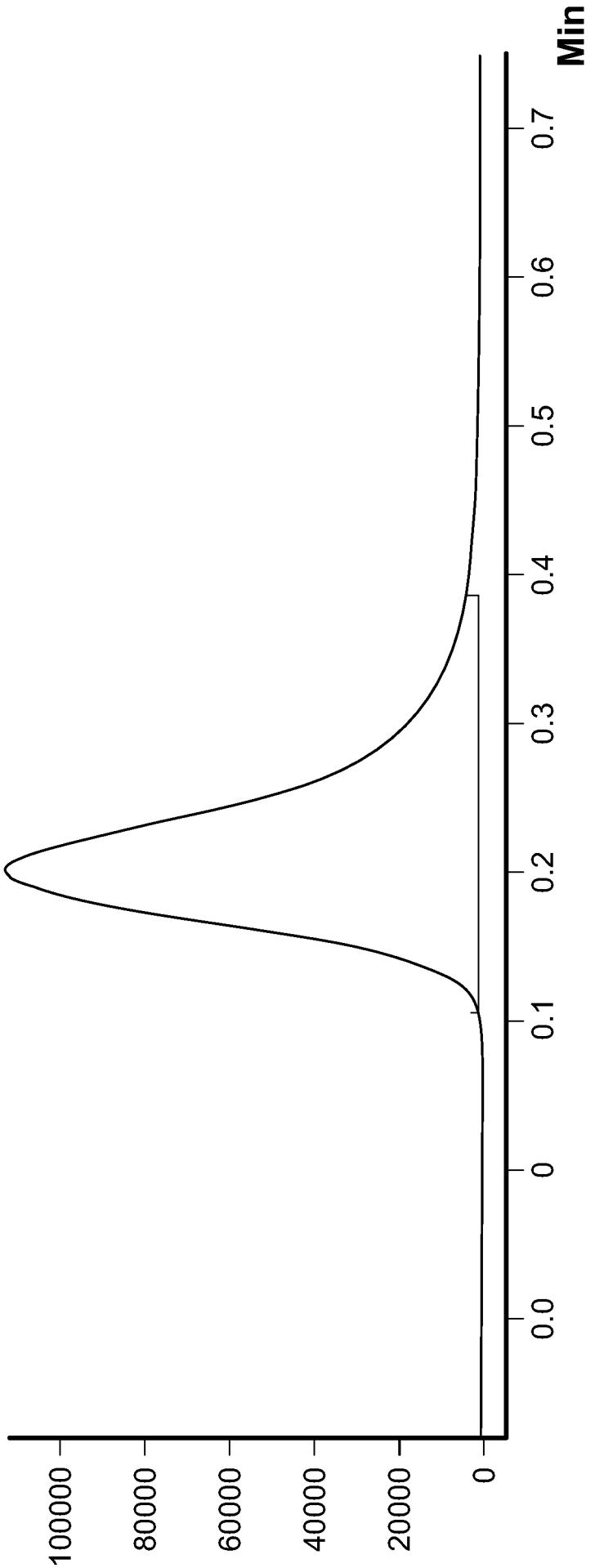


FIG. 26C

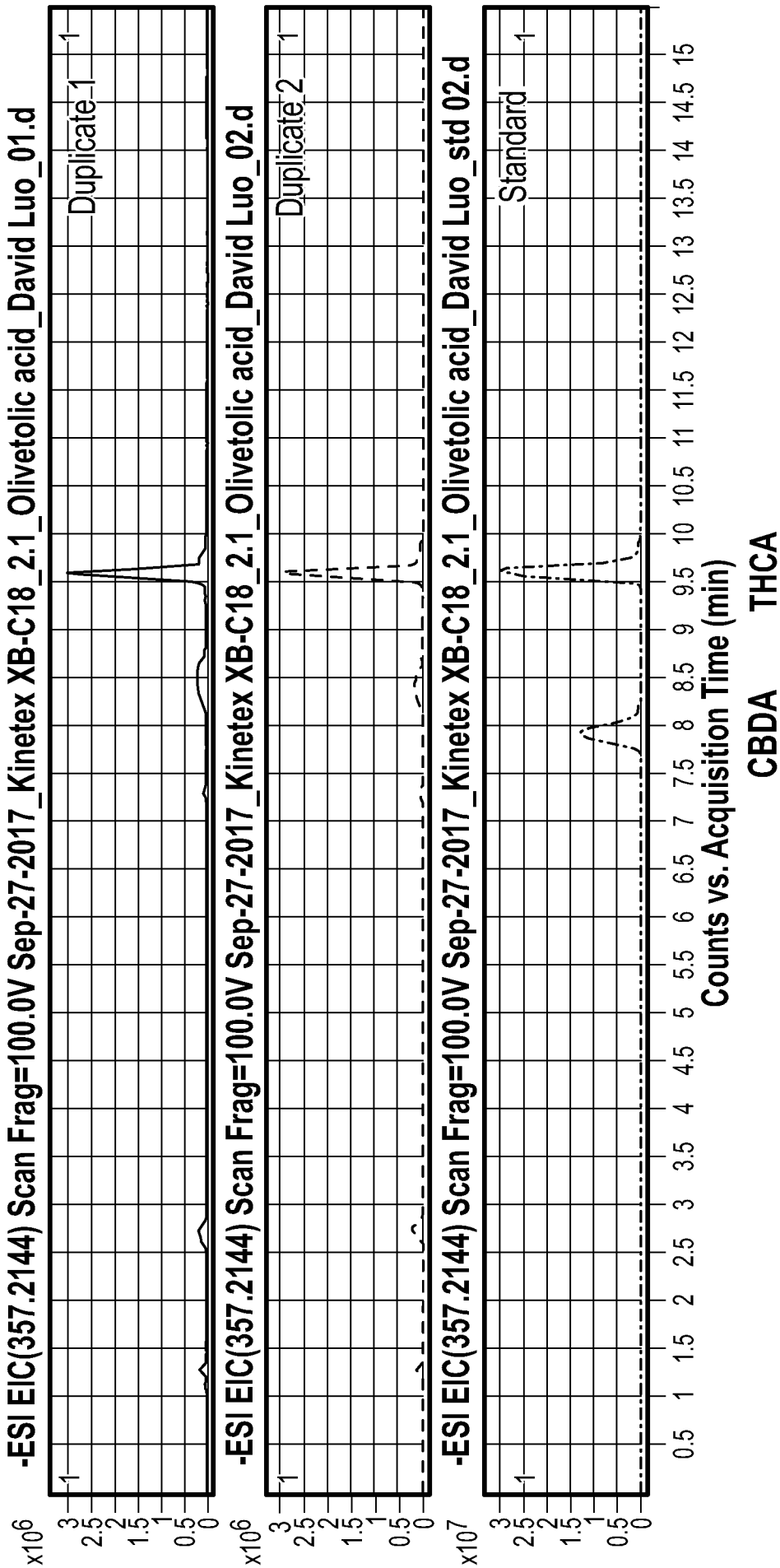

THCA synthase with N-terminal truncation and ProA signal sequence
-ESI EIC(357.2144) Scan Frag=100.0V Sep-27-2017_Kinetex XB-C18_2.1_Olivetolic acid_David Luo_01.d
Duplicate 1
-ESI EIC(357.2144) Scan Frag=100.0V Sep-27-2017_Kinetex XB-C18_2.1_Olivetolic acid_David Luo_02.d
Duplicate 2
-ESI EIC(357.2144) Scan Frag=100.0V Sep-27-2017_Kinetex XB-C18_2.1_Olivetolic acid_David Luo_std 02.d
Standard
x10^6
x10^6
x10^7
Counts vs. Acquisition Time (min)
CBDA
THCA

FIG. 27

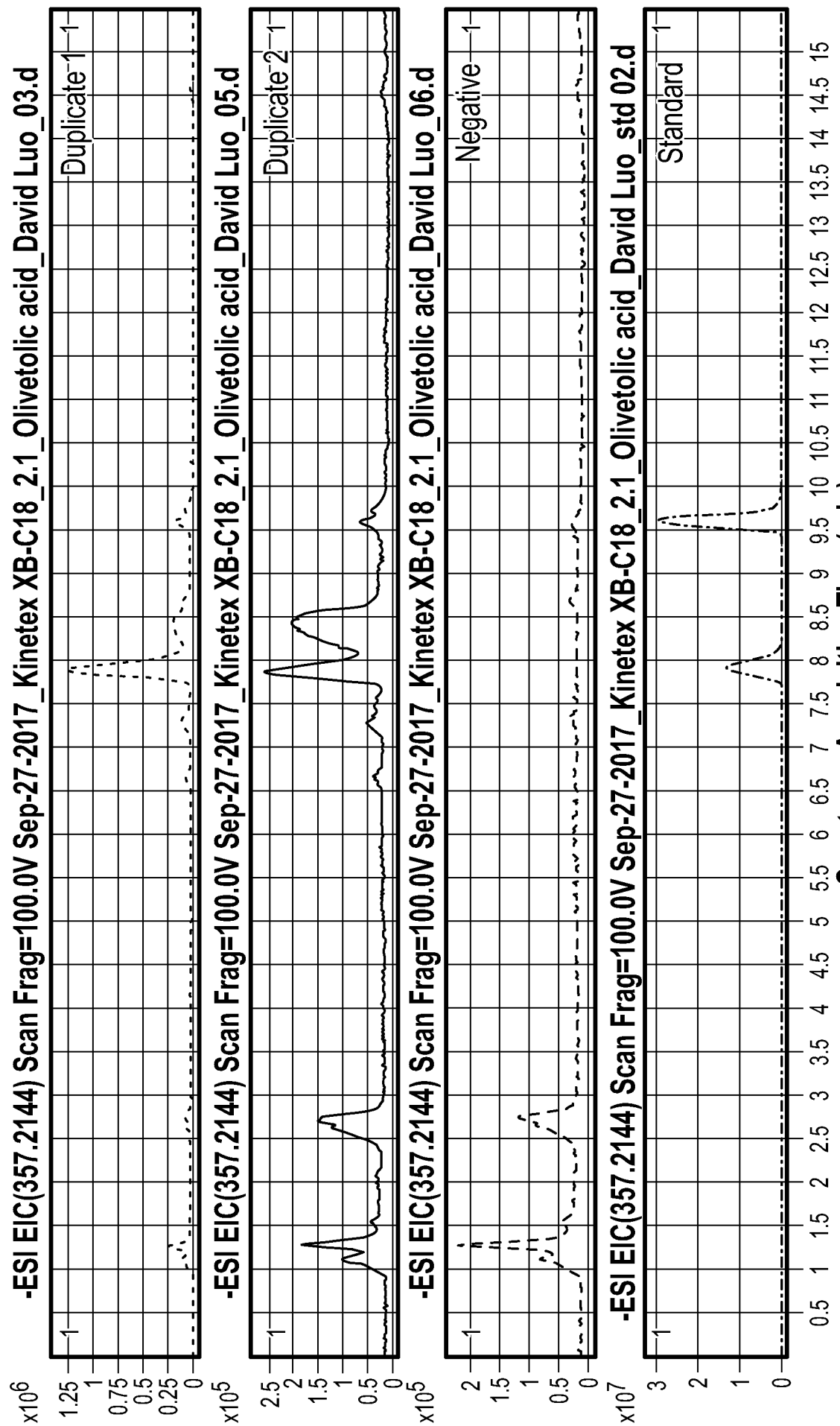

CBDA synthase with N-terminal truncation and ProA signal sequence
-ESI EIC(357.2144) Scan Frag=100.0V Sep-27-2017_Kinetex XB-C18_2.1_Olivetolic acid_David Luo_03.d
Duplicate 1
-ESI EIC(357.2144) Scan Frag=100.0V Sep-27-2017_Kinetex XB-C18_2.1_Olivetolic acid_David Luo_05.d
Duplicate 2
-ESI EIC(357.2144) Scan Frag=100.0V Sep-27-2017_Kinetex XB-C18_2.1_Olivetolic acid_David Luo_06.d
Negative
-ESI EIC(357.2144) Scan Frag=100.0V Sep-27-2017_Kinetex XB-C18_2.1_Olivetolic acid_David Luo_std 02.d
Standard
Counts vs. Acquisition Time (min)
CBDA
THCA

FIG. 28

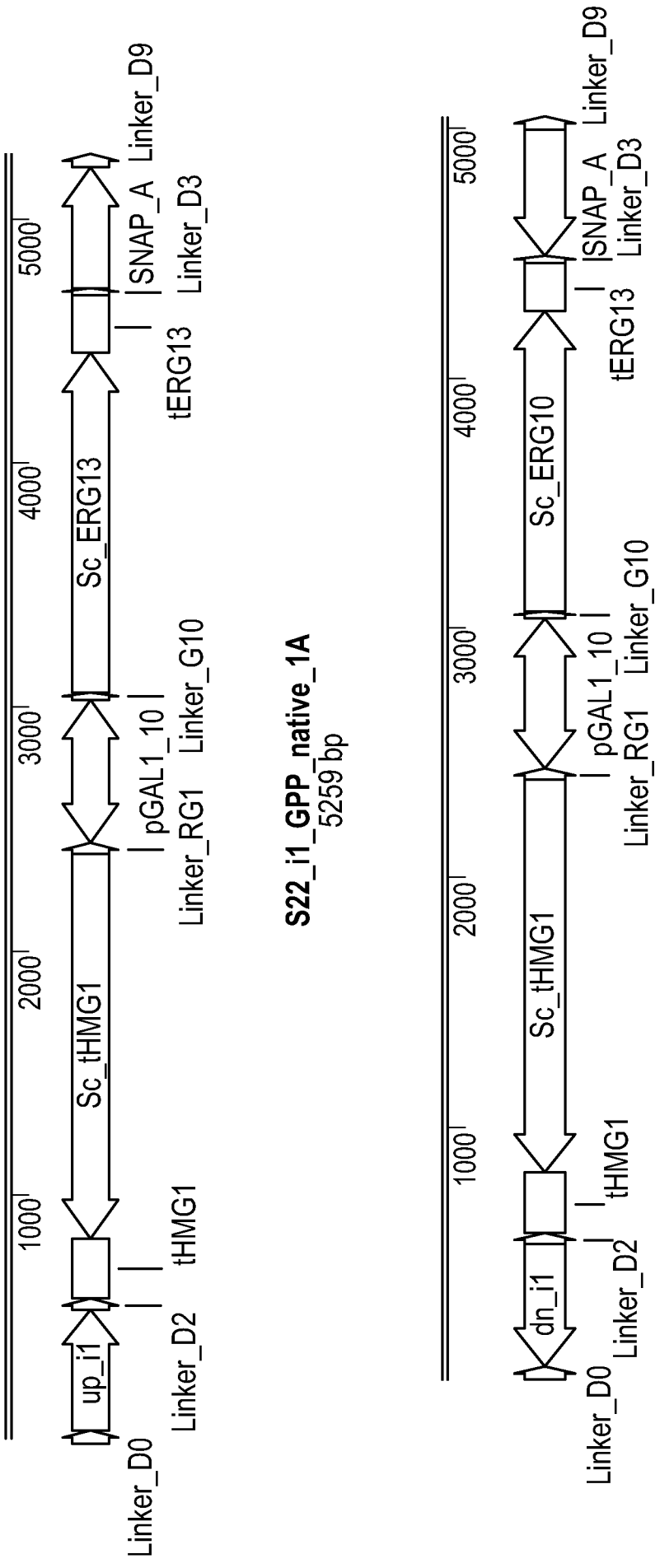

1000
2000
3000
4000
5000
Linker_D0
up_i1
Linker_D2
tHMG1
Sc_tHMG1
Linker_RG1
pGAL1_10
Linker_G10
Sc_ERG13
tERG13
Linker_D3
SNAP_A
Linker_D9
S22_i1_GPP_native_1A
5259 bp
1000
2000
3000
4000
5000
Linker_D0
dn_i1
Linker_D2
tHMG1
Sc_tHMG1
Linker_RG1
pGAL1_10
Linker_G10
Sc_ERG10
tERG13
Linker_D3
SNAP_A
Linker_D9
S23_i1_GPP_native_1B
5028 bp

FIG. 29A

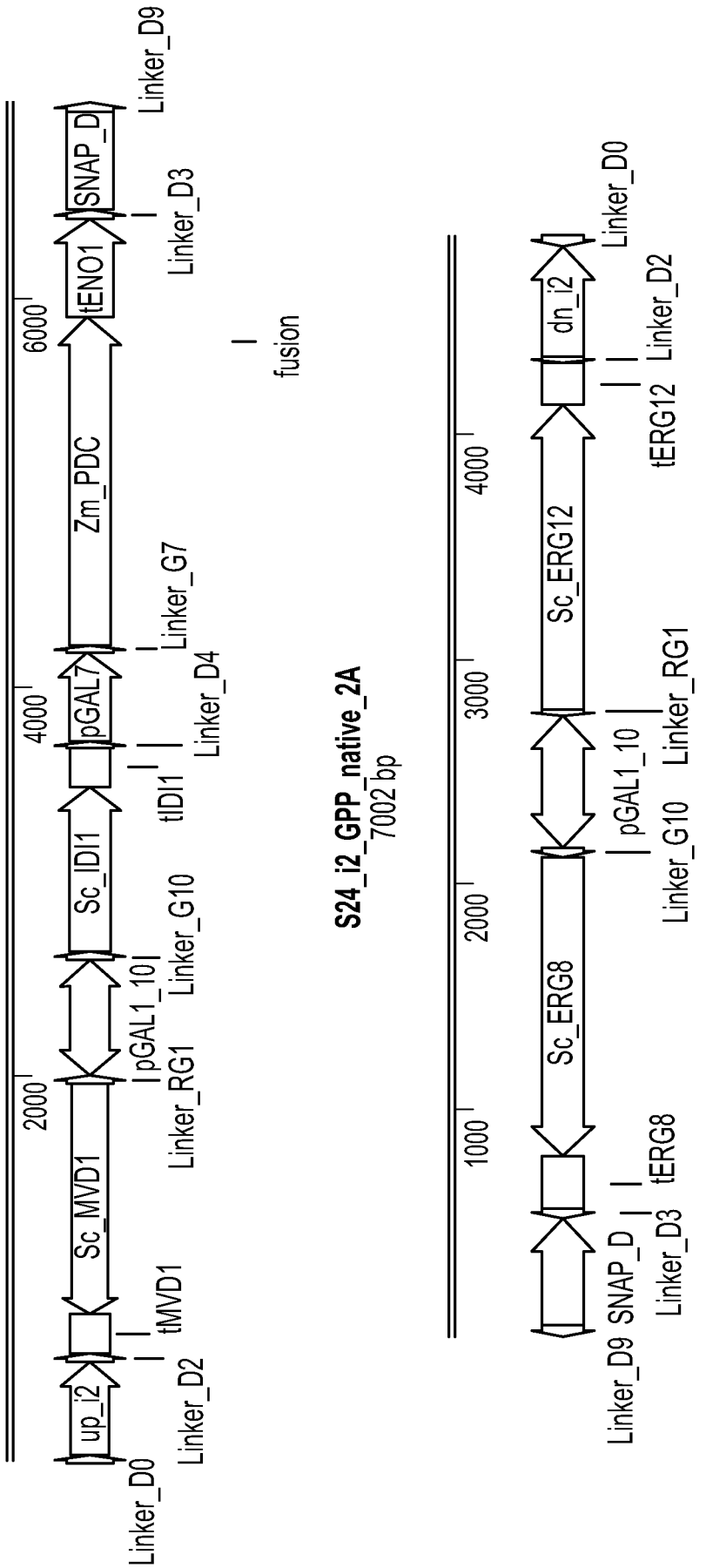

Linker_D0
up_i2
Linker_D2
tMVD1
Sc_MVD1
Linker_RG1
pGAL1_10
Linker_G10
Sc_IDI1
tIDI1
Linker_D4
pGAL7
Linker_G7
Zm_PDC
tENO1
Linker_D3
SNAP_D
Linker_D9
2000
4000
6000
fusion
S24_i2_GPP_native_2A
7002 bp
Linker_D9
SNAP_D
Linker_D3
tERG8
Sc_ERG8
Linker_G10
pGAL1_10
Linker_RG1
Sc_ERG12
tERG12
Linker_D2
dn_i2
Linker_D0
1000
2000
3000
4000
S25_i2_GPP_native_2B
4897 bp

FIG. 29B

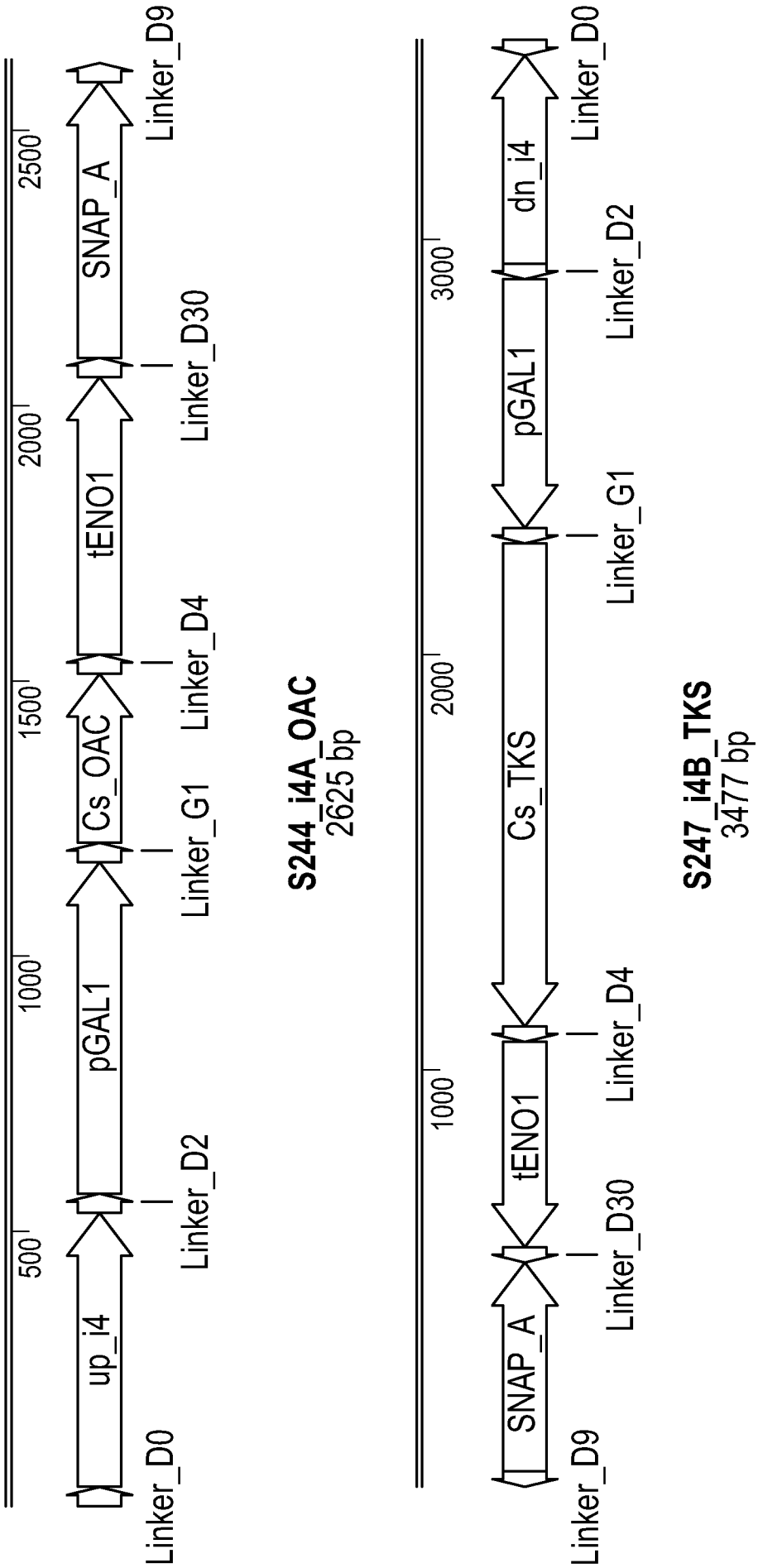

500
1000
1500
2000
2500
Linker_D0
up_i4
Linker_D2
pGAL1
Linker_G1
Cs_OAC
Linker_D4
tENO1
Linker_D30
SNAP_A
Linker_D9
S244_i4A_OAC
2625 bp
1000
2000
3000
Linker_D9
SNAP_A
Linker_D30
tENO1
Linker_D4
Cs_TKS
Linker_G1
pGAL1
Linker_D2
dn_i4
Linker_D0
S247_i4B_TKS
3477 bp

FIG. 30A

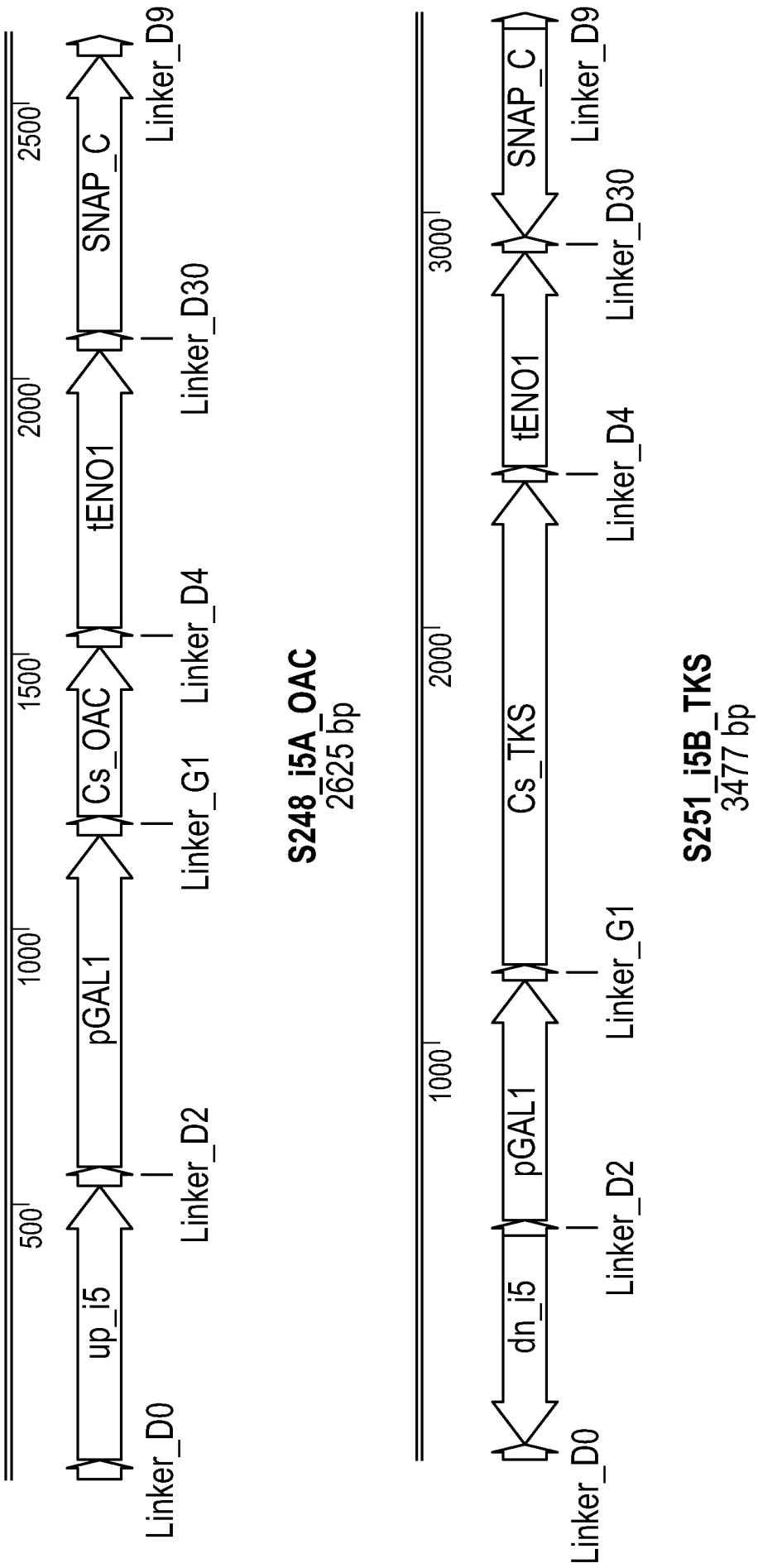

500
1000
1500
2000
2500
Linker_D0
up_i5
Linker_D2
pGAL1
Linker_G1
Cs_OAC
Linker_D4
tENO1
Linker_D30
SNAP_C
Linker_D9
S248_i5A_OAC
2625 bp
1000
2000
3000
Linker_D0
dn_i5
Linker_D2
pGAL1
Linker_G1
Cs_TKS
Linker_D4
tENO1
Linker_D30
SNAP_C
Linker_D9
S251_i5B_TKS
3477 bp

FIG. 30B

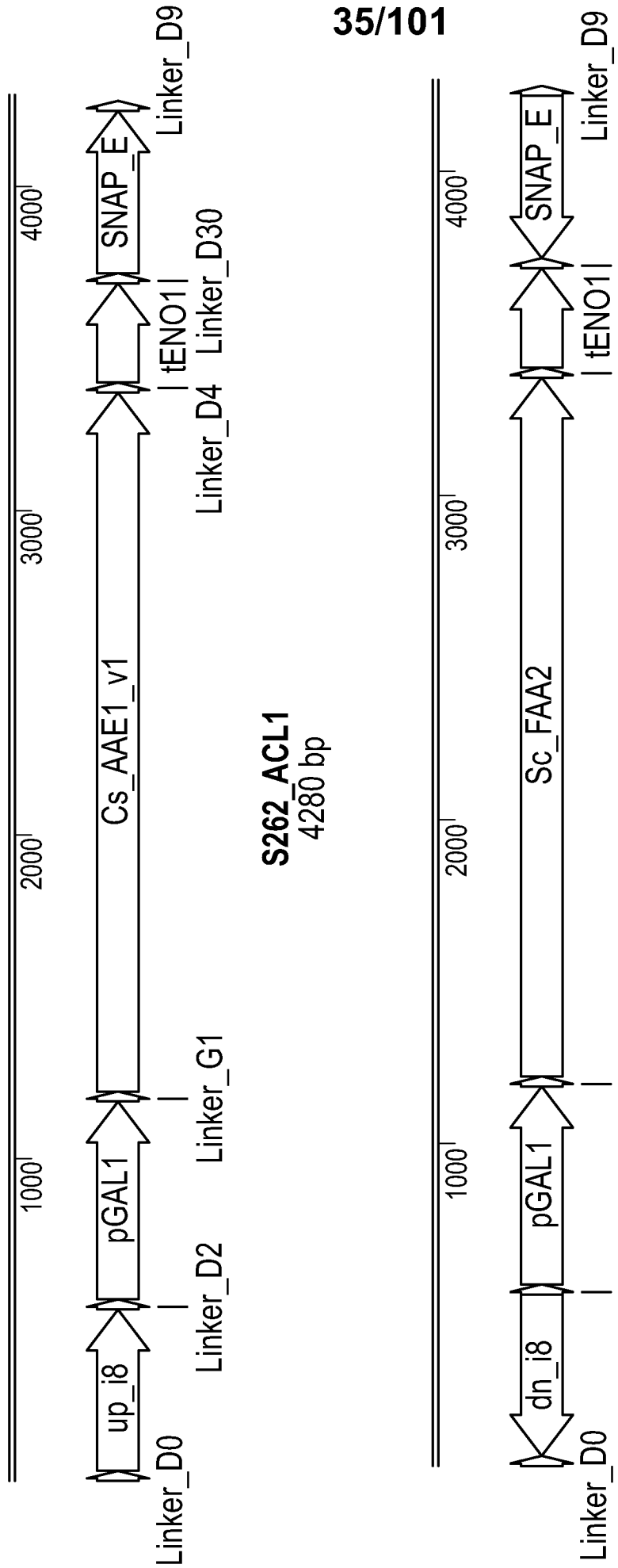

1000
2000
3000
4000
Linker_D0
up_i8
Linker_D2
pGAL1
Linker_G1
Cs_AAE1_v1
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S262_ACL1
4280 bp
1000
2000
3000
4000
Linker_D0
dn_i8
Linker_D2
pGAL1
Linker_G1
Sc_FAA2
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S281_ACL20
4352 bp

FIG. 30C

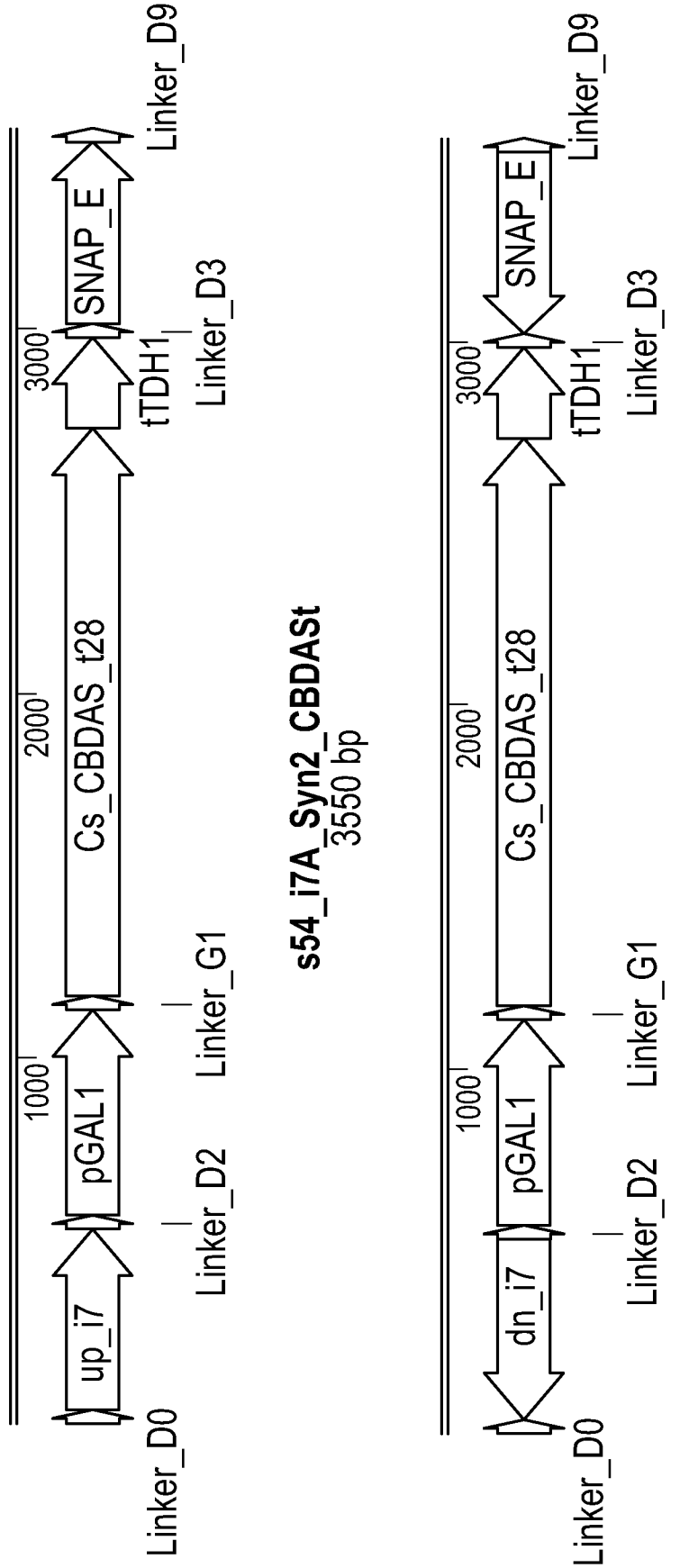
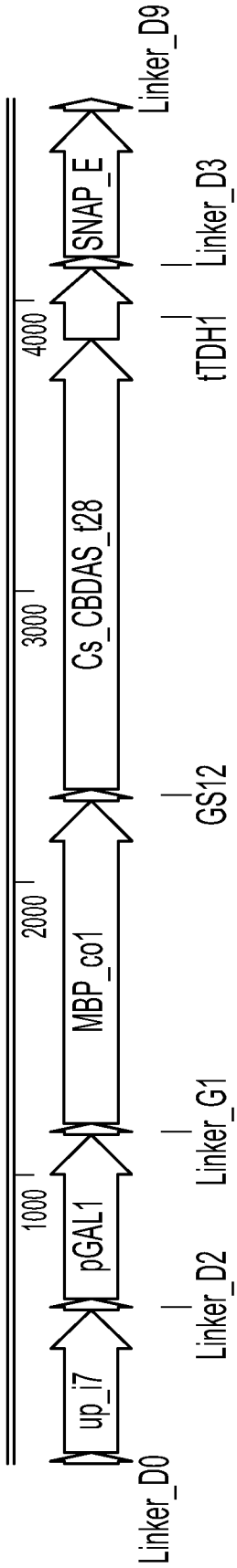

Linker_D0
up_i7
Linker_D2
pGAL1
Linker_G1
Cs_CBDAS_t28
tTDH1
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s54_i7A_Syn2_CBDASt
3550 bp
Linker_D0
dn_i7
Linker_D2
pGAL1
Linker_G1
Cs_CBDAS_t28
tTDH1
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s68_i7B_Syn2_CBDASt
3550 bp

FIG. 31


Linker_D0
up_i7
Linker_D2
pGAL1
Linker_G1
MBP_co1
GS12
Cs_CBDAS_t28
tTDH1
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
4000
s56_i7A_Syn4_MBP-CBDASt
4693 bp

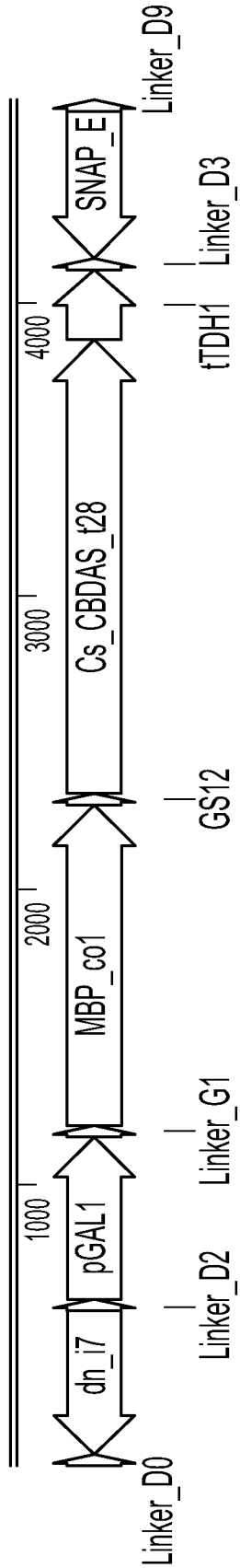

Linker_D0
dn_i7
Linker_D2
pGAL1
Linker_G1
MBP_co1
GS12
Cs_CBDAS_t28
tTDH1
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
4000
s70_i7B_Syn4_MBP-CBDASt
4693 bp

FIG. 32

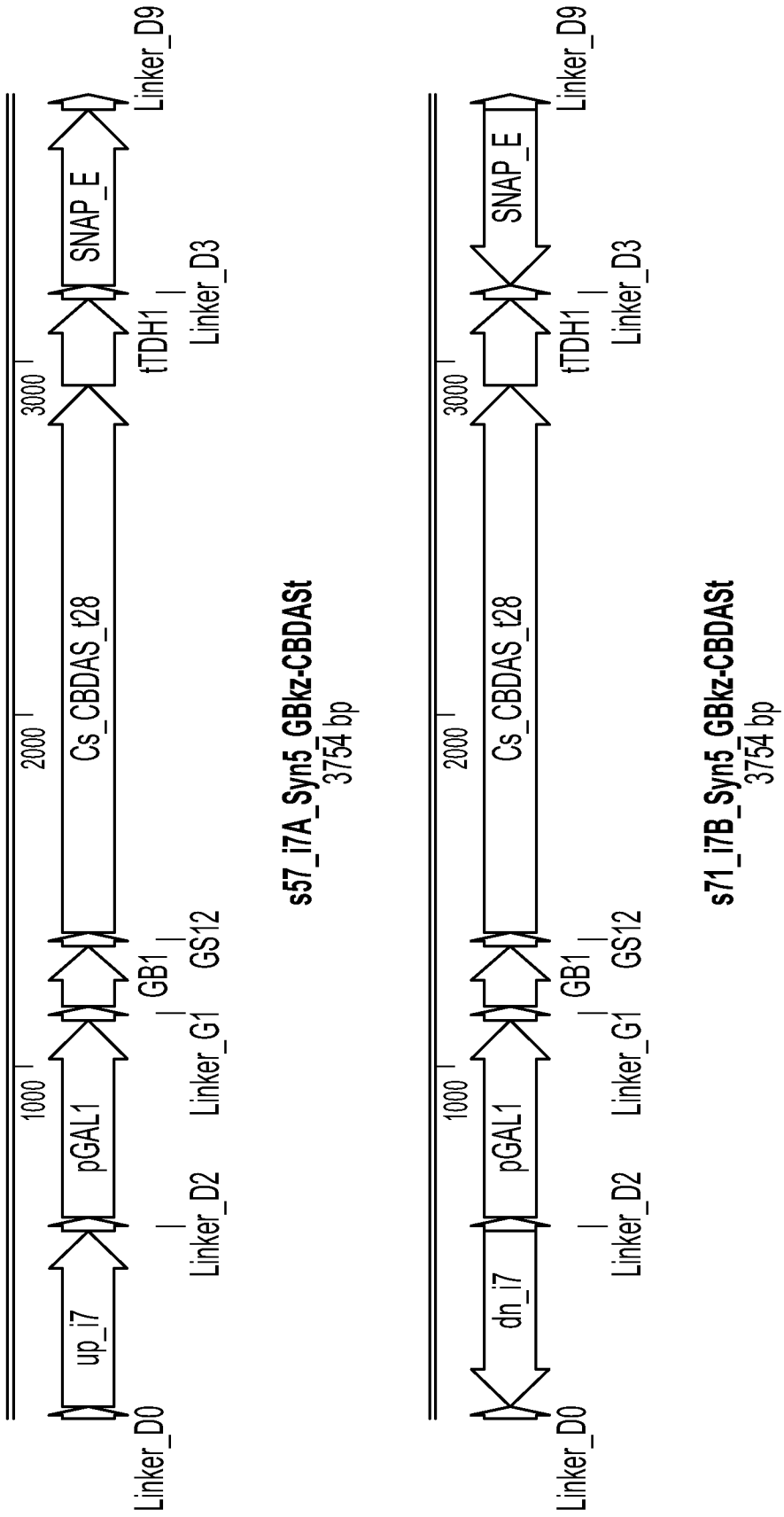
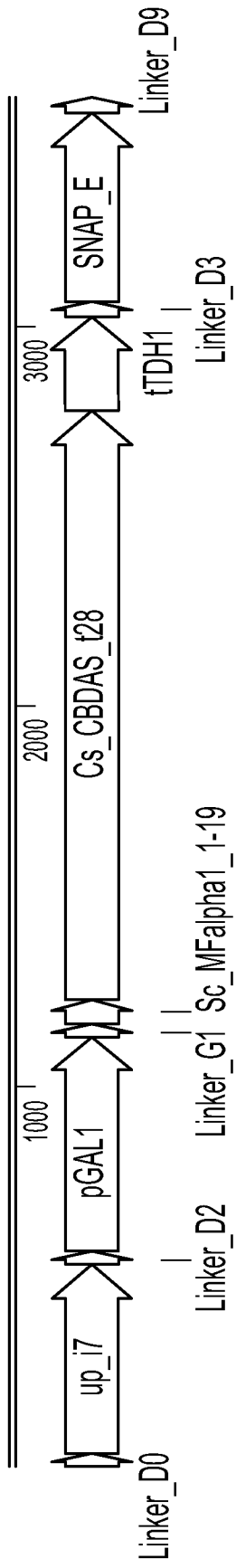
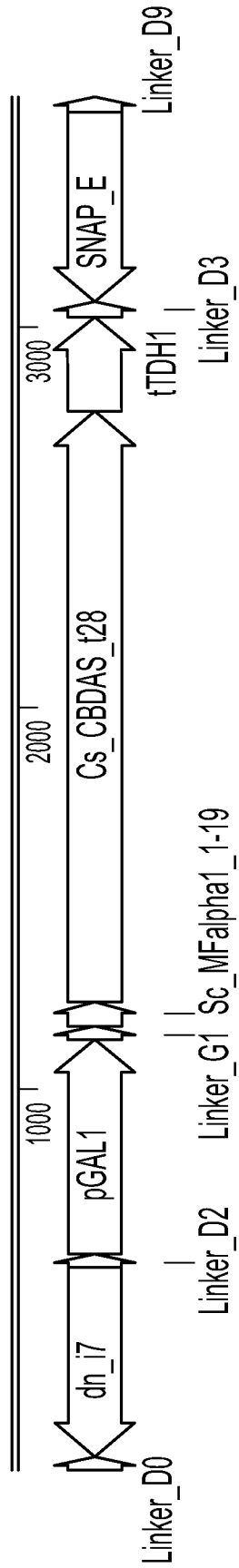

Linker_D0
up_i7
Linker_D2
pGAL1
Linker_G1
GB1
GS12
Cs_CBDAS_t28
tTDH1
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s57_i7A_Syn5_GBkz-CBDASt
3754 bp
Linker_D0
dn_i7
Linker_D2
pGAL1
Linker_G1
GB1
GS12
Cs_CBDAS_t28
tTDH1
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s71_i7B_Syn5_GBkz-CBDASt
3754 bp

FIG. 33



s58_i7A_Syn7_MFer-CBDASt
3601 bp



s72_i7B_Syn7_MFer-CBDASt
3601 bp

FIG. 34

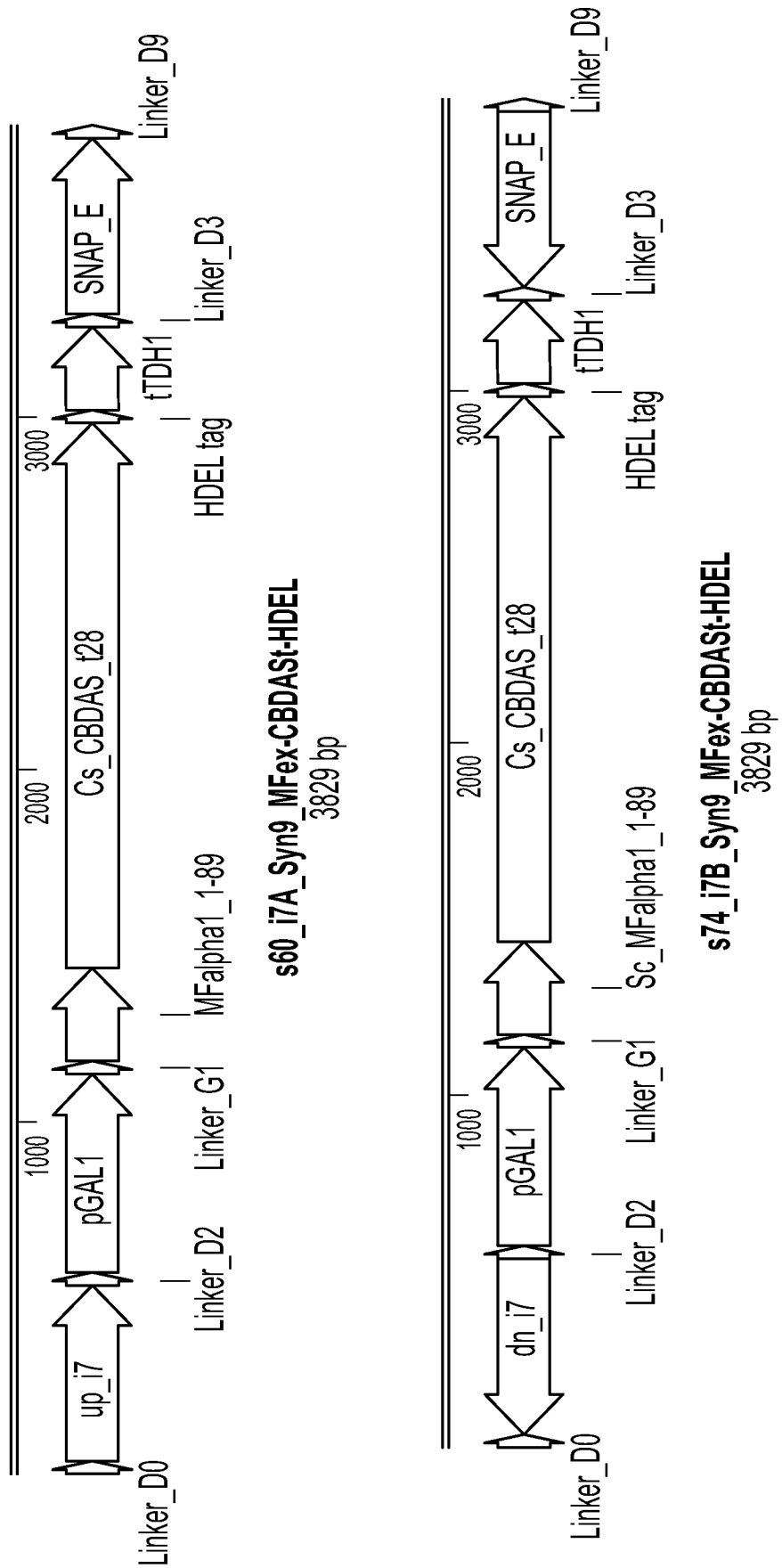

Linker_D0
up_i7
Linker_D2
pGAL1
Linker_G1
MFalpha1_1-89
Cs_CBDAS_t28
HDEL tag
tTDH1
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s60_i7A_Syn9_MFex-CBDASt-HDEL
3829 bp
Linker_D0
dn_i7
Linker_D2
pGAL1
Linker_G1
Sc_MFalpha1_1-89
Cs_CBDAS_t28
HDEL tag
tTDH1
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s74_i7B_Syn9_MFex-CBDASt-HDEL
3829 bp

FIG. 35

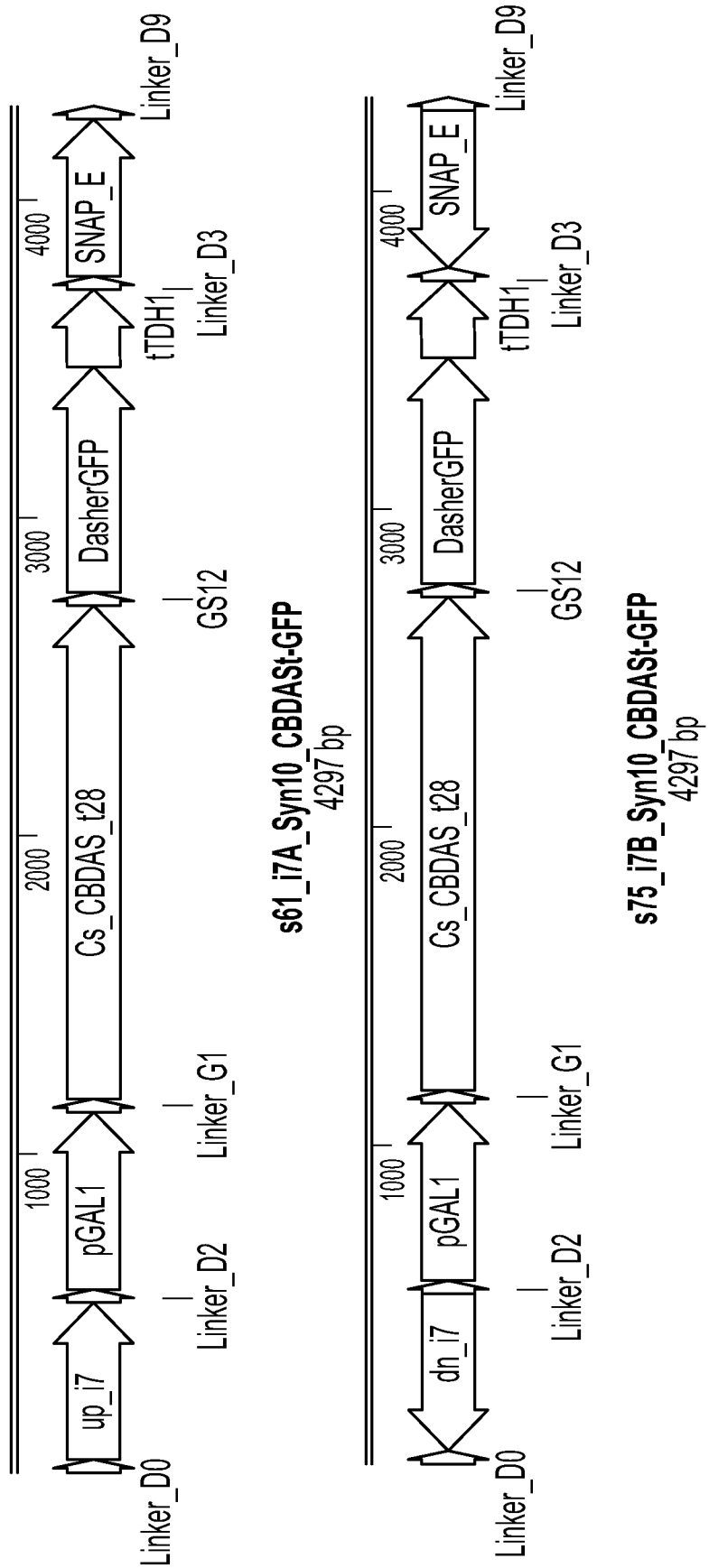

1000
2000
3000
4000
Linker_D0
up_i7
Linker_D2
pGAL1
Linker_G1
Cs_CBDAS_t28
GS12
DasherGFP
tTDH1
Linker_D3
SNAP_E
Linker_D9
s61_i7A_Syn10_CBDASt-GFP
4297 bp
Linker_D0
dn_i7
Linker_D2
pGAL1
Linker_G1
Cs_CBDAS_t28
GS12
DasherGFP
tTDH1
Linker_D3
SNAP_E
Linker_D9
s75_i7B_Syn10_CBDASt-GFP
4297 bp

FIG. 36

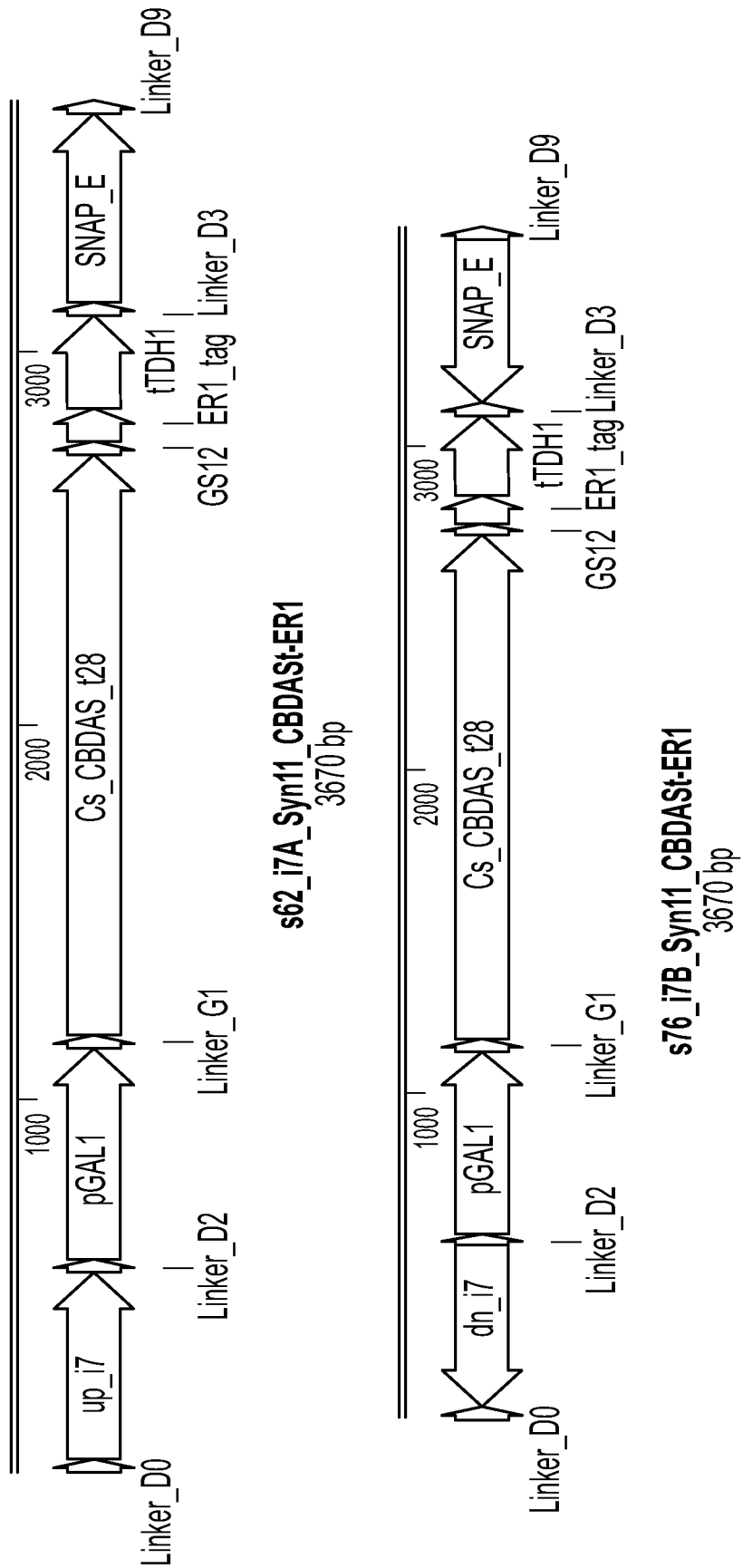

Linker_D0
up_i7
Linker_D2
pGAL1
1000
Linker_G1
Cs_CBDAS_t28
2000
GS12
ER1_tag
tTDH1
3000
Linker_D3
SNAP_E
Linker_D9
s62_i7A_Syn11_CBDASt-ER1
3670 bp
Linker_D0
dn_i7
Linker_D2
pGAL1
1000
Linker_G1
Cs_CBDAS_t28
2000
GS12
ER1_tag
tTDH1
3000
Linker_D3
SNAP_E
Linker_D9
s76_i7B_Syn11_CBDASt-ER1
3670 bp

FIG. 37

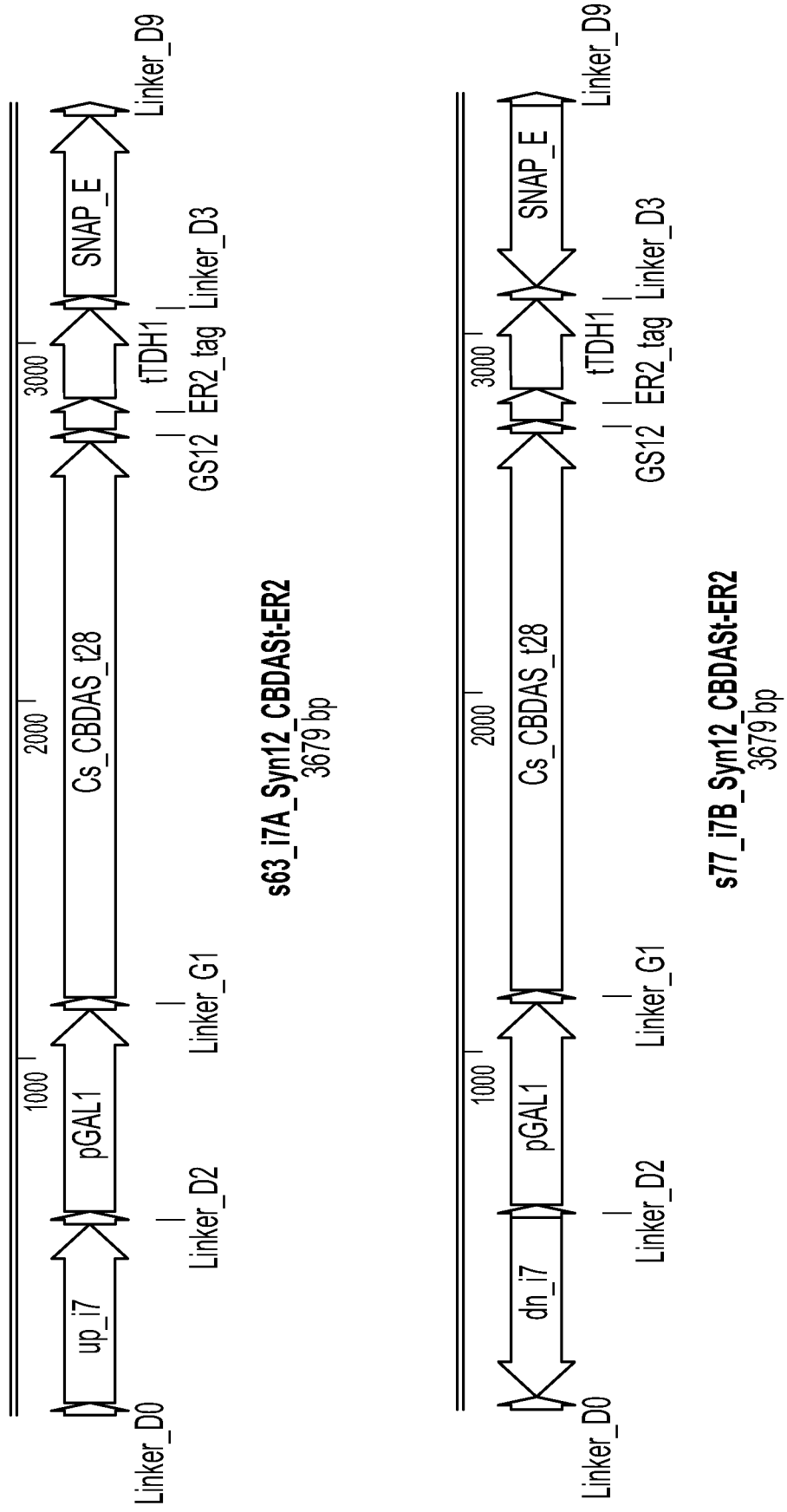

Linker_D0
up_i7
Linker_D2
pGAL1
Linker_G1
Cs_CBDAS_t28
GS12
tTDH1
ER2_tag
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s63_i7A_Syn12_CBDASt-ER2
3679 bp
Linker_D0
dn_i7
Linker_D2
pGAL1
Linker_G1
Cs_CBDAS_t28
GS12
tTDH1
ER2_tag
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s77_i7B_Syn12_CBDASt-ER2
3679 bp

FIG. 38

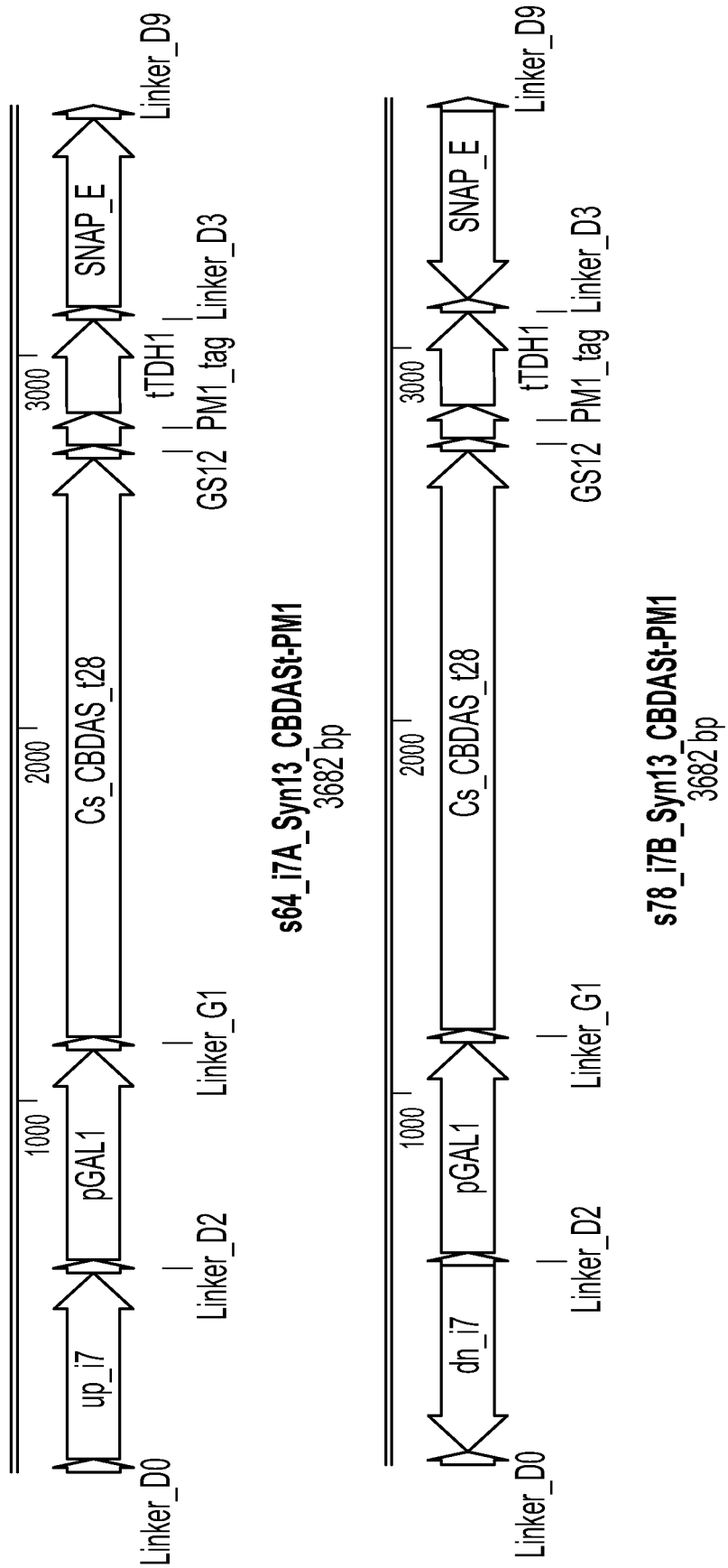

Linker_D0
up_i7
Linker_D2
pGAL1
Linker_G1
Cs_CBDAS_t28
tTDH1
GS12
PM1_tag
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s64_i7A_Syn13_CBDASt-PM1
3682 bp
Linker_D0
dn_i7
Linker_D2
pGAL1
Linker_G1
Cs_CBDAS_t28
tTDH1
GS12
PM1_tag
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s78_i7B_Syn13_CBDASt-PM1
3682 bp

FIG. 39

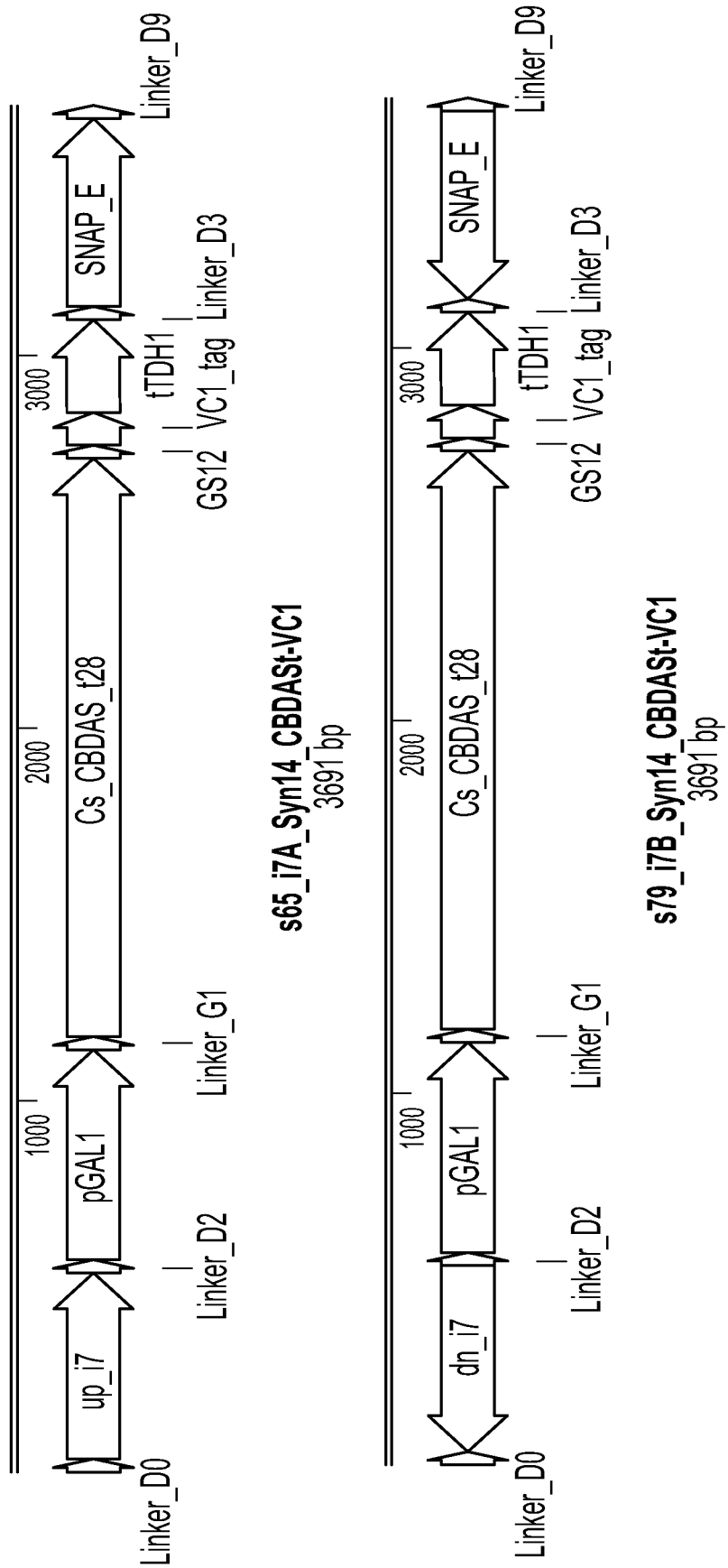

Linker_D0
up_i7
Linker_D2
pGAL1
Linker_G1
Cs_CBDAS_t28
tTDH1
GS12
VC1_tag
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s65_i7A_Syn14_CBDASt-VC1
3691 bp
Linker_D0
dn_i7
Linker_D2
pGAL1
Linker_G1
Cs_CBDAS_t28
tTDH1
GS12
VC1_tag
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s79_i7B_Syn14_CBDASt-VC1
3691 bp

FIG. 40

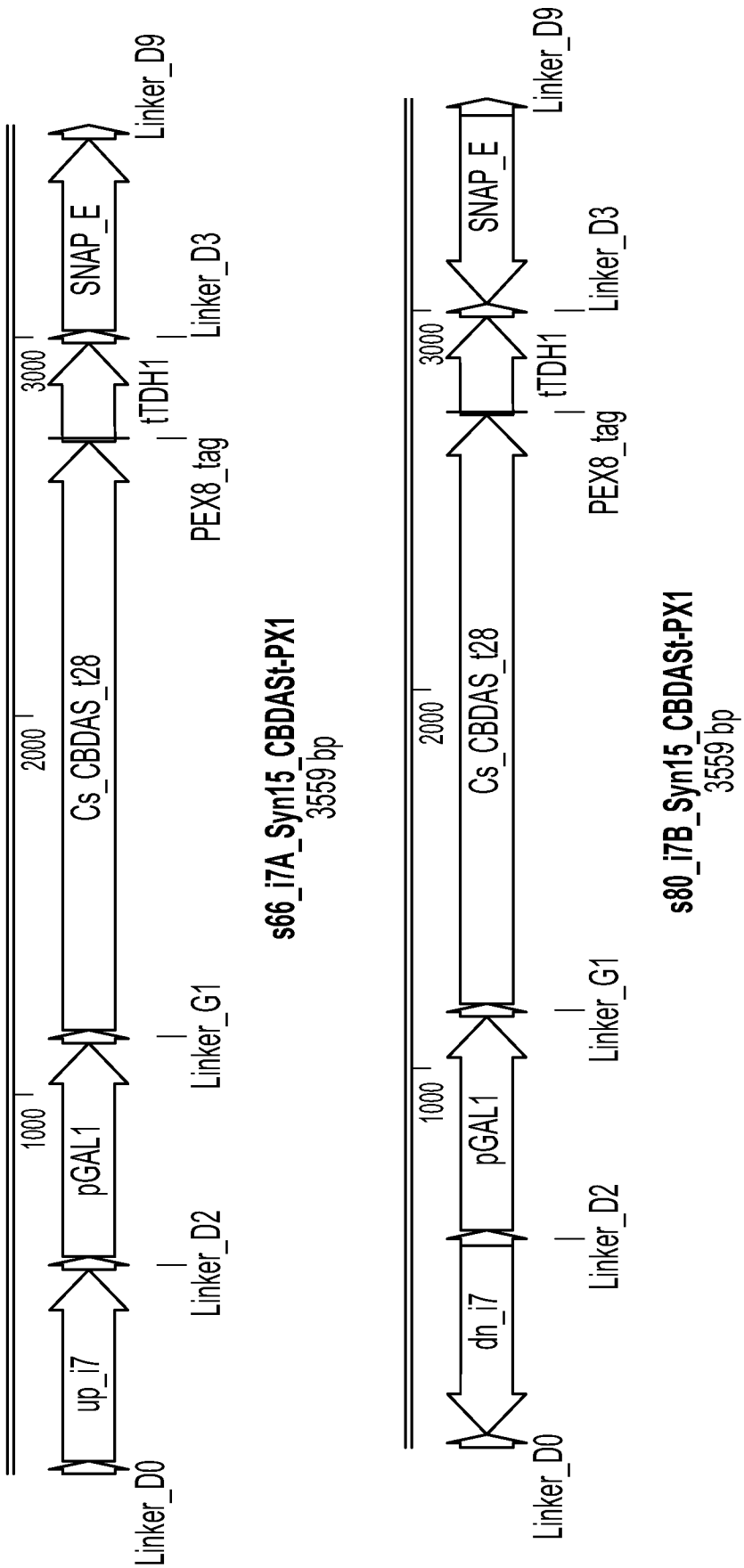

Linker_D0
up_i7
Linker_D2
pGAL1
Linker_G1
Cs_CBDAS_t28
PEX8_tag
tTDH1
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s66_i7A_Syn15_CBDASt-PX1
3559 bp
Linker_D0
dn_i7
Linker_D2
pGAL1
Linker_G1
Cs_CBDAS_t28
PEX8_tag
tTDH1
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s80_i7B_Syn15_CBDASt-PX1
3559 bp

FIG. 41

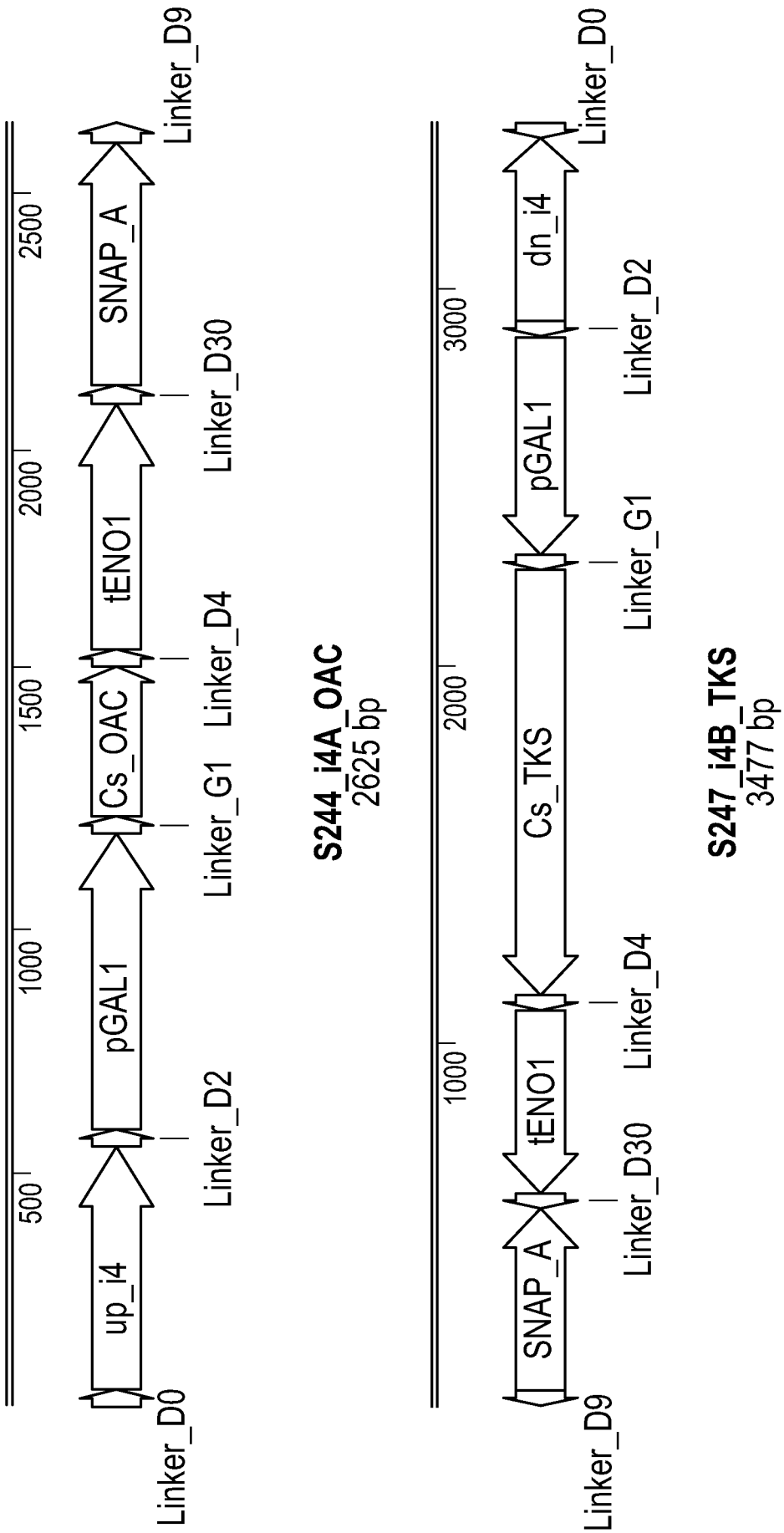

500
1000
1500
2000
2500
Linker_D0
up_i4
Linker_D2
pGAL1
Linker_G1
Cs_OAC
Linker_D4
tENO1
Linker_D30
SNAP_A
Linker_D9
S244_i4A_OAC
2625 bp
1000
2000
3000
Linker_D9
SNAP_A
Linker_D30
tENO1
Linker_D4
Cs_TKS
Linker_G1
pGAL1
Linker_D2
dn_i4
Linker_D0
S247_i4B_TKS
3477 bp

FIG. 42A

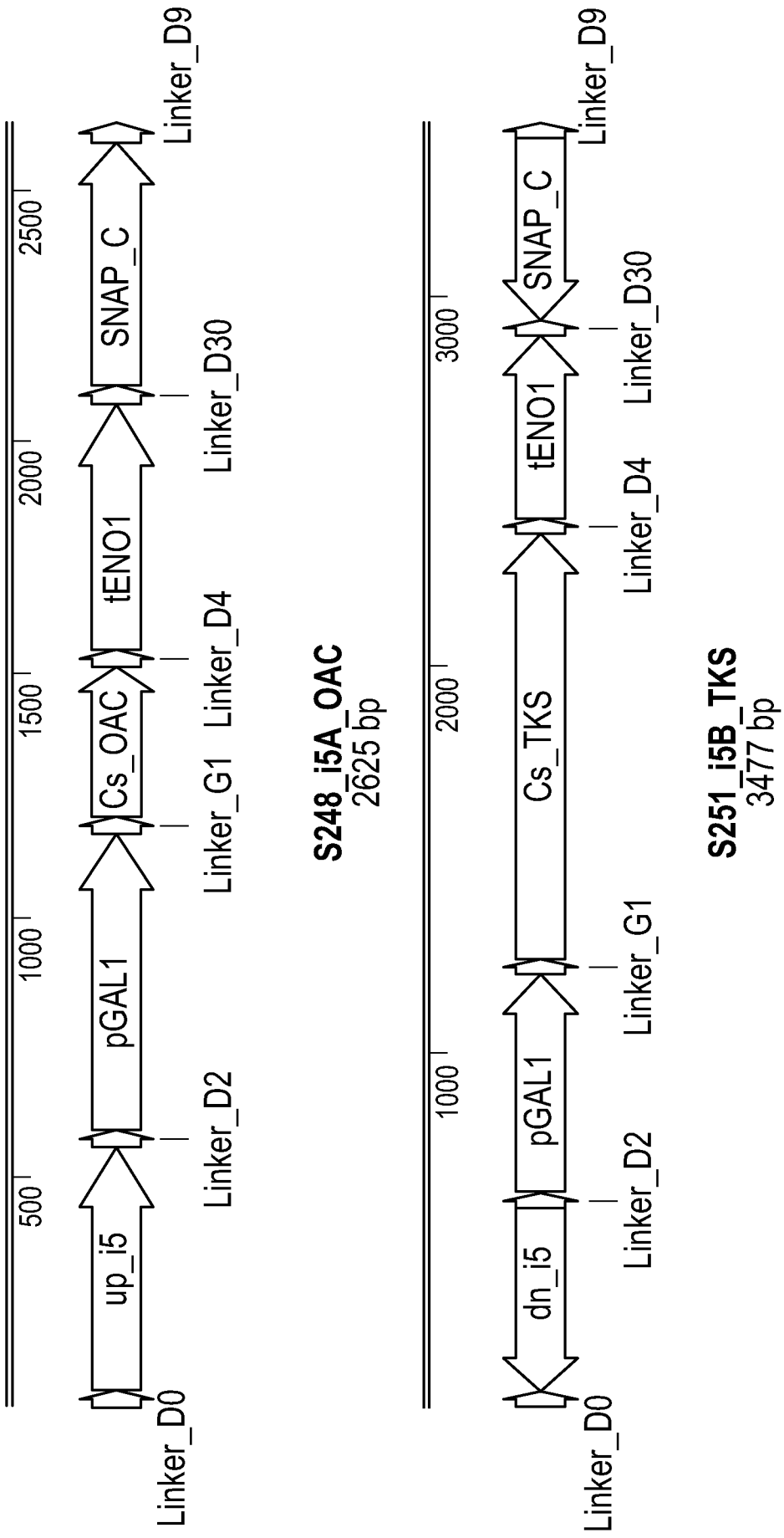

500
1000
1500
2000
2500
Linker_D0
up_i5
Linker_D2
pGAL1
Linker_G1
Cs_OAC
Linker_D4
tENO1
Linker_D30
SNAP_C
Linker_D9
S248_i5A_OAC
2625 bp
1000
2000
3000
Linker_D0
dn_i5
Linker_D2
pGAL1
Linker_G1
Cs_TKS
Linker_D4
tENO1
Linker_D30
SNAP_C
Linker_D9
S251_i5B_TKS
3477 bp

FIG. 42B

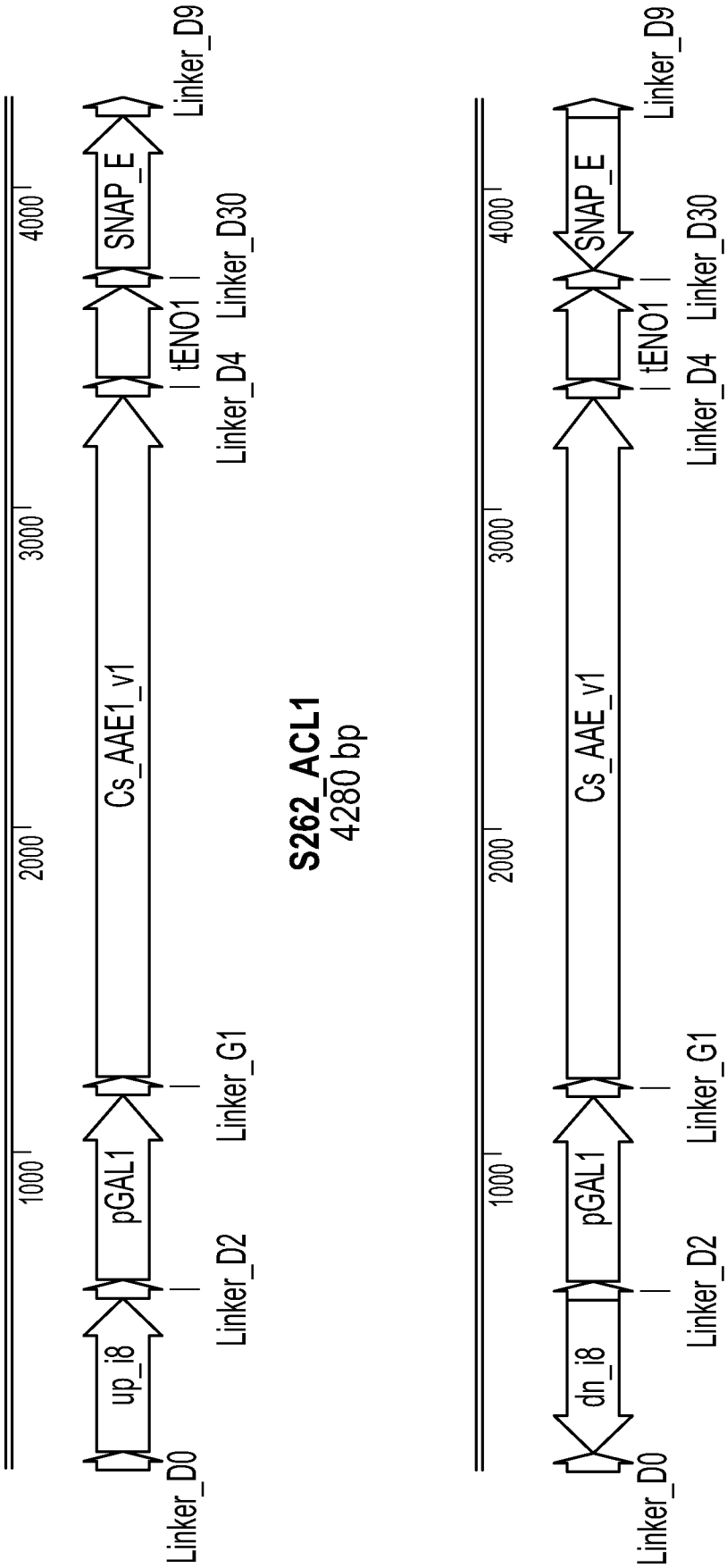

1000
2000
3000
4000
Linker_D0
up_i8
Linker_D2
pGAL1
Linker_G1
Cs_AAE1_v1
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S262_ACL1
4280 bp
1000
2000
3000
4000
Linker_D0
dn_i8
Linker_D2
pGAL1
Linker_G1
Cs_AAE_v1
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S274_ACL13
4280 bp

FIG. 42C

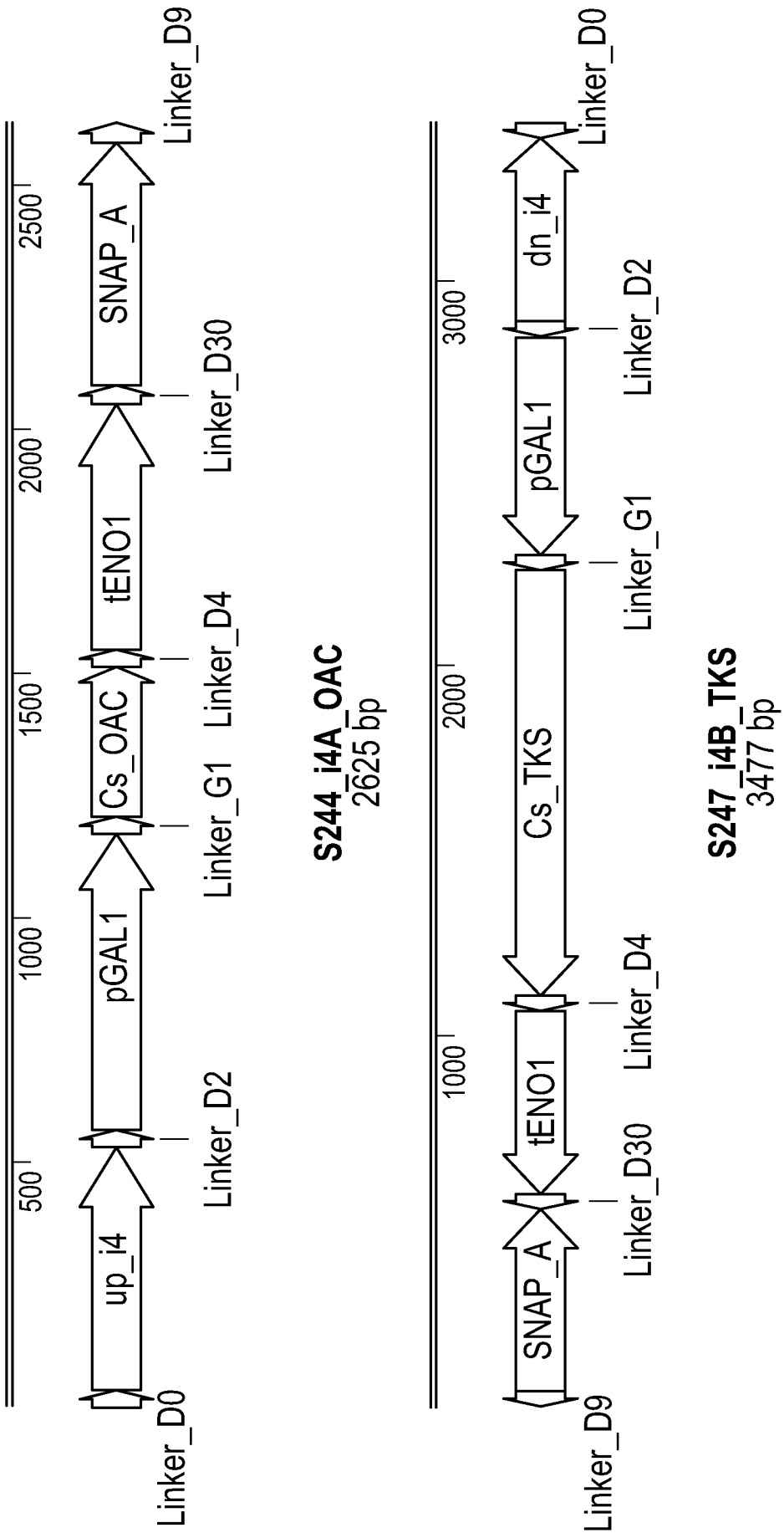

500
1000
1500
2000
2500
Linker_D0
up_i4
Linker_D2
pGAL1
Linker_G1
Cs_OAC
Linker_D4
tENO1
Linker_D30
SNAP_A
Linker_D9
S244_i4A_OAC
2625 bp
1000
2000
3000
Linker_D9
SNAP_A
Linker_D30
tENO1
Linker_D4
Cs_TKS
Linker_G1
pGAL1
Linker_D2
dn_i4
Linker_D0
S247_i4B_TKS
3477 bp

FIG. 43A

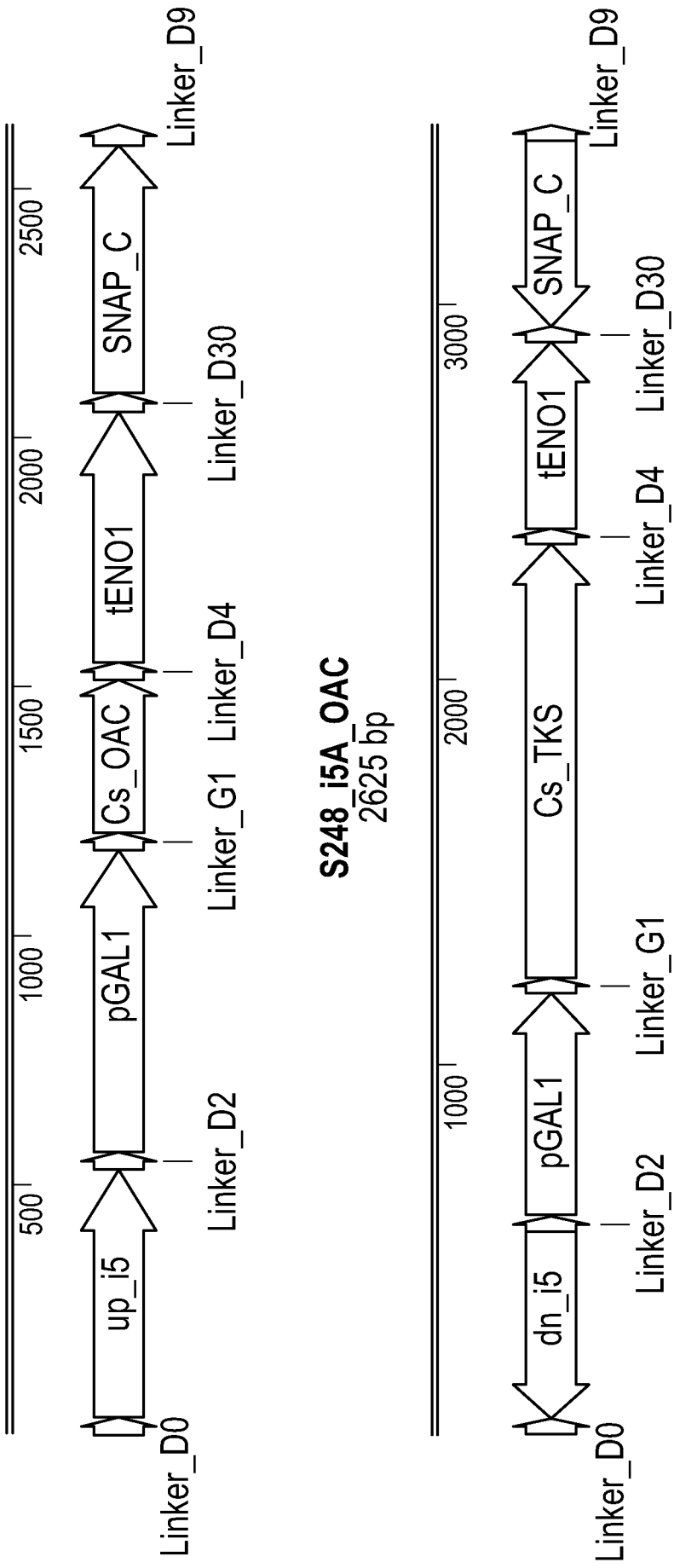

500
1000
1500
2000
2500
Linker_D0
up_i5
Linker_D2
pGAL1
Linker_G1
Cs_OAC
Linker_D4
tENO1
Linker_D30
SNAP_C
Linker_D9
S248_i5A_OAC
2625 bp
1000
2000
3000
Linker_D0
dn_i5
Linker_D2
pGAL1
Linker_G1
Cs_TKS
Linker_D4
tENO1
Linker_D30
SNAP_C
Linker_D9
S251_i5B_TKS
3477 bp

FIG. 43B

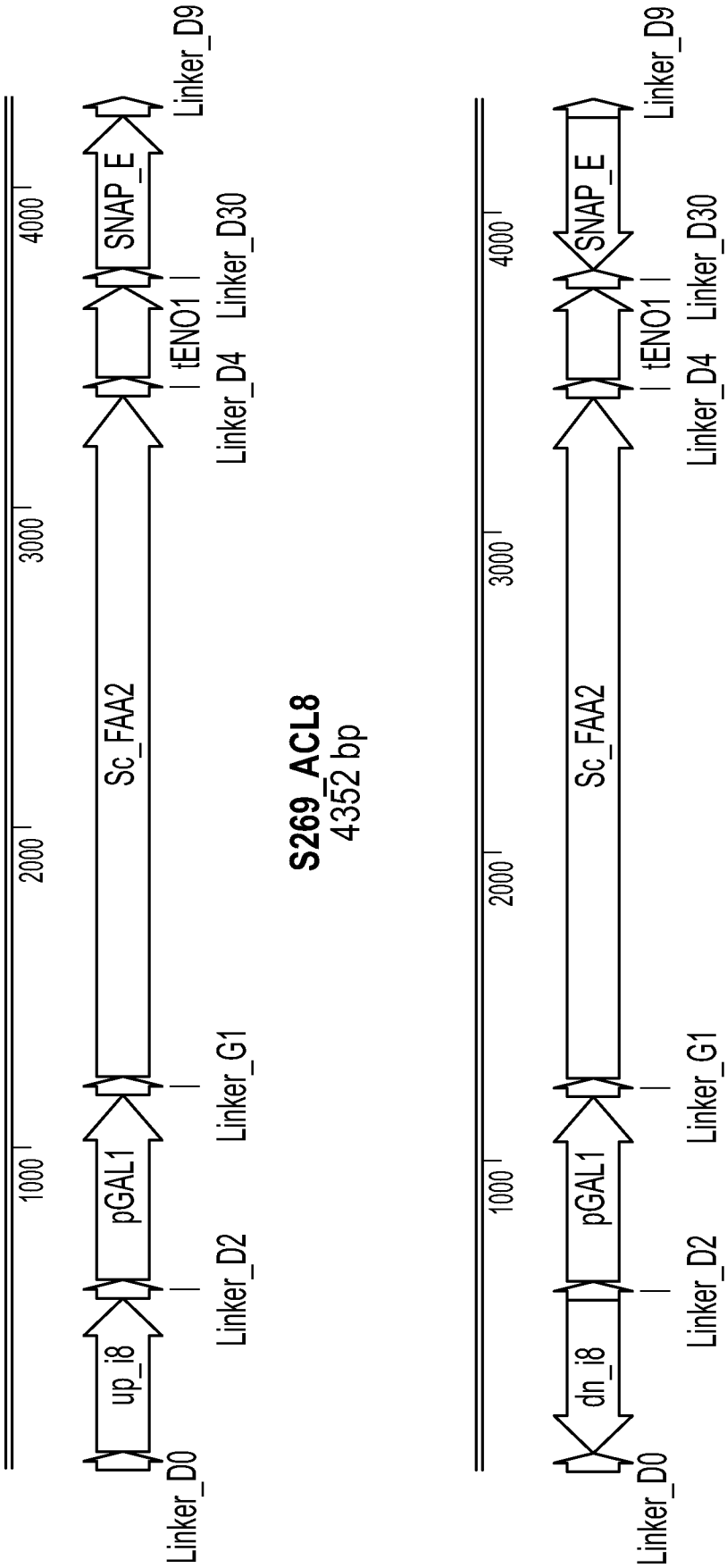

1000
2000
3000
4000
Linker_D0
up_i8
Linker_D2
pGAL1
Linker_G1
Sc_FAA2
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S269_ACL8
4352 bp
1000
2000
3000
4000
Linker_D0
dn_i8
Linker_D2
pGAL1
Linker_G1
Sc_FAA2
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S281_ACL20
4352 bp

FIG. 43C

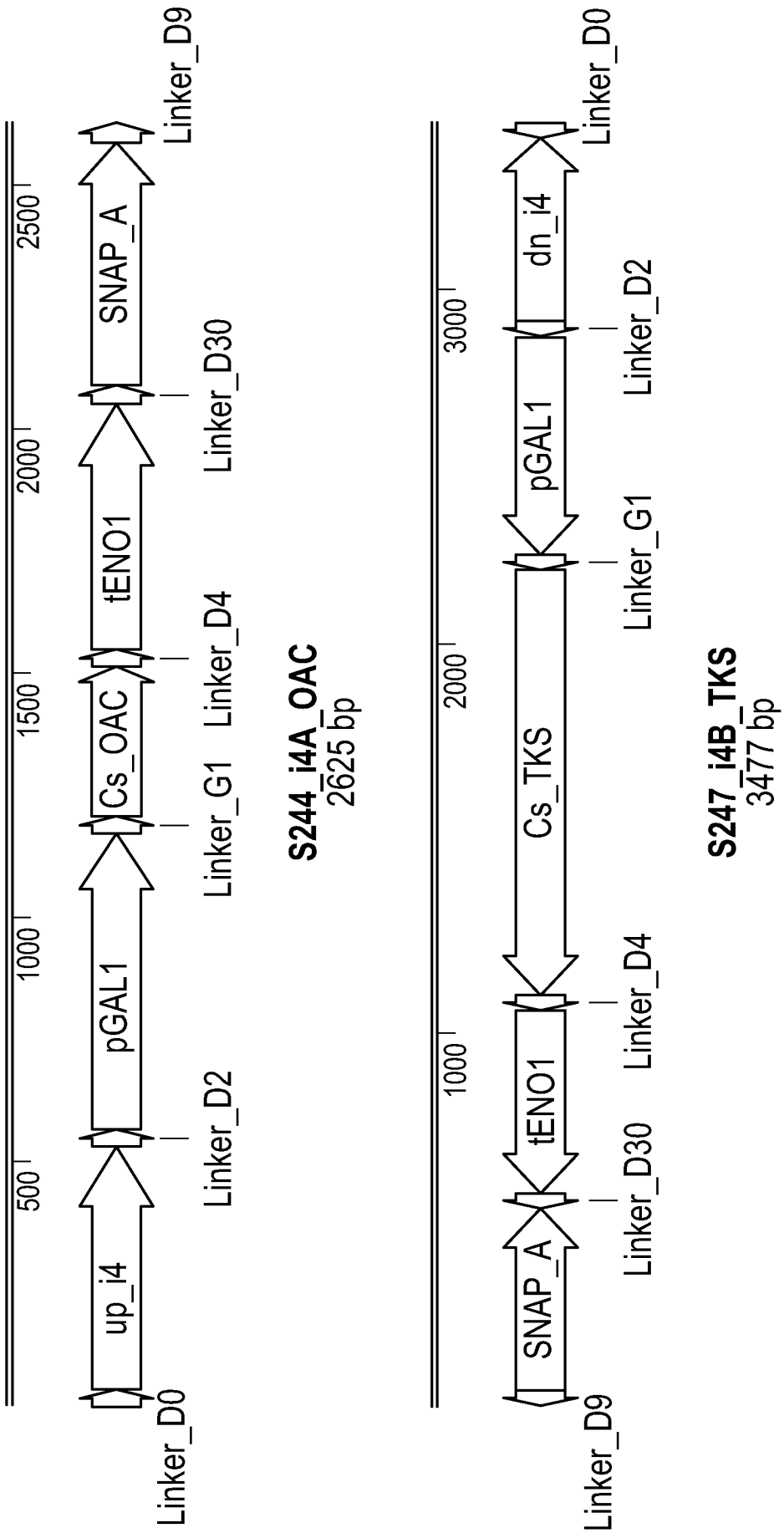

500
1000
1500
2000
2500
Linker_D0
up_i4
Linker_D2
pGAL1
Linker_G1
Cs_OAC
Linker_D4
tENO1
Linker_D30
SNAP_A
Linker_D9
S244_i4A_OAC
2625 bp
1000
2000
3000
Linker_D9
SNAP_A
Linker_D30
tENO1
Linker_D4
Cs_TKS
Linker_G1
pGAL1
Linker_D2
dn_i4
Linker_D0
S247_i4B_TKS
3477 bp

FIG. 44A

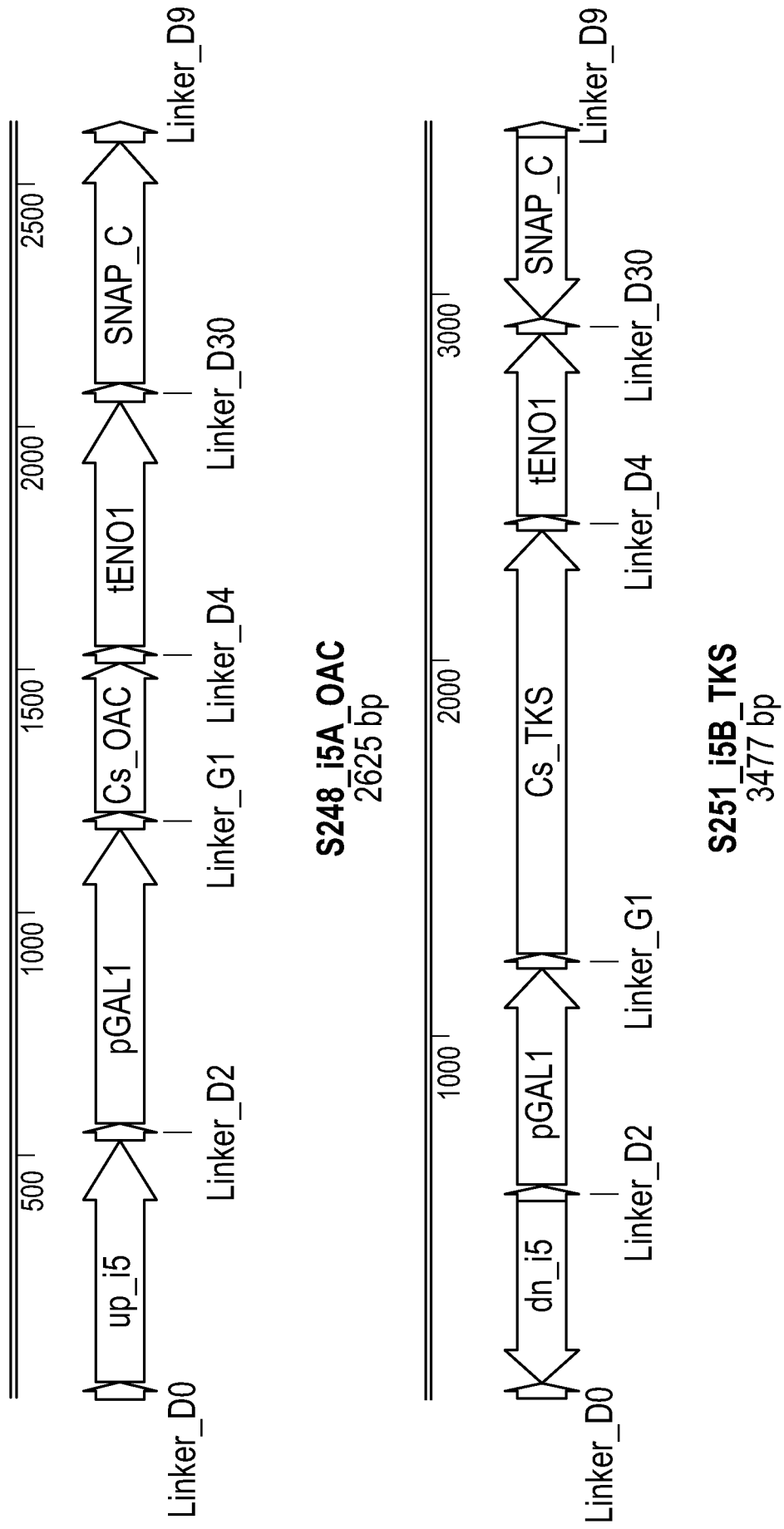

500
1000
1500
2000
2500
up_i5
pGAL1
Cs_OAC
tENO1
SNAP_C
Linker_D0
Linker_D2
Linker_G1
Linker_D4
Linker_D30
Linker_D9
S248_i5A_OAC
2625 bp
1000
2000
3000
dn_i5
pGAL1
Cs_TKS
tENO1
SNAP_C
Linker_D0
Linker_D2
Linker_G1
Linker_D4
Linker_D30
Linker_D9
S251_i5B_TKS
3477 bp

FIG. 44B

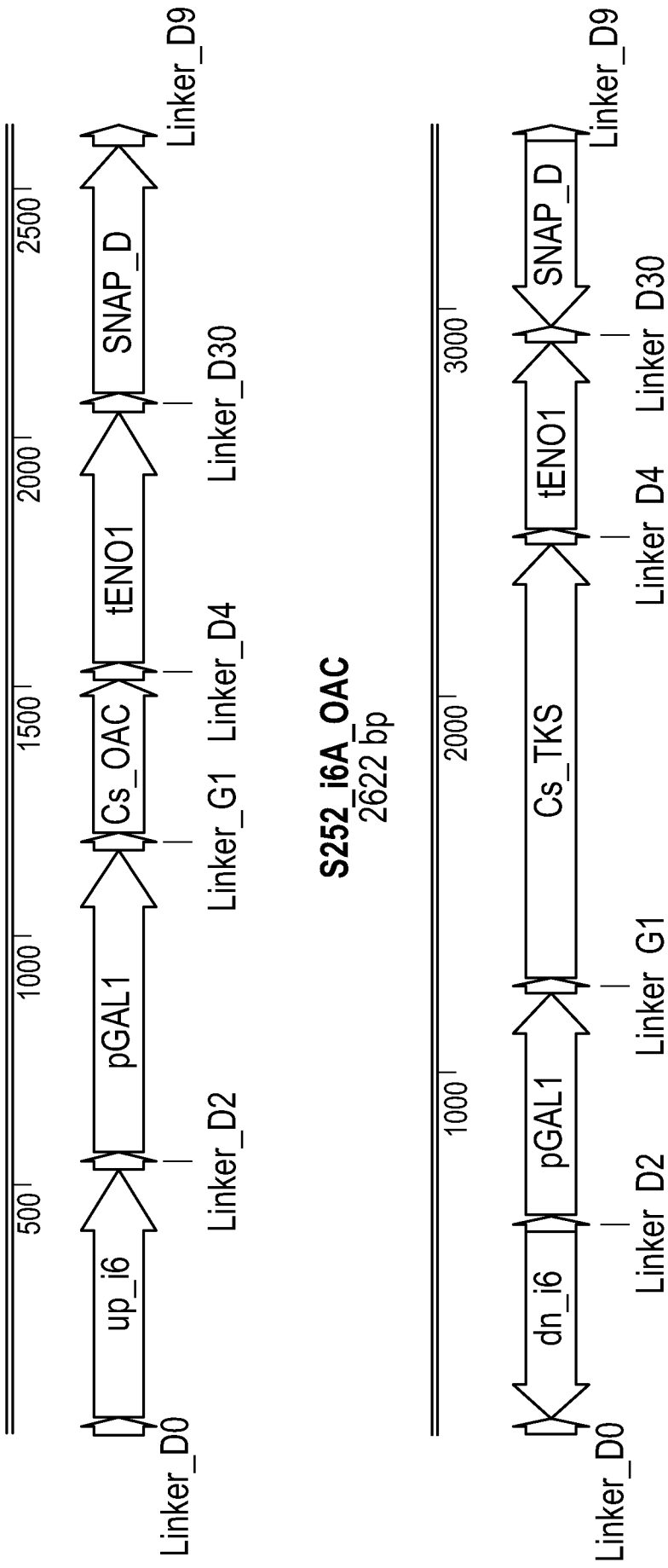

500
1000
1500
2000
2500
Linker_D0
up_i6
Linker_D2
pGAL1
Linker_G1
Cs_OAC
Linker_D4
tENO1
Linker_D30
SNAP_D
Linker_D9
S252_i6A_OAC
2622 bp
1000
2000
3000
Linker_D0
dn_i6
Linker_D2
pGAL1
Linker_G1
Cs_TKS
Linker_D4
tENO1
Linker_D30
SNAP_D
Linker_D9
S255_i6B_TKS
3474 bp

FIG. 44C

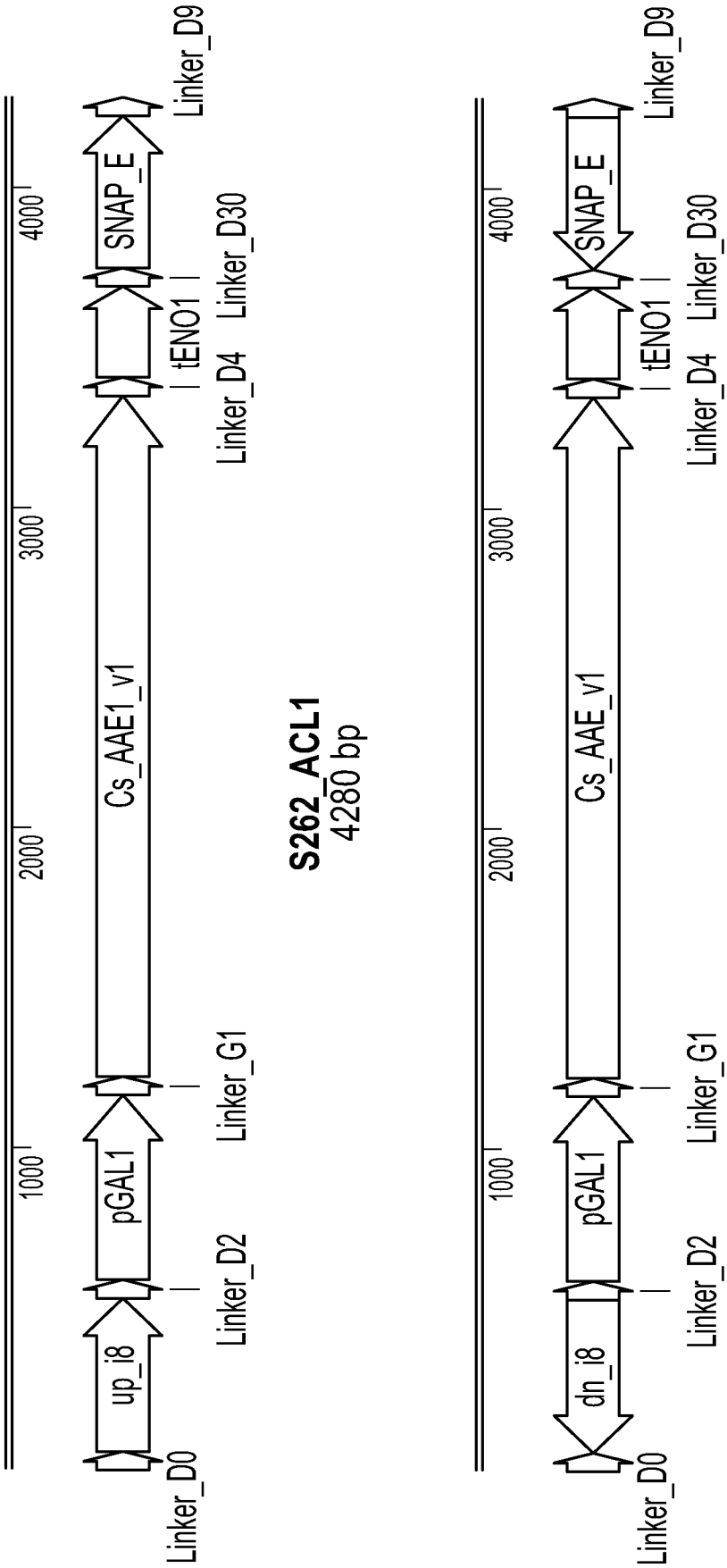

1000
2000
3000
4000
Linker_D0
up_i8
Linker_D2
pGAL1
Linker_G1
Cs_AAE1_v1
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S262_ACL1
4280 bp
1000
2000
3000
4000
Linker_D0
dn_i8
Linker_D2
pGAL1
Linker_G1
Cs_AAE_v1
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S274_ACL13
4280 bp

FIG. 45

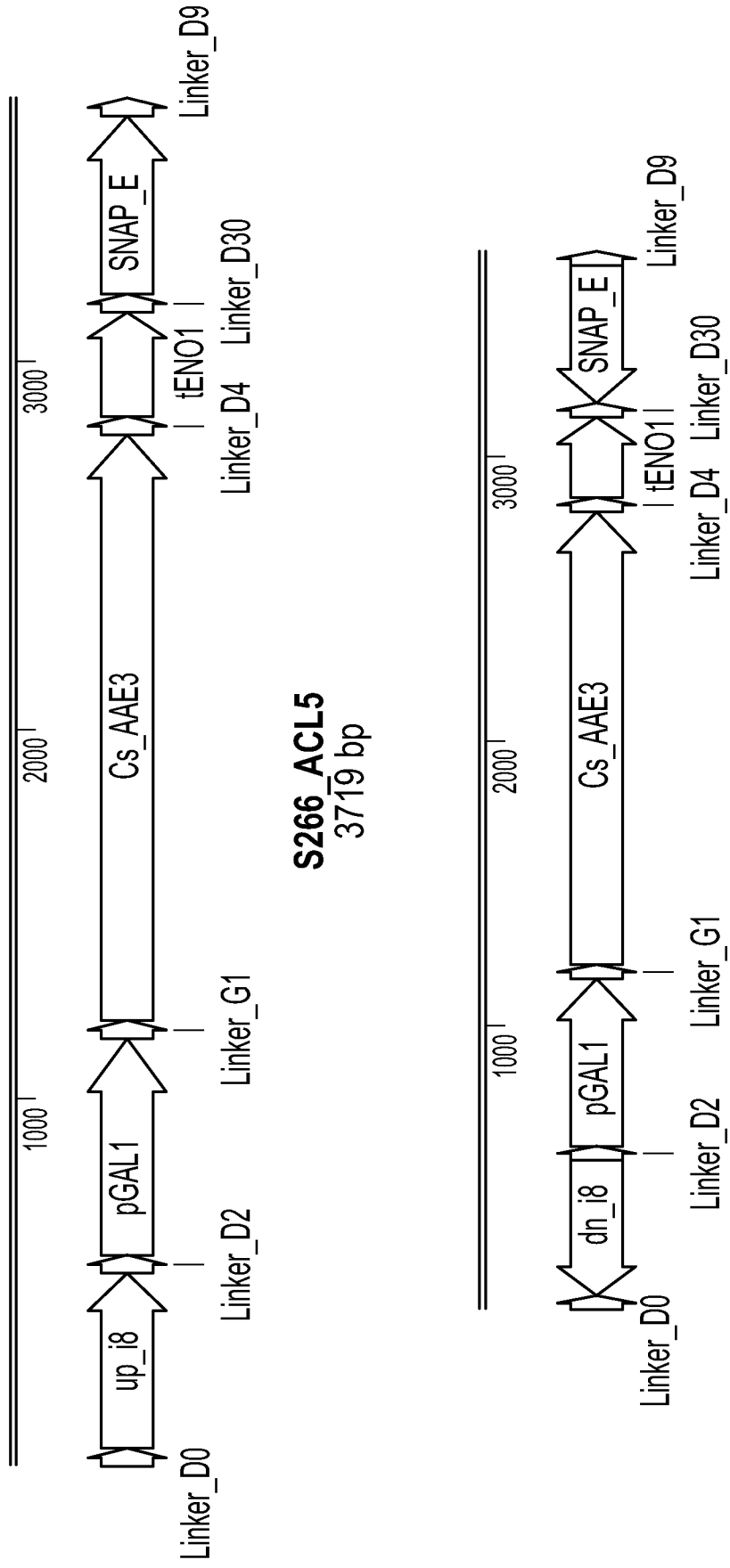

1000
2000
3000
up_i8
pGAL1
Cs_AAE3
SNAP_E
Linker_D0
Linker_D2
Linker_G1
tENO1
Linker_D4
Linker_D30
Linker_D9
S266_ACL5
3719 bp
1000
2000
3000
dn_i8
pGAL1
Cs_AAE3
SNAP_E
Linker_D0
Linker_D2
Linker_G1
tENO1
Linker_D4
Linker_D30
Linker_D9
S278_ACL17
3719 bp

FIG. 46

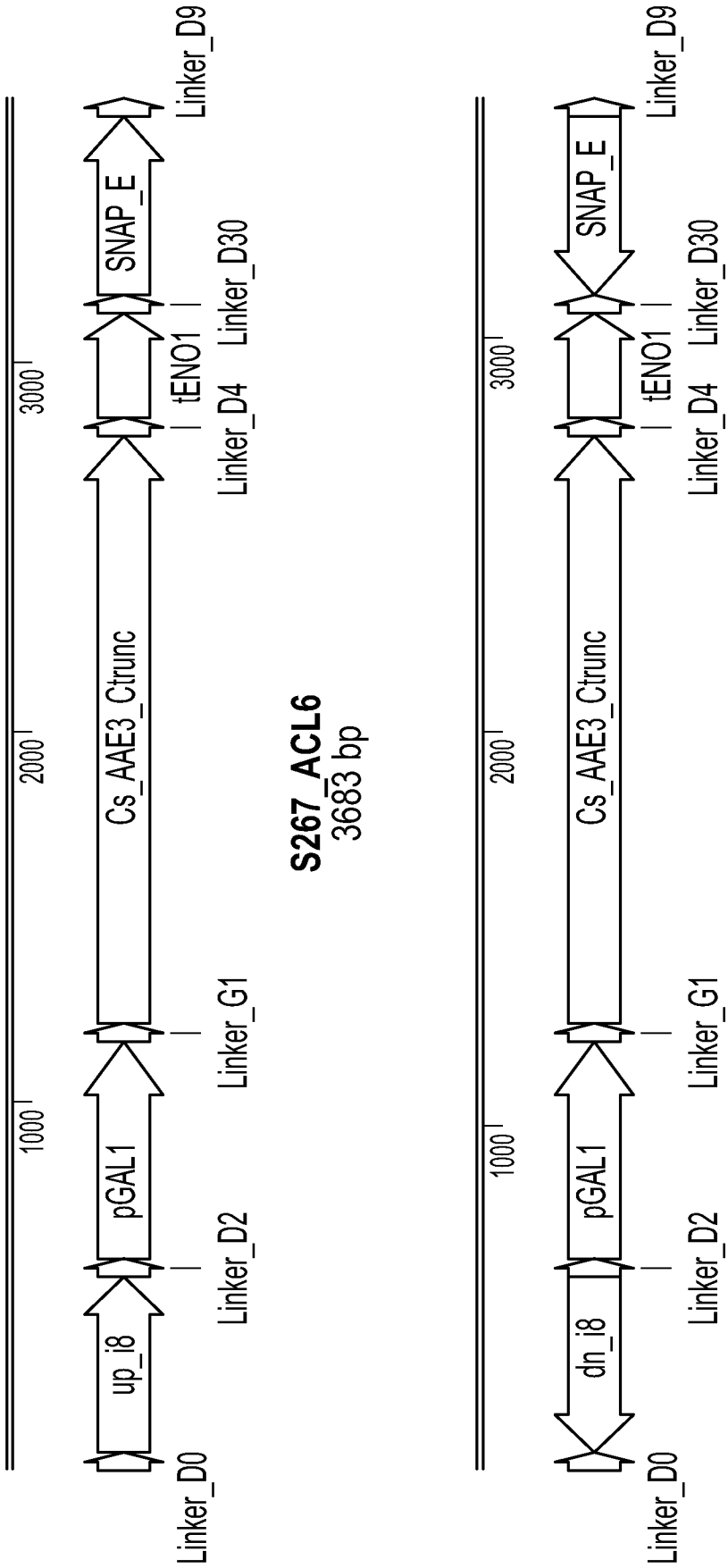

Linker_D0
up_i8
Linker_D2
pGAL1
Linker_G1
Cs_AAE3_Ctrunc
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
1000
2000
3000
S267_ACL6
3683 bp
Linker_D0
dn_i8
Linker_D2
pGAL1
Linker_G1
Cs_AAE3_Ctrunc
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
1000
2000
3000
S279_ACL18
3683 bp

FIG. 47

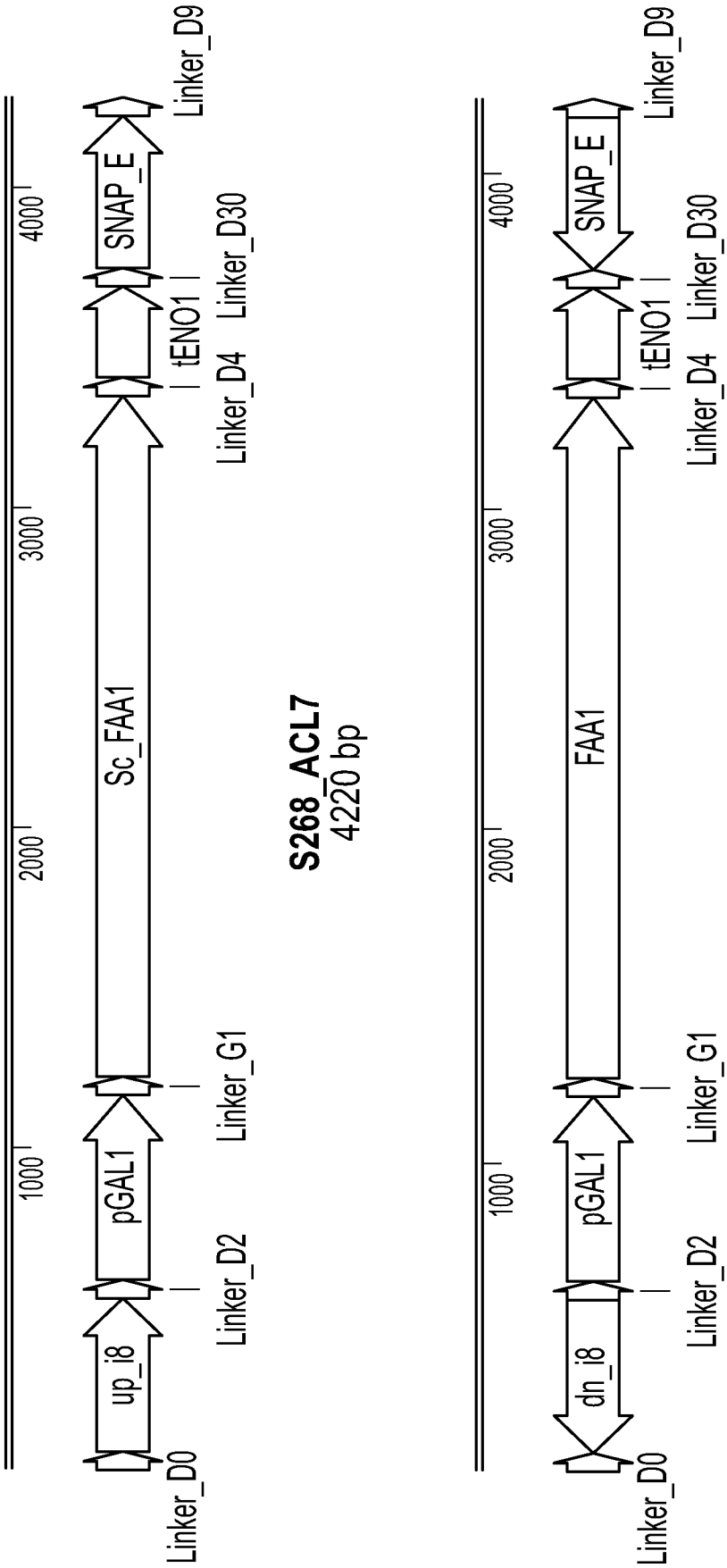

1000
2000
3000
4000
Linker_D0
up_i8
Linker_D2
pGAL1
Linker_G1
Sc_FAA1
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S268_ACL7
4220 bp
1000
2000
3000
4000
Linker_D0
dn_i8
Linker_D2
pGAL1
Linker_G1
FAA1
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S280_ACL19
4220 bp

FIG. 48

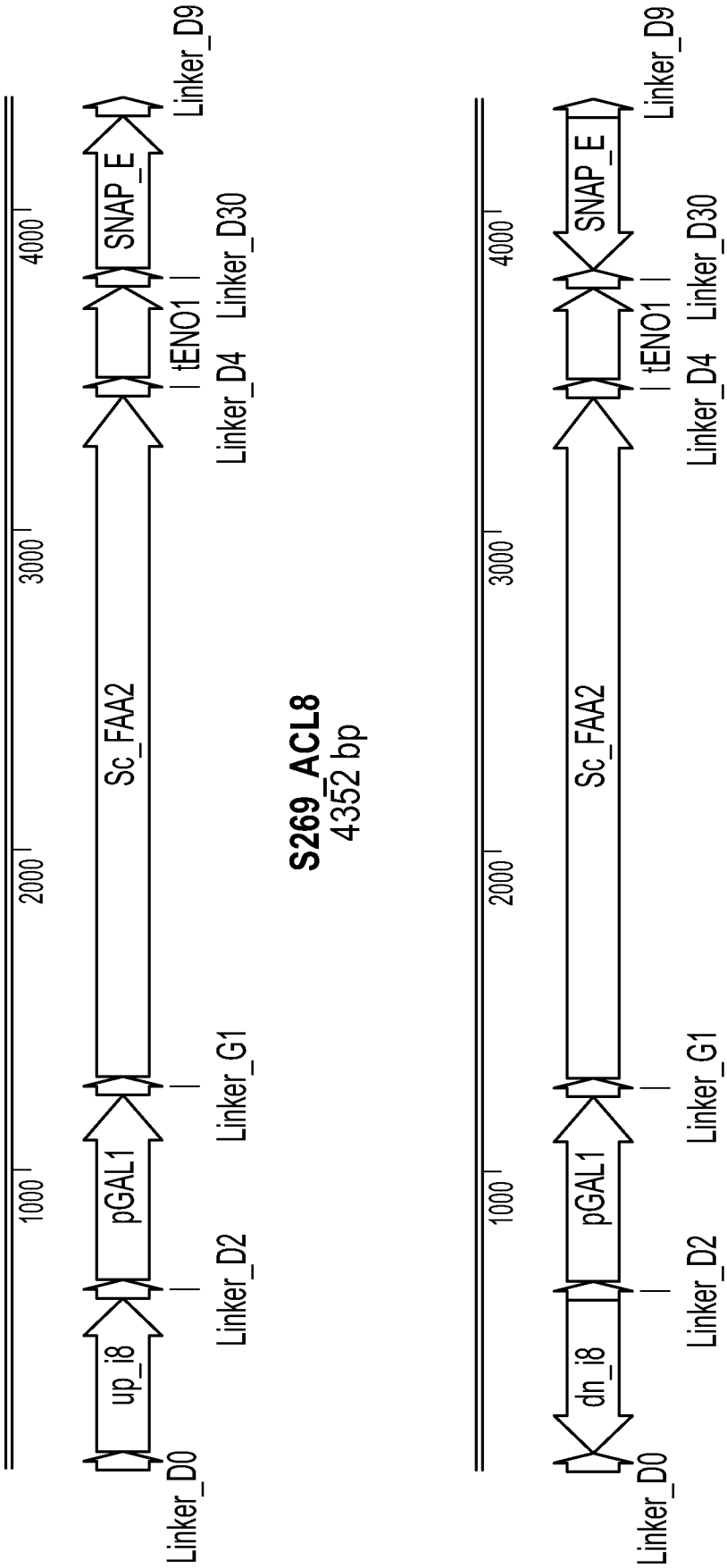

1000
2000
3000
4000
Linker_D0
up_i8
Linker_D2
pGAL1
Linker_G1
Sc_FAA2
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S269_ACL8
4352 bp
1000
2000
3000
4000
Linker_D0
dn_i8
Linker_D2
pGAL1
Linker_G1
Sc_FAA2
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S281_ACL20
4352 bp

FIG. 49

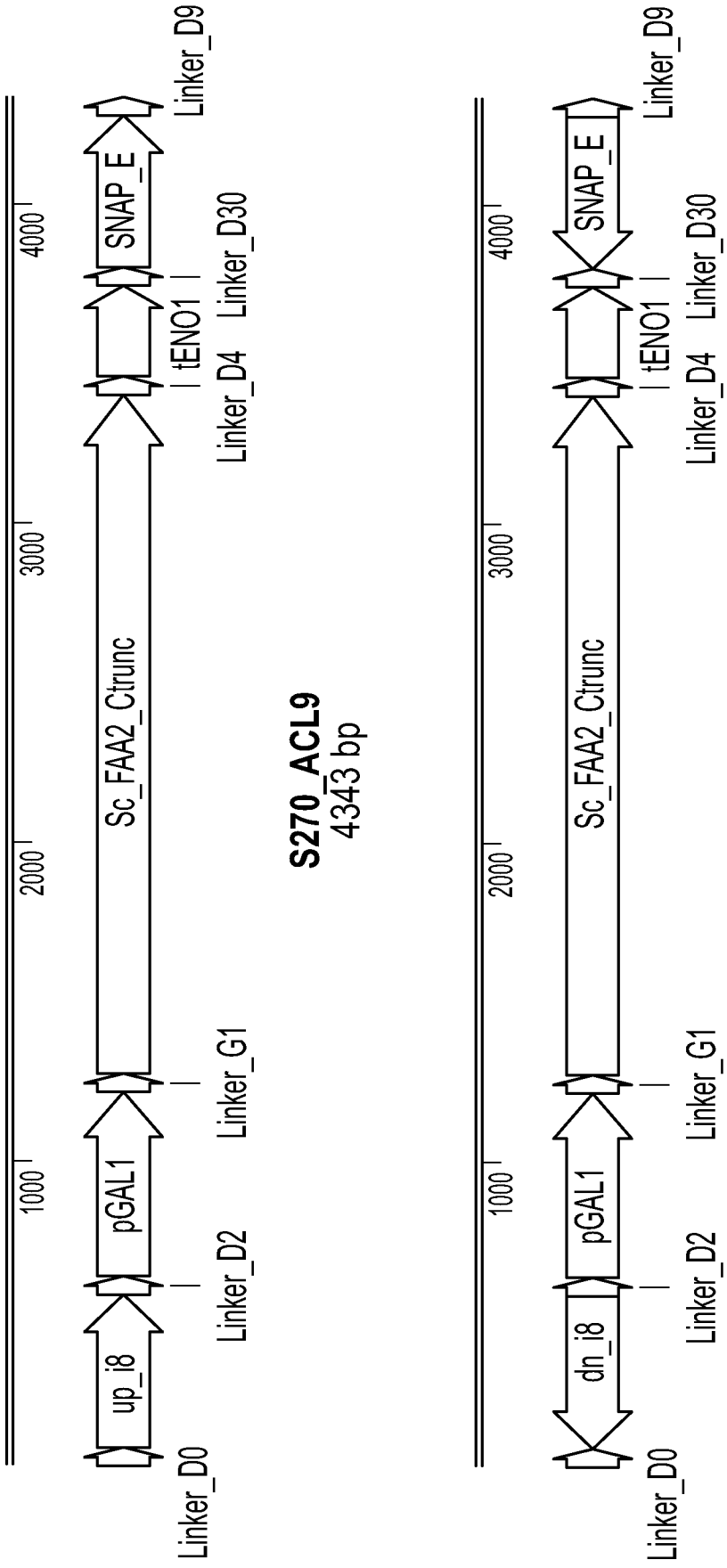

1000
2000
3000
4000
Linker_D0
up_i8
Linker_D2
pGAL1
Linker_G1
Sc_FAA2_Ctrunc
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S270_ACL9
4343 bp
1000
2000
3000
4000
Linker_D0
dn_i8
Linker_D2
pGAL1
Linker_G1
Sc_FAA2_Ctrunc
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S282_ACL21
4343 bp

FIG. 50

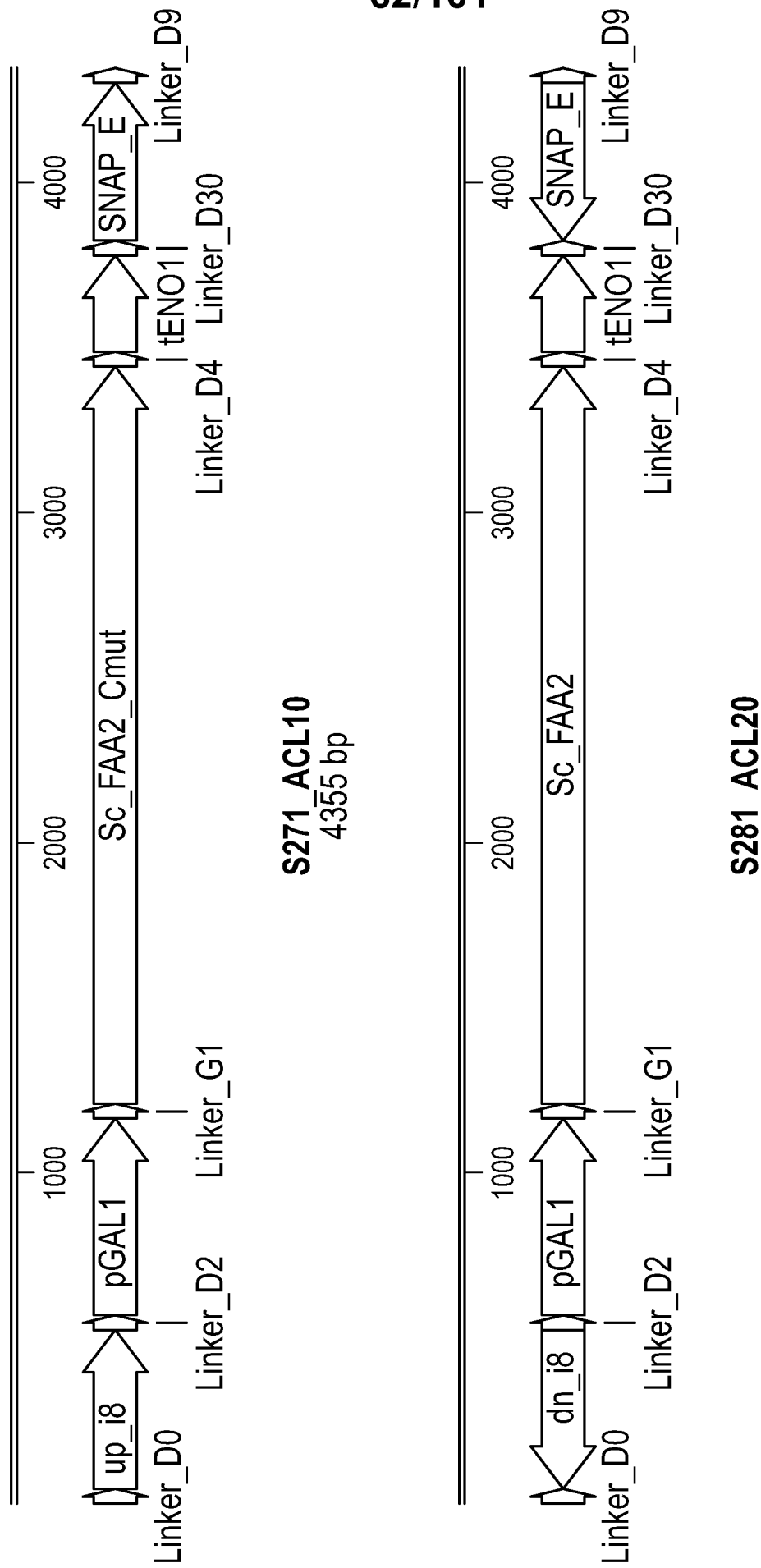

1000
2000
3000
4000
up_i8
pGAL1
Sc_FAA2_Cmut
SNAP_E
Linker_D0
Linker_D2
Linker_G1
Linker_D4
tENO1
Linker_D30
Linker_D9
S271_ACL10
4355 bp
dn_i8
Sc_FAA2
S281_ACL20
4352 bp

FIG. 51

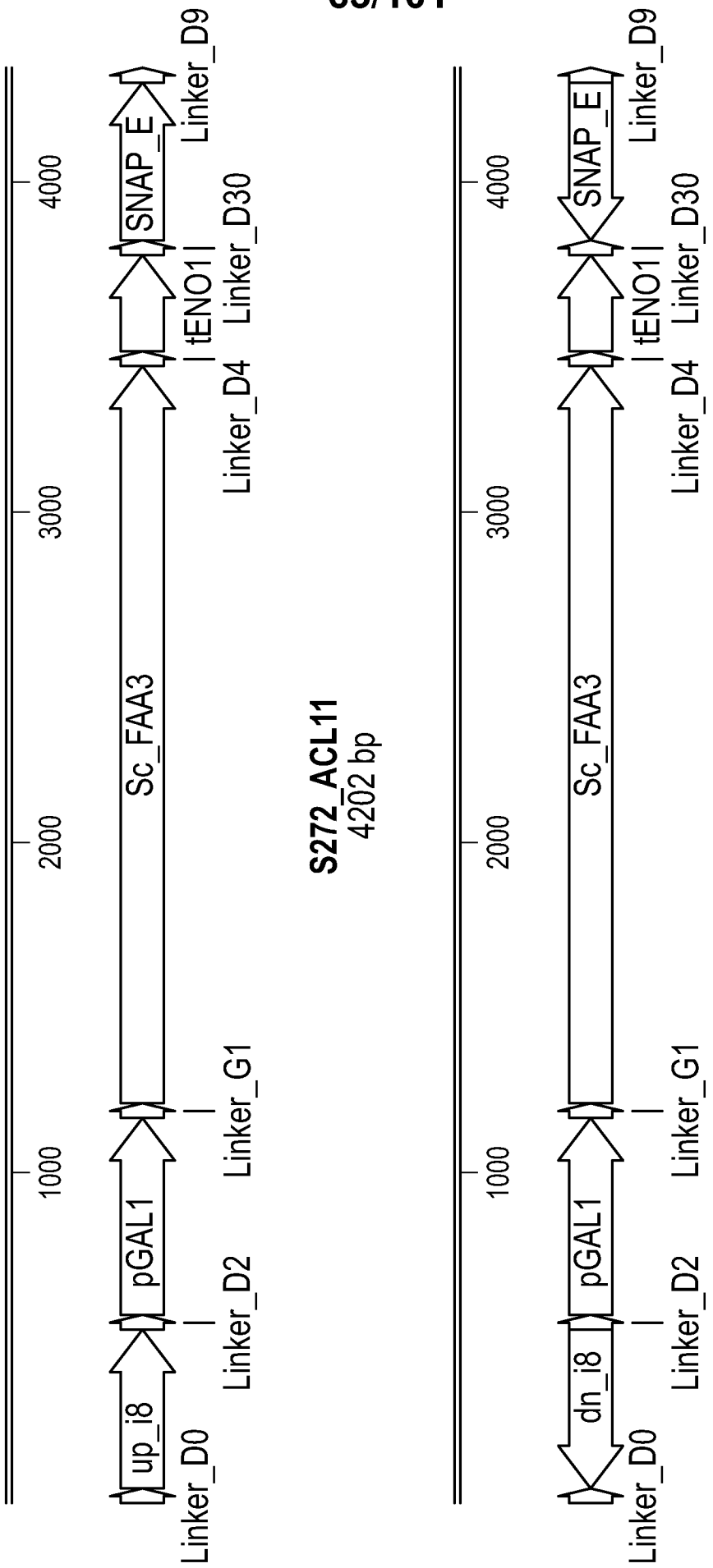

1000
2000
3000
4000
Linker_D0
up_i8
Linker_D2
pGAL1
Linker_G1
Sc_FAA3
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S272_ACL11
4202 bp
dn_i8
S284_ACL23
4202 bp

FIG. 52

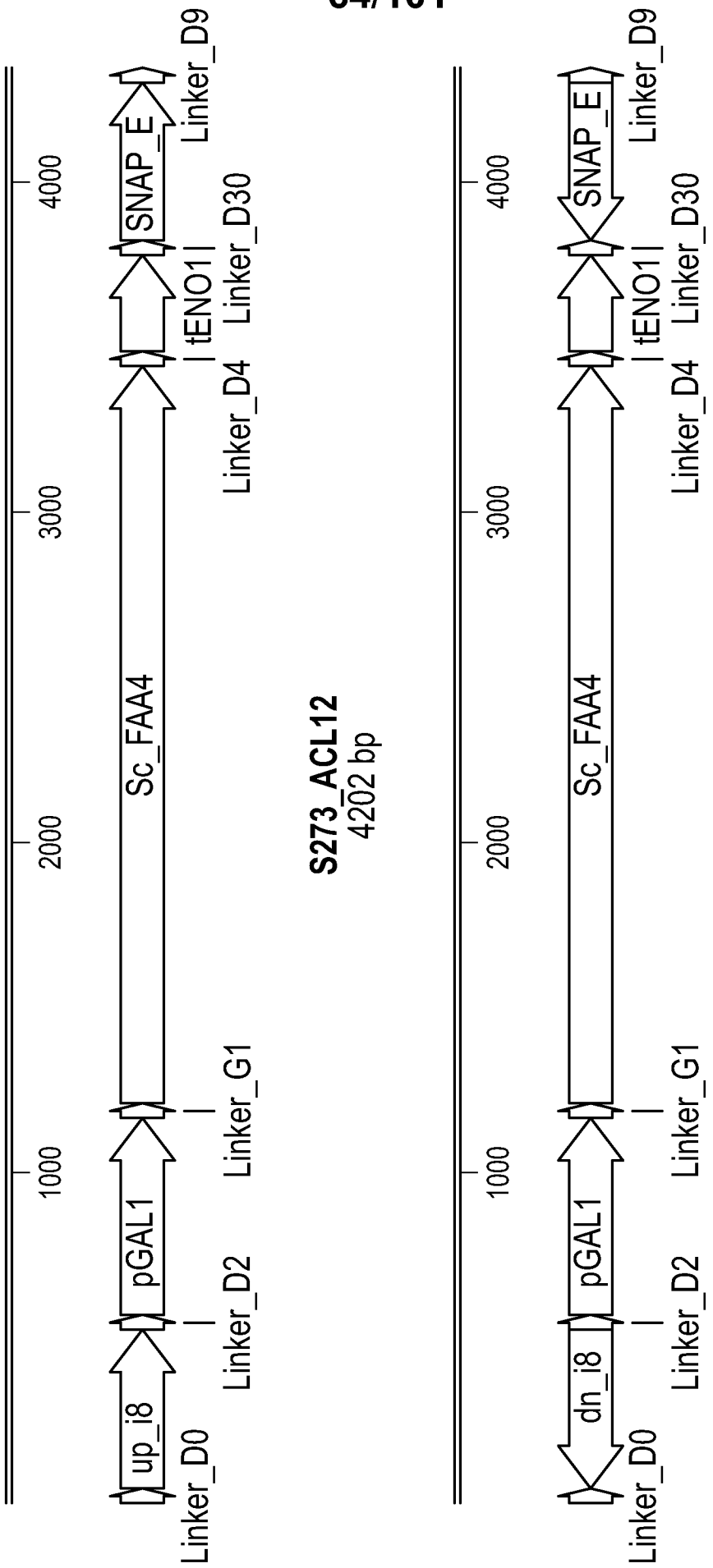

1000
2000
3000
4000
Linker_D0
up_i8
Linker_D2
pGAL1
Linker_G1
Sc_FAA4
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S273_ACL12
4202 bp
1000
2000
3000
4000
Linker_D0
dn_i8
Linker_D2
pGAL1
Linker_G1
Sc_FAA4
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
S285_ACL24
4202 bp

FIG. 53

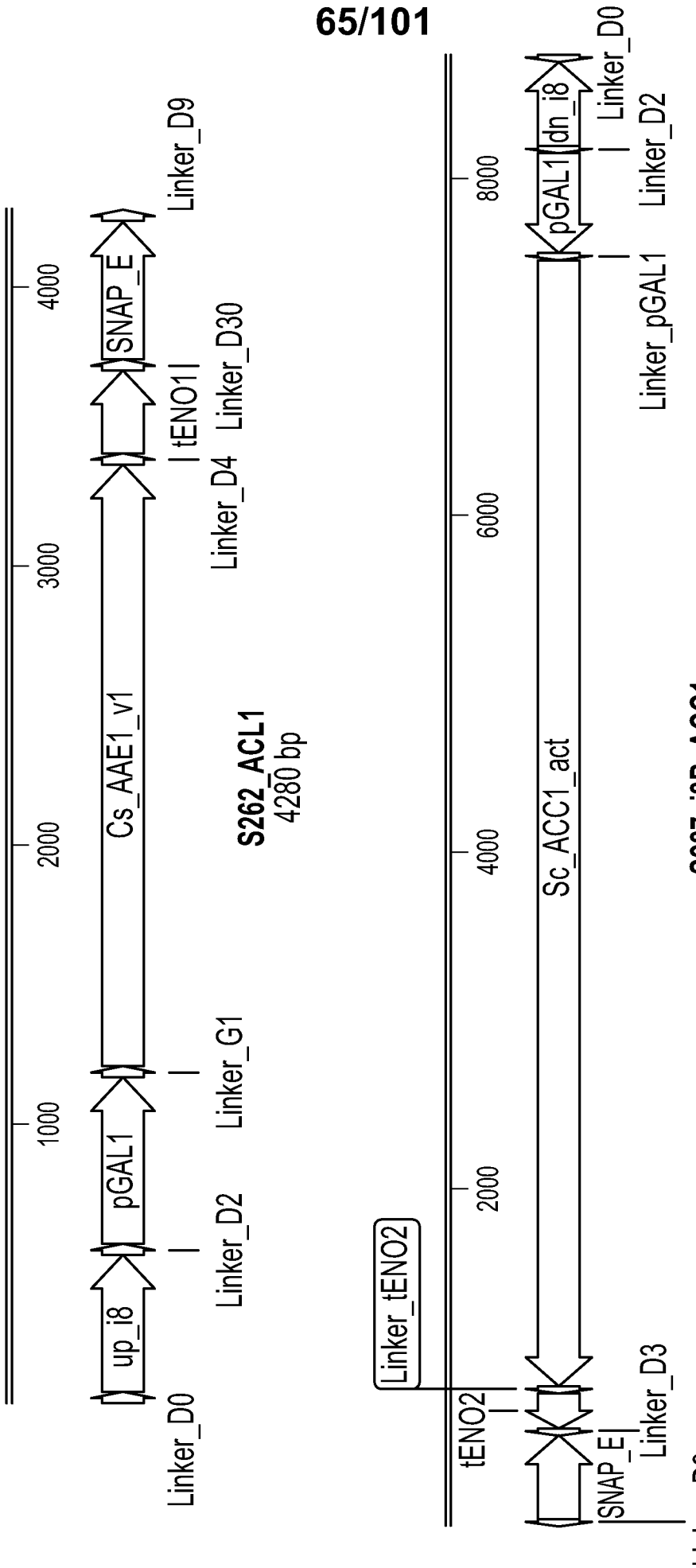

1000
2000
3000
4000
up_i8
pGAL1
Cs_AAE1_v1
SNAP_E
Linker_D0
Linker_D2
Linker_G1
Linker_D4
tENO1
Linker_D30
Linker_D9
S262_ACL1
4280 bp
2000
4000
6000
8000
Linker_tENO2
tENO2
Sc_ACC1_act
SNAP_E
Linker_D3
Linker_D9
pGAL1
dn_i8
Linker_pGAL1
Linker_D2
Linker_D0
S287_i8B_ACC1
8732 bp

FIG. 54

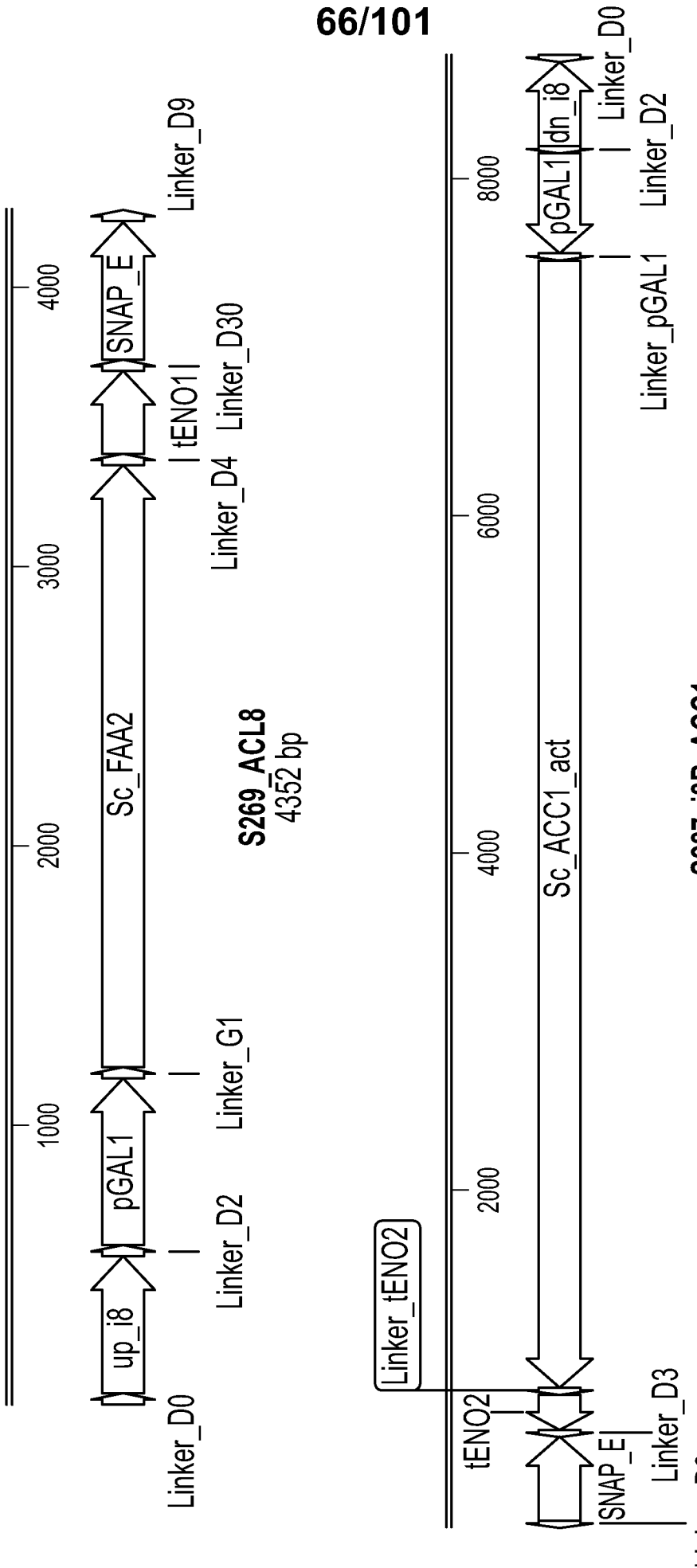

1000
2000
3000
4000
up_i8
Linker_D0
pGAL1
Linker_D2
Sc_FAA2
Linker_G1
SNAP_E
Linker_D9
tENO1
Linker_D4
Linker_D30
S269_ACL8
4352 bp
2000
4000
6000
8000
Linker_tENO2
tENO2
SNAP_E
Linker_D3
Linker_D9
Sc_ACC1_act
pGAL1
Linker_pGAL1
dn_i8
Linker_D0
Linker_D2
S287_i8B_ACC1
8732 bp

FIG. 55

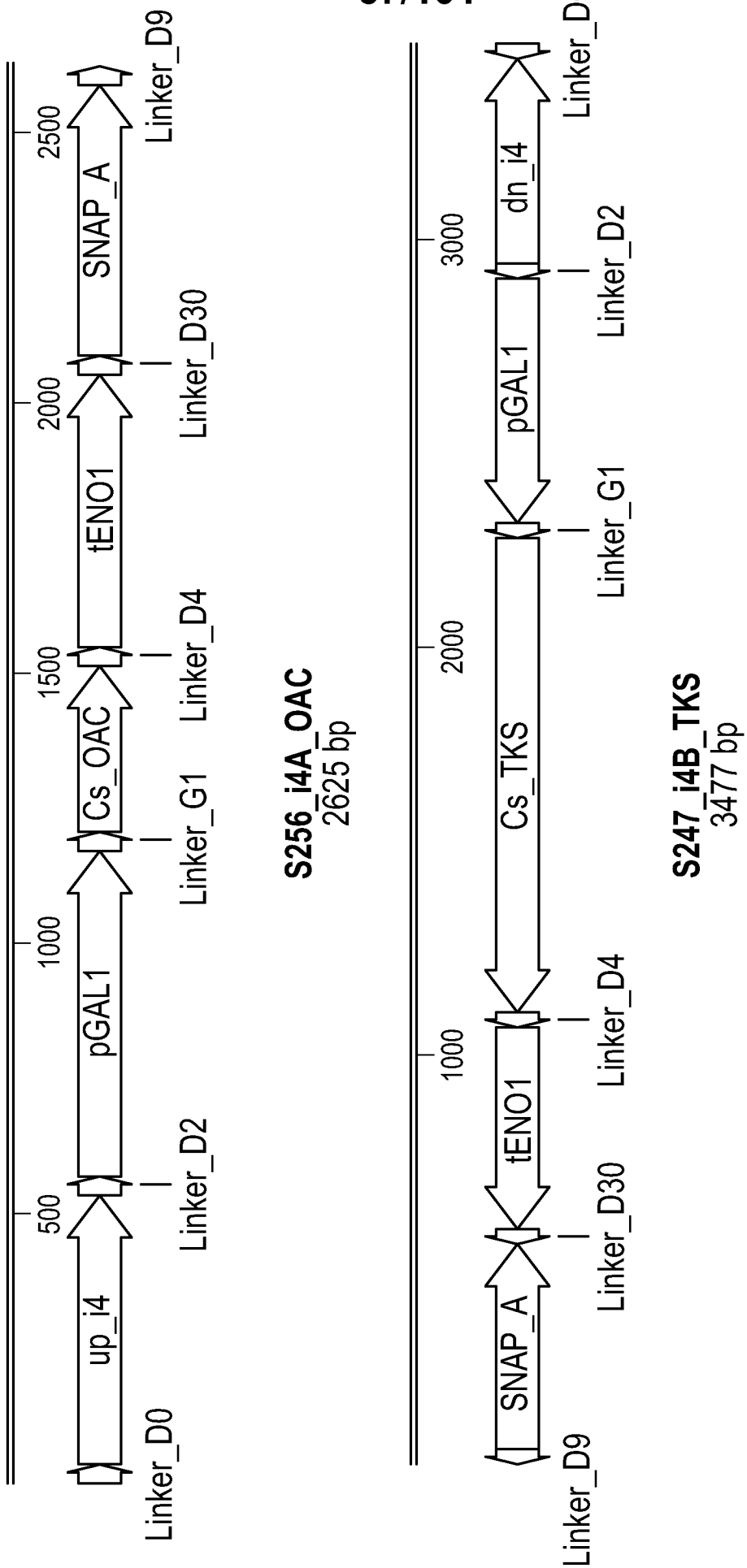

500
1000
1500
2000
2500
Linker_D0
up_i4
Linker_D2
pGAL1
Linker_G1
Cs_OAC
Linker_D4
tENO1
Linker_D30
SNAP_A
Linker_D9
S256_i4A_OAC
2625 bp
1000
2000
3000
Linker_D9
SNAP_A
Linker_D30
tENO1
Linker_D4
Cs_TKS
Linker_G1
pGAL1
Linker_D2
dn_i4
Linker_D0
S247_i4B_TKS
3477 bp

FIG. 56A

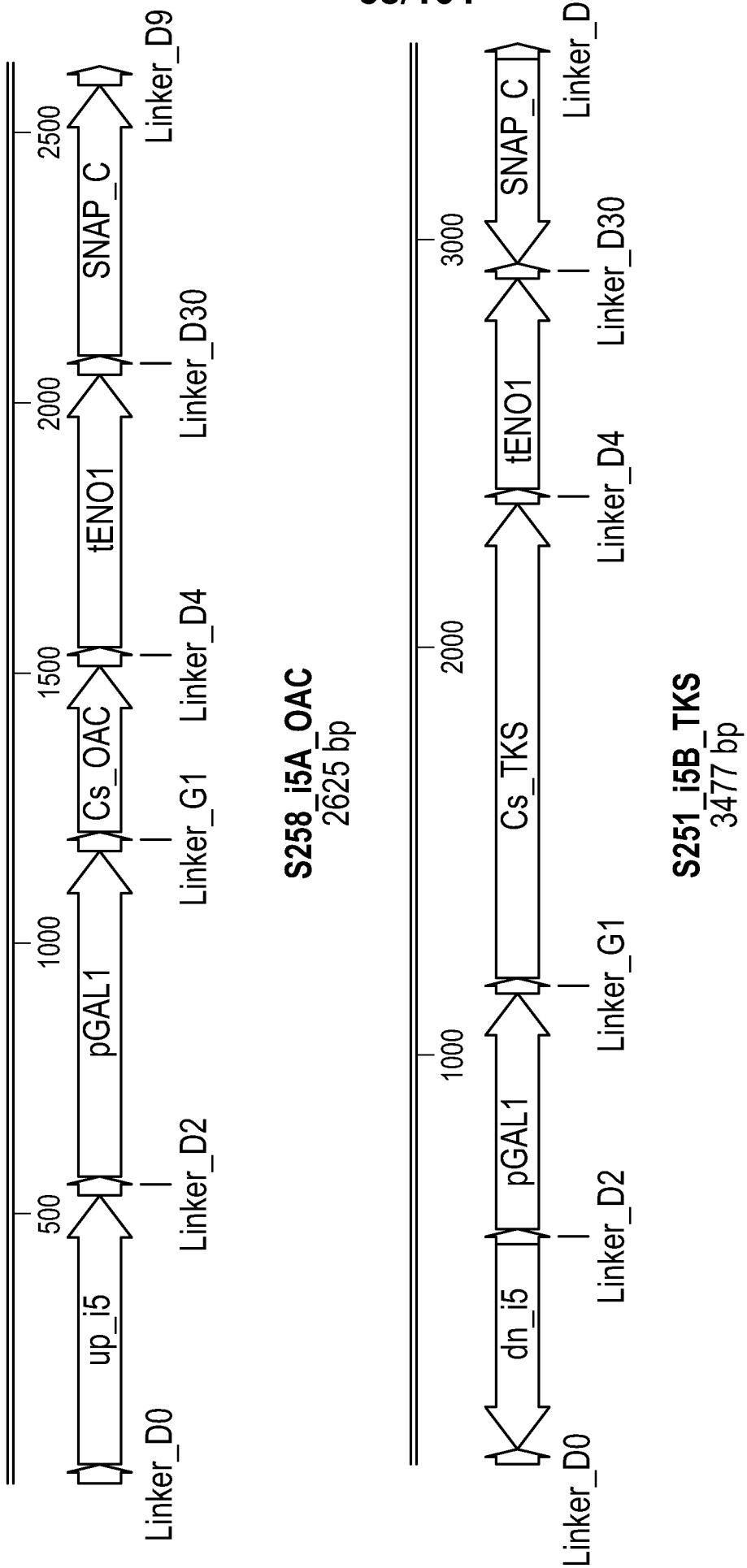

500
1000
1500
2000
2500
up_i5
pGAL1
Cs_OAC
tENO1
SNAP_C
Linker_D0
Linker_D2
Linker_G1
Linker_D4
Linker_D30
Linker_D9
S258_i5A_OAC
2625 bp
1000
2000
3000
dn_i5
pGAL1
Cs_TKS
tENO1
SNAP_C
Linker_D0
Linker_D2
Linker_G1
Linker_D4
Linker_D30
Linker_D9
S251_i5B_TKS
3477 bp

FIG. 56B

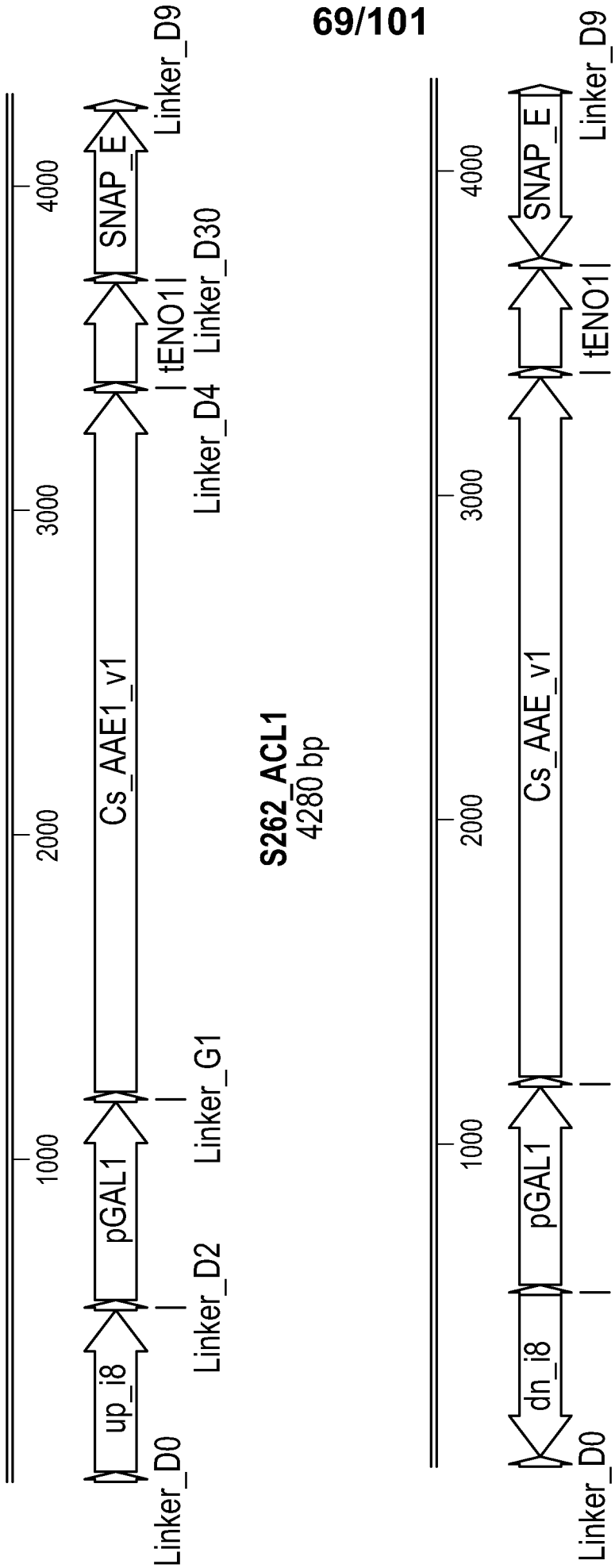

Linker_D0
up_i8
Linker_D2
pGAL1
Linker_G1
Cs_AAE1_v1
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
1000
2000
3000
4000
S262_ACL1
4280 bp
Linker_D0
dn_i8
Linker_D2
pGAL1
Linker_G1
Cs_AAE_v1
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
1000
2000
3000
4000
S274_ACL13
4280 bp

FIG. 56C

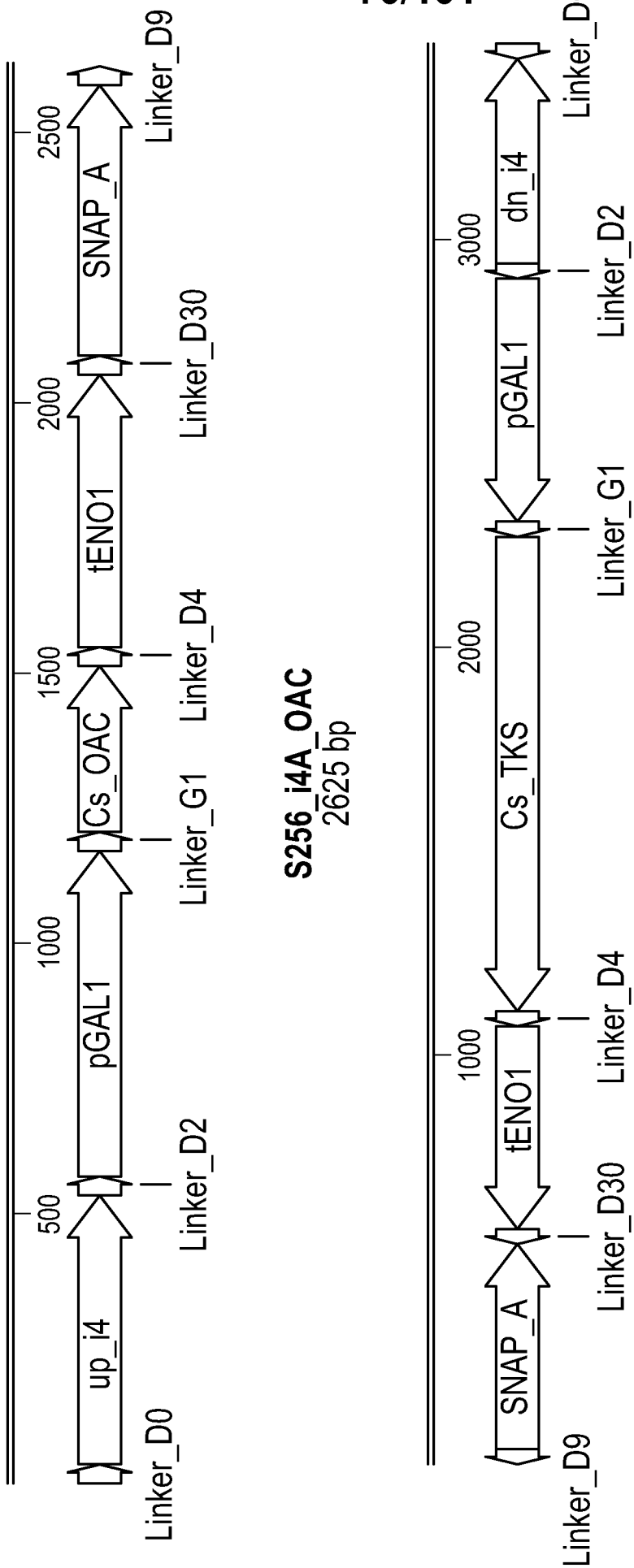

500
1000
1500
2000
2500
up_i4
pGAL1
Cs_OAC
tENO1
SNAP_A
Linker_D0
Linker_D2
Linker_G1
Linker_D4
Linker_D30
Linker_D9
S256_i4A_OAC
2625 bp
1000
2000
3000
SNAP_A
tENO1
Cs_TKS
pGAL1
dn_i4
Linker_D9
Linker_D30
Linker_D4
Linker_G1
Linker_D2
Linker_D0
S247_i4B_TKS
3477 bp

FIG. 57A

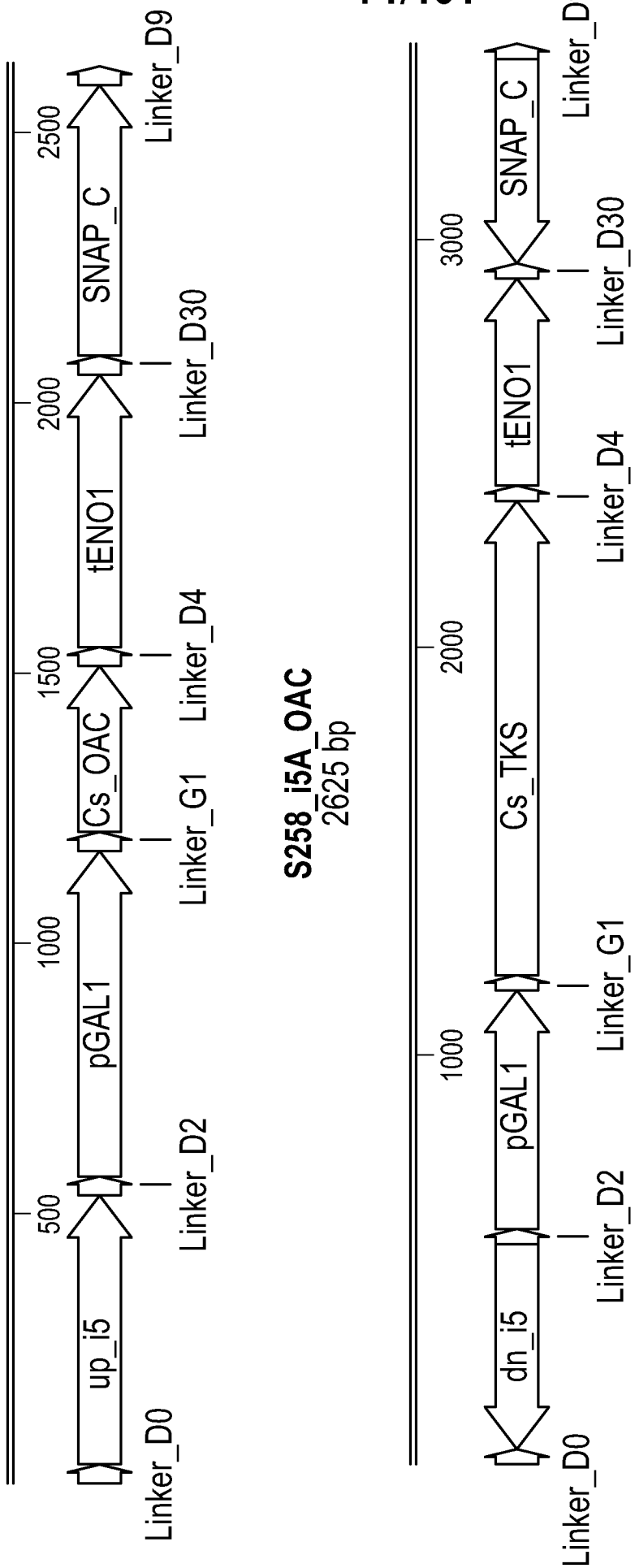

500
1000
1500
2000
2500
Linker_D0
up_i5
Linker_D2
pGAL1
Linker_G1
Cs_OAC
Linker_D4
tENO1
Linker_D30
SNAP_C
Linker_D9
S258_i5A_OAC
2625 bp
1000
2000
3000
Linker_D0
dn_i5
Linker_D2
pGAL1
Linker_G1
Cs_TKS
Linker_D4
tENO1
Linker_D30
SNAP_C
Linker_D9
S251_i5B_TKS
3477 bp

FIG. 57B

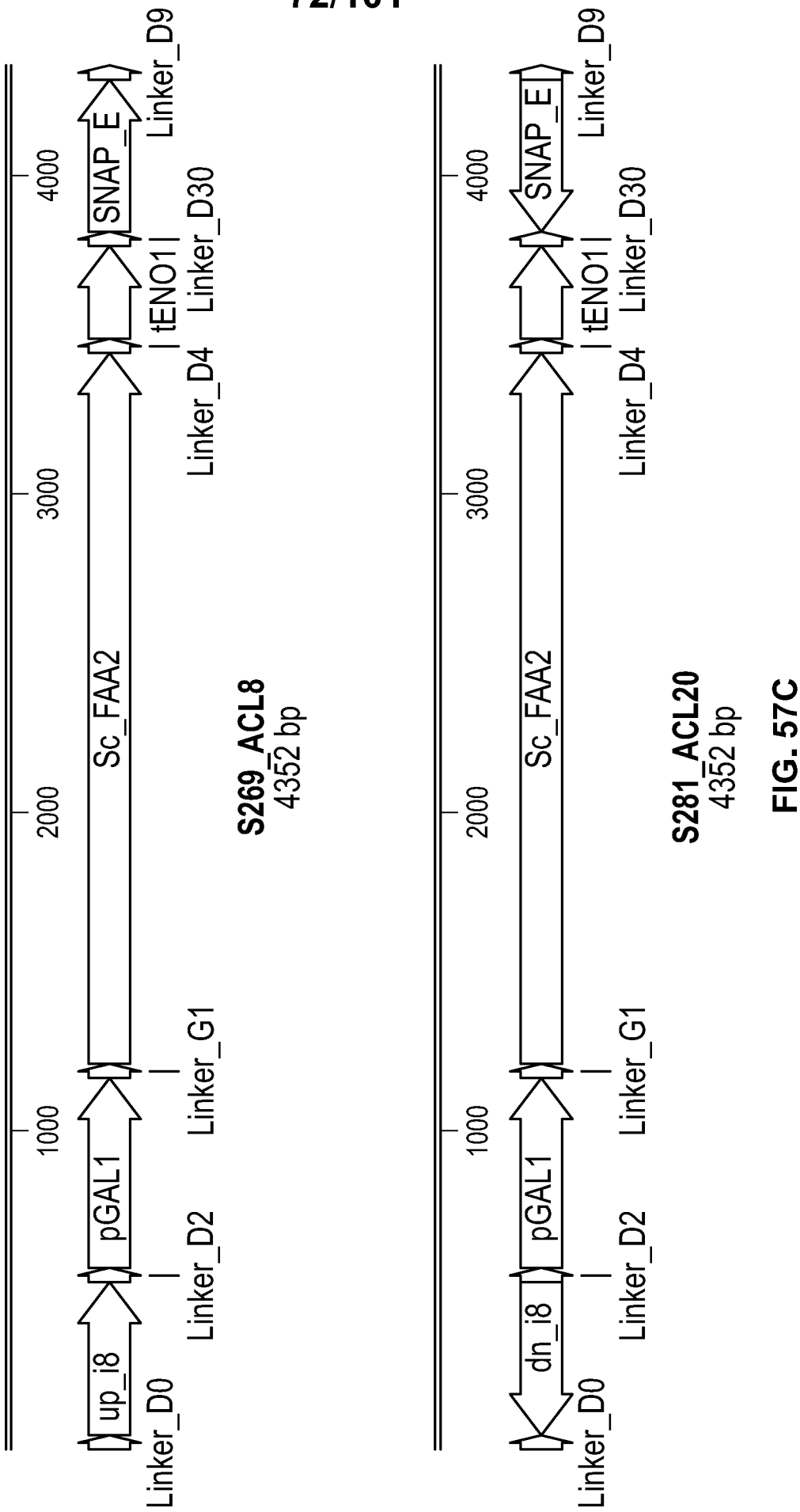
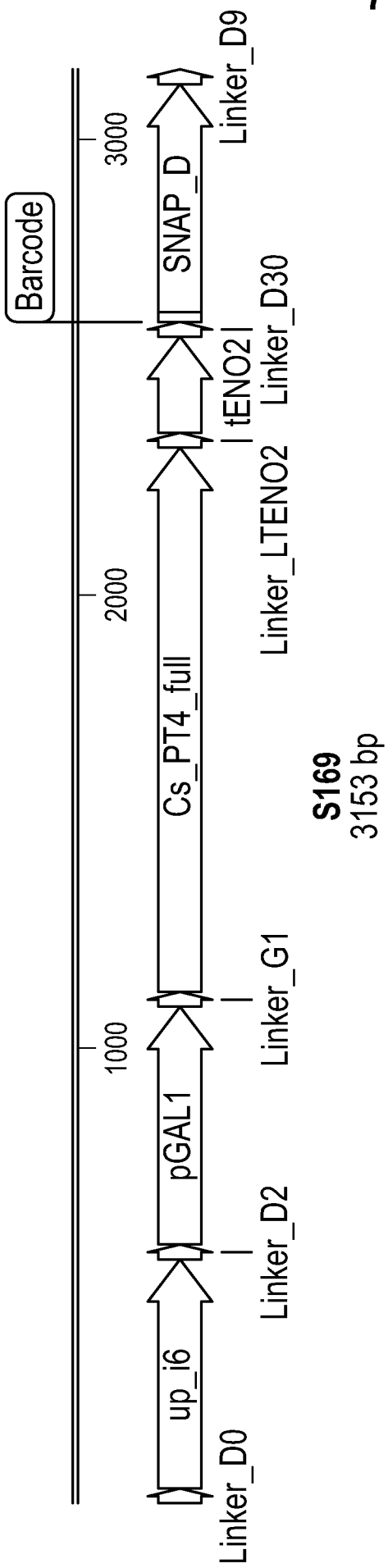
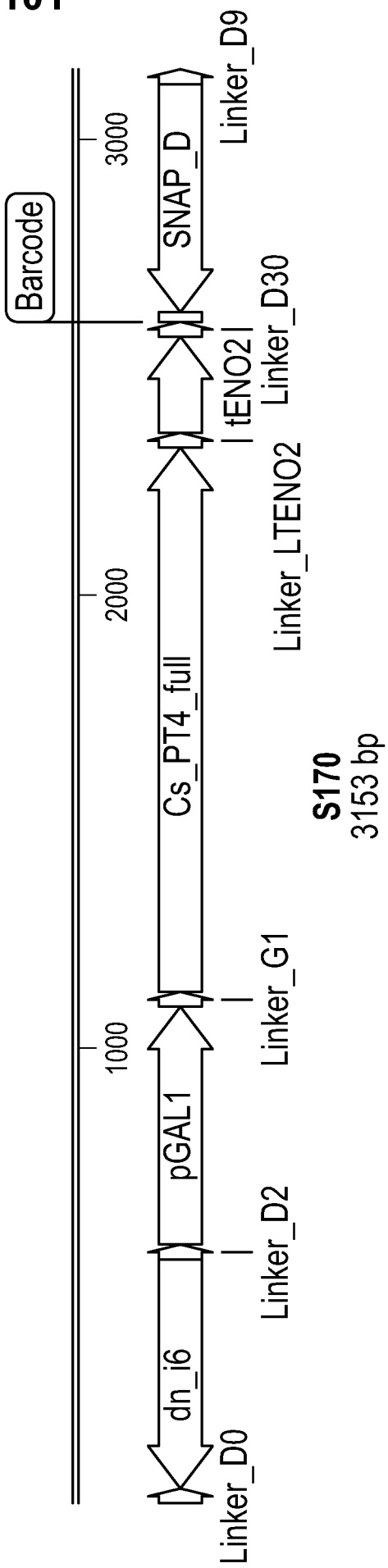

Linker_D0
up_i8
Linker_D2
pGAL1
Linker_G1
Sc_FAA2
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
1000
2000
3000
4000
S269_ACL8
4352 bp
Linker_D0
dn_i8
Linker_D2
pGAL1
Linker_G1
Sc_FAA2
Linker_D4
tENO1
Linker_D30
SNAP_E
Linker_D9
1000
2000
3000
4000
S281_ACL20
4352 bp

FIG. 57C

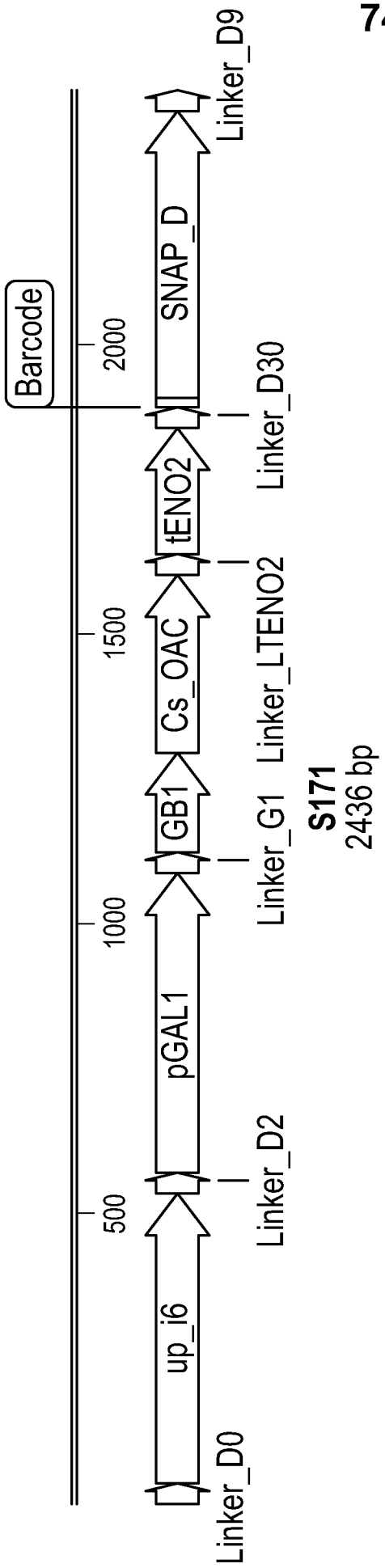


S169
3153 bp



S170
3153 bp

FIG. 58


Barcode
2000
SNAP_D
Linker_D9
1500
tENO2
Linker_D30
Cs_OAC
Linker_LTENO2
1000
GB1
Linker_G1
pGAL1
Linker_D2
500
up_i6
Linker_D0
S171
2436 bp

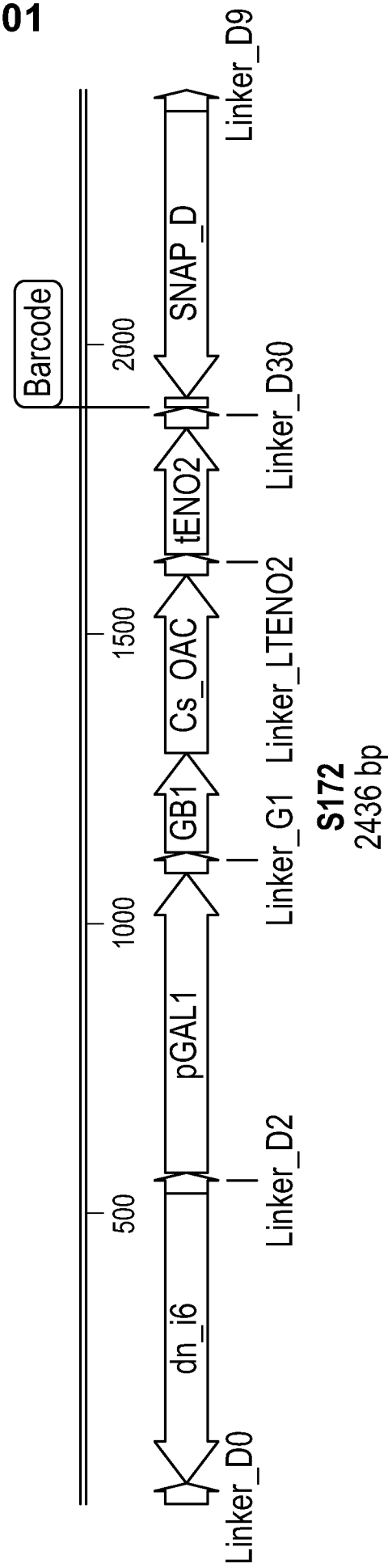

Barcode
2000
SNAP_D
Linker_D9
1500
tENO2
Linker_D30
Cs_OAC
Linker_LTENO2
1000
GB1
Linker_G1
pGAL1
Linker_D2
500
dn_i6
Linker_D0
S172
2436 bp

FIG. 59

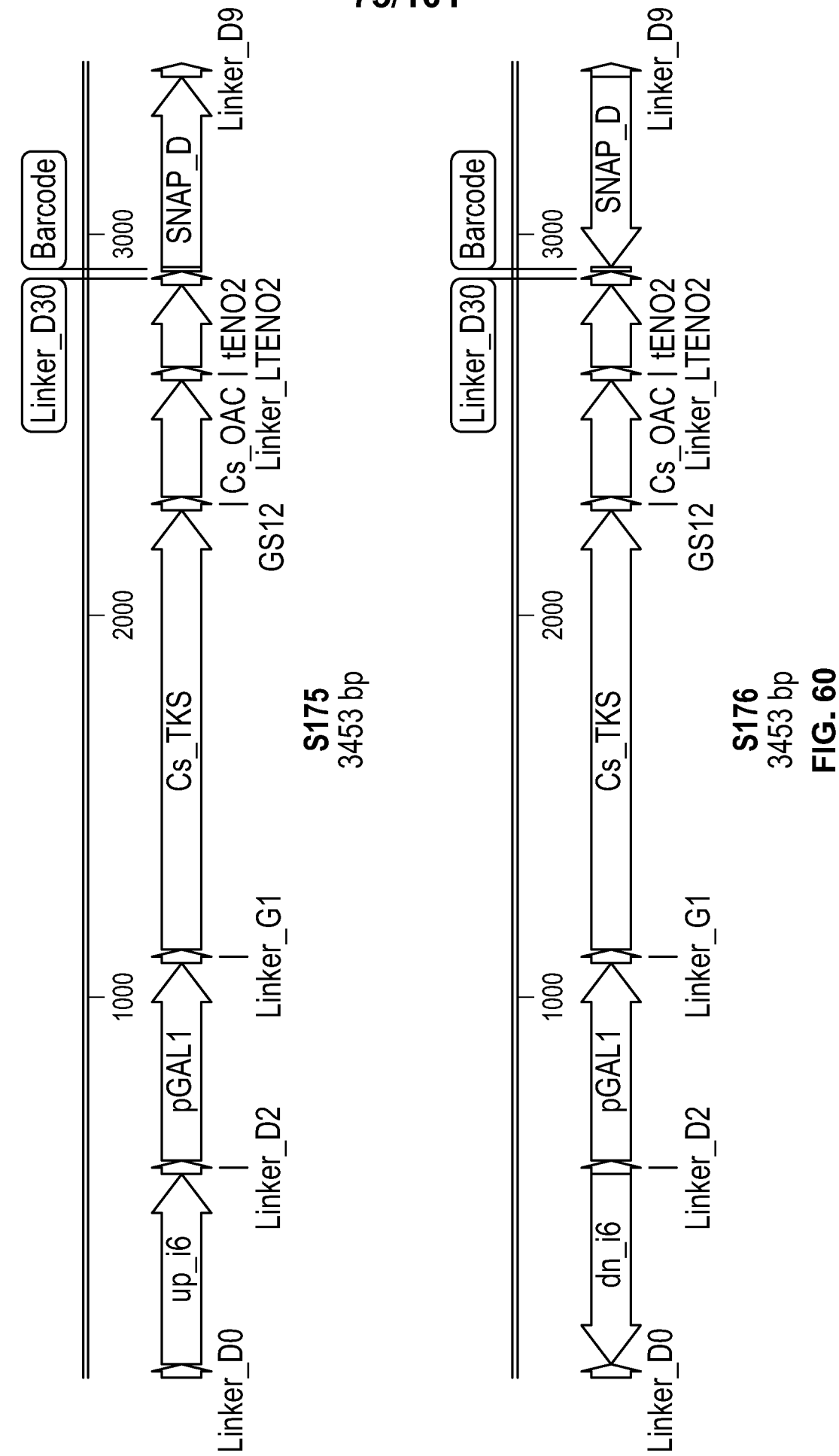

Linker_D0
up_i6
Linker_D2
pGAL1
Linker_G1
Cs_TKS
GS12
Cs_OAC
tENO2
Linker_LTENO2
Linker_D30
Barcode
Linker_D9
SNAP_D
1000
2000
3000
S175
3453 bp
Linker_D0
dn_i6
Linker_D2
pGAL1
Linker_G1
Cs_TKS
GS12
Cs_OAC
tENO2
Linker_LTENO2
Linker_D30
Barcode
Linker_D9
SNAP_D
1000
2000
3000
S176
3453 bp

FIG. 60

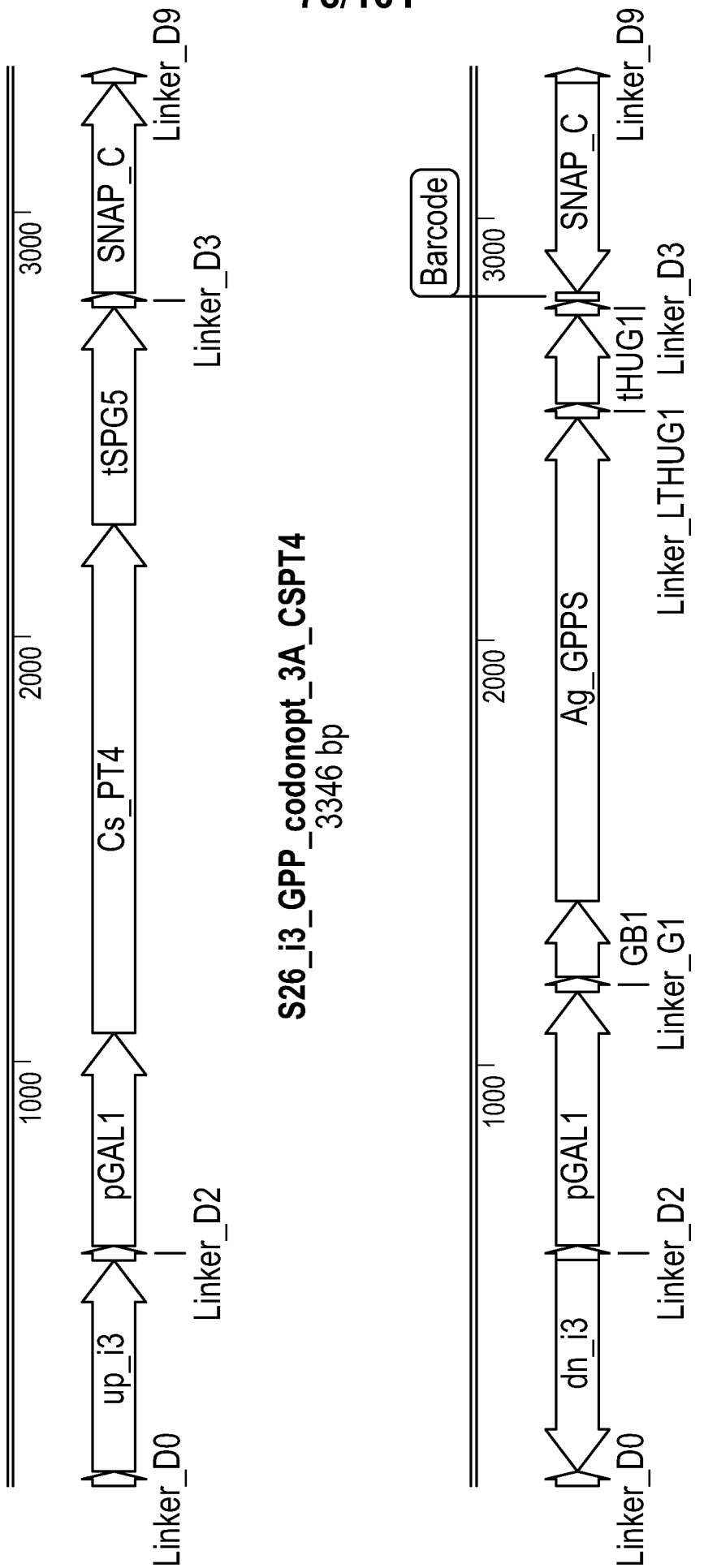
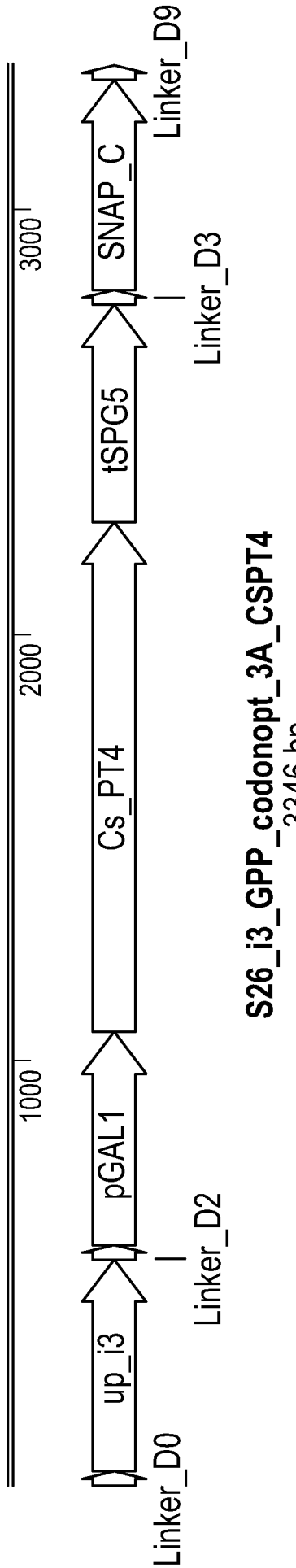
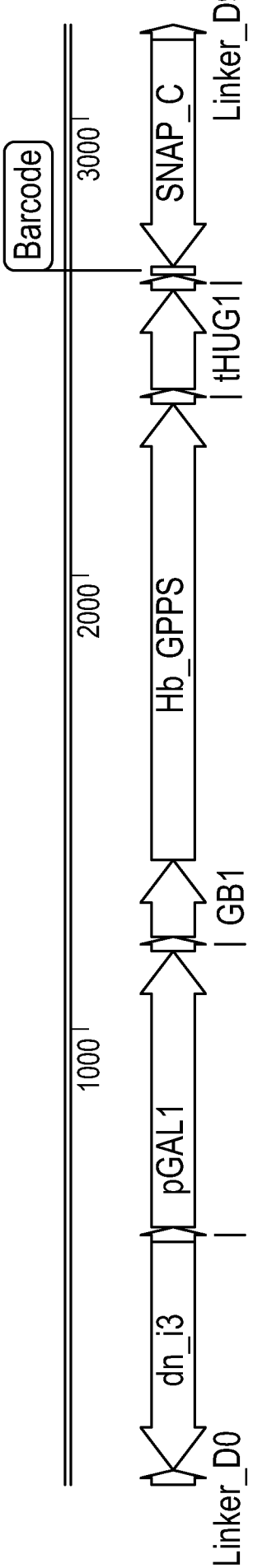

1000
2000
3000
Linker_D0
up_i3
Linker_D2
pGAL1
Cs_PT4
Linker_D3
tSPG5
SNAP_C
Linker_D9
S26_i3_GPP_codonopt_3A_CSPT4
3346 bp
Barcode
dn_i3
GB1
Linker_G1
Ag_GPPS
tHUG1
Linker_LTHUG1
S151
3363 bp

FIG. 61



S26_i3_GPP_codonopt_3A_CSPT4
3346 bp



S156
3210 bp

FIG. 62

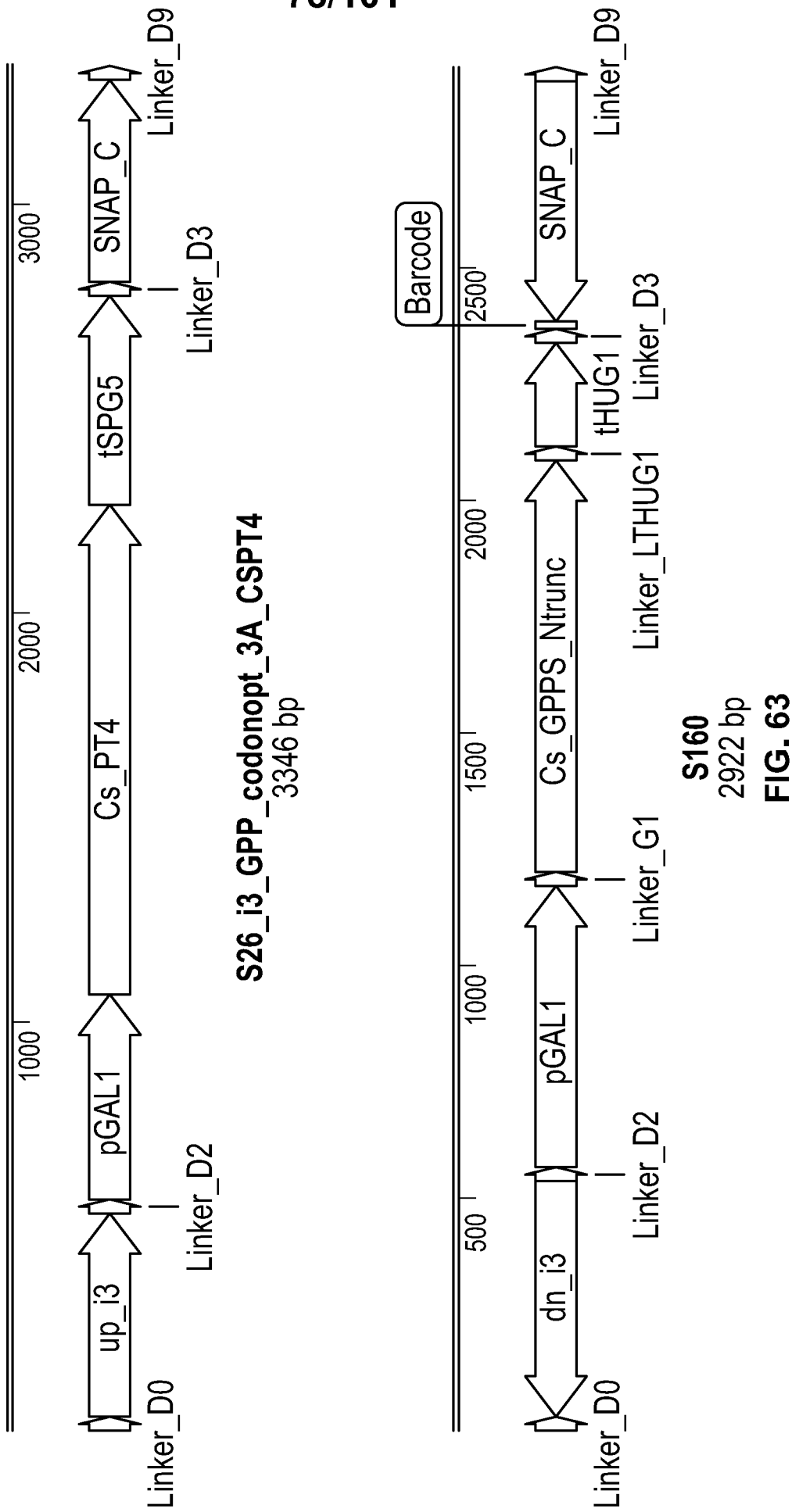

Linker_D0
up_i3
Linker_D2
pGAL1
Cs_PT4
tSPG5
Linker_D3
SNAP_C
Linker_D9
1000
2000
3000
S26_i3_GPP_codonopt_3A_CSPT4
3346 bp
Linker_D0
dn_i3
Linker_D2
pGAL1
Linker_G1
Cs_GPPS_Ntrunc
Linker_LTHUG1
tHUG1
Linker_D3
Barcode
SNAP_C
Linker_D9
500
1000
1500
2000
2500
S160
2922 bp

FIG. 63

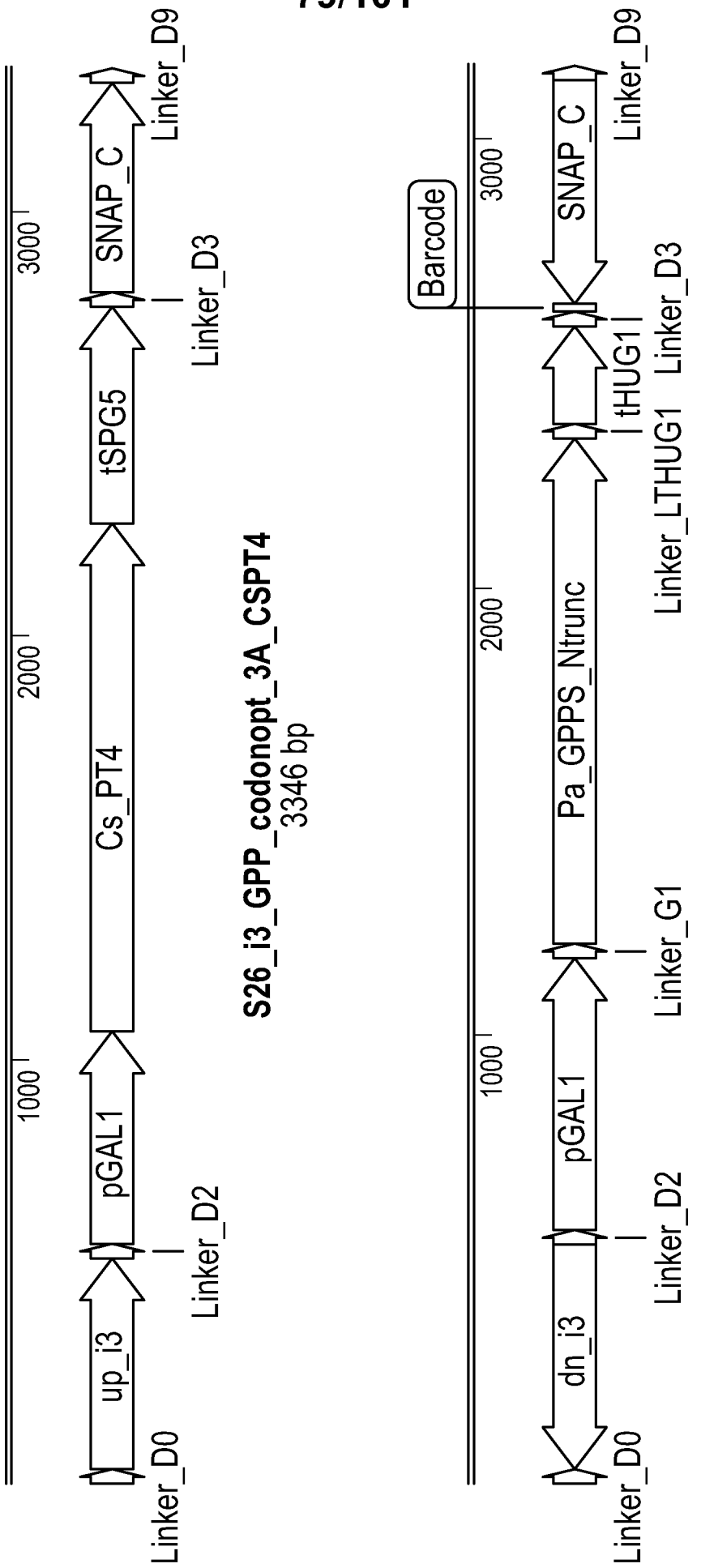

Linker_D0
up_i3
Linker_D2
pGAL1
Cs_PT4
tSPG5
Linker_D3
SNAP_C
Linker_D9
1000
2000
3000
S26_i3_GPP_codonopt_3A_CSPT4
3346 bp
Linker_D0
dn_i3
Linker_D2
pGAL1
Linker_G1
Pa_GPPS_Ntrunc
Linker_LTHUG1
tHUG1
Linker_D3
Barcode
SNAP_C
Linker_D9
1000
2000
3000
S162
3183 bp

FIG. 64

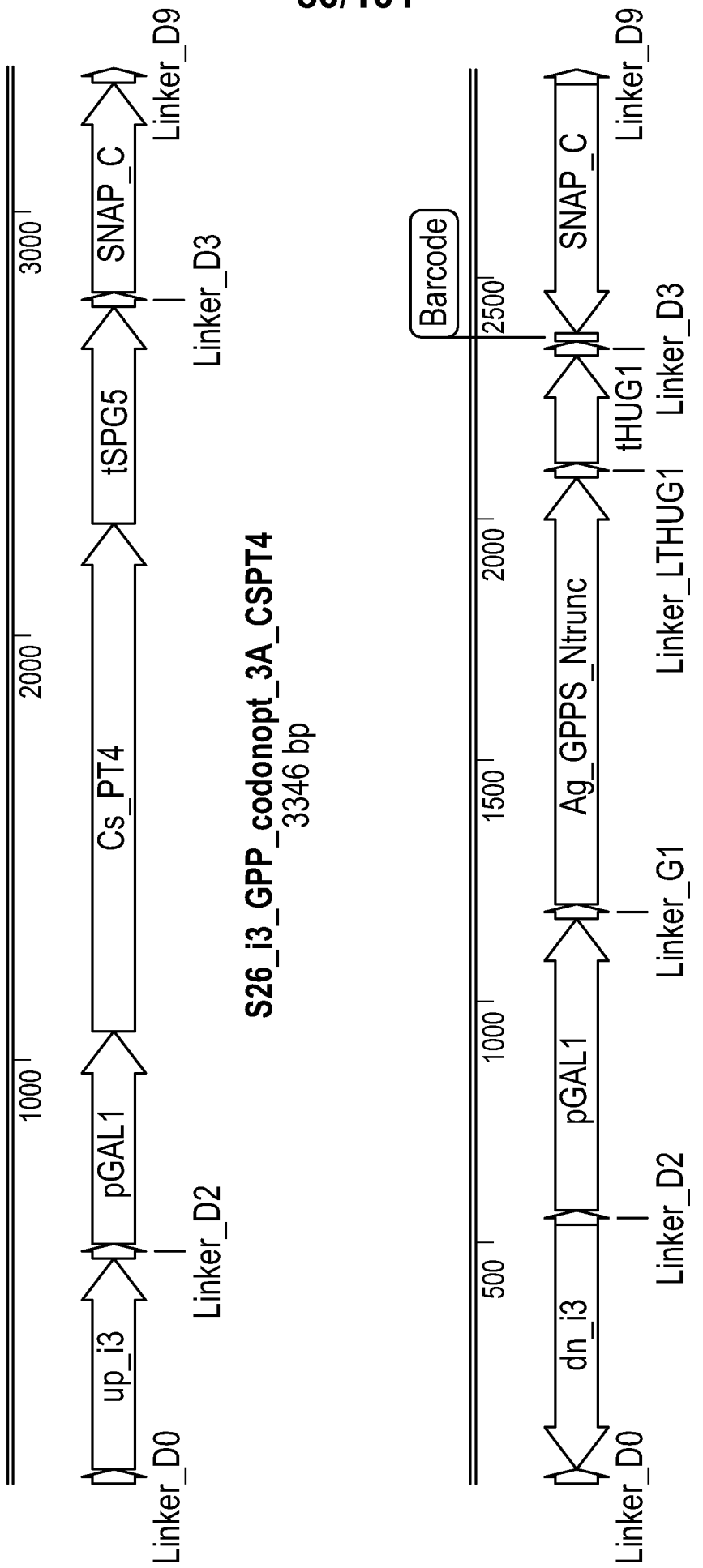

Linker_D0
up_i3
Linker_D2
pGAL1
Cs_PT4
tSPG5
Linker_D3
SNAP_C
Linker_D9
1000
2000
3000
S26_i3_GPP_codonopt_3A_CSPT4
3346 bp
Linker_D0
dn_i3
Linker_D2
pGAL1
Ag_GPPS_Ntrunc
Linker_G1
tHUG1
Linker_LTHUG1
Barcode
Linker_D3
SNAP_C
Linker_D9
500
1000
1500
2000
2500
S163
2937 bp

FIG. 65

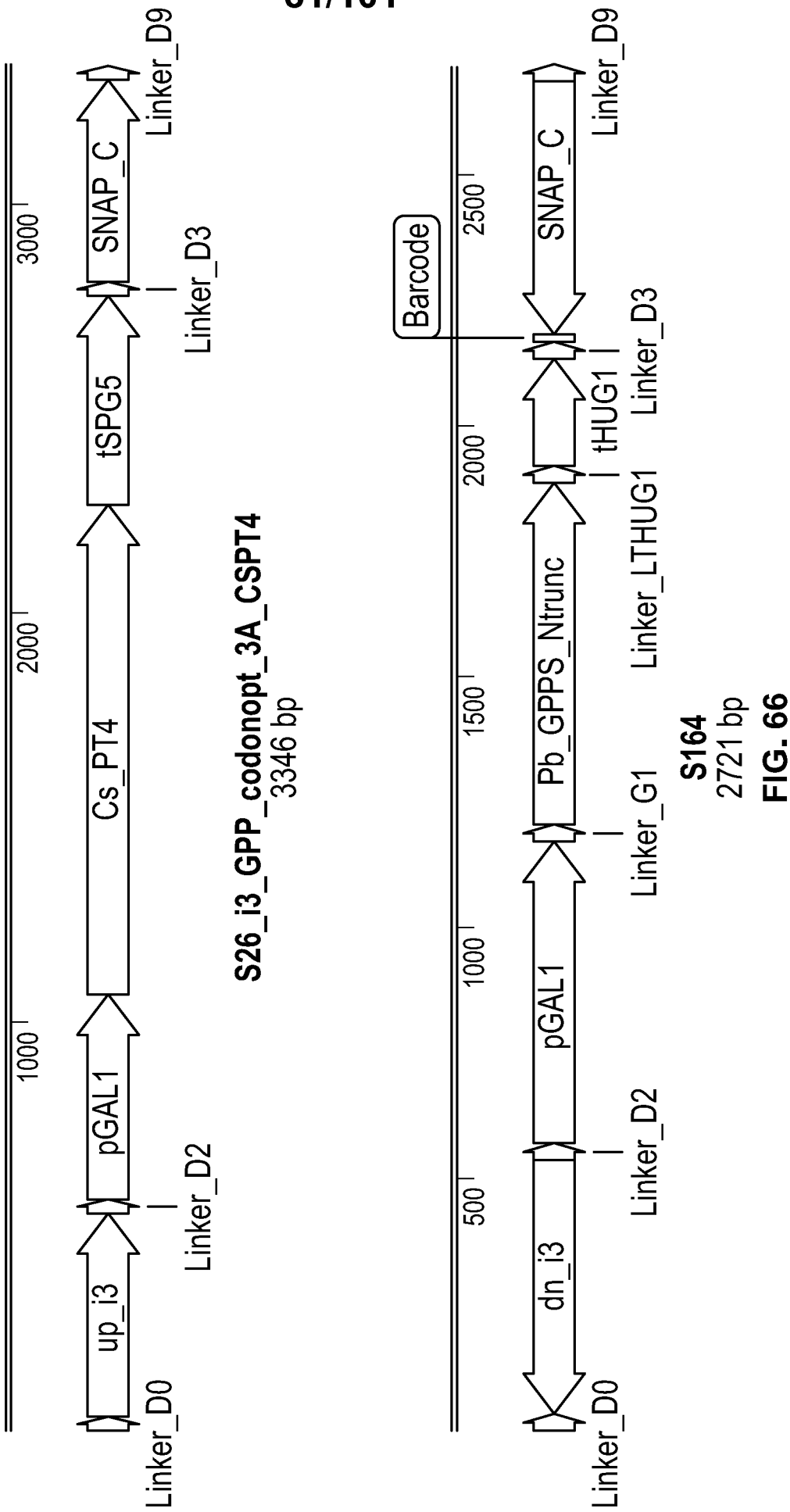

Linker_D0
up_i3
Linker_D2
pGAL1
Cs_PT4
tSPG5
Linker_D3
SNAP_C
Linker_D9
1000
2000
3000
S26_i3_GPP_codonopt_3A_CSPT4
3346 bp
Linker_D0
dn_i3
Linker_D2
pGAL1
Pb_GPPS_Ntrunc
Linker_G1
Barcode
Linker_LTHUG1
tHUG1
Linker_D3
SNAP_C
Linker_D9
500
1000
1500
2000
2500
S164
2721 bp

FIG. 66

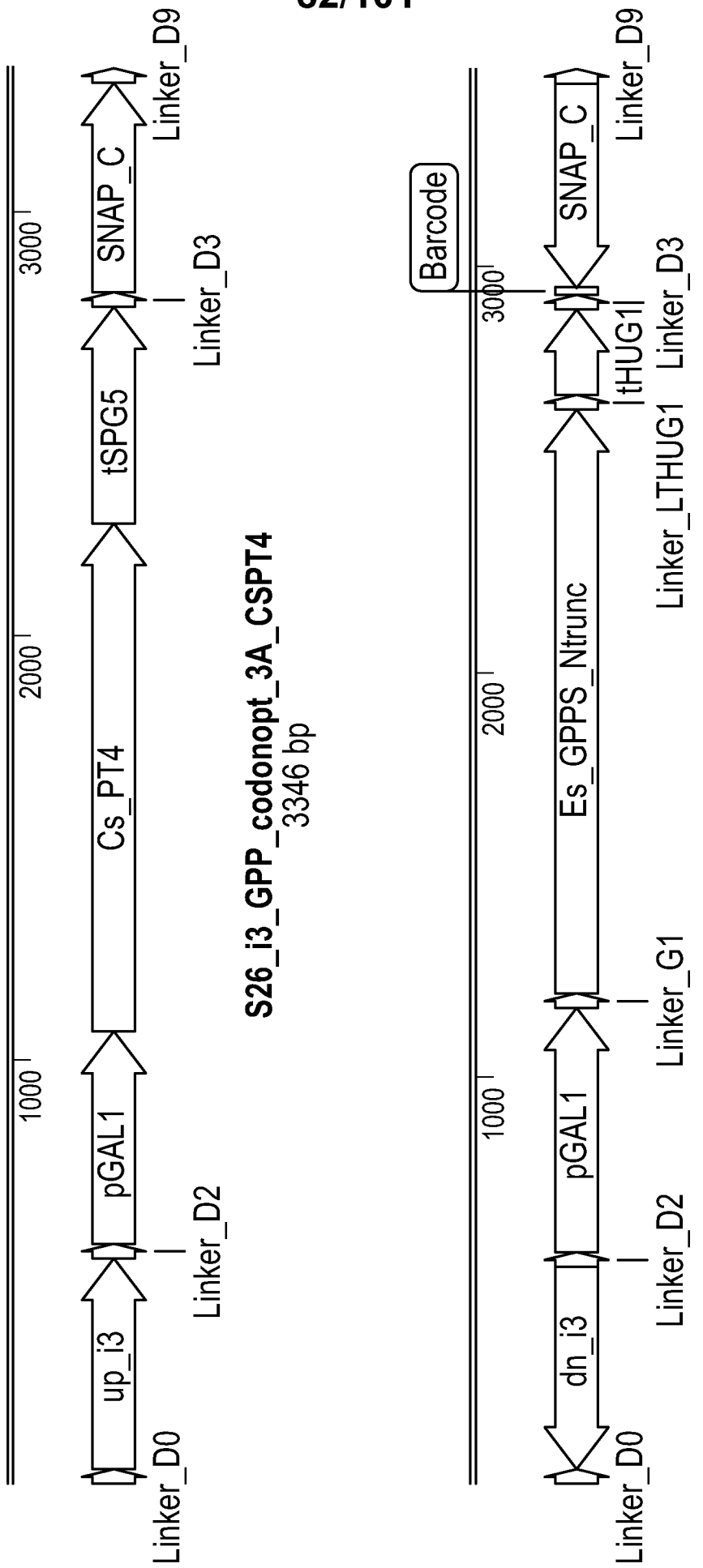

1000
2000
3000
Linker_D0
up_i3
Linker_D2
pGAL1
Cs_PT4
Linker_D3
tSPG5
SNAP_C
Linker_D9
S26_i3_GPP_codonopt_3A_CSPT4
3346 bp
Barcode
dn_i3
Es_GPPS_Ntrunc
tHUG1
Linker_LTHUG1
S166
3483 bp

FIG. 67

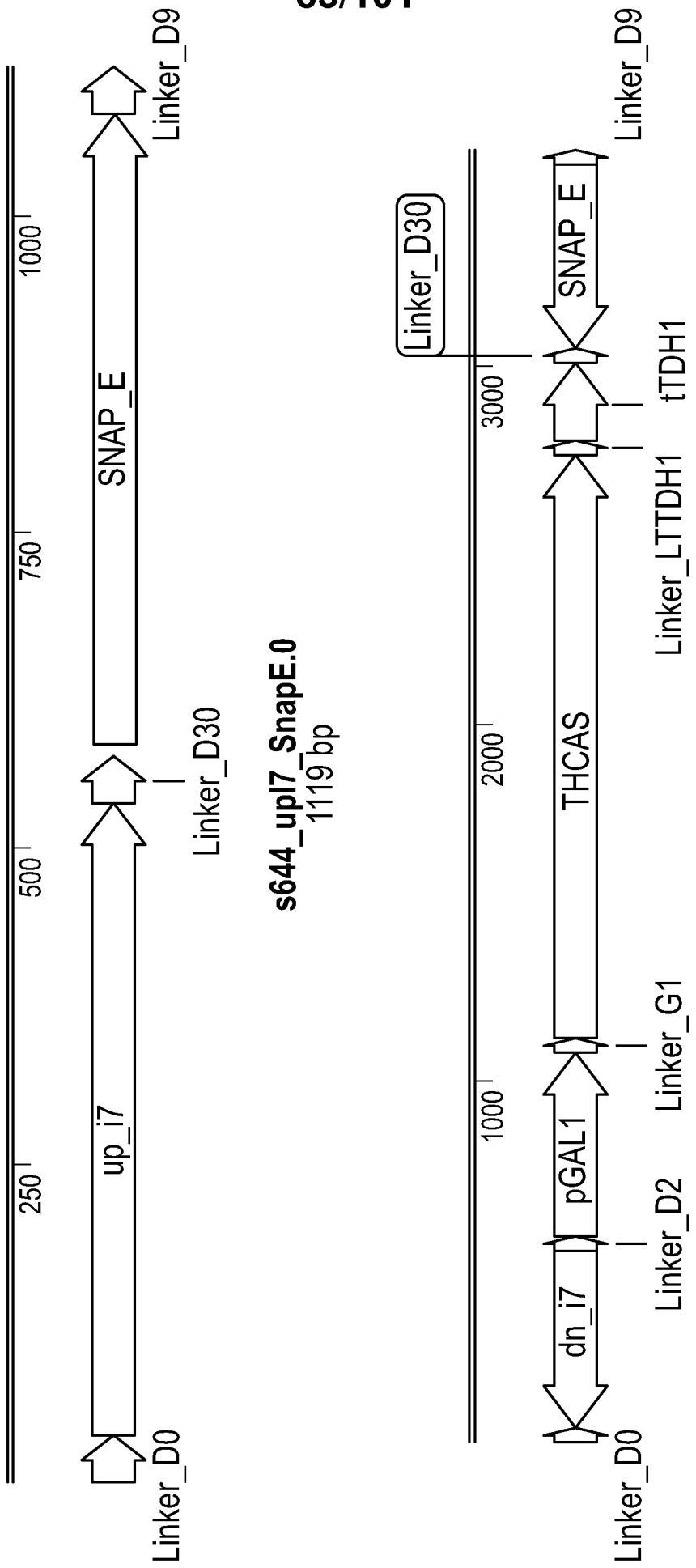
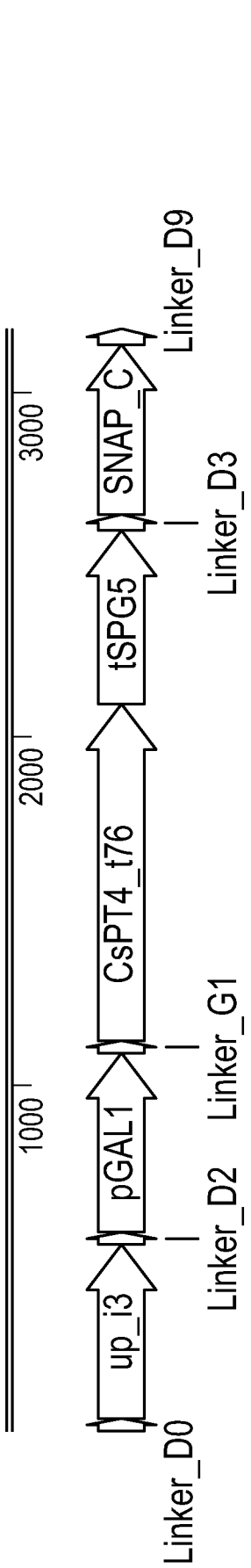
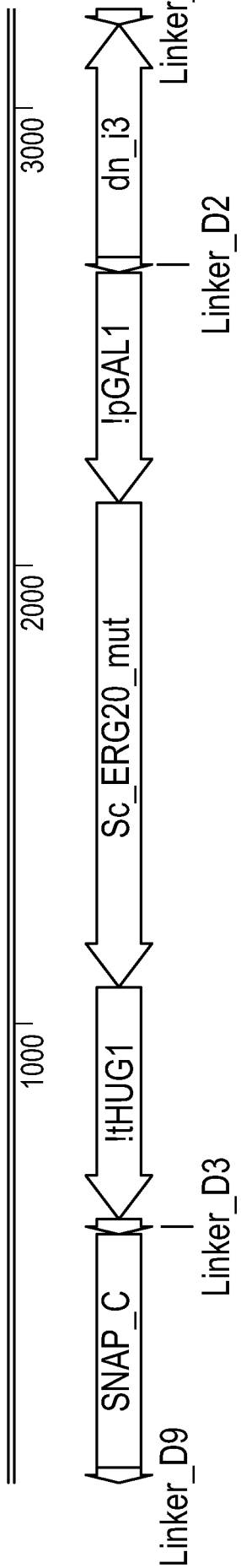

Linker_D0
up_i7
SNAP_E
Linker_D9
Linker_D30
250
500
750
1000
s644_upI7_SnapE.0
1119 bp
Linker_D0
dn_i7
pGAL1
Linker_D2
Linker_G1
THCAS
Linker_LTTDH1
tTDH1
SNAP_E
Linker_D9
Linker_D30
1000
2000
3000
s483_i7b_THCASfull.1
3596 bp

FIG. 68



i3B_CsPT4_t76
3175 bp



S9_i3_GPP_codonopt_3B_0
3212 bp

FIG. 69

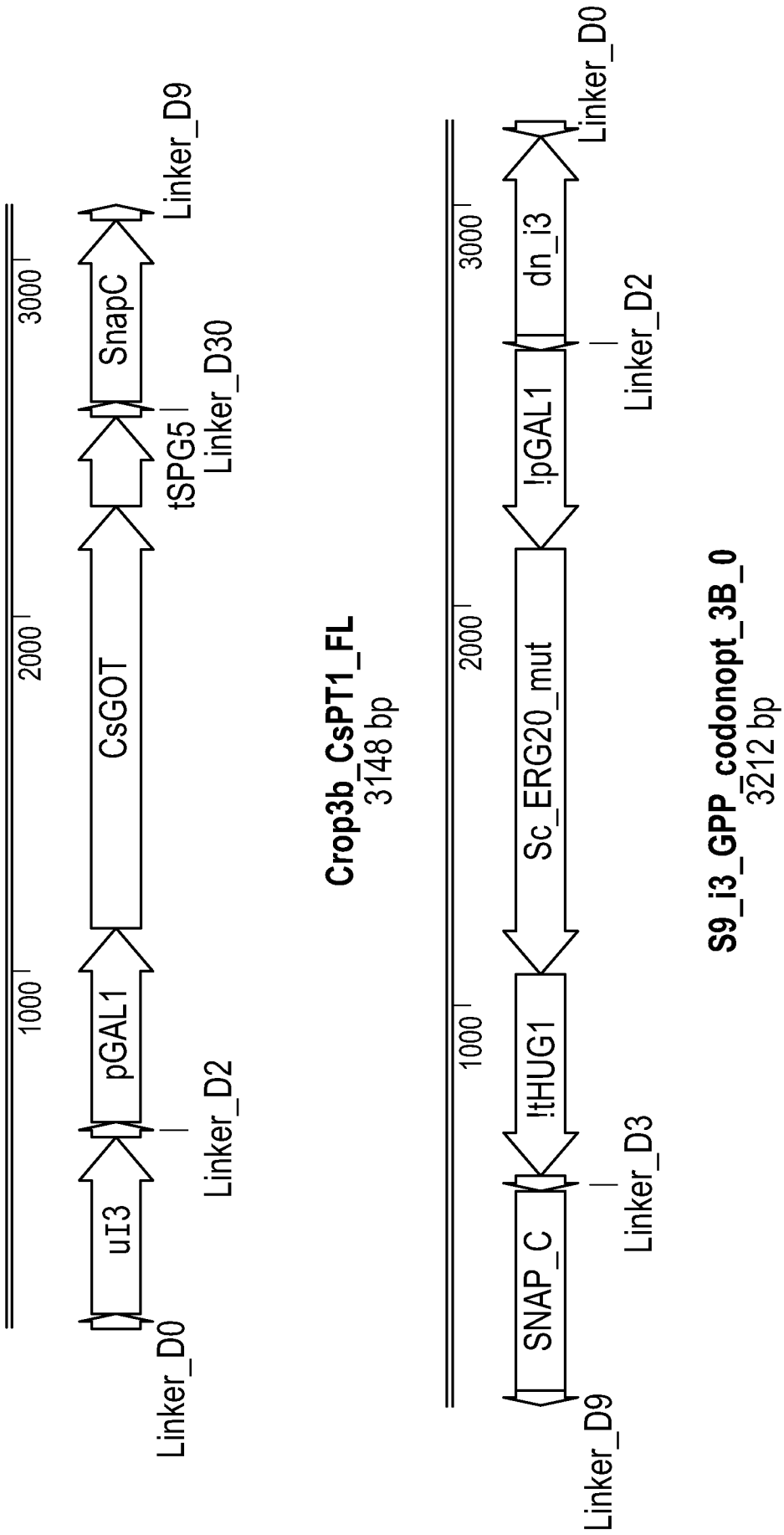

1000
2000
3000
Linker_D0
uI3
Linker_D2
pGAL1
CsGOT
tSPG5
Linker_D30
SnapC
Linker_D9
Crop3b_CsPT1_FL
3148 bp
1000
2000
3000
Linker_D9
SNAP_C
Linker_D3
!tHUG1
Sc_ERG20_mut
!pGAL1
Linker_D2
dn_i3
Linker_D0
S9_i3_GPP_codonopt_3B_0
3212 bp

FIG. 70

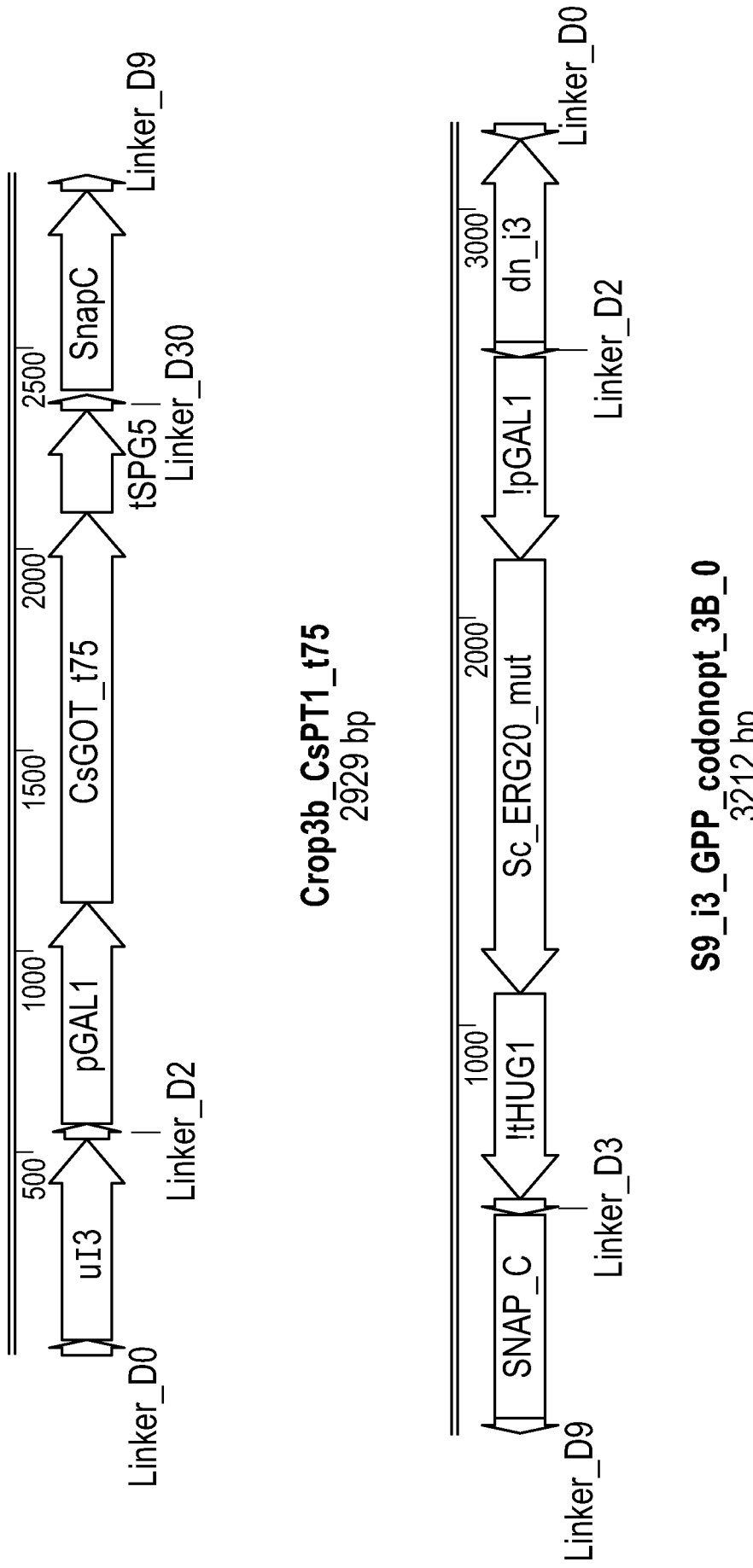

Linker_D0
uI3
500
Linker_D2
pGAL1
1000
CsGOT_t75
1500
tSPG5
2000
Linker_D30
SnapC
2500
Linker_D9
Crop3b_CsPT1_t75
2929 bp
Linker_D9
SNAP_C
Linker_D3
!tHUG1
1000
Sc_ERG20_mut
2000
!pGAL1
Linker_D2
dn_i3
3000
Linker_D0
S9_i3_GPP_codonopt_3B_0
3212 bp

FIG. 71

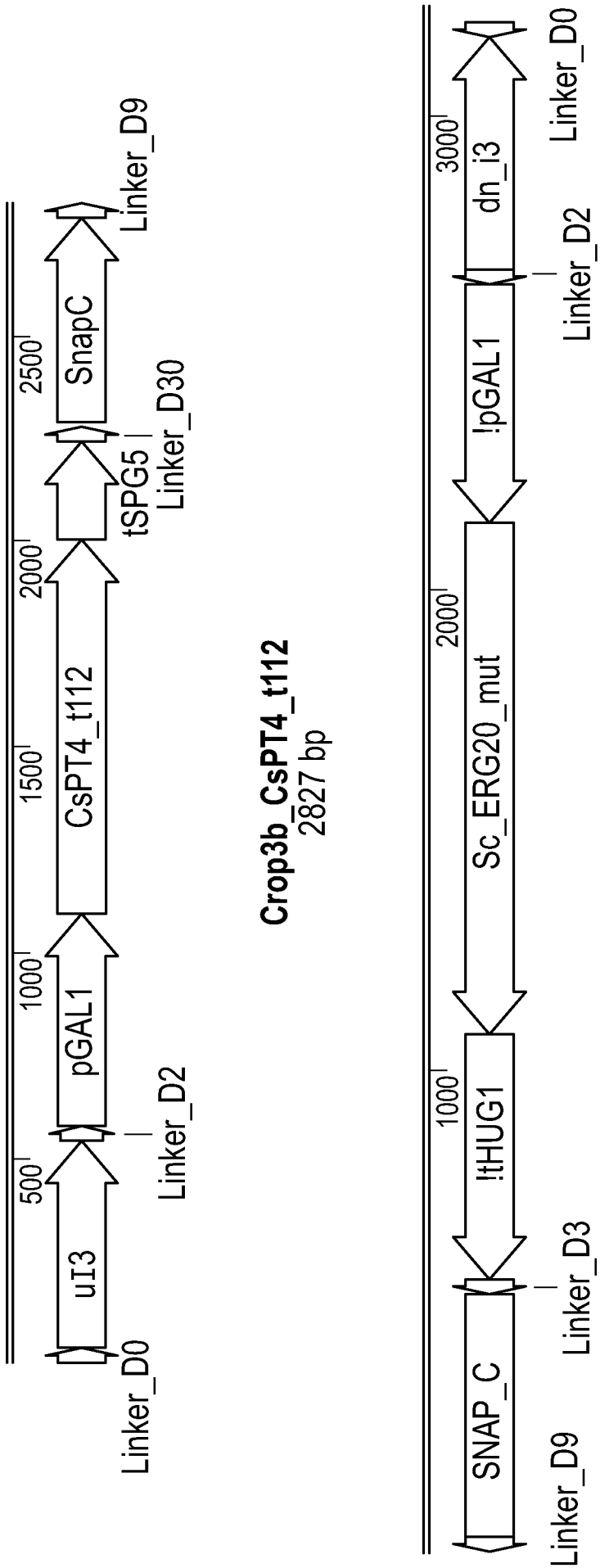

Linker_D0
uI3
500
Linker_D2
pGAL1
1000
CsPT4_t112
1500
tSPG5
Linker_D30
2000
SnapC
2500
Linker_D9
Crop3b_CsPT4_t112
2827 bp
Linker_D9
SNAP_C
Linker_D3
!tHUG1
1000
Sc_ERG20_mut
2000
!pGAL1
Linker_D2
dn_i3
3000
Linker_D0
S9_i3_GPP_codonopt_3B_0
3212 bp

FIG. 72

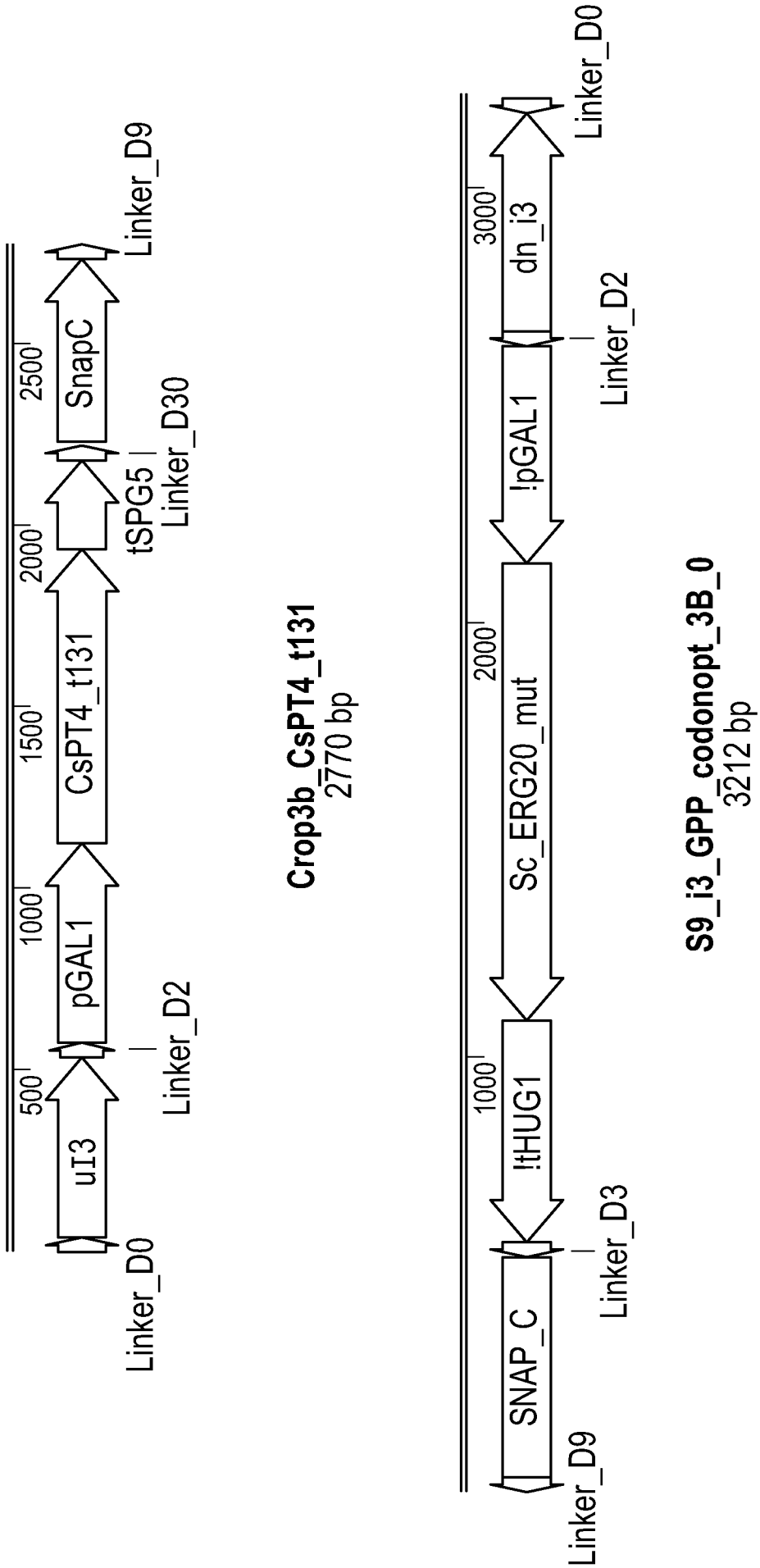

Linker_D0
uI3
500
Linker_D2
pGAL1
1000
CsPT4_t131
1500
tSPG5
2000
Linker_D30
SnapC
2500
Linker_D9
Crop3b_CsPT4_t131
2770 bp
Linker_D9
SNAP_C
Linker_D3
!tHUG1
1000
Sc_ERG20_mut
2000
!pGAL1
Linker_D2
dn_i3
3000
Linker_D0
S9_i3_GPP_codonopt_3B_0
3212 bp

FIG. 73

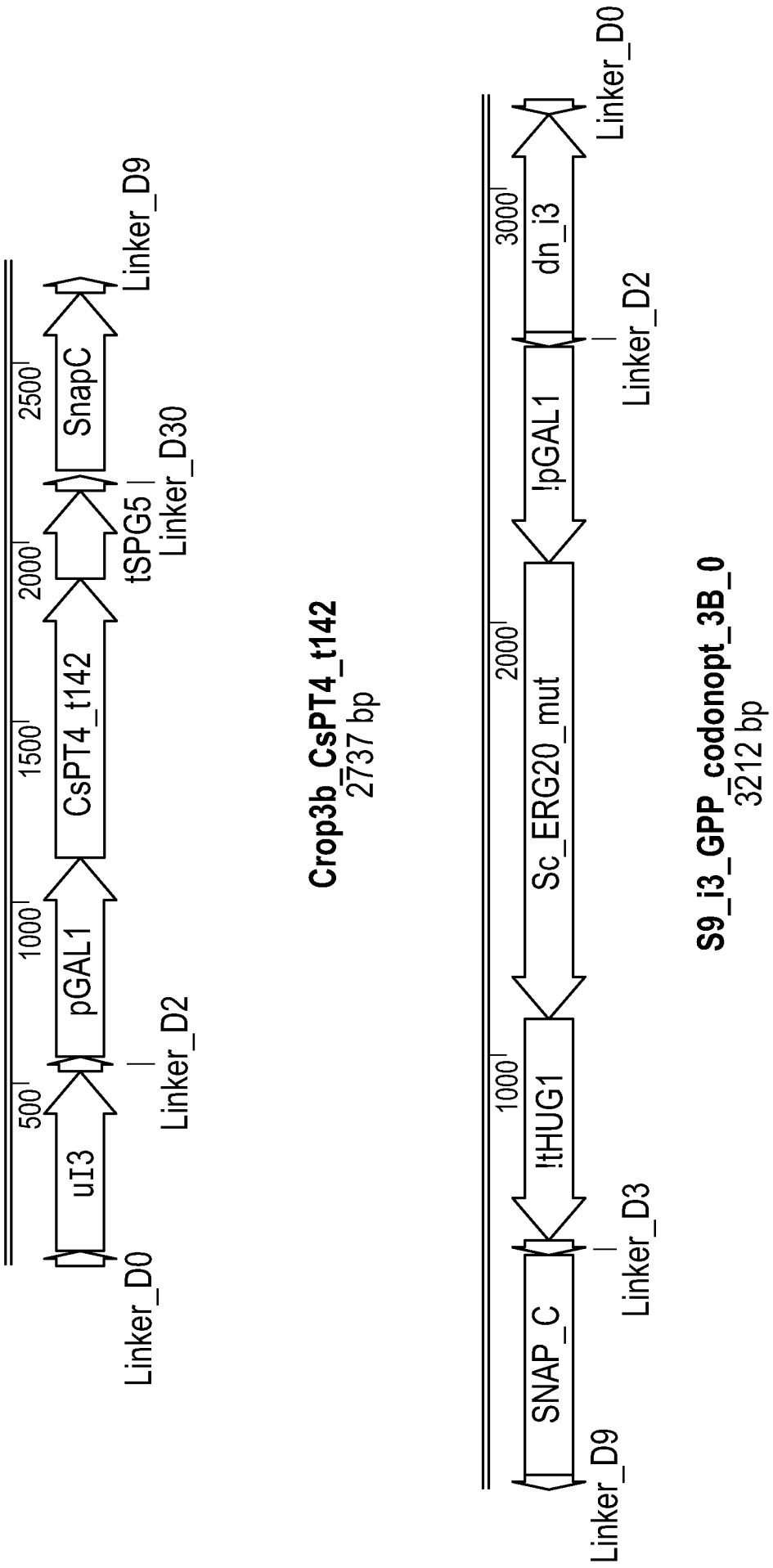

Linker_D0
uI3
500
Linker_D2
pGAL1
1000
CsPT4_t142
1500
tSPG5
2000
Linker_D30
SnapC
2500
Linker_D9
Crop3b_CsPT4_t142
2737 bp
Linker_D9
SNAP_C
Linker_D3
!tHUG1
1000
Sc_ERG20_mut
2000
!pGAL1
Linker_D2
dn_i3
3000
Linker_D0
S9_i3_GPP_codonopt_3B_0
3212 bp

FIG. 74

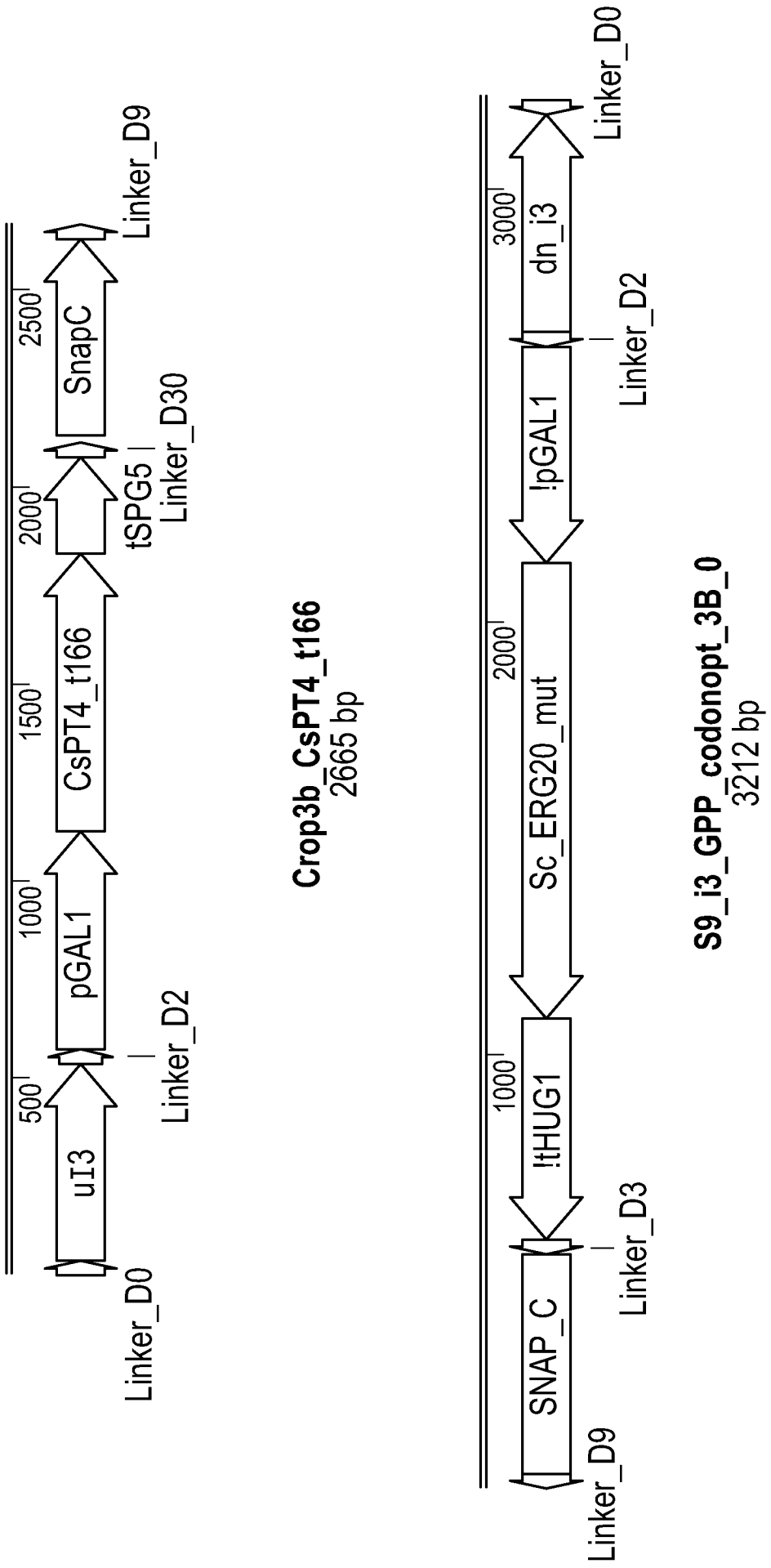

Linker_D0
uI3
500
Linker_D2
pGAL1
1000
CsPT4_t166
1500
tSPG5
Linker_D30
2000
SnapC
2500
Linker_D9
Crop3b_CsPT4_t166
2665 bp
Linker_D0
dn_i3
3000
Linker_D2
!pGAL1
Sc_ERG20_mut
2000
!tHUG1
1000
Linker_D3
SNAP_C
Linker_D9
S9_i3_GPP_codonopt_3B_0
3212 bp

FIG. 75

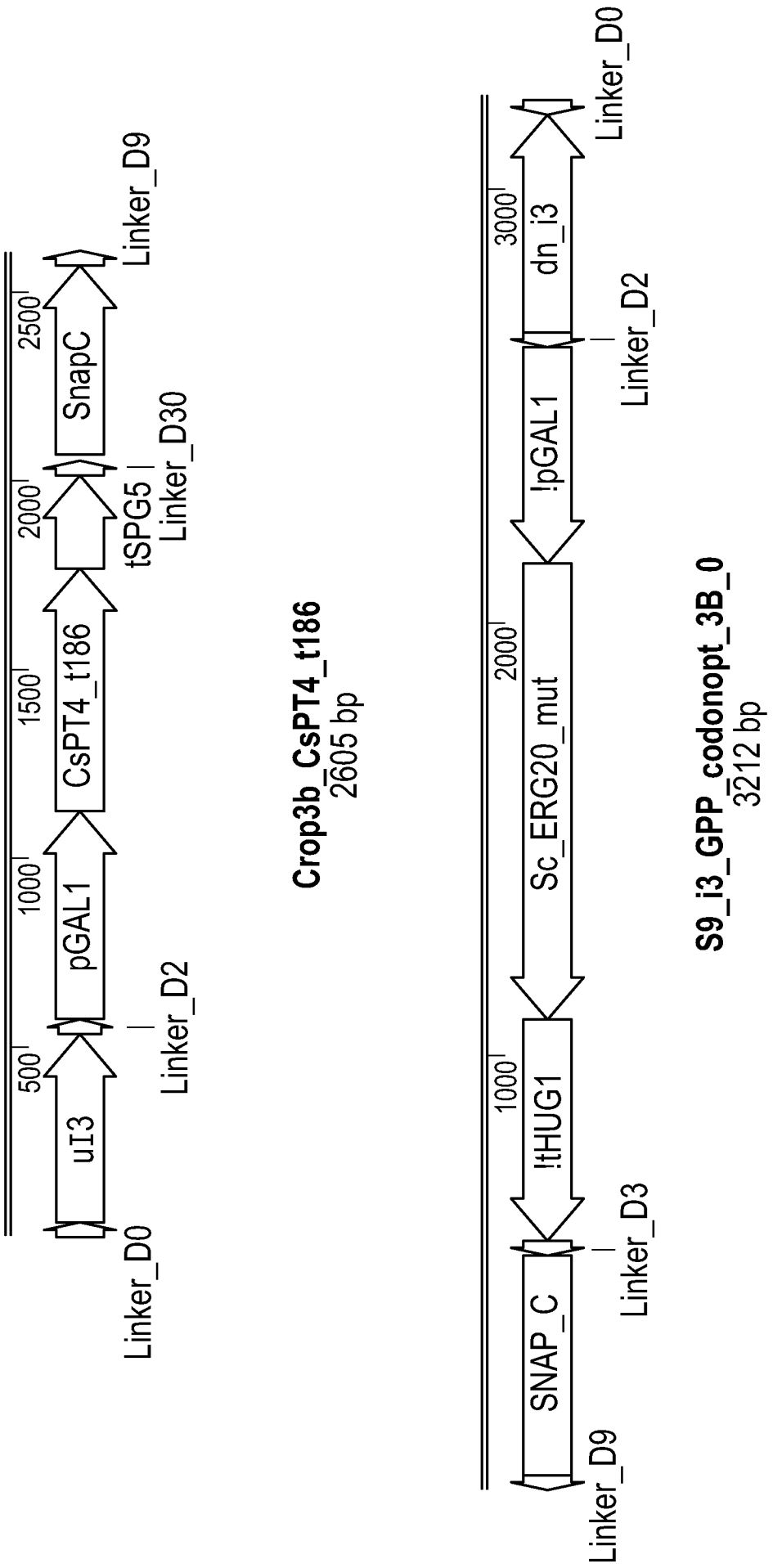

Linker_D0
uI3
500
Linker_D2
pGAL1
1000
CsPT4_t186
1500
tSPG5
Linker_D30
2000
SnapC
2500
Linker_D9
Crop3b_CsPT4_t186
2605 bp
Linker_D0
dn_i3
3000
Linker_D2
!pGAL1
Sc_ERG20_mut
2000
!tHUG1
1000
Linker_D3
SNAP_C
Linker_D9
S9_i3_GPP_codonopt_3B_0
3212 bp

FIG. 76

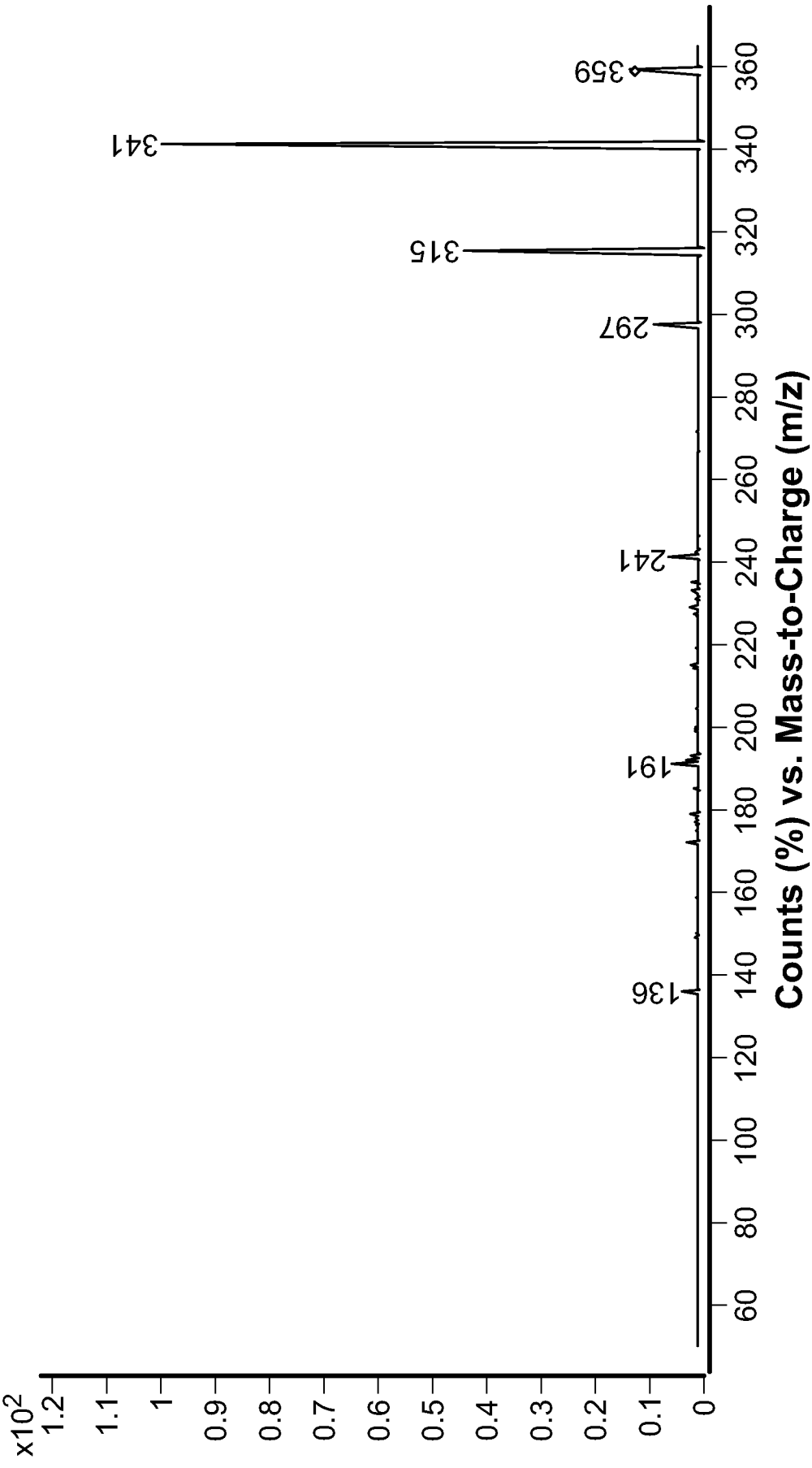

x10^2
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
60
80
100
120
140
160
180
200
220
240
260
280
300
320
340
360
136
191
241
297
315
341
359
Counts (%) vs. Mass-to-Charge (m/z)

FIG. 77

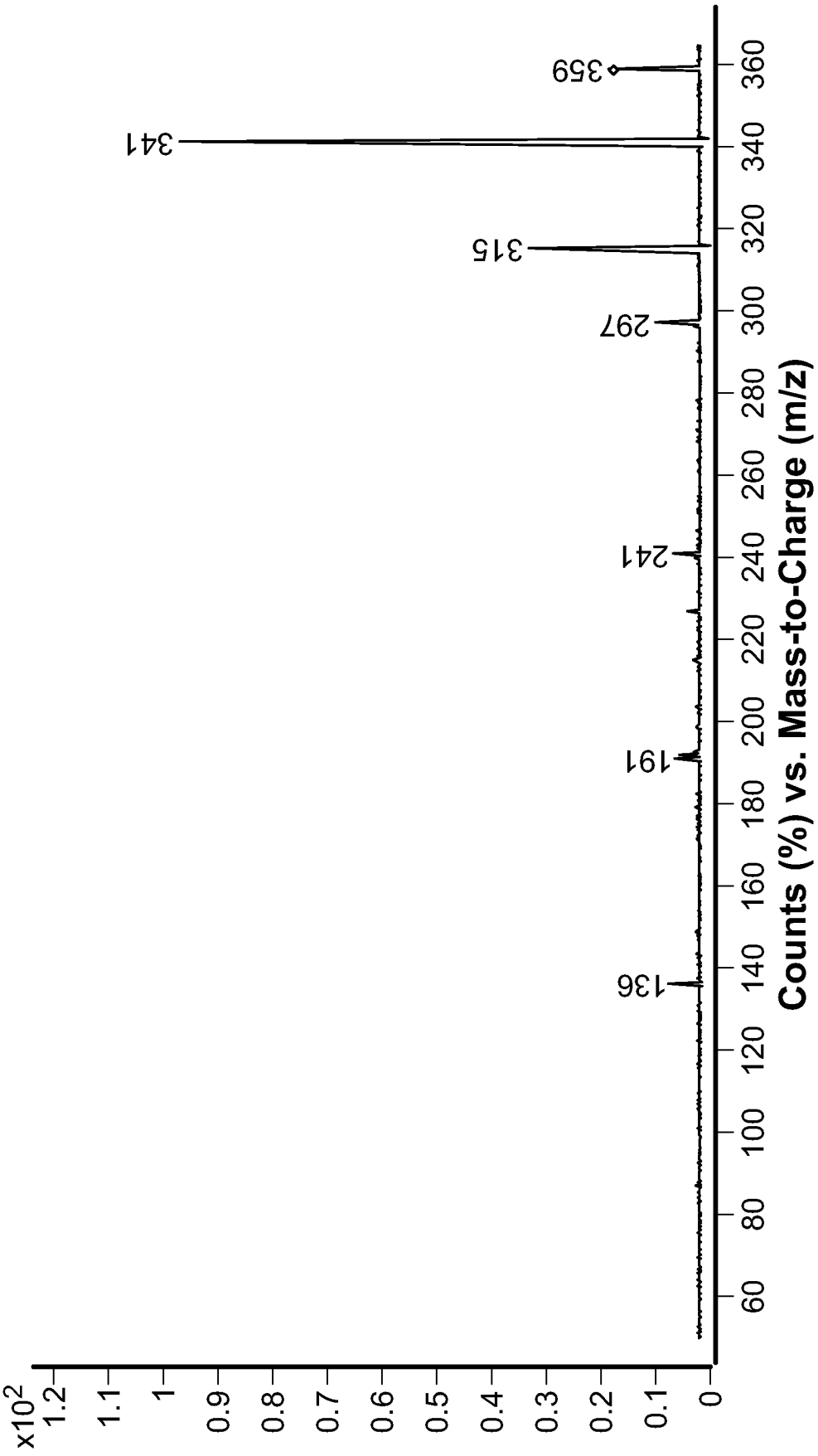

x10^2
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
60
80
100
120
140
160
180
200
220
240
260
280
300
320
340
360
136
191
241
297
315
341
359
Counts (%) vs. Mass-to-Charge (m/z)

FIG. 78

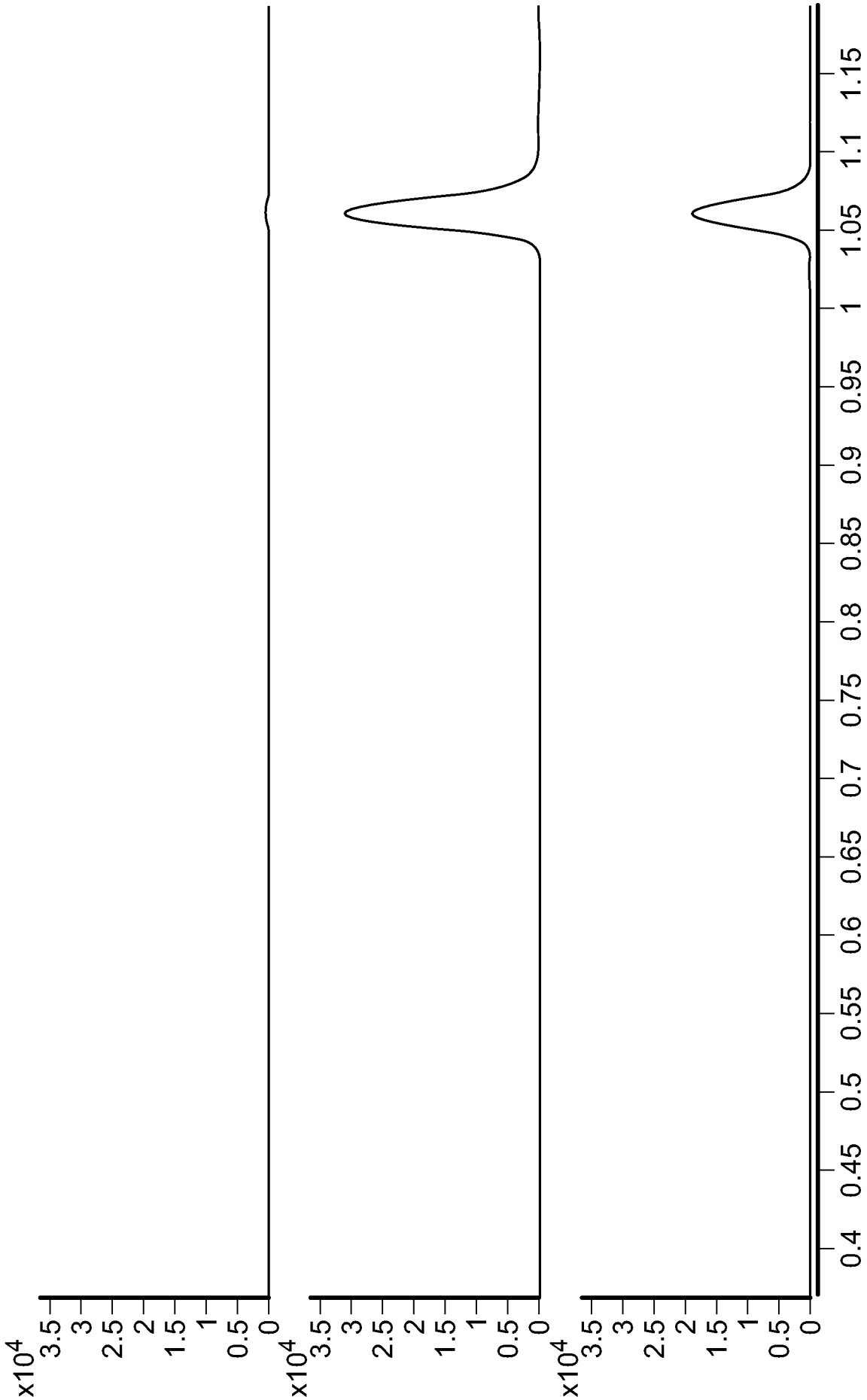

x10^4
3.5
3
2.5
2
1.5
1
0.5
0
0.4
0.45
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
0.95
1
1.05
1.1
1.15
Counts vs. Acquisition Time (min)

FIG. 79

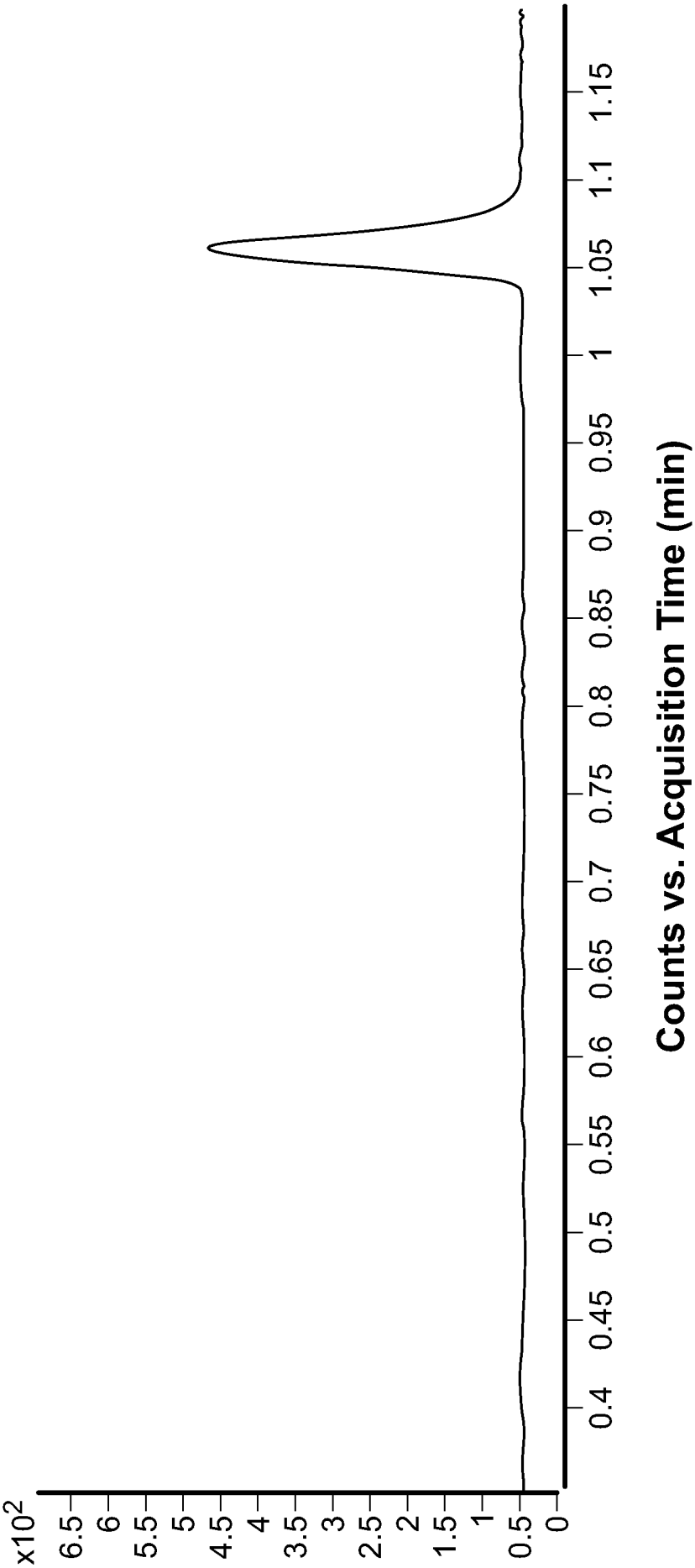

x10^2
6.5
6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0.4
0.45
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
0.95
1
1.05
1.1
1.15
Counts vs. Acquisition Time (min)

FIG. 80

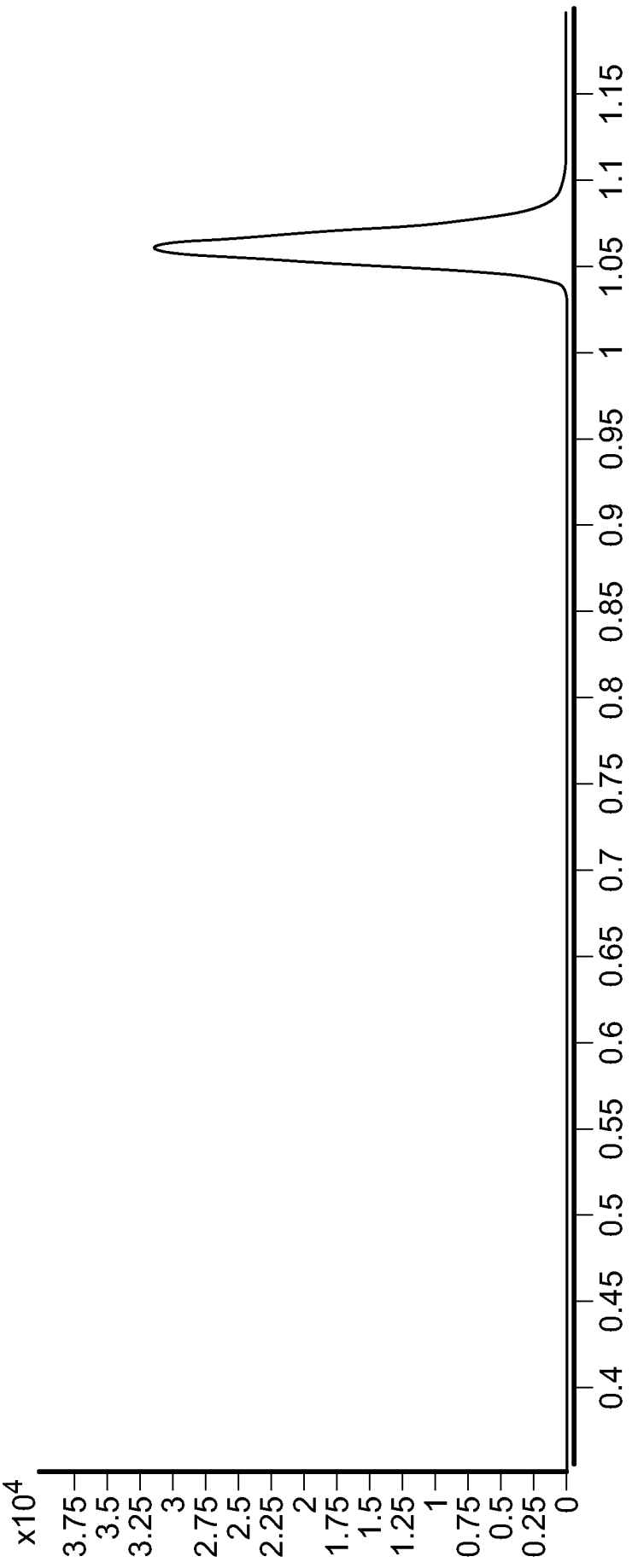

x10^4
3.75
3.5
3.25
3
2.75
2.5
2.25
2
1.75
1.5
1.25
1
0.75
0.5
0.25
0
0.4
0.45
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
0.95
1
1.05
1.1
1.15
Counts vs. Acquisition Time (min)

FIG. 81

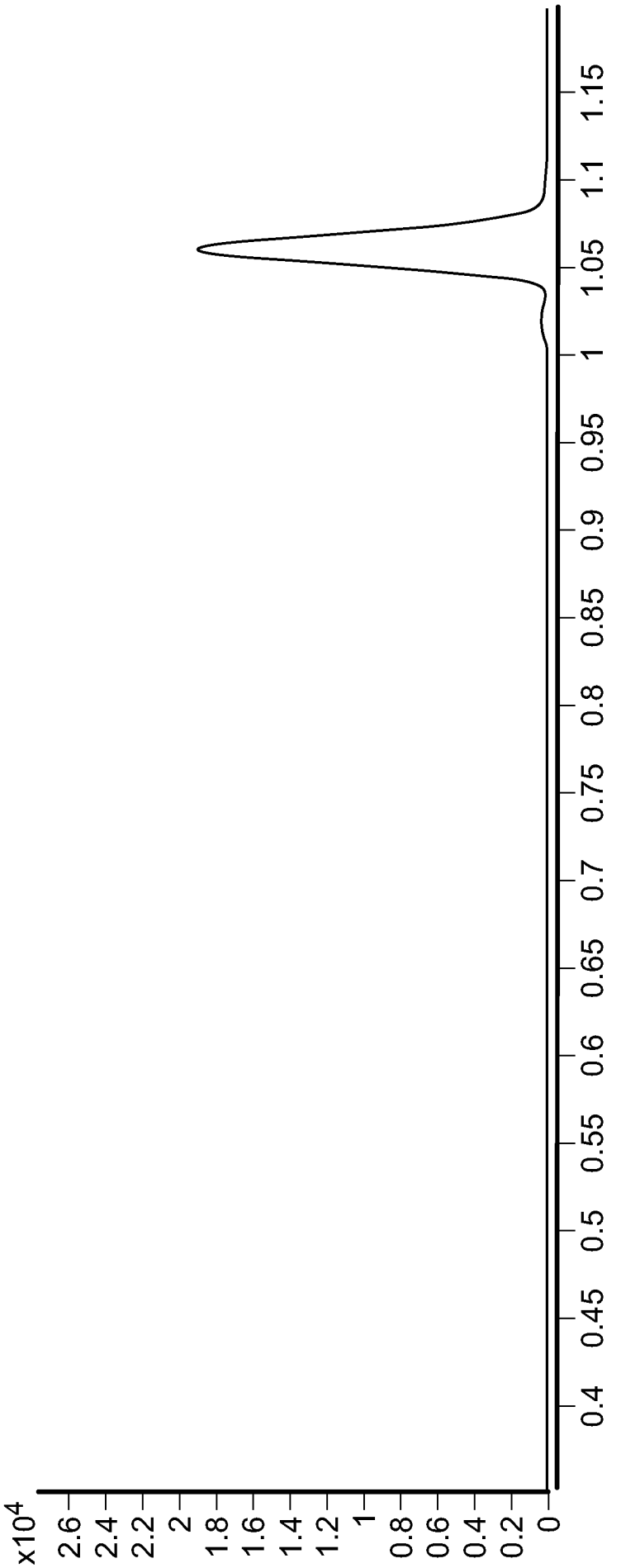

x10^4
2.6
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0.4
0.45
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
0.95
1
1.05
1.1
1.15
Counts vs. Acquisition Time (min)

FIG. 82

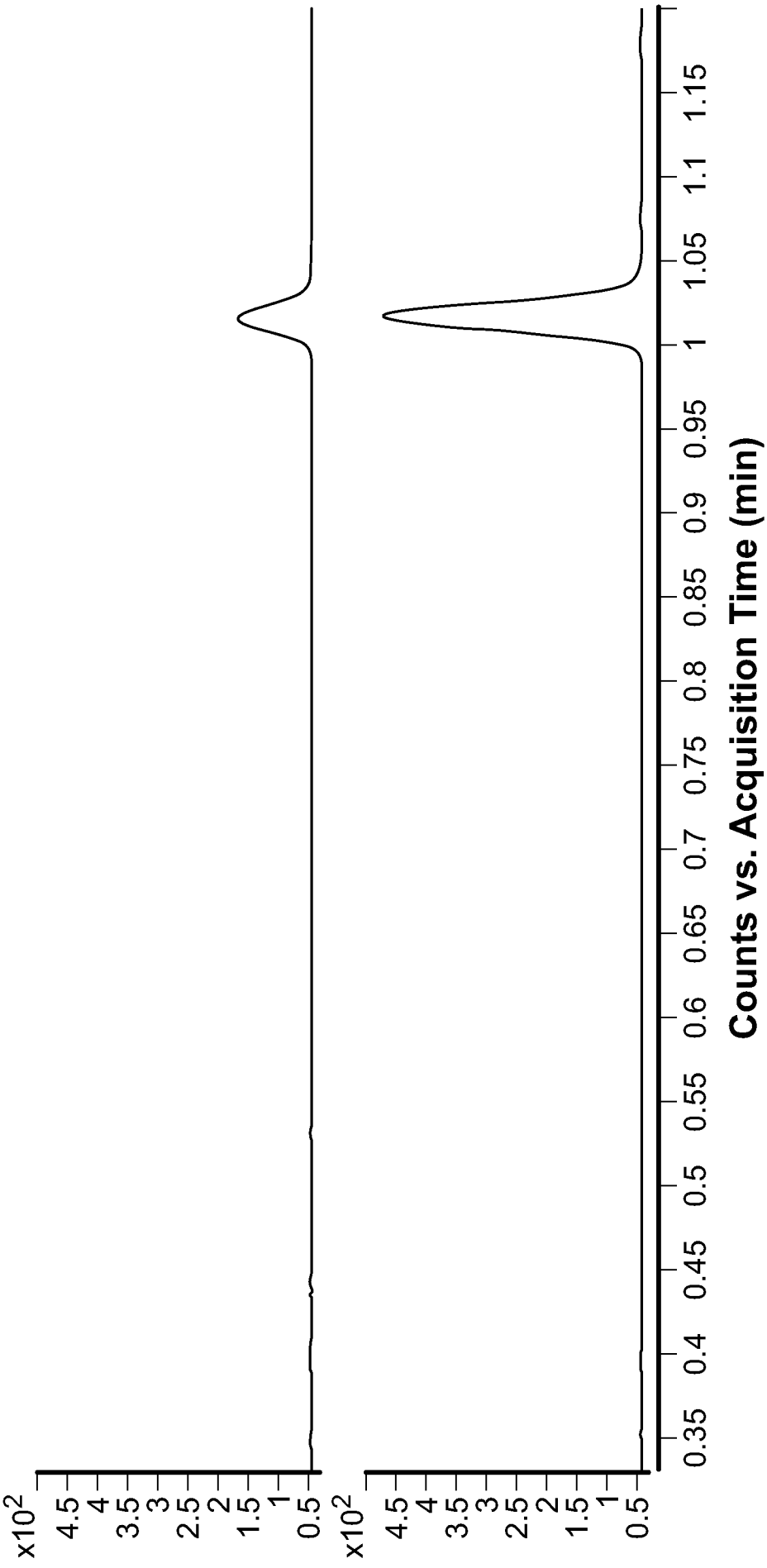

x10^2
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0.35
0.4
0.45
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
0.95
1
1.05
1.1
1.15
Counts vs. Acquisition Time (min)

FIG. 83

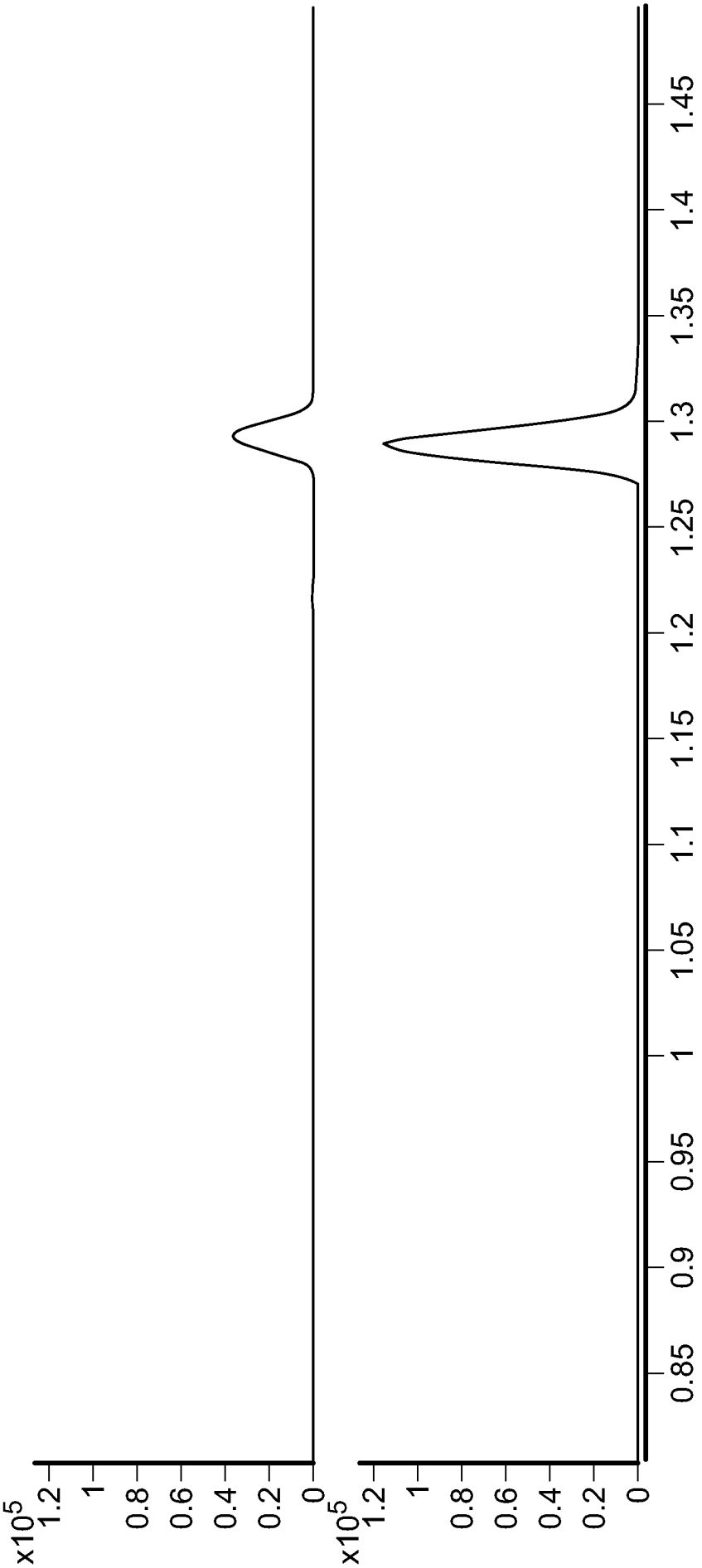

x10^5
1.2
1
0.8
0.6
0.4
0.2
0
0.85
0.9
0.95
1
1.05
1.1
1.15
1.2
1.25
1.3
1.35
1.4
1.45
Counts vs. Acquisition Time (min)

FIG. 84

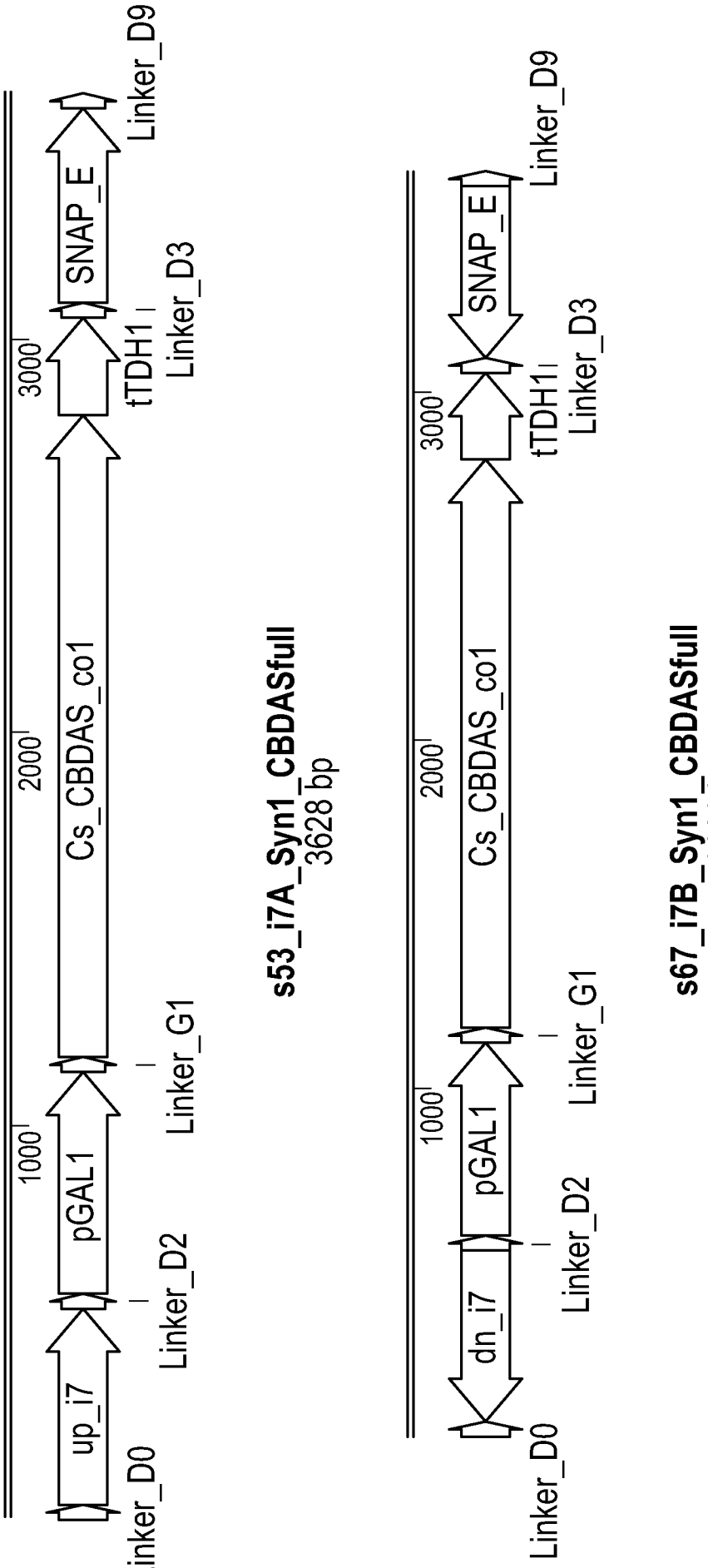

Linker_D0
up_i7
Linker_D2
pGAL1
Linker_G1
Cs_CBDAS_co1
tTDH1
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s53_i7A_Syn1_CBDASfull
3628 bp
Linker_D0
dn_i7
Linker_D2
pGAL1
Linker_G1
Cs_CBDAS_co1
tTDH1
Linker_D3
SNAP_E
Linker_D9
1000
2000
3000
s67_i7B_Syn1_CBDASfull
3628 bp

FIG. 85

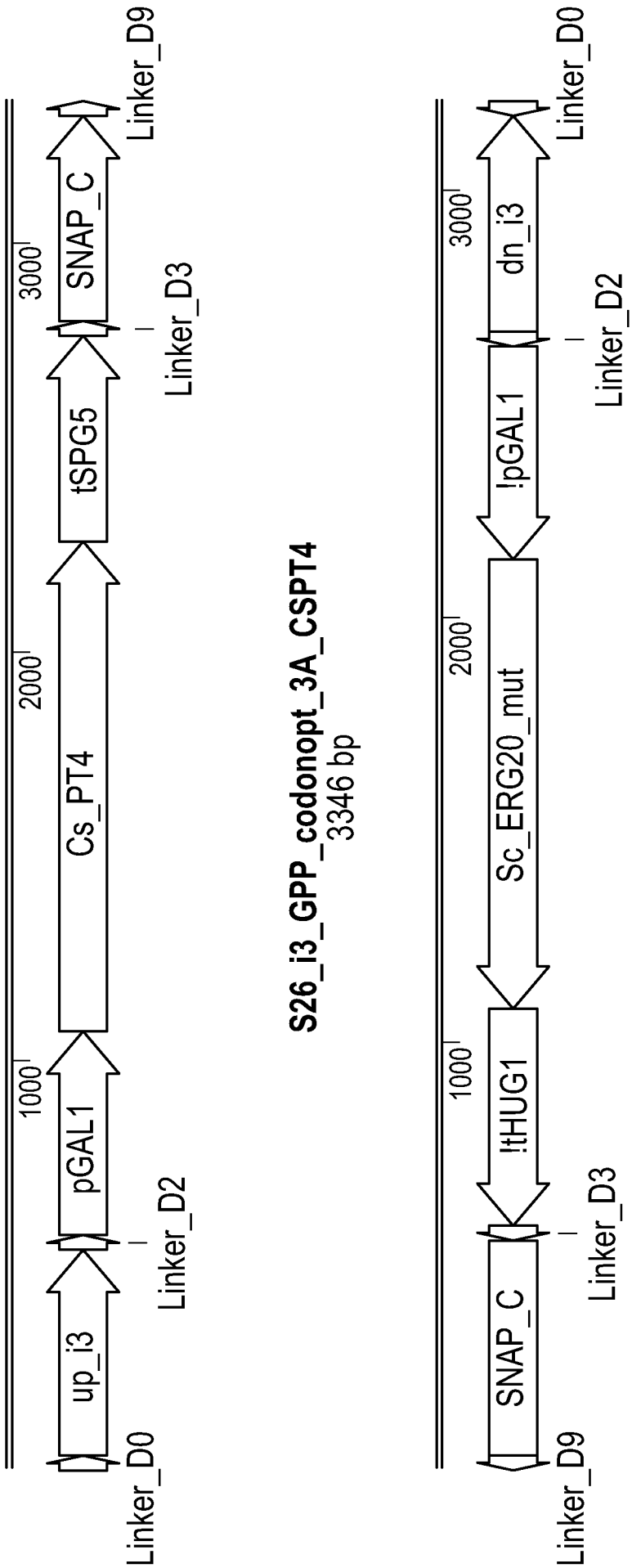

Linker_D0
up_i3
Linker_D2
pGAL1
Cs_PT4
tSPG5
Linker_D3
SNAP_C
Linker_D9
1000
2000
3000
S26_i3_GPP_codonopt_3A_CSPT4
3346 bp
Linker_D9
SNAP_C
Linker_D3
!tHUG1
Sc_ERG20_mut
!pGAL1
Linker_D2
dn_i3
Linker_D0
1000
2000
3000
S9_i3_GPP_codonopt_3B_0
3212 bp

FIG. 86

SEQUENCE LISTING

<110> The Regents of the University of California
Keasling, Jay D.
d'Espaux, Leo
Wong, Jeff
Luo, Xiaozhou
Reiter, Michael
Denby, Charles
Lechner, Anna

<120> MICROORGANISMS AND METHODS FOR PRODUCING CANNABINOIDS AND CANNABINOID DERIVATIVES

<130> DEMT-001/00WO 326934-2003

<150> US 62/569,532
<151> 2017-10-07

<150> US 62/491,114
<151> 2017-04-27

<160> 225

<170> PatentIn version 3.5

<210> 1
<211> 776
<212> PRT
<213> Cannabis sativa

<400> 1

Met Ala Leu Glu Leu Pro His Leu Leu Pro Tyr Lys Leu Val Lys Gly
1 5 10 15

Gln Thr Leu Val Ala Gln Ala Ala Arg Ala Glu Leu Ala Ser Ser Ser
20 25 30

Ser Ser Ser Val Ile Leu Lys Ser Asn Phe Ile Asn Asn Asn Tyr Ile
35 40 45

Asn Tyr Cys Asn Asn Asn Asn Asn Asn Glu Arg Arg Leu Val Val Arg
50 55 60

Arg Asp Trp Glu Thr Met Ala Ser Ser Pro Ser His Ser Arg Asn Asn
65 70 75 80

```
Asn Asp Ile Arg Thr Ile Asn His Leu Arg His Val Asp Ser Met Ala
                85                  90                  95

Thr Met Pro Ser Gly Ala Gly Lys Ile Pro Arg Leu Asn Ala Val Ile
            100                 105                 110

Leu Gly Glu Ala Leu Ala Thr Glu Glu Asn Asp Leu Val Phe Pro Thr
        115                 120                 125

Asp Glu Phe Ser Gln Gln Ala His Val Pro Ser Pro Gln Lys Tyr Leu
    130                 135                 140

Glu Met Tyr Lys Arg Ser Ile Glu Asp Pro Ala Gly Phe Trp Ser Glu
145                 150                 155                 160

Ile Ala Ser Gln Phe Tyr Trp Lys Gln Lys Trp Asp Asp Ser Val Tyr
                165                 170                 175

Ser Glu Asn Leu Asp Val Ser Lys Gly Arg Val Asn Ile Glu Trp Phe
            180                 185                 190

Lys Gly Gly Ile Thr Asn Ile Cys Tyr Asn Cys Leu Asp Lys Asn Val
        195                 200                 205

Glu Ala Gly Leu Gly Asp Lys Ile Ala Leu Tyr Trp Glu Gly Asn Asp
    210                 215                 220

Thr Gly Phe Asp Asp Ser Leu Thr Tyr Ser Gln Leu Leu His Lys Val
225                 230                 235                 240

Cys Gln Leu Ala Asn Tyr Leu Lys Asp Met Gly Val Gln Lys Gly Asp
                245                 250                 255

Ala Val Val Ile Tyr Leu Pro Met Leu Leu Glu Leu Pro Ile Thr Met
            260                 265                 270

Leu Ala Cys Ala Arg Ile Gly Ala Val His Ser Val Val Phe Ala Gly
        275                 280                 285
```

```
Phe Ser Ala Glu Ser Leu Ser Gln Arg Ile Ile Asp Cys Lys Pro Lys
    290                 295                 300

Val Val Ile Thr Cys Asn Ala Val Lys Arg Gly Pro Lys Ile Ile His
305                 310                 315                 320

Leu Lys Asp Ile Val Asp Ala Ala Leu Val Glu Ser Ala Lys Thr Gly
                325                 330                 335

Val Pro Ile Asp Thr Cys Leu Val Tyr Glu Asn Gln Leu Ala Met Lys
            340                 345                 350

Arg Asp Ile Thr Lys Trp Gln Asp Gly Arg Asp Ile Trp Trp Gln Asp
        355                 360                 365

Val Ile Pro Lys Tyr Pro Thr Glu Cys Ala Val Glu Trp Val Asp Ala
    370                 375                 380

Glu Asp Pro Leu Phe Leu Leu Tyr Thr Ser Gly Ser Thr Gly Lys Pro
385                 390                 395                 400

Lys Gly Val Leu His Thr Thr Gly Gly Tyr Met Val Tyr Thr Ala Thr
                405                 410                 415

Thr Phe Lys Tyr Ala Phe Asp Tyr Lys Pro Ser Asp Val Tyr Trp Cys
            420                 425                 430

Thr Ala Asp Cys Gly Trp Ile Thr Gly His Ser Tyr Val Thr Tyr Gly
        435                 440                 445

Pro Leu Leu Asn Gly Ala Ser Cys Ile Val Phe Glu Gly Ala Pro Asn
    450                 455                 460

Tyr Pro Asp Ser Gly Arg Cys Trp Asp Ile Val Asp Lys Tyr Lys Val
465                 470                 475                 480

Thr Ile Phe Tyr Thr Ala Pro Thr Leu Val Arg Ser Leu Met Arg Asp
                485                 490                 495
```

```
Gly Asp Glu Tyr Val Thr Arg Tyr Ser Arg Lys Ser Leu Arg Ile Leu
            500                 505                 510

Gly Ser Val Gly Glu Pro Ile Asn Pro Ser Ala Trp Arg Trp Phe Tyr
        515                 520                 525

Asn Val Val Gly Asp Ser Arg Cys Pro Ile Ser Asp Thr Trp Trp Gln
    530                 535                 540

Thr Glu Thr Gly Gly Phe Met Ile Thr Pro Leu Pro Gly Ala Trp Pro
545                 550                 555                 560

Gln Lys Pro Gly Ser Ala Thr Phe Pro Phe Phe Gly Val Lys Pro Val
                565                 570                 575

Ile Val Asp Glu Lys Gly Val Glu Ile Glu Gly Glu Cys Ser Gly Tyr
            580                 585                 590

Leu Cys Val Lys Gly Ser Trp Pro Gly Ala Phe Arg Thr Leu Tyr Gly
        595                 600                 605

Asp Tyr Glu Arg Tyr Glu Thr Thr Tyr Phe Lys Pro Phe Thr Gly Tyr
    610                 615                 620

Tyr Phe Thr Gly Asp Gly Cys Ser Arg Asp Lys Asp Gly Tyr His Trp
625                 630                 635                 640

Leu Thr Gly Arg Val Asp Asp Val Ile Asn Val Ser Gly His Arg Ile
                645                 650                 655

Gly Thr Ala Glu Val Glu Ser Ala Leu Val Ser His Pro Lys Cys Ala
            660                 665                 670

Glu Ala Ala Val Val Gly Ile Glu His Glu Val Lys Gly Gln Ala Ile
        675                 680                 685

Tyr Ala Phe Val Thr Leu Val Glu Gly Glu Pro Tyr Ser Glu Glu Leu
    690                 695                 700
```

```
Arg Lys Ser Leu Ile Leu Ser Val Arg Lys Gln Ile Gly Ala Phe Ala
705                 710                 715                 720


Ala Pro Glu Arg Ile His Trp Ala Pro Gly Leu Pro Lys Thr Arg Ser
                725                 730                 735


Gly Lys Ile Met Arg Arg Ile Leu Arg Lys Ile Ala Ser Gly Gln Leu
            740                 745                 750


Asp Glu Leu Gly Asp Thr Ser Thr Leu Ala Asp Pro Asn Val Val Glu
        755                 760                 765


Gln Leu Ile Ser Leu Ser Asn Cys
    770                 775


<210>  2
<211>  579
<212>  PRT
<213>  Streptomyces sp.

<400>  2

Met Glu Leu Ala Leu Pro Ala Glu Leu Ala Pro Thr Leu Pro Glu Ala
1               5                   10                  15


Leu Arg Leu Arg Ser Glu Gln Gln Pro Asp Thr Val Ala Tyr Val Phe
            20                  25                  30


Leu Arg Asp Gly Glu Thr Pro Glu Glu Thr Leu Thr Tyr Gly Arg Leu
        35                  40                  45


Asp Arg Ala Ala Arg Ala Arg Ala Ala Ala Leu Glu Ala Ala Gly Leu
    50                  55                  60


Ala Gly Gly Thr Ala Val Leu Leu Tyr Pro Ser Gly Leu Glu Phe Val
65                  70                  75                  80


Ala Ala Leu Leu Gly Cys Met Tyr Ala Gly Thr Ala Gly Ala Pro Val
                85                  90                  95
```

```
Gln Val Pro Thr Arg Arg Arg Gly Met Glu Arg Ala Arg Arg Ile Ala
            100                 105                 110

Asp Asp Ala Gly Ala Lys Thr Ile Leu Thr Thr Thr Ala Val Lys Arg
        115                 120                 125

Glu Val Glu Glu His Phe Ala Asp Leu Leu Thr Gly Leu Thr Val Ile
    130                 135                 140

Asp Thr Glu Ser Leu Pro Asp Val Pro Asp Asp Ala Pro Ala Val Arg
145                 150                 155                 160

Leu Pro Gly Pro Asp Asp Val Ala Leu Leu Gln Tyr Thr Ser Gly Ser
                165                 170                 175

Thr Gly Asp Pro Lys Gly Val Glu Val Thr His Ala Asn Phe Arg Ala
            180                 185                 190

Asn Val Ala Glu Thr Val Glu Leu Trp Pro Val Arg Ser Asp Gly Thr
        195                 200                 205

Val Val Asn Trp Leu Pro Leu Phe His Asp Met Gly Leu Met Phe Gly
    210                 215                 220

Val Val Met Pro Leu Phe Thr Gly Val Pro Ala Tyr Leu Met Ala Pro
225                 230                 235                 240

Gln Ser Phe Ile Arg Arg Pro Ala Arg Trp Leu Glu Ala Ile Ser Arg
                245                 250                 255

Phe Arg Gly Thr His Ala Ala Ala Pro Ser Phe Ala Tyr Glu Leu Cys
            260                 265                 270

Val Arg Ser Val Ala Asp Thr Gly Leu Pro Ala Gly Leu Asp Leu Ser
        275                 280                 285

Ser Trp Arg Val Ala Val Asn Gly Ala Glu Pro Val Arg Trp Thr Ala
    290                 295                 300
```

```
Val Ala Asp Phe Thr Glu Ala Tyr Ala Pro Ala Gly Phe Arg Pro Gln
305                 310                 315                 320


Ala Met Cys Pro Gly Tyr Gly Leu Ala Glu Asn Thr Leu Lys Leu Ser
                325                 330                 335


Gly Ser Pro Glu Asp Arg Pro Pro Thr Leu Leu Arg Ala Asp Ala Ala
            340                 345                 350


Ala Leu Gln Asp Gly Arg Val Val Pro Leu Thr Gly Pro Gly Thr Asp
        355                 360                 365


Gly Val Arg Leu Val Gly Ser Gly Val Thr Val Pro Ser Ser Arg Val
    370                 375                 380


Ala Val Val Asp Pro Gly Thr Gly Thr Glu Gln Pro Ala Gly Arg Val
385                 390                 395                 400


Gly Glu Ile Trp Ile Asn Gly Pro Cys Val Ala Arg Gly Tyr His Gly
                405                 410                 415


Arg Pro Ala Glu Ser Ala Glu Ser Phe Gly Ala Arg Ile Ala Gly Gln
            420                 425                 430


Glu Ala Arg Gly Thr Trp Leu Arg Thr Gly Asp Leu Gly Phe Leu His
        435                 440                 445


Asp Gly Glu Val Phe Val Ala Gly Arg Leu Lys Asp Val Val Ile His
    450                 455                 460


Gln Gly Arg Asn Phe Tyr Pro Gln Asp Ile Glu Leu Ser Ala Glu Val
465                 470                 475                 480


Ser Asp Arg Ala Leu His Pro Asn Cys Ala Ala Ala Phe Ala Leu Asp
                485                 490                 495


Asp Gly Arg Thr Glu Arg Leu Val Leu Leu Val Glu Ala Asp Gly Arg
            500                 505                 510
```

```
Ala Leu Arg Asn Gly Gly Ala Asp Ala Leu Arg Ala Arg Val His Asp
        515                 520                 525


Ala Val Trp Asp Arg Gln Arg Leu Arg Ile Asp Glu Ile Val Leu Leu
    530                 535                 540


Arg Arg Gly Ala Leu Pro Lys Thr Ser Ser Gly Lys Val Gln Arg Arg
545                 550                 555                 560


Leu Ala Arg Ser Arg Tyr Leu Asp Gly Glu Phe Gly Pro Ala Pro Ala
                565                 570                 575


Arg Glu Ala
```

<210> 3
<211> 1671
<212> PRT
<213> Aspergillus sp.

<400> 3

```
Met Val Ile Gln Gly Lys Arg Leu Ala Ala Ser Ser Ile Gln Leu Leu
1               5                   10                  15


Ala Ser Ser Leu Asp Ala Lys Lys Leu Cys Tyr Glu Tyr Asp Glu Arg
            20                  25                  30


Gln Ala Pro Gly Val Thr Gln Ile Thr Glu Glu Ala Pro Thr Glu Gln
        35                  40                  45


Pro Pro Leu Ser Thr Pro Pro Ser Leu Pro Gln Thr Pro Asn Ile Ser
    50                  55                  60


Pro Ile Ser Ala Ser Lys Ile Val Ile Asp Asp Val Ala Leu Ser Arg
65                  70                  75                  80


Val Gln Ile Val Gln Ala Leu Val Ala Arg Lys Leu Lys Thr Ala Ile
                85                  90                  95


Ala Gln Leu Pro Thr Ser Lys Ser Ile Lys Glu Leu Ser Gly Gly Arg
```

100 105 110

Ser Ser Leu Gln Asn Glu Leu Val Gly Asp Ile His Asn Glu Phe Ser
115 120 125

Ser Ile Pro Asp Ala Pro Glu Gln Ile Leu Leu Arg Asp Phe Gly Asp
130 135 140

Ala Asn Pro Thr Val Gln Leu Gly Lys Thr Ser Ser Ala Ala Val Ala
145 150 155 160

Lys Leu Ile Ser Ser Lys Met Pro Ser Asp Phe Asn Ala Asn Ala Ile
165 170 175

Arg Ala His Leu Ala Asn Lys Trp Gly Leu Gly Pro Leu Arg Gln Thr
180 185 190

Ala Val Leu Leu Tyr Ala Ile Ala Ser Glu Pro Pro Ser Arg Leu Ala
195 200 205

Ser Ser Ser Ala Ala Glu Glu Tyr Trp Asp Asn Val Ser Ser Met Tyr
210 215 220

Ala Glu Ser Cys Gly Ile Thr Leu Arg Pro Arg Gln Asp Thr Met Asn
225 230 235 240

Glu Asp Ala Met Ala Ser Ser Ala Ile Asp Pro Ala Val Val Ala Glu
245 250 255

Phe Ser Lys Gly His Arg Arg Leu Gly Val Gln Gln Phe Gln Ala Leu
260 265 270

Ala Glu Tyr Leu Gln Ile Asp Leu Ser Gly Ser Gln Ala Ser Gln Ser
275 280 285

Asp Ala Leu Val Ala Glu Leu Gln Gln Lys Val Asp Leu Trp Thr Ala
290 295 300

Glu Met Thr Pro Glu Phe Leu Ala Gly Ile Ser Pro Met Leu Asp Val

305 310 315 320

Lys Lys Ser Arg Arg Tyr Gly Ser Trp Trp Asn Met Ala Arg Gln Asp
325 330 335

Val Leu Ala Phe Tyr Arg Arg Pro Ser Tyr Ser Glu Phe Val Asp Asp
340 345 350

Ala Leu Ala Phe Lys Val Phe Leu Asn Arg Leu Cys Asn Arg Ala Asp
355 360 365

Glu Ala Leu Leu Asn Met Val Arg Ser Leu Ser Cys Asp Ala Tyr Phe
370 375 380

Lys Gln Gly Ser Leu Pro Gly Tyr His Ala Ala Ser Arg Leu Leu Glu
385 390 395 400

Gln Ala Ile Thr Ser Thr Val Ala Asp Cys Pro Lys Ala Arg Leu Ile
405 410 415

Leu Pro Ala Val Gly Pro His Thr Thr Ile Thr Lys Asp Gly Thr Ile
420 425 430

Glu Tyr Ala Glu Ala Pro Arg Gln Gly Val Ser Gly Pro Thr Ala Tyr
435 440 445

Ile Gln Ser Leu Arg Gln Gly Ala Ser Phe Ile Gly Leu Lys Ser Ala
450 455 460

Asp Val Asp Thr Gln Ser Asn Leu Thr Asp Ala Leu Leu Asp Ala Met
465 470 475 480

Cys Leu Ala Leu His Asn Gly Ile Ser Phe Val Gly Lys Thr Phe Leu
485 490 495

Val Thr Gly Ala Gly Gln Gly Ser Ile Gly Ala Gly Val Val Arg Leu
500 505 510

Leu Leu Glu Gly Gly Ala Arg Val Leu Val Thr Thr Ser Arg Glu Pro

515 520 525

Ala Thr Thr Ser Arg Tyr Phe Gln Gln Met Tyr Asp Asn His Gly Ala
530 535 540

Lys Phe Ser Glu Leu Arg Val Val Pro Cys Asn Leu Ala Ser Ala Gln
545 550 555 560

Asp Cys Glu Gly Leu Ile Arg His Val Tyr Asp Pro Arg Gly Leu Asn
565 570 575

Trp Asp Leu Asp Ala Ile Leu Pro Phe Ala Ala Ala Ser Asp Tyr Ser
580 585 590

Thr Glu Met His Asp Ile Arg Gly Gln Ser Glu Leu Gly His Arg Leu
595 600 605

Met Leu Val Asn Val Phe Arg Val Leu Gly His Ile Val His Cys Lys
610 615 620

Arg Asp Ala Gly Val Asp Cys His Pro Thr Gln Val Leu Leu Pro Leu
625 630 635 640

Ser Pro Asn His Gly Ile Phe Gly Gly Asp Gly Met Tyr Pro Glu Ser
645 650 655

Lys Leu Ala Leu Glu Ser Leu Phe His Arg Ile Arg Ser Glu Ser Trp
660 665 670

Ser Asp Gln Leu Ser Ile Cys Gly Val Arg Ile Gly Trp Thr Arg Ser
675 680 685

Thr Gly Leu Met Thr Ala His Asp Ile Ile Ala Glu Thr Val Glu Glu
690 695 700

His Gly Ile Arg Thr Phe Ser Val Ala Glu Met Ala Leu Asn Ile Ala
705 710 715 720

Met Leu Leu Thr Pro Asp Phe Val Ala His Cys Glu Asp Gly Pro Leu

725 730 735

Asp Ala Asp Phe Thr Gly Ser Leu Gly Thr Leu Gly Ser Ile Pro Gly
740 745 750

Phe Leu Ala Gln Leu His Gln Lys Val Gln Leu Ala Ala Glu Val Ile
755 760 765

Arg Ala Val Gln Ala Glu Asp Glu His Glu Arg Phe Leu Ser Pro Gly
770 775 780

Thr Lys Pro Thr Leu Gln Ala Pro Val Ala Pro Met His Pro Arg Ser
785 790 795 800

Ser Leu Arg Val Gly Tyr Pro Arg Leu Pro Asp Tyr Glu Gln Glu Ile
805 810 815

Arg Pro Leu Ser Pro Arg Leu Glu Arg Leu Gln Asp Pro Ala Asn Ala
820 825 830

Val Val Val Val Gly Tyr Ser Glu Leu Gly Pro Trp Gly Ser Ala Arg
835 840 845

Leu Arg Trp Glu Ile Glu Ser Gln Gly Gln Trp Thr Ser Ala Gly Tyr
850 855 860

Val Glu Leu Ala Trp Leu Met Asn Leu Ile Arg His Val Asn Asp Glu
865 870 875 880

Ser Tyr Val Gly Trp Val Asp Thr Gln Thr Gly Lys Pro Val Arg Asp
885 890 895

Gly Glu Ile Gln Ala Leu Tyr Gly Asp His Ile Asp Asn His Thr Gly
900 905 910

Ile Arg Pro Ile Gln Ser Thr Ser Tyr Asn Pro Glu Arg Met Glu Val
915 920 925

Leu Gln Glu Val Ala Val Glu Glu Asp Leu Pro Glu Phe Glu Val Ser

```
    930                 935                 940

Gln Leu Thr Ala Asp Ala Met Arg Leu Arg His Gly Ala Asn Val Ser
945                 950                 955                 960

Ile Arg Pro Ser Gly Asn Pro Asp Ala Cys His Val Lys Leu Lys Arg
                965                 970                 975

Gly Ala Val Ile Leu Val Pro Lys Thr Val Pro Phe Val Trp Gly Ser
            980                 985                 990

Cys Ala Gly Glu Leu Pro Lys Gly  Trp Thr Pro Ala Lys  Tyr Gly Ile
        995                 1000                1005

Pro Glu  Asn Leu Ile His Gln  Val Asp Pro Val Thr  Leu Tyr Thr
    1010                1015                1020

Ile Cys  Cys Val Ala Glu Ala  Phe Tyr Ser Ala Gly  Ile Thr His
    1025                1030                1035

Pro Leu  Glu Val Phe Arg His  Ile His Leu Ser Glu  Leu Gly Asn
    1040                1045                1050

Phe Ile  Gly Ser Ser Met Gly  Gly Pro Thr Lys Thr  Arg Gln Leu
    1055                1060                1065

Tyr Arg  Asp Val Tyr Phe Asp  His Glu Ile Pro Ser  Asp Val Leu
    1070                1075                1080

Gln Asp  Thr Tyr Leu Asn Thr  Pro Ala Ala Trp Val  Asn Met Leu
    1085                1090                1095

Leu Leu  Gly Cys Thr Gly Pro  Ile Lys Thr Pro Val  Gly Ala Cys
    1100                1105                1110

Ala Thr  Gly Val Glu Ser Ile  Asp Ser Gly Tyr Glu  Ser Ile Met
    1115                1120                1125

Ala Gly  Lys Thr Lys Met Cys  Leu Val Gly Gly Tyr  Asp Asp Leu
```

```
    1130                1135                1140

Gln Glu  Glu Ala Ser Tyr Gly  Phe Ala Gln Leu Lys  Ala Thr Val
    1145                1150                1155

Asn Val  Glu Glu Glu Ile Ala  Cys Gly Arg Gln Pro  Ser Glu Met
    1160                1165                1170

Ser Arg  Pro Met Ala Glu Ser  Arg Ala Gly Phe Val  Glu Ala His
    1175                1180                1185

Gly Cys  Gly Val Gln Leu Leu  Cys Arg Gly Asp Ile  Ala Leu Gln
    1190                1195                1200

Met Gly  Leu Pro Ile Tyr Ala  Val Ile Ala Ser Ser  Ala Met Ala
    1205                1210                1215

Ala Asp  Lys Ile Gly Ser Ser  Val Pro Ala Pro Gly  Gln Gly Ile
    1220                1225                1230

Leu Ser  Phe Ser Arg Glu Arg  Ala Arg Ser Ser Met  Ile Ser Val
    1235                1240                1245

Thr Ser  Arg Pro Ser Ser Arg  Ser Ser Thr Ser Ser  Glu Val Ser
    1250                1255                1260

Asp Lys  Ser Ser Leu Thr Ser  Ile Thr Ser Ile Ser  Asn Pro Ala
    1265                1270                1275

Pro Arg  Ala Gln Arg Ala Arg  Ser Thr Thr Asp Met  Ala Pro Leu
    1280                1285                1290

Arg Ala  Ala Leu Ala Thr Trp  Gly Leu Thr Ile Asp  Asp Leu Asp
    1295                1300                1305

Val Ala  Ser Leu His Gly Thr  Ser Thr Arg Gly Asn  Asp Leu Asn
    1310                1315                1320

Glu Pro  Glu Val Ile Glu Thr  Gln Met Arg His Leu  Gly Arg Thr
```

1325 1330 1335

Pro Gly Arg Pro Leu Trp Ala Ile Cys Gln Lys Ser Val Thr Gly
1340 1345 1350

His Pro Lys Ala Pro Ala Ala Ala Trp Met Leu Asn Gly Cys Leu
1355 1360 1365

Gln Val Leu Asp Ser Gly Leu Val Pro Gly Asn Arg Asn Leu Asp
1370 1375 1380

Thr Leu Asp Glu Ala Leu Arg Ser Ala Ser His Leu Cys Phe Pro
1385 1390 1395

Thr Arg Thr Val Gln Leu Arg Glu Val Lys Ala Phe Leu Leu Thr
1400 1405 1410

Ser Phe Gly Phe Gly Gln Lys Gly Gly Gln Val Val Gly Val Ala
1415 1420 1425

Pro Lys Tyr Phe Phe Ala Thr Leu Pro Arg Pro Glu Val Glu Gly
1430 1435 1440

Tyr Tyr Arg Lys Val Arg Val Arg Thr Glu Ala Gly Asp Arg Ala
1445 1450 1455

Tyr Ala Ala Ala Val Met Ser Gln Ala Val Val Lys Ile Gln Thr
1460 1465 1470

Gln Asn Pro Tyr Asp Glu Pro Asp Ala Pro Arg Ile Phe Leu Asp
1475 1480 1485

Pro Leu Ala Arg Ile Ser Gln Asp Pro Ser Thr Gly Gln Tyr Arg
1490 1495 1500

Phe Arg Ser Asp Ala Thr Pro Ala Leu Asp Asp Asp Ala Leu Pro
1505 1510 1515

Pro Pro Gly Glu Pro Thr Glu Leu Val Lys Gly Ile Ser Ser Ala

1520 1525 1530

Trp Ile Glu Glu Lys Val Arg Pro His Met Ser Pro Gly Gly Thr
1535 1540 1545

Val Gly Val Asp Leu Val Pro Leu Ala Ser Phe Asp Ala Tyr Lys
1550 1555 1560

Asn Ala Ile Phe Val Glu Arg Asn Tyr Thr Val Arg Glu Arg Asp
1565 1570 1575

Trp Ala Glu Lys Ser Ala Asp Val Arg Ala Ala Tyr Ala Ser Arg
1580 1585 1590

Trp Cys Ala Lys Glu Ala Val Phe Lys Cys Leu Gln Thr His Ser
1595 1600 1605

Gln Gly Ala Gly Ala Ala Met Lys Glu Ile Glu Ile Glu His Gly
1610 1615 1620

Gly Asn Gly Ala Pro Lys Val Lys Leu Arg Gly Ala Ala Gln Thr
1625 1630 1635

Ala Ala Arg Gln Arg Gly Leu Glu Gly Val Gln Leu Ser Ile Ser
1640 1645 1650

Tyr Gly Asp Asp Ala Val Ile Ala Val Ala Leu Gly Leu Met Ser
1655 1660 1665

Gly Ala Ser
1670

<210> 4
<211> 1888
<212> PRT
<213> Aspergillus sp.

<400> 4

Met Gly Ser Val Ser Arg Glu His Glu Ser Ile Pro Ile Gln Ala Ala
1 5 10 15

```
Gln Arg Gly Ala Ala Arg Ile Cys Ala Ala Phe Gly Gly Gln Gly Ser
            20                  25                  30

Asn Asn Leu Asp Val Leu Lys Gly Leu Leu Glu Leu Tyr Lys Arg Tyr
        35                  40                  45

Gly Pro Asp Leu Asp Glu Leu Leu Asp Val Ala Ser Asn Thr Leu Ser
    50                  55                  60

Gln Leu Ala Ser Ser Pro Ala Ala Ile Asp Val His Glu Pro Trp Gly
65                  70                  75                  80

Phe Asp Leu Arg Gln Trp Leu Thr Thr Pro Glu Val Ala Pro Ser Lys
                85                  90                  95

Glu Ile Leu Ala Leu Pro Pro Arg Ser Phe Pro Leu Asn Thr Leu Leu
            100                 105                 110

Ser Leu Ala Leu Tyr Cys Ala Thr Cys Arg Glu Leu Glu Leu Asp Pro
        115                 120                 125

Gly Gln Phe Arg Ser Leu Leu His Ser Ser Thr Gly His Ser Gln Gly
    130                 135                 140

Ile Leu Ala Ala Val Ala Ile Thr Gln Ala Glu Ser Trp Pro Thr Phe
145                 150                 155                 160

Tyr Asp Ala Cys Arg Thr Val Leu Gln Ile Ser Phe Trp Ile Gly Leu
                165                 170                 175

Glu Ala Tyr Leu Phe Thr Pro Ser Ser Ala Ala Ser Asp Ala Met Ile
            180                 185                 190

Gln Asp Cys Ile Glu His Gly Glu Gly Leu Leu Ser Ser Met Leu Ser
        195                 200                 205

Val Ser Gly Leu Ser Arg Ser Gln Val Glu Arg Val Ile Glu His Val
    210                 215                 220
```

```
Asn Lys Gly Leu Gly Glu Cys Asn Arg Trp Val His Leu Ala Leu Val
225                 230                 235                 240

Asn Ser His Glu Lys Phe Val Leu Ala Gly Pro Pro Gln Ser Leu Trp
                245                 250                 255

Ala Val Cys Leu His Val Arg Arg Ile Arg Ala Asp Asn Asp Leu Asp
            260                 265                 270

Gln Ser Arg Ile Leu Phe Arg Asn Arg Lys Pro Ile Val Asp Ile Leu
        275                 280                 285

Phe Leu Pro Ile Ser Ala Pro Phe His Thr Pro Tyr Leu Asp Gly Val
    290                 295                 300

Gln Asp Arg Val Ile Glu Ala Leu Ser Ser Ala Ser Leu Ala Leu His
305                 310                 315                 320

Ser Ile Lys Ile Pro Leu Tyr His Thr Gly Thr Gly Ser Asn Leu Gln
                325                 330                 335

Glu Leu Gln Pro His Gln Leu Ile Pro Thr Leu Ile Arg Ala Ile Thr
            340                 345                 350

Val Asp Gln Leu Asp Trp Pro Leu Val Cys Arg Gly Leu Asn Ala Thr
        355                 360                 365

His Val Leu Asp Phe Gly Pro Gly Gln Thr Cys Ser Leu Ile Gln Glu
    370                 375                 380

Leu Thr Gln Gly Thr Gly Val Ser Val Ile Gln Leu Thr Thr Gln Ser
385                 390                 395                 400

Gly Pro Lys Pro Val Gly Gly His Leu Ala Ala Val Asn Trp Glu Ala
                405                 410                 415

Glu Phe Gly Leu Arg Leu His Ala Asn Val His Gly Ala Ala Lys Leu
            420                 425                 430
```

```
His Asn Arg Met Thr Thr Leu Leu Gly Lys Pro Pro Val Met Val Ala
        435                 440                 445

Gly Met Thr Pro Thr Thr Val Arg Trp Asp Phe Val Ala Ala Val Ala
    450                 455                 460

Gln Ala Gly Tyr His Val Glu Leu Ala Gly Gly Gly Tyr His Ala Glu
465                 470                 475                 480

Arg Gln Phe Glu Ala Glu Ile Arg Arg Leu Ala Thr Ala Ile Pro Ala
                485                 490                 495

Asp His Gly Ile Thr Cys Asn Leu Leu Tyr Ala Lys Pro Thr Thr Phe
            500                 505                 510

Ser Trp Gln Ile Ser Val Ile Lys Asp Leu Val Arg Gln Gly Val Pro
        515                 520                 525

Val Glu Gly Ile Thr Ile Gly Ala Gly Ile Pro Ser Pro Glu Val Val
    530                 535                 540

Gln Glu Cys Val Gln Ser Ile Gly Leu Lys His Ile Ser Phe Lys Pro
545                 550                 555                 560

Gly Ser Phe Glu Ala Ile His Gln Val Ile Gln Ile Ala Arg Thr His
                565                 570                 575

Pro Asn Phe Leu Ile Gly Leu Gln Trp Thr Ala Gly Arg Gly Gly Gly
            580                 585                 590

His His Ser Trp Glu Asp Phe His Gly Pro Ile Leu Ala Thr Tyr Ala
        595                 600                 605

Gln Ile Arg Ser Cys Pro Asn Ile Leu Leu Val Val Gly Ser Gly Phe
    610                 615                 620

Gly Gly Gly Pro Asp Thr Phe Pro Tyr Leu Thr Gly Gln Trp Ala Gln
625                 630                 635                 640
```

```
Ala Phe Gly Tyr Pro Cys Met Pro Phe Asp Gly Val Leu Leu Gly Ser
                645                 650                 655

Arg Met Met Val Ala Arg Glu Ala His Thr Ser Ala Gln Ala Lys Arg
            660                 665                 670

Leu Ile Ile Asp Ala Gln Gly Val Gly Asp Ala Asp Trp His Lys Ser
        675                 680                 685

Phe Asp Glu Pro Thr Gly Gly Val Val Thr Val Asn Ser Glu Phe Gly
    690                 695                 700

Gln Pro Ile His Val Leu Ala Thr Arg Gly Val Met Leu Trp Lys Glu
705                 710                 715                 720

Leu Asp Asn Arg Val Phe Ser Ile Lys Asp Thr Ser Lys Arg Leu Glu
                725                 730                 735

Tyr Leu Arg Asn His Arg Gln Glu Ile Val Ser Arg Leu Asn Ala Asp
            740                 745                 750

Phe Ala Arg Pro Trp Phe Ala Val Asp Gly His Gly Gln Asn Val Glu
        755                 760                 765

Leu Glu Asp Met Thr Tyr Leu Glu Val Leu Arg Arg Leu Cys Asp Leu
    770                 775                 780

Thr Tyr Val Ser His Gln Lys Arg Trp Val Asp Pro Ser Tyr Arg Ile
785                 790                 795                 800

Leu Leu Leu Asp Phe Val His Leu Leu Arg Glu Arg Phe Gln Cys Ala
                805                 810                 815

Ile Asp Asn Pro Gly Glu Tyr Pro Leu Asp Ile Ile Val Arg Val Glu
            820                 825                 830

Glu Ser Leu Lys Asp Lys Ala Tyr Arg Thr Leu Tyr Pro Glu Asp Val
        835                 840                 845
```

```
Ser Leu Leu Met His Leu Phe Ser Arg Arg Asp Ile Lys Pro Val Pro
    850                 855                 860

Phe Ile Pro Arg Leu Asp Glu Arg Phe Glu Thr Trp Phe Lys Lys Asp
865                 870                 875                 880

Ser Leu Trp Gln Ser Glu Asp Val Glu Ala Val Ile Gly Gln Asp Val
                885                 890                 895

Gln Arg Ile Phe Ile Ile Gln Gly Pro Met Ala Val Gln Tyr Ser Ile
            900                 905                 910

Ser Asp Asp Glu Ser Val Lys Asp Ile Leu His Asn Ile Cys Asn His
        915                 920                 925

Tyr Val Glu Ala Leu Gln Ala Asp Ser Arg Glu Thr Ser Ile Gly Asp
    930                 935                 940

Val His Ser Ile Thr Gln Lys Pro Leu Ser Ala Phe Pro Gly Leu Lys
945                 950                 955                 960

Val Thr Thr Asn Arg Val Gln Gly Leu Tyr Lys Phe Glu Lys Val Gly
                965                 970                 975

Ala Val Pro Glu Met Asp Val Leu Phe Glu His Ile Val Gly Leu Ser
            980                 985                 990

Lys Ser Trp Ala Arg Thr Cys Leu  Met Ser Lys Ser Val  Phe Arg Asp
        995                 1000                1005

Gly Ser  Arg Leu His Asn Pro  Ile Arg Ala Ala Leu  Gln Leu Gln
    1010                1015                1020

Arg Gly  Asp Thr Ile Glu Val  Leu Leu Thr Ala Asp  Ser Glu Ile
    1025                1030                1035

Arg Lys  Ile Arg Leu Ile Ser  Pro Thr Gly Asp Gly  Gly Ser Thr
    1040                1045                1050
```

```
Ser Lys  Val Val Leu Glu Ile  Val Ser Asn Asp Gly  Gln Arg Val
    1055                1060                1065

Phe Ala  Thr Leu Ala Pro Asn  Ile Pro Leu Ser Pro  Glu Pro Ser
    1070                1075                1080

Val Val  Phe Cys Phe Lys Val  Asp Gln Lys Pro Asn  Glu Trp Thr
    1085                1090                1095

Leu Glu  Glu Asp Ala Ser Gly  Arg Ala Glu Arg Ile  Lys Ala Leu
    1100                1105                1110

Tyr Met  Ser Leu Trp Asn Leu  Gly Phe Pro Asn Lys  Ala Ser Val
    1115                1120                1125

Leu Gly  Leu Asn Ser Gln Phe  Thr Gly Glu Glu Leu  Met Ile Thr
    1130                1135                1140

Thr Asp  Lys Ile Arg Asp Phe  Glu Arg Val Leu Arg  Gln Thr Ser
    1145                1150                1155

Pro Leu  Gln Leu Gln Ser Trp  Asn Pro Gln Gly Cys  Val Pro Ile
    1160                1165                1170

Asp Tyr  Cys Val Val Ile Ala  Trp Ser Ala Leu Thr  Lys Pro Leu
    1175                1180                1185

Met Val  Ser Ser Leu Lys Cys  Asp Leu Leu Asp Leu  Leu His Ser
    1190                1195                1200

Ala Ile  Ser Phe His Tyr Ala  Pro Ser Val Lys Pro  Leu Arg Val
    1205                1210                1215

Gly Asp  Ile Val Lys Thr Ser  Ser Arg Ile Leu Ala  Val Ser Val
    1220                1225                1230

Arg Pro  Arg Gly Thr Met Leu  Thr Val Ser Ala Asp  Ile Gln Arg
    1235                1240                1245
```

```
Gln Gly  Gln His Val Val Thr  Val Lys Ser Asp Phe  Phe Leu Gly
    1250                 1255                 1260

Gly Pro  Val Leu Ala Cys Glu  Thr Pro Phe Glu Leu  Thr Glu Glu
    1265                 1270                 1275

Pro Glu  Met Val Val His Val  Asp Ser Glu Val Arg  Arg Ala Ile
    1280                 1285                 1290

Leu His  Ser Arg Lys Trp Leu  Met Arg Glu Asp Arg  Ala Leu Asp
    1295                 1300                 1305

Leu Leu  Gly Arg Gln Leu Leu  Phe Arg Leu Lys Ser  Glu Lys Leu
    1310                 1315                 1320

Phe Arg  Pro Asp Gly Gln Leu  Ala Leu Leu Gln Val  Thr Gly Ser
    1325                 1330                 1335

Val Phe  Ser Tyr Ser Pro Asp  Gly Ser Thr Thr Ala  Phe Gly Arg
    1340                 1345                 1350

Val Tyr  Phe Glu Ser Glu Ser  Cys Thr Gly Asn Val  Val Met Asp
    1355                 1360                 1365

Phe Leu  His Arg Tyr Gly Ala  Pro Arg Ala Gln Leu  Leu Glu Leu
    1370                 1375                 1380

Gln His  Pro Gly Trp Thr Gly  Thr Ser Thr Val Ala  Val Arg Gly
    1385                 1390                 1395

Pro Arg  Arg Ser Gln Ser Tyr  Ala Arg Val Ser Leu  Asp His Asn
    1400                 1405                 1410

Pro Ile  His Val Cys Pro Ala  Phe Ala Arg Tyr Ala  Gly Leu Ser
    1415                 1420                 1425

Gly Pro  Ile Val His Gly Met  Glu Thr Ser Ala Met  Met Arg Arg
    1430                 1435                 1440
```

```
Ile Ala  Glu Trp Ala Ile Gly  Asp Ala Asp Arg Ser  Arg Phe Arg
    1445                1450                1455

Ser Trp  His Ile Thr Leu Gln  Ala Pro Val His Pro  Asn Asp Pro
    1460                1465                1470

Leu Arg  Val Glu Leu Gln His  Lys Ala Met Glu Asp  Gly Glu Met
    1475                1480                1485

Val Leu  Lys Val Gln Ala Phe  Asn Glu Arg Thr Glu  Glu Arg Val
    1490                1495                1500

Ala Glu  Ala Asp Ala His Val  Glu Gln Glu Thr Thr  Ala Tyr Val
    1505                1510                1515

Phe Cys  Gly Gln Gly Ser Gln  Arg Gln Gly Met Gly  Met Asp Leu
    1520                1525                1530

Tyr Val  Asn Cys Pro Glu Ala  Lys Ala Leu Trp Ala  Arg Ala Asp
    1535                1540                1545

Lys His  Leu Trp Glu Lys Tyr  Gly Phe Ser Ile Leu  His Ile Val
    1550                1555                1560

Gln Asn  Asn Pro Pro Ala Leu  Thr Val His Phe Gly  Ser Gln Arg
    1565                1570                1575

Gly Arg  Arg Ile Arg Ala Asn  Tyr Leu Arg Met Met  Gly Gln Pro
    1580                1585                1590

Pro Ile  Asp Gly Arg His Pro  Pro Ile Leu Lys Gly  Leu Thr Arg
    1595                1600                1605

Asn Ser  Thr Ser Tyr Thr Phe  Ser Tyr Ser Gln Gly  Leu Leu Met
    1610                1615                1620

Ser Thr  Gln Phe Ala Gln Pro  Ala Leu Ala Leu Met  Glu Met Ala
    1625                1630                1635
```

```
Gln Phe  Glu Trp Leu Lys Ala  Gln Gly Val Val Gln  Lys Gly Ala
    1640                1645                1650

Arg Phe  Ala Gly His Ser Leu  Gly Glu Tyr Ala Ala  Leu Gly Ala
    1655                1660                1665

Cys Ala  Ser Phe Leu Ser Phe  Glu Asp Leu Ile Ser  Leu Ile Phe
    1670                1675                1680

Tyr Arg  Gly Leu Lys Met Gln  Asn Ala Leu Pro Arg  Asp Ala Asn
    1685                1690                1695

Gly His  Thr Asp Tyr Gly Met  Leu Ala Ala Asp Pro  Ser Arg Ile
    1700                1705                1710

Gly Lys  Gly Phe Glu Glu Ala  Ser Leu Lys Cys Leu  Val His Ile
    1715                1720                1725

Ile Gln  Gln Glu Thr Gly Trp  Phe Val Glu Val Val  Asn Tyr Asn
    1730                1735                1740

Ile Asn  Ser Gln Gln Tyr Val  Cys Ala Gly His Phe  Arg Ala Leu
    1745                1750                1755

Trp Met  Leu Gly Lys Ile Cys  Asp Asp Leu Ser Cys  His Pro Gln
    1760                1765                1770

Pro Glu  Thr Val Glu Gly Gln  Glu Leu Arg Ala Met  Val Trp Lys
    1775                1780                1785

His Val  Pro Thr Val Glu Gln  Val Pro Arg Glu Asp  Arg Met Glu
    1790                1795                1800

Arg Gly  Arg Ala Thr Ile Pro  Leu Pro Gly Ile Asp  Ile Pro Tyr
    1805                1810                1815

His Ser  Thr Met Leu Arg Gly  Glu Ile Glu Pro Tyr  Arg Glu Tyr
    1820                1825                1830
```

```
Leu Ser  Glu Arg Ile Lys Val  Gly Asp Val Lys Pro  Cys Glu Leu
    1835                1840                1845


Val Gly  Arg Trp Ile Pro Asn  Val Val Gly Gln Pro  Phe Ser Val
    1850                1855                1860


Asp Lys  Ser Tyr Val Gln Leu  Val His Gly Ile Thr  Gly Ser Pro
    1865                1870                1875


Arg Leu  His Ser Leu Leu Gln  Gln Met Ala
    1880                1885


<210>  5
<211>  377
<212>  PRT
<213>  Mentha x piperita

<400>  5

Met Ser Ala Leu Val Asn Pro Val Ala Lys Trp Pro Gln Thr Ile Gly
1               5                   10                  15


Val Lys Asp Val His Gly Gly Arg Arg Arg Arg Ser Arg Ser Thr Leu
            20                  25                  30


Phe Gln Ser His Pro Leu Arg Thr Glu Met Pro Phe Ser Leu Tyr Phe
        35                  40                  45


Ser Ser Pro Leu Lys Ala Pro Ala Thr Phe Ser Val Ser Ala Val Tyr
    50                  55                  60


Thr Lys Glu Gly Ser Glu Ile Arg Asp Lys Asp Pro Ala Pro Ser Thr
65                  70                  75                  80


Ser Pro Ala Phe Asp Phe Asp Gly Tyr Met Leu Arg Lys Ala Lys Ser
                85                  90                  95


Val Asn Lys Ala Leu Glu Ala Ala Val Gln Met Lys Glu Pro Leu Lys
            100                 105                 110
```

```
Ile His Glu Ser Met Arg Tyr Ser Leu Leu Ala Gly Gly Lys Arg Val
        115                 120                 125

Arg Pro Met Leu Cys Ile Ala Ala Cys Glu Leu Val Gly Gly Asp Glu
    130                 135                 140

Ser Thr Ala Met Pro Ala Ala Cys Ala Val Glu Met Ile His Thr Met
145                 150                 155                 160

Ser Leu Met His Asp Asp Leu Pro Cys Met Asp Asn Asp Asp Leu Arg
                165                 170                 175

Arg Gly Lys Pro Thr Asn His Met Ala Phe Gly Glu Ser Val Ala Val
            180                 185                 190

Leu Ala Gly Asp Ala Leu Leu Ser Phe Ala Phe Glu His Val Ala Ala
        195                 200                 205

Ala Thr Lys Gly Ala Pro Pro Glu Arg Ile Val Arg Val Leu Gly Glu
    210                 215                 220

Leu Ala Val Ser Ile Gly Ser Glu Gly Leu Val Ala Gly Gln Val Val
225                 230                 235                 240

Asp Val Cys Ser Glu Gly Met Ala Glu Val Gly Leu Asp His Leu Glu
                245                 250                 255

Phe Ile His His His Lys Thr Ala Ala Leu Leu Gln Gly Ser Val Val
            260                 265                 270

Leu Gly Ala Ile Leu Gly Gly Gly Lys Glu Glu Glu Val Ala Lys Leu
        275                 280                 285

Arg Lys Phe Ala Asn Cys Ile Gly Leu Leu Phe Gln Val Val Asp Asp
    290                 295                 300

Ile Leu Asp Val Thr Lys Ser Ser Lys Glu Leu Gly Lys Thr Ala Gly
305                 310                 315                 320
```

```
Lys Asp Leu Val Ala Asp Lys Thr Thr Tyr Pro Lys Leu Ile Gly Val
                325                 330                 335


Glu Lys Ser Lys Glu Phe Ala Asp Arg Leu Asn Arg Glu Ala Gln Glu
            340                 345                 350


Gln Leu Leu His Phe His Pro His Arg Ala Ala Pro Leu Ile Ala Leu
        355                 360                 365


Ala Asn Tyr Ile Ala Tyr Arg Asp Asn
    370                 375


<210>  6
<211>  313
<212>  PRT
<213>  Mentha x piperita

<400>  6

Met Ala Ile Asn Leu Ser His Ile Asn Ser Lys Thr Cys Phe Pro Leu
1               5                   10                  15


Lys Thr Arg Ser Asp Leu Ser Arg Ser Ser Ser Ala Arg Cys Met Pro
            20                  25                  30


Thr Ala Ala Ala Ala Ala Phe Pro Thr Ile Ala Thr Ala Ala Gln Ser
        35                  40                  45


Gln Pro Tyr Trp Ala Ala Ile Glu Ala Asp Ile Glu Arg Tyr Leu Lys
    50                  55                  60


Lys Ser Ile Thr Ile Arg Pro Pro Glu Thr Val Phe Gly Pro Met His
65                  70                  75                  80


His Leu Thr Phe Ala Ala Pro Ala Thr Ala Ala Ser Thr Leu Cys Leu
                85                  90                  95


Ala Ala Cys Glu Leu Val Gly Gly Asp Arg Ser Gln Ala Met Ala Ala
            100                 105                 110
```

```
Ala Ala Ala Ile His Leu Val His Ala Ala Ala Tyr Val His Glu His
        115                 120                 125


Leu Pro Leu Thr Asp Gly Ser Arg Pro Val Ser Lys Pro Ala Ile Gln
    130                 135                 140


His Lys Tyr Gly Pro Asn Val Glu Leu Leu Thr Gly Asp Gly Ile Val
145                 150                 155                 160


Pro Phe Gly Phe Glu Leu Leu Ala Gly Ser Val Asp Pro Ala Arg Thr
                165                 170                 175


Asp Asp Pro Asp Arg Ile Leu Arg Val Ile Ile Glu Ile Ser Arg Ala
            180                 185                 190


Gly Gly Pro Glu Gly Met Ile Ser Gly Leu His Arg Glu Glu Glu Ile
        195                 200                 205


Val Asp Gly Asn Thr Ser Leu Asp Phe Ile Glu Tyr Val Cys Lys Lys
    210                 215                 220


Lys Tyr Gly Glu Met His Ala Cys Gly Ala Ala Cys Gly Ala Ile Leu
225                 230                 235                 240


Gly Gly Ala Ala Glu Glu Glu Ile Gln Lys Leu Arg Asn Phe Gly Leu
                245                 250                 255


Tyr Gln Gly Thr Leu Arg Gly Met Met Glu Met Lys Asn Ser His Gln
            260                 265                 270


Leu Ile Asp Glu Asn Ile Ile Gly Lys Leu Lys Glu Leu Ala Leu Glu
        275                 280                 285


Glu Leu Gly Gly Phe His Gly Lys Asn Ala Glu Leu Met Ser Ser Leu
    290                 295                 300


Val Ala Glu Pro Ser Leu Tyr Ala Ala
305                 310
```

<210> 7
<211> 352
<212> PRT
<213> Saccharomyces sp.

<400> 7

```
Met Ala Ser Glu Lys Glu Ile Arg Arg Glu Arg Phe Leu Asn Val Phe
1               5                  10                  15


Pro Lys Leu Val Glu Glu Leu Asn Ala Ser Leu Leu Ala Tyr Gly Met
            20                  25                  30


Pro Lys Glu Ala Cys Asp Trp Tyr Ala His Ser Leu Asn Tyr Asn Thr
        35                  40                  45


Pro Gly Gly Lys Leu Asn Arg Gly Leu Ser Val Val Asp Thr Tyr Ala
    50                  55                  60


Ile Leu Ser Asn Lys Thr Val Glu Gln Leu Gly Gln Glu Glu Tyr Glu
65                  70                  75                  80


Lys Val Ala Ile Leu Gly Trp Cys Ile Glu Leu Leu Gln Ala Tyr Phe
                85                  90                  95


Leu Val Ala Asp Asp Met Met Asp Lys Ser Ile Thr Arg Arg Gly Gln
            100                 105                 110


Pro Cys Trp Tyr Lys Val Pro Glu Val Gly Glu Ile Ala Ile Asn Asp
        115                 120                 125


Ala Phe Met Leu Glu Ala Ala Ile Tyr Lys Leu Leu Lys Ser His Phe
    130                 135                 140


Arg Asn Glu Lys Tyr Tyr Ile Asp Ile Thr Glu Leu Phe His Glu Val
145                 150                 155                 160


Thr Phe Gln Thr Glu Leu Gly Gln Leu Met Asp Leu Ile Thr Ala Pro
                165                 170                 175


Glu Asp Lys Val Asp Leu Ser Lys Phe Ser Leu Lys Lys His Ser Phe
```

180 185 190

Ile Val Thr Phe Lys Thr Ala Tyr Tyr Ser Phe Tyr Leu Pro Val Ala
195 200 205

Leu Ala Met Tyr Val Ala Gly Ile Thr Asp Glu Lys Asp Leu Lys Gln
210 215 220

Ala Arg Asp Val Leu Ile Pro Leu Gly Glu Tyr Phe Gln Ile Gln Asp
225 230 235 240

Asp Tyr Leu Asp Cys Phe Gly Thr Pro Glu Gln Ile Gly Lys Ile Gly
245 250 255

Thr Asp Ile Gln Asp Asn Lys Cys Ser Trp Val Ile Asn Lys Ala Leu
260 265 270

Glu Leu Ala Ser Ala Glu Gln Arg Lys Thr Leu Asp Glu Asn Tyr Gly
275 280 285

Lys Lys Asp Ser Val Ala Glu Ala Lys Cys Lys Lys Ile Phe Asn Asp
290 295 300

Leu Lys Ile Glu Gln Leu Tyr His Glu Tyr Glu Glu Ser Ile Ala Lys
305 310 315 320

Asp Leu Lys Ala Lys Ile Ser Gln Val Asp Glu Ser Arg Gly Phe Lys
325 330 335

Ala Asp Val Leu Thr Ala Phe Leu Asn Lys Val Tyr Lys Arg Ser Lys
340 345 350

<210> 8
<211> 352
<212> PRT
<213> Artificial Sequence

<220>
<223> Mutated Erg20 farnesylpyrophosphate synthase (K197G)

<400> 8

```
Met Ala Ser Glu Lys Glu Ile Arg Arg Glu Arg Phe Leu Asn Val Phe
1               5                  10                  15

Pro Lys Leu Val Glu Glu Leu Asn Ala Ser Leu Leu Ala Tyr Gly Met
            20                  25                  30

Pro Lys Glu Ala Cys Asp Trp Tyr Ala His Ser Leu Asn Tyr Asn Thr
        35                  40                  45

Pro Gly Gly Lys Leu Asn Arg Gly Leu Ser Val Val Asp Thr Tyr Ala
    50                  55                  60

Ile Leu Ser Asn Lys Thr Val Glu Gln Leu Gly Gln Glu Glu Tyr Glu
65                  70                  75                  80

Lys Val Ala Ile Leu Gly Trp Cys Ile Glu Leu Leu Gln Ala Tyr Phe
                85                  90                  95

Leu Val Ala Asp Asp Met Met Asp Lys Ser Ile Thr Arg Arg Gly Gln
            100                 105                 110

Pro Cys Trp Tyr Lys Val Pro Glu Val Gly Glu Ile Ala Ile Asn Asp
        115                 120                 125

Ala Phe Met Leu Glu Ala Ala Ile Tyr Lys Leu Leu Lys Ser His Phe
    130                 135                 140

Arg Asn Glu Lys Tyr Tyr Ile Asp Ile Thr Glu Leu Phe His Glu Val
145                 150                 155                 160

Thr Phe Gln Thr Glu Leu Gly Gln Leu Met Asp Leu Ile Thr Ala Pro
                165                 170                 175

Glu Asp Lys Val Asp Leu Ser Lys Phe Ser Leu Lys Lys His Ser Phe
            180                 185                 190

Ile Val Thr Phe Gly Thr Ala Tyr Tyr Ser Phe Tyr Leu Pro Val Ala
        195                 200                 205
```

```
Leu Ala Met Tyr Val Ala Gly Ile Thr Asp Glu Lys Asp Leu Lys Gln
    210                 215                 220


Ala Arg Asp Val Leu Ile Pro Leu Gly Glu Tyr Phe Gln Ile Gln Asp
225                 230                 235                 240


Asp Tyr Leu Asp Cys Phe Gly Thr Pro Glu Gln Ile Gly Lys Ile Gly
                245                 250                 255


Thr Asp Ile Gln Asp Asn Lys Cys Ser Trp Val Ile Asn Lys Ala Leu
            260                 265                 270


Glu Leu Ala Ser Ala Glu Gln Arg Lys Thr Leu Asp Glu Asn Tyr Gly
        275                 280                 285


Lys Lys Asp Ser Val Ala Glu Ala Lys Cys Lys Lys Ile Phe Asn Asp
    290                 295                 300


Leu Lys Ile Glu Gln Leu Tyr His Glu Tyr Glu Glu Ser Ile Ala Lys
305                 310                 315                 320


Asp Leu Lys Ala Lys Ile Ser Gln Val Asp Glu Ser Arg Gly Phe Lys
                325                 330                 335


Ala Asp Val Leu Thr Ala Phe Leu Asn Lys Val Tyr Lys Arg Ser Lys
            340                 345                 350


<210>  9
<211>  496
<212>  PRT
<213>  Rhizophagus irregularis

<400>  9

Met Arg Ala Gln Ala His Leu Gly Gly Gly Leu Lys Arg Ile Glu Thr
1               5                   10                  15


Gln His Gln Lys Gly Lys Leu Thr Ala Arg Glu Arg Ala Glu Leu Leu
            20                  25                  30
```

```
Leu Asp Pro Gly Ser Phe Asn Glu Tyr Asp Thr Phe Val Glu His Gln
        35                  40                  45

Cys Thr Asp Phe Gly Met Asp Lys Asn Lys Ile Ile Gly Asp Gly Val
    50                  55                  60

Val Thr Gly His Gly Thr Ile Asn Gly Arg Arg Val Phe Thr Phe Ser
65                  70                  75                  80

Gln Asp Phe Thr Ala Phe Gly Gly Ser Leu Ser Lys Met His Ala Gln
                85                  90                  95

Lys Ile Cys Lys Ile Met Asp Lys Ala Met Leu Val Gly Ala Pro Val
            100                 105                 110

Ile Gly Leu Asn Asp Ser Gly Gly Ala Arg Ile Gln Glu Gly Val Asp
        115                 120                 125

Ser Leu Ala Gly Tyr Ala Asp Ile Phe Gln Arg Asn Val Leu Ser Ser
    130                 135                 140

Gly Val Val Pro Gln Leu Ser Leu Ile Met Gly Pro Cys Ala Gly Gly
145                 150                 155                 160

Ala Val Tyr Ser Pro Ala Leu Thr Asp Phe Thr Phe Met Val Arg Asp
                165                 170                 175

Thr Ser Tyr Leu Phe Val Thr Gly Pro Glu Val Val Lys Ala Val Cys
            180                 185                 190

Asn Glu Asp Val Thr Gln Glu Glu Leu Gly Gly Ala Asn Thr His Thr
        195                 200                 205

Val Ile Ser Gly Val Ala His Ala Ala Phe Glu Asn Asp Ile Glu Ala
    210                 215                 220

Ile Gln Arg Ile Arg Asp Phe Met Asp Phe Leu Pro Leu Ser Asn Arg
225                 230                 235                 240
```

```
Glu Gln Ala Pro Thr Arg Tyr Ser Asp Asp Pro Ile Asp Arg Glu Asp
                245                 250                 255

Pro Ser Leu Asn His Ile Ile Pro Val Asp Ser Thr Lys Ala Tyr Asp
            260                 265                 270

Met Arg Glu Ile Ile Thr Arg Leu Ile Asp Asp Gly His Phe Phe Glu
        275                 280                 285

Ile Met Pro Asp Tyr Ala Lys Asn Ile Val Val Gly Phe Ala Arg Met
    290                 295                 300

Gly Gly Lys Thr Val Ser Ile Val Gly Asn Gln Pro Leu Val Ser Ser
305                 310                 315                 320

Gly Val Leu Asp Ile Asn Ser Ser Val Lys Ala Ala Arg Phe Val Arg
                325                 330                 335

Phe Cys Asp Ala Phe Asn Ile Pro Ile Ile Thr Leu Val Asp Val Pro
            340                 345                 350

Gly Phe Leu Pro Gly Thr Ala Gln Glu His Asn Gly Ile Ile Arg His
        355                 360                 365

Gly Ala Lys Leu Leu Tyr Ala Tyr Ala Glu Ala Thr Val Pro Lys Ile
    370                 375                 380

Thr Ile Ile Thr Arg Lys Ala Tyr Gly Gly Ala Tyr Asp Val Met Ser
385                 390                 395                 400

Ser Lys His Leu Arg Gly Asp Met Asn Tyr Ser Trp Pro Thr Gly Glu
                405                 410                 415

Ile Ala Val Met Gly Ala Lys Gly Ala Val Glu Ile Ile Phe Arg His
            420                 425                 430

Val Glu Asp Arg Thr Gln Ser Glu His Glu Tyr Ile Asp Lys Phe Ala
        435                 440                 445
```

```
Asn Pro Ile Pro Ala Ala Gln Arg Gly Tyr Ile Asp Asp Ile Ile Leu
    450                 455                 460


Pro Ala Ala Thr Arg Lys Arg Ile Ile Glu Asp Leu Phe Val Leu Ser
465                 470                 475                 480


His Lys Gln Leu Pro Leu Ile Tyr Lys Lys His Asp Asn Cys Pro Leu
                485                 490                 495


<210>  10
<211>  101
<212>  PRT
<213>  Cannabis sativa

<400>  10

Met Ala Val Lys His Leu Ile Val Leu Lys Phe Lys Asp Glu Ile Thr
1               5                   10                  15


Glu Ala Gln Lys Glu Glu Phe Phe Lys Thr Tyr Val Asn Leu Val Asn
            20                  25                  30


Ile Ile Pro Ala Met Lys Asp Val Tyr Trp Gly Lys Asp Val Thr Gln
        35                  40                  45


Lys Asn Lys Glu Glu Gly Tyr Thr His Ile Val Glu Val Thr Phe Glu
    50                  55                  60


Ser Val Glu Thr Ile Gln Asp Tyr Ile Ile His Pro Ala His Val Gly
65                  70                  75                  80


Phe Gly Asp Val Tyr Arg Ser Phe Trp Glu Lys Leu Leu Ile Phe Asp
                85                  90                  95


Tyr Thr Pro Arg Lys
            100


<210>  11
<211>  385
<212>  PRT
<213>  Cannabis sativa
```

<400> 11

```
Met Asn His Leu Arg Ala Glu Gly Pro Ala Ser Val Leu Ala Ile Gly
1               5                  10                  15

Thr Ala Asn Pro Glu Asn Ile Leu Leu Gln Asp Glu Phe Pro Asp Tyr
            20                  25                  30

Tyr Phe Arg Val Thr Lys Ser Glu His Met Thr Gln Leu Lys Glu Lys
        35                  40                  45

Phe Arg Lys Ile Cys Asp Lys Ser Met Ile Arg Lys Arg Asn Cys Phe
    50                  55                  60

Leu Asn Glu Glu His Leu Lys Gln Asn Pro Arg Leu Val Glu His Glu
65                  70                  75                  80

Met Gln Thr Leu Asp Ala Arg Gln Asp Met Leu Val Val Glu Val Pro
                85                  90                  95

Lys Leu Gly Lys Asp Ala Cys Ala Lys Ala Ile Lys Glu Trp Gly Gln
            100                 105                 110

Pro Lys Ser Lys Ile Thr His Leu Ile Phe Thr Ser Ala Ser Thr Thr
        115                 120                 125

Asp Met Pro Gly Ala Asp Tyr His Cys Ala Lys Leu Leu Gly Leu Ser
    130                 135                 140

Pro Ser Val Lys Arg Val Met Met Tyr Gln Leu Gly Cys Tyr Gly Gly
145                 150                 155                 160

Gly Thr Val Leu Arg Ile Ala Lys Asp Ile Ala Glu Asn Asn Lys Gly
                165                 170                 175

Ala Arg Val Leu Ala Val Cys Cys Asp Ile Met Ala Cys Leu Phe Arg
            180                 185                 190

Gly Pro Ser Glu Ser Asp Leu Glu Leu Leu Val Gly Gln Ala Ile Phe
        195                 200                 205
```

```
Gly Asp Gly Ala Ala Ala Val Ile Val Gly Ala Glu Pro Asp Glu Ser
    210                 215                 220

Val Gly Glu Arg Pro Ile Phe Glu Leu Val Ser Thr Gly Gln Thr Ile
225                 230                 235                 240

Leu Pro Asn Ser Glu Gly Thr Ile Gly Gly His Ile Arg Glu Ala Gly
                245                 250                 255

Leu Ile Phe Asp Leu His Lys Asp Val Pro Met Leu Ile Ser Asn Asn
            260                 265                 270

Ile Glu Lys Cys Leu Ile Glu Ala Phe Thr Pro Ile Gly Ile Ser Asp
        275                 280                 285

Trp Asn Ser Ile Phe Trp Ile Thr His Pro Gly Gly Lys Ala Ile Leu
    290                 295                 300

Asp Lys Val Glu Glu Lys Leu His Leu Lys Ser Asp Lys Phe Val Asp
305                 310                 315                 320

Ser Arg His Val Leu Ser Glu His Gly Asn Met Ser Ser Ser Thr Val
                325                 330                 335

Leu Phe Val Met Asp Glu Leu Arg Lys Arg Ser Leu Glu Glu Gly Lys
            340                 345                 350

Ser Thr Thr Gly Asp Gly Phe Glu Trp Gly Val Leu Phe Gly Phe Gly
        355                 360                 365

Pro Gly Leu Thr Val Glu Arg Val Val Val Arg Ser Val Pro Ile Lys
    370                 375                 380

Tyr
385
```

<210> 12
<211> 327

<212> PRT
<213> Artificial Sequence

<220>
<223> Truncated geranyl pyrophosphate olivetolic acid geranyltransferase (GOT)

<400> 12

```
Met Asn Ser Ile Arg Ala Ala Thr Thr Asn Gln Thr Glu Pro Pro Glu
1               5                  10                  15

Ser Asp Asn His Ser Val Ala Thr Lys Ile Leu Asn Phe Gly Lys Ala
            20                  25                  30

Cys Trp Lys Leu Gln Arg Pro Tyr Thr Ile Ile Ala Phe Thr Ser Cys
        35                  40                  45

Ala Cys Gly Leu Phe Gly Lys Glu Leu Leu His Asn Thr Asn Leu Ile
    50                  55                  60

Ser Trp Ser Leu Met Phe Lys Ala Phe Phe Phe Leu Val Ala Ile Leu
65                  70                  75                  80

Cys Ile Ala Ser Phe Thr Thr Thr Ile Asn Gln Ile Tyr Asp Leu His
                85                  90                  95

Ile Asp Arg Ile Asn Lys Pro Asp Leu Pro Leu Ala Ser Gly Glu Ile
            100                 105                 110

Ser Val Asn Thr Ala Trp Ile Met Ser Ile Ile Val Ala Leu Phe Gly
        115                 120                 125

Leu Ile Ile Thr Ile Lys Met Lys Gly Gly Pro Leu Tyr Ile Phe Gly
    130                 135                 140

Tyr Cys Phe Gly Ile Phe Gly Gly Ile Val Tyr Ser Val Pro Pro Phe
145                 150                 155                 160

Arg Trp Lys Gln Asn Pro Ser Thr Ala Phe Leu Leu Asn Phe Leu Ala
                165                 170                 175
```

```
His Ile Ile Thr Asn Phe Thr Phe Tyr Tyr Ala Ser Arg Ala Ala Leu
            180                 185                 190


Gly Leu Pro Phe Glu Leu Arg Pro Ser Phe Thr Phe Leu Leu Ala Phe
        195                 200                 205


Met Lys Ser Met Gly Ser Ala Leu Ala Leu Ile Lys Asp Ala Ser Asp
    210                 215                 220


Val Glu Gly Asp Thr Lys Phe Gly Ile Ser Thr Leu Ala Ser Lys Tyr
225                 230                 235                 240


Gly Ser Arg Asn Leu Thr Leu Phe Cys Ser Gly Ile Val Leu Leu Ser
                245                 250                 255


Tyr Val Ala Ala Ile Leu Ala Gly Ile Ile Trp Pro Gln Ala Phe Asn
            260                 265                 270


Ser Asn Val Met Leu Leu Ser His Ala Ile Leu Ala Phe Trp Leu Ile
        275                 280                 285


Leu Gln Thr Arg Asp Phe Ala Leu Thr Asn Tyr Asp Pro Glu Ala Gly
    290                 295                 300


Arg Arg Phe Tyr Glu Phe Met Trp Lys Leu Tyr Tyr Ala Glu Tyr Leu
305                 310                 315                 320


Val Tyr Val Phe Ile Gly Ser
                325
```

<210> 13
<211> 332
<212> PRT
<213> Artificial Sequence

<220>
<223> Engineered geranyl pyrophosphate olivetolic acid
       geranyltransferase (GOT)

<400> 13

```
Met Lys Asp Gln Arg Gly Asn Ser Ile Arg Ala Ser Ala Gln Ile Glu
1               5                   10                  15

Asp Arg Pro Pro Glu Ser Gly Asn Leu Ser Ala Leu Thr Asn Val Lys
            20                  25                  30

Asp Phe Val Ser Val Cys Trp Glu Tyr Val Arg Pro Tyr Thr Ala Lys
        35                  40                  45

Gly Val Ile Ile Cys Ser Ser Cys Leu Phe Gly Arg Glu Leu Leu Glu
    50                  55                  60

Asn Pro Asn Leu Phe Ser Trp Pro Leu Ile Phe Lys Ala Phe Phe Phe
65                  70                  75                  80

Leu Val Ala Ile Leu Cys Ile Ala Ser Phe Thr Thr Thr Ile Asn Gln
                85                  90                  95

Ile Tyr Asp Leu His Ile Asp Arg Ile Asn Lys Pro Asp Leu Pro Leu
            100                 105                 110

Ala Ser Gly Glu Ile Ser Val Asn Thr Ala Trp Ile Met Ser Ile Ile
        115                 120                 125

Val Ala Leu Phe Gly Leu Ile Ile Thr Ile Lys Met Lys Gly Gly Pro
    130                 135                 140

Leu Tyr Ile Phe Gly Tyr Cys Phe Gly Ile Phe Gly Gly Ile Val Tyr
145                 150                 155                 160

Ser Val Pro Pro Phe Arg Trp Lys Gln Asn Pro Ser Thr Ala Phe Leu
                165                 170                 175

Leu Asn Phe Leu Ala His Ile Ile Thr Asn Phe Thr Phe Tyr Tyr Ala
            180                 185                 190

Ser Arg Ala Ala Leu Gly Leu Pro Phe Glu Leu Arg Pro Ser Phe Thr
        195                 200                 205
```

Phe Leu Leu Ala Phe Met Lys Ser Met Gly Ser Ala Leu Ala Leu Ile
210 215 220

Lys Asp Ala Ser Asp Val Glu Gly Asp Thr Lys Phe Gly Ile Ser Thr
225 230 235 240

Leu Ala Ser Lys Tyr Gly Ser Arg Asn Leu Thr Leu Phe Cys Ser Gly
245 250 255

Ile Val Leu Leu Ser Tyr Val Ala Ala Ile Leu Ala Gly Ile Ile Trp
260 265 270

Pro Gln Ala Phe Asn Ser Asn Val Met Leu Leu Ser His Ala Ile Leu
275 280 285

Ala Phe Trp Leu Ile Leu Gln Thr Arg Asp Phe Ala Leu Thr Asn Tyr
290 295 300

Asp Pro Glu Ala Gly Arg Arg Phe Tyr Glu Phe Met Trp Lys Leu Tyr
305 310 315 320

Tyr Ala Glu Tyr Leu Val Tyr Val Phe Ile Gly Ser
325 330

<210> 14
<211> 547
<212> PRT
<213> Artificial Sequence

<220>
<223> Mutant tetrahydro cannabinolic acid synthase (THCAS)

<400> 14

Met Asn Cys Ser Ala Phe Ser Phe Trp Phe Val Cys Lys Ile Ile Phe
1 5 10 15

Phe Phe Leu Ser Phe His Ile Gln Ile Ser Ile Ala Asn Pro Arg Glu
20 25 30

Asn Phe Leu Lys Cys Phe Ser Lys His Ile Pro Asn Asn Val Ala Asn
35 40 45

```
Pro Lys Leu Val Tyr Thr Gln His Asp Gln Leu Tyr Met Ser Ile Leu
    50                  55                  60

Asn Ser Thr Ile Gln Asn Leu Arg Phe Ile Ser Asp Thr Thr Pro Lys
65                  70                  75                  80

Pro Leu Val Ile Val Thr Pro Ser Asn Asn Ser His Ile Gln Ala Thr
                85                  90                  95

Ile Leu Cys Ser Lys Lys Val Gly Leu Gln Ile Arg Thr Arg Ser Gly
            100                 105                 110

Gly His Asp Ala Glu Gly Met Ser Tyr Ile Ser Gln Val Pro Phe Val
        115                 120                 125

Val Val Asp Leu Arg Asn Met His Ser Ile Lys Ile Asp Val His Ser
    130                 135                 140

Gln Thr Ala Trp Val Glu Ala Gly Ala Thr Leu Gly Glu Val Tyr Tyr
145                 150                 155                 160

Trp Ile Asn Glu Lys Asn Glu Asn Leu Ser Phe Pro Gly Gly Tyr Cys
                165                 170                 175

Pro Thr Val Gly Val Gly Gly His Phe Ser Gly Gly Gly Tyr Gly Ala
            180                 185                 190

Leu Met Arg Asn Tyr Gly Leu Ala Ala Asp Asn Ile Ile Asp Ala His
        195                 200                 205

Leu Val Asn Val Asp Gly Lys Val Leu Asp Arg Lys Ser Met Gly Glu
    210                 215                 220

Asp Leu Phe Trp Ala Ile Arg Gly Gly Gly Gly Glu Asn Phe Gly Ile
225                 230                 235                 240

Ile Ala Ala Trp Lys Ile Lys Leu Val Ala Val Pro Ser Lys Ser Thr
                245                 250                 255
```

```
Ile Phe Ser Val Lys Lys Asn Met Glu Ile His Gly Leu Val Lys Leu
            260                 265                 270

Phe Asn Lys Trp Gln Asn Ile Ala Tyr Lys Tyr Asp Lys Asp Leu Val
        275                 280                 285

Leu Met Thr His Phe Ile Thr Lys Asn Ile Thr Asp Asn His Gly Lys
    290                 295                 300

Asn Lys Thr Thr Val His Gly Tyr Phe Ser Ser Ile Phe His Gly Gly
305                 310                 315                 320

Val Asp Ser Leu Val Asp Leu Met Asn Lys Ser Phe Pro Glu Leu Gly
                325                 330                 335

Ile Lys Lys Thr Asp Cys Lys Glu Phe Ser Trp Ile Asp Thr Thr Ile
            340                 345                 350

Phe Tyr Ser Gly Val Val Asn Phe Asn Thr Ala Asn Phe Lys Lys Glu
        355                 360                 365

Ile Leu Leu Asp Arg Ser Ala Gly Lys Lys Thr Ala Phe Ser Ile Lys
    370                 375                 380

Leu Asp Tyr Val Lys Lys Pro Ile Pro Glu Thr Ala Met Val Lys Ile
385                 390                 395                 400

Leu Glu Lys Leu Tyr Glu Glu Asp Val Gly Ala Gly Met Tyr Val Leu
                405                 410                 415

Tyr Pro Tyr Gly Gly Ile Met Glu Glu Ile Ser Glu Ser Ala Ile Pro
            420                 425                 430

Phe Pro His Arg Ala Gly Ile Met Tyr Glu Leu Trp Tyr Thr Ala Ser
        435                 440                 445

Trp Glu Lys Gln Glu Asp Asn Glu Lys His Ile Asn Trp Val Arg Ser
    450                 455                 460
```

```
Val Tyr Asn Phe Thr Thr Pro Tyr Val Ser Gln Asn Pro Arg Leu Ala
465                 470                 475                 480


Tyr Leu Asn Tyr Arg Asp Leu Asp Leu Gly Lys Thr Asn His Ala Ser
                485                 490                 495


Pro Asn Asn Tyr Thr Gln Ala Arg Ile Trp Gly Glu Lys Tyr Phe Gly
            500                 505                 510


Lys Asn Phe Asn Arg Leu Val Lys Val Lys Thr Lys Val Asp Pro Asn
        515                 520                 525


Asn Phe Phe Arg Asn Glu Gln Ser Ile Pro Pro Leu Pro Pro His His
    530                 535                 540


His Gly Ser
545


<210>  15
<211>  520
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Truncated tetrahydro cannabinolic acid synthase (THCAS)

<400>  15

Met Asn Pro Arg Glu Asn Phe Leu Lys Cys Phe Ser Lys His Ile Pro
1               5                   10                  15


Asn Asn Val Ala Asn Pro Lys Leu Val Tyr Thr Gln His Asp Gln Leu
            20                  25                  30


Tyr Met Ser Ile Leu Asn Ser Thr Ile Gln Asn Leu Arg Phe Ile Ser
        35                  40                  45


Asp Thr Thr Pro Lys Pro Leu Val Ile Val Thr Pro Ser Asn Asn Ser
    50                  55                  60
```

```
His Ile Gln Ala Thr Ile Leu Cys Ser Lys Lys Val Gly Leu Gln Ile
65                  70                  75                  80

Arg Thr Arg Ser Gly Gly His Asp Ala Glu Gly Met Ser Tyr Ile Ser
                85                  90                  95

Gln Val Pro Phe Val Val Val Asp Leu Arg Asn Met His Ser Ile Lys
            100                 105                 110

Ile Asp Val His Ser Gln Thr Ala Trp Val Glu Ala Gly Ala Thr Leu
        115                 120                 125

Gly Glu Val Tyr Tyr Trp Ile Asn Glu Lys Asn Glu Asn Leu Ser Phe
    130                 135                 140

Pro Gly Gly Tyr Cys Pro Thr Val Gly Val Gly Gly His Phe Ser Gly
145                 150                 155                 160

Gly Gly Tyr Gly Ala Leu Met Arg Asn Tyr Gly Leu Ala Ala Asp Asn
                165                 170                 175

Ile Ile Asp Ala His Leu Val Asn Val Asp Gly Lys Val Leu Asp Arg
            180                 185                 190

Lys Ser Met Gly Glu Asp Leu Phe Trp Ala Ile Arg Gly Gly Gly Gly
        195                 200                 205

Glu Asn Phe Gly Ile Ile Ala Ala Trp Lys Ile Lys Leu Val Ala Val
    210                 215                 220

Pro Ser Lys Ser Thr Ile Phe Ser Val Lys Lys Asn Met Glu Ile His
225                 230                 235                 240

Gly Leu Val Lys Leu Phe Asn Lys Trp Gln Asn Ile Ala Tyr Lys Tyr
                245                 250                 255

Asp Lys Asp Leu Val Leu Met Thr His Phe Ile Thr Lys Asn Ile Thr
            260                 265                 270
```

```
Asp Asn His Gly Lys Asn Lys Thr Thr Val His Gly Tyr Phe Ser Ser
        275                 280                 285

Ile Phe His Gly Gly Val Asp Ser Leu Val Asp Leu Met Asn Lys Ser
    290                 295                 300

Phe Pro Glu Leu Gly Ile Lys Lys Thr Asp Cys Lys Glu Phe Ser Trp
305                 310                 315                 320

Ile Asp Thr Thr Ile Phe Tyr Ser Gly Val Val Asn Phe Asn Thr Ala
                325                 330                 335

Asn Phe Lys Lys Glu Ile Leu Leu Asp Arg Ser Ala Gly Lys Lys Thr
            340                 345                 350

Ala Phe Ser Ile Lys Leu Asp Tyr Val Lys Lys Pro Ile Pro Glu Thr
        355                 360                 365

Ala Met Val Lys Ile Leu Glu Lys Leu Tyr Glu Glu Asp Val Gly Ala
    370                 375                 380

Gly Met Tyr Val Leu Tyr Pro Tyr Gly Gly Ile Met Glu Glu Ile Ser
385                 390                 395                 400

Glu Ser Ala Ile Pro Phe Pro His Arg Ala Gly Ile Met Tyr Glu Leu
                405                 410                 415

Trp Tyr Thr Ala Ser Trp Glu Lys Gln Glu Asp Asn Glu Lys His Ile
            420                 425                 430

Asn Trp Val Arg Ser Val Tyr Asn Phe Thr Thr Pro Tyr Val Ser Gln
        435                 440                 445

Asn Pro Arg Leu Ala Tyr Leu Asn Tyr Arg Asp Leu Asp Leu Gly Lys
    450                 455                 460

Thr Asn His Ala Ser Pro Asn Asn Tyr Thr Gln Ala Arg Ile Trp Gly
465                 470                 475                 480
```

Glu Lys Tyr Phe Gly Lys Asn Phe Asn Arg Leu Val Lys Val Lys Thr
485 490 495

Lys Val Asp Pro Asn Asn Phe Phe Arg Asn Glu Gln Ser Ile Pro Pro
500 505 510

Leu Pro Pro His His His Gly Ser
515 520

<210> 16
<211> 519
<212> PRT
<213> Artificial Sequence

<220>
<223> Truncated cannabidiolic acid synthase (CBDAS)

<400> 16

Met Asn Pro Arg Glu Asn Phe Leu Lys Cys Phe Ser Gln Tyr Ile Pro
1 5 10 15

Asn Asn Ala Thr Asn Leu Lys Leu Val Tyr Thr Gln Asn Asn Pro Leu
20 25 30

Tyr Met Ser Val Leu Asn Ser Thr Ile His Asn Leu Arg Phe Thr Ser
35 40 45

Asp Thr Thr Pro Lys Pro Leu Val Ile Val Thr Pro Ser His Val Ser
50 55 60

His Ile Gln Gly Thr Ile Leu Cys Ser Lys Lys Val Gly Leu Gln Ile
65 70 75 80

Arg Thr Arg Ser Gly Gly His Asp Ser Glu Gly Met Ser Tyr Ile Ser
85 90 95

Gln Val Pro Phe Val Ile Val Asp Leu Arg Asn Met Arg Ser Ile Lys
100 105 110

Ile Asp Val His Ser Gln Thr Ala Trp Val Glu Ala Gly Ala Thr Leu
115 120 125

```
Gly Glu Val Tyr Tyr Trp Val Asn Glu Lys Asn Glu Asn Leu Ser Leu
    130                 135                 140

Ala Ala Gly Tyr Cys Pro Thr Val Cys Ala Gly Gly His Phe Gly Gly
145                 150                 155                 160

Gly Gly Tyr Gly Pro Leu Met Arg Asn Tyr Gly Leu Ala Ala Asp Asn
                165                 170                 175

Ile Ile Asp Ala His Leu Val Asn Val His Gly Lys Val Leu Asp Arg
            180                 185                 190

Lys Ser Met Gly Glu Asp Leu Phe Trp Ala Leu Arg Gly Gly Gly Ala
        195                 200                 205

Glu Ser Phe Gly Ile Ile Val Ala Trp Lys Ile Arg Leu Val Ala Val
    210                 215                 220

Pro Lys Ser Thr Met Phe Ser Val Lys Lys Ile Met Glu Ile His Glu
225                 230                 235                 240

Leu Val Lys Leu Val Asn Lys Trp Gln Asn Ile Ala Tyr Lys Tyr Asp
                245                 250                 255

Lys Asp Leu Leu Leu Met Thr His Phe Ile Thr Arg Asn Ile Thr Asp
            260                 265                 270

Asn Gln Gly Lys Asn Lys Thr Ala Ile His Thr Tyr Phe Ser Ser Val
        275                 280                 285

Phe Leu Gly Gly Val Asp Ser Leu Val Asp Leu Met Asn Lys Ser Phe
    290                 295                 300

Pro Glu Leu Gly Ile Lys Lys Thr Asp Cys Arg Gln Leu Ser Trp Ile
305                 310                 315                 320

Asp Thr Ile Ile Phe Tyr Ser Gly Val Val Asn Tyr Asp Thr Asp Asn
                325                 330                 335
```

```
Phe Asn Lys Glu Ile Leu Leu Asp Arg Ser Ala Gly Gln Asn Gly Ala
            340                 345                 350

Phe Lys Ile Lys Leu Asp Tyr Val Lys Lys Pro Ile Pro Glu Ser Val
        355                 360                 365

Phe Val Gln Ile Leu Glu Lys Leu Tyr Glu Glu Asp Ile Gly Ala Gly
    370                 375                 380

Met Tyr Ala Leu Tyr Pro Tyr Gly Gly Ile Met Asp Glu Ile Ser Glu
385                 390                 395                 400

Ser Ala Ile Pro Phe Pro His Arg Ala Gly Ile Leu Tyr Glu Leu Trp
                405                 410                 415

Tyr Ile Cys Ser Trp Glu Lys Gln Glu Asp Asn Glu Lys His Leu Asn
            420                 425                 430

Trp Ile Arg Asn Ile Tyr Asn Phe Met Thr Pro Tyr Val Ser Lys Asn
        435                 440                 445

Pro Arg Leu Ala Tyr Leu Asn Tyr Arg Asp Leu Asp Ile Gly Ile Asn
    450                 455                 460

Asp Pro Lys Asn Pro Asn Asn Tyr Thr Gln Ala Arg Ile Trp Gly Glu
465                 470                 475                 480

Lys Tyr Phe Gly Lys Asn Phe Asp Arg Leu Val Lys Val Lys Thr Leu
                485                 490                 495

Val Asp Pro Asn Asn Phe Phe Arg Asn Glu Gln Ser Ile Pro Pro Leu
            500                 505                 510

Pro Arg His Arg His Gly Ser
        515
```

<210> 17
<211> 502

<212> PRT
<213> Artificial Sequence

<220>
<223> Truncated 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMGR)

<400> 17

```
Met Val Leu Thr Asn Lys Thr Val Ile Ser Gly Ser Lys Val Lys Ser
1               5                  10                  15


Leu Ser Ser Ala Gln Ser Ser Ser Ser Gly Pro Ser Ser Ser Ser Glu
            20                  25                  30


Glu Asp Asp Ser Arg Asp Ile Glu Ser Leu Asp Lys Lys Ile Arg Pro
        35                  40                  45


Leu Glu Glu Leu Glu Ala Leu Leu Ser Ser Gly Asn Thr Lys Gln Leu
    50                  55                  60


Lys Asn Lys Glu Val Ala Ala Leu Val Ile His Gly Lys Leu Pro Leu
65                  70                  75                  80


Tyr Ala Leu Glu Lys Lys Leu Gly Asp Thr Thr Arg Ala Val Ala Val
                85                  90                  95


Arg Arg Lys Ala Leu Ser Ile Leu Ala Glu Ala Pro Val Leu Ala Ser
            100                 105                 110


Asp Arg Leu Pro Tyr Lys Asn Tyr Asp Tyr Asp Arg Val Phe Gly Ala
        115                 120                 125


Cys Cys Glu Asn Val Ile Gly Tyr Met Pro Leu Pro Val Gly Val Ile
    130                 135                 140


Gly Pro Leu Val Ile Asp Gly Thr Ser Tyr His Ile Pro Met Ala Thr
145                 150                 155                 160


Thr Glu Gly Cys Leu Val Ala Ser Ala Met Arg Gly Cys Lys Ala Ile
                165                 170                 175
```

```
Asn Ala Gly Gly Gly Ala Thr Thr Val Leu Thr Lys Asp Gly Met Thr
            180                 185                 190


Arg Gly Pro Val Val Arg Phe Pro Thr Leu Lys Arg Ser Gly Ala Cys
        195                 200                 205


Lys Ile Trp Leu Asp Ser Glu Glu Gly Gln Asn Ala Ile Lys Lys Ala
    210                 215                 220


Phe Asn Ser Thr Ser Arg Phe Ala Arg Leu Gln His Ile Gln Thr Cys
225                 230                 235                 240


Leu Ala Gly Asp Leu Leu Phe Met Arg Phe Arg Thr Thr Thr Gly Asp
                245                 250                 255


Ala Met Gly Met Asn Met Ile Ser Lys Gly Val Glu Tyr Ser Leu Lys
            260                 265                 270


Gln Met Val Glu Glu Tyr Gly Trp Glu Asp Met Glu Val Val Ser Val
        275                 280                 285


Ser Gly Asn Tyr Cys Thr Asp Lys Lys Pro Ala Ala Ile Asn Trp Ile
    290                 295                 300


Glu Gly Arg Gly Lys Ser Val Val Ala Glu Ala Thr Ile Pro Gly Asp
305                 310                 315                 320


Val Val Arg Lys Val Leu Lys Ser Asp Val Ser Ala Leu Val Glu Leu
                325                 330                 335


Asn Ile Ala Lys Asn Leu Val Gly Ser Ala Met Ala Gly Ser Val Gly
            340                 345                 350


Gly Phe Asn Ala His Ala Ala Asn Leu Val Thr Ala Val Phe Leu Ala
        355                 360                 365


Leu Gly Gln Asp Pro Ala Gln Asn Val Glu Ser Ser Asn Cys Ile Thr
    370                 375                 380
```

```
Leu Met Lys Glu Val Asp Gly Asp Leu Arg Ile Ser Val Ser Met Pro
385                 390                 395                 400


Ser Ile Glu Val Gly Thr Ile Gly Gly Gly Thr Val Leu Glu Pro Gln
                405                 410                 415


Gly Ala Met Leu Asp Leu Leu Gly Val Arg Gly Pro His Ala Thr Ala
            420                 425                 430


Pro Gly Thr Asn Ala Arg Gln Leu Ala Arg Ile Val Ala Cys Ala Val
        435                 440                 445


Leu Ala Gly Glu Leu Ser Leu Cys Ala Ala Leu Ala Ala Gly His Leu
    450                 455                 460


Val Gln Ser His Met Thr His Asn Arg Lys Pro Ala Glu Pro Thr Lys
465                 470                 475                 480


Pro Asn Asn Leu Asp Ala Thr Asp Ile Asn Arg Leu Lys Asp Gly Ser
                485                 490                 495


Val Thr Cys Ile Lys Ser
            500


<210>  18
<211>  284
<212>  PRT
<213>  Ralstonia sp.

<400>  18

Met Ser Ile Arg Thr Val Gly Ile Val Gly Ala Gly Thr Met Gly Asn
1               5                   10                  15


Gly Ile Ala Gln Ala Cys Ala Val Val Gly Leu Asn Val Val Met Val
            20                  25                  30


Asp Ile Ser Asp Ala Ala Val Gln Lys Gly Val Ala Thr Val Ala Ser
        35                  40                  45


Ser Leu Asp Arg Leu Ile Lys Lys Glu Lys Leu Thr Glu Ala Asp Lys
```

50 55 60

Ala Ser Ala Leu Ala Arg Ile Lys Gly Ser Thr Ser Tyr Asp Asp Leu
65 70 75 80

Lys Ala Thr Asp Ile Val Ile Glu Ala Ala Thr Glu Asn Tyr Asp Leu
85 90 95

Lys Val Lys Ile Leu Lys Gln Ile Asp Gly Ile Val Gly Glu Asn Val
100 105 110

Ile Ile Ala Ser Asn Thr Ser Ser Ile Ser Ile Thr Lys Leu Ala Ala
115 120 125

Val Thr Ser Arg Ala Asp Arg Phe Ile Gly Met His Phe Phe Asn Pro
130 135 140

Val Pro Val Met Ala Leu Val Glu Leu Ile Arg Gly Leu Gln Thr Ser
145 150 155 160

Asp Thr Thr His Ala Ala Val Glu Ala Leu Ser Lys Gln Leu Gly Lys
165 170 175

Tyr Pro Ile Thr Val Lys Asn Ser Pro Gly Phe Val Val Asn Arg Ile
180 185 190

Leu Cys Pro Met Ile Asn Glu Ala Phe Cys Val Leu Gly Glu Gly Leu
195 200 205

Ala Ser Pro Glu Glu Ile Asp Glu Gly Met Lys Leu Gly Cys Asn His
210 215 220

Pro Ile Gly Pro Leu Ala Leu Ala Asp Met Ile Gly Leu Asp Thr Met
225 230 235 240

Leu Ala Val Met Glu Val Leu Tyr Thr Glu Phe Ala Asp Pro Lys Tyr
245 250 255

Arg Pro Ala Met Leu Met Arg Glu Met Val Ala Ala Gly Tyr Leu Gly

260 265 270

Arg Lys Thr Gly Arg Gly Val Tyr Val Tyr Ser Lys
275 280

<210> 19
<211> 261
<212> PRT
<213> Clostridium sp.

<400> 19

Met Glu Leu Asn Asn Val Ile Leu Glu Lys Glu Gly Lys Val Ala Val
1 5 10 15

Val Thr Ile Asn Arg Pro Lys Ala Leu Asn Ala Leu Asn Ser Asp Thr
20 25 30

Leu Lys Glu Met Asp Tyr Val Ile Gly Glu Ile Glu Asn Asp Ser Glu
35 40 45

Val Leu Ala Val Ile Leu Thr Gly Ala Gly Glu Lys Ser Phe Val Ala
50 55 60

Gly Ala Asp Ile Ser Glu Met Lys Glu Met Asn Thr Ile Glu Gly Arg
65 70 75 80

Lys Phe Gly Ile Leu Gly Asn Lys Val Phe Arg Arg Leu Glu Leu Leu
85 90 95

Glu Lys Pro Val Ile Ala Ala Val Asn Gly Phe Ala Leu Gly Gly Gly
100 105 110

Cys Glu Ile Ala Met Ser Cys Asp Ile Arg Ile Ala Ser Ser Asn Ala
115 120 125

Arg Phe Gly Gln Pro Glu Val Gly Leu Gly Ile Thr Pro Gly Phe Gly
130 135 140

Gly Thr Gln Arg Leu Ser Arg Leu Val Gly Met Gly Met Ala Lys Gln
145 150 155 160

```
Leu Ile Phe Thr Ala Gln Asn Ile Lys Ala Asp Glu Ala Leu Arg Ile
                165                 170                 175

Gly Leu Val Asn Lys Val Val Glu Pro Ser Glu Leu Met Asn Thr Ala
            180                 185                 190

Lys Glu Ile Ala Asn Lys Ile Val Ser Asn Ala Pro Val Ala Val Lys
        195                 200                 205

Leu Ser Lys Gln Ala Ile Asn Arg Gly Met Gln Cys Asp Ile Asp Thr
    210                 215                 220

Ala Leu Ala Phe Glu Ser Glu Ala Phe Gly Glu Cys Phe Ser Thr Glu
225                 230                 235                 240

Asp Gln Lys Asp Ala Met Thr Ala Phe Ile Glu Lys Arg Lys Ile Glu
                245                 250                 255

Gly Phe Lys Asn Arg
            260


<210>  20
<211>  397
<212>  PRT
<213>  Treponema sp.

<400>  20

Met Ile Val Lys Pro Met Val Arg Asn Asn Ile Cys Leu Asn Ala His
1               5                   10                  15

Pro Gln Gly Cys Lys Lys Gly Val Glu Asp Gln Ile Glu Tyr Thr Lys
            20                  25                  30

Lys Arg Ile Thr Ala Glu Val Lys Ala Gly Ala Lys Ala Pro Lys Asn
        35                  40                  45

Val Leu Val Leu Gly Cys Ser Asn Gly Tyr Gly Leu Ala Ser Arg Ile
    50                  55                  60
```

```
Thr Ala Ala Phe Gly Tyr Gly Ala Ala Thr Ile Gly Val Ser Phe Glu
65                  70                  75                  80

Lys Ala Gly Ser Glu Thr Lys Tyr Gly Thr Pro Gly Trp Tyr Asn Asn
                85                  90                  95

Leu Ala Phe Asp Glu Ala Ala Lys Arg Glu Gly Leu Tyr Ser Val Thr
            100                 105                 110

Ile Asp Gly Asp Ala Phe Ser Asp Glu Ile Lys Ala Gln Val Ile Glu
        115                 120                 125

Glu Ala Lys Lys Lys Gly Ile Lys Phe Asp Leu Ile Val Tyr Ser Leu
    130                 135                 140

Ala Ser Pro Val Arg Thr Asp Pro Asp Thr Gly Ile Met His Lys Ser
145                 150                 155                 160

Val Leu Lys Pro Phe Gly Lys Thr Phe Thr Gly Lys Thr Val Asp Pro
                165                 170                 175

Phe Thr Gly Glu Leu Lys Glu Ile Ser Ala Glu Pro Ala Asn Asp Glu
            180                 185                 190

Glu Ala Ala Ala Thr Val Lys Val Met Gly Gly Glu Asp Trp Glu Arg
        195                 200                 205

Trp Ile Lys Gln Leu Ser Lys Glu Gly Leu Leu Glu Glu Gly Cys Ile
    210                 215                 220

Thr Leu Ala Tyr Ser Tyr Ile Gly Pro Glu Ala Thr Gln Ala Leu Tyr
225                 230                 235                 240

Arg Lys Gly Thr Ile Gly Lys Ala Lys Glu His Leu Glu Ala Thr Ala
                245                 250                 255

His Arg Leu Asn Lys Glu Asn Pro Ser Ile Arg Ala Phe Val Ser Val
            260                 265                 270
```

```
Asn Lys Gly Leu Val Thr Arg Ala Ser Ala Val Ile Pro Val Ile Pro
        275                 280                 285


Leu Tyr Leu Ala Ser Leu Phe Lys Val Met Lys Glu Lys Gly Asn His
    290                 295                 300


Glu Gly Cys Ile Glu Gln Ile Thr Arg Leu Tyr Ala Glu Arg Leu Tyr
305                 310                 315                 320


Arg Lys Asp Gly Thr Ile Pro Val Asp Glu Glu Asn Arg Ile Arg Ile
                325                 330                 335


Asp Asp Trp Glu Leu Glu Glu Asp Val Gln Lys Ala Val Ser Ala Leu
            340                 345                 350


Met Glu Lys Val Thr Gly Glu Asn Ala Glu Ser Leu Thr Asp Leu Ala
        355                 360                 365


Gly Tyr Arg His Asp Phe Leu Ala Ser Asn Gly Phe Asp Val Glu Gly
    370                 375                 380


Ile Asn Tyr Glu Ala Glu Val Glu Arg Phe Asp Arg Ile
385                 390                 395


<210>  21
<211>  394
<212>  PRT
<213>  Ralstonia sp.

<400>  21

Met Thr Arg Glu Val Val Val Val Ser Gly Val Arg Thr Ala Ile Gly
1               5                   10                  15


Thr Phe Gly Gly Ser Leu Lys Asp Val Ala Pro Ala Glu Leu Gly Ala
            20                  25                  30


Leu Val Val Arg Glu Ala Leu Ala Arg Ala Gln Val Ser Gly Asp Asp
        35                  40                  45
```

```
Val Gly His Val Val Phe Gly Asn Val Ile Gln Thr Glu Pro Arg Asp
    50                  55                  60

Met Tyr Leu Gly Arg Val Ala Ala Val Asn Gly Gly Val Thr Ile Asn
65                  70                  75                  80

Ala Pro Ala Leu Thr Val Asn Arg Leu Cys Gly Ser Gly Leu Gln Ala
                85                  90                  95

Ile Val Ser Ala Ala Gln Thr Ile Leu Leu Gly Asp Thr Asp Val Ala
            100                 105                 110

Ile Gly Gly Gly Ala Glu Ser Met Ser Arg Ala Pro Tyr Leu Ala Pro
        115                 120                 125

Ala Ala Arg Trp Gly Ala Arg Met Gly Asp Ala Gly Leu Val Asp Met
    130                 135                 140

Met Leu Gly Ala Leu His Asp Pro Phe His Arg Ile His Met Gly Val
145                 150                 155                 160

Thr Ala Glu Asn Val Ala Lys Glu Tyr Asp Ile Ser Arg Ala Gln Gln
                165                 170                 175

Asp Glu Ala Ala Leu Glu Ser His Arg Arg Ala Ser Ala Ala Ile Lys
            180                 185                 190

Ala Gly Tyr Phe Lys Asp Gln Ile Val Pro Val Val Ser Lys Gly Arg
        195                 200                 205

Lys Gly Asp Val Thr Phe Asp Thr Asp Glu His Val Arg His Asp Ala
    210                 215                 220

Thr Ile Asp Asp Met Thr Lys Leu Arg Pro Val Phe Val Lys Glu Asn
225                 230                 235                 240

Gly Thr Val Thr Ala Gly Asn Ala Ser Gly Leu Asn Asp Ala Ala Ala
                245                 250                 255
```

Ala Val Val Met Met Glu Arg Ala Glu Ala Glu Arg Arg Gly Leu Lys
260 265 270

Pro Leu Ala Arg Leu Val Ser Tyr Gly His Ala Gly Val Asp Pro Lys
275 280 285

Ala Met Gly Ile Gly Pro Val Pro Ala Thr Lys Ile Ala Leu Glu Arg
290 295 300

Ala Gly Leu Gln Val Ser Asp Leu Asp Val Ile Glu Ala Asn Glu Ala
305 310 315 320

Phe Ala Ala Gln Ala Cys Ala Val Thr Lys Ala Leu Gly Leu Asp Pro
325 330 335

Ala Lys Val Asn Pro Asn Gly Ser Gly Ile Ser Leu Gly His Pro Ile
340 345 350

Gly Ala Thr Gly Ala Leu Ile Thr Val Lys Ala Leu His Glu Leu Asn
355 360 365

Arg Val Gln Gly Arg Tyr Ala Leu Val Thr Met Cys Ile Gly Gly Gly
370 375 380

Gln Gly Ile Ala Ala Ile Phe Glu Arg Ile
385 390

<210> 22
<211> 813
<212> PRT
<213> Enterococcus sp.

<400> 22

Met Lys Glu Val Val Met Ile Asp Ala Ala Arg Thr Pro Ile Gly Lys
1 5 10 15

Tyr Arg Gly Ser Leu Ser Pro Phe Thr Ala Val Glu Leu Gly Thr Leu
20 25 30

Val Thr Lys Gly Leu Leu Asp Lys Thr Lys Leu Lys Lys Asp Lys Ile

```
        35                  40                  45

Asp Gln Val Ile Phe Gly Asn Val Leu Gln Ala Gly Asn Gly Gln Asn
    50                  55                  60

Val Ala Arg Gln Ile Ala Leu Asn Ser Gly Leu Pro Val Asp Val Pro
65                  70                  75                  80

Ala Met Thr Ile Asn Glu Val Cys Gly Ser Gly Met Lys Ala Val Ile
                85                  90                  95

Leu Ala Arg Gln Leu Ile Gln Leu Gly Glu Ala Glu Leu Val Ile Ala
            100                 105                 110

Gly Gly Thr Glu Ser Met Ser Gln Ala Pro Met Leu Lys Pro Tyr Gln
        115                 120                 125

Ser Glu Thr Asn Glu Tyr Gly Glu Pro Ile Ser Ser Met Val Asn Asp
    130                 135                 140

Gly Leu Thr Asp Ala Phe Ser Asn Ala His Met Gly Leu Thr Ala Glu
145                 150                 155                 160

Lys Val Ala Thr Gln Phe Ser Val Ser Arg Glu Glu Gln Asp Arg Tyr
                165                 170                 175

Ala Leu Ser Ser Gln Leu Lys Ala Ala His Ala Val Glu Ala Gly Val
            180                 185                 190

Phe Ser Glu Glu Ile Ile Pro Val Lys Ile Ser Asp Glu Asp Val Leu
        195                 200                 205

Ser Glu Asp Glu Ala Val Arg Gly Asn Ser Thr Leu Glu Lys Leu Gly
    210                 215                 220

Thr Leu Arg Thr Val Phe Ser Glu Glu Gly Thr Val Thr Ala Gly Asn
225                 230                 235                 240

Ala Ser Pro Leu Asn Asp Gly Ala Ser Val Val Ile Leu Ala Ser Lys
```

245 250 255

Glu Tyr Ala Glu Asn Asn Asn Leu Pro Tyr Leu Ala Thr Ile Lys Glu
260 265 270

Val Ala Glu Val Gly Ile Asp Pro Ser Ile Met Gly Ile Ala Pro Ile
275 280 285

Lys Ala Ile Gln Lys Leu Thr Asp Arg Ser Gly Met Asn Leu Ser Thr
290 295 300

Ile Asp Leu Phe Glu Ile Asn Glu Ala Phe Ala Ala Ser Ser Ile Val
305 310 315 320

Val Ser Gln Glu Leu Gln Leu Asp Glu Glu Lys Val Asn Ile Tyr Gly
325 330 335

Gly Ala Ile Ala Leu Gly His Pro Ile Gly Ala Ser Gly Ala Arg Ile
340 345 350

Leu Thr Thr Leu Ala Tyr Gly Leu Leu Arg Glu Gln Lys Arg Tyr Gly
355 360 365

Ile Ala Ser Leu Cys Ile Gly Gly Gly Leu Gly Leu Ala Val Leu Leu
370 375 380

Glu Ala Asn Met Glu Gln Thr His Lys Asp Val Gln Lys Lys Lys Phe
385 390 395 400

Tyr Gln Leu Thr Pro Ser Glu Arg Arg Ser Gln Leu Ile Glu Lys Asn
405 410 415

Val Leu Thr Gln Glu Thr Ala Leu Ile Phe Gln Glu Gln Thr Leu Ser
420 425 430

Glu Glu Leu Ser Asp His Met Ile Glu Asn Gln Val Ser Glu Val Glu
435 440 445

Ile Pro Met Gly Ile Ala Gln Asn Phe Gln Ile Asn Gly Lys Lys Lys

```
    450                 455                 460

Trp Ile Pro Met Ala Thr Glu Glu Pro Ser Val Ile Ala Ala Ala Ser
465                 470                 475                 480

Asn Gly Ala Lys Ile Cys Gly Asn Ile Cys Ala Glu Thr Pro Gln Arg
                485                 490                 495

Leu Met Arg Gly Gln Ile Val Leu Ser Gly Lys Ser Glu Tyr Gln Ala
            500                 505                 510

Val Ile Asn Ala Val Asn His Arg Lys Glu Glu Leu Ile Leu Cys Ala
        515                 520                 525

Asn Glu Ser Tyr Pro Ser Ile Val Lys Arg Gly Gly Gly Val Gln Asp
    530                 535                 540

Ile Ser Thr Arg Glu Phe Met Gly Ser Phe His Ala Tyr Leu Ser Ile
545                 550                 555                 560

Asp Phe Leu Val Asp Val Lys Asp Ala Met Gly Ala Asn Met Ile Asn
                565                 570                 575

Ser Ile Leu Glu Ser Val Ala Asn Lys Leu Arg Glu Trp Phe Pro Glu
            580                 585                 590

Glu Glu Ile Leu Phe Ser Ile Leu Ser Asn Phe Ala Thr Glu Ser Leu
        595                 600                 605

Ala Ser Ala Cys Cys Glu Ile Pro Phe Glu Arg Leu Gly Arg Asn Lys
    610                 615                 620

Glu Ile Gly Glu Gln Ile Ala Lys Lys Ile Gln Gln Ala Gly Glu Tyr
625                 630                 635                 640

Ala Lys Leu Asp Pro Tyr Arg Ala Ala Thr His Asn Lys Gly Ile Met
                645                 650                 655

Asn Gly Ile Glu Ala Val Val Ala Ala Thr Gly Asn Asp Thr Arg Ala
```

660 665 670

Val Ser Ala Ser Ile His Ala Tyr Ala Ala Arg Asn Gly Leu Tyr Gln
675 680 685

Gly Leu Thr Asp Trp Gln Ile Lys Gly Asp Lys Leu Val Gly Lys Leu
690 695 700

Thr Val Pro Leu Ala Val Ala Thr Val Gly Gly Ala Ser Asn Ile Leu
705 710 715 720

Pro Lys Ala Lys Ala Ser Leu Ala Met Leu Asp Ile Asp Ser Ala Lys
725 730 735

Glu Leu Ala Gln Val Ile Ala Ala Val Gly Leu Ala Gln Asn Leu Ala
740 745 750

Ala Leu Arg Ala Leu Val Thr Glu Gly Ile Gln Lys Gly His Met Gly
755 760 765

Leu Gln Ala Arg Ser Leu Ala Ile Ser Ile Gly Ala Ile Gly Glu Glu
770 775 780

Ile Glu Gln Val Ala Lys Lys Leu Arg Glu Ala Glu Lys Met Asn Gln
785 790 795 800

Gln Thr Ala Ile Gln Ile Leu Glu Lys Ile Arg Glu Lys
805 810

<210> 23
<211> 389
<212> PRT
<213> Lactobacillus plantarum

<400> 23

Met Lys Ile Gly Ile Asp Lys Leu His Phe Ala Thr Ser His Leu Tyr
1 5 10 15

Val Asp Met Ala Glu Leu Ala Thr Ala Arg Gln Ala Glu Pro Asp Lys
20 25 30

```
Tyr Leu Ile Gly Ile Gly Gln Ser Lys Met Ala Val Ile Pro Pro Ser
        35                  40                  45

Gln Asp Val Val Thr Leu Ala Ala Asn Ala Ala Ala Pro Met Leu Thr
    50                  55                  60

Ala Thr Asp Ile Ala Ala Ile Asp Leu Leu Val Val Gly Thr Glu Ser
65                  70                  75                  80

Gly Ile Asp Asn Ser Lys Ala Ser Ala Ile Tyr Val Ala Lys Leu Leu
                85                  90                  95

Gly Leu Ser Gln Arg Val Arg Thr Ile Glu Met Lys Glu Ala Cys Tyr
            100                 105                 110

Ala Ala Thr Ala Gly Val Gln Leu Ala Gln Asp His Val Arg Val His
        115                 120                 125

Pro Asp Lys Lys Ala Leu Val Ile Gly Ser Asp Val Ala Arg Tyr Gly
    130                 135                 140

Leu Asn Thr Pro Gly Glu Pro Thr Gln Gly Gly Gly Ala Val Ala Met
145                 150                 155                 160

Leu Ile Ser Ala Asp Pro Lys Val Leu Val Leu Gly Thr Glu Ser Ser
                165                 170                 175

Leu Leu Ser Glu Asp Val Met Asp Phe Trp Arg Pro Leu Tyr His Thr
            180                 185                 190

Glu Ala Leu Val Asp Gly Lys Tyr Ser Ser Asn Ile Tyr Ile Asp Tyr
        195                 200                 205

Phe Gln Asp Val Phe Lys Asn Tyr Leu Gln Thr Thr Gln Thr Ser Pro
    210                 215                 220

Asp Thr Leu Thr Ala Leu Val Phe His Leu Pro Tyr Thr Lys Met Gly
225                 230                 235                 240
```

```
Leu Lys Ala Leu Arg Ser Val Leu Pro Leu Val Asp Ala Glu Lys Gln
                245                 250                 255

Ala Gln Trp Leu Ala His Phe Glu His Ala Arg Gln Leu Asn Arg Gln
            260                 265                 270

Val Gly Asn Leu Tyr Thr Gly Ser Leu Tyr Leu Ser Leu Leu Ser Gln
        275                 280                 285

Leu Leu Thr Asp Pro Gln Leu Gln Pro Gly Asn Arg Leu Gly Leu Phe
    290                 295                 300

Ser Tyr Gly Ser Gly Ala Glu Gly Glu Phe Tyr Thr Gly Val Ile Gln
305                 310                 315                 320

Pro Asp Tyr Gln Thr Gly Leu Asp His Gly Leu Pro Gln Arg Leu Ala
                325                 330                 335

Arg Arg Arg Arg Val Ser Val Ala Glu Tyr Glu Ala Leu Phe Ser His
            340                 345                 350

Gln Leu Gln Trp Arg Ala Asp Asp Gln Ser Val Ser Tyr Ala Asp Asp
        355                 360                 365

Pro His Arg Phe Val Leu Thr Gly Gln Lys Asn Glu Gln Arg Gln Tyr
    370                 375                 380

Leu Asp Gln Gln Val
385


<210>  24
<211>  491
<212>  PRT
<213>  Saccharomyces cerevisiae

<400>  24

Met Lys Leu Ser Thr Lys Leu Cys Trp Cys Gly Ile Lys Gly Arg Leu
1               5                   10                  15
```

```
Arg Pro Gln Lys Gln Gln Gln Leu His Asn Thr Asn Leu Gln Met Thr
            20                  25                  30

Glu Leu Lys Lys Gln Lys Thr Ala Glu Gln Lys Thr Arg Pro Gln Asn
        35                  40                  45

Val Gly Ile Lys Gly Ile Gln Ile Tyr Ile Pro Thr Gln Cys Val Asn
    50                  55                  60

Gln Ser Glu Leu Glu Lys Phe Asp Gly Val Ser Gln Gly Lys Tyr Thr
65                  70                  75                  80

Ile Gly Leu Gly Gln Thr Asn Met Ser Phe Val Asn Asp Arg Glu Asp
                85                  90                  95

Ile Tyr Ser Met Ser Leu Thr Val Leu Ser Lys Leu Ile Lys Ser Tyr
            100                 105                 110

Asn Ile Asp Thr Asn Lys Ile Gly Arg Leu Glu Val Gly Thr Glu Thr
        115                 120                 125

Leu Ile Asp Lys Ser Lys Ser Val Lys Ser Val Leu Met Gln Leu Phe
    130                 135                 140

Gly Glu Asn Thr Asp Val Glu Gly Ile Asp Thr Leu Asn Ala Cys Tyr
145                 150                 155                 160

Gly Gly Thr Asn Ala Leu Phe Asn Ser Leu Asn Trp Ile Glu Ser Asn
                165                 170                 175

Ala Trp Asp Gly Arg Asp Ala Ile Val Val Cys Gly Asp Ile Ala Ile
            180                 185                 190

Tyr Asp Lys Gly Ala Ala Arg Pro Thr Gly Gly Ala Gly Thr Val Ala
        195                 200                 205

Met Trp Ile Gly Pro Asp Ala Pro Ile Val Phe Asp Ser Val Arg Ala
    210                 215                 220
```

```
Ser Tyr Met Glu His Ala Tyr Asp Phe Tyr Lys Pro Asp Phe Thr Ser
225                 230                 235                 240

Glu Tyr Pro Tyr Val Asp Gly His Phe Ser Leu Thr Cys Tyr Val Lys
                245                 250                 255

Ala Leu Asp Gln Val Tyr Lys Ser Tyr Ser Lys Lys Ala Ile Ser Lys
            260                 265                 270

Gly Leu Val Ser Asp Pro Ala Gly Ser Asp Ala Leu Asn Val Leu Lys
        275                 280                 285

Tyr Phe Asp Tyr Asn Val Phe His Val Pro Thr Cys Lys Leu Val Thr
    290                 295                 300

Lys Ser Tyr Gly Arg Leu Leu Tyr Asn Asp Phe Arg Ala Asn Pro Gln
305                 310                 315                 320

Leu Phe Pro Glu Val Asp Ala Glu Leu Ala Thr Arg Asp Tyr Asp Glu
                325                 330                 335

Ser Leu Thr Asp Lys Asn Ile Glu Lys Thr Phe Val Asn Val Ala Lys
            340                 345                 350

Pro Phe His Lys Glu Arg Val Ala Gln Ser Leu Ile Val Pro Thr Asn
        355                 360                 365

Thr Gly Asn Met Tyr Thr Ala Ser Val Tyr Ala Ala Phe Ala Ser Leu
    370                 375                 380

Leu Asn Tyr Val Gly Ser Asp Asp Leu Gln Gly Lys Arg Val Gly Leu
385                 390                 395                 400

Phe Ser Tyr Gly Ser Gly Leu Ala Ala Ser Leu Tyr Ser Cys Lys Ile
                405                 410                 415

Val Gly Asp Val Gln His Ile Ile Lys Glu Leu Asp Ile Thr Asn Lys
            420                 425                 430
```

```
Leu Ala Lys Arg Ile Thr Glu Thr Pro Lys Asp Tyr Glu Ala Ala Ile
        435                 440                 445


Glu Leu Arg Glu Asn Ala His Leu Lys Lys Asn Phe Lys Pro Gln Gly
    450                 455                 460


Ser Ile Glu His Leu Gln Ser Gly Val Tyr Tyr Leu Thr Asn Ile Asp
465                 470                 475                 480


Asp Lys Phe Arg Arg Ser Tyr Asp Val Lys Lys
                485                 490


<210>  25
<211>  398
<212>  PRT
<213>  Saccharomyces cerevisiae

<400>  25

Met Ser Gln Asn Val Tyr Ile Val Ser Thr Ala Arg Thr Pro Ile Gly
1               5                   10                  15


Ser Phe Gln Gly Ser Leu Ser Ser Lys Thr Ala Val Glu Leu Gly Ala
            20                  25                  30


Val Ala Leu Lys Gly Ala Leu Ala Lys Val Pro Glu Leu Asp Ala Ser
        35                  40                  45


Lys Asp Phe Asp Glu Ile Ile Phe Gly Asn Val Leu Ser Ala Asn Leu
    50                  55                  60


Gly Gln Ala Pro Ala Arg Gln Val Ala Leu Ala Ala Gly Leu Ser Asn
65                  70                  75                  80


His Ile Val Ala Ser Thr Val Asn Lys Val Cys Ala Ser Ala Met Lys
                85                  90                  95


Ala Ile Ile Leu Gly Ala Gln Ser Ile Lys Cys Gly Asn Ala Asp Val
            100                 105                 110
```

```
Val Val Ala Gly Gly Cys Glu Ser Met Thr Asn Ala Pro Tyr Tyr Met
        115                 120                 125

Pro Ala Ala Arg Ala Gly Ala Lys Phe Gly Gln Thr Val Leu Val Asp
    130                 135                 140

Gly Val Glu Arg Asp Gly Leu Asn Asp Ala Tyr Asp Gly Leu Ala Met
145                 150                 155                 160

Gly Val His Ala Glu Lys Cys Ala Arg Asp Trp Asp Ile Thr Arg Glu
                165                 170                 175

Gln Gln Asp Asn Phe Ala Ile Glu Ser Tyr Gln Lys Ser Gln Lys Ser
            180                 185                 190

Gln Lys Glu Gly Lys Phe Asp Asn Glu Ile Val Pro Val Thr Ile Lys
        195                 200                 205

Gly Phe Arg Gly Lys Pro Asp Thr Gln Val Thr Lys Asp Glu Glu Pro
    210                 215                 220

Ala Arg Leu His Val Glu Lys Leu Arg Ser Ala Arg Thr Val Phe Gln
225                 230                 235                 240

Lys Glu Asn Gly Thr Val Thr Ala Ala Asn Ala Ser Pro Ile Asn Asp
                245                 250                 255

Gly Ala Ala Ala Val Ile Leu Val Ser Glu Lys Val Leu Lys Glu Lys
            260                 265                 270

Asn Leu Lys Pro Leu Ala Ile Ile Lys Gly Trp Gly Glu Ala Ala His
        275                 280                 285

Gln Pro Ala Asp Phe Thr Trp Ala Pro Ser Leu Ala Val Pro Lys Ala
    290                 295                 300

Leu Lys His Ala Gly Ile Glu Asp Ile Asn Ser Val Asp Tyr Phe Glu
305                 310                 315                 320
```

Phe Asn Glu Ala Phe Ser Val Val Gly Leu Val Asn Thr Lys Ile Leu
325 330 335

Lys Leu Asp Pro Ser Lys Val Asn Val Tyr Gly Gly Ala Val Ala Leu
340 345 350

Gly His Pro Leu Gly Cys Ser Gly Ala Arg Val Val Val Thr Leu Leu
355 360 365

Ser Ile Leu Gln Gln Glu Gly Gly Lys Ile Gly Val Ala Ala Ile Cys
370 375 380

Asn Gly Gly Gly Gly Ala Ser Ser Ile Val Ile Glu Lys Ile
385 390 395

<210> 26
<211> 247
<212> PRT
<213> Bacteroides sp.

<400> 26

Met Ser Asp Asp Lys Lys Ile Gly Ser Tyr Lys Phe Ile Ala Glu Pro
1 5 10 15

Phe His Val Asp Phe Asn Gly Arg Leu Thr Met Gly Val Leu Gly Asn
20 25 30

His Leu Leu Asn Cys Ala Gly Phe His Ala Ser Glu Arg Gly Phe Gly
35 40 45

Ile Ala Thr Leu Asn Glu Asp Asn Tyr Thr Trp Val Leu Ser Arg Leu
50 55 60

Ala Ile Asp Leu Glu Glu Met Pro Tyr Gln Tyr Glu Glu Phe Thr Val
65 70 75 80

Gln Thr Trp Val Glu Asn Val Tyr Arg Leu Phe Thr Asp Arg Asn Phe
85 90 95

Ala Ile Ile Asp Lys Asp Gly Lys Lys Ile Gly Tyr Ala Arg Ser Val

100 105 110

Trp Ala Met Ile Asn Leu Asn Thr Arg Lys Pro Ala Asp Leu Leu Thr
115 120 125

Leu His Gly Gly Ser Ile Val Asp Tyr Val Cys Asp Glu Pro Cys Pro
130 135 140

Ile Glu Lys Pro Ser Arg Ile Lys Val Ala Thr Asp Gln Pro Cys Ala
145 150 155 160

Lys Leu Thr Ala Lys Tyr Ser Asp Ile Asp Ile Asn Gly His Val Asn
165 170 175

Ser Ile Arg Tyr Ile Glu His Ile Leu Asp Leu Phe Pro Ile Asp Leu
180 185 190

Tyr Lys Ser Lys Arg Ile Gln Arg Phe Glu Met Ala Tyr Val Ala Glu
195 200 205

Ser Tyr Tyr Gly Asp Glu Leu Ser Phe Phe Glu Glu Glu Val Ser Glu
210 215 220

Asn Glu Tyr His Val Glu Ile Lys Lys Asn Gly Ser Glu Val Val Cys
225 230 235 240

Arg Ala Lys Val Lys Phe Val
245

<210> 27
<211> 247
<212> PRT
<213> Bacteroides thetaiotaomicron

<400> 27

Met Ser Glu Glu Asn Lys Ile Gly Thr Tyr Gln Phe Val Ala Glu Pro
1 5 10 15

Phe His Val Asp Phe Asn Gly Arg Leu Thr Met Gly Val Leu Gly Asn
20 25 30

```
His Leu Leu Asn Cys Ala Gly Phe His Ala Ser Asp Arg Gly Phe Gly
        35                  40                  45

Ile Ala Thr Leu Asn Glu Asp Asn Tyr Thr Trp Val Leu Ser Arg Leu
    50                  55                  60

Ala Ile Glu Leu Asp Glu Met Pro Tyr Gln Tyr Glu Lys Phe Ser Val
65                  70                  75                  80

Gln Thr Trp Val Glu Asn Val Tyr Arg Leu Phe Thr Asp Arg Asn Phe
                85                  90                  95

Ala Val Ile Asp Lys Asp Gly Lys Lys Ile Gly Tyr Ala Arg Ser Val
            100                 105                 110

Trp Ala Met Ile Asn Leu Asn Thr Arg Lys Pro Ala Asp Leu Leu Ala
        115                 120                 125

Leu His Gly Gly Ser Ile Val Asp Tyr Ile Cys Asp Glu Pro Cys Pro
    130                 135                 140

Ile Glu Lys Pro Ser Arg Ile Lys Val Thr Ser Asn Gln Pro Val Ala
145                 150                 155                 160

Thr Leu Thr Ala Lys Tyr Ser Asp Ile Asp Ile Asn Gly His Val Asn
                165                 170                 175

Ser Ile Arg Tyr Ile Glu His Ile Leu Asp Leu Phe Pro Ile Glu Leu
            180                 185                 190

Tyr Gln Thr Lys Arg Ile Arg Arg Phe Glu Met Ala Tyr Val Ala Glu
        195                 200                 205

Ser Tyr Phe Gly Asp Glu Leu Ser Phe Phe Cys Asp Glu Val Ser Glu
    210                 215                 220

Asn Glu Phe His Val Glu Val Lys Lys Asn Gly Ser Glu Val Val Cys
225                 230                 235                 240
```

Arg Ser Lys Val Ile Phe Glu
245

<210> 28
<211> 238
<212> PRT
<213> Bryantella formatexigen

<400> 28

Met Ile Tyr Met Ala Tyr Gln Tyr Arg Ser Arg Ile Arg Tyr Ser Glu
1 5 10 15

Ile Gly Glu Asp Lys Lys Leu Thr Leu Pro Gly Leu Val Asn Tyr Phe
20 25 30

Gln Asp Cys Ser Thr Phe Gln Ser Glu Ala Leu Gly Ile Gly Leu Asp
35 40 45

Thr Leu Gly Ala Arg Gln Arg Ala Trp Leu Leu Ala Ser Trp Lys Ile
50 55 60

Val Ile Asp Arg Leu Pro Arg Leu Gly Glu Glu Val Val Thr Glu Thr
65 70 75 80

Trp Pro Tyr Gly Phe Lys Gly Phe Gln Gly Asn Arg Asn Phe Arg Met
85 90 95

Leu Asp Gln Glu Gly His Thr Leu Ala Ala Ala Ala Ser Val Trp Ile
100 105 110

Tyr Leu Asn Val Glu Ser Gly His Pro Cys Arg Ile Asp Gly Asp Val
115 120 125

Leu Glu Ala Tyr Glu Leu Glu Glu Glu Leu Pro Leu Gly Pro Phe Ser
130 135 140

Arg Lys Ile Pro Val Pro Glu Glu Ser Thr Glu Arg Asp Ser Phe Leu
145 150 155 160

```
Val Met Arg Ser His Leu Asp Thr Asn His His Val Asn Asn Gly Gln
                165                 170                 175


Tyr Ile Leu Met Ala Glu Glu Tyr Leu Pro Glu Gly Phe Lys Val Lys
            180                 185                 190


Gln Ile Arg Val Glu Tyr Arg Lys Ala Ala Val Leu His Asp Thr Ile
        195                 200                 205


Val Pro Phe Val Cys Thr Glu Pro Gln Arg Cys Thr Val Ser Leu Cys
    210                 215                 220


Gly Ser Asp Glu Lys Pro Phe Ala Val Val Glu Phe Ser Glu
225                 230                 235


<210>  29
<211>  246
<212>  PRT
<213>  Lactobacillus brevis

<400>  29

Met Ala Ala Asn Glu Phe Ser Glu Thr His Arg Val Val Tyr Tyr Glu
1               5                   10                  15


Ala Asp Asp Thr Gly Gln Leu Thr Leu Ala Met Leu Ile Asn Leu Phe
            20                  25                  30


Val Leu Val Ser Glu Asp Gln Asn Asp Ala Leu Gly Leu Ser Thr Ala
        35                  40                  45


Phe Val Gln Ser His Gly Val Gly Trp Val Val Thr Gln Tyr His Leu
    50                  55                  60


His Ile Asp Glu Leu Pro Arg Thr Gly Ala Gln Val Thr Ile Lys Thr
65                  70                  75                  80


Arg Ala Thr Ala Tyr Asn Arg Tyr Phe Ala Tyr Arg Glu Tyr Trp Leu
                85                  90                  95
```

```
Leu Asp Asp Ala Gly Gln Val Leu Ala Tyr Gly Glu Gly Ile Trp Val
            100                 105                 110

Thr Met Ser Tyr Ala Thr Arg Lys Ile Thr Thr Ile Pro Ala Glu Val
        115                 120                 125

Met Ala Pro Tyr His Ser Glu Glu Gln Thr Arg Leu Pro Arg Leu Pro
    130                 135                 140

Arg Pro Asp His Phe Asp Glu Ala Val Asn Gln Thr Leu Lys Pro Tyr
145                 150                 155                 160

Thr Val Arg Tyr Phe Asp Ile Asp Gly Asn Gly His Val Asn Asn Ala
                165                 170                 175

His Tyr Phe Asp Trp Met Leu Asp Val Leu Pro Ala Thr Phe Leu Arg
            180                 185                 190

Ala His His Pro Thr Asp Val Lys Ile Arg Phe Glu Asn Glu Val Gln
        195                 200                 205

Tyr Gly His Gln Val Thr Ser Glu Leu Ser Gln Ala Ala Ala Leu Thr
    210                 215                 220

Thr Gln His Met Ile Lys Val Gly Asp Leu Thr Ala Val Lys Ala Thr
225                 230                 235                 240

Ile Gln Trp Asp Asn Arg
                245
```

<210> 30
<211> 261
<212> PRT
<213> Lactobacillus plantarum

<400> 30

```
Met Ala Thr Leu Gly Ala Asn Ala Ser Leu Tyr Ser Glu Gln His Arg
1               5                   10                  15

Ile Thr Tyr Tyr Glu Cys Asp Arg Thr Gly Arg Ala Thr Leu Thr Thr
```

```
            20                  25                  30

Leu Ile Asp Ile Ala Val Leu Ala Ser Glu Asp Gln Ser Asp Ala Leu
        35                  40                  45

Gly Leu Thr Thr Glu Met Val Gln Ser His Gly Val Gly Trp Val Val
    50                  55                  60

Thr Gln Tyr Ala Ile Asp Ile Thr Arg Met Pro Arg Gln Asp Glu Val
65                  70                  75                  80

Val Thr Ile Ala Val Arg Gly Ser Ala Tyr Asn Pro Tyr Phe Ala Tyr
                85                  90                  95

Arg Glu Phe Trp Ile Arg Asp Ala Asp Gly Gln Gln Leu Ala Tyr Ile
            100                 105                 110

Thr Ser Ile Trp Val Met Met Ser Gln Thr Thr Arg Arg Ile Val Lys
        115                 120                 125

Ile Leu Pro Glu Leu Val Ala Pro Tyr Gln Ser Glu Val Val Lys Arg
    130                 135                 140

Ile Pro Arg Leu Pro Arg Pro Ile Ser Phe Glu Ala Thr Asp Thr Thr
145                 150                 155                 160

Ile Thr Lys Pro Tyr His Val Arg Phe Phe Asp Ile Asp Pro Asn Arg
                165                 170                 175

His Val Asn Asn Ala His Tyr Phe Asp Trp Leu Val Asp Thr Leu Pro
            180                 185                 190

Ala Thr Phe Leu Leu Gln His Asp Leu Val His Val Asp Val Arg Tyr
        195                 200                 205

Glu Asn Glu Val Lys Tyr Gly Gln Thr Val Thr Ala His Ala Asn Ile
    210                 215                 220

Leu Pro Ser Glu Val Ala Asp Gln Val Thr Thr Ser His Leu Ile Glu
```

225 230 235 240

Val Asp Asp Glu Lys Cys Cys Glu Val Thr Ile Gln Trp Arg Thr Leu
245 250 255

Pro Glu Pro Ile Gln
260

<210> 31
<211> 250
<212> PRT
<213> Streptococcus dysgalactiae

<400> 31

Met Gly Leu Ser Tyr Arg Glu Asp Ile Lys Leu Pro Phe Glu Leu Cys
1 5 10 15

Asp Val Lys Ser Asp Ile Lys Phe Pro Leu Leu Leu Asp Tyr Cys Leu
20 25 30

Thr Val Ser Gly Arg Gln Ser Ala Gln Leu Gly Arg Ser Asn Asp Tyr
35 40 45

Leu Leu Glu Gln Tyr Gly Leu Ile Trp Ile Val Thr Asp Tyr Glu Ala
50 55 60

Thr Ile His Arg Leu Pro His Phe Gln Glu Thr Ile Thr Ile Glu Thr
65 70 75 80

Lys Ala Leu Ser Tyr Asn Lys Phe Phe Cys Tyr Arg Gln Phe Tyr Ile
85 90 95

Tyr Asp Gln Glu Gly Gly Leu Leu Val Asp Ile Leu Ala Tyr Phe Ala
100 105 110

Leu Leu Asn Pro Asp Thr Arg Lys Val Ala Thr Ile Pro Glu Asp Leu
115 120 125

Val Ala Pro Phe Glu Thr Asp Phe Val Lys Lys Leu His Arg Val Pro
130 135 140

```
Lys Met Pro Leu Leu Glu Gln Ser Ile Asp Arg Asp Tyr Tyr Val Arg
145                 150                 155                 160


Tyr Phe Asp Ile Asp Met Asn Gly His Val Asn Asn Ser Lys Tyr Leu
                165                 170                 175


Asp Trp Met Tyr Asp Val Leu Gly Cys Glu Phe Leu Lys Thr His Gln
            180                 185                 190


Pro Leu Lys Met Thr Leu Lys Tyr Val Lys Glu Val Ser Pro Gly Gly
        195                 200                 205


Gln Ile Thr Ser Ser Tyr His Leu Asp Gln Leu Thr Ser Tyr His Gln
    210                 215                 220


Ile Thr Ser Asp Gly Gln Leu Asn Ala Gln Ala Met Ile Glu Trp Arg
225                 230                 235                 240


Ala Ile Lys Gln Thr Glu Ser Glu Ile Asp
                245                 250


<210>  32
<211>  620
<212>  PRT
<213>  Escherichia coli

<400>  32

Met Ser Phe Asp Ile Ala Lys Tyr Pro Thr Leu Ala Leu Val Asp Ser
1               5                   10                  15


Thr Gln Glu Leu Arg Leu Leu Pro Lys Glu Ser Leu Pro Lys Leu Cys
            20                  25                  30


Asp Glu Leu Arg Arg Tyr Leu Leu Asp Ser Val Ser Arg Ser Ser Gly
        35                  40                  45


His Phe Ala Ser Gly Leu Gly Thr Val Glu Leu Thr Val Ala Leu His
    50                  55                  60
```

```
Tyr Val Tyr Asn Thr Pro Phe Asp Gln Leu Ile Trp Asp Val Gly His
65                  70                  75                  80

Gln Ala Tyr Pro His Lys Ile Leu Thr Gly Arg Arg Asp Lys Ile Gly
                85                  90                  95

Thr Ile Arg Gln Lys Gly Gly Leu His Pro Phe Pro Trp Arg Gly Glu
            100                 105                 110

Ser Glu Tyr Asp Val Leu Ser Val Gly His Ser Ser Thr Ser Ile Ser
        115                 120                 125

Ala Gly Ile Gly Ile Ala Val Ala Ala Glu Lys Glu Gly Lys Asn Arg
    130                 135                 140

Arg Thr Val Cys Val Ile Gly Asp Gly Ala Ile Thr Ala Gly Met Ala
145                 150                 155                 160

Phe Glu Ala Met Asn His Ala Gly Asp Ile Arg Pro Asp Met Leu Val
                165                 170                 175

Ile Leu Asn Asp Asn Glu Met Ser Ile Ser Glu Asn Val Gly Ala Leu
            180                 185                 190

Asn Asn His Leu Ala Gln Leu Leu Ser Gly Lys Leu Tyr Ser Ser Leu
        195                 200                 205

Arg Glu Gly Gly Lys Lys Val Phe Ser Gly Val Pro Pro Ile Lys Glu
    210                 215                 220

Leu Leu Lys Arg Thr Glu Glu His Ile Lys Gly Met Val Val Pro Gly
225                 230                 235                 240

Thr Leu Phe Glu Glu Leu Gly Phe Asn Tyr Ile Gly Pro Val Asp Gly
                245                 250                 255

His Asp Val Leu Gly Leu Ile Thr Thr Leu Lys Asn Met Arg Asp Leu
            260                 265                 270
```

```
Lys Gly Pro Gln Phe Leu His Ile Met Thr Lys Lys Gly Arg Gly Tyr
        275                 280                 285

Glu Pro Ala Glu Lys Asp Pro Ile Thr Phe His Ala Val Pro Lys Phe
    290                 295                 300

Asp Pro Ser Ser Gly Cys Leu Pro Lys Ser Ser Gly Gly Leu Pro Ser
305                 310                 315                 320

Tyr Ser Lys Ile Phe Gly Asp Trp Leu Cys Glu Thr Ala Ala Lys Asp
                325                 330                 335

Asn Lys Leu Met Ala Ile Thr Pro Ala Met Arg Glu Gly Ser Gly Met
            340                 345                 350

Val Glu Phe Ser Arg Lys Phe Pro Asp Arg Tyr Phe Asp Val Ala Ile
        355                 360                 365

Ala Glu Gln His Ala Val Thr Phe Ala Ala Gly Leu Ala Ile Gly Gly
    370                 375                 380

Tyr Lys Pro Ile Val Ala Ile Tyr Ser Thr Phe Leu Gln Arg Ala Tyr
385                 390                 395                 400

Asp Gln Val Leu His Asp Val Ala Ile Gln Lys Leu Pro Val Leu Phe
                405                 410                 415

Ala Ile Asp Arg Ala Gly Ile Val Gly Ala Asp Gly Gln Thr His Gln
            420                 425                 430

Gly Ala Phe Asp Leu Ser Tyr Leu Arg Cys Ile Pro Glu Met Val Ile
        435                 440                 445

Met Thr Pro Ser Asp Glu Asn Glu Cys Arg Gln Met Leu Tyr Thr Gly
    450                 455                 460

Tyr His Tyr Asn Asp Gly Pro Ser Ala Val Arg Tyr Pro Arg Gly Asn
465                 470                 475                 480
```

```
Ala Val Gly Val Glu Leu Thr Pro Leu Glu Lys Leu Pro Ile Gly Lys
                485                 490                 495


Gly Ile Val Lys Arg Arg Gly Glu Lys Leu Ala Ile Leu Asn Phe Gly
            500                 505                 510


Thr Leu Met Pro Glu Ala Ala Lys Val Ala Glu Ser Leu Asn Ala Thr
        515                 520                 525


Leu Val Asp Met Arg Phe Val Lys Pro Leu Asp Glu Ala Leu Ile Leu
    530                 535                 540


Glu Met Ala Ala Ser His Glu Ala Leu Val Thr Val Glu Glu Asn Ala
545                 550                 555                 560


Ile Met Gly Gly Ala Gly Ser Gly Val Asn Glu Val Leu Met Ala His
                565                 570                 575


Arg Lys Pro Val Pro Val Leu Asn Ile Gly Leu Pro Asp Phe Phe Ile
            580                 585                 590


Pro Gln Gly Thr Gln Glu Glu Met Arg Ala Glu Leu Gly Leu Asp Ala
        595                 600                 605


Ala Gly Met Glu Ala Lys Ile Lys Ala Trp Leu Ala
    610                 615                 620


<210>  33
<211>  398
<212>  PRT
<213>  Escherichia coli

<400>  33

Met Lys Gln Leu Thr Ile Leu Gly Ser Thr Gly Ser Ile Gly Cys Ser
1               5                   10                  15


Thr Leu Asp Val Val Arg His Asn Pro Glu His Phe Arg Val Val Ala
            20                  25                  30
```

```
Leu Val Ala Gly Lys Asn Val Thr Arg Met Val Glu Gln Cys Leu Glu
        35                  40                  45

Phe Ser Pro Arg Tyr Ala Val Met Asp Asp Glu Ala Ser Ala Lys Leu
    50                  55                  60

Leu Lys Thr Met Leu Gln Gln Gln Gly Ser Arg Thr Glu Val Leu Ser
65                  70                  75                  80

Gly Gln Gln Ala Ala Cys Asp Met Ala Ala Leu Glu Asp Val Asp Gln
                85                  90                  95

Val Met Ala Ala Ile Val Gly Ala Ala Gly Leu Leu Pro Thr Leu Ala
            100                 105                 110

Ala Ile Arg Ala Gly Lys Thr Ile Leu Leu Ala Asn Lys Glu Ser Leu
        115                 120                 125

Val Thr Cys Gly Arg Leu Phe Met Asp Ala Val Lys Gln Ser Lys Ala
    130                 135                 140

Gln Leu Leu Pro Val Asp Ser Glu His Asn Ala Ile Phe Gln Ser Leu
145                 150                 155                 160

Pro Gln Pro Ile Gln His Asn Leu Gly Tyr Ala Asp Leu Glu Gln Asn
                165                 170                 175

Gly Val Val Ser Ile Leu Leu Thr Gly Ser Gly Gly Pro Phe Arg Glu
            180                 185                 190

Thr Pro Leu Arg Asp Leu Ala Thr Met Thr Pro Asp Gln Ala Cys Arg
        195                 200                 205

His Pro Asn Trp Ser Met Gly Arg Lys Ile Ser Val Asp Ser Ala Thr
    210                 215                 220

Met Met Asn Lys Gly Leu Glu Tyr Ile Glu Ala Arg Trp Leu Phe Asn
225                 230                 235                 240
```

```
Ala Ser Ala Ser Gln Met Glu Val Leu Ile His Pro Gln Ser Val Ile
                245                 250                 255


His Ser Met Val Arg Tyr Gln Asp Gly Ser Val Leu Ala Gln Leu Gly
            260                 265                 270


Glu Pro Asp Met Arg Thr Pro Ile Ala His Thr Met Ala Trp Pro Asn
        275                 280                 285


Arg Val Asn Ser Gly Val Lys Pro Leu Asp Phe Cys Lys Leu Ser Ala
    290                 295                 300


Leu Thr Phe Ala Ala Pro Asp Tyr Asp Arg Tyr Pro Cys Leu Lys Leu
305                 310                 315                 320


Ala Met Glu Ala Phe Glu Gln Gly Gln Ala Ala Thr Thr Ala Leu Asn
                325                 330                 335


Ala Ala Asn Glu Ile Thr Val Ala Ala Phe Leu Ala Gln Gln Ile Arg
            340                 345                 350


Phe Thr Asp Ile Ala Ala Leu Asn Leu Ser Val Leu Glu Lys Met Asp
        355                 360                 365


Met Arg Glu Pro Gln Cys Val Asp Asp Val Leu Ser Val Asp Ala Asn
    370                 375                 380


Ala Arg Glu Val Ala Arg Lys Glu Val Met Arg Leu Ala Ser
385                 390                 395


<210>  34
<211>  236
<212>  PRT
<213>  Escherichia coli

<400>  34

Met Ala Thr Thr His Leu Asp Val Cys Ala Val Val Pro Ala Ala Gly
1               5                   10                  15


Phe Gly Arg Arg Met Gln Thr Glu Cys Pro Lys Gln Tyr Leu Ser Ile
```

```
            20                  25                  30

Gly Asn Gln Thr Ile Leu Glu His Ser Val His Ala Leu Leu Ala His
        35                  40                  45

Pro Arg Val Lys Arg Val Val Ile Ala Ile Ser Pro Gly Asp Ser Arg
    50                  55                  60

Phe Ala Gln Leu Pro Leu Ala Asn His Pro Arg Ile Thr Val Val Asp
65                  70                  75                  80

Gly Gly Glu Glu Arg Ala Asp Ser Val Leu Ala Gly Leu Lys Ala Ala
                85                  90                  95

Gly Asp Ala Gln Trp Val Leu Val His Asp Ala Ala Arg Pro Cys Leu
            100                 105                 110

His Gln Asp Asp Leu Ala Arg Leu Leu Ala Leu Ser Glu Thr Ser Arg
        115                 120                 125

Thr Gly Gly Ile Leu Ala Ala Pro Val Arg Asp Thr Met Lys Arg Ala
    130                 135                 140

Glu Pro Gly Lys Asn Ala Ile Ala His Thr Val Asp Arg Asn Gly Leu
145                 150                 155                 160

Trp His Ala Leu Thr Pro Gln Phe Phe Pro Arg Glu Leu Leu His Asp
                165                 170                 175

Cys Leu Thr Arg Ala Leu Asn Glu Gly Ala Thr Ile Thr Asp Glu Ala
            180                 185                 190

Ser Ala Leu Glu Tyr Cys Gly Phe His Pro Gln Leu Val Glu Gly Arg
        195                 200                 205

Ala Asp Asn Ile Lys Val Thr Arg Pro Glu Asp Leu Ala Leu Ala Glu
    210                 215                 220

Phe Tyr Leu Thr Arg Thr Ile His Gln Glu Asn Thr
```

225 230 235

<210> 35
<211> 283
<212> PRT
<213> Escherichia coli

<400> 35

Met Arg Thr Gln Trp Pro Ser Pro Ala Lys Leu Asn Leu Phe Leu Tyr
1 5 10 15

Ile Thr Gly Gln Arg Ala Asp Gly Tyr His Thr Leu Gln Thr Leu Phe
20 25 30

Gln Phe Leu Asp Tyr Gly Asp Thr Ile Ser Ile Glu Leu Arg Asp Asp
35 40 45

Gly Asp Ile Arg Leu Leu Thr Pro Val Glu Gly Val Glu His Glu Asp
50 55 60

Asn Leu Ile Val Arg Ala Ala Arg Leu Leu Met Lys Thr Ala Ala Asp
65 70 75 80

Ser Gly Arg Leu Pro Thr Gly Ser Gly Ala Asn Ile Ser Ile Asp Lys
85 90 95

Arg Leu Pro Met Gly Gly Gly Leu Gly Gly Gly Ser Ser Asn Ala Ala
100 105 110

Thr Val Leu Val Ala Leu Asn His Leu Trp Gln Cys Gly Leu Ser Met
115 120 125

Asp Glu Leu Ala Glu Met Gly Leu Thr Leu Gly Ala Asp Val Pro Val
130 135 140

Phe Val Arg Gly His Ala Ala Phe Ala Glu Gly Val Gly Glu Ile Leu
145 150 155 160

Thr Pro Val Asp Pro Pro Glu Lys Trp Tyr Leu Val Ala His Pro Gly
165 170 175

```
Val Ser Ile Pro Thr Pro Val Ile Phe Lys Asp Pro Glu Leu Pro Arg
            180                 185                 190

Asn Thr Pro Lys Arg Ser Ile Glu Thr Leu Leu Lys Cys Glu Phe Ser
        195                 200                 205

Asn Asp Cys Glu Val Ile Ala Arg Lys Arg Phe Arg Glu Val Asp Ala
    210                 215                 220

Val Leu Ser Trp Leu Leu Glu Tyr Ala Pro Ser Arg Leu Thr Gly Thr
225                 230                 235                 240

Gly Ala Cys Val Phe Ala Glu Phe Asp Thr Glu Ser Glu Ala Arg Gln
                245                 250                 255

Val Leu Glu Gln Ala Pro Glu Trp Leu Asn Gly Phe Val Ala Lys Gly
            260                 265                 270

Ala Asn Leu Ser Pro Leu His Arg Ala Met Leu
        275                 280


<210>  36
<211>  159
<212>  PRT
<213>  Escherichia coli

<400>  36

Met Arg Ile Gly His Gly Phe Asp Val His Ala Phe Gly Gly Glu Gly
1               5                   10                  15

Pro Ile Ile Ile Gly Gly Val Arg Ile Pro Tyr Glu Lys Gly Leu Leu
            20                  25                  30

Ala His Ser Asp Gly Asp Val Val Leu His Ala Leu Thr Asp Ala Leu
        35                  40                  45

Leu Gly Ala Ala Ala Leu Gly Asp Ile Gly Lys Leu Phe Pro Asp Thr
    50                  55                  60
```

```
Asp Pro Ala Phe Lys Gly Ala Asp Ser Arg Glu Leu Leu Arg Glu Ala
65                  70                  75                  80


Trp Arg Arg Ile Gln Ala Lys Gly Tyr Ala Leu Gly Asn Val Asp Val
                85                  90                  95


Thr Ile Ile Ala Gln Ala Pro Arg Met Leu Pro His Ile Pro Gln Met
            100                 105                 110


Arg Val Phe Ile Ala Glu Asp Leu Gly Cys His Met Asp Asp Val Asn
        115                 120                 125


Val Lys Ala Thr Thr Thr Glu Lys Leu Gly Phe Thr Gly Arg Gly Glu
    130                 135                 140


Gly Ile Ala Cys Glu Ala Val Ala Leu Leu Ile Lys Ala Thr Lys
145                 150                 155


<210>  37
<211>  372
<212>  PRT
<213>  Escherichia coli

<400>  37

Met His Asn Gln Ala Pro Ile Gln Arg Arg Lys Ser Thr Arg Ile Tyr
1               5                   10                  15


Val Gly Asn Val Pro Ile Gly Asp Gly Ala Pro Ile Ala Val Gln Ser
            20                  25                  30


Met Thr Asn Thr Arg Thr Thr Asp Val Glu Ala Thr Val Asn Gln Ile
        35                  40                  45


Lys Ala Leu Glu Arg Val Gly Ala Asp Ile Val Arg Val Ser Val Pro
    50                  55                  60


Thr Met Asp Ala Ala Glu Ala Phe Lys Leu Ile Lys Gln Gln Val Asn
65                  70                  75                  80
```

```
Val Pro Leu Val Ala Asp Ile His Phe Asp Tyr Arg Ile Ala Leu Lys
                85                  90                  95

Val Ala Glu Tyr Gly Val Asp Cys Leu Arg Ile Asn Pro Gly Asn Ile
            100                 105                 110

Gly Asn Glu Glu Arg Ile Arg Met Val Val Asp Cys Ala Arg Asp Lys
        115                 120                 125

Asn Ile Pro Ile Arg Ile Gly Val Asn Ala Gly Ser Leu Glu Lys Asp
    130                 135                 140

Leu Gln Glu Lys Tyr Gly Glu Pro Thr Pro Gln Ala Leu Leu Glu Ser
145                 150                 155                 160

Ala Met Arg His Val Asp His Leu Asp Arg Leu Asn Phe Asp Gln Phe
                165                 170                 175

Lys Val Ser Val Lys Ala Ser Asp Val Phe Leu Ala Val Glu Ser Tyr
            180                 185                 190

Arg Leu Leu Ala Lys Gln Ile Asp Gln Pro Leu His Leu Gly Ile Thr
        195                 200                 205

Glu Ala Gly Gly Ala Arg Ser Gly Ala Val Lys Ser Ala Ile Gly Leu
    210                 215                 220

Gly Leu Leu Leu Ser Glu Gly Ile Gly Asp Thr Leu Arg Val Ser Leu
225                 230                 235                 240

Ala Ala Asp Pro Val Glu Glu Ile Lys Val Gly Phe Asp Ile Leu Lys
                245                 250                 255

Ser Leu Arg Ile Arg Ser Arg Gly Ile Asn Phe Ile Ala Cys Pro Thr
            260                 265                 270

Cys Ser Arg Gln Glu Phe Asp Val Ile Gly Thr Val Asn Ala Leu Glu
        275                 280                 285
```

```
Gln Arg Leu Glu Asp Ile Ile Thr Pro Met Asp Val Ser Ile Ile Gly
    290                 295                 300

Cys Val Val Asn Gly Pro Gly Glu Ala Leu Val Ser Thr Leu Gly Val
305                 310                 315                 320

Thr Gly Gly Asn Lys Lys Ser Gly Leu Tyr Glu Asp Gly Val Arg Lys
                325                 330                 335

Asp Arg Leu Asp Asn Asn Asp Met Ile Asp Gln Leu Glu Ala Arg Ile
            340                 345                 350

Arg Ala Lys Ala Ser Gln Leu Asp Glu Ala Arg Arg Ile Asp Val Gln
        355                 360                 365

Gln Val Glu Lys
    370
```

<210> 38
<211> 316
<212> PRT
<213> Escherichia coli

<400> 38

```
Met Gln Ile Leu Leu Ala Asn Pro Arg Gly Phe Cys Ala Gly Val Asp
1               5                   10                  15

Arg Ala Ile Ser Ile Val Glu Asn Ala Leu Ala Ile Tyr Gly Ala Pro
            20                  25                  30

Ile Tyr Val Arg His Glu Val Val His Asn Arg Tyr Val Val Asp Ser
        35                  40                  45

Leu Arg Glu Arg Gly Ala Ile Phe Ile Glu Gln Ile Ser Glu Val Pro
    50                  55                  60

Asp Gly Ala Ile Leu Ile Phe Ser Ala His Gly Val Ser Gln Ala Val
65                  70                  75                  80

Arg Asn Glu Ala Lys Ser Arg Asp Leu Thr Val Phe Asp Ala Thr Cys
```

85 90 95

Pro Leu Val Thr Lys Val His Met Glu Val Ala Arg Ala Ser Arg Arg
100 105 110

Gly Glu Glu Ser Ile Leu Ile Gly His Ala Gly His Pro Glu Val Glu
115 120 125

Gly Thr Met Gly Gln Tyr Ser Asn Pro Glu Gly Gly Met Tyr Leu Val
130 135 140

Glu Ser Pro Asp Asp Val Trp Lys Leu Thr Val Lys Asn Glu Glu Lys
145 150 155 160

Leu Ser Phe Met Thr Gln Thr Thr Leu Ser Val Asp Asp Thr Ser Asp
165 170 175

Val Ile Asp Ala Leu Arg Lys Arg Phe Pro Lys Ile Val Gly Pro Arg
180 185 190

Lys Asp Asp Ile Cys Tyr Ala Thr Thr Asn Arg Gln Glu Ala Val Arg
195 200 205

Ala Leu Ala Glu Gln Ala Glu Val Val Leu Val Val Gly Ser Lys Asn
210 215 220

Ser Ser Asn Ser Asn Arg Leu Ala Glu Leu Ala Gln Arg Met Gly Lys
225 230 235 240

Arg Ala Phe Leu Ile Asp Asp Ala Thr Asp Ile Gln Glu Glu Trp Val
245 250 255

Lys Glu Ala Lys Cys Val Gly Val Thr Ala Gly Ala Ser Ala Pro Asp
260 265 270

Ile Leu Val Gln Asn Val Val Ala Arg Leu Gln Gln Leu Gly Gly Gly
275 280 285

Glu Ala Ile Pro Leu Glu Gly Arg Glu Glu Asn Ile Val Phe Glu Val

290 295 300

Pro Lys Glu Leu Arg Val Asp Ile Arg Glu Val Asp
305 310 315

<210> 39
<211> 182
<212> PRT
<213> Escherichia coli

<400> 39

Met Gln Thr Glu His Val Ile Leu Leu Asn Ala Gln Gly Val Pro Thr
1 5 10 15

Gly Thr Leu Glu Lys Tyr Ala Ala His Thr Ala Asp Thr Arg Leu His
20 25 30

Leu Ala Phe Ser Ser Trp Leu Phe Asn Ala Lys Gly Gln Leu Leu Val
35 40 45

Thr Arg Arg Ala Leu Ser Lys Lys Ala Trp Pro Gly Val Trp Thr Asn
50 55 60

Ser Val Cys Gly His Pro Gln Leu Gly Glu Ser Asn Glu Asp Ala Val
65 70 75 80

Ile Arg Arg Cys Arg Tyr Glu Leu Gly Val Glu Ile Thr Pro Pro Glu
85 90 95

Ser Ile Tyr Pro Asp Phe Arg Tyr Arg Ala Thr Asp Pro Ser Gly Ile
100 105 110

Val Glu Asn Glu Val Cys Pro Val Phe Ala Ala Arg Thr Thr Ser Ala
115 120 125

Leu Gln Ile Asn Asp Asp Glu Val Met Asp Tyr Gln Trp Cys Asp Leu
130 135 140

Ala Asp Val Leu His Gly Ile Asp Ala Thr Pro Trp Ala Phe Ser Pro
145 150 155 160

```
Trp Met Val Met Gln Ala Thr Asn Arg Glu Ala Arg Lys Arg Leu Ser
                165                 170                 175


Ala Phe Thr Gln Leu Lys
            180


<210>  40
<211>  299
<212>  PRT
<213>  Escherichia coli

<400>  40

Met Asp Phe Pro Gln Gln Leu Glu Ala Cys Val Lys Gln Ala Asn Gln
1               5                   10                  15


Ala Leu Ser Arg Phe Ile Ala Pro Leu Pro Phe Gln Asn Thr Pro Val
            20                  25                  30


Val Glu Thr Met Gln Tyr Gly Ala Leu Leu Gly Gly Lys Arg Leu Arg
        35                  40                  45


Pro Phe Leu Val Tyr Ala Thr Gly His Met Phe Gly Val Ser Thr Asn
    50                  55                  60


Thr Leu Asp Ala Pro Ala Ala Ala Val Glu Cys Ile His Ala Tyr Ser
65                  70                  75                  80


Leu Ile His Asp Asp Leu Pro Ala Met Asp Asp Asp Asp Leu Arg Arg
                85                  90                  95


Gly Leu Pro Thr Cys His Val Lys Phe Gly Glu Ala Asn Ala Ile Leu
            100                 105                 110


Ala Gly Asp Ala Leu Gln Thr Leu Ala Phe Ser Ile Leu Ser Asp Ala
        115                 120                 125


Asp Met Pro Glu Val Ser Asp Arg Asp Arg Ile Ser Met Ile Ser Glu
    130                 135                 140
```

```
Leu Ala Ser Ala Ser Gly Ile Ala Gly Met Cys Gly Gly Gln Ala Leu
145                 150                 155                 160


Asp Leu Asp Ala Glu Gly Lys His Val Pro Leu Asp Ala Leu Glu Arg
                165                 170                 175


Ile His Arg His Lys Thr Gly Ala Leu Ile Arg Ala Ala Val Arg Leu
            180                 185                 190


Gly Ala Leu Ser Ala Gly Asp Lys Gly Arg Arg Ala Leu Pro Val Leu
        195                 200                 205


Asp Lys Tyr Ala Glu Ser Ile Gly Leu Ala Phe Gln Val Gln Asp Asp
    210                 215                 220


Ile Leu Asp Val Val Gly Asp Thr Ala Thr Leu Gly Lys Arg Gln Gly
225                 230                 235                 240


Ala Asp Gln Gln Leu Gly Lys Ser Thr Tyr Pro Ala Leu Leu Gly Leu
                245                 250                 255


Glu Gln Ala Arg Lys Lys Ala Arg Asp Leu Ile Asp Asp Ala Arg Gln
            260                 265                 270


Ser Leu Lys Gln Leu Ala Glu Gln Ser Leu Asp Thr Ser Ala Leu Glu
        275                 280                 285


Ala Leu Ala Asp Tyr Ile Ile Gln Arg Asn Lys
    290                 295


<210>  41
<211>  1083
<212>  DNA
<213>  Saccharomyces cerevisiae

<400>  41
atgaagctac taaccttccc aggtcaaggg acctccatct ccatttcgat attaaaagcg       60

ataataagaa acaaatcaag agaattccaa acaatactga gtcagaacgg caaggaatca      120

aatgatctat tgcagtacat cttccagaac ccttccagcc ccggaagcat tgcagtctgc      180
```

tccaaccttt tctatcaatt gtaccagata ctctcgaatc cttctgatcc tcaagatcaa 240

gcaccaaaaa atatgactaa gatcgattcc cccgacaaga aagacaatga acaatgttac 300

cttttgggtc actcgctagg cgagttaaca tgtctgagtg ttaattcact gttttcgtta 360

aaggatcttt ttgatattgc taattttaga aataagttaa tggtaacatc tactgaaaag 420

tacttagtag cccacaatat caacagatcc aacaaatttg aaatgtgggc actctcttct 480

ccgagggcca cagatttacc gcaagaagtg caaaaactac taaattcccc taatttatta 540

tcatcttcac aaaataccat ttctgtagca aatgcaaatt cagtaaagca atgtgtagtc 600

accggtctgg ttgatgattt agagtcctta agaacagaat tgaacttaag gttcccgcgt 660

ttaagaatta cagaattaac taacccatac aacatcccct tccataatag cactgtgttg 720

aggcccgttc aggaaccact ctatgactac atttgggata tattaaagaa aaacggaact 780

cacacgttga tggagttgaa ccatccaata atagctaact tagatggtaa tatatcttac 840

tatattcatc atgccctaga tagattcgtt aagtgttcaa gcaggactgt gcaattcacc 900

atgtgttatg ataccataaa ctctggaacc ccagtggaaa ttgataagag tatttgcttt 960

ggcccgggca atgtgattta taaccttatt cggagaaatt gtccccaagt ggacactata 1020

gaatacacct ctttagcaac tatagacgct tatcacaagg cggcagagga gaacaaagat 1080

tga 1083

<210> 42
<211> 360
<212> PRT
<213> Saccharomyces cerevisiae

<400> 42

Met Lys Leu Leu Thr Phe Pro Gly Gln Gly Thr Ser Ile Ser Ile Ser
1 5 10 15

Ile Leu Lys Ala Ile Ile Arg Asn Lys Ser Arg Glu Phe Gln Thr Ile
20 25 30

Leu Ser Gln Asn Gly Lys Glu Ser Asn Asp Leu Leu Gln Tyr Ile Phe
35 40 45

```
Gln Asn Pro Ser Ser Pro Gly Ser Ile Ala Val Cys Ser Asn Leu Phe
    50                  55                  60

Tyr Gln Leu Tyr Gln Ile Leu Ser Asn Pro Ser Asp Pro Gln Asp Gln
65                  70                  75                  80

Ala Pro Lys Asn Met Thr Lys Ile Asp Ser Pro Asp Lys Lys Asp Asn
                85                  90                  95

Glu Gln Cys Tyr Leu Leu Gly His Ser Leu Gly Glu Leu Thr Cys Leu
            100                 105                 110

Ser Val Asn Ser Leu Phe Ser Leu Lys Asp Leu Phe Asp Ile Ala Asn
        115                 120                 125

Phe Arg Asn Lys Leu Met Val Thr Ser Thr Glu Lys Tyr Leu Val Ala
    130                 135                 140

His Asn Ile Asn Arg Ser Asn Lys Phe Glu Met Trp Ala Leu Ser Ser
145                 150                 155                 160

Pro Arg Ala Thr Asp Leu Pro Gln Glu Val Gln Lys Leu Leu Asn Ser
                165                 170                 175

Pro Asn Leu Leu Ser Ser Ser Gln Asn Thr Ile Ser Val Ala Asn Ala
            180                 185                 190

Asn Ser Val Lys Gln Cys Val Val Thr Gly Leu Val Asp Asp Leu Glu
        195                 200                 205

Ser Leu Arg Thr Glu Leu Asn Leu Arg Phe Pro Arg Leu Arg Ile Thr
    210                 215                 220

Glu Leu Thr Asn Pro Tyr Asn Ile Pro Phe His Asn Ser Thr Val Leu
225                 230                 235                 240

Arg Pro Val Gln Glu Pro Leu Tyr Asp Tyr Ile Trp Asp Ile Leu Lys
                245                 250                 255
```

```
Lys Asn Gly Thr His Thr Leu Met Glu Leu Asn His Pro Ile Ile Ala
            260                 265                 270


Asn Leu Asp Gly Asn Ile Ser Tyr Tyr Ile His His Ala Leu Asp Arg
        275                 280                 285


Phe Val Lys Cys Ser Ser Arg Thr Val Gln Phe Thr Met Cys Tyr Asp
    290                 295                 300


Thr Ile Asn Ser Gly Thr Pro Val Glu Ile Asp Lys Ser Ile Cys Phe
305                 310                 315                 320


Gly Pro Gly Asn Val Ile Tyr Asn Leu Ile Arg Arg Asn Cys Pro Gln
                325                 330                 335


Val Asp Thr Ile Glu Tyr Thr Ser Leu Ala Thr Ile Asp Ala Tyr His
            340                 345                 350


Lys Ala Ala Glu Glu Asn Lys Asp
        355                 360


<210>  43
<211>  1191
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  Artificial beta-ketothiolase (BktB) nucleotide sequence

<400>  43
atgactagag aagttgtcgt cgtttccggt gtccgtaccg ctatcggtac tttcggtggt       60

tccttaaagg atgttgctcc tgctgaattg ggtgctttag ttgttagaga agctttggcc      120

agagcccaag tctccggtga cgacgttggt cacgtcgttt tcggtaacgt catccaaact      180

gaaccacgtg acatgtactt gggtagagtc gccgctgtta acggtggtgt caccatcaac      240

gctcctgcct taactgttaa cagattatgt ggttccggtt tacaagctat tgtctctgcc      300

gcccaaacta tcttgttggg tgatactgac gttgctattg gtggtggtgc tgaatctatg      360

tctagagctc catacttggc tccagctgcc cgttggggtg ctagaatggg tgacgccggt      420
```

ttggtcgata tgatgttggg tgccttgcat gatcctttcc acagaatcca catgggtgtt 480

accgctgaaa acgttgctaa ggaatacgat atctctagag ctcaacaaga tgaagccgct 540

ttagaatctc acagacgtgc ctccgccgct attaaggctg gttacttcaa ggaccaaatt 600

gttccagttg tctctaaggg tcgtaaaggt gatgttacct ttgatactga cgaacacgtt 660

agacacgacg ccactattga cgatatgact aaattaagac cagtctttgt taaggagaat 720

ggtaccgtta ctgctggtaa cgcttctggt ttgaacgatg ccgccgctgc cgttgttatg 780

atggaaagag ctgaagccga aagacgtggt ttaaagccat tggccagatt agtctcctac 840

ggtcacgctg gtgtcgaccc aaaggctatg ggtatcggtc cagttcctgc tactaagatt 900

gctttagaaa gagctggttt gcaagtttct gacttggacg tcatcgaagc caacgaagcc 960

ttcgctgctc aagcttgtgc tgtcaccaag gctttgggtt tggatccagc taaagttaac 1020

cctaatggtt ctggtatttc cttgggtcac ccaatcggtg ctaccggtgc tttaatcact 1080

gttaaagcct tacacgaatt gaacagagtt caaggtagat acgctttggt cactatgtgc 1140

atcggtggtg gtcaaggtat cgctgctatc ttcgaaagaa tcggatccta a 1191

<210> 44
<211> 396
<212> PRT
<213> Artificial Sequence

<220>
<223> Engineered beta-ketothiolase (BktB)

<400> 44

Met Thr Arg Glu Val Val Val Val Ser Gly Val Arg Thr Ala Ile Gly
1 5 10 15

Thr Phe Gly Gly Ser Leu Lys Asp Val Ala Pro Ala Glu Leu Gly Ala
20 25 30

Leu Val Val Arg Glu Ala Leu Ala Arg Ala Gln Val Ser Gly Asp Asp
35 40 45

Val Gly His Val Val Phe Gly Asn Val Ile Gln Thr Glu Pro Arg Asp
50 55 60

```
Met Tyr Leu Gly Arg Val Ala Ala Val Asn Gly Gly Val Thr Ile Asn
65                  70                  75                  80

Ala Pro Ala Leu Thr Val Asn Arg Leu Cys Gly Ser Gly Leu Gln Ala
                85                  90                  95

Ile Val Ser Ala Ala Gln Thr Ile Leu Leu Gly Asp Thr Asp Val Ala
            100                 105                 110

Ile Gly Gly Gly Ala Glu Ser Met Ser Arg Ala Pro Tyr Leu Ala Pro
        115                 120                 125

Ala Ala Arg Trp Gly Ala Arg Met Gly Asp Ala Gly Leu Val Asp Met
    130                 135                 140

Met Leu Gly Ala Leu His Asp Pro Phe His Arg Ile His Met Gly Val
145                 150                 155                 160

Thr Ala Glu Asn Val Ala Lys Glu Tyr Asp Ile Ser Arg Ala Gln Gln
                165                 170                 175

Asp Glu Ala Ala Leu Glu Ser His Arg Arg Ala Ser Ala Ala Ile Lys
            180                 185                 190

Ala Gly Tyr Phe Lys Asp Gln Ile Val Pro Val Val Ser Lys Gly Arg
        195                 200                 205

Lys Gly Asp Val Thr Phe Asp Thr Asp Glu His Val Arg His Asp Ala
    210                 215                 220

Thr Ile Asp Asp Met Thr Lys Leu Arg Pro Val Phe Val Lys Glu Asn
225                 230                 235                 240

Gly Thr Val Thr Ala Gly Asn Ala Ser Gly Leu Asn Asp Ala Ala Ala
                245                 250                 255

Ala Val Val Met Met Glu Arg Ala Glu Ala Glu Arg Arg Gly Leu Lys
            260                 265                 270
```

```
Pro Leu Ala Arg Leu Val Ser Tyr Gly His Ala Gly Val Asp Pro Lys
        275                 280                 285


Ala Met Gly Ile Gly Pro Val Pro Ala Thr Lys Ile Ala Leu Glu Arg
    290                 295                 300


Ala Gly Leu Gln Val Ser Asp Leu Asp Val Ile Glu Ala Asn Glu Ala
305                 310                 315                 320


Phe Ala Ala Gln Ala Cys Ala Val Thr Lys Ala Leu Gly Leu Asp Pro
                325                 330                 335


Ala Lys Val Asn Pro Asn Gly Ser Gly Ile Ser Leu Gly His Pro Ile
            340                 345                 350


Gly Ala Thr Gly Ala Leu Ile Thr Val Lys Ala Leu His Glu Leu Asn
        355                 360                 365


Arg Val Gln Gly Arg Tyr Ala Leu Val Thr Met Cys Ile Gly Gly Gly
    370                 375                 380


Gln Gly Ile Ala Ala Ile Phe Glu Arg Ile Gly Ser
385                 390                 395


<210>  45
<211>  858
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  Artificial PaaH1: 3-hydroxyacyl-CoA dehydrogenase nucleotide
       sequence

<400>  45
atgtccatca gaactgtcgg tattgttggt gctggtacta tgggtaacgg tattgctcaa       60

gcctgtgctg tcgtcggttt gaacgtcgtc atggtcgaca tttctgacgc tgctgttcaa      120

aagggtgttg ctactgtcgc ttcctctttg gacagattaa ttaagaagga aaagttgacc      180

gaagccgaca aggcctctgc cttggccaga attaagggtt ccacttctta tgacgacttg      240

aaagctaccg acattgttat cgaagctgct actgaaaact acgatttgaa agttaagatc      300
```

ttgaagcaaa ttgatggtat cgtcggtgag aacgtcatta ttgcttctaa cacttcctcc 360

atttctatca ctaaattagc cgccgtcacc tctagagccg acagatttat cggtatgcac 420

ttctttaatc cagttccagt catggctttg gtcgaattaa ttagaggttt gcaaacctcc 480

gacaccaccc acgccgccgt tgaagctttg tctaagcaat tgggtaagta cccaatcacc 540

gttaaaaatt ccccaggttt cgttgtcaac cgtattttgt gcccaatgat caatgaagct 600

ttctgtgtct tgggtgaggg tttggcctcc ccagaagaaa tcgatgaagg tatgaagtta 660

ggttgtaacc accctattgg tcctttagcc ttggccgaca tgatcggttt agacactatg 720

ttggccgtta tggaagtctt gtacactgaa ttcgctgacc caaagtacag accagctatg 780

ttaatgagag aaatggttgc tgccggttat ttgggtagaa agactggtcg tggtgtttat 840

gtctactcta aagggatc 858

<210> 46
<211> 286
<212> PRT
<213> Artificial Sequence

<220>
<223> Engineered PaaH1: 3-hydroxyacyl-CoA dehydrogenase

<400> 46

Met Ser Ile Arg Thr Val Gly Ile Val Gly Ala Gly Thr Met Gly Asn
1 5 10 15

Gly Ile Ala Gln Ala Cys Ala Val Val Gly Leu Asn Val Val Met Val
20 25 30

Asp Ile Ser Asp Ala Ala Val Gln Lys Gly Val Ala Thr Val Ala Ser
35 40 45

Ser Leu Asp Arg Leu Ile Lys Lys Glu Lys Leu Thr Glu Ala Asp Lys
50 55 60

Ala Ser Ala Leu Ala Arg Ile Lys Gly Ser Thr Ser Tyr Asp Asp Leu
65 70 75 80

```
Lys Ala Thr Asp Ile Val Ile Glu Ala Ala Thr Glu Asn Tyr Asp Leu
                85                  90                  95

Lys Val Lys Ile Leu Lys Gln Ile Asp Gly Ile Val Gly Glu Asn Val
            100                 105                 110

Ile Ile Ala Ser Asn Thr Ser Ser Ile Ser Ile Thr Lys Leu Ala Ala
        115                 120                 125

Val Thr Ser Arg Ala Asp Arg Phe Ile Gly Met His Phe Phe Asn Pro
    130                 135                 140

Val Pro Val Met Ala Leu Val Glu Leu Ile Arg Gly Leu Gln Thr Ser
145                 150                 155                 160

Asp Thr Thr His Ala Ala Val Glu Ala Leu Ser Lys Gln Leu Gly Lys
                165                 170                 175

Tyr Pro Ile Thr Val Lys Asn Ser Pro Gly Phe Val Val Asn Arg Ile
            180                 185                 190

Leu Cys Pro Met Ile Asn Glu Ala Phe Cys Val Leu Gly Glu Gly Leu
        195                 200                 205

Ala Ser Pro Glu Glu Ile Asp Glu Gly Met Lys Leu Gly Cys Asn His
    210                 215                 220

Pro Ile Gly Pro Leu Ala Leu Ala Asp Met Ile Gly Leu Asp Thr Met
225                 230                 235                 240

Leu Ala Val Met Glu Val Leu Tyr Thr Glu Phe Ala Asp Pro Lys Tyr
                245                 250                 255

Arg Pro Ala Met Leu Met Arg Glu Met Val Ala Ala Gly Tyr Leu Gly
            260                 265                 270

Arg Lys Thr Gly Arg Gly Val Tyr Val Tyr Ser Lys Gly Ile
        275                 280                 285
```

<210> 47
<211> 792
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial crotonase (Crt) nucleotide sequence

<400> 47
atggaattga acaacgttat tttggaaaag gaaggtaagg tcgctgtcgt tactatcaac 60

agaccaaagg ctttaaacgc tttgaactct gacaccttga aagaaatgga ttatgttatc 120

ggtgaaatcg aaaatgactc tgaagttttg gccgttatct tgactggtgc tggtgaaaaa 180

tctttcgttg ctggtgctga catttctgaa atgaaggaga tgaataccat tgaaggtaga 240

aagttcggta tcttgggtaa caaggttttt agaagattgg aattgttgga aaaaccagtc 300

atcgctgctg ttaacggttt cgctttaggt ggtggttgtg aaatcgctat gtcctgtgac 360

attcgtatcg cctcctccaa tgctagattc ggtcaaccag aagttggttt aggtattact 420

ccaggtttcg gtggtaccca aagattgtct agattggtcg gtatgggtat ggctaagcaa 480

ttaattttca ctgctcaaaa cattaaggct gatgaagcct tacgtattgg tttggtcaac 540

aaggtcgttg aaccatctga attgatgaat accgctaagg aaattgctaa caaaattgtt 600

tctaatgccc cagttgctgt caagttgtcc aagcaagcta ttaacagagg tatgcaatgt 660

gatattgaca ctgctttggc tttcgaatcc gaagcttttg gtgaatgttt ttctaccgaa 720

gatcaaaagg atgctatgac cgctttcatc gagaagagaa agatcgaagg tttcaaaaac 780

agaggatcct aa 792

<210> 48
<211> 263
<212> PRT
<213> Artificial Sequence

<220>
<223> Engineered crotonase (Crt)

<400> 48

Met Glu Leu Asn Asn Val Ile Leu Glu Lys Glu Gly Lys Val Ala Val
1 5 10 15

```
Val Thr Ile Asn Arg Pro Lys Ala Leu Asn Ala Leu Asn Ser Asp Thr
            20                  25                  30

Leu Lys Glu Met Asp Tyr Val Ile Gly Glu Ile Glu Asn Asp Ser Glu
        35                  40                  45

Val Leu Ala Val Ile Leu Thr Gly Ala Gly Glu Lys Ser Phe Val Ala
    50                  55                  60

Gly Ala Asp Ile Ser Glu Met Lys Glu Met Asn Thr Ile Glu Gly Arg
65                  70                  75                  80

Lys Phe Gly Ile Leu Gly Asn Lys Val Phe Arg Arg Leu Glu Leu Leu
                85                  90                  95

Glu Lys Pro Val Ile Ala Ala Val Asn Gly Phe Ala Leu Gly Gly Gly
            100                 105                 110

Cys Glu Ile Ala Met Ser Cys Asp Ile Arg Ile Ala Ser Ser Asn Ala
        115                 120                 125

Arg Phe Gly Gln Pro Glu Val Gly Leu Gly Ile Thr Pro Gly Phe Gly
    130                 135                 140

Gly Thr Gln Arg Leu Ser Arg Leu Val Gly Met Gly Met Ala Lys Gln
145                 150                 155                 160

Leu Ile Phe Thr Ala Gln Asn Ile Lys Ala Asp Glu Ala Leu Arg Ile
                165                 170                 175

Gly Leu Val Asn Lys Val Val Glu Pro Ser Glu Leu Met Asn Thr Ala
            180                 185                 190

Lys Glu Ile Ala Asn Lys Ile Val Ser Asn Ala Pro Val Ala Val Lys
        195                 200                 205

Leu Ser Lys Gln Ala Ile Asn Arg Gly Met Gln Cys Asp Ile Asp Thr
    210                 215                 220
```

Ala Leu Ala Phe Glu Ser Glu Ala Phe Gly Glu Cys Phe Ser Thr Glu
225 230 235 240

Asp Gln Lys Asp Ala Met Thr Ala Phe Ile Glu Lys Arg Lys Ile Glu
245 250 255

Gly Phe Lys Asn Arg Gly Ser
260

<210> 49
<211> 1200
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial Ter: trans-2-enoyl-CoA reductase nucleotide sequence

<400> 49
atgattgtca aaccaatggt tcgtaacaac atttgtttaa atgcccaccc acaaggttgt 60

aagaagggtg ttgaagatca aatcgaatac actaaaaaga gaattaccgc tgaagttaaa 120

gctggtgcta aggccccaaa gaacgttttg gttttgggtt gttccaacgg ttacggtttg 180

gcctccagaa ttactgctgc ttttggttac ggtgccgcta ccatcggtgt ctctttcgaa 240

aaggccggtt ccgaaactaa gtacggtact ccaggttggt acaataactt ggctttcgat 300

gaagctgcta agagagaagg tttgtattcc gttactattg acggtgatgc cttttctgac 360

gaaatcaaag ctcaagtcat cgaagaagcc aaaaagaaag gtatcaagtt cgatttgatt 420

gtctactctt tagcctctcc tgttagaact gatccagata ctggtattat gcacaaatcc 480

gttttgaagc cattcggtaa gaccttcact ggtaaaactg tcgatccttt cactggtgaa 540

ttaaaggaaa tctctgctga acctgccaac gacgaagaag ctgctgccac tgttaaggtt 600

atgggtggtg aagactggga aagatggatc aagcaattat ctaaggaagg tttgttggaa 660

gaaggttgta tcaccttggc ttactcttac atcggtccag aagctaccca agctttgtac 720

agaaagggta ccattggtaa ggctaaagaa cacttggagg ctactgctca tagattgaac 780

aaggaaaatc catccatcag agcctttgtt tccgtcaata aaggtttggt cactagagcc 840

tctgccgtca ttccagttat ccctttatac ttggcttctt tgtttaaagt catgaaggaa 900

aagggtaacc atgaaggttg tatcgaacaa atcactcgtt tgtacgctga acgtttatac 960

agaaaggacg gtaccatccc tgtcgatgaa gaaaacagaa tcagaatcga cgattgggaa 1020

ttggaagaag atgttcaaaa agccgtttcc gccttgatgg aaaaggtcac cggtgaaaat 1080

gccgaatcct tgactgactt agctggttac agacatgact ttttagcttc taatggtttc 1140

gatgttgaag gtattaacta tgaggctgaa gtcgaaagat ttgacagaat cggatcctaa 1200

<210> 50
<211> 399
<212> PRT
<213> Artificial Sequence

<220>
<223> Engineered Ter: trans-2-enoyl-CoA reductase

<400> 50

Met Ile Val Lys Pro Met Val Arg Asn Asn Ile Cys Leu Asn Ala His
1 5 10 15

Pro Gln Gly Cys Lys Lys Gly Val Glu Asp Gln Ile Glu Tyr Thr Lys
20 25 30

Lys Arg Ile Thr Ala Glu Val Lys Ala Gly Ala Lys Ala Pro Lys Asn
35 40 45

Val Leu Val Leu Gly Cys Ser Asn Gly Tyr Gly Leu Ala Ser Arg Ile
50 55 60

Thr Ala Ala Phe Gly Tyr Gly Ala Ala Thr Ile Gly Val Ser Phe Glu
65 70 75 80

Lys Ala Gly Ser Glu Thr Lys Tyr Gly Thr Pro Gly Trp Tyr Asn Asn
85 90 95

Leu Ala Phe Asp Glu Ala Ala Lys Arg Glu Gly Leu Tyr Ser Val Thr
100 105 110

Ile Asp Gly Asp Ala Phe Ser Asp Glu Ile Lys Ala Gln Val Ile Glu
115 120 125

```
Glu Ala Lys Lys Lys Gly Ile Lys Phe Asp Leu Ile Val Tyr Ser Leu
    130                 135                 140

Ala Ser Pro Val Arg Thr Asp Pro Asp Thr Gly Ile Met His Lys Ser
145                 150                 155                 160

Val Leu Lys Pro Phe Gly Lys Thr Phe Thr Gly Lys Thr Val Asp Pro
                165                 170                 175

Phe Thr Gly Glu Leu Lys Glu Ile Ser Ala Glu Pro Ala Asn Asp Glu
            180                 185                 190

Glu Ala Ala Ala Thr Val Lys Val Met Gly Gly Glu Asp Trp Glu Arg
        195                 200                 205

Trp Ile Lys Gln Leu Ser Lys Glu Gly Leu Leu Glu Glu Gly Cys Ile
    210                 215                 220

Thr Leu Ala Tyr Ser Tyr Ile Gly Pro Glu Ala Thr Gln Ala Leu Tyr
225                 230                 235                 240

Arg Lys Gly Thr Ile Gly Lys Ala Lys Glu His Leu Glu Ala Thr Ala
                245                 250                 255

His Arg Leu Asn Lys Glu Asn Pro Ser Ile Arg Ala Phe Val Ser Val
            260                 265                 270

Asn Lys Gly Leu Val Thr Arg Ala Ser Ala Val Ile Pro Val Ile Pro
        275                 280                 285

Leu Tyr Leu Ala Ser Leu Phe Lys Val Met Lys Glu Lys Gly Asn His
    290                 295                 300

Glu Gly Cys Ile Glu Gln Ile Thr Arg Leu Tyr Ala Glu Arg Leu Tyr
305                 310                 315                 320

Arg Lys Asp Gly Thr Ile Pro Val Asp Glu Glu Asn Arg Ile Arg Ile
                325                 330                 335
```

Asp Asp Trp Glu Leu Glu Glu Asp Val Gln Lys Ala Val Ser Ala Leu
340 345 350

Met Glu Lys Val Thr Gly Glu Asn Ala Glu Ser Leu Thr Asp Leu Ala
355 360 365

Gly Tyr Arg His Asp Phe Leu Ala Ser Asn Gly Phe Asp Val Glu Gly
370 375 380

Ile Asn Tyr Glu Ala Glu Val Glu Arg Phe Asp Arg Ile Gly Ser
385 390 395

<210> 51
<211> 812
<212> DNA
<213> Artificial Sequence

<220>
<223> Truncated 3-hydroxy-3-methyl-glutaryl-CoA reductase (tHMG1)

<400> 51
atggcgcgtg accaattggt gaaaactgaa gtcaccaaga agtcttttac tgctcctgta 60

caaaaggctt ctacaccagt tttaaccaat aaaacagtca tttctggatc gaaagtcaaa 120

agtttatcat ctgcgcaatc gagctcatca ggaccttcat catctagtga ggaagatgat 180

tcccgcgata ttgaaagctt ggataagaaa atacgtcctt tagaagaatt agaagcatta 240

ttaagtagtg gaaatacaaa acaattgaag aacaaagagg tcgctgcctt ggttattcac 300

ggtaagttac ctttgtacgc tttggagaaa aaattaggtg atactacgag agcggttgcg 360

gtacgtagga aggctctttc aattttggca gaagctcctg tattagcatc tgatcgttta 420

ccatataaaa attatgacta cgaccgcgta tttggcgctt gttgtgaaaa tgttataggt 480

tacatgcctt tgcccgttgg tgttataggc cccttggtta tcgatggtac atcttatcat 540

ataccaatgg caactacaga gggttgtttg gtagcttctg ccatgcgtgg ctgtaaggca 600

atcaatgctg gcggtggtgc aacaactgtt ttaactaagg atggtatgac aagaggccca 660

gtagtccgtt tcccaacttt gaaaagatct ggtgcctgta agatatggtt agactcagaa 720

gagggacaaa acgcaattaa aaaagctttt aactctacat caagatttgc acgtctgcaa 780

catattcaaa cttgtctagc aggagatttg tt 812

<210> 52
<211> 270
<212> PRT
<213> Artificial Sequence

<220>
<223> Truncated 3-hydroxy-3-methyl-glutaryl-CoA reductase (tHMG1)

<400> 52

```
Met Ala Arg Asp Gln Leu Val Lys Thr Glu Val Thr Lys Lys Ser Phe
1               5                  10                  15


Thr Ala Pro Val Gln Lys Ala Ser Thr Pro Val Leu Thr Asn Lys Thr
            20                  25                  30


Val Ile Ser Gly Ser Lys Val Lys Ser Leu Ser Ser Ala Gln Ser Ser
        35                  40                  45


Ser Ser Gly Pro Ser Ser Ser Ser Glu Glu Asp Asp Ser Arg Asp Ile
    50                  55                  60


Glu Ser Leu Asp Lys Lys Ile Arg Pro Leu Glu Glu Leu Glu Ala Leu
65                  70                  75                  80


Leu Ser Ser Gly Asn Thr Lys Gln Leu Lys Asn Lys Glu Val Ala Ala
                85                  90                  95


Leu Val Ile His Gly Lys Leu Pro Leu Tyr Ala Leu Glu Lys Lys Leu
            100                 105                 110


Gly Asp Thr Thr Arg Ala Val Ala Val Arg Arg Lys Ala Leu Ser Ile
        115                 120                 125


Leu Ala Glu Ala Pro Val Leu Ala Ser Asp Arg Leu Pro Tyr Lys Asn
    130                 135                 140


Tyr Asp Tyr Asp Arg Val Phe Gly Ala Cys Cys Glu Asn Val Ile Gly
145                 150                 155                 160
```

Tyr Met Pro Leu Pro Val Gly Val Ile Gly Pro Leu Val Ile Asp Gly
165 170 175

Thr Ser Tyr His Ile Pro Met Ala Thr Thr Glu Gly Cys Leu Val Ala
180 185 190

Ser Ala Met Arg Gly Cys Lys Ala Ile Asn Ala Gly Gly Gly Ala Thr
195 200 205

Thr Val Leu Thr Lys Asp Gly Met Thr Arg Gly Pro Val Val Arg Phe
210 215 220

Pro Thr Leu Lys Arg Ser Gly Ala Cys Lys Ile Trp Leu Asp Ser Glu
225 230 235 240

Glu Gly Gln Asn Ala Ile Lys Lys Ala Phe Asn Ser Thr Ser Arg Phe
245 250 255

Ala Arg Leu Gln His Ile Gln Thr Cys Leu Ala Gly Asp Leu
260 265 270

<210> 53
<211> 2412
<212> DNA
<213> Enterococcus sp.

<400> 53
atgaagactg tcgttatcat agatgccttg agaacaccaa tcggtaaata caaaggttca 60

ttatcccaag tttccgccgt tgacttaggt actcatgtta ctacacaatt gttgaagaga 120

cactccacaa tcagtgaaga aatcgatcaa gtcatattcg gtaacgtatt gcaagctggt 180

aatggtcaaa acccagccag acaaatagct atcaattctg gtttatcaca tgaaattcct 240

gctatgacag taaacgaagt ttgtggttca ggcatgaaag cagtcatttt ggccaagcaa 300

ttgatacaat taggtgaagc agaagtttta atcgccggtg gtatagaaaa catgagtcaa 360

gctccaaaat tgcaaagatt caattacgaa actgaatctt acgatgcacc tttctcttcg 420

atgatgtatg atggtttgac tgacgctttt tctggtcaag caatgggttt aacagctgaa 480

aatgtcgcag aaaagtacca tgtaaccaga gaagaacaag atcaattttc cgttcacagt 540

caattaaaag ctgcacaagc acaagccgaa ggtattttcg ccgacgaaat agctccattg 600

gaagtttctg gtacattagt cgaaaaggat gaaggtatta gacctaactc cagtgttgaa 660

aaattgggta ctttgaagac agtattcaag gaagacggta cagttaccgc tggtaatgcc 720

tctaccatta acgatggtgc tagtgcattg attatagctt ctcaagaata tgccgaagct 780

catggtttgc catacttagc tatcattaga gatagtgtag aagttggtat tgacccagca 840

tacatgggta tctctcctat aaaagcaatc caaaagttgt tagccagaaa ccaattgacc 900

actgaagaaa ttgatttgta cgaaattaac gaagcatttg ccgctacatc aatcgttgtc 960

caaagagaat tggcattgcc agaagaaaag gttaacatct atggtggtgg tatctccttg 1020

ggtcacgcta taggtgcaac cggtgccaga ttgttgactt ccttaagtta ccaattgaac 1080

caaaaggaaa agaaatacgg tgttgcttct ttatgcattg gtggtggttt gggtttagca 1140

atgttgttag aaagaccaca acaaaagaaa aattctagat tctaccaaat gtcccctgaa 1200

gaaagattgg cctcattgtt aaatgaaggt caaatttccg cagatactaa gaaagaattt 1260

gaaaacaccg ctttatcttc acaaatcgca aaccatatga tcgaaaacca aatctctgaa 1320

acagaagttc caatgggtgt cggtttgcac ttaactgtcg atgaaacaga ctatttggta 1380

ccaatggcta ccgaagaacc tagtgttatc gcagccttat ctaatggtgc taagatagca 1440

caaggtttta agactgttaa ccaacaaaga ttgatgagag gtcaaatcgt attctacgat 1500

gttgctgacc cagaatcatt aatcgataag ttgcaagtaa gagaagccga agtttttcaa 1560

caagctgaat tgtcttaccc ttcaatagtt aagagaggtg gtggtttgag agatttgcaa 1620

tacagaactt ttgacgaatc cttcgtcagt gtagatttct tagttgatgt caaggacgcc 1680

atgggtgcta atattgttaa cgcaatgttg gaaggtgtcg ccgaattgtt tagagaatgg 1740

ttcgctgaac aaaagatttt gttttctatc ttgtcaaact acgctacaga atctgtagtt 1800

accatgaaaa ctgcaattcc agtttccaga ttgagtaagg gttctaacgg tagagaaatc 1860

gctgaaaaga ttgttttggc atcaagatat gcctccttag acccttacag agctgttact 1920

cataataagg gtataatgaa cggtatcgaa gctgtcgtat tagcaaccgg taatgatact 1980

agagcagtat ctgcctcatg tcacgcattc gccgttaagg aaggtagata ccaaggtttg 2040

acatcatgga ccttggatgg tgaacaatta attggtgaaa tatccgttcc attggcttta 2100

gcaactgttg gtggtgctac aaaagtcttg cctaagagtc aagctgcagc cgatttgtta 2160

gccgtcactg acgctaagga attgtctaga gttgtcgctg cagtaggttt agctcaaaat 2220

ttggccgctt taagagcatt ggtttcagaa ggtattcaaa aaggtcatat ggctttgcaa 2280

gcaagatcct tagccatgac agttggtgct accggtaaag aagtcgaagc cgtagctcaa 2340

caattaaaaa gacaaaagac aatgaaccaa gacagagcaa tggctatatt aaacgatttg 2400

agaaagcaat aa 2412

<210> 54
<211> 803
<212> PRT
<213> Enterococcus sp.

<400> 54

Met Lys Thr Val Val Ile Ile Asp Ala Leu Arg Thr Pro Ile Gly Lys
1 5 10 15

Tyr Lys Gly Ser Leu Ser Gln Val Ser Ala Val Asp Leu Gly Thr His
20 25 30

Val Thr Thr Gln Leu Leu Lys Arg His Ser Thr Ile Ser Glu Glu Ile
35 40 45

Asp Gln Val Ile Phe Gly Asn Val Leu Gln Ala Gly Asn Gly Gln Asn
50 55 60

Pro Ala Arg Gln Ile Ala Ile Asn Ser Gly Leu Ser His Glu Ile Pro
65 70 75 80

Ala Met Thr Val Asn Glu Val Cys Gly Ser Gly Met Lys Ala Val Ile
85 90 95

Leu Ala Lys Gln Leu Ile Gln Leu Gly Glu Ala Glu Val Leu Ile Ala
100 105 110

Gly Gly Ile Glu Asn Met Ser Gln Ala Pro Lys Leu Gln Arg Phe Asn
115 120 125

```
Tyr Glu Thr Glu Ser Tyr Asp Ala Pro Phe Ser Ser Met Met Tyr Asp
    130                 135                 140

Gly Leu Thr Asp Ala Phe Ser Gly Gln Ala Met Gly Leu Thr Ala Glu
145                 150                 155                 160

Asn Val Ala Glu Lys Tyr His Val Thr Arg Glu Glu Gln Asp Gln Phe
                165                 170                 175

Ser Val His Ser Gln Leu Lys Ala Ala Gln Ala Gln Ala Glu Gly Ile
            180                 185                 190

Phe Ala Asp Glu Ile Ala Pro Leu Glu Val Ser Gly Thr Leu Val Glu
        195                 200                 205

Lys Asp Glu Gly Ile Arg Pro Asn Ser Ser Val Glu Lys Leu Gly Thr
    210                 215                 220

Leu Lys Thr Val Phe Lys Glu Asp Gly Thr Val Thr Ala Gly Asn Ala
225                 230                 235                 240

Ser Thr Ile Asn Asp Gly Ala Ser Ala Leu Ile Ile Ala Ser Gln Glu
                245                 250                 255

Tyr Ala Glu Ala His Gly Leu Pro Tyr Leu Ala Ile Ile Arg Asp Ser
            260                 265                 270

Val Glu Val Gly Ile Asp Pro Ala Tyr Met Gly Ile Ser Pro Ile Lys
        275                 280                 285

Ala Ile Gln Lys Leu Leu Ala Arg Asn Gln Leu Thr Thr Glu Glu Ile
    290                 295                 300

Asp Leu Tyr Glu Ile Asn Glu Ala Phe Ala Ala Thr Ser Ile Val Val
305                 310                 315                 320

Gln Arg Glu Leu Ala Leu Pro Glu Glu Lys Val Asn Ile Tyr Gly Gly
                325                 330                 335
```

```
Gly Ile Ser Leu Gly His Ala Ile Gly Ala Thr Gly Ala Arg Leu Leu
            340                 345                 350

Thr Ser Leu Ser Tyr Gln Leu Asn Gln Lys Glu Lys Lys Tyr Gly Val
        355                 360                 365

Ala Ser Leu Cys Ile Gly Gly Gly Leu Gly Leu Ala Met Leu Leu Glu
    370                 375                 380

Arg Pro Gln Gln Lys Lys Asn Ser Arg Phe Tyr Gln Met Ser Pro Glu
385                 390                 395                 400

Glu Arg Leu Ala Ser Leu Leu Asn Glu Gly Gln Ile Ser Ala Asp Thr
                405                 410                 415

Lys Lys Glu Phe Glu Asn Thr Ala Leu Ser Ser Gln Ile Ala Asn His
            420                 425                 430

Met Ile Glu Asn Gln Ile Ser Glu Thr Glu Val Pro Met Gly Val Gly
        435                 440                 445

Leu His Leu Thr Val Asp Glu Thr Asp Tyr Leu Val Pro Met Ala Thr
    450                 455                 460

Glu Glu Pro Ser Val Ile Ala Ala Leu Ser Asn Gly Ala Lys Ile Ala
465                 470                 475                 480

Gln Gly Phe Lys Thr Val Asn Gln Gln Arg Leu Met Arg Gly Gln Ile
                485                 490                 495

Val Phe Tyr Asp Val Ala Asp Pro Glu Ser Leu Ile Asp Lys Leu Gln
            500                 505                 510

Val Arg Glu Ala Glu Val Phe Gln Gln Ala Glu Leu Ser Tyr Pro Ser
        515                 520                 525

Ile Val Lys Arg Gly Gly Gly Leu Arg Asp Leu Gln Tyr Arg Thr Phe
    530                 535                 540
```

```
Asp Glu Ser Phe Val Ser Val Asp Phe Leu Val Asp Val Lys Asp Ala
545                 550                 555                 560


Met Gly Ala Asn Ile Val Asn Ala Met Leu Glu Gly Val Ala Glu Leu
                565                 570                 575


Phe Arg Glu Trp Phe Ala Glu Gln Lys Ile Leu Phe Ser Ile Leu Ser
            580                 585                 590


Asn Tyr Ala Thr Glu Ser Val Val Thr Met Lys Thr Ala Ile Pro Val
        595                 600                 605


Ser Arg Leu Ser Lys Gly Ser Asn Gly Arg Glu Ile Ala Glu Lys Ile
    610                 615                 620


Val Leu Ala Ser Arg Tyr Ala Ser Leu Asp Pro Tyr Arg Ala Val Thr
625                 630                 635                 640


His Asn Lys Gly Ile Met Asn Gly Ile Glu Ala Val Val Leu Ala Thr
                645                 650                 655


Gly Asn Asp Thr Arg Ala Val Ser Ala Ser Cys His Ala Phe Ala Val
            660                 665                 670


Lys Glu Gly Arg Tyr Gln Gly Leu Thr Ser Trp Thr Leu Asp Gly Glu
        675                 680                 685


Gln Leu Ile Gly Glu Ile Ser Val Pro Leu Ala Leu Ala Thr Val Gly
    690                 695                 700


Gly Ala Thr Lys Val Leu Pro Lys Ser Gln Ala Ala Ala Asp Leu Leu
705                 710                 715                 720


Ala Val Thr Asp Ala Lys Glu Leu Ser Arg Val Val Ala Ala Val Gly
                725                 730                 735


Leu Ala Gln Asn Leu Ala Ala Leu Arg Ala Leu Val Ser Glu Gly Ile
            740                 745                 750
```

Gln Lys Gly His Met Ala Leu Gln Ala Arg Ser Leu Ala Met Thr Val
755 760 765

Gly Ala Thr Gly Lys Glu Val Glu Ala Val Ala Gln Gln Leu Lys Arg
770 775 780

Gln Lys Thr Met Asn Gln Asp Arg Ala Met Ala Ile Leu Asn Asp Leu
785 790 795 800

Arg Lys Gln

<210> 55
<211> 1152
<212> DNA
<213> Enterococcus sp.

<400> 55
atgacaattg ggattgataa aattagtttt tttgtgcccc cttattatat tgatatgacg 60

gcactggctg aagccagaaa tgtagaccct ggaaaatttc atattggtat tgggcaagac 120

caaatggcgg tgaacccaat cagccaagat attgtgacat ttgcagccaa tgccgcagaa 180

gcgatcttga ccaaagaaga taaagaggcc attgatatgg tgattgtcgg gactgagtcc 240

agtatcgatg agtcaaaagc ggccgcagtt gtcttacatc gtttaatggg gattcaacct 300

ttcgctcgct ctttcgaaat caaggaagct tgttacggag caacagcagg cttacagtta 360

gctaagaatc acgtagcctt acatccagat aaaaaagtct tggtcgtagc ggcagatatt 420

gcaaaatatg gcttaaattc tggcggtgag cctacacaag gagctggggc ggttgcaatg 480

ttagttgcta gtgaaccgcg cattttggct ttaaaagagg ataatgtgat gctgacgcaa 540

gatatctatg acttttggcg tccaacaggc cacccgtatc ctatggtcga tggtcctttg 600

tcaaacgaaa cctacatcca atcttttgcc caagtctggg atgaacataa aaaacgaacc 660

ggtcttgatt ttgcagatta tgatgcttta gcgttccata ttccttacac aaaaatgggc 720

aaaaaagcct tattagcaaa aatctccgac caaactgaag cagaacagga acgaatttta 780

gcccgttatg aagaaagtat cgtctatagt cgtcgcgtag gaaacttgta tacgggttca 840

ctttatctgg gactcatttc ccttttagaa aatgcaacga ctttaaccgc aggcaatcaa 900

attggtttat tcagttatgg ttctggtgct gtcgctgaat ttttcactgg tgaattagta 960

gctggttatc aaaatcattt acaaaaagaa actcatttag cactgctgga taatcggaca 1020

gaactttcta tcgctgaata tgaagccatg tttgcagaaa ctttagacac agacattgat 1080

caaacgttag aagatgaatt aaaatatagt atttctgcta ttaataatac cgttcgttct 1140

tatcgaaact aa 1152

<210> 56
<211> 383
<212> PRT
<213> Enterococcus sp.

<400> 56

Met Thr Ile Gly Ile Asp Lys Ile Ser Phe Phe Val Pro Pro Tyr Tyr
1 5 10 15

Ile Asp Met Thr Ala Leu Ala Glu Ala Arg Asn Val Asp Pro Gly Lys
20 25 30

Phe His Ile Gly Ile Gly Gln Asp Gln Met Ala Val Asn Pro Ile Ser
35 40 45

Gln Asp Ile Val Thr Phe Ala Ala Asn Ala Ala Glu Ala Ile Leu Thr
50 55 60

Lys Glu Asp Lys Glu Ala Ile Asp Met Val Ile Val Gly Thr Glu Ser
65 70 75 80

Ser Ile Asp Glu Ser Lys Ala Ala Ala Val Val Leu His Arg Leu Met
85 90 95

Gly Ile Gln Pro Phe Ala Arg Ser Phe Glu Ile Lys Glu Ala Cys Tyr
100 105 110

Gly Ala Thr Ala Gly Leu Gln Leu Ala Lys Asn His Val Ala Leu His
115 120 125

Pro Asp Lys Lys Val Leu Val Val Ala Ala Asp Ile Ala Lys Tyr Gly
130 135 140

```
Leu Asn Ser Gly Gly Glu Pro Thr Gln Gly Ala Gly Ala Val Ala Met
145                 150                 155                 160

Leu Val Ala Ser Glu Pro Arg Ile Leu Ala Leu Lys Glu Asp Asn Val
                165                 170                 175

Met Leu Thr Gln Asp Ile Tyr Asp Phe Trp Arg Pro Thr Gly His Pro
            180                 185                 190

Tyr Pro Met Val Asp Gly Pro Leu Ser Asn Glu Thr Tyr Ile Gln Ser
        195                 200                 205

Phe Ala Gln Val Trp Asp Glu His Lys Lys Arg Thr Gly Leu Asp Phe
    210                 215                 220

Ala Asp Tyr Asp Ala Leu Ala Phe His Ile Pro Tyr Thr Lys Met Gly
225                 230                 235                 240

Lys Lys Ala Leu Leu Ala Lys Ile Ser Asp Gln Thr Glu Ala Glu Gln
                245                 250                 255

Glu Arg Ile Leu Ala Arg Tyr Glu Glu Ser Ile Val Tyr Ser Arg Arg
            260                 265                 270

Val Gly Asn Leu Tyr Thr Gly Ser Leu Tyr Leu Gly Leu Ile Ser Leu
        275                 280                 285

Leu Glu Asn Ala Thr Thr Leu Thr Ala Gly Asn Gln Ile Gly Leu Phe
    290                 295                 300

Ser Tyr Gly Ser Gly Ala Val Ala Glu Phe Phe Thr Gly Glu Leu Val
305                 310                 315                 320

Ala Gly Tyr Gln Asn His Leu Gln Lys Glu Thr His Leu Ala Leu Leu
                325                 330                 335

Asp Asn Arg Thr Glu Leu Ser Ile Ala Glu Tyr Glu Ala Met Phe Ala
            340                 345                 350
```

Glu Thr Leu Asp Thr Asp Ile Asp Gln Thr Leu Glu Asp Glu Leu Lys
355 360 365

Tyr Ser Ile Ser Ala Ile Asn Asn Thr Val Arg Ser Tyr Arg Asn
370 375 380

<210> 57
<211> 867
<212> DNA
<213> Saccharomyces sp.

<400> 57
atgactgccg acaacaatag tatgccccat ggtgcagtat ctagttacgc caaattagtg 60

caaaaccaaa cacctgaaga cattttggaa gagtttcctg aaattattcc attacaacaa 120

agacctaata cccgatctag tgagacgtca aatgacgaaa gcggagaaac atgtttttct 180

ggtcatgatg aggagcaaat taagttaatg aatgaaaatt gtattgtttt ggattgggac 240

gataatgcta ttggtgccgg taccaagaaa gtttgtcatt taatggaaaa tattgaaaag 300

ggtttactac atcgtgcatt ctccgtcttt attttcaatg aacaaggtga attactttta 360

caacaaagag ccactgaaaa aataactttc cctgatcttt ggactaacac atgctgctct 420

catccactat gtattgatga cgaattaggt ttgaagggta agctagacga taagattaag 480

ggcgctatta ctgcggcggt gagaaaacta gatcatgaat taggtattcc agaagatgaa 540

actaagacaa ggggtaagtt tcacttttta aacagaatcc attacatggc accaagcaat 600

gaaccatggg gtgaacatga aattgattac atcctatttt ataagatcaa cgctaaagaa 660

aacttgactg tcaacccaaa cgtcaatgaa gttagagact tcaaatgggt ttcaccaaat 720

gatttgaaaa ctatgtttgc tgacccaagt tacaagttta cgccttggtt taagattatt 780

tgcgagaatt acttattcaa ctggtgggag caattagatg acctttctga agtggaaaat 840

gacaggcaaa ttcatagaat gctataa 867

<210> 58
<211> 288
<212> PRT
<213> Saccharomyces sp.

<400> 58

```
Met Thr Ala Asp Asn Asn Ser Met Pro His Gly Ala Val Ser Ser Tyr
1               5                  10                  15

Ala Lys Leu Val Gln Asn Gln Thr Pro Glu Asp Ile Leu Glu Glu Phe
            20                  25                  30

Pro Glu Ile Ile Pro Leu Gln Gln Arg Pro Asn Thr Arg Ser Ser Glu
        35                  40                  45

Thr Ser Asn Asp Glu Ser Gly Glu Thr Cys Phe Ser Gly His Asp Glu
    50                  55                  60

Glu Gln Ile Lys Leu Met Asn Glu Asn Cys Ile Val Leu Asp Trp Asp
65                  70                  75                  80

Asp Asn Ala Ile Gly Ala Gly Thr Lys Lys Val Cys His Leu Met Glu
                85                  90                  95

Asn Ile Glu Lys Gly Leu Leu His Arg Ala Phe Ser Val Phe Ile Phe
            100                 105                 110

Asn Glu Gln Gly Glu Leu Leu Leu Gln Gln Arg Ala Thr Glu Lys Ile
        115                 120                 125

Thr Phe Pro Asp Leu Trp Thr Asn Thr Cys Cys Ser His Pro Leu Cys
    130                 135                 140

Ile Asp Asp Glu Leu Gly Leu Lys Gly Lys Leu Asp Asp Lys Ile Lys
145                 150                 155                 160

Gly Ala Ile Thr Ala Ala Val Arg Lys Leu Asp His Glu Leu Gly Ile
                165                 170                 175

Pro Glu Asp Glu Thr Lys Thr Arg Gly Lys Phe His Phe Leu Asn Arg
            180                 185                 190

Ile His Tyr Met Ala Pro Ser Asn Glu Pro Trp Gly Glu His Glu Ile
        195                 200                 205
```

```
Asp Tyr Ile Leu Phe Tyr Lys Ile Asn Ala Lys Glu Asn Leu Thr Val
    210                 215                 220

Asn Pro Asn Val Asn Glu Val Arg Asp Phe Lys Trp Val Ser Pro Asn
225                 230                 235                 240

Asp Leu Lys Thr Met Phe Ala Asp Pro Ser Tyr Lys Phe Thr Pro Trp
                245                 250                 255

Phe Lys Ile Ile Cys Glu Asn Tyr Leu Phe Asn Trp Trp Glu Gln Leu
            260                 265                 270

Asp Asp Leu Ser Glu Val Glu Asn Asp Arg Gln Ile His Arg Met Leu
        275                 280                 285
```

<210> 59
<211> 1059
<212> DNA
<213> Artificial Sequence

<220>
<223> Mutant farnesyl pyrophosphate synthase (Erg20mut, F96W, N127W)

<400> 59

```
atggcttcag aaaaagaaat taggagagag agattcttga acgttttccc taaattagta       60

gaggaattga acgcatcgct tttggcttac ggtatgccta aggaagcatg tgactggtat      120

gcccactcat tgaactacaa cactccaggc ggtaagctaa atagaggttt gtccgttgtg      180

gacacgtatg ctattctctc caacaagacc gttgaacaat tggggcaaga agaatacgaa      240

aaggttgcca ttctaggttg gtgcattgag ttgttgcagg cttactggtt ggtcgccgat      300

gatatgatgg acaagtccat taccagaaga ggccaaccat gttggtacaa ggttcctgaa      360

gttggggaaa ttgccatctg ggacgcattc atgttagagg ctgctatcta caagcttttg      420

aaatctcact tcagaaacga aaaatactac atagatatca ccgaattgtt ccatgaggtc      480

accttccaaa ccgaattggg ccaattgatg gacttaatca ctgcacctga agacaaagtc      540

gacttgagta agttctccct aaagaagcac tccttcatag ttactttcaa gactgcttac      600

tattctttct acttgcctgt cgcattggcc atgtacgttg ccggtatcac ggatgaaaag      660
```

gatttgaaac aagccagaga tgtcttgatt ccattgggtg aatacttcca aattcaagat 720

gactacttag actgcttcgg taccccagaa cagatcggta agatcggtac agatatccaa 780

gataacaaat gttcttgggt aatcaacaag gcattggaac ttgcttccgc agaacaaaga 840

aagactttag acgaaaatta cggtaagaag gactcagtcg cagaagccaa atgcaaaaag 900

attttcaatg acttgaaaat tgaacagcta taccacgaat atgaagagtc tattgccaag 960

gatttgaagg ccaaaatttc tcaggtcgat gagtctcgtg gcttcaaagc tgatgtctta 1020

actgcgttct tgaacaaagt ttacaagaga agcaaatag 1059

<210> 60
<211> 352
<212> PRT
<213> Artificial Sequence

<220>
<223> Mutant farnesyl pyrophosphate synthase (Erg20mut, F96W, N127W)

<400> 60

Met Ala Ser Glu Lys Glu Ile Arg Arg Glu Arg Phe Leu Asn Val Phe
1 5 10 15

Pro Lys Leu Val Glu Glu Leu Asn Ala Ser Leu Leu Ala Tyr Gly Met
20 25 30

Pro Lys Glu Ala Cys Asp Trp Tyr Ala His Ser Leu Asn Tyr Asn Thr
35 40 45

Pro Gly Gly Lys Leu Asn Arg Gly Leu Ser Val Val Asp Thr Tyr Ala
50 55 60

Ile Leu Ser Asn Lys Thr Val Glu Gln Leu Gly Gln Glu Glu Tyr Glu
65 70 75 80

Lys Val Ala Ile Leu Gly Trp Cys Ile Glu Leu Leu Gln Ala Tyr Trp
85 90 95

Leu Val Ala Asp Asp Met Met Asp Lys Ser Ile Thr Arg Arg Gly Gln
100 105 110

```
Pro Cys Trp Tyr Lys Val Pro Glu Val Gly Glu Ile Ala Ile Trp Asp
        115                 120                 125

Ala Phe Met Leu Glu Ala Ala Ile Tyr Lys Leu Leu Lys Ser His Phe
    130                 135                 140

Arg Asn Glu Lys Tyr Tyr Ile Asp Ile Thr Glu Leu Phe His Glu Val
145                 150                 155                 160

Thr Phe Gln Thr Glu Leu Gly Gln Leu Met Asp Leu Ile Thr Ala Pro
                165                 170                 175

Glu Asp Lys Val Asp Leu Ser Lys Phe Ser Leu Lys Lys His Ser Phe
            180                 185                 190

Ile Val Thr Phe Lys Thr Ala Tyr Tyr Ser Phe Tyr Leu Pro Val Ala
        195                 200                 205

Leu Ala Met Tyr Val Ala Gly Ile Thr Asp Glu Lys Asp Leu Lys Gln
    210                 215                 220

Ala Arg Asp Val Leu Ile Pro Leu Gly Glu Tyr Phe Gln Ile Gln Asp
225                 230                 235                 240

Asp Tyr Leu Asp Cys Phe Gly Thr Pro Glu Gln Ile Gly Lys Ile Gly
                245                 250                 255

Thr Asp Ile Gln Asp Asn Lys Cys Ser Trp Val Ile Asn Lys Ala Leu
            260                 265                 270

Glu Leu Ala Ser Ala Glu Gln Arg Lys Thr Leu Asp Glu Asn Tyr Gly
        275                 280                 285

Lys Lys Asp Ser Val Ala Glu Ala Lys Cys Lys Lys Ile Phe Asn Asp
    290                 295                 300

Leu Lys Ile Glu Gln Leu Tyr His Glu Tyr Glu Glu Ser Ile Ala Lys
305                 310                 315                 320
```

Asp Leu Lys Ala Lys Ile Ser Gln Val Asp Glu Ser Arg Gly Phe Lys
325 330 335

Ala Asp Val Leu Thr Ala Phe Leu Asn Lys Val Tyr Lys Arg Ser Lys
340 345 350

<210> 61
<211> 1356
<212> DNA
<213> Saccharomyces sp.

<400> 61
atgtcagagt tgagagcctt cagtgcccca gggaaagcgt tactagctgg tggatattta 60

gttttagata caaaatatga agcatttgta gtcggattat cggcaagaat gcatgctgta 120

gcccatcctt acggttcatt gcaagggtct gataagtttg aagtgcgtgt gaaaagtaaa 180

caatttaaag atggggagtg gctgtaccat ataagtccta aaagtggctt cattcctgtt 240

tcgataggcg gatctaagaa ccctttcatt gaaaaagtta tcgctaacgt atttagctac 300

tttaaaccta acatggacga ctactgcaat agaaacttgt tcgttattga tattttctct 360

gatgatgcct accattctca ggaggatagc gttaccgaac atcgtggcaa cagaagattg 420

agttttcatt cgcacagaat tgaagaagtt cccaaaacag ggctgggctc ctcggcaggt 480

ttagtcacag ttttaactac agctttggcc tccttttttg tatcggacct ggaaaataat 540

gtagacaaat atagagaagt tattcataat ttagcacaag ttgctcattg tcaagctcag 600

ggtaaaattg gaagcgggtt tgatgtagcg gcggcagcat atggatctat cagatataga 660

agattcccac ccgcattaat ctctaatttg ccagatattg gaagtgctac ttacggcagt 720

aaactggcgc atttggttga tgaagaagac tggaatatta cgattaaaag taaccattta 780

ccttcgggat taactttatg gatgggcgat attaagaatg gttcagaaac agtaaaactg 840

gtccagaagg taaaaaattg gtatgattcg catatgccag aaagcttgaa aatatataca 900

gaactcgatc atgcaaattc tagatttatg gatggactat ctaaactaga tcgcttacac 960

gagactcatg acgattacag cgatcagata tttgagtctc ttgagaggaa tgactgtacc 1020

tgtcaaaagt atcctgaaat cacagaagtt agagatgcag ttgccacaat tagacgttcc 1080

tttagaaaaa taactaaaga atctggtgcc gatatcgaac ctcccgtaca aactagctta 1140

ttggatgatt gccagacctt aaaaggagtt cttacttgct taatacctgg tgctggtggt 1200

tatgacgcca ttgcagtgat tactaagcaa gatgttgatc ttagggctca aaccgctaat 1260

gacaaaagat tttctaaggt tcaatggctg gatgtaactc aggctgactg gggtgttagg 1320

aaagaaaaag atccggaaac ttatcttgat aaataa 1356

<210> 62
<211> 451
<212> PRT
<213> Saccharomyces sp.

<400> 62

Met Ser Glu Leu Arg Ala Phe Ser Ala Pro Gly Lys Ala Leu Leu Ala
1 5 10 15

Gly Gly Tyr Leu Val Leu Asp Thr Lys Tyr Glu Ala Phe Val Val Gly
20 25 30

Leu Ser Ala Arg Met His Ala Val Ala His Pro Tyr Gly Ser Leu Gln
35 40 45

Gly Ser Asp Lys Phe Glu Val Arg Val Lys Ser Lys Gln Phe Lys Asp
50 55 60

Gly Glu Trp Leu Tyr His Ile Ser Pro Lys Ser Gly Phe Ile Pro Val
65 70 75 80

Ser Ile Gly Gly Ser Lys Asn Pro Phe Ile Glu Lys Val Ile Ala Asn
85 90 95

Val Phe Ser Tyr Phe Lys Pro Asn Met Asp Asp Tyr Cys Asn Arg Asn
100 105 110

Leu Phe Val Ile Asp Ile Phe Ser Asp Asp Ala Tyr His Ser Gln Glu
115 120 125

Asp Ser Val Thr Glu His Arg Gly Asn Arg Arg Leu Ser Phe His Ser
130 135 140

```
His Arg Ile Glu Glu Val Pro Lys Thr Gly Leu Gly Ser Ser Ala Gly
145                 150                 155                 160

Leu Val Thr Val Leu Thr Thr Ala Leu Ala Ser Phe Phe Val Ser Asp
                165                 170                 175

Leu Glu Asn Asn Val Asp Lys Tyr Arg Glu Val Ile His Asn Leu Ala
            180                 185                 190

Gln Val Ala His Cys Gln Ala Gln Gly Lys Ile Gly Ser Gly Phe Asp
        195                 200                 205

Val Ala Ala Ala Ala Tyr Gly Ser Ile Arg Tyr Arg Arg Phe Pro Pro
    210                 215                 220

Ala Leu Ile Ser Asn Leu Pro Asp Ile Gly Ser Ala Thr Tyr Gly Ser
225                 230                 235                 240

Lys Leu Ala His Leu Val Asp Glu Glu Asp Trp Asn Ile Thr Ile Lys
                245                 250                 255

Ser Asn His Leu Pro Ser Gly Leu Thr Leu Trp Met Gly Asp Ile Lys
            260                 265                 270

Asn Gly Ser Glu Thr Val Lys Leu Val Gln Lys Val Lys Asn Trp Tyr
        275                 280                 285

Asp Ser His Met Pro Glu Ser Leu Lys Ile Tyr Thr Glu Leu Asp His
    290                 295                 300

Ala Asn Ser Arg Phe Met Asp Gly Leu Ser Lys Leu Asp Arg Leu His
305                 310                 315                 320

Glu Thr His Asp Asp Tyr Ser Asp Gln Ile Phe Glu Ser Leu Glu Arg
                325                 330                 335

Asn Asp Cys Thr Cys Gln Lys Tyr Pro Glu Ile Thr Glu Val Arg Asp
            340                 345                 350
```

```
Ala Val Ala Thr Ile Arg Arg Ser Phe Arg Lys Ile Thr Lys Glu Ser
        355                 360                 365


Gly Ala Asp Ile Glu Pro Pro Val Gln Thr Ser Leu Leu Asp Asp Cys
    370                 375                 380


Gln Thr Leu Lys Gly Val Leu Thr Cys Leu Ile Pro Gly Ala Gly Gly
385                 390                 395                 400


Tyr Asp Ala Ile Ala Val Ile Thr Lys Gln Asp Val Asp Leu Arg Ala
                405                 410                 415


Gln Thr Ala Asn Asp Lys Arg Phe Ser Lys Val Gln Trp Leu Asp Val
            420                 425                 430


Thr Gln Ala Asp Trp Gly Val Arg Lys Glu Lys Asp Pro Glu Thr Tyr
        435                 440                 445


Leu Asp Lys
    450


<210>  63
<211>  1332
<212>  DNA
<213>  Saccharomyces sp.

<400>  63
atgtcattac cgttcttaac ttctgcaccg ggaaaggtta ttatttttgg tgaacactct       60

gctgtgtaca acaagcctgc cgtcgctgct agtgtgtctg cgttgagaac ctacctgcta      120

ataagcgagt catctgcacc agatactatt gaattggact tcccggacat tagctttaat      180

cataagtggt ccatcaatga tttcaatgcc atcaccgagg atcaagtaaa ctcccaaaaa      240

ttggccaagg ctcaacaagc caccgatggc ttgtctcagg aactcgttag tcttttggat      300

ccgttgttag ctcaactatc cgaatccttc cactaccatg cagcgttttg tttcctgtat      360

atgtttgttt gcctatgccc ccatgccaag aatattaagt tttctttaaa gtctacttta      420

cccatcggtg ctgggttggg ctcaagcgcc tctatttctg tatcactggc cttagctatg      480
```

gcctacttgg gggggttaat aggatctaat gacttggaaa agctgtcaga aaacgataag 540

catatagtga atcaatgggc cttcataggt gaaaagtgta ttcacggtac cccttcagga 600

atagataacg ctgtggccac ttatggtaat gccctgctat ttgaaaaaga ctcacataat 660

ggaacaataa acacaaacaa ttttaagttc ttagatgatt tcccagccat tccaatgatc 720

ctaacctata ctagaattcc aaggtctaca aaagatcttg ttgctcgcgt tcgtgtgttg 780

gtcaccgaga aatttcctga agttatgaag ccaattctag atgccatggg tgaatgtgcc 840

ctacaaggct tagagatcat gactaagtta agtaaatgta aaggcaccga tgacgaggct 900

gtagaaacta ataatgaact gtatgaacaa ctattggaat tgataagaat aaatcatgga 960

ctgcttgtct caatcggtgt ttctcatcct ggattagaac ttattaaaaa tctgagcgat 1020

gatttgagaa ttggctccac aaaacttacc ggtgctggtg gcggcggttg ctctttgact 1080

ttgttacgaa gagacattac tcaagagcaa attgacagct tcaaaaagaa attgcaagat 1140

gattttagtt acgagacatt tgaaacagac ttgggtggga ctggctgctg tttgttaagc 1200

gcaaaaaatt tgaataaaga tcttaaaatc aaatccctag tattccaatt atttgaaaat 1260

aaaactacca caaagcaaca aattgacgat ctattattgc caggaaacac gaatttacca 1320

tggacttcat aa 1332

<210> 64
<211> 443
<212> PRT
<213> Saccharomyces sp.

<400> 64

Met Ser Leu Pro Phe Leu Thr Ser Ala Pro Gly Lys Val Ile Ile Phe
1 5 10 15

Gly Glu His Ser Ala Val Tyr Asn Lys Pro Ala Val Ala Ala Ser Val
20 25 30

Ser Ala Leu Arg Thr Tyr Leu Leu Ile Ser Glu Ser Ser Ala Pro Asp
35 40 45

Thr Ile Glu Leu Asp Phe Pro Asp Ile Ser Phe Asn His Lys Trp Ser
50 55 60

```
Ile Asn Asp Phe Asn Ala Ile Thr Glu Asp Gln Val Asn Ser Gln Lys
65                  70                  75                  80

Leu Ala Lys Ala Gln Gln Ala Thr Asp Gly Leu Ser Gln Glu Leu Val
                85                  90                  95

Ser Leu Leu Asp Pro Leu Leu Ala Gln Leu Ser Glu Ser Phe His Tyr
            100                 105                 110

His Ala Ala Phe Cys Phe Leu Tyr Met Phe Val Cys Leu Cys Pro His
        115                 120                 125

Ala Lys Asn Ile Lys Phe Ser Leu Lys Ser Thr Leu Pro Ile Gly Ala
    130                 135                 140

Gly Leu Gly Ser Ser Ala Ser Ile Ser Val Ser Leu Ala Leu Ala Met
145                 150                 155                 160

Ala Tyr Leu Gly Gly Leu Ile Gly Ser Asn Asp Leu Glu Lys Leu Ser
                165                 170                 175

Glu Asn Asp Lys His Ile Val Asn Gln Trp Ala Phe Ile Gly Glu Lys
            180                 185                 190

Cys Ile His Gly Thr Pro Ser Gly Ile Asp Asn Ala Val Ala Thr Tyr
        195                 200                 205

Gly Asn Ala Leu Leu Phe Glu Lys Asp Ser His Asn Gly Thr Ile Asn
    210                 215                 220

Thr Asn Asn Phe Lys Phe Leu Asp Asp Phe Pro Ala Ile Pro Met Ile
225                 230                 235                 240

Leu Thr Tyr Thr Arg Ile Pro Arg Ser Thr Lys Asp Leu Val Ala Arg
                245                 250                 255

Val Arg Val Leu Val Thr Glu Lys Phe Pro Glu Val Met Lys Pro Ile
            260                 265                 270
```

```
Leu Asp Ala Met Gly Glu Cys Ala Leu Gln Gly Leu Glu Ile Met Thr
        275                 280                 285


Lys Leu Ser Lys Cys Lys Gly Thr Asp Asp Glu Ala Val Glu Thr Asn
    290                 295                 300


Asn Glu Leu Tyr Glu Gln Leu Leu Glu Leu Ile Arg Ile Asn His Gly
305                 310                 315                 320


Leu Leu Val Ser Ile Gly Val Ser His Pro Gly Leu Glu Leu Ile Lys
                325                 330                 335


Asn Leu Ser Asp Asp Leu Arg Ile Gly Ser Thr Lys Leu Thr Gly Ala
            340                 345                 350


Gly Gly Gly Gly Cys Ser Leu Thr Leu Leu Arg Arg Asp Ile Thr Gln
        355                 360                 365


Glu Gln Ile Asp Ser Phe Lys Lys Lys Leu Gln Asp Asp Phe Ser Tyr
    370                 375                 380


Glu Thr Phe Glu Thr Asp Leu Gly Gly Thr Gly Cys Cys Leu Leu Ser
385                 390                 395                 400


Ala Lys Asn Leu Asn Lys Asp Leu Lys Ile Lys Ser Leu Val Phe Gln
                405                 410                 415


Leu Phe Glu Asn Lys Thr Thr Thr Lys Gln Gln Ile Asp Asp Leu Leu
            420                 425                 430


Leu Pro Gly Asn Thr Asn Leu Pro Trp Thr Ser
        435                 440


<210>  65
<211>  1191
<212>  DNA
<213>  Saccharomyces sp.

<400>  65
```

atgaccgttt acacagcatc cgttaccgca cccgtcaaca tcgcaaccct taagtattgg 60

gggaaaaggg acacgaagtt gaatctgccc accaattcgt ccatatcagt gactttatcg 120

caagatgacc tcagaacgtt gacctctgcg gctactgcac ctgagtttga acgcgacact 180

ttgtggttaa atggagaacc acacagcatc gacaatgaaa gaactcaaaa ttgtctgcgc 240

gacctacgcc aattaagaaa ggaaatggaa tcgaaggacg cctcattgcc cacattatct 300

caatggaaac tccacattgt ctccgaaaat aactttccta cagcagctgg tttagcttcc 360

tccgctgctg gctttgctgc attggtctct gcaattgcta agttatacca attaccacag 420

tcaacttcag aaatatctag aatagcaaga aaggggtctg gttcagcttg tagatcgttg 480

tttggcggat acgtggcctg ggaaatggga aaagctgaag atggtcatga ttccatggca 540

gtacaaatcg cagacagctc tgactggcct cagatgaaag cttgtgtcct agttgtcagc 600

gatattaaaa aggatgtgag ttccactcag ggtatgcaat tgaccgtggc aacctccgaa 660

ctatttaaag aaagaattga acatgtcgta ccaaagagat ttgaagtcat gcgtaaagcc 720

attgttgaaa aagatttcgc cacctttgca aaggaaacaa tgatggattc caactctttc 780

catgccacat gtttggactc tttccctcca atattctaca tgaatgacac ttccaagcgt 840

atcatcagtt ggtgccacac cattaatcag ttttacggag aaacaatcgt tgcatacacg 900

tttgatgcag gtccaaatgc tgtgttgtac tacttagctg aaaatgagtc gaaactcttt 960

gcatttatct ataaattgtt tggctctgtt cctggatggg acaagaaatt tactactgag 1020

cagcttgagg ctttcaacca tcaatttgaa tcatctaact ttactgcacg tgaattggat 1080

cttgagttgc aaaaggatgt tgccagagtg attttaactc aagtcggttc aggcccacaa 1140

gaaacaaacg aatctttgat tgacgcaaag actggtctac caaaggaata a 1191

<210> 66
<211> 396
<212> PRT
<213> Saccharomyces sp.

<400> 66

Met Thr Val Tyr Thr Ala Ser Val Thr Ala Pro Val Asn Ile Ala Thr
1 5 10 15

```
Leu Lys Tyr Trp Gly Lys Arg Asp Thr Lys Leu Asn Leu Pro Thr Asn
            20                  25                  30

Ser Ser Ile Ser Val Thr Leu Ser Gln Asp Asp Leu Arg Thr Leu Thr
        35                  40                  45

Ser Ala Ala Thr Ala Pro Glu Phe Glu Arg Asp Thr Leu Trp Leu Asn
    50                  55                  60

Gly Glu Pro His Ser Ile Asp Asn Glu Arg Thr Gln Asn Cys Leu Arg
65                  70                  75                  80

Asp Leu Arg Gln Leu Arg Lys Glu Met Glu Ser Lys Asp Ala Ser Leu
                85                  90                  95

Pro Thr Leu Ser Gln Trp Lys Leu His Ile Val Ser Glu Asn Asn Phe
            100                 105                 110

Pro Thr Ala Ala Gly Leu Ala Ser Ser Ala Ala Gly Phe Ala Ala Leu
        115                 120                 125

Val Ser Ala Ile Ala Lys Leu Tyr Gln Leu Pro Gln Ser Thr Ser Glu
    130                 135                 140

Ile Ser Arg Ile Ala Arg Lys Gly Ser Gly Ser Ala Cys Arg Ser Leu
145                 150                 155                 160

Phe Gly Gly Tyr Val Ala Trp Glu Met Gly Lys Ala Glu Asp Gly His
                165                 170                 175

Asp Ser Met Ala Val Gln Ile Ala Asp Ser Ser Asp Trp Pro Gln Met
            180                 185                 190

Lys Ala Cys Val Leu Val Val Ser Asp Ile Lys Lys Asp Val Ser Ser
        195                 200                 205

Thr Gln Gly Met Gln Leu Thr Val Ala Thr Ser Glu Leu Phe Lys Glu
    210                 215                 220
```

```
Arg Ile Glu His Val Val Pro Lys Arg Phe Glu Val Met Arg Lys Ala
225                 230                 235                 240


Ile Val Glu Lys Asp Phe Ala Thr Phe Ala Lys Glu Thr Met Met Asp
                245                 250                 255


Ser Asn Ser Phe His Ala Thr Cys Leu Asp Ser Phe Pro Pro Ile Phe
            260                 265                 270


Tyr Met Asn Asp Thr Ser Lys Arg Ile Ile Ser Trp Cys His Thr Ile
        275                 280                 285


Asn Gln Phe Tyr Gly Glu Thr Ile Val Ala Tyr Thr Phe Asp Ala Gly
    290                 295                 300


Pro Asn Ala Val Leu Tyr Tyr Leu Ala Glu Asn Glu Ser Lys Leu Phe
305                 310                 315                 320


Ala Phe Ile Tyr Lys Leu Phe Gly Ser Val Pro Gly Trp Asp Lys Lys
                325                 330                 335


Phe Thr Thr Glu Gln Leu Glu Ala Phe Asn His Gln Phe Glu Ser Ser
            340                 345                 350


Asn Phe Thr Ala Arg Glu Leu Asp Leu Glu Leu Gln Lys Asp Val Ala
        355                 360                 365


Arg Val Ile Leu Thr Gln Val Gly Ser Gly Pro Gln Glu Thr Asn Glu
    370                 375                 380


Ser Leu Ile Asp Ala Lys Thr Gly Leu Pro Lys Glu
385                 390                 395


<210>  67
<211>  2751
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  Engineered phosphomevalonate kinase/mevalonate kinase
       (Erg8-T2A-Erg12)
```

<400> 67
atgtcagagt tgagagcctt cagtgcccca gggaaagcgt tactagctgg tggatattta 60

gttttagata caaaatatga agcatttgta gtcggattat cggcaagaat gcatgctgta 120

gcccatcctt acggttcatt gcaagggtct gataagtttg aagtgcgtgt gaaaagtaaa 180

caatttaaag atggggagtg gctgtaccat ataagtccta aaagtggctt cattcctgtt 240

tcgataggcg gatctaagaa ccctttcatt gaaaaagtta tcgctaacgt atttagctac 300

tttaaaccta acatggacga ctactgcaat agaaacttgt tcgttattga tattttctct 360

gatgatgcct accattctca ggaggatagc gttaccgaac atcgtggcaa cagaagattg 420

agttttcatt cgcacagaat tgaagaagtt cccaaaacag ggctgggctc ctcggcaggt 480

ttagtcacag ttttaactac agctttggcc tccttttttg tatcggacct ggaaaataat 540

gtagacaaat atagagaagt tattcataat ttagcacaag ttgctcattg tcaagctcag 600

ggtaaaattg gaagcgggtt tgatgtagcg gcggcagcat atggatctat cagatataga 660

agattcccac ccgcattaat ctctaatttg ccagatattg gaagtgctac ttacggcagt 720

aaactggcgc atttggttga tgaagaagac tggaatatta cgattaaaag taaccattta 780

ccttcgggat taactttatg gatgggcgat attaagaatg gttcagaaac agtaaaactg 840

gtccagaagg taaaaaattg gtatgattcg catatgccag aaagcttgaa aatatataca 900

gaactcgatc atgcaaattc tagatttatg gatggactat ctaaactaga tcgcttacac 960

gagactcatg acgattacag cgatcagata tttgagtctc ttgagaggaa tgactgtacc 1020

tgtcaaaagt atcctgaaat cacagaagtt agagatgcag ttgccacaat tagacgttcc 1080

tttagaaaaa taactaaaga atctggtgcc gatatcgaac ctcccgtaca aactagctta 1140

ttggatgatt gccagacctt aaaaggagtt cttacttgct taatacctgg tgctggtggt 1200

tatgacgcca ttgcagtgat tactaagcaa gatgttgatc ttagggctca aaccgctaat 1260

gacaaaagat tttctaaggt tcaatggctg gatgtaactc aggctgactg gggtgttagg 1320

aaagaaaaag atccggaaac ttatcttgat aaaaagcttg agggcagagg aagtcttcta 1380

acatgcggtg acgtggagga gaatcccggc cctgctagca tgtcattacc gttcttaact 1440

tctgcaccgg gaaaggttat tatttttggt gaacactctg ctgtgtacaa caagcctgcc 1500

gtcgctgcta gtgtgtctgc gttgagaacc tacctgctaa taagcgagtc atctgcacca 1560

gatactattg aattggactt cccggacatt agctttaatc ataagtggtc catcaatgat 1620

ttcaatgcca tcaccgagga tcaagtaaac tcccaaaaat tggccaaggc tcaacaagcc 1680

accgatggct tgtctcagga actcgttagt cttttggatc cgttgttagc tcaactatcc 1740

gaatccttcc actaccatgc agcgttttgt ttcctgtata tgtttgtttg cctatgcccc 1800

catgccaaga atattaagtt ttctttaaag tctactttac ccatcggtgc tgggttgggc 1860

tcaagcgcct ctatttctgt atcactggcc ttagctatgg cctacttggg ggggttaata 1920

ggatctaatg acttggaaaa gctgtcagaa aacgataagc atatagtgaa tcaatgggcc 1980

ttcataggtg aaaagtgtat tcacggtacc ccttcaggaa tagataacgc tgtggccact 2040

tatggtaatg ccctgctatt tgaaaaagac tcacataatg gaacaataaa cacaaacaat 2100

tttaagttct tagatgattt cccagccatt ccaatgatcc taacctatac tagaattcca 2160

aggtctacaa aagatcttgt tgctcgcgtt cgtgtgttgg tcaccgagaa atttcctgaa 2220

gttatgaagc caattctaga tgccatgggt gaatgtgccc tacaaggctt agagatcatg 2280

actaagttaa gtaaatgtaa aggcaccgat gacgaggctg tagaaactaa taatgaactg 2340

tatgaacaac tattggaatt gataagaata aatcatggac tgcttgtctc aatcggtgtt 2400

tctcatcctg gattagaact tattaaaaat ctgagcgatg atttgagaat tggctccaca 2460

aaacttaccg gtgctggtgg cggcggttgc tctttgactt tgttacgaag agacattact 2520

caagagcaaa ttgacagctt caaaaagaaa ttgcaagatg attttagtta cgagacattt 2580

gaaacagact tgggtgggac tggctgctgt ttgttaagcg caaaaaattt gaataaagat 2640

cttaaaatca aatccctagt attccaatta tttgaaaata aaactaccac aaagcaacaa 2700

attgacgatc tattattgcc aggaaacacg aatttaccat ggacttcata a 2751

<210> 68
<211> 916
<212> PRT
<213> Artificial Sequence

<220>
<223> Engineered phosphomevalonate kinase/mevalonate kinase
(Erg8-T2A-Erg12)

<400> 68

```
Met Ser Glu Leu Arg Ala Phe Ser Ala Pro Gly Lys Ala Leu Leu Ala
1               5                   10                  15

Gly Gly Tyr Leu Val Leu Asp Thr Lys Tyr Glu Ala Phe Val Val Gly
            20                  25                  30

Leu Ser Ala Arg Met His Ala Val Ala His Pro Tyr Gly Ser Leu Gln
        35                  40                  45

Gly Ser Asp Lys Phe Glu Val Arg Val Lys Ser Lys Gln Phe Lys Asp
    50                  55                  60

Gly Glu Trp Leu Tyr His Ile Ser Pro Lys Ser Gly Phe Ile Pro Val
65                  70                  75                  80

Ser Ile Gly Gly Ser Lys Asn Pro Phe Ile Glu Lys Val Ile Ala Asn
                85                  90                  95

Val Phe Ser Tyr Phe Lys Pro Asn Met Asp Asp Tyr Cys Asn Arg Asn
            100                 105                 110

Leu Phe Val Ile Asp Ile Phe Ser Asp Asp Ala Tyr His Ser Gln Glu
        115                 120                 125

Asp Ser Val Thr Glu His Arg Gly Asn Arg Arg Leu Ser Phe His Ser
    130                 135                 140

His Arg Ile Glu Glu Val Pro Lys Thr Gly Leu Gly Ser Ser Ala Gly
145                 150                 155                 160

Leu Val Thr Val Leu Thr Thr Ala Leu Ala Ser Phe Phe Val Ser Asp
                165                 170                 175

Leu Glu Asn Asn Val Asp Lys Tyr Arg Glu Val Ile His Asn Leu Ala
            180                 185                 190

Gln Val Ala His Cys Gln Ala Gln Gly Lys Ile Gly Ser Gly Phe Asp
        195                 200                 205
```

```
Val Ala Ala Ala Ala Tyr Gly Ser Ile Arg Tyr Arg Arg Phe Pro Pro
    210                 215                 220

Ala Leu Ile Ser Asn Leu Pro Asp Ile Gly Ser Ala Thr Tyr Gly Ser
225                 230                 235                 240

Lys Leu Ala His Leu Val Asp Glu Glu Asp Trp Asn Ile Thr Ile Lys
                245                 250                 255

Ser Asn His Leu Pro Ser Gly Leu Thr Leu Trp Met Gly Asp Ile Lys
            260                 265                 270

Asn Gly Ser Glu Thr Val Lys Leu Val Gln Lys Val Lys Asn Trp Tyr
        275                 280                 285

Asp Ser His Met Pro Glu Ser Leu Lys Ile Tyr Thr Glu Leu Asp His
    290                 295                 300

Ala Asn Ser Arg Phe Met Asp Gly Leu Ser Lys Leu Asp Arg Leu His
305                 310                 315                 320

Glu Thr His Asp Asp Tyr Ser Asp Gln Ile Phe Glu Ser Leu Glu Arg
                325                 330                 335

Asn Asp Cys Thr Cys Gln Lys Tyr Pro Glu Ile Thr Glu Val Arg Asp
            340                 345                 350

Ala Val Ala Thr Ile Arg Arg Ser Phe Arg Lys Ile Thr Lys Glu Ser
        355                 360                 365

Gly Ala Asp Ile Glu Pro Pro Val Gln Thr Ser Leu Leu Asp Asp Cys
    370                 375                 380

Gln Thr Leu Lys Gly Val Leu Thr Cys Leu Ile Pro Gly Ala Gly Gly
385                 390                 395                 400

Tyr Asp Ala Ile Ala Val Ile Thr Lys Gln Asp Val Asp Leu Arg Ala
                405                 410                 415
```

```
Gln Thr Ala Asn Asp Lys Arg Phe Ser Lys Val Gln Trp Leu Asp Val
            420                 425                 430

Thr Gln Ala Asp Trp Gly Val Arg Lys Glu Lys Asp Pro Glu Thr Tyr
        435                 440                 445

Leu Asp Lys Lys Leu Glu Gly Arg Gly Ser Leu Leu Thr Cys Gly Asp
    450                 455                 460

Val Glu Glu Asn Pro Gly Pro Ala Ser Met Ser Leu Pro Phe Leu Thr
465                 470                 475                 480

Ser Ala Pro Gly Lys Val Ile Ile Phe Gly Glu His Ser Ala Val Tyr
                485                 490                 495

Asn Lys Pro Ala Val Ala Ala Ser Val Ser Ala Leu Arg Thr Tyr Leu
            500                 505                 510

Leu Ile Ser Glu Ser Ser Ala Pro Asp Thr Ile Glu Leu Asp Phe Pro
        515                 520                 525

Asp Ile Ser Phe Asn His Lys Trp Ser Ile Asn Asp Phe Asn Ala Ile
    530                 535                 540

Thr Glu Asp Gln Val Asn Ser Gln Lys Leu Ala Lys Ala Gln Gln Ala
545                 550                 555                 560

Thr Asp Gly Leu Ser Gln Glu Leu Val Ser Leu Leu Asp Pro Leu Leu
                565                 570                 575

Ala Gln Leu Ser Glu Ser Phe His Tyr His Ala Ala Phe Cys Phe Leu
            580                 585                 590

Tyr Met Phe Val Cys Leu Cys Pro His Ala Lys Asn Ile Lys Phe Ser
        595                 600                 605

Leu Lys Ser Thr Leu Pro Ile Gly Ala Gly Leu Gly Ser Ser Ala Ser
    610                 615                 620
```

```
Ile Ser Val Ser Leu Ala Leu Ala Met Ala Tyr Leu Gly Gly Leu Ile
625                 630                 635                 640


Gly Ser Asn Asp Leu Glu Lys Leu Ser Glu Asn Asp Lys His Ile Val
                645                 650                 655


Asn Gln Trp Ala Phe Ile Gly Glu Lys Cys Ile His Gly Thr Pro Ser
            660                 665                 670


Gly Ile Asp Asn Ala Val Ala Thr Tyr Gly Asn Ala Leu Leu Phe Glu
        675                 680                 685


Lys Asp Ser His Asn Gly Thr Ile Asn Thr Asn Asn Phe Lys Phe Leu
    690                 695                 700


Asp Asp Phe Pro Ala Ile Pro Met Ile Leu Thr Tyr Thr Arg Ile Pro
705                 710                 715                 720


Arg Ser Thr Lys Asp Leu Val Ala Arg Val Arg Val Leu Val Thr Glu
                725                 730                 735


Lys Phe Pro Glu Val Met Lys Pro Ile Leu Asp Ala Met Gly Glu Cys
            740                 745                 750


Ala Leu Gln Gly Leu Glu Ile Met Thr Lys Leu Ser Lys Cys Lys Gly
        755                 760                 765


Thr Asp Asp Glu Ala Val Glu Thr Asn Asn Glu Leu Tyr Glu Gln Leu
    770                 775                 780


Leu Glu Leu Ile Arg Ile Asn His Gly Leu Leu Val Ser Ile Gly Val
785                 790                 795                 800


Ser His Pro Gly Leu Glu Leu Ile Lys Asn Leu Ser Asp Asp Leu Arg
                805                 810                 815


Ile Gly Ser Thr Lys Leu Thr Gly Ala Gly Gly Gly Gly Cys Ser Leu
            820                 825                 830
```

Thr Leu Leu Arg Arg Asp Ile Thr Gln Glu Gln Ile Asp Ser Phe Lys
835 840 845

Lys Lys Leu Gln Asp Asp Phe Ser Tyr Glu Thr Phe Glu Thr Asp Leu
850 855 860

Gly Gly Thr Gly Cys Cys Leu Leu Ser Ala Lys Asn Leu Asn Lys Asp
865 870 875 880

Leu Lys Ile Lys Ser Leu Val Phe Gln Leu Phe Glu Asn Lys Thr Thr
885 890 895

Thr Lys Gln Gln Ile Asp Asp Leu Leu Leu Pro Gly Asn Thr Asn Leu
900 905 910

Pro Trp Thr Ser
915

<210> 69
<211> 756
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial neryl pyrophosphate (NPP) synthase (NPPS) nucleotide sequence

<400> 69
atgtgctcac ttaatttgca aacggaaaag ctatgctatg aagacaatga caatgacttg 60

gacgaggaac tgatgccgaa gcacatagcg ctaatcatgg atggtaatag acgttgggca 120

aaagacaagg gcttagaagt gtacgaaggg cacaaacata taatcccgaa actaaaagaa 180

atatgtgaca tatcctccaa gttggggatt cagatcatca cagcgttcgc gttctccaca 240

gagaactgga agagatccaa ggaggaagtc gatttcctat tgcagatgtt tgaagaaatc 300

tatgacgaat ttagccgttc tggggtgaga gtgagtatca tcggatgcaa aagcgatttg 360

ccgatgaccc ttcaaaaatg tatcgcattg acagaggaaa cgacgaaagg caataaggga 420

ttacacctgg tcatagcact taactacggt gggtattacg atatcctaca agcaacgaag 480

tccattgtaa acaaggctat gaatggttta ttggacgttg aagacatcaa taaaaatctg 540

ttcgaccaag aattagaaag caaatgccct aaccctgact tgctgatcag aactggggga 600

gaacagaggg tctctaattt tcttctatgg caattggctt atactgagtt ctattttacc 660

aatactttat tccctgactt tggtgaagag gacctgaaag aagccatcat gaattttcaa 720

cagagacacc gtagattcgg aggacatact tattga 756

<210> 70
<211> 251
<212> PRT
<213> Solanum sp.

<400> 70

Met Cys Ser Leu Asn Leu Gln Thr Glu Lys Leu Cys Tyr Glu Asp Asn
1 5 10 15

Asp Asn Asp Leu Asp Glu Glu Leu Met Pro Lys His Ile Ala Leu Ile
20 25 30

Met Asp Gly Asn Arg Arg Trp Ala Lys Asp Lys Gly Leu Glu Val Tyr
35 40 45

Glu Gly His Lys His Ile Ile Pro Lys Leu Lys Glu Ile Cys Asp Ile
50 55 60

Ser Ser Lys Leu Gly Ile Gln Ile Ile Thr Ala Phe Ala Phe Ser Thr
65 70 75 80

Glu Asn Trp Lys Arg Ser Lys Glu Glu Val Asp Phe Leu Leu Gln Met
85 90 95

Phe Glu Glu Ile Tyr Asp Glu Phe Ser Arg Ser Gly Val Arg Val Ser
100 105 110

Ile Ile Gly Cys Lys Ser Asp Leu Pro Met Thr Leu Gln Lys Cys Ile
115 120 125

Ala Leu Thr Glu Glu Thr Thr Lys Gly Asn Lys Gly Leu His Leu Val
130 135 140

```
Ile Ala Leu Asn Tyr Gly Gly Tyr Tyr Asp Ile Leu Gln Ala Thr Lys
145                 150                 155                 160


Ser Ile Val Asn Lys Ala Met Asn Gly Leu Leu Asp Val Glu Asp Ile
                165                 170                 175


Asn Lys Asn Leu Phe Asp Gln Glu Leu Glu Ser Lys Cys Pro Asn Pro
            180                 185                 190


Asp Leu Leu Ile Arg Thr Gly Gly Glu Gln Arg Val Ser Asn Phe Leu
        195                 200                 205


Leu Trp Gln Leu Ala Tyr Thr Glu Phe Tyr Phe Thr Asn Thr Leu Phe
    210                 215                 220


Pro Asp Phe Gly Glu Glu Asp Leu Lys Glu Ala Ile Met Asn Phe Gln
225                 230                 235                 240


Gln Arg His Arg Arg Phe Gly Gly His Thr Tyr
                245                 250


<210>  71
<211>  1113
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  Artificial geranylgeranyl pyrophosphate synthase large subunit
       (GPPSlsu) nucleotide sequence

<400>  71
atgagcaccg tgaatctgac ctgggtgcag acgtgctcta tgttcaacca gggcgggcgt       60

tcccgttcat tgtcaacctt caacttaaat ctgtaccatc cattgaagaa aacgcctttc      120

tctatccaga cacctaagca gaaaaggcca acttcccccc tctcatctat cagtgccgta      180

ttaacggagc aggaagcagt aaaggagggt gacgaggaaa aaagcatatt taacttcaaa      240

tcttatatgg ttcagaaagc taatagcgtg aatcaggcac tagattctgc ggtgttattg      300

agagacccca ttatgataca tgaatctatg cgttactctt tgcttgcggg cggcaagcgt      360
```

gtcagaccga tgttatgctt aagtgcgtgc gagttagtag gaggtaaaga gtctgtagca 420

atgcccgcag catgtgctgt agaaatgata cacacaatgt cactgattca cgatgatctt 480

ccttgcatgg ataacgacga tcttcgtaga ggtaagccaa ccaaccacaa ggtattcggg 540

gaagacgtgg cagttttagc aggagacgcg ctactagcgt tcgcgtttga acacatggca 600

gttagcacag taggagttcc agcagcaaaa atagttaggg ctataggaga gttagcaaag 660

tccatcggta gcgagggcct tgttgccgga caggtagttg atatcgatag tgaagggttg 720

gctaacgtgg gactagaaca actggagttc atccacctac acaagacagg ggcactgctt 780

gaagcgagtg ttgtacttgg ggctattctg gggggaggaa cagatgagga ggtagaaaaa 840

ctacgtagtt ttgccaggtg tataggacta ctatttcaag ttgtagatga tatccttgac 900

gtcacgaaga gtagtcaaga gttaggaaaa acagcaggga aagatctagt tgccgataaa 960

gtaacctacc ccaggctaat gggtatcgat aaatctcgtg agttcgccga acaattaaat 1020

actgaggcta agcaacattt aagcgggttt gatcctatta aggctgcgcc gctgattgct 1080

ctagcaaact atattgcata tagacagaac tga 1113

<210> 72
<211> 370
<212> PRT
<213> Cannabis sativa

<400> 72

Met Ser Thr Val Asn Leu Thr Trp Val Gln Thr Cys Ser Met Phe Asn
1 5 10 15

Gln Gly Gly Arg Ser Arg Ser Leu Ser Thr Phe Asn Leu Asn Leu Tyr
20 25 30

His Pro Leu Lys Lys Thr Pro Phe Ser Ile Gln Thr Pro Lys Gln Lys
35 40 45

Arg Pro Thr Ser Pro Phe Ser Ser Ile Ser Ala Val Leu Thr Glu Gln
50 55 60

Glu Ala Val Lys Glu Gly Asp Glu Glu Lys Ser Ile Phe Asn Phe Lys
65 70 75 80

```
Ser Tyr Met Val Gln Lys Ala Asn Ser Val Asn Gln Ala Leu Asp Ser
                85                  90                  95

Ala Val Leu Leu Arg Asp Pro Ile Met Ile His Glu Ser Met Arg Tyr
            100                 105                 110

Ser Leu Leu Ala Gly Gly Lys Arg Val Arg Pro Met Leu Cys Leu Ser
        115                 120                 125

Ala Cys Glu Leu Val Gly Gly Lys Glu Ser Val Ala Met Pro Ala Ala
    130                 135                 140

Cys Ala Val Glu Met Ile His Thr Met Ser Leu Ile His Asp Asp Leu
145                 150                 155                 160

Pro Cys Met Asp Asn Asp Asp Leu Arg Arg Gly Lys Pro Thr Asn His
                165                 170                 175

Lys Val Phe Gly Glu Asp Val Ala Val Leu Ala Gly Asp Ala Leu Leu
            180                 185                 190

Ala Phe Ala Phe Glu His Met Ala Val Ser Thr Val Gly Val Pro Ala
        195                 200                 205

Ala Lys Ile Val Arg Ala Ile Gly Glu Leu Ala Lys Ser Ile Gly Ser
    210                 215                 220

Glu Gly Leu Val Ala Gly Gln Val Val Asp Ile Asp Ser Glu Gly Leu
225                 230                 235                 240

Ala Asn Val Gly Leu Glu Gln Leu Glu Phe Ile His Leu His Lys Thr
                245                 250                 255

Gly Ala Leu Leu Glu Ala Ser Val Val Leu Gly Ala Ile Leu Gly Gly
            260                 265                 270

Gly Thr Asp Glu Glu Val Glu Lys Leu Arg Ser Phe Ala Arg Cys Ile
        275                 280                 285
```

```
Gly Leu Leu Phe Gln Val Val Asp Asp Ile Leu Asp Val Thr Lys Ser
    290                 295                 300

Ser Gln Glu Leu Gly Lys Thr Ala Gly Lys Asp Leu Val Ala Asp Lys
305                 310                 315                 320

Val Thr Tyr Pro Arg Leu Met Gly Ile Asp Lys Ser Arg Glu Phe Ala
                325                 330                 335

Glu Gln Leu Asn Thr Glu Ala Lys Gln His Leu Ser Gly Phe Asp Pro
            340                 345                 350

Ile Lys Ala Ala Pro Leu Ile Ala Leu Ala Asn Tyr Ile Ala Tyr Arg
        355                 360                 365

Gln Asn
    370
```

<210> 73
<211> 948
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial geranylgeranyl pyrophosphate synthase small subunit (GPPSssu) nucleotide sequence

<400> 73
atggctgttt acaacctttc aatcaactgt tctcccagat tcgtccatca tgtatacgtg 60

ccccatttta catgtaaatc aaataagagc ctgagccatg tccccatgag aatcacgatg 120

tcaaagcagc atcatcactc atactttgcc tctacaacgg cagatgtcga tgcccatcta 180

aaacaatcaa tcacaattaa acccccgttg tctgtccacg aagccatgta taactttatc 240

ttcagtacgc caccgaattt ggcgccatca ttatgtgtcg cagcatgtga attggttggg 300

ggtcaccagg gacaggcgat ggcagcggcc agcgcattaa gggtaataca tgctagcatc 360

gttacccacg atcaccttcc gttaacggga aggccaaacc ccacctcacc ggaggccgct 420

acgcacaatt cctataatcc aaacatacag ttgttattac ctgacgccat tacacccttc 480

gggtttgagc tattagcgtc cagtgatgat cttacacaca acaagagtga gagagttctt 540

agggtgatcg ttgaatttac gaggactttc ggttccagag gcactataga cgcccaatac 600

cacgaaaagt tggctagtag gtttgatgtg gatagccatg aggcaaagac cgtaggatgg 660

gggcattacc catcattgaa aaaggaggga gccatgcacg catgtgctgc tgcctgcgga 720

gcaatattgg gtgaggctca tgaagaagaa gtggaaaaat tgcgtacatt cgggctgtat 780

gtcggcatga tccaaggtta tgcgaacaga ttcatcatga gcagtacaga ggagaaaaaa 840

gaggctgaca ggataattga ggagcttacc aatttagcgc gtcaggagct gaaatacttc 900

gatggaagga acctagaacc gttttcaaca ttcttgttcc gtttgtag 948

<210> 74
<211> 315
<212> PRT
<213> Cannabis sativa

<400> 74

Met Ala Val Tyr Asn Leu Ser Ile Asn Cys Ser Pro Arg Phe Val His
1 5 10 15

His Val Tyr Val Pro His Phe Thr Cys Lys Ser Asn Lys Ser Leu Ser
20 25 30

His Val Pro Met Arg Ile Thr Met Ser Lys Gln His His His Ser Tyr
35 40 45

Phe Ala Ser Thr Thr Ala Asp Val Asp Ala His Leu Lys Gln Ser Ile
50 55 60

Thr Ile Lys Pro Pro Leu Ser Val His Glu Ala Met Tyr Asn Phe Ile
65 70 75 80

Phe Ser Thr Pro Pro Asn Leu Ala Pro Ser Leu Cys Val Ala Ala Cys
85 90 95

Glu Leu Val Gly Gly His Gln Gly Gln Ala Met Ala Ala Ala Ser Ala
100 105 110

```
Leu Arg Val Ile His Ala Ser Ile Val Thr His Asp His Leu Pro Leu
        115                 120                 125

Thr Gly Arg Pro Asn Pro Thr Ser Pro Glu Ala Ala Thr His Asn Ser
    130                 135                 140

Tyr Asn Pro Asn Ile Gln Leu Leu Leu Pro Asp Ala Ile Thr Pro Phe
145                 150                 155                 160

Gly Phe Glu Leu Leu Ala Ser Ser Asp Asp Leu Thr His Asn Lys Ser
                165                 170                 175

Glu Arg Val Leu Arg Val Ile Val Glu Phe Thr Arg Thr Phe Gly Ser
            180                 185                 190

Arg Gly Thr Ile Asp Ala Gln Tyr His Glu Lys Leu Ala Ser Arg Phe
        195                 200                 205

Asp Val Asp Ser His Glu Ala Lys Thr Val Gly Trp Gly His Tyr Pro
    210                 215                 220

Ser Leu Lys Lys Glu Gly Ala Met His Ala Cys Ala Ala Ala Cys Gly
225                 230                 235                 240

Ala Ile Leu Gly Glu Ala His Glu Glu Glu Val Glu Lys Leu Arg Thr
                245                 250                 255

Phe Gly Leu Tyr Val Gly Met Ile Gln Gly Tyr Ala Asn Arg Phe Ile
            260                 265                 270

Met Ser Ser Thr Glu Glu Lys Lys Glu Ala Asp Arg Ile Ile Glu Glu
        275                 280                 285

Leu Thr Asn Leu Ala Arg Gln Glu Leu Lys Tyr Phe Asp Gly Arg Asn
    290                 295                 300

Leu Glu Pro Phe Ser Thr Phe Leu Phe Arg Leu
305                 310                 315
```

<210> 75
<211> 1161
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial tetraketide synthase (TKS) nucleotide sequence

<400> 75
atgaatcatt taagagctga aggtccagcc tccgttttgg ccatcggtac cgctaaccct 60

gaaaacattt tgttgcaaga cgaattccca gactactact tcagagtcac taagtccgaa 120

cacatgaccc aattgaagga gaagttcaga aagatttgtg acaagtccat gattagaaag 180

agaaactgtt tcttgaacga agaacacttg aagcaaaacc caagattggt tgaacatgaa 240

atgcaaactt tggacgctag acaagacatg ttggttgttg aagtccctaa gttgggtaag 300

gatgcctgtg ctaaggccat taaagaatgg ggtcaaccta agtccaagat tacccacttg 360

attttcacct ctgcctccac cactgacatg cctggtgctg attaccactg cgctaagtta 420

ttgggtttgt ctccatccgt taagagagtt atgatgtacc aattgggttg ctacggtggt 480

ggtactgttt taagaattgc taaggatatt gctgaaaaca acaagggtgc cagagtctta 540

gctgtctgct gtgacattat ggcttgttta ttcagaggtc catctgaatc cgacttggaa 600

ttgttggttg gtcaagctat cttcggtgac ggtgctgctg ccgttattgt tggtgctgaa 660

ccagacgaat ccgttggtga aagaccaatt tttgaattgg tttccaccgg tcaaactatt 720

ttgccaaatt ccgaaggtac catcggtggt catatcagag aagccggttt gatcttcgac 780

ttacataagg atgtcccaat gttgatctct aacaacattg aaaagtgttt gatcgaagct 840

tttaccccaa ttggtatttc tgactggaac tctatcttct ggattaccca tcctggtggt 900

aaggctattt tggataaggt cgaggaaaaa ttgcacttga agtctgacaa gttcgttgac 960

tctagacacg tcttgtccga acatggtaat atgtcctctt ccaccgtttt attcgttatg 1020

gatgagttga gaaagagatc cttagaagaa ggtaagtcca ccaccggtga tggttttgag 1080

tggggtgttt tgttcggttt cggtccaggt ttgaccgtcg aaagagttgt tgttagatct 1140

gtcccaatta agtacggatc c 1161

<210> 76
<211> 387

<212> PRT
<213> Artificial Sequence

<220>
<223> Artificial tetraketide synthase (TKS)

<400> 76

Met Asn His Leu Arg Ala Glu Gly Pro Ala Ser Val Leu Ala Ile Gly
1 5 10 15

Thr Ala Asn Pro Glu Asn Ile Leu Leu Gln Asp Glu Phe Pro Asp Tyr
20 25 30

Tyr Phe Arg Val Thr Lys Ser Glu His Met Thr Gln Leu Lys Glu Lys
35 40 45

Phe Arg Lys Ile Cys Asp Lys Ser Met Ile Arg Lys Arg Asn Cys Phe
50 55 60

Leu Asn Glu Glu His Leu Lys Gln Asn Pro Arg Leu Val Glu His Glu
65 70 75 80

Met Gln Thr Leu Asp Ala Arg Gln Asp Met Leu Val Val Glu Val Pro
85 90 95

Lys Leu Gly Lys Asp Ala Cys Ala Lys Ala Ile Lys Glu Trp Gly Gln
100 105 110

Pro Lys Ser Lys Ile Thr His Leu Ile Phe Thr Ser Ala Ser Thr Thr
115 120 125

Asp Met Pro Gly Ala Asp Tyr His Cys Ala Lys Leu Leu Gly Leu Ser
130 135 140

Pro Ser Val Lys Arg Val Met Met Tyr Gln Leu Gly Cys Tyr Gly Gly
145 150 155 160

Gly Thr Val Leu Arg Ile Ala Lys Asp Ile Ala Glu Asn Asn Lys Gly
165 170 175

```
Ala Arg Val Leu Ala Val Cys Cys Asp Ile Met Ala Cys Leu Phe Arg
            180                 185                 190


Gly Pro Ser Glu Ser Asp Leu Glu Leu Leu Val Gly Gln Ala Ile Phe
        195                 200                 205


Gly Asp Gly Ala Ala Ala Val Ile Val Gly Ala Glu Pro Asp Glu Ser
    210                 215                 220


Val Gly Glu Arg Pro Ile Phe Glu Leu Val Ser Thr Gly Gln Thr Ile
225                 230                 235                 240


Leu Pro Asn Ser Glu Gly Thr Ile Gly Gly His Ile Arg Glu Ala Gly
                245                 250                 255


Leu Ile Phe Asp Leu His Lys Asp Val Pro Met Leu Ile Ser Asn Asn
            260                 265                 270


Ile Glu Lys Cys Leu Ile Glu Ala Phe Thr Pro Ile Gly Ile Ser Asp
        275                 280                 285


Trp Asn Ser Ile Phe Trp Ile Thr His Pro Gly Gly Lys Ala Ile Leu
    290                 295                 300


Asp Lys Val Glu Glu Lys Leu His Leu Lys Ser Asp Lys Phe Val Asp
305                 310                 315                 320


Ser Arg His Val Leu Ser Glu His Gly Asn Met Ser Ser Ser Thr Val
                325                 330                 335


Leu Phe Val Met Asp Glu Leu Arg Lys Arg Ser Leu Glu Glu Gly Lys
            340                 345                 350


Ser Thr Thr Gly Asp Gly Phe Glu Trp Gly Val Leu Phe Gly Phe Gly
        355                 360                 365


Pro Gly Leu Thr Val Glu Arg Val Val Val Arg Ser Val Pro Ile Lys
    370                 375                 380
```

Tyr Gly Ser
385

<210> 77
<211> 309
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial olivetolic acid cyclase (OAC) nucleotide sequence

<400> 77
atggccgtca agcacttgat cgttttgaag ttcaaggatg aaatcactga agctcaaaag 60

gaagaattct tcaaaaccta cgtcaactta gtcaatatta ttccagccat gaaggacgtc 120

tattggggta aggacgttac tcaaaagaat aaggaggaag gttatactca tatcgttgag 180

gtcactttcg aatctgttga gactattcaa gactacatca tccacccagc ccacgttggt 240

ttcggtgatg tttatcgttc cttctgggaa aaattgttga tcttcgacta caccctaga 300

aagggatcc 309

<210> 78
<211> 103
<212> PRT
<213> Artificial Sequence

<220>
<223> Artificial olivetolic acid cyclase (OAC)

<400> 78

Met Ala Val Lys His Leu Ile Val Leu Lys Phe Lys Asp Glu Ile Thr
1 5 10 15

Glu Ala Gln Lys Glu Glu Phe Phe Lys Thr Tyr Val Asn Leu Val Asn
20 25 30

Ile Ile Pro Ala Met Lys Asp Val Tyr Trp Gly Lys Asp Val Thr Gln
35 40 45

Lys Asn Lys Glu Glu Gly Tyr Thr His Ile Val Glu Val Thr Phe Glu
50 55 60

Ser Val Glu Thr Ile Gln Asp Tyr Ile Ile His Pro Ala His Val Gly
65 70 75 80

Phe Gly Asp Val Tyr Arg Ser Phe Trp Glu Lys Leu Leu Ile Phe Asp
85 90 95

Tyr Thr Pro Arg Lys Gly Ser
100

<210> 79
<211> 1578
<212> DNA
<213> Artificial Sequence

<220>
<223> Fusion tetraketide synthase-olivetolic acid cyclase (TKS-OAC)

<400> 79
atgaatcatt taagagctga aggtccagcc tccgttttgg ccatcggtac cgctaaccct 60

gaaaacattt tgttgcaaga cgaattccca gactactact tcagagtcac taagtccgaa 120

cacatgaccc aattgaagga gaagttcaga aagatttgtg acaagtccat gattagaaag 180

agaaactgtt tcttgaacga agaacacttg aagcaaaacc caagattggt tgaacatgaa 240

atgcaaactt tggacgctag acaagacatg ttggttgttg aagtccctaa gttgggtaag 300

gatgcctgtg ctaaggccat taaagaatgg ggtcaaccta agtccaagat tacccacttg 360

attttcacct ctgcctccac cactgacatg cctggtgctg attaccactg cgctaagtta 420

ttgggtttgt ctccatccgt taagagagtt atgatgtacc aattgggttg ctacggtggt 480

ggtactgttt taagaattgc taaggatatt gctgaaaaca acaagggtgc cagagtctta 540

gctgtctgct gtgacattat ggcttgttta ttcagaggtc catctgaatc cgacttggaa 600

ttgttggttg gtcaagctat cttcggtgac ggtgctgctg ccgttattgt tggtgctgaa 660

ccagacgaat ccgttggtga aagaccaatt tttgaattgg tttccaccgg tcaaactatt 720

ttgccaaatt ccgaaggtac catcggtggt catatcagag aagccggttt gatcttcgac 780

ttacataagg atgtcccaat gttgatctct aacaacattg aaaagtgttt gatcgaagct 840

tttaccccaa ttggtatttc tgactggaac tctatcttct ggattaccca tcctggtggt 900

aaggctattt tggataaggt cgaggaaaaa ttgcacttga agtctgacaa gttcgttgac 960

tctagacacg tcttgtccga acatggtaat atgtcctctt ccaccgtttt attcgttatg 1020

gatgagttga gaaagagatc cttagaagaa ggtaagtcca ccaccggtga tggttttgag 1080

tggggtgttt tgttcggttt cggtccaggt ttgaccgtcg aaagagttgt tgttagatct 1140

gtcccaatta agtacgcagc cacaagcggt tctacgggct ccacgggctc taccggcagt 1200

gggaggagca ctgggtcaac gggatcaaca ggtagtggaa gatcacacat ggttgccgtc 1260

aagcacttga tcgttttgaa gttcaaggat gaaatcactg aagctcaaaa ggaagaattc 1320

ttcaaaacct acgtcaactt agtcaatatt attccagcca tgaaggacgt ctattggggt 1380

aaggacgtta ctcaaaagaa taaggaggaa ggttatactc atatcgttga ggtcactttc 1440

gaatctgttg agactattca agactacatc atccacccag cccacgttgg tttcggtgat 1500

gtttatcgtt ccttctggga aaaattgttg atcttcgact acacccctag aaagggtaac 1560

tcgagagctt ttgattaa 1578

<210> 80
<211> 525
<212> PRT
<213> Artificial Sequence

<220>
<223> Fusion tetraketide synthase-olivetolic acid cyclase (TKS-OAC)

<400> 80

Met Asn His Leu Arg Ala Glu Gly Pro Ala Ser Val Leu Ala Ile Gly
1 5 10 15

Thr Ala Asn Pro Glu Asn Ile Leu Leu Gln Asp Glu Phe Pro Asp Tyr
20 25 30

Tyr Phe Arg Val Thr Lys Ser Glu His Met Thr Gln Leu Lys Glu Lys
35 40 45

Phe Arg Lys Ile Cys Asp Lys Ser Met Ile Arg Lys Arg Asn Cys Phe
50 55 60

Leu Asn Glu Glu His Leu Lys Gln Asn Pro Arg Leu Val Glu His Glu
65 70 75 80

```
Met Gln Thr Leu Asp Ala Arg Gln Asp Met Leu Val Val Glu Val Pro
                85                  90                  95

Lys Leu Gly Lys Asp Ala Cys Ala Lys Ala Ile Lys Glu Trp Gly Gln
            100                 105                 110

Pro Lys Ser Lys Ile Thr His Leu Ile Phe Thr Ser Ala Ser Thr Thr
        115                 120                 125

Asp Met Pro Gly Ala Asp Tyr His Cys Ala Lys Leu Leu Gly Leu Ser
    130                 135                 140

Pro Ser Val Lys Arg Val Met Met Tyr Gln Leu Gly Cys Tyr Gly Gly
145                 150                 155                 160

Gly Thr Val Leu Arg Ile Ala Lys Asp Ile Ala Glu Asn Asn Lys Gly
                165                 170                 175

Ala Arg Val Leu Ala Val Cys Cys Asp Ile Met Ala Cys Leu Phe Arg
            180                 185                 190

Gly Pro Ser Glu Ser Asp Leu Glu Leu Leu Val Gly Gln Ala Ile Phe
        195                 200                 205

Gly Asp Gly Ala Ala Ala Val Ile Val Gly Ala Glu Pro Asp Glu Ser
    210                 215                 220

Val Gly Glu Arg Pro Ile Phe Glu Leu Val Ser Thr Gly Gln Thr Ile
225                 230                 235                 240

Leu Pro Asn Ser Glu Gly Thr Ile Gly Gly His Ile Arg Glu Ala Gly
                245                 250                 255

Leu Ile Phe Asp Leu His Lys Asp Val Pro Met Leu Ile Ser Asn Asn
            260                 265                 270

Ile Glu Lys Cys Leu Ile Glu Ala Phe Thr Pro Ile Gly Ile Ser Asp
        275                 280                 285
```

```
Trp Asn Ser Ile Phe Trp Ile Thr His Pro Gly Gly Lys Ala Ile Leu
    290                 295                 300

Asp Lys Val Glu Glu Lys Leu His Leu Lys Ser Asp Lys Phe Val Asp
305                 310                 315                 320

Ser Arg His Val Leu Ser Glu His Gly Asn Met Ser Ser Ser Thr Val
                325                 330                 335

Leu Phe Val Met Asp Glu Leu Arg Lys Arg Ser Leu Glu Glu Gly Lys
            340                 345                 350

Ser Thr Thr Gly Asp Gly Phe Glu Trp Gly Val Leu Phe Gly Phe Gly
        355                 360                 365

Pro Gly Leu Thr Val Glu Arg Val Val Val Arg Ser Val Pro Ile Lys
    370                 375                 380

Tyr Ala Ala Thr Ser Gly Ser Thr Gly Ser Thr Gly Ser Thr Gly Ser
385                 390                 395                 400

Gly Arg Ser Thr Gly Ser Thr Gly Ser Thr Gly Ser Gly Arg Ser His
                405                 410                 415

Met Val Ala Val Lys His Leu Ile Val Leu Lys Phe Lys Asp Glu Ile
            420                 425                 430

Thr Glu Ala Gln Lys Glu Glu Phe Phe Lys Thr Tyr Val Asn Leu Val
        435                 440                 445

Asn Ile Ile Pro Ala Met Lys Asp Val Tyr Trp Gly Lys Asp Val Thr
    450                 455                 460

Gln Lys Asn Lys Glu Glu Gly Tyr Thr His Ile Val Glu Val Thr Phe
465                 470                 475                 480

Glu Ser Val Glu Thr Ile Gln Asp Tyr Ile Ile His Pro Ala His Val
                485                 490                 495
```

Gly Phe Gly Asp Val Tyr Arg Ser Phe Trp Glu Lys Leu Leu Ile Phe
500 505 510

Asp Tyr Thr Pro Arg Lys Gly Asn Ser Arg Ala Phe Asp
515 520 525

<210> 81
<211> 1185
<212> DNA
<213> Cannabis sativa

<400> 81
atgggattgt ccagcgtgtg caccttctca ttccaaacca actaccatac acttctcaat 60

ccgcacaata ataacccgaa aaccagctta ttatgttata gacacccgaa gacgcctatt 120

aagtacagtt ataacaactt tcctagcaag cattgctcta ctaaaagttt tcatctgcaa 180

aacaagtgct ctgagtcctt gagtatagca aagaatagca ttagagctgc aacgacaaat 240

caaaccgagc cgccggagtc tgataaccat agtgtggcga ccaagatact aaattttggc 300

aaagcgtgtt ggaagctaca acgaccttat actattatcg cgtttacgag ttgtgcatgt 360

gggctgttcg ggaaagagct cttgcacaat acaaacttaa tcagttggag tttgatgttc 420

aaagcatttt tttttctcgt cgctatctta tgtatcgcgt catttaccac gaccataaat 480

caaatatacg atctgcatat cgatcgtatc aataagcccg acctcccact ggcctcaggt 540

gaaatttccg ttaacacggc gtggattatg agtataatcg tagcactatt tggacttatt 600

ataaccatca aaatgaaggg cggtcctcta tacatttttg gatattgttt tgggattttt 660

ggaggtatag tctattccgt ccccccattc agatggaaac aaaacccgtc caccgctttc 720

cttttaaatt tcttggcaca tatcatcaca aacttcacgt tttactatgc cagccgagcc 780

gcactgggac tcccgttcga gttgcgtccg tcattcacct tccttttagc ttttatgaaa 840

tctatgggaa gcgctttagc tttaattaag gacgcgagcg acgtggaagg ggacacgaaa 900

ttcggtataa gcacgctggc ttcaaaatat ggaagtcgta atctcactct attttgttct 960

gggattgtac tcctaagtta cgtagctgcg atactcgcag gcattatatg gccacaagct 1020

ttcaactcca acgtaatgtt gctatcacat gcaatcttgg ccttctggct catccttcaa 1080

actagagatt ttgcactaac gaactacgat ccagaagcgg gtcgtcgatt ttacgaattt 1140

atgtggaaac tgtactatgc tgagtacctc gtctatgtgt tcata 1185

<210> 82
<211> 395
<212> PRT
<213> Cannabis sativa

<400> 82

Met Gly Leu Ser Ser Val Cys Thr Phe Ser Phe Gln Thr Asn Tyr His
1 5 10 15

Thr Leu Leu Asn Pro His Asn Asn Asn Pro Lys Thr Ser Leu Leu Cys
20 25 30

Tyr Arg His Pro Lys Thr Pro Ile Lys Tyr Ser Tyr Asn Asn Phe Pro
35 40 45

Ser Lys His Cys Ser Thr Lys Ser Phe His Leu Gln Asn Lys Cys Ser
50 55 60

Glu Ser Leu Ser Ile Ala Lys Asn Ser Ile Arg Ala Ala Thr Thr Asn
65 70 75 80

Gln Thr Glu Pro Pro Glu Ser Asp Asn His Ser Val Ala Thr Lys Ile
85 90 95

Leu Asn Phe Gly Lys Ala Cys Trp Lys Leu Gln Arg Pro Tyr Thr Ile
100 105 110

Ile Ala Phe Thr Ser Cys Ala Cys Gly Leu Phe Gly Lys Glu Leu Leu
115 120 125

His Asn Thr Asn Leu Ile Ser Trp Ser Leu Met Phe Lys Ala Phe Phe
130 135 140

Phe Leu Val Ala Ile Leu Cys Ile Ala Ser Phe Thr Thr Thr Ile Asn
145 150 155 160

```
Gln Ile Tyr Asp Leu His Ile Asp Arg Ile Asn Lys Pro Asp Leu Pro
                165                 170                 175

Leu Ala Ser Gly Glu Ile Ser Val Asn Thr Ala Trp Ile Met Ser Ile
            180                 185                 190

Ile Val Ala Leu Phe Gly Leu Ile Ile Thr Ile Lys Met Lys Gly Gly
        195                 200                 205

Pro Leu Tyr Ile Phe Gly Tyr Cys Phe Gly Ile Phe Gly Gly Ile Val
    210                 215                 220

Tyr Ser Val Pro Pro Phe Arg Trp Lys Gln Asn Pro Ser Thr Ala Phe
225                 230                 235                 240

Leu Leu Asn Phe Leu Ala His Ile Ile Thr Asn Phe Thr Phe Tyr Tyr
                245                 250                 255

Ala Ser Arg Ala Ala Leu Gly Leu Pro Phe Glu Leu Arg Pro Ser Phe
            260                 265                 270

Thr Phe Leu Leu Ala Phe Met Lys Ser Met Gly Ser Ala Leu Ala Leu
        275                 280                 285

Ile Lys Asp Ala Ser Asp Val Glu Gly Asp Thr Lys Phe Gly Ile Ser
    290                 295                 300

Thr Leu Ala Ser Lys Tyr Gly Ser Arg Asn Leu Thr Leu Phe Cys Ser
305                 310                 315                 320

Gly Ile Val Leu Leu Ser Tyr Val Ala Ala Ile Leu Ala Gly Ile Ile
                325                 330                 335

Trp Pro Gln Ala Phe Asn Ser Asn Val Met Leu Leu Ser His Ala Ile
            340                 345                 350

Leu Ala Phe Trp Leu Ile Leu Gln Thr Arg Asp Phe Ala Leu Thr Asn
        355                 360                 365
```

Tyr Asp Pro Glu Ala Gly Arg Arg Phe Tyr Glu Phe Met Trp Lys Leu
370 375 380

Tyr Tyr Ala Glu Tyr Leu Val Tyr Val Phe Ile
385 390 395

<210> 83
<211> 921
<212> DNA
<213> Streptomyces sp.

<400> 83
atgtctgagg cggcagacgt agagagagta tacgctgcta tggaggaagc ggctggatta 60

ttgggggtgg cttgtgccag agacaagata tatccgttac tgtctacttt ccaggacact 120

cttgtagaag gagggagtgt ggtggtgttt agtatggcat caggccgtca ttcaacagag 180

ctagatttca gtatatctgt gccaacaagt cacggtgatc catacgcaac cgtagtcgag 240

aagggtcttt tcccggcaac agggcatcct gtagatgatt tgcttgccga cacacagaag 300

cacctgcccg tctccatgtt cgcaatcgat ggtgaggtga ccggaggatt taaaaagact 360

tacgctttct tcccgactga caatatgcca ggagttgccg agttgagtgc aataccatcc 420

atgccgccag cagtcgcgga gaacgccgaa ttgttcgccc gttacggctt ggacaaagtc 480

caaatgacta gtatggacta taaaaagagg caggtgaatc tatatttcag cgaactttct 540

gcccaaacct tggaggcgga gagcgtttta gcccttgtta gggagttagg gctacacgtc 600

ccgaatgagt tgggtttgaa attttgtaag cgtagctttt cagtatatcc gacgctgaac 660

tgggaaactg gaaagattga caggctatgc tttgcagtga tttctaatga ccctacgctt 720

gtaccttcct cagacgaggg cgacatcgag aaattccaca actatgccac aaaagctccg 780

tatgcctacg tcggcgaaaa acgtactcta gtatacggtt tgactctgag tcccaaggaa 840

gagtattaca agctaggagc gtactatcat atcactgatg tgcaacgtgg cttgctgaaa 900

gccttcgact ccttagagga c 921

<210> 84
<211> 307
<212> PRT
<213> Streptomyces sp.

<400> 84

```
Met Ser Glu Ala Ala Asp Val Glu Arg Val Tyr Ala Ala Met Glu Glu
1               5                  10                  15

Ala Ala Gly Leu Leu Gly Val Ala Cys Ala Arg Asp Lys Ile Tyr Pro
            20                  25                  30

Leu Leu Ser Thr Phe Gln Asp Thr Leu Val Glu Gly Gly Ser Val Val
        35                  40                  45

Val Phe Ser Met Ala Ser Gly Arg His Ser Thr Glu Leu Asp Phe Ser
    50                  55                  60

Ile Ser Val Pro Thr Ser His Gly Asp Pro Tyr Ala Thr Val Val Glu
65                  70                  75                  80

Lys Gly Leu Phe Pro Ala Thr Gly His Pro Val Asp Asp Leu Leu Ala
                85                  90                  95

Asp Thr Gln Lys His Leu Pro Val Ser Met Phe Ala Ile Asp Gly Glu
            100                 105                 110

Val Thr Gly Gly Phe Lys Lys Thr Tyr Ala Phe Phe Pro Thr Asp Asn
        115                 120                 125

Met Pro Gly Val Ala Glu Leu Ser Ala Ile Pro Ser Met Pro Pro Ala
    130                 135                 140

Val Ala Glu Asn Ala Glu Leu Phe Ala Arg Tyr Gly Leu Asp Lys Val
145                 150                 155                 160

Gln Met Thr Ser Met Asp Tyr Lys Lys Arg Gln Val Asn Leu Tyr Phe
                165                 170                 175

Ser Glu Leu Ser Ala Gln Thr Leu Glu Ala Glu Ser Val Leu Ala Leu
            180                 185                 190

Val Arg Glu Leu Gly Leu His Val Pro Asn Glu Leu Gly Leu Lys Phe
        195                 200                 205
```

Cys Lys Arg Ser Phe Ser Val Tyr Pro Thr Leu Asn Trp Glu Thr Gly
210 215 220

Lys Ile Asp Arg Leu Cys Phe Ala Val Ile Ser Asn Asp Pro Thr Leu
225 230 235 240

Val Pro Ser Ser Asp Glu Gly Asp Ile Glu Lys Phe His Asn Tyr Ala
245 250 255

Thr Lys Ala Pro Tyr Ala Tyr Val Gly Glu Lys Arg Thr Leu Val Tyr
260 265 270

Gly Leu Thr Leu Ser Pro Lys Glu Glu Tyr Tyr Lys Leu Gly Ala Tyr
275 280 285

Tyr His Ile Thr Asp Val Gln Arg Gly Leu Leu Lys Ala Phe Asp Ser
290 295 300

Leu Glu Asp
305

<210> 85
<211> 1635
<212> DNA
<213> Cannabis sativa

<400> 85
atgaactgtt ccgcgtttag tttctggttc gtgtgcaaga tcatcttctt ttttctaagc 60

ttcaacattc aaatcagcat cgcgaatcct caggagaact tcctgaagtg tttctcagaa 120

tacataccaa ataatcccgc caatcctaaa tttatatata cccaacatga tcagctatac 180

atgagtgtat tgaactctac gattcagaat ctaagattca catctgatac aacgccgaaa 240

cctctagtaa tcgtgacacc gtctaatgtc tcccatattc aagcttctat cttgtgctca 300

aagaaagtcg gtcttcaaat aaggacacgt tctggcgggc atgacgccga gggcatgtca 360

tatatcagcc aagtaccatt tgtagtcgtg gatttaagaa acatgcattc tataaaaatc 420

gacgttcact cccaaacggc atgggtggaa gctggagcga cactggggga ggtgtactac 480

tggatcaatg aaaagaacga aaatttttcc ttccccggag gatattgtcc gacagttggg 540

gtgggggcc acttctctgg cggcgggtac ggcgctctga tgcgtaatta tggactggcc 600

gcagataaca taatcgacgc gcatttggtg aacgttgacg ggaaggtttt ggataggaag 660

tctatgggag aggacctatt ctgggcaatt agaggcggag gaggagagaa ttttggtatt 720

attgctgcat ggaagattaa attggttgcg gtgccgagta aaagtaccat cttttccgtc 780

aagaaaaaca tggagattca cggactagtt aagctgttta ataaatggca aaacatcgcc 840

tataagtacg acaaagattt ggttctgatg acgcatttca taactaagaa tataactgat 900

aatcacggca agaataagac cactgtgcac ggttatttta gttcaatatt ccatggcggc 960

gttgactccc ttgtcgattt gatgaataag agcttccctg aattgggtat caagaagaca 1020

gactgcaaag aattctcctg gattgatacg actatcttct attcaggggt cgtgaatttc 1080

aacactgcga atttcaaaaa ggagatattg ttagaccgtt ccgcgggaaa aaaaactgcg 1140

ttttctatta aactagatta tgtgaaaaaa ccgattcctg agacagccat ggttaagatt 1200

cttgaaaaat tgtatgaaga ggatgtcggg gtcggtatgt acgtccttta cccttacgga 1260

ggaatcatgg aagaaatatc cgaatctgca attcctttcc cgcatcgtgc cggtattatg 1320

tatgagctat ggtacaccgc tagctgggag aagcaggaag ataacgagaa gcatatcaat 1380

tgggtgaggt ctgtgtataa ttttacaaca ccatacgtca gtcaaaaccc tagattggcc 1440

tatcttaact atcgtgatct ggacttggga aaaacaaatc cagaatcccc aaataactac 1500

actcaagccc gtatatgggg cgagaagtac ttcggcaaaa atttcaatag actggtcaaa 1560

gttaagacga aagcagaccc taataatttc ttccgtaacg aacaatcaat tcccccgctt 1620

ccgccacacc atcac 1635

<210> 86
<211> 545
<212> PRT
<213> Cannabis sativa

<400> 86

Met Asn Cys Ser Ala Phe Ser Phe Trp Phe Val Cys Lys Ile Ile Phe
1 5 10 15

```
Phe Phe Leu Ser Phe Asn Ile Gln Ile Ser Ile Ala Asn Pro Gln Glu
            20                  25                  30

Asn Phe Leu Lys Cys Phe Ser Glu Tyr Ile Pro Asn Asn Pro Ala Asn
        35                  40                  45

Pro Lys Phe Ile Tyr Thr Gln His Asp Gln Leu Tyr Met Ser Val Leu
    50                  55                  60

Asn Ser Thr Ile Gln Asn Leu Arg Phe Thr Ser Asp Thr Thr Pro Lys
65                  70                  75                  80

Pro Leu Val Ile Val Thr Pro Ser Asn Val Ser His Ile Gln Ala Ser
                85                  90                  95

Ile Leu Cys Ser Lys Lys Val Gly Leu Gln Ile Arg Thr Arg Ser Gly
            100                 105                 110

Gly His Asp Ala Glu Gly Met Ser Tyr Ile Ser Gln Val Pro Phe Val
        115                 120                 125

Val Val Asp Leu Arg Asn Met His Ser Ile Lys Ile Asp Val His Ser
    130                 135                 140

Gln Thr Ala Trp Val Glu Ala Gly Ala Thr Leu Gly Glu Val Tyr Tyr
145                 150                 155                 160

Trp Ile Asn Glu Lys Asn Glu Asn Phe Ser Phe Pro Gly Gly Tyr Cys
                165                 170                 175

Pro Thr Val Gly Val Gly Gly His Phe Ser Gly Gly Gly Tyr Gly Ala
            180                 185                 190

Leu Met Arg Asn Tyr Gly Leu Ala Ala Asp Asn Ile Ile Asp Ala His
        195                 200                 205

Leu Val Asn Val Asp Gly Lys Val Leu Asp Arg Lys Ser Met Gly Glu
    210                 215                 220
```

```
Asp Leu Phe Trp Ala Ile Arg Gly Gly Gly Gly Glu Asn Phe Gly Ile
225                 230                 235                 240


Ile Ala Ala Trp Lys Ile Lys Leu Val Ala Val Pro Ser Lys Ser Thr
                245                 250                 255


Ile Phe Ser Val Lys Lys Asn Met Glu Ile His Gly Leu Val Lys Leu
            260                 265                 270


Phe Asn Lys Trp Gln Asn Ile Ala Tyr Lys Tyr Asp Lys Asp Leu Val
        275                 280                 285


Leu Met Thr His Phe Ile Thr Lys Asn Ile Thr Asp Asn His Gly Lys
    290                 295                 300


Asn Lys Thr Thr Val His Gly Tyr Phe Ser Ser Ile Phe His Gly Gly
305                 310                 315                 320


Val Asp Ser Leu Val Asp Leu Met Asn Lys Ser Phe Pro Glu Leu Gly
                325                 330                 335


Ile Lys Lys Thr Asp Cys Lys Glu Phe Ser Trp Ile Asp Thr Thr Ile
            340                 345                 350


Phe Tyr Ser Gly Val Val Asn Phe Asn Thr Ala Asn Phe Lys Lys Glu
        355                 360                 365


Ile Leu Leu Asp Arg Ser Ala Gly Lys Lys Thr Ala Phe Ser Ile Lys
    370                 375                 380


Leu Asp Tyr Val Lys Lys Pro Ile Pro Glu Thr Ala Met Val Lys Ile
385                 390                 395                 400


Leu Glu Lys Leu Tyr Glu Glu Asp Val Gly Val Gly Met Tyr Val Leu
                405                 410                 415


Tyr Pro Tyr Gly Gly Ile Met Glu Glu Ile Ser Glu Ser Ala Ile Pro
            420                 425                 430
```

Phe Pro His Arg Ala Gly Ile Met Tyr Glu Leu Trp Tyr Thr Ala Ser
435 440 445

Trp Glu Lys Gln Glu Asp Asn Glu Lys His Ile Asn Trp Val Arg Ser
450 455 460

Val Tyr Asn Phe Thr Thr Pro Tyr Val Ser Gln Asn Pro Arg Leu Ala
465 470 475 480

Tyr Leu Asn Tyr Arg Asp Leu Asp Leu Gly Lys Thr Asn Pro Glu Ser
485 490 495

Pro Asn Asn Tyr Thr Gln Ala Arg Ile Trp Gly Glu Lys Tyr Phe Gly
500 505 510

Lys Asn Phe Asn Arg Leu Val Lys Val Lys Thr Lys Ala Asp Pro Asn
515 520 525

Asn Phe Phe Arg Asn Glu Gln Ser Ile Pro Pro Leu Pro Pro His His
530 535 540

His
545

<210> 87
<211> 1632
<212> DNA
<213> Cannabis sativa

<400> 87
atgaaatgtt ctactttcag tttttggttc gtgtgtaaga tcatcttttt ctttttcagc 60

ttcaatatac agacaagtat cgccaatcca agagaaaatt tcttaaaatg tttttcacag 120

tacatcccta ataacgccac taacctgaaa ttagtgtaca cccaaaataa tcctctttat 180

atgtctgttt taaactccac gatccataat ttaaggttta catcagatac gacaccaaag 240

cccttggtaa tcgtgactcc cagccacgtg agccacatac aggggaccat cctgtgctct 300

aagaaagtag gcttgcagat caggacaaga tccggtggac acgacagtga gggaatgtcc 360

tatatttcac aagtcccctt cgttatagta gatctgagga acatgaggtc cattaagatt 420

gatgtgcact cacaaacggc ttgggttgaa gctggagcca cattgggaga ggtttattac 480

tgggtgaatg agaagaacga gaacctttca ttagcagcgg gatattgtcc cacggtgtgc 540

gcaggtgggc atttcggggg aggagggtac ggccctttga tgagaaatta cgggctagcg 600

gcagacaaca tcatcgacgc ccatctggtg aacgtgcatg gaaaagtact ggacagaaag 660

tcaatgggcg aggacctgtt ttgggctttg agaggggggcg gtgcagagtc atttggcatc 720

atagttgcat ggaaaatcag acttgttgcc gtcccaaagt ccacaatgtt ctctgttaag 780

aaaatcatgg agatacacga attggtgaaa ttagtgaata aatggcaaaa catagcgtac 840

aagtacgaca aagacttact gctgatgaca cactttatca cccgtaatat tacagataat 900

cagggtaaga acaaaaccgc gatccataca tatttttcat ccgtttttct aggcggtgtc 960

gattcattag tagatctgat gaacaaatct ttccccgaac ttggtatcaa aaagactgat 1020

tgcagacagt tatcatggat tgatacaata attttctatt ctggtgtcgt aaattacgat 1080

accgataatt ttaataagga aatactatta gatcgttccg ctgggcagaa tggtgcattc 1140

aagataaaac ttgattatgt caaaaagccc attccagaga gtgtctttgt gcagatcctt 1200

gagaagttgt atgaagaaga cattggtgca gggatgtacg cgctatatcc gtacgggggt 1260

attatggacg agatttctga gagcgccata ccattcccac acagagcagg aattttatac 1320

gagttatggt atatctgctc atgggaaaaa caggaagaca acgagaagca cttaaactgg 1380

atacgtaata tctataattt tatgacccca tacgtatcaa aaaatccgcg tcttgcgtac 1440

cttaactaca gggacctgga cataggtata aacgacccaa aaaatcccaa taattacacc 1500

caagctagaa tctgggggga gaagtatttc ggtaagaact ttgaccgttt ggtaaaagtc 1560

aaaactctgg tcgatccgaa caatttcttc cgtaacgagc aatccatacc tccgctaccg 1620

agacatagac at 1632

<210> 88
<211> 544
<212> PRT
<213> Cannabis sativa

<400> 88

Met Lys Cys Ser Thr Phe Ser Phe Trp Phe Val Cys Lys Ile Ile Phe
1 5 10 15

```
Phe Phe Phe Ser Phe Asn Ile Gln Thr Ser Ile Ala Asn Pro Arg Glu
            20                  25                  30

Asn Phe Leu Lys Cys Phe Ser Gln Tyr Ile Pro Asn Asn Ala Thr Asn
        35                  40                  45

Leu Lys Leu Val Tyr Thr Gln Asn Asn Pro Leu Tyr Met Ser Val Leu
    50                  55                  60

Asn Ser Thr Ile His Asn Leu Arg Phe Thr Ser Asp Thr Thr Pro Lys
65                  70                  75                  80

Pro Leu Val Ile Val Thr Pro Ser His Val Ser His Ile Gln Gly Thr
                85                  90                  95

Ile Leu Cys Ser Lys Lys Val Gly Leu Gln Ile Arg Thr Arg Ser Gly
            100                 105                 110

Gly His Asp Ser Glu Gly Met Ser Tyr Ile Ser Gln Val Pro Phe Val
        115                 120                 125

Ile Val Asp Leu Arg Asn Met Arg Ser Ile Lys Ile Asp Val His Ser
    130                 135                 140

Gln Thr Ala Trp Val Glu Ala Gly Ala Thr Leu Gly Glu Val Tyr Tyr
145                 150                 155                 160

Trp Val Asn Glu Lys Asn Glu Asn Leu Ser Leu Ala Ala Gly Tyr Cys
                165                 170                 175

Pro Thr Val Cys Ala Gly Gly His Phe Gly Gly Gly Gly Tyr Gly Pro
            180                 185                 190

Leu Met Arg Asn Tyr Gly Leu Ala Ala Asp Asn Ile Ile Asp Ala His
        195                 200                 205

Leu Val Asn Val His Gly Lys Val Leu Asp Arg Lys Ser Met Gly Glu
    210                 215                 220
```

```
Asp Leu Phe Trp Ala Leu Arg Gly Gly Gly Ala Glu Ser Phe Gly Ile
225                 230                 235                 240


Ile Val Ala Trp Lys Ile Arg Leu Val Ala Val Pro Lys Ser Thr Met
                245                 250                 255


Phe Ser Val Lys Lys Ile Met Glu Ile His Glu Leu Val Lys Leu Val
            260                 265                 270


Asn Lys Trp Gln Asn Ile Ala Tyr Lys Tyr Asp Lys Asp Leu Leu Leu
        275                 280                 285


Met Thr His Phe Ile Thr Arg Asn Ile Thr Asp Asn Gln Gly Lys Asn
    290                 295                 300


Lys Thr Ala Ile His Thr Tyr Phe Ser Ser Val Phe Leu Gly Gly Val
305                 310                 315                 320


Asp Ser Leu Val Asp Leu Met Asn Lys Ser Phe Pro Glu Leu Gly Ile
                325                 330                 335


Lys Lys Thr Asp Cys Arg Gln Leu Ser Trp Ile Asp Thr Ile Ile Phe
            340                 345                 350


Tyr Ser Gly Val Val Asn Tyr Asp Thr Asp Asn Phe Asn Lys Glu Ile
        355                 360                 365


Leu Leu Asp Arg Ser Ala Gly Gln Asn Gly Ala Phe Lys Ile Lys Leu
    370                 375                 380


Asp Tyr Val Lys Lys Pro Ile Pro Glu Ser Val Phe Val Gln Ile Leu
385                 390                 395                 400


Glu Lys Leu Tyr Glu Glu Asp Ile Gly Ala Gly Met Tyr Ala Leu Tyr
                405                 410                 415


Pro Tyr Gly Gly Ile Met Asp Glu Ile Ser Glu Ser Ala Ile Pro Phe
            420                 425                 430
```

Pro His Arg Ala Gly Ile Leu Tyr Glu Leu Trp Tyr Ile Cys Ser Trp
435 440 445

Glu Lys Gln Glu Asp Asn Glu Lys His Leu Asn Trp Ile Arg Asn Ile
450 455 460

Tyr Asn Phe Met Thr Pro Tyr Val Ser Lys Asn Pro Arg Leu Ala Tyr
465 470 475 480

Leu Asn Tyr Arg Asp Leu Asp Ile Gly Ile Asn Asp Pro Lys Asn Pro
485 490 495

Asn Asn Tyr Thr Gln Ala Arg Ile Trp Gly Glu Lys Tyr Phe Gly Lys
500 505 510

Asn Phe Asp Arg Leu Val Lys Val Lys Thr Leu Val Asp Pro Asn Asn
515 520 525

Phe Phe Arg Asn Glu Gln Ser Ile Pro Pro Leu Pro Arg His Arg His
530 535 540

<210> 89
<211> 2160
<212> DNA
<213> Cannabis sativa

<400> 89
atgggaaaaa actacaaaag tctggactcc gtcgtcgcgt cagacttcat tgccctaggc 60

ataacatcag aggtagcgga aaccttacac ggcagactag ccgagattgt ttgtaactac 120

ggggcggcta ctccccagac ttggatcaat atagccaatc acatattaag ccccgatttg 180

ccgttttccc ttcaccaaat gttgttctac ggctgctata aggactttgg accagcgccc 240

cccgcgtgga ttcctgatcc ggagaaagtt aaatccacga atcttggggc attactagaa 300

aaacgtggca aagaattcct aggagttaaa tataaggacc ccatatcttc cttttcacac 360

tttcaagaat tttcagttag aaacccagag gtttactgga ggacagtatt aatggatgag 420

atgaagataa gctttagtaa ggatccggag tgtattctgc gtagagatga cattaacaat 480

cctggcggaa gtgaatggct gcctggtggg tacctgaata gtgctaagaa ctgtttaaac 540

gtcaactcta ataaaaaatt gaatgataca atgattgtat ggagagacga agggaacgat 600

gacctaccat tgaacaagct gactctagat cagctacgta aacgtgtatg gttggtcggg 660

tacgcgctgg aggagatggg attagaaaaa ggatgcgcaa ttgctatcga catgcctatg 720

catgtggacg cggtagtcat ttacttggcc attgtcctag cgggttacgt cgtcgtttca 780

attgcagaca gcttttctgc acccgaaatc agtacccgtc tgcgtttgtc taaagctaag 840

gcaatattta cccaagacca tataattaga ggcaagaagc gtataccgtt gtacagtagg 900

gttgtagagg caaagtcacc catggctatt gtgataccat gctctggctc taatatagga 960

gcggagctta gagatggtga catctcctgg gattactttc ttgaacgtgc taaggagttt 1020

aaaaactgtg aatttactgc aagagagcag cccgtggatg catacacaaa catattgttc 1080

tccagcggta ctacgggaga acctaaagca ataccttgga cacaagctac acccccttaaa 1140

gcggccgctg acggatggtc ccacctggat atcaggaagg gtgacgtcat agtttggccg 1200

actaacctgg gatggatgat gggcccttgg ctggtttacg ctagccttct gaatggggcc 1260

agcattgcat tgtacaatgg ctcaccgctt gtatcaggct tcgcgaagtt cgtacaggac 1320

gccaaggtaa caatgctagg cgtagttccg tccatagtta ggtcttggaa gagcacgaac 1380

tgcgttagtg gctacgattg gagcactatt cgttgtttca gctcttctgg cgaggccagc 1440

aacgttgatg aatatttgtg gttgatgggg agagcgaact acaaacctgt tattgagatg 1500

tgcggcggaa ctgagattgg gggagcattc tccgccggtt cttttctaca agcccaaagt 1560

ttatcctctt ttagcagcca gtgcatgggc tgtacactat acattctgga caaaaatggt 1620

tatccgatgc cgaaaaacaa gcccggcatc ggagaactgg ccctaggacc cgtgatgttc 1680

ggcgctagta agacgttgtt gaatgggaat caccacgacg tttattttaa gggaatgcca 1740

actttgaatg gcgaagtact tcgtagacac ggagacatct ttgagttgac ttcaaacggt 1800

tactaccacg ctcatggacg tgccgatgat acgatgaaca ttgggggaat taaaatttca 1860

tccatagaaa tagaacgtgt gtgtaacgaa gtcgatgatc gtgtattcga gactacagcg 1920

atcggtgtcc caccgttggg tgggggacca gaacaattgg taatcttttt tgttctgaaa 1980

gactccaacg atacgaccat cgacctaaat cagctgaggc tatcctttaa tctgggcttg 2040

cagaaaaagc taaatccttt attcaaagtc actagagttg ttcctttatc ttcattacca 2100

agaactgcaa caaataaaat aatgcgtaga gttctaaggc agcagtttag tcatttcgaa 2160

<210> 90
<211> 720
<212> PRT
<213> Cannabis sativa

<400> 90

Met Gly Lys Asn Tyr Lys Ser Leu Asp Ser Val Val Ala Ser Asp Phe
1 5 10 15

Ile Ala Leu Gly Ile Thr Ser Glu Val Ala Glu Thr Leu His Gly Arg
20 25 30

Leu Ala Glu Ile Val Cys Asn Tyr Gly Ala Ala Thr Pro Gln Thr Trp
35 40 45

Ile Asn Ile Ala Asn His Ile Leu Ser Pro Asp Leu Pro Phe Ser Leu
50 55 60

His Gln Met Leu Phe Tyr Gly Cys Tyr Lys Asp Phe Gly Pro Ala Pro
65 70 75 80

Pro Ala Trp Ile Pro Asp Pro Glu Lys Val Lys Ser Thr Asn Leu Gly
85 90 95

Ala Leu Leu Glu Lys Arg Gly Lys Glu Phe Leu Gly Val Lys Tyr Lys
100 105 110

Asp Pro Ile Ser Ser Phe Ser His Phe Gln Glu Phe Ser Val Arg Asn
115 120 125

Pro Glu Val Tyr Trp Arg Thr Val Leu Met Asp Glu Met Lys Ile Ser
130 135 140

Phe Ser Lys Asp Pro Glu Cys Ile Leu Arg Arg Asp Asp Ile Asn Asn
145 150 155 160

```
Pro Gly Gly Ser Glu Trp Leu Pro Gly Gly Tyr Leu Asn Ser Ala Lys
                165                 170                 175

Asn Cys Leu Asn Val Asn Ser Asn Lys Lys Leu Asn Asp Thr Met Ile
            180                 185                 190

Val Trp Arg Asp Glu Gly Asn Asp Asp Leu Pro Leu Asn Lys Leu Thr
        195                 200                 205

Leu Asp Gln Leu Arg Lys Arg Val Trp Leu Val Gly Tyr Ala Leu Glu
    210                 215                 220

Glu Met Gly Leu Glu Lys Gly Cys Ala Ile Ala Ile Asp Met Pro Met
225                 230                 235                 240

His Val Asp Ala Val Val Ile Tyr Leu Ala Ile Val Leu Ala Gly Tyr
                245                 250                 255

Val Val Val Ser Ile Ala Asp Ser Phe Ser Ala Pro Glu Ile Ser Thr
            260                 265                 270

Arg Leu Arg Leu Ser Lys Ala Lys Ala Ile Phe Thr Gln Asp His Ile
        275                 280                 285

Ile Arg Gly Lys Lys Arg Ile Pro Leu Tyr Ser Arg Val Val Glu Ala
    290                 295                 300

Lys Ser Pro Met Ala Ile Val Ile Pro Cys Ser Gly Ser Asn Ile Gly
305                 310                 315                 320

Ala Glu Leu Arg Asp Gly Asp Ile Ser Trp Asp Tyr Phe Leu Glu Arg
                325                 330                 335

Ala Lys Glu Phe Lys Asn Cys Glu Phe Thr Ala Arg Glu Gln Pro Val
            340                 345                 350

Asp Ala Tyr Thr Asn Ile Leu Phe Ser Ser Gly Thr Thr Gly Glu Pro
        355                 360                 365
```

```
Lys Ala Ile Pro Trp Thr Gln Ala Thr Pro Leu Lys Ala Ala Ala Asp
    370                 375                 380

Gly Trp Ser His Leu Asp Ile Arg Lys Gly Asp Val Ile Val Trp Pro
385                 390                 395                 400

Thr Asn Leu Gly Trp Met Met Gly Pro Trp Leu Val Tyr Ala Ser Leu
                405                 410                 415

Leu Asn Gly Ala Ser Ile Ala Leu Tyr Asn Gly Ser Pro Leu Val Ser
            420                 425                 430

Gly Phe Ala Lys Phe Val Gln Asp Ala Lys Val Thr Met Leu Gly Val
        435                 440                 445

Val Pro Ser Ile Val Arg Ser Trp Lys Ser Thr Asn Cys Val Ser Gly
    450                 455                 460

Tyr Asp Trp Ser Thr Ile Arg Cys Phe Ser Ser Ser Gly Glu Ala Ser
465                 470                 475                 480

Asn Val Asp Glu Tyr Leu Trp Leu Met Gly Arg Ala Asn Tyr Lys Pro
                485                 490                 495

Val Ile Glu Met Cys Gly Gly Thr Glu Ile Gly Gly Ala Phe Ser Ala
            500                 505                 510

Gly Ser Phe Leu Gln Ala Gln Ser Leu Ser Ser Phe Ser Ser Gln Cys
        515                 520                 525

Met Gly Cys Thr Leu Tyr Ile Leu Asp Lys Asn Gly Tyr Pro Met Pro
    530                 535                 540

Lys Asn Lys Pro Gly Ile Gly Glu Leu Ala Leu Gly Pro Val Met Phe
545                 550                 555                 560

Gly Ala Ser Lys Thr Leu Leu Asn Gly Asn His His Asp Val Tyr Phe
                565                 570                 575
```

Lys Gly Met Pro Thr Leu Asn Gly Glu Val Leu Arg Arg His Gly Asp
580 585 590

Ile Phe Glu Leu Thr Ser Asn Gly Tyr Tyr His Ala His Gly Arg Ala
595 600 605

Asp Asp Thr Met Asn Ile Gly Gly Ile Lys Ile Ser Ser Ile Glu Ile
610 615 620

Glu Arg Val Cys Asn Glu Val Asp Asp Arg Val Phe Glu Thr Thr Ala
625 630 635 640

Ile Gly Val Pro Pro Leu Gly Gly Gly Pro Glu Gln Leu Val Ile Phe
645 650 655

Phe Val Leu Lys Asp Ser Asn Asp Thr Thr Ile Asp Leu Asn Gln Leu
660 665 670

Arg Leu Ser Phe Asn Leu Gly Leu Gln Lys Lys Leu Asn Pro Leu Phe
675 680 685

Lys Val Thr Arg Val Val Pro Leu Ser Ser Leu Pro Arg Thr Ala Thr
690 695 700

Asn Lys Ile Met Arg Arg Val Leu Arg Gln Gln Phe Ser His Phe Glu
705 710 715 720

<210> 91
<211> 1599
<212> DNA
<213> Cannabis sativa

<400> 91
atggaaaagt ctggttatgg tagagatggt atctacaggt ctttaagacc accattgcat 60

ttgccaaaca acaacaactt gtccatggtc agtttcttgt tcagaaactc ttcttcctac 120

ccacaaaaac cagccttgat tgactctgaa actaatcaaa tcttgtcctt ctcccacttc 180

aaatccaccg ttattaaggt ttctcacggt ttcttgaact tgggtatcaa gaagaatgac 240

tggttgatct acgctccaaa ctctattcat ttcccagttt gctttttggg tattattgct 300

tctggtgcta ttgctactac ttccaaccca ttatacaccg tcagtgaatt gtctaagcaa 360

gtcaaggatt ctaacccaaa gttgattatc accgttccac aattattgga aaaggtcaag 420

ggtttcaact tgccaaccat tttgattggt ccagactcag aacaagaatc ctcttcagat 480

aaggttatga ccttcaacga tttggttaac ttgggtggtt cttctggttc tgaatttcca 540

atcgttgatg acttcaagca atctgatact gctgctttgt tgtactcttc tggtactact 600

ggtatgtcta aaggttggtt gactcacaag aactttatcg cctcttcttt gatggttacc 660

atggaacaag acttggttgg tgaaatggat aacgttttct tgtgcttctt gccaatgttc 720

catgttttcg gtttggccat tattacctac gctcaattgc aaagaggtaa cactgttatt 780

tccgccagat tcgatttgga aaagatgttg aaggacgtcg aaaagtatgt tactcatttg 840

tggtggcctc cagttatttt ggctttgtct aaaaactcca tggttaagtt caacttgtca 900

tccatcaagt acattggttc aggtgctgct ccattgggta aggatttgat ggaagaatgt 960

tctaaatggc catacggtat agttgctcaa ggttacggta tgactgaaac ttgtggtatc 1020

gtttctatgg aagatatcag aggtggtaag agaaattctg gttcagctgg tatgttggct 1080

tcaggtgttg aagctcaaat agtttctgtt gataccttga aaccattgcc accaaatcaa 1140

ttgggtgaaa tttgggttaa gggtccaaat atgatgcaag gttacttcaa caatccacaa 1200

gctaccaagt tgaccattga taagaaaggt tgggttcata ctggtgactt gggttacttt 1260

gatgaagatg gtcacttgta ctgggacaga atcaaagaat tgattaagta caagggtttt 1320

caagtcgctc cagctgaatt ggaaggtttg ttggtttctc atccagaaat attggatgcc 1380

tggattccat ttccagatgc tgaagctggt gaagttccag ttgcttattg gagatcacca 1440

aactcttcat tgactgaaaa cgacgtcaag aagttcattg ctggtcaagt tgcttctttc 1500

aagagattga gaaaggtcac cttcatcaac tctgttccaa aatctgcttc cggtaagatc 1560

ttgagaagag aattgatcca aaaggtcaga tccaatatg 1599

<210> 92
<211> 533
<212> PRT
<213> Cannabis sativa

<400> 92

```
Met Glu Lys Ser Gly Tyr Gly Arg Asp Gly Ile Tyr Arg Ser Leu Arg
1               5                  10                  15

Pro Pro Leu His Leu Pro Asn Asn Asn Asn Leu Ser Met Val Ser Phe
            20                  25                  30

Leu Phe Arg Asn Ser Ser Ser Tyr Pro Gln Lys Pro Ala Leu Ile Asp
        35                  40                  45

Ser Glu Thr Asn Gln Ile Leu Ser Phe Ser His Phe Lys Ser Thr Val
    50                  55                  60

Ile Lys Val Ser His Gly Phe Leu Asn Leu Gly Ile Lys Lys Asn Asp
65                  70                  75                  80

Trp Leu Ile Tyr Ala Pro Asn Ser Ile His Phe Pro Val Cys Phe Leu
                85                  90                  95

Gly Ile Ile Ala Ser Gly Ala Ile Ala Thr Thr Ser Asn Pro Leu Tyr
            100                 105                 110

Thr Val Ser Glu Leu Ser Lys Gln Val Lys Asp Ser Asn Pro Lys Leu
        115                 120                 125

Ile Ile Thr Val Pro Gln Leu Leu Glu Lys Val Lys Gly Phe Asn Leu
    130                 135                 140

Pro Thr Ile Leu Ile Gly Pro Asp Ser Glu Gln Glu Ser Ser Ser Asp
145                 150                 155                 160

Lys Val Met Thr Phe Asn Asp Leu Val Asn Leu Gly Gly Ser Ser Gly
                165                 170                 175

Ser Glu Phe Pro Ile Val Asp Asp Phe Lys Gln Ser Asp Thr Ala Ala
            180                 185                 190

Leu Leu Tyr Ser Ser Gly Thr Thr Gly Met Ser Lys Gly Trp Leu Thr
        195                 200                 205
```

```
His Lys Asn Phe Ile Ala Ser Ser Leu Met Val Thr Met Glu Gln Asp
    210                 215                 220

Leu Val Gly Glu Met Asp Asn Val Phe Leu Cys Phe Leu Pro Met Phe
225                 230                 235                 240

His Val Phe Gly Leu Ala Ile Ile Thr Tyr Ala Gln Leu Gln Arg Gly
                245                 250                 255

Asn Thr Val Ile Ser Ala Arg Phe Asp Leu Glu Lys Met Leu Lys Asp
            260                 265                 270

Val Glu Lys Tyr Val Thr His Leu Trp Trp Pro Pro Val Ile Leu Ala
        275                 280                 285

Leu Ser Lys Asn Ser Met Val Lys Phe Asn Leu Ser Ser Ile Lys Tyr
    290                 295                 300

Ile Gly Ser Gly Ala Ala Pro Leu Gly Lys Asp Leu Met Glu Glu Cys
305                 310                 315                 320

Ser Lys Trp Pro Tyr Gly Ile Val Ala Gln Gly Tyr Gly Met Thr Glu
                325                 330                 335

Thr Cys Gly Ile Val Ser Met Glu Asp Ile Arg Gly Gly Lys Arg Asn
            340                 345                 350

Ser Gly Ser Ala Gly Met Leu Ala Ser Gly Val Glu Ala Gln Ile Val
        355                 360                 365

Ser Val Asp Thr Leu Lys Pro Leu Pro Pro Asn Gln Leu Gly Glu Ile
    370                 375                 380

Trp Val Lys Gly Pro Asn Met Met Gln Gly Tyr Phe Asn Asn Pro Gln
385                 390                 395                 400

Ala Thr Lys Leu Thr Ile Asp Lys Lys Gly Trp Val His Thr Gly Asp
                405                 410                 415
```

Leu Gly Tyr Phe Asp Glu Asp Gly His Leu Tyr Trp Asp Arg Ile Lys
420 425 430

Glu Leu Ile Lys Tyr Lys Gly Phe Gln Val Ala Pro Ala Glu Leu Glu
435 440 445

Gly Leu Leu Val Ser His Pro Glu Ile Leu Asp Ala Trp Ile Pro Phe
450 455 460

Pro Asp Ala Glu Ala Gly Glu Val Pro Val Ala Tyr Trp Arg Ser Pro
465 470 475 480

Asn Ser Ser Leu Thr Glu Asn Asp Val Lys Lys Phe Ile Ala Gly Gln
485 490 495

Val Ala Ser Phe Lys Arg Leu Arg Lys Val Thr Phe Ile Asn Ser Val
500 505 510

Pro Lys Ser Ala Ser Gly Lys Ile Leu Arg Arg Glu Leu Ile Gln Lys
515 520 525

Val Arg Ser Asn Met
530

<210> 93
<211> 744
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial acetoacetyl-CoA reductase (PhaB) nucleotide sequence

<400> 93
atgacgcaga gaatcgccta tgtaacgggt gggatgggtg ggataggaac cgccatatgt 60

cagagactag caaaggacgg attcagggtt gtagccggtt gcggtcctaa tagtccaaga 120

agagagaaat ggttggaaca gcaaaaagct ctaggatttg attttatagc atcagaaggg 180

aatgttgctg actgggattc tacaaagacg gcatttgaca aagtgaaatc tgaagtcggc 240

gaggtcgatg tcctaattaa caacgccggc atcaccagag atgtggtttt caggaagatg 300

actagggctg actgggacgc cgtgatagac acaaatttga cgagcttgtt caacgtcaca 360

aagcaagtaa ttgacggcat ggcagatcgt gggtggggaa ggatagtcaa tatctccagc 420

gtcaacggtc agaaaggcca gttcggacag actaactact ccacagcgaa ggctggctta 480

cacggattca cgatggcctt ggcccaagag gtggctacta aaggggtgac tgtgaacaca 540

gtgtcaccag gatacatcgc gacggatatg gtcaaagcta ttagacaaga tgtcctggac 600

aagattgttg ccactattcc cgtaaagagg cttgggttac cagaagagat agcttcaatt 660

tgcgcttggc tatctagtga ggaatcaggg ttcagcactg gggcggactt ttcattaaac 720

ggtggattac acatgggagg atcc 744

<210> 94
<211> 248
<212> PRT
<213> Artificial Sequence

<220>
<223> Mutant acetoacetyl-CoA reductase (PhaB)

<400> 94

Met Thr Gln Arg Ile Ala Tyr Val Thr Gly Gly Met Gly Gly Ile Gly
1 5 10 15

Thr Ala Ile Cys Gln Arg Leu Ala Lys Asp Gly Phe Arg Val Val Ala
20 25 30

Gly Cys Gly Pro Asn Ser Pro Arg Arg Glu Lys Trp Leu Glu Gln Gln
35 40 45

Lys Ala Leu Gly Phe Asp Phe Ile Ala Ser Glu Gly Asn Val Ala Asp
50 55 60

Trp Asp Ser Thr Lys Thr Ala Phe Asp Lys Val Lys Ser Glu Val Gly
65 70 75 80

Glu Val Asp Val Leu Ile Asn Asn Ala Gly Ile Thr Arg Asp Val Val
85 90 95

Phe Arg Lys Met Thr Arg Ala Asp Trp Asp Ala Val Ile Asp Thr Asn
100 105 110

Leu Thr Ser Leu Phe Asn Val Thr Lys Gln Val Ile Asp Gly Met Ala
115 120 125

Asp Arg Gly Trp Gly Arg Ile Val Asn Ile Ser Ser Val Asn Gly Gln
130 135 140

Lys Gly Gln Phe Gly Gln Thr Asn Tyr Ser Thr Ala Lys Ala Gly Leu
145 150 155 160

His Gly Phe Thr Met Ala Leu Ala Gln Glu Val Ala Thr Lys Gly Val
165 170 175

Thr Val Asn Thr Val Ser Pro Gly Tyr Ile Ala Thr Asp Met Val Lys
180 185 190

Ala Ile Arg Gln Asp Val Leu Asp Lys Ile Val Ala Thr Ile Pro Val
195 200 205

Lys Arg Leu Gly Leu Pro Glu Glu Ile Ala Ser Ile Cys Ala Trp Leu
210 215 220

Ser Ser Glu Glu Ser Gly Phe Ser Thr Gly Ala Asp Phe Ser Leu Asn
225 230 235 240

Gly Gly Leu His Met Gly Gly Ser
245

<210> 95
<211> 408
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial (R)-specific enoyl-CoA hydratase (PhaJ)

<400> 95
atgtctgccc agagtctgga agtcggtcaa aaagcaagac tgtcaaaaag atttggggcg 60

gcagaggtag cggcgttcgc ggcgctgtct gaggatttta atccactgca cttagatcct 120

gcgttcgccg cgacaacagc attcgagagg cccatcgtgc acggcatgct acttgcctct 180

ttgttctcag gtctactggg tcaacagtta cctgggaaag gaagcatcta tctgggacag 240

tcattgtctt ttaagctgcc cgtcttcgtc ggcgatgagg tgacagcaga agtagaagtc 300

acagcattga gggaagacaa gcctattgcg acccttacta ctcgtatttt tactcagggc 360

ggagccttag cagtgacagg agaagctgta gtaaaactac caggatcc 408

<210> 96
<211> 136
<212> PRT
<213> Artificial Sequence

<220>
<223> Mutant (R)-specific enoyl-CoA hydratase (PhaJ)

<400> 96

Met Ser Ala Gln Ser Leu Glu Val Gly Gln Lys Ala Arg Leu Ser Lys
1 5 10 15

Arg Phe Gly Ala Ala Glu Val Ala Ala Phe Ala Ala Leu Ser Glu Asp
20 25 30

Phe Asn Pro Leu His Leu Asp Pro Ala Phe Ala Ala Thr Thr Ala Phe
35 40 45

Glu Arg Pro Ile Val His Gly Met Leu Leu Ala Ser Leu Phe Ser Gly
50 55 60

Leu Leu Gly Gln Gln Leu Pro Gly Lys Gly Ser Ile Tyr Leu Gly Gln
65 70 75 80

Ser Leu Ser Phe Lys Leu Pro Val Phe Val Gly Asp Glu Val Thr Ala
85 90 95

Glu Val Glu Val Thr Ala Leu Arg Glu Asp Lys Pro Ile Ala Thr Leu
100 105 110

Thr Thr Arg Ile Phe Thr Gln Gly Gly Ala Leu Ala Val Thr Gly Glu
115 120 125

Ala Val Val Lys Leu Pro Gly Ser
130 135

<210> 97
<211> 2233
<212> PRT
<213> Artificial Sequence

<220>
<223> Mutated acetyl-CoA carboxylase (ACC1) (S659A, S1157A)

<400> 97

Met Ser Glu Glu Ser Leu Phe Glu Ser Ser Pro Gln Lys Met Glu Tyr
1 5 10 15

Glu Ile Thr Asn Tyr Ser Glu Arg His Thr Glu Leu Pro Gly His Phe
20 25 30

Ile Gly Leu Asn Thr Val Asp Lys Leu Glu Glu Ser Pro Leu Arg Asp
35 40 45

Phe Val Lys Ser His Gly Gly His Thr Val Ile Ser Lys Ile Leu Ile
50 55 60

Ala Asn Asn Gly Ile Ala Ala Val Lys Glu Ile Arg Ser Val Arg Lys
65 70 75 80

Trp Ala Tyr Glu Thr Phe Gly Asp Asp Arg Thr Val Gln Phe Val Ala
85 90 95

Met Ala Thr Pro Glu Asp Leu Glu Ala Asn Ala Glu Tyr Ile Arg Met
100 105 110

Ala Asp Gln Tyr Ile Glu Val Pro Gly Gly Thr Asn Asn Asn Asn Tyr
115 120 125

Ala Asn Val Asp Leu Ile Val Asp Ile Ala Glu Arg Ala Asp Val Asp
130 135 140

Ala Val Trp Ala Gly Trp Gly His Ala Ser Glu Asn Pro Leu Leu Pro
145 150 155 160

```
Glu Lys Leu Ser Gln Ser Lys Arg Lys Val Ile Phe Ile Gly Pro Pro
                165                 170                 175

Gly Asn Ala Met Arg Ser Leu Gly Asp Lys Ile Ser Ser Thr Ile Val
            180                 185                 190

Ala Gln Ser Ala Lys Val Pro Cys Ile Pro Trp Ser Gly Thr Gly Val
        195                 200                 205

Asp Thr Val His Val Asp Glu Lys Thr Gly Leu Val Ser Val Asp Asp
    210                 215                 220

Asp Ile Tyr Gln Lys Gly Cys Cys Thr Ser Pro Glu Asp Gly Leu Gln
225                 230                 235                 240

Lys Ala Lys Arg Ile Gly Phe Pro Val Met Ile Lys Ala Ser Glu Gly
                245                 250                 255

Gly Gly Gly Lys Gly Ile Arg Gln Val Glu Arg Glu Glu Asp Phe Ile
            260                 265                 270

Ala Leu Tyr His Gln Ala Ala Asn Glu Ile Pro Gly Ser Pro Ile Phe
        275                 280                 285

Ile Met Lys Leu Ala Gly Arg Ala Arg His Leu Glu Val Gln Leu Leu
    290                 295                 300

Ala Asp Gln Tyr Gly Thr Asn Ile Ser Leu Phe Gly Arg Asp Cys Ser
305                 310                 315                 320

Val Gln Arg Arg His Gln Lys Ile Ile Glu Glu Ala Pro Val Thr Ile
                325                 330                 335

Ala Lys Ala Glu Thr Phe His Glu Met Glu Lys Ala Ala Val Arg Leu
            340                 345                 350

Gly Lys Leu Val Gly Tyr Val Ser Ala Gly Thr Val Glu Tyr Leu Tyr
        355                 360                 365
```

```
Ser His Asp Asp Gly Lys Phe Tyr Phe Leu Glu Leu Asn Pro Arg Leu
    370                 375                 380

Gln Val Glu His Pro Thr Thr Glu Met Val Ser Gly Val Asn Leu Pro
385                 390                 395                 400

Ala Ala Gln Leu Gln Ile Ala Met Gly Ile Pro Met His Arg Ile Ser
                405                 410                 415

Asp Ile Arg Thr Leu Tyr Gly Met Asn Pro His Ser Ala Ser Glu Ile
            420                 425                 430

Asp Phe Glu Phe Lys Thr Gln Asp Ala Thr Lys Lys Gln Arg Arg Pro
        435                 440                 445

Ile Pro Lys Gly His Cys Thr Ala Cys Arg Ile Thr Ser Glu Asp Pro
    450                 455                 460

Asn Asp Gly Phe Lys Pro Ser Gly Gly Thr Leu His Glu Leu Asn Phe
465                 470                 475                 480

Arg Ser Ser Ser Asn Val Trp Gly Tyr Phe Ser Val Gly Asn Asn Gly
                485                 490                 495

Asn Ile His Ser Phe Ser Asp Ser Gln Phe Gly His Ile Phe Ala Phe
            500                 505                 510

Gly Glu Asn Arg Gln Ala Ser Arg Lys His Met Val Val Ala Leu Lys
        515                 520                 525

Glu Leu Ser Ile Arg Gly Asp Phe Arg Thr Thr Val Glu Tyr Leu Ile
    530                 535                 540

Lys Leu Leu Glu Thr Glu Asp Phe Glu Asp Asn Thr Ile Thr Thr Gly
545                 550                 555                 560

Trp Leu Asp Asp Leu Ile Thr His Lys Met Thr Ala Glu Lys Pro Asp
                565                 570                 575
```

```
Pro Thr Leu Ala Val Ile Cys Gly Ala Ala Thr Lys Ala Phe Leu Ala
            580                 585                 590

Ser Glu Glu Ala Arg His Lys Tyr Ile Glu Ser Leu Gln Lys Gly Gln
        595                 600                 605

Val Leu Ser Lys Asp Leu Leu Gln Thr Met Phe Pro Val Asp Phe Ile
    610                 615                 620

His Glu Gly Lys Arg Tyr Lys Phe Thr Val Ala Lys Ser Gly Asn Asp
625                 630                 635                 640

Arg Tyr Thr Leu Phe Ile Asn Gly Ser Lys Cys Asp Ile Ile Leu Arg
                645                 650                 655

Gln Leu Ala Asp Gly Gly Leu Leu Ile Ala Ile Gly Gly Lys Ser His
            660                 665                 670

Thr Ile Tyr Trp Lys Glu Glu Val Ala Ala Thr Arg Leu Ser Val Asp
        675                 680                 685

Ser Met Thr Thr Leu Leu Glu Val Glu Asn Asp Pro Thr Gln Leu Arg
    690                 695                 700

Thr Pro Ser Pro Gly Lys Leu Val Lys Phe Leu Val Glu Asn Gly Glu
705                 710                 715                 720

His Ile Ile Lys Gly Gln Pro Tyr Ala Glu Ile Glu Val Met Lys Met
                725                 730                 735

Gln Met Pro Leu Val Ser Gln Glu Asn Gly Ile Val Gln Leu Leu Lys
            740                 745                 750

Gln Pro Gly Ser Thr Ile Val Ala Gly Asp Ile Met Ala Ile Met Thr
        755                 760                 765

Leu Asp Asp Pro Ser Lys Val Lys His Ala Leu Pro Phe Glu Gly Met
    770                 775                 780
```

```
Leu Pro Asp Phe Gly Ser Pro Val Ile Glu Gly Thr Lys Pro Ala Tyr
785                 790                 795                 800

Lys Phe Lys Ser Leu Val Ser Thr Leu Glu Asn Ile Leu Lys Gly Tyr
                805                 810                 815

Asp Asn Gln Val Ile Met Asn Ala Ser Leu Gln Gln Leu Ile Glu Val
            820                 825                 830

Leu Arg Asn Pro Lys Leu Pro Tyr Ser Glu Trp Lys Leu His Ile Ser
        835                 840                 845

Ala Leu His Ser Arg Leu Pro Ala Lys Leu Asp Glu Gln Met Glu Glu
    850                 855                 860

Leu Val Ala Arg Ser Leu Arg Arg Gly Ala Val Phe Pro Ala Arg Gln
865                 870                 875                 880

Leu Ser Lys Leu Ile Asp Met Ala Val Lys Asn Pro Glu Tyr Asn Pro
                885                 890                 895

Asp Lys Leu Leu Gly Ala Val Val Glu Pro Leu Ala Asp Ile Ala His
            900                 905                 910

Lys Tyr Ser Asn Gly Leu Glu Ala His Glu His Ser Ile Phe Val His
        915                 920                 925

Phe Leu Glu Glu Tyr Tyr Glu Val Glu Lys Leu Phe Asn Gly Pro Asn
    930                 935                 940

Val Arg Glu Glu Asn Ile Ile Leu Lys Leu Arg Asp Glu Asn Pro Lys
945                 950                 955                 960

Asp Leu Asp Lys Val Ala Leu Thr Val Leu Ser His Ser Lys Val Ser
                965                 970                 975

Ala Lys Asn Asn Leu Ile Leu Ala Ile Leu Lys His Tyr Gln Pro Leu
            980                 985                 990
```

```
Cys Lys Leu Ser Ser Lys Val Ser  Ala Ile Phe Ser Thr  Pro Leu Gln
        995                 1000                1005


His Ile  Val Glu Leu Glu Ser  Lys Ala Thr Ala Lys  Val Ala Leu
    1010                1015                1020


Gln Ala  Arg Glu Ile Leu Ile  Gln Gly Ala Leu Pro  Ser Val Lys
    1025                1030                1035


Glu Arg  Thr Glu Gln Ile Glu  His Ile Leu Lys Ser  Ser Val Val
    1040                1045                1050


Lys Val  Ala Tyr Gly Ser Ser  Asn Pro Lys Arg Ser  Glu Pro Asp
    1055                1060                1065


Leu Asn  Ile Leu Lys Asp Leu  Ile Asp Ser Asn Tyr  Val Val Phe
    1070                1075                1080


Asp Val  Leu Leu Gln Phe Leu  Thr His Gln Asp Pro  Val Val Thr
    1085                1090                1095


Ala Ala  Ala Ala Gln Val Tyr  Ile Arg Arg Ala Tyr  Arg Ala Tyr
    1100                1105                1110


Thr Ile  Gly Asp Ile Arg Val  His Glu Gly Val Thr  Val Pro Ile
    1115                1120                1125


Val Glu  Trp Lys Phe Gln Leu  Pro Ser Ala Ala Phe  Ser Thr Phe
    1130                1135                1140


Pro Thr  Val Lys Ser Lys Met  Gly Met Asn Arg Ala  Val Ser Val
    1145                1150                1155


Ala Asp  Leu Ser Tyr Val Ala  Asn Ser Gln Ser Ser  Pro Leu Arg
    1160                1165                1170


Glu Gly  Ile Leu Met Ala Val  Asp His Leu Asp Asp  Val Asp Glu
    1175                1180                1185
```

```
Ile Leu  Ser Gln Ser Leu Glu  Val Ile Pro Arg His  Gln Ser Ser
    1190                1195                1200

Ser Asn  Gly Pro Ala Pro Asp  Arg Ser Gly Ser Ser  Ala Ser Leu
    1205                1210                1215

Ser Asn  Val Ala Asn Val Cys  Val Ala Ser Thr Glu  Gly Phe Glu
    1220                1225                1230

Ser Glu  Glu Glu Ile Leu Val  Arg Leu Arg Glu Ile  Leu Asp Leu
    1235                1240                1245

Asn Lys  Gln Glu Leu Ile Asn  Ala Ser Ile Arg Arg  Ile Thr Phe
    1250                1255                1260

Met Phe  Gly Phe Lys Asp Gly  Ser Tyr Pro Lys Tyr  Tyr Thr Phe
    1265                1270                1275

Asn Gly  Pro Asn Tyr Asn Glu  Asn Glu Thr Ile Arg  His Ile Glu
    1280                1285                1290

Pro Ala  Leu Ala Phe Gln Leu  Glu Leu Gly Arg Leu  Ser Asn Phe
    1295                1300                1305

Asn Ile  Lys Pro Ile Phe Thr  Asp Asn Arg Asn Ile  His Val Tyr
    1310                1315                1320

Glu Ala  Val Ser Lys Thr Ser  Pro Leu Asp Lys Arg  Phe Phe Thr
    1325                1330                1335

Arg Gly  Ile Ile Arg Thr Gly  His Ile Arg Asp Asp  Ile Ser Ile
    1340                1345                1350

Gln Glu  Tyr Leu Thr Ser Glu  Ala Asn Arg Leu Met  Ser Asp Ile
    1355                1360                1365

Leu Asp  Asn Leu Glu Val Thr  Asp Thr Ser Asn Ser  Asp Leu Asn
    1370                1375                1380
```

```
His Ile  Phe Ile Asn Phe Ile  Ala Val Phe Asp Ile  Ser Pro Glu
    1385                1390                1395

Asp Val  Glu Ala Ala Phe Gly  Gly Phe Leu Glu Arg  Phe Gly Lys
    1400                1405                1410

Arg Leu  Leu Arg Leu Arg Val  Ser Ser Ala Glu Ile  Arg Ile Ile
    1415                1420                1425

Ile Lys  Asp Pro Gln Thr Gly  Ala Pro Val Pro Leu  Arg Ala Leu
    1430                1435                1440

Ile Asn  Asn Val Ser Gly Tyr  Val Ile Lys Thr Glu  Met Tyr Thr
    1445                1450                1455

Glu Val  Lys Asn Ala Lys Gly  Glu Trp Val Phe Lys  Ser Leu Gly
    1460                1465                1470

Lys Pro  Gly Ser Met His Leu  Arg Pro Ile Ala Thr  Pro Tyr Pro
    1475                1480                1485

Val Lys  Glu Trp Leu Gln Pro  Lys Arg Tyr Lys Ala  His Leu Met
    1490                1495                1500

Gly Thr  Thr Tyr Val Tyr Asp  Phe Pro Glu Leu Phe  Arg Gln Ala
    1505                1510                1515

Ser Ser  Ser Gln Trp Lys Asn  Phe Ser Ala Asp Val  Lys Leu Thr
    1520                1525                1530

Asp Asp  Phe Phe Ile Ser Asn  Glu Leu Ile Glu Asp  Glu Asn Gly
    1535                1540                1545

Glu Leu  Thr Glu Val Glu Arg  Glu Pro Gly Ala Asn  Ala Ile Gly
    1550                1555                1560

Met Val  Ala Phe Lys Ile Thr  Val Lys Thr Pro Glu  Tyr Pro Arg
    1565                1570                1575
```

```
Gly Arg  Gln Phe Val Val Val  Ala Asn Asp Ile Thr  Phe Lys Ile
    1580                1585                1590

Gly Ser  Phe Gly Pro Gln Glu  Asp Glu Phe Phe Asn  Lys Val Thr
    1595                1600                1605

Glu Tyr  Ala Arg Lys Arg Gly  Ile Pro Arg Ile Tyr  Leu Ala Ala
    1610                1615                1620

Asn Ser  Gly Ala Arg Ile Gly  Met Ala Glu Glu Ile  Val Pro Leu
    1625                1630                1635

Phe Gln  Val Ala Trp Asn Asp  Ala Ala Asn Pro Asp  Lys Gly Phe
    1640                1645                1650

Gln Tyr  Leu Tyr Leu Thr Ser  Glu Gly Met Glu Thr  Leu Lys Lys
    1655                1660                1665

Phe Asp  Lys Glu Asn Ser Val  Leu Thr Glu Arg Thr  Val Ile Asn
    1670                1675                1680

Gly Glu  Glu Arg Phe Val Ile  Lys Thr Ile Ile Gly  Ser Glu Asp
    1685                1690                1695

Gly Leu  Gly Val Glu Cys Leu  Arg Gly Ser Gly Leu  Ile Ala Gly
    1700                1705                1710

Ala Thr  Ser Arg Ala Tyr His  Asp Ile Phe Thr Ile  Thr Leu Val
    1715                1720                1725

Thr Cys  Arg Ser Val Gly Ile  Gly Ala Tyr Leu Val  Arg Leu Gly
    1730                1735                1740

Gln Arg  Ala Ile Gln Val Glu  Gly Gln Pro Ile Ile  Leu Thr Gly
    1745                1750                1755

Ala Pro  Ala Ile Asn Lys Met  Leu Gly Arg Glu Val  Tyr Thr Ser
    1760                1765                1770
```

```
Asn Leu  Gln Leu Gly Gly Thr  Gln Ile Met Tyr Asn  Asn Gly Val
    1775                 1780                 1785

Ser His  Leu Thr Ala Val Asp  Asp Leu Ala Gly Val  Glu Lys Ile
    1790                 1795                 1800

Val Glu  Trp Met Ser Tyr Val  Pro Ala Lys Arg Asn  Met Pro Val
    1805                 1810                 1815

Pro Ile  Leu Glu Thr Lys Asp  Thr Trp Asp Arg Pro  Val Asp Phe
    1820                 1825                 1830

Thr Pro  Thr Asn Asp Glu Thr  Tyr Asp Val Arg Trp  Met Ile Glu
    1835                 1840                 1845

Gly Arg  Glu Thr Glu Ser Gly  Phe Glu Tyr Gly Leu  Phe Asp Lys
    1850                 1855                 1860

Gly Ser  Phe Phe Glu Thr Leu  Ser Gly Trp Ala Lys  Gly Val Val
    1865                 1870                 1875

Val Gly  Arg Ala Arg Leu Gly  Gly Ile Pro Leu Gly  Val Ile Gly
    1880                 1885                 1890

Val Glu  Thr Arg Thr Val Glu  Asn Leu Ile Pro Ala  Asp Pro Ala
    1895                 1900                 1905

Asn Pro  Asn Ser Ala Glu Thr  Leu Ile Gln Glu Pro  Gly Gln Val
    1910                 1915                 1920

Trp His  Pro Asn Ser Ala Phe  Lys Thr Ala Gln Ala  Ile Asn Asp
    1925                 1930                 1935

Phe Asn  Asn Gly Glu Gln Leu  Pro Met Met Ile Leu  Ala Asn Trp
    1940                 1945                 1950

Arg Gly  Phe Ser Gly Gly Gln  Arg Asp Met Phe Asn  Glu Val Leu
    1955                 1960                 1965
```

```
Lys Tyr  Gly Ser Phe Ile Val  Asp Ala Leu Val Asp  Tyr Lys Gln
    1970                1975                1980

Pro Ile  Ile Ile Tyr Ile Pro  Pro Thr Gly Glu Leu  Arg Gly Gly
    1985                1990                1995

Ser Trp  Val Val Val Asp Pro  Thr Ile Asn Ala Asp  Gln Met Glu
    2000                2005                2010

Met Tyr  Ala Asp Val Asn Ala  Arg Ala Gly Val Leu  Glu Pro Gln
    2015                2020                2025

Gly Met  Val Gly Ile Lys Phe  Arg Arg Glu Lys Leu  Leu Asp Thr
    2030                2035                2040

Met Asn  Arg Leu Asp Asp Lys  Tyr Arg Glu Leu Arg  Ser Gln Leu
    2045                2050                2055

Ser Asn  Lys Ser Leu Ala Pro  Glu Val His Gln Gln  Ile Ser Lys
    2060                2065                2070

Gln Leu  Ala Asp Arg Glu Arg  Glu Leu Leu Pro Ile  Tyr Gly Gln
    2075                2080                2085

Ile Ser  Leu Gln Phe Ala Asp  Leu His Asp Arg Ser  Ser Arg Met
    2090                2095                2100

Val Ala  Lys Gly Val Ile Ser  Lys Glu Leu Glu Trp  Thr Glu Ala
    2105                2110                2115

Arg Arg  Phe Phe Phe Trp Arg  Leu Arg Arg Arg Leu  Asn Glu Glu
    2120                2125                2130

Tyr Leu  Ile Lys Arg Leu Ser  His Gln Val Gly Glu  Ala Ser Arg
    2135                2140                2145

Leu Glu  Lys Ile Ala Arg Ile  Arg Ser Trp Tyr Pro  Ala Ser Val
    2150                2155                2160
```

Asp His Glu Asp Asp Arg Gln Val Ala Thr Trp Ile Glu Glu Asn
2165 2170 2175

Tyr Lys Thr Leu Asp Asp Lys Leu Lys Gly Leu Lys Leu Glu Ser
2180 2185 2190

Phe Ala Gln Asp Leu Ala Lys Lys Ile Arg Ser Asp His Asp Asn
2195 2200 2205

Ala Ile Asp Gly Leu Ser Glu Val Ile Lys Met Leu Ser Thr Asp
2210 2215 2220

Asp Lys Glu Lys Leu Leu Lys Thr Leu Lys
2225 2230

<210> 98
<211> 321
<212> PRT
<213> Artificial Sequence

<220>
<223> Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsGOTt75

<400> 98

Met Ala Ala Thr Thr Asn Gln Thr Glu Pro Pro Glu Ser Asp Asn His
1 5 10 15

Ser Val Ala Thr Lys Ile Leu Asn Phe Gly Lys Ala Cys Trp Lys Leu
20 25 30

Gln Arg Pro Tyr Thr Ile Ile Ala Phe Thr Ser Cys Ala Cys Gly Leu
35 40 45

Phe Gly Lys Glu Leu Leu His Asn Thr Asn Leu Ile Ser Trp Ser Leu
50 55 60

Met Phe Lys Ala Phe Phe Phe Leu Val Ala Ile Leu Cys Ile Ala Ser
65 70 75 80

```
Phe Thr Thr Thr Ile Asn Gln Ile Tyr Asp Leu His Ile Asp Arg Ile
                85                  90                  95

Asn Lys Pro Asp Leu Pro Leu Ala Ser Gly Glu Ile Ser Val Asn Thr
            100                 105                 110

Ala Trp Ile Met Ser Ile Ile Val Ala Leu Phe Gly Leu Ile Ile Thr
        115                 120                 125

Ile Lys Met Lys Gly Gly Pro Leu Tyr Ile Phe Gly Tyr Cys Phe Gly
    130                 135                 140

Ile Phe Gly Gly Ile Val Tyr Ser Val Pro Pro Phe Arg Trp Lys Gln
145                 150                 155                 160

Asn Pro Ser Thr Ala Phe Leu Leu Asn Phe Leu Ala His Ile Ile Thr
                165                 170                 175

Asn Phe Thr Phe Tyr Tyr Ala Ser Arg Ala Ala Leu Gly Leu Pro Phe
            180                 185                 190

Glu Leu Arg Pro Ser Phe Thr Phe Leu Leu Ala Phe Met Lys Ser Met
        195                 200                 205

Gly Ser Ala Leu Ala Leu Ile Lys Asp Ala Ser Asp Val Glu Gly Asp
    210                 215                 220

Thr Lys Phe Gly Ile Ser Thr Leu Ala Ser Lys Tyr Gly Ser Arg Asn
225                 230                 235                 240

Leu Thr Leu Phe Cys Ser Gly Ile Val Leu Leu Ser Tyr Val Ala Ala
                245                 250                 255

Ile Leu Ala Gly Ile Ile Trp Pro Gln Ala Phe Asn Ser Asn Val Met
            260                 265                 270

Leu Leu Ser His Ala Ile Leu Ala Phe Trp Leu Ile Leu Gln Thr Arg
        275                 280                 285
```

Asp Phe Ala Leu Thr Asn Tyr Asp Pro Glu Ala Gly Arg Arg Phe Tyr
290 295 300

Glu Phe Met Trp Lys Leu Tyr Tyr Ala Glu Tyr Leu Val Tyr Val Phe
305 310 315 320

Ile

<210> 99
<211> 363
<212> PRT
<213> Artificial Sequence

<220>
<223> Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsGOTt33

<400> 99

Met Ser His Pro Lys Thr Pro Ile Lys Tyr Ser Tyr Asn Asn Phe Pro
1 5 10 15

Ser Lys His Cys Ser Thr Lys Ser Phe His Leu Gln Asn Lys Cys Ser
20 25 30

Glu Ser Leu Ser Ile Ala Lys Asn Ser Ile Arg Ala Ala Thr Thr Asn
35 40 45

Gln Thr Glu Pro Pro Glu Ser Asp Asn His Ser Val Ala Thr Lys Ile
50 55 60

Leu Asn Phe Gly Lys Ala Cys Trp Lys Leu Gln Arg Pro Tyr Thr Ile
65 70 75 80

Ile Ala Phe Thr Ser Cys Ala Cys Gly Leu Phe Gly Lys Glu Leu Leu
85 90 95

His Asn Thr Asn Leu Ile Ser Trp Ser Leu Met Phe Lys Ala Phe Phe
100 105 110

```
Phe Leu Val Ala Ile Leu Cys Ile Ala Ser Phe Thr Thr Thr Ile Asn
        115                 120                 125


Gln Ile Tyr Asp Leu His Ile Asp Arg Ile Asn Lys Pro Asp Leu Pro
    130                 135                 140


Leu Ala Ser Gly Glu Ile Ser Val Asn Thr Ala Trp Ile Met Ser Ile
145                 150                 155                 160


Ile Val Ala Leu Phe Gly Leu Ile Ile Thr Ile Lys Met Lys Gly Gly
                165                 170                 175


Pro Leu Tyr Ile Phe Gly Tyr Cys Phe Gly Ile Phe Gly Gly Ile Val
            180                 185                 190


Tyr Ser Val Pro Pro Phe Arg Trp Lys Gln Asn Pro Ser Thr Ala Phe
        195                 200                 205


Leu Leu Asn Phe Leu Ala His Ile Ile Thr Asn Phe Thr Phe Tyr Tyr
    210                 215                 220


Ala Ser Arg Ala Ala Leu Gly Leu Pro Phe Glu Leu Arg Pro Ser Phe
225                 230                 235                 240


Thr Phe Leu Leu Ala Phe Met Lys Ser Met Gly Ser Ala Leu Ala Leu
                245                 250                 255


Ile Lys Asp Ala Ser Asp Val Glu Gly Asp Thr Lys Phe Gly Ile Ser
            260                 265                 270


Thr Leu Ala Ser Lys Tyr Gly Ser Arg Asn Leu Thr Leu Phe Cys Ser
        275                 280                 285


Gly Ile Val Leu Leu Ser Tyr Val Ala Ala Ile Leu Ala Gly Ile Ile
    290                 295                 300


Trp Pro Gln Ala Phe Asn Ser Asn Val Met Leu Leu Ser His Ala Ile
305                 310                 315                 320
```

Leu Ala Phe Trp Leu Ile Leu Gln Thr Arg Asp Phe Ala Leu Thr Asn
325 330 335

Tyr Asp Pro Glu Ala Gly Arg Arg Phe Tyr Glu Phe Met Trp Lys Leu
340 345 350

Tyr Tyr Ala Glu Tyr Leu Val Tyr Val Phe Ile
355 360

<210> 100
<211> 323
<212> PRT
<213> Artificial Sequence

<220>
<223> Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4t

<400> 100

Met Ser Ala Gly Ser Asp Gln Ile Glu Gly Ser Pro His His Glu Ser
1 5 10 15

Asp Asn Ser Ile Ala Thr Lys Ile Leu Asn Phe Gly His Thr Cys Trp
20 25 30

Lys Leu Gln Arg Pro Tyr Val Val Lys Gly Met Ile Ser Ile Ala Cys
35 40 45

Gly Leu Phe Gly Arg Glu Leu Phe Asn Asn Arg His Leu Phe Ser Trp
50 55 60

Gly Leu Met Trp Lys Ala Phe Phe Ala Leu Val Pro Ile Leu Ser Phe
65 70 75 80

Asn Phe Phe Ala Ala Ile Met Asn Gln Ile Tyr Asp Val Asp Ile Asp
85 90 95

Arg Ile Asn Lys Pro Asp Leu Pro Leu Val Ser Gly Glu Met Ser Ile
100 105 110

Glu Thr Ala Trp Ile Leu Ser Ile Ile Val Ala Leu Thr Gly Leu Ile

115 120 125

Val Thr Ile Lys Leu Lys Ser Ala Pro Leu Phe Val Phe Ile Tyr Ile
130 135 140

Phe Gly Ile Phe Ala Gly Phe Ala Tyr Ser Val Pro Pro Ile Arg Trp
145 150 155 160

Lys Gln Tyr Pro Phe Thr Asn Phe Leu Ile Thr Ile Ser Ser His Val
165 170 175

Gly Leu Ala Phe Thr Ser Tyr Ser Ala Thr Thr Ser Ala Leu Gly Leu
180 185 190

Pro Phe Val Trp Arg Pro Ala Phe Ser Phe Ile Ile Ala Phe Met Thr
195 200 205

Val Met Gly Met Thr Ile Ala Phe Ala Lys Asp Ile Ser Asp Ile Glu
210 215 220

Gly Asp Ala Lys Tyr Gly Val Ser Thr Val Ala Thr Lys Leu Gly Ala
225 230 235 240

Arg Asn Met Thr Phe Val Val Ser Gly Val Leu Leu Leu Asn Tyr Leu
245 250 255

Val Ser Ile Ser Ile Gly Ile Ile Trp Pro Gln Val Phe Lys Ser Asn
260 265 270

Ile Met Ile Leu Ser His Ala Ile Leu Ala Phe Cys Leu Ile Phe Gln
275 280 285

Thr Arg Glu Leu Ala Leu Ala Asn Tyr Ala Ser Ala Pro Ser Arg Gln
290 295 300

Phe Phe Glu Phe Ile Trp Leu Leu Tyr Tyr Ala Glu Tyr Phe Val Tyr
305 310 315 320

Val Phe Ile

<210> 101
<211> 323
<212> PRT
<213> Artificial Sequence

<220>
<223> Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT7t

<400> 101

Met Ser Thr Asp Thr Ala Asn Gln Thr Glu Pro Pro Glu Ser Asn Thr
1 5 10 15

Lys Tyr Ser Val Val Thr Lys Ile Leu Ser Phe Gly His Thr Cys Trp
20 25 30

Lys Leu Gln Arg Pro Tyr Thr Phe Ile Gly Val Ile Ser Cys Ala Cys
35 40 45

Gly Leu Phe Gly Arg Glu Leu Phe His Asn Thr Asn Leu Leu Ser Trp
50 55 60

Ser Leu Met Leu Lys Ala Phe Ser Ser Leu Met Val Ile Leu Ser Val
65 70 75 80

Asn Leu Cys Thr Asn Ile Ile Asn Gln Ile Thr Asp Leu Asp Ile Asp
85 90 95

Arg Ile Asn Lys Pro Asp Leu Pro Leu Ala Ser Gly Glu Met Ser Ile
100 105 110

Glu Thr Ala Trp Ile Met Ser Ile Ile Val Ala Leu Thr Gly Leu Ile
115 120 125

Leu Thr Ile Lys Leu Asn Cys Gly Pro Leu Phe Ile Ser Leu Tyr Cys
130 135 140

Val Ser Ile Leu Val Gly Ala Leu Tyr Ser Val Pro Pro Phe Arg Trp
145 150 155 160

```
Lys Gln Asn Pro Asn Thr Ala Phe Ser Ser Tyr Phe Met Gly Leu Val
                165                 170                 175

Ile Val Asn Phe Thr Cys Tyr Tyr Ala Ser Arg Ala Ala Phe Gly Leu
            180                 185                 190

Pro Phe Glu Met Ser Pro Pro Phe Thr Phe Ile Leu Ala Phe Val Lys
        195                 200                 205

Ser Met Gly Ser Ala Leu Phe Leu Cys Lys Asp Val Ser Asp Ile Glu
    210                 215                 220

Gly Asp Ser Lys His Gly Ile Ser Thr Leu Ala Thr Arg Tyr Gly Ala
225                 230                 235                 240

Lys Asn Ile Thr Phe Leu Cys Ser Gly Ile Val Leu Leu Thr Tyr Val
                245                 250                 255

Ser Ala Ile Leu Ala Ala Ile Ile Trp Pro Gln Ala Phe Lys Ser Asn
            260                 265                 270

Val Met Leu Leu Ser His Ala Thr Leu Ala Phe Trp Leu Ile Phe Gln
        275                 280                 285

Thr Arg Glu Phe Ala Leu Thr Asn Tyr Asn Pro Glu Ala Gly Arg Lys
    290                 295                 300

Phe Tyr Glu Phe Met Trp Lys Leu His Tyr Ala Glu Tyr Leu Val Tyr
305                 310                 315                 320

Val Phe Ile
```

<210> 102
<211> 315
<212> PRT
<213> Artificial Sequence

<220>

<223> Truncated geranyl pyrophosphate olivetolic acid geranyltransferase HlPT1Lt

<400> 102

```
Met Asp Arg Pro Pro Glu Ser Gly Asn Leu Ser Ala Leu Thr Asn Val
1               5                  10                  15

Lys Asp Phe Val Ser Val Cys Trp Glu Tyr Val Arg Pro Tyr Thr Ala
            20                  25                  30

Lys Gly Val Ile Ile Cys Ser Ser Cys Leu Phe Gly Arg Glu Leu Leu
        35                  40                  45

Glu Asn Pro Asn Leu Phe Ser Trp Pro Leu Ile Phe Arg Ala Leu Leu
    50                  55                  60

Gly Met Leu Ala Ile Leu Gly Ser Cys Phe Tyr Thr Ala Gly Ile Asn
65                  70                  75                  80

Gln Ile Phe Asp Met Asp Ile Asp Arg Ile Asn Lys Pro Asp Leu Pro
                85                  90                  95

Leu Val Ser Gly Arg Ile Ser Val Glu Ser Ala Trp Leu Leu Thr Leu
            100                 105                 110

Ser Pro Ala Ile Ile Gly Phe Ile Leu Ile Leu Lys Leu Asn Ser Gly
        115                 120                 125

Pro Leu Leu Thr Ser Leu Tyr Cys Leu Ala Ile Leu Ser Gly Thr Ile
    130                 135                 140

Tyr Ser Val Pro Pro Phe Arg Trp Lys Lys Asn Pro Ile Thr Ala Phe
145                 150                 155                 160

Leu Cys Ile Leu Met Ile His Ala Gly Leu Asn Phe Ser Val Tyr Tyr
                165                 170                 175

Ala Ser Arg Ala Ala Leu Gly Leu Ala Phe Val Trp Ser Pro Ser Phe
            180                 185                 190
```

```
Ser Phe Ile Thr Ala Phe Ile Thr Phe Met Thr Leu Thr Leu Ala Ser
        195                 200                 205


Ser Lys Asp Leu Ser Asp Ile Asn Gly Asp Arg Lys Phe Gly Val Glu
    210                 215                 220


Thr Phe Ala Thr Lys Leu Gly Ala Lys Asn Ile Thr Leu Leu Gly Thr
225                 230                 235                 240


Gly Leu Leu Leu Leu Asn Tyr Val Ala Ala Ile Ser Thr Ala Ile Ile
                245                 250                 255


Trp Pro Lys Ala Phe Lys Ser Asn Ile Met Leu Leu Ser His Ala Ile
            260                 265                 270


Leu Ala Phe Ser Leu Phe Phe Gln Ala Arg Glu Leu Asp Arg Thr Asn
        275                 280                 285


Tyr Thr Pro Glu Ala Cys Lys Ser Phe Tyr Glu Phe Ile Trp Ile Leu
    290                 295                 300


Phe Ser Ala Glu Tyr Val Val Tyr Leu Phe Ile
305                 310                 315


<210>  103
<211>  329
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Truncated geranyl pyrophosphate olivetolic acid
       geranyltransferase HlPT2t

<400>  103

Met Gly His Leu Pro Arg Pro Asn Ser Leu Thr Ala Trp Ser His Gln
1               5                   10                  15


Ser Glu Phe Pro Ser Thr Ile Val Thr Lys Gly Ser Asn Phe Gly His
            20                  25                  30
```

```
Ala Ser Trp Lys Phe Val Arg Pro Ile Pro Phe Val Ala Val Ser Ile
        35                  40                  45

Ile Cys Thr Ser Leu Phe Gly Ala Glu Leu Leu Lys Asn Pro Asn Leu
    50                  55                  60

Phe Ser Trp Gln Leu Met Phe Asp Ala Phe Gln Gly Leu Val Val Ile
65                  70                  75                  80

Leu Leu Tyr His Ile Tyr Ile Asn Gly Leu Asn Gln Ile Tyr Asp Leu
                85                  90                  95

Glu Ser Asp Arg Ile Asn Lys Pro Asp Leu Pro Leu Ala Ala Glu Glu
            100                 105                 110

Met Ser Val Lys Ser Ala Trp Phe Leu Thr Ile Phe Ser Ala Val Ala
        115                 120                 125

Ser Leu Leu Leu Met Ile Lys Leu Lys Cys Gly Leu Phe Leu Thr Cys
    130                 135                 140

Met Tyr Cys Cys Tyr Leu Val Ile Gly Ala Met Tyr Ser Val Pro Pro
145                 150                 155                 160

Phe Arg Trp Lys Met Asn Thr Phe Thr Ser Thr Leu Trp Asn Phe Ser
                165                 170                 175

Glu Ile Gly Ile Gly Ile Asn Phe Leu Ile Asn Tyr Ala Ser Arg Ala
            180                 185                 190

Thr Leu Gly Leu Pro Phe Gln Trp Arg Pro Pro Phe Thr Phe Ile Ile
        195                 200                 205

Gly Phe Val Ser Thr Leu Ser Ile Ile Leu Ser Ile Leu Lys Asp Val
    210                 215                 220

Pro Asp Val Glu Gly Asp Lys Lys Val Gly Met Ser Thr Leu Pro Val
225                 230                 235                 240
```

```
Ile Phe Gly Ala Arg Thr Ile Val Leu Val Gly Ser Gly Phe Phe Leu
                245                 250                 255


Leu Asn Tyr Val Ala Ala Ile Gly Val Ala Ile Met Trp Pro Gln Ala
            260                 265                 270


Phe Lys Gly Tyr Ile Met Ile Pro Ala His Ala Ile Phe Ala Ser Ala
        275                 280                 285


Leu Ile Phe Lys Thr Trp Leu Leu Asp Lys Ala Asn Tyr Ala Lys Glu
    290                 295                 300


Ala Ser Asp Ser Tyr Tyr His Phe Leu Trp Phe Leu Met Ile Ala Glu
305                 310                 315                 320


Tyr Ile Leu Tyr Pro Phe Ile Ser Thr
                325


<210>  104
<211>  514
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Truncated tetrahydrocannabinolic acid synthase THCASt28

<400>  104

Met Asn Phe Leu Lys Cys Phe Ser Glu Tyr Ile Pro Asn Asn Pro Ala
1               5                   10                  15


Asn Pro Lys Phe Ile Tyr Thr Gln His Asp Gln Leu Tyr Met Ser Val
            20                  25                  30


Leu Asn Ser Thr Ile Gln Asn Leu Arg Phe Thr Ser Asp Thr Thr Pro
        35                  40                  45


Lys Pro Leu Val Ile Val Thr Pro Ser Asn Val Ser His Ile Gln Ala
    50                  55                  60


Ser Ile Leu Cys Ser Lys Lys Val Gly Leu Gln Ile Arg Thr Arg Ser
65                  70                  75                  80
```

```
Gly Gly His Asp Ala Glu Gly Met Ser Tyr Ile Ser Gln Val Pro Phe
                85                  90                  95

Val Val Val Asp Leu Arg Asn Met His Ser Ile Lys Ile Asp Val His
            100                 105                 110

Ser Gln Thr Ala Trp Val Glu Ala Gly Ala Thr Leu Gly Glu Val Tyr
        115                 120                 125

Tyr Trp Ile Asn Glu Lys Asn Glu Asn Phe Ser Phe Pro Gly Gly Tyr
    130                 135                 140

Cys Pro Thr Val Gly Val Gly Gly His Phe Ser Gly Gly Gly Tyr Gly
145                 150                 155                 160

Ala Leu Met Arg Asn Tyr Gly Leu Ala Ala Asp Asn Ile Ile Asp Ala
                165                 170                 175

His Leu Val Asn Val Asp Gly Lys Val Leu Asp Arg Lys Ser Met Gly
            180                 185                 190

Glu Asp Leu Phe Trp Ala Ile Arg Gly Gly Gly Gly Glu Asn Phe Gly
        195                 200                 205

Ile Ile Ala Ala Trp Lys Ile Lys Leu Val Ala Val Pro Ser Lys Ser
    210                 215                 220

Thr Ile Phe Ser Val Lys Lys Asn Met Glu Ile His Gly Leu Val Lys
225                 230                 235                 240

Leu Phe Asn Lys Trp Gln Asn Ile Ala Tyr Lys Tyr Asp Lys Asp Leu
                245                 250                 255

Val Leu Met Thr His Phe Ile Thr Lys Asn Ile Thr Asp Asn His Gly
            260                 265                 270

Lys Asn Lys Thr Thr Val His Gly Tyr Phe Ser Ser Ile Phe His Gly
        275                 280                 285
```

```
Gly Val Asp Ser Leu Val Asp Leu Met Asn Lys Ser Phe Pro Glu Leu
    290                 295                 300

Gly Ile Lys Lys Thr Asp Cys Lys Glu Phe Ser Trp Ile Asp Thr Thr
305                 310                 315                 320

Ile Phe Tyr Ser Gly Val Val Asn Phe Asn Thr Ala Asn Phe Lys Lys
                325                 330                 335

Glu Ile Leu Leu Asp Arg Ser Ala Gly Lys Lys Thr Ala Phe Ser Ile
            340                 345                 350

Lys Leu Asp Tyr Val Lys Lys Pro Ile Pro Glu Thr Ala Met Val Lys
        355                 360                 365

Ile Leu Glu Lys Leu Tyr Glu Glu Asp Val Gly Val Gly Met Tyr Val
    370                 375                 380

Leu Tyr Pro Tyr Gly Gly Ile Met Glu Glu Ile Ser Glu Ser Ala Ile
385                 390                 395                 400

Pro Phe Pro His Arg Ala Gly Ile Met Tyr Glu Leu Trp Tyr Thr Ala
                405                 410                 415

Ser Trp Glu Lys Gln Glu Asp Asn Glu Lys His Ile Asn Trp Val Arg
            420                 425                 430

Ser Val Tyr Asn Phe Thr Thr Pro Tyr Val Ser Gln Asn Pro Arg Leu
        435                 440                 445

Ala Tyr Leu Asn Tyr Arg Asp Leu Asp Leu Gly Lys Thr Asn Pro Glu
    450                 455                 460

Ser Pro Asn Asn Tyr Thr Gln Ala Arg Ile Trp Gly Glu Lys Tyr Phe
465                 470                 475                 480

Gly Lys Asn Phe Asn Arg Leu Val Lys Val Lys Thr Lys Ala Asp Pro
                485                 490                 495
```

Asn Asn Phe Phe Arg Asn Glu Gln Ser Ile Pro Pro Leu Pro Pro His
500 505 510

His His

<210> 105
<211> 517
<212> PRT
<213> Artificial Sequence

<220>
<223> Truncated cannabidiolic acid synthase CBDASt28*

<400> 105

Met Asn Pro Arg Glu Asn Phe Leu Lys Cys Phe Ser Gln Tyr Ile Pro
1 5 10 15

Asn Asn Ala Thr Asn Leu Lys Leu Val Tyr Thr Gln Asn Asn Pro Leu
20 25 30

Tyr Met Ser Val Leu Asn Ser Thr Ile His Asn Leu Arg Phe Thr Ser
35 40 45

Asp Thr Thr Pro Lys Pro Leu Val Ile Val Thr Pro Ser His Val Ser
50 55 60

His Ile Gln Gly Thr Ile Leu Cys Ser Lys Lys Val Gly Leu Gln Ile
65 70 75 80

Arg Thr Arg Ser Gly Gly His Asp Ser Glu Gly Met Ser Tyr Ile Ser
85 90 95

Gln Val Pro Phe Val Ile Val Asp Leu Arg Asn Met Arg Ser Ile Lys
100 105 110

Ile Asp Val His Ser Gln Thr Ala Trp Val Glu Ala Gly Ala Thr Leu
115 120 125

```
Gly Glu Val Tyr Tyr Trp Val Asn Glu Lys Asn Glu Asn Leu Ser Leu
    130                 135                 140

Ala Ala Gly Tyr Cys Pro Thr Val Cys Ala Gly Gly His Phe Gly Gly
145                 150                 155                 160

Gly Gly Tyr Gly Pro Leu Met Arg Asn Tyr Gly Leu Ala Ala Asp Asn
                165                 170                 175

Ile Ile Asp Ala His Leu Val Asn Val His Gly Lys Val Leu Asp Arg
            180                 185                 190

Lys Ser Met Gly Glu Asp Leu Phe Trp Ala Leu Arg Gly Gly Gly Ala
        195                 200                 205

Glu Ser Phe Gly Ile Ile Val Ala Trp Lys Ile Arg Leu Val Ala Val
    210                 215                 220

Pro Lys Ser Thr Met Phe Ser Val Lys Lys Ile Met Glu Ile His Glu
225                 230                 235                 240

Leu Val Lys Leu Val Asn Lys Trp Gln Asn Ile Ala Tyr Lys Tyr Asp
                245                 250                 255

Lys Asp Leu Leu Leu Met Thr His Phe Ile Thr Arg Asn Ile Thr Asp
            260                 265                 270

Asn Gln Gly Lys Asn Lys Thr Ala Ile His Thr Tyr Phe Ser Ser Val
        275                 280                 285

Phe Leu Gly Gly Val Asp Ser Leu Val Asp Leu Met Asn Lys Ser Phe
    290                 295                 300

Pro Glu Leu Gly Ile Lys Lys Thr Asp Cys Arg Gln Leu Ser Trp Ile
305                 310                 315                 320

Asp Thr Ile Ile Phe Tyr Ser Gly Val Val Asn Tyr Asp Thr Asp Asn
                325                 330                 335
```

```
Phe Asn Lys Glu Ile Leu Leu Asp Arg Ser Ala Gly Gln Asn Gly Ala
            340                 345                 350

Phe Lys Ile Lys Leu Asp Tyr Val Lys Lys Pro Ile Pro Glu Ser Val
        355                 360                 365

Phe Val Gln Ile Leu Glu Lys Leu Tyr Glu Glu Asp Ile Gly Ala Gly
    370                 375                 380

Met Tyr Ala Leu Tyr Pro Tyr Gly Gly Ile Met Asp Glu Ile Ser Glu
385                 390                 395                 400

Ser Ala Ile Pro Phe Pro His Arg Ala Gly Ile Leu Tyr Glu Leu Trp
                405                 410                 415

Tyr Ile Cys Ser Trp Glu Lys Gln Glu Asp Asn Glu Lys His Leu Asn
            420                 425                 430

Trp Ile Arg Asn Ile Tyr Asn Phe Met Thr Pro Tyr Val Ser Lys Asn
        435                 440                 445

Pro Arg Leu Ala Tyr Leu Asn Tyr Arg Asp Leu Asp Ile Gly Ile Asn
    450                 455                 460

Asp Pro Lys Asn Pro Asn Asn Tyr Thr Gln Ala Arg Ile Trp Gly Glu
465                 470                 475                 480

Lys Tyr Phe Gly Lys Asn Phe Asp Arg Leu Val Lys Val Lys Thr Leu
                485                 490                 495

Val Asp Pro Asn Asn Phe Phe Arg Asn Glu Gln Ser Ile Pro Pro Leu
            500                 505                 510

Pro Arg His Arg His
        515
```

<210> 106
<211> 2051
<212> PRT
<213> Artificial Sequence

<220>
<223> Mutated fatty acid synthase (FAS1, I306A, R1834K)

<400> 106

```
Met Asp Ala Tyr Ser Thr Arg Pro Leu Thr Leu Ser His Gly Ser Leu
1               5                   10                  15


Glu His Val Leu Leu Val Pro Thr Ala Ser Phe Phe Ile Ala Ser Gln
            20                  25                  30


Leu Gln Glu Gln Phe Asn Lys Ile Leu Pro Glu Pro Thr Glu Gly Phe
        35                  40                  45


Ala Ala Asp Asp Glu Pro Thr Thr Pro Ala Glu Leu Val Gly Lys Phe
    50                  55                  60


Leu Gly Tyr Val Ser Ser Leu Val Glu Pro Ser Lys Val Gly Gln Phe
65                  70                  75                  80


Asp Gln Val Leu Asn Leu Cys Leu Thr Glu Phe Glu Asn Cys Tyr Leu
                85                  90                  95


Glu Gly Asn Asp Ile His Ala Leu Ala Ala Lys Leu Leu Gln Glu Asn
            100                 105                 110


Asp Thr Thr Leu Val Lys Thr Lys Glu Leu Ile Lys Asn Tyr Ile Thr
        115                 120                 125


Ala Arg Ile Met Ala Lys Arg Pro Phe Asp Lys Lys Ser Asn Ser Ala
    130                 135                 140


Leu Phe Arg Ala Val Gly Glu Gly Asn Ala Gln Leu Val Ala Ile Phe
145                 150                 155                 160


Gly Gly Gln Gly Asn Thr Asp Asp Tyr Phe Glu Glu Leu Arg Asp Leu
                165                 170                 175


Tyr Gln Thr Tyr His Val Leu Val Gly Asp Leu Ile Lys Phe Ser Ala
            180                 185                 190
```

```
Glu Thr Leu Ser Glu Leu Ile Arg Thr Thr Leu Asp Ala Glu Lys Val
        195                 200                 205

Phe Thr Gln Gly Leu Asn Ile Leu Glu Trp Leu Glu Asn Pro Ser Asn
    210                 215                 220

Thr Pro Asp Lys Asp Tyr Leu Leu Ser Ile Pro Ile Ser Cys Pro Leu
225                 230                 235                 240

Ile Gly Val Ile Gln Leu Ala His Tyr Val Val Thr Ala Lys Leu Leu
                245                 250                 255

Gly Phe Thr Pro Gly Glu Leu Arg Ser Tyr Leu Lys Gly Ala Thr Gly
            260                 265                 270

His Ser Gln Gly Leu Val Thr Ala Val Ala Ile Ala Glu Thr Asp Ser
        275                 280                 285

Trp Glu Ser Phe Phe Val Ser Val Arg Lys Ala Ile Thr Val Leu Phe
    290                 295                 300

Phe Gly Gly Val Arg Cys Tyr Glu Ala Tyr Pro Asn Thr Ser Leu Pro
305                 310                 315                 320

Pro Ser Ile Leu Glu Asp Ser Leu Glu Asn Asn Glu Gly Val Pro Ser
                325                 330                 335

Pro Met Leu Ser Ile Ser Asn Leu Thr Gln Glu Gln Val Gln Asp Tyr
            340                 345                 350

Val Asn Lys Thr Asn Ser His Leu Pro Ala Gly Lys Gln Val Glu Ile
        355                 360                 365

Ser Leu Val Asn Gly Ala Lys Asn Leu Val Val Ser Gly Pro Pro Gln
    370                 375                 380

Ser Leu Tyr Gly Leu Asn Leu Thr Leu Arg Lys Ala Lys Ala Pro Ser
385                 390                 395                 400
```

```
Gly Leu Asp Gln Ser Arg Ile Pro Phe Ser Glu Arg Lys Leu Lys Phe
                405                 410                 415

Ser Asn Arg Phe Leu Pro Val Ala Ser Pro Phe His Ser His Leu Leu
            420                 425                 430

Val Pro Ala Ser Asp Leu Ile Asn Lys Asp Leu Val Lys Asn Asn Val
        435                 440                 445

Ser Phe Asn Ala Lys Asp Ile Gln Ile Pro Val Tyr Asp Thr Phe Asp
    450                 455                 460

Gly Ser Asp Leu Arg Val Leu Ser Gly Ser Ile Ser Glu Arg Ile Val
465                 470                 475                 480

Asp Cys Ile Ile Arg Leu Pro Val Lys Trp Glu Thr Thr Thr Gln Phe
                485                 490                 495

Lys Ala Thr His Ile Leu Asp Phe Gly Pro Gly Gly Ala Ser Gly Leu
            500                 505                 510

Gly Val Leu Thr His Arg Asn Lys Asp Gly Thr Gly Val Arg Val Ile
        515                 520                 525

Val Ala Gly Thr Leu Asp Ile Asn Pro Asp Asp Asp Tyr Gly Phe Lys
    530                 535                 540

Gln Glu Ile Phe Asp Val Thr Ser Asn Gly Leu Lys Lys Asn Pro Asn
545                 550                 555                 560

Trp Leu Glu Glu Tyr His Pro Lys Leu Ile Lys Asn Lys Ser Gly Lys
                565                 570                 575

Ile Phe Val Glu Thr Lys Phe Ser Lys Leu Ile Gly Arg Pro Pro Leu
            580                 585                 590

Leu Val Pro Gly Met Thr Pro Cys Thr Val Ser Pro Asp Phe Val Ala
        595                 600                 605
```

```
Ala Thr Thr Asn Ala Gly Tyr Thr Ile Glu Leu Ala Gly Gly Gly Tyr
    610                 615                 620

Phe Ser Ala Ala Gly Met Thr Ala Ala Ile Asp Ser Val Val Ser Gln
625                 630                 635                 640

Ile Glu Lys Gly Ser Thr Phe Gly Ile Asn Leu Ile Tyr Val Asn Pro
                645                 650                 655

Phe Met Leu Gln Trp Gly Ile Pro Leu Ile Lys Glu Leu Arg Ser Lys
            660                 665                 670

Gly Tyr Pro Ile Gln Phe Leu Thr Ile Gly Ala Gly Val Pro Ser Leu
        675                 680                 685

Glu Val Ala Ser Glu Tyr Ile Glu Thr Leu Gly Leu Lys Tyr Leu Gly
    690                 695                 700

Leu Lys Pro Gly Ser Ile Asp Ala Ile Ser Gln Val Ile Asn Ile Ala
705                 710                 715                 720

Lys Ala His Pro Asn Phe Pro Ile Ala Leu Gln Trp Thr Gly Gly Arg
                725                 730                 735

Gly Gly Gly His His Ser Phe Glu Asp Ala His Thr Pro Met Leu Gln
            740                 745                 750

Met Tyr Ser Lys Ile Arg Arg His Pro Asn Ile Met Leu Ile Phe Gly
        755                 760                 765

Ser Gly Phe Gly Ser Ala Asp Asp Thr Tyr Pro Tyr Leu Thr Gly Glu
    770                 775                 780

Trp Ser Thr Lys Phe Asp Tyr Pro Pro Met Pro Phe Asp Gly Phe Leu
785                 790                 795                 800

Phe Gly Ser Arg Val Met Ile Ala Lys Glu Val Lys Thr Ser Pro Asp
                805                 810                 815
```

```
Ala Lys Lys Cys Ile Ala Ala Cys Thr Gly Val Pro Asp Asp Lys Trp
            820                 825                 830

Glu Gln Thr Tyr Lys Lys Pro Thr Gly Gly Ile Val Thr Val Arg Ser
        835                 840                 845

Glu Met Gly Glu Pro Ile His Lys Ile Ala Thr Arg Gly Val Met Leu
    850                 855                 860

Trp Lys Glu Phe Asp Glu Thr Ile Phe Asn Leu Pro Lys Asn Lys Leu
865                 870                 875                 880

Val Pro Thr Leu Glu Ala Lys Arg Asp Tyr Ile Ile Ser Arg Leu Asn
                885                 890                 895

Ala Asp Phe Gln Lys Pro Trp Phe Ala Thr Val Asn Gly Gln Ala Arg
            900                 905                 910

Asp Leu Ala Thr Met Thr Tyr Glu Glu Val Ala Lys Arg Leu Val Glu
        915                 920                 925

Leu Met Phe Ile Arg Ser Thr Asn Ser Trp Phe Asp Val Thr Trp Arg
    930                 935                 940

Thr Phe Thr Gly Asp Phe Leu Arg Arg Val Glu Glu Arg Phe Thr Lys
945                 950                 955                 960

Ser Lys Thr Leu Ser Leu Ile Gln Ser Tyr Ser Leu Leu Asp Lys Pro
                965                 970                 975

Asp Glu Ala Ile Glu Lys Val Phe Asn Ala Tyr Pro Ala Ala Arg Glu
            980                 985                 990

Gln Phe Leu Asn Ala Gln Asp Ile  Asp His Phe Leu Ser  Met Cys Gln
        995                 1000                1005

Asn Pro  Met Gln Lys Pro Val  Pro Phe Val Pro Val  Leu Asp Arg
    1010                1015                1020
```

```
Arg Phe  Glu Ile Phe Phe Lys  Lys Asp Ser Leu Trp  Gln Ser Glu
    1025                1030                1035

His Leu  Glu Ala Val Val Asp  Gln Asp Val Gln Arg  Thr Cys Ile
    1040                1045                1050

Leu His  Gly Pro Val Ala Ala  Gln Phe Thr Lys Val  Ile Asp Glu
    1055                1060                1065

Pro Ile  Lys Ser Ile Met Asp  Gly Ile His Asp Gly  His Ile Lys
    1070                1075                1080

Lys Leu  Leu His Gln Tyr Tyr  Gly Asp Asp Glu Ser  Lys Ile Pro
    1085                1090                1095

Ala Val  Glu Tyr Phe Gly Gly  Glu Ser Pro Val Asp  Val Gln Ser
    1100                1105                1110

Gln Val  Asp Ser Ser Ser Val  Ser Glu Asp Ser Ala  Val Phe Lys
    1115                1120                1125

Ala Thr  Ser Ser Thr Asp Glu  Glu Ser Trp Phe Lys  Ala Leu Ala
    1130                1135                1140

Gly Ser  Glu Ile Asn Trp Arg  His Ala Ser Phe Leu  Cys Ser Phe
    1145                1150                1155

Ile Thr  Gln Asp Lys Met Phe  Val Ser Asn Pro Ile  Arg Lys Val
    1160                1165                1170

Phe Lys  Pro Ser Gln Gly Met  Val Val Glu Ile Ser  Asn Gly Asn
    1175                1180                1185

Thr Ser  Ser Lys Thr Val Val  Thr Leu Ser Glu Pro  Val Gln Gly
    1190                1195                1200

Glu Leu  Lys Pro Thr Val Ile  Leu Lys Leu Leu Lys  Glu Asn Ile
    1205                1210                1215
```

```
Ile Gln  Met Glu Met Ile Glu  Asn Arg Thr Met Asp  Gly Lys Pro
    1220                1225                1230

Val Ser  Leu Pro Leu Leu Tyr  Asn Phe Asn Pro Asp  Asn Gly Phe
    1235                1240                1245

Ala Pro  Ile Ser Glu Val Met  Glu Asp Arg Asn Gln  Arg Ile Lys
    1250                1255                1260

Glu Met  Tyr Trp Lys Leu Trp  Ile Asp Glu Pro Phe  Asn Leu Asp
    1265                1270                1275

Phe Asp  Pro Arg Asp Val Ile  Lys Gly Lys Asp Phe  Glu Ile Thr
    1280                1285                1290

Ala Lys  Glu Val Tyr Asp Phe  Thr His Ala Val Gly  Asn Asn Cys
    1295                1300                1305

Glu Asp  Phe Val Ser Arg Pro  Asp Arg Thr Met Leu  Ala Pro Met
    1310                1315                1320

Asp Phe  Ala Ile Val Val Gly  Trp Arg Ala Ile Ile  Lys Ala Ile
    1325                1330                1335

Phe Pro  Asn Thr Val Asp Gly  Asp Leu Leu Lys Leu  Val His Leu
    1340                1345                1350

Ser Asn  Gly Tyr Lys Met Ile  Pro Gly Ala Lys Pro  Leu Gln Val
    1355                1360                1365

Gly Asp  Val Val Ser Thr Thr  Ala Val Ile Glu Ser  Val Val Asn
    1370                1375                1380

Gln Pro  Thr Gly Lys Ile Val  Asp Val Val Gly Thr  Leu Ser Arg
    1385                1390                1395

Asn Gly  Lys Pro Val Met Glu  Val Thr Ser Ser Phe  Phe Tyr Arg
    1400                1405                1410
```

```
Gly Asn  Tyr Thr Asp Phe Glu  Asn Thr Phe Gln Lys  Thr Val Glu
    1415                1420                1425

Pro Val  Tyr Gln Met His Ile  Lys Thr Ser Lys Asp  Ile Ala Val
    1430                1435                1440

Leu Arg  Ser Lys Glu Trp Phe  Gln Leu Asp Asp Glu  Asp Phe Asp
    1445                1450                1455

Leu Leu  Asn Lys Thr Leu Thr  Phe Glu Thr Glu Thr  Glu Val Thr
    1460                1465                1470

Phe Lys  Asn Ala Asn Ile Phe  Ser Ser Val Lys Cys  Phe Gly Pro
    1475                1480                1485

Ile Lys  Val Glu Leu Pro Thr  Lys Glu Thr Val Glu  Ile Gly Ile
    1490                1495                1500

Val Asp  Tyr Glu Ala Gly Ala  Ser His Gly Asn Pro  Val Val Asp
    1505                1510                1515

Phe Leu  Lys Arg Asn Gly Ser  Thr Leu Glu Gln Lys  Val Asn Leu
    1520                1525                1530

Glu Asn  Pro Ile Pro Ile Ala  Val Leu Asp Ser Tyr  Thr Pro Ser
    1535                1540                1545

Thr Asn  Glu Pro Tyr Ala Arg  Val Ser Gly Asp Leu  Asn Pro Ile
    1550                1555                1560

His Val  Ser Arg His Phe Ala  Ser Tyr Ala Asn Leu  Pro Gly Thr
    1565                1570                1575

Ile Thr  His Gly Met Phe Ser  Ser Ala Ser Val Arg  Ala Leu Ile
    1580                1585                1590

Glu Asn  Trp Ala Ala Asp Ser  Val Ser Ser Arg Val  Arg Gly Tyr
    1595                1600                1605
```

```
Thr Cys  Gln Phe Val Asp Met  Val Leu Pro Asn Thr  Ala Leu Lys
    1610                1615                1620

Thr Ser  Ile Gln His Val Gly  Met Ile Asn Gly Arg  Lys Leu Ile
    1625                1630                1635

Lys Phe  Glu Thr Arg Asn Glu  Asp Asp Val Val Val  Leu Thr Gly
    1640                1645                1650

Glu Ala  Glu Ile Glu Gln Pro  Val Thr Thr Phe Val  Phe Thr Gly
    1655                1660                1665

Gln Gly  Ser Gln Glu Gln Gly  Met Gly Met Asp Leu  Tyr Lys Thr
    1670                1675                1680

Ser Lys  Ala Ala Gln Asp Val  Trp Asn Arg Ala Asp  Asn His Phe
    1685                1690                1695

Lys Asp  Thr Tyr Gly Phe Ser  Ile Leu Asp Ile Val  Ile Asn Asn
    1700                1705                1710

Pro Val  Asn Leu Thr Ile His  Phe Gly Gly Glu Lys  Gly Lys Arg
    1715                1720                1725

Ile Arg  Glu Asn Tyr Ser Ala  Met Ile Phe Glu Thr  Ile Val Asp
    1730                1735                1740

Gly Lys  Leu Lys Thr Glu Lys  Ile Phe Lys Glu Ile  Asn Glu His
    1745                1750                1755

Ser Thr  Ser Tyr Thr Phe Arg  Ser Glu Lys Gly Leu  Leu Ser Ala
    1760                1765                1770

Thr Gln  Phe Thr Gln Pro Ala  Leu Thr Leu Met Glu  Lys Ala Ala
    1775                1780                1785

Phe Glu  Asp Leu Lys Ser Lys  Gly Leu Ile Pro Ala  Asp Ala Thr
    1790                1795                1800
```

```
Phe Ala  Gly His Ser Leu Gly  Glu Tyr Ala Ala Leu  Ala Ser Leu
    1805                1810                1815

Ala Asp  Val Met Ser Ile Glu  Ser Leu Val Glu Val  Val Phe Tyr
    1820                1825                1830

Phe Gly  Met Thr Met Gln Val  Ala Val Pro Arg Asp  Glu Leu Gly
    1835                1840                1845

Arg Ser  Asn Tyr Gly Met Ile  Ala Ile Asn Pro Gly  Arg Val Ala
    1850                1855                1860

Ala Ser  Phe Ser Gln Glu Ala  Leu Gln Tyr Val Val  Glu Arg Val
    1865                1870                1875

Gly Lys  Arg Thr Gly Trp Leu  Val Glu Ile Val Asn  Tyr Asn Val
    1880                1885                1890

Glu Asn  Gln Gln Tyr Val Ala  Ala Gly Asp Leu Arg  Ala Leu Asp
    1895                1900                1905

Thr Val  Thr Asn Val Leu Asn  Phe Ile Lys Leu Gln  Lys Ile Asp
    1910                1915                1920

Ile Ile  Glu Leu Gln Lys Ser  Leu Ser Leu Glu Glu  Val Glu Gly
    1925                1930                1935

His Leu  Phe Glu Ile Ile Asp  Glu Ala Ser Lys Lys  Ser Ala Val
    1940                1945                1950

Lys Pro  Arg Pro Leu Lys Leu  Glu Arg Gly Phe Ala  Cys Ile Pro
    1955                1960                1965

Leu Val  Gly Ile Ser Val Pro  Phe His Ser Thr Tyr  Leu Met Asn
    1970                1975                1980

Gly Val  Lys Pro Phe Lys Ser  Phe Leu Lys Lys Asn  Ile Ile Lys
    1985                1990                1995
```

```
Glu Asn  Val Lys Val Ala Arg  Leu Ala Gly Lys Tyr  Ile Pro Asn
    2000                 2005                 2010


Leu Thr  Ala Lys Pro Phe Gln  Val Thr Lys Glu Tyr  Phe Gln Asp
    2015                 2020                 2025


Val Tyr  Asp Leu Thr Gly Ser  Glu Pro Ile Lys Glu  Ile Ile Asp
    2030                 2035                 2040


Asn Trp  Glu Lys Tyr Glu Gln  Ser
    2045                 2050


<210>  107
<211>  1887
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Mutated fatty acid synthase (FAS2, G1250S)

<400>  107

Met Lys Pro Glu Val Glu Gln Glu Leu Ala His Ile Leu Leu Thr Glu
1               5                  10                  15


Leu Leu Ala Tyr Gln Phe Ala Ser Pro Val Arg Trp Ile Glu Thr Gln
            20                  25                  30


Asp Val Phe Leu Lys Asp Phe Asn Thr Glu Arg Val Val Glu Ile Gly
        35                  40                  45


Pro Ser Pro Thr Leu Ala Gly Met Ala Gln Arg Thr Leu Lys Asn Lys
    50                  55                  60


Tyr Glu Ser Tyr Asp Ala Ala Leu Ser Leu His Arg Glu Ile Leu Cys
65                  70                  75                  80


Tyr Ser Lys Asp Ala Lys Glu Ile Tyr Tyr Thr Pro Asp Pro Ser Glu
                85                  90                  95
```

```
Leu Ala Ala Lys Glu Glu Pro Ala Lys Glu Glu Ala Pro Ala Pro Thr
            100                 105                 110


Pro Ala Ala Ser Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Val
        115                 120                 125


Ala Ala Ala Ala Pro Ala Ala Ala Ala Ala Glu Ile Ala Asp Glu Pro
    130                 135                 140


Val Lys Ala Ser Leu Leu Leu His Val Leu Val Ala His Lys Leu Lys
145                 150                 155                 160


Lys Ser Leu Asp Ser Ile Pro Met Ser Lys Thr Ile Lys Asp Leu Val
                165                 170                 175


Gly Gly Lys Ser Thr Val Gln Asn Glu Ile Leu Gly Asp Leu Gly Lys
            180                 185                 190


Glu Phe Gly Thr Thr Pro Glu Lys Pro Glu Glu Thr Pro Leu Glu Glu
        195                 200                 205


Leu Ala Glu Thr Phe Gln Asp Thr Phe Ser Gly Ala Leu Gly Lys Gln
    210                 215                 220


Ser Ser Ser Leu Leu Ser Arg Leu Ile Ser Ser Lys Met Pro Gly Gly
225                 230                 235                 240


Phe Thr Ile Thr Val Ala Arg Lys Tyr Leu Gln Thr Arg Trp Gly Leu
                245                 250                 255


Pro Ser Gly Arg Gln Asp Gly Val Leu Leu Val Ala Leu Ser Asn Glu
            260                 265                 270


Pro Ala Ala Arg Leu Gly Ser Glu Ala Asp Ala Lys Ala Phe Leu Asp
        275                 280                 285


Ser Met Ala Gln Lys Tyr Ala Ser Ile Val Gly Val Asp Leu Ser Ser
    290                 295                 300
```

```
Ala Ala Ser Ala Ser Gly Ala Ala Gly Ala Gly Ala Ala Ala Gly Ala
305                 310                 315                 320


Ala Met Ile Asp Ala Gly Ala Leu Glu Glu Ile Thr Lys Asp His Lys
                325                 330                 335


Val Leu Ala Arg Gln Gln Leu Gln Val Leu Ala Arg Tyr Leu Lys Met
            340                 345                 350


Asp Leu Asp Asn Gly Glu Arg Lys Phe Leu Lys Glu Lys Asp Thr Val
        355                 360                 365


Ala Glu Leu Gln Ala Gln Leu Asp Tyr Leu Asn Ala Glu Leu Gly Glu
    370                 375                 380


Phe Phe Val Asn Gly Val Ala Thr Ser Phe Ser Arg Lys Lys Ala Arg
385                 390                 395                 400


Thr Phe Asp Ser Ser Trp Asn Trp Ala Lys Gln Ser Leu Leu Ser Leu
                405                 410                 415


Tyr Phe Glu Ile Ile His Gly Val Leu Lys Asn Val Asp Arg Glu Val
            420                 425                 430


Val Ser Glu Ala Ile Asn Ile Met Asn Arg Ser Asn Asp Ala Leu Ile
        435                 440                 445


Lys Phe Met Glu Tyr His Ile Ser Asn Thr Asp Glu Thr Lys Gly Glu
    450                 455                 460


Asn Tyr Gln Leu Val Lys Thr Leu Gly Glu Gln Leu Ile Glu Asn Cys
465                 470                 475                 480


Lys Gln Val Leu Asp Val Asp Pro Val Tyr Lys Asp Val Ala Lys Pro
                485                 490                 495


Thr Gly Pro Lys Thr Ala Ile Asp Lys Asn Gly Asn Ile Thr Tyr Ser
            500                 505                 510
```

```
Glu Glu Pro Arg Glu Lys Val Arg Lys Leu Ser Gln Tyr Val Gln Glu
        515                 520                 525

Met Ala Leu Gly Gly Pro Ile Thr Lys Glu Ser Gln Pro Thr Ile Glu
    530                 535                 540

Glu Asp Leu Thr Arg Val Tyr Lys Ala Ile Ser Ala Gln Ala Asp Lys
545                 550                 555                 560

Gln Asp Ile Ser Ser Ser Thr Arg Val Glu Phe Glu Lys Leu Tyr Ser
                565                 570                 575

Asp Leu Met Lys Phe Leu Glu Ser Ser Lys Glu Ile Asp Pro Ser Gln
            580                 585                 590

Thr Thr Gln Leu Ala Gly Met Asp Val Glu Asp Ala Leu Asp Lys Asp
        595                 600                 605

Ser Thr Lys Glu Val Ala Ser Leu Pro Asn Lys Ser Thr Ile Ser Lys
    610                 615                 620

Thr Val Ser Ser Thr Ile Pro Arg Glu Thr Ile Pro Phe Leu His Leu
625                 630                 635                 640

Arg Lys Lys Thr Pro Ala Gly Asp Trp Lys Tyr Asp Arg Gln Leu Ser
                645                 650                 655

Ser Leu Phe Leu Asp Gly Leu Glu Lys Ala Ala Phe Asn Gly Val Thr
            660                 665                 670

Phe Lys Asp Lys Tyr Val Leu Ile Thr Gly Ala Gly Lys Gly Ser Ile
        675                 680                 685

Gly Ala Glu Val Leu Gln Gly Leu Leu Gln Gly Gly Ala Lys Val Val
    690                 695                 700

Val Thr Thr Ser Arg Phe Ser Lys Gln Val Thr Asp Tyr Tyr Gln Ser
705                 710                 715                 720
```

```
Ile Tyr Ala Lys Tyr Gly Ala Lys Gly Ser Thr Leu Ile Val Val Pro
                725                 730                 735

Phe Asn Gln Gly Ser Lys Gln Asp Val Glu Ala Leu Ile Glu Phe Ile
            740                 745                 750

Tyr Asp Thr Glu Lys Asn Gly Gly Leu Gly Trp Asp Leu Asp Ala Ile
        755                 760                 765

Ile Pro Phe Ala Ala Ile Pro Glu Gln Gly Ile Glu Leu Glu His Ile
    770                 775                 780

Asp Ser Lys Ser Glu Phe Ala His Arg Ile Met Leu Thr Asn Ile Leu
785                 790                 795                 800

Arg Met Met Gly Cys Val Lys Lys Gln Lys Ser Ala Arg Gly Ile Glu
                805                 810                 815

Thr Arg Pro Ala Gln Val Ile Leu Pro Met Ser Pro Asn His Gly Thr
            820                 825                 830

Phe Gly Gly Asp Gly Met Tyr Ser Glu Ser Lys Leu Ser Leu Glu Thr
        835                 840                 845

Leu Phe Asn Arg Trp His Ser Glu Ser Trp Ala Asn Gln Leu Thr Val
    850                 855                 860

Cys Gly Ala Ile Ile Gly Trp Thr Arg Gly Thr Gly Leu Met Ser Ala
865                 870                 875                 880

Asn Asn Ile Ile Ala Glu Gly Ile Glu Lys Met Gly Val Arg Thr Phe
                885                 890                 895

Ser Gln Lys Glu Met Ala Phe Asn Leu Leu Gly Leu Leu Thr Pro Glu
            900                 905                 910

Val Val Glu Leu Cys Gln Lys Ser Pro Val Met Ala Asp Leu Asn Gly
        915                 920                 925
```

```
Gly Leu Gln Phe Val Pro Glu Leu Lys Glu Phe Thr Ala Lys Leu Arg
    930                 935                 940

Lys Glu Leu Val Glu Thr Ser Glu Val Arg Lys Ala Val Ser Ile Glu
945                 950                 955                 960

Thr Ala Leu Glu His Lys Val Val Asn Gly Asn Ser Ala Asp Ala Ala
                965                 970                 975

Tyr Ala Gln Val Glu Ile Gln Pro Arg Ala Asn Ile Gln Leu Asp Phe
            980                 985                 990

Pro Glu Leu Lys Pro Tyr Lys Gln  Val Lys Gln Ile Ala  Pro Ala Glu
        995                 1000                1005

Leu Glu  Gly Leu Leu Asp Leu  Glu Arg Val Ile Val  Val Thr Gly
    1010                1015                1020

Phe Ala  Glu Val Gly Pro Trp  Gly Ser Ala Arg Thr  Arg Trp Glu
    1025                1030                1035

Met Glu  Ala Phe Gly Glu Phe  Ser Leu Glu Gly Cys  Val Glu Met
    1040                1045                1050

Ala Trp  Ile Met Gly Phe Ile  Ser Tyr His Asn Gly  Asn Leu Lys
    1055                1060                1065

Gly Arg  Pro Tyr Thr Gly Trp  Val Asp Ser Lys Thr  Lys Glu Pro
    1070                1075                1080

Val Asp  Asp Lys Asp Val Lys  Ala Lys Tyr Glu Thr  Ser Ile Leu
    1085                1090                1095

Glu His  Ser Gly Ile Arg Leu  Ile Glu Pro Glu Leu  Phe Asn Gly
    1100                1105                1110

Tyr Asn  Pro Glu Lys Lys Glu  Met Ile Gln Glu Val  Ile Val Glu
    1115                1120                1125
```

```
Glu Asp  Leu Glu Pro Phe Glu  Ala Ser Lys Glu Thr  Ala Glu Gln
    1130                1135                1140


Phe Lys  His Gln His Gly Asp  Lys Val Asp Ile Phe  Glu Ile Pro
    1145                1150                1155


Glu Thr  Gly Glu Tyr Ser Val  Lys Leu Leu Lys Gly  Ala Thr Leu
    1160                1165                1170


Tyr Ile  Pro Lys Ala Leu Arg  Phe Asp Arg Leu Val  Ala Gly Gln
    1175                1180                1185


Ile Pro  Thr Gly Trp Asn Ala  Lys Thr Tyr Gly Ile  Ser Asp Asp
    1190                1195                1200


Ile Ile  Ser Gln Val Asp Pro  Ile Thr Leu Phe Val  Leu Val Ser
    1205                1210                1215


Val Val  Glu Ala Phe Ile Ala  Ser Gly Ile Thr Asp  Pro Tyr Glu
    1220                1225                1230


Met Tyr  Lys Tyr Val His Val  Ser Glu Val Gly Asn  Cys Ser Gly
    1235                1240                1245


Ser Ser  Met Gly Gly Val Ser  Ala Leu Arg Gly Met  Phe Lys Asp
    1250                1255                1260


Arg Phe  Lys Asp Glu Pro Val  Gln Asn Asp Ile Leu  Gln Glu Ser
    1265                1270                1275


Phe Ile  Asn Thr Met Ser Ala  Trp Val Asn Met Leu  Leu Ile Ser
    1280                1285                1290


Ser Ser  Gly Pro Ile Lys Thr  Pro Val Gly Ala Cys  Ala Thr Ser
    1295                1300                1305


Val Glu  Ser Val Asp Ile Gly  Val Glu Thr Ile Leu  Ser Gly Lys
    1310                1315                1320
```

```
Ala Arg  Ile Cys Ile Val Gly  Gly Tyr Asp Asp Phe  Gln Glu Glu
    1325                1330                1335

Gly Ser  Phe Glu Phe Gly Asn  Met Lys Ala Thr Ser  Asn Thr Leu
    1340                1345                1350

Glu Glu  Phe Glu His Gly Arg  Thr Pro Ala Glu Met  Ser Arg Pro
    1355                1360                1365

Ala Thr  Thr Thr Arg Asn Gly  Phe Met Glu Ala Gln  Gly Ala Gly
    1370                1375                1380

Ile Gln  Ile Ile Met Gln Ala  Asp Leu Ala Leu Lys  Met Gly Val
    1385                1390                1395

Pro Ile  Tyr Gly Ile Val Ala  Met Ala Ala Thr Ala  Thr Asp Lys
    1400                1405                1410

Ile Gly  Arg Ser Val Pro Ala  Pro Gly Lys Gly Ile  Leu Thr Thr
    1415                1420                1425

Ala Arg  Glu His His Ser Ser  Val Lys Tyr Ala Ser  Pro Asn Leu
    1430                1435                1440

Asn Met  Lys Tyr Arg Lys Arg  Gln Leu Val Thr Arg  Glu Ala Gln
    1445                1450                1455

Ile Lys  Asp Trp Val Glu Asn  Glu Leu Glu Ala Leu  Lys Leu Glu
    1460                1465                1470

Ala Glu  Glu Ile Pro Ser Glu  Asp Gln Asn Glu Phe  Leu Leu Glu
    1475                1480                1485

Arg Thr  Arg Glu Ile His Asn  Glu Ala Glu Ser Gln  Leu Arg Ala
    1490                1495                1500

Ala Gln  Gln Gln Trp Gly Asn  Asp Phe Tyr Lys Arg  Asp Pro Arg
    1505                1510                1515
```

```
Ile Ala  Pro Leu Arg Gly Ala  Leu Ala Thr Tyr Gly  Leu Thr Ile
    1520                1525                1530

Asp Asp  Leu Gly Val Ala Ser  Phe His Gly Thr Ser  Thr Lys Ala
    1535                1540                1545

Asn Asp  Lys Asn Glu Ser Ala  Thr Ile Asn Glu Met  Met Lys His
    1550                1555                1560

Leu Gly  Arg Ser Glu Gly Asn  Pro Val Ile Gly Val  Phe Gln Lys
    1565                1570                1575

Phe Leu  Thr Gly His Pro Lys  Gly Ala Ala Gly Ala  Trp Met Met
    1580                1585                1590

Asn Gly  Ala Leu Gln Ile Leu  Asn Ser Gly Ile Ile  Pro Gly Asn
    1595                1600                1605

Arg Asn  Ala Asp Asn Val Asp  Lys Ile Leu Glu Gln  Phe Glu Tyr
    1610                1615                1620

Val Leu  Tyr Pro Ser Lys Thr  Leu Lys Thr Asp Gly  Val Arg Ala
    1625                1630                1635

Val Ser  Ile Thr Ser Phe Gly  Phe Gly Gln Lys Gly  Gly Gln Ala
    1640                1645                1650

Ile Val  Val His Pro Asp Tyr  Leu Tyr Gly Ala Ile  Thr Glu Asp
    1655                1660                1665

Arg Tyr  Asn Glu Tyr Val Ala  Lys Val Ser Ala Arg  Glu Lys Ser
    1670                1675                1680

Ala Tyr  Lys Phe Phe His Asn  Gly Met Ile Tyr Asn  Lys Leu Phe
    1685                1690                1695

Val Ser  Lys Glu His Ala Pro  Tyr Thr Asp Glu Leu  Glu Glu Asp
    1700                1705                1710
```

```
Val Tyr  Leu Asp Pro Leu Ala  Arg Val Ser Lys Asp  Lys Lys Ser
    1715                 1720                 1725


Gly Ser  Leu Thr Phe Asn Ser  Lys Asn Ile Gln Ser  Lys Asp Ser
    1730                 1735                 1740


Tyr Ile  Asn Ala Asn Thr Ile  Glu Thr Ala Lys Met  Ile Glu Asn
    1745                 1750                 1755


Met Thr  Lys Glu Lys Val Ser  Asn Gly Gly Val Gly  Val Asp Val
    1760                 1765                 1770


Glu Leu  Ile Thr Ser Ile Asn  Val Glu Asn Asp Thr  Phe Ile Glu
    1775                 1780                 1785


Arg Asn  Phe Thr Pro Gln Glu  Ile Glu Tyr Cys Ser  Ala Gln Pro
    1790                 1795                 1800


Ser Val  Gln Ser Ser Phe Ala  Gly Thr Trp Ser Ala  Lys Glu Ala
    1805                 1810                 1815


Val Phe  Lys Ser Leu Gly Val  Lys Ser Leu Gly Gly  Gly Ala Ala
    1820                 1825                 1830


Leu Lys  Asp Ile Glu Ile Val  Arg Val Asn Lys Asn  Ala Pro Ala
    1835                 1840                 1845


Val Glu  Leu His Gly Asn Ala  Lys Lys Ala Ala Glu  Glu Ala Gly
    1850                 1855                 1860


Val Thr  Asp Val Lys Val Ser  Ile Ser His Asp Asp  Leu Gln Ala
    1865                 1870                 1875


Val Ala  Val Ala Val Ser Thr  Lys Lys
    1880                 1885
```

<210> 108
<211> 371
<212> PRT
<213> Artificial Sequence

<220>
<223> MBPtag

<400> 108

```
Met Lys Ile Glu Glu Gly Lys Leu Val Ile Trp Ile Asn Gly Asp Lys
1               5                  10                  15


Gly Tyr Asn Gly Leu Ala Glu Val Gly Lys Lys Phe Glu Lys Asp Thr
            20                  25                  30


Gly Ile Lys Val Thr Val Glu His Pro Asp Lys Leu Glu Glu Lys Phe
        35                  40                  45


Pro Gln Val Ala Ala Thr Gly Asp Gly Pro Asp Ile Ile Phe Trp Ala
    50                  55                  60


His Asp Arg Phe Gly Gly Tyr Ala Gln Ser Gly Leu Leu Ala Glu Ile
65                  70                  75                  80


Thr Pro Asp Lys Ala Phe Gln Asp Lys Leu Tyr Pro Phe Thr Trp Asp
                85                  90                  95


Ala Val Arg Tyr Asn Gly Lys Leu Ile Ala Tyr Pro Ile Ala Val Glu
            100                 105                 110


Ala Leu Ser Leu Ile Tyr Asn Lys Asp Leu Leu Pro Asn Pro Pro Lys
        115                 120                 125


Thr Trp Glu Glu Ile Pro Ala Leu Asp Lys Glu Leu Lys Ala Lys Gly
    130                 135                 140


Lys Ser Ala Leu Met Phe Asn Leu Gln Glu Pro Tyr Phe Thr Trp Pro
145                 150                 155                 160


Leu Ile Ala Ala Asp Gly Gly Tyr Ala Phe Lys Tyr Glu Asn Gly Lys
                165                 170                 175


Tyr Asp Ile Lys Asp Val Gly Val Asp Asn Ala Gly Ala Lys Ala Gly
            180                 185                 190
```

```
Leu Thr Phe Leu Val Asp Leu Ile Lys Asn Lys His Met Asn Ala Asp
        195                 200                 205

Thr Asp Tyr Ser Ile Ala Glu Ala Ala Phe Asn Lys Gly Glu Thr Ala
    210                 215                 220

Met Thr Ile Asn Gly Pro Trp Ala Trp Ser Asn Ile Asp Thr Ser Lys
225                 230                 235                 240

Val Asn Tyr Gly Val Thr Val Leu Pro Thr Phe Lys Gly Gln Pro Ser
                245                 250                 255

Lys Pro Phe Val Gly Val Leu Ser Ala Gly Ile Asn Ala Ala Ser Pro
            260                 265                 270

Asn Lys Glu Leu Ala Lys Glu Phe Leu Glu Asn Tyr Leu Leu Thr Asp
        275                 280                 285

Glu Gly Leu Glu Ala Val Asn Lys Asp Lys Pro Leu Gly Ala Val Ala
    290                 295                 300

Leu Lys Ser Tyr Glu Glu Glu Leu Ala Lys Asp Pro Arg Ile Ala Ala
305                 310                 315                 320

Thr Met Glu Asn Ala Gln Lys Gly Glu Ile Met Pro Asn Ile Pro Gln
                325                 330                 335

Met Ser Ala Phe Trp Tyr Ala Val Arg Thr Ala Val Ile Asn Ala Ala
            340                 345                 350

Ser Gly Arg Gln Thr Val Asp Glu Ala Leu Lys Asp Ala Gln Thr Arg
        355                 360                 365

Ile Thr Lys
    370
```

<210> 109
<211> 24

<212> PRT
<213> Artificial Sequence

<220>
<223> ProA tag

<400> 109

Met Ile Phe Asp Gly Thr Thr Met Ser Ile Ala Ile Gly Leu Leu Ser
1 5 10 15

Thr Leu Gly Ile Gly Ala Glu Ala
20

<210> 110
<211> 398
<212> PRT
<213> Cannibis sativa

<400> 110

Met Gly Leu Ser Leu Val Cys Thr Phe Ser Phe Gln Thr Asn Tyr His
1 5 10 15

Thr Leu Leu Asn Pro His Asn Lys Asn Pro Lys Asn Ser Leu Leu Ser
20 25 30

Tyr Gln His Pro Lys Thr Pro Ile Ile Lys Ser Ser Tyr Asp Asn Phe
35 40 45

Pro Ser Lys Tyr Cys Leu Thr Lys Asn Phe His Leu Leu Gly Leu Asn
50 55 60

Ser His Asn Arg Ile Ser Ser Gln Ser Arg Ser Ile Arg Ala Gly Ser
65 70 75 80

Asp Gln Ile Glu Gly Ser Pro His His Glu Ser Asp Asn Ser Ile Ala
85 90 95

Thr Lys Ile Leu Asn Phe Gly His Thr Cys Trp Lys Leu Gln Arg Pro
100 105 110

Tyr Val Val Lys Gly Met Ile Ser Ile Ala Cys Gly Leu Phe Gly Arg

115 120 125

Glu Leu Phe Asn Asn Arg His Leu Phe Ser Trp Gly Leu Met Trp Lys
130 135 140

Ala Phe Phe Ala Leu Val Pro Ile Leu Ser Phe Asn Phe Phe Ala Ala
145 150 155 160

Ile Met Asn Gln Ile Tyr Asp Val Asp Ile Asp Arg Ile Asn Lys Pro
165 170 175

Asp Leu Pro Leu Val Ser Gly Glu Met Ser Ile Glu Thr Ala Trp Ile
180 185 190

Leu Ser Ile Ile Val Ala Leu Thr Gly Leu Ile Val Thr Ile Lys Leu
195 200 205

Lys Ser Ala Pro Leu Phe Val Phe Ile Tyr Ile Phe Gly Ile Phe Ala
210 215 220

Gly Phe Ala Tyr Ser Val Pro Pro Ile Arg Trp Lys Gln Tyr Pro Phe
225 230 235 240

Thr Asn Phe Leu Ile Thr Ile Ser Ser His Val Gly Leu Ala Phe Thr
245 250 255

Ser Tyr Ser Ala Thr Thr Ser Ala Leu Gly Leu Pro Phe Val Trp Arg
260 265 270

Pro Ala Phe Ser Phe Ile Ile Ala Phe Met Thr Val Met Gly Met Thr
275 280 285

Ile Ala Phe Ala Lys Asp Ile Ser Asp Ile Glu Gly Asp Ala Lys Tyr
290 295 300

Gly Val Ser Thr Val Ala Thr Lys Leu Gly Ala Arg Asn Met Thr Phe
305 310 315 320

Val Val Ser Gly Val Leu Leu Leu Asn Tyr Leu Val Ser Ile Ser Ile

325 330 335

Gly Ile Ile Trp Pro Gln Val Phe Lys Ser Asn Ile Met Ile Leu Ser
340 345 350

His Ala Ile Leu Ala Phe Cys Leu Ile Phe Gln Thr Arg Glu Leu Ala
355 360 365

Leu Ala Asn Tyr Ala Ser Ala Pro Ser Arg Gln Phe Phe Glu Phe Ile
370 375 380

Trp Leu Leu Tyr Tyr Ala Glu Tyr Phe Val Tyr Val Phe Ile
385 390 395

<210> 111
<211> 1197
<212> DNA
<213> Cannibis sativa

<400> 111
atgggtttat ctttggtctg caccttctcc tttcaaacta actaccacac tttattgaat 60

ccacataata agaatcctaa gaactcttta ttgtcctacc aacacccaaa gactcctatt 120

atcaagtcct cttacgataa cttcccatct aagtactgtt tgactaagaa tttccatttg 180

ttgggtttga attctcacaa cagaatttcc tcccaatccc gttctattag agccggttct 240

gatcaaatcg aaggttcccc tcatcatgag tccgataact ccattgctac taaaatttta 300

aatttcggtc atacttgttg gaagttgcaa cgtccttacg ttgtcaaggg tatgatctct 360

attgcttgtg gtttgttcgg tagagaattg tttaacaaca gacacttgtt ctcttggggt 420

ttgatgtgga aagctttctt cgctttggtc ccaattttgt ctttcaattt cttcgccgcc 480

atcatgaacc aaatctacga tgttgatatc gaccgtatca acaagccaga cttaccttta 540

gtttccggtg aaatgtccat tgaaactgct tggatcttgt ctatcattgt tgccttgact 600

ggtttaattg ttactattaa gttgaagtcc gctccattgt ttgtcttcat ctacatcttc 660

ggtatcttcg ctggtttcgc ttactccgtc ccacctatta gatggaaaca atatcctttt 720

accaatttct tgatcactat ttcctctcat gttggtttgg ctttcacttc ttactctgcc 780

accacttctg ctttaggttt gcctttcgtt tggcgtcctg ccttctcttt cattattgct 840

ttcatgactg tcatgggtat gactattgcc tttgctaaag acatttctga tatcgaaggt 900

gatgctaagt acggtgtctc taccgttgct accaagttag gtgctagaaa tatgactttt 960

gttgtttctg gtgtcttatt gttgaactac ttggtttcta tctctattgg tatcatttgg 1020

ccacaagttt tcaagtctaa cattatgatc ttgtctcatg ctattttggc tttctgtttg 1080

atctttcaaa ctcgtgaatt agccttagcc aattatgcct ctgccccatc ccgtcaattt 1140

ttcgaattca tctggttgtt atactatgcc gaatacttcg tttacgtctt catttaa 1197

<210> 112
<211> 522
<212> PRT
<213> Artificial Sequence

<220>
<223> Truncated acyl activating enzyme AAE (CsAAE3 truncation)

<400> 112

Met Glu Lys Ser Gly Tyr Gly Arg Asp Gly Ile Tyr Arg Ser Leu Arg
1 5 10 15

Pro Pro Leu His Leu Pro Asn Asn Asn Asn Leu Ser Met Val Ser Phe
20 25 30

Leu Phe Arg Asn Ser Ser Ser Tyr Pro Gln Lys Pro Ala Leu Ile Asp
35 40 45

Ser Glu Thr Asn Gln Ile Leu Ser Phe Ser His Phe Lys Ser Thr Val
50 55 60

Ile Lys Val Ser His Gly Phe Leu Asn Leu Gly Ile Lys Lys Asn Asp
65 70 75 80

Trp Leu Ile Tyr Ala Pro Asn Ser Ile His Phe Pro Val Cys Phe Leu
85 90 95

Gly Ile Ile Ala Ser Gly Ala Ile Ala Thr Thr Ser Asn Pro Leu Tyr
100 105 110

```
Thr Val Ser Glu Leu Ser Lys Gln Val Lys Asp Ser Asn Pro Lys Leu
        115                 120                 125

Ile Ile Thr Val Pro Gln Leu Leu Glu Lys Val Lys Gly Phe Asn Leu
    130                 135                 140

Pro Thr Ile Leu Ile Gly Pro Asp Ser Glu Gln Glu Ser Ser Ser Asp
145                 150                 155                 160

Lys Val Met Thr Phe Asn Asp Leu Val Asn Leu Gly Gly Ser Ser Gly
                165                 170                 175

Ser Glu Phe Pro Ile Val Asp Asp Phe Lys Gln Ser Asp Thr Ala Ala
            180                 185                 190

Leu Leu Tyr Ser Ser Gly Thr Thr Gly Met Ser Lys Gly Trp Leu Thr
        195                 200                 205

His Lys Asn Phe Ile Ala Ser Ser Leu Met Val Thr Met Glu Gln Asp
    210                 215                 220

Leu Val Gly Glu Met Asp Asn Val Phe Leu Cys Phe Leu Pro Met Phe
225                 230                 235                 240

His Val Phe Gly Leu Ala Ile Ile Thr Tyr Ala Gln Leu Gln Arg Gly
                245                 250                 255

Asn Thr Val Ile Ser Ala Arg Phe Asp Leu Glu Lys Met Leu Lys Asp
            260                 265                 270

Val Glu Lys Tyr Val Thr His Leu Trp Trp Pro Pro Val Ile Leu Ala
        275                 280                 285

Leu Ser Lys Asn Ser Met Val Lys Phe Asn Leu Ser Ser Ile Lys Tyr
    290                 295                 300

Ile Gly Ser Gly Ala Ala Pro Leu Gly Lys Asp Leu Met Glu Glu Cys
305                 310                 315                 320
```

```
Ser Lys Trp Pro Tyr Gly Ile Val Ala Gln Gly Tyr Gly Met Thr Glu
                325                 330                 335

Thr Cys Gly Ile Val Ser Met Glu Asp Ile Arg Gly Gly Lys Arg Asn
            340                 345                 350

Ser Gly Ser Ala Gly Met Leu Ala Ser Gly Val Glu Ala Gln Ile Val
        355                 360                 365

Ser Val Asp Thr Leu Lys Pro Leu Pro Pro Asn Gln Leu Gly Glu Ile
    370                 375                 380

Trp Val Lys Gly Pro Asn Met Met Gln Gly Tyr Phe Asn Asn Pro Gln
385                 390                 395                 400

Ala Thr Lys Leu Thr Ile Asp Lys Lys Gly Trp Val His Thr Gly Asp
                405                 410                 415

Leu Gly Tyr Phe Asp Glu Asp Gly His Leu Tyr Trp Asp Arg Ile Lys
            420                 425                 430

Glu Leu Ile Lys Tyr Lys Gly Phe Gln Val Ala Pro Ala Glu Leu Glu
        435                 440                 445

Gly Leu Leu Val Ser His Pro Glu Ile Leu Asp Ala Trp Ile Pro Phe
    450                 455                 460

Pro Asp Ala Glu Ala Gly Glu Val Pro Val Ala Tyr Trp Arg Ser Pro
465                 470                 475                 480

Asn Ser Ser Leu Thr Glu Asn Asp Val Lys Lys Phe Ile Ala Gly Gln
                485                 490                 495

Val Ala Ser Phe Lys Arg Leu Arg Lys Val Thr Phe Ile Asn Ser Val
            500                 505                 510

Pro Lys Ser Ala Ser Gly Lys Ile Leu Arg
        515                 520
```

<210> 113
<211> 526
<212> PRT
<213> Artificial Sequence

<220>
<223> Truncated 3-hydroxy-3-methyl-glutaryl-CoA reductase (tHMG1)

<400> 113

```
Met Ser Asp Gln Leu Val Lys Thr Glu Val Thr Lys Lys Ser Phe Thr
1               5                  10                  15


Ala Pro Val Gln Lys Ala Ser Thr Pro Val Leu Thr Asn Lys Thr Val
            20                  25                  30


Ile Ser Gly Ser Lys Val Lys Ser Leu Ser Ser Ala Gln Ser Ser Ser
        35                  40                  45


Ser Gly Pro Ser Ser Ser Ser Glu Glu Asp Asp Ser Arg Asp Ile Glu
    50                  55                  60


Ser Leu Asp Lys Lys Ile Arg Pro Leu Glu Glu Leu Glu Ala Leu Leu
65                  70                  75                  80


Ser Ser Gly Asn Thr Lys Gln Leu Lys Asn Lys Glu Val Ala Ala Leu
                85                  90                  95


Val Ile His Gly Lys Leu Pro Leu Tyr Ala Leu Glu Lys Lys Leu Gly
            100                 105                 110


Asp Thr Thr Arg Ala Val Ala Val Arg Arg Lys Ala Leu Ser Ile Leu
        115                 120                 125


Ala Glu Ala Pro Val Leu Ala Ser Asp Arg Leu Pro Tyr Lys Asn Tyr
    130                 135                 140


Asp Tyr Asp Arg Val Phe Gly Ala Cys Cys Glu Asn Val Ile Gly Tyr
145                 150                 155                 160


Met Pro Leu Pro Val Gly Val Ile Gly Pro Leu Val Ile Asp Gly Thr
                165                 170                 175
```

```
Ser Tyr His Ile Pro Met Ala Thr Thr Glu Gly Cys Leu Val Ala Ser
            180                 185                 190

Ala Met Arg Gly Cys Lys Ala Ile Asn Ala Gly Gly Gly Ala Thr Thr
        195                 200                 205

Val Leu Thr Lys Asp Gly Met Thr Arg Gly Pro Val Val Arg Phe Pro
    210                 215                 220

Thr Leu Lys Arg Ser Gly Ala Cys Lys Ile Trp Leu Asp Ser Glu Glu
225                 230                 235                 240

Gly Gln Asn Ala Ile Lys Lys Ala Phe Asn Ser Thr Ser Arg Phe Ala
                245                 250                 255

Arg Leu Gln His Ile Gln Thr Cys Leu Ala Gly Asp Leu Leu Phe Met
            260                 265                 270

Arg Phe Arg Thr Thr Thr Gly Asp Ala Met Gly Met Asn Met Ile Ser
        275                 280                 285

Lys Gly Val Glu Tyr Ser Leu Lys Gln Met Val Glu Glu Tyr Gly Trp
    290                 295                 300

Glu Asp Met Glu Val Val Ser Val Ser Gly Asn Tyr Cys Thr Asp Lys
305                 310                 315                 320

Lys Pro Ala Ala Ile Asn Trp Ile Glu Gly Arg Gly Lys Ser Val Val
                325                 330                 335

Ala Glu Ala Thr Ile Pro Gly Asp Val Val Arg Lys Val Leu Lys Ser
            340                 345                 350

Asp Val Ser Ala Leu Val Glu Leu Asn Ile Ala Lys Asn Leu Val Gly
        355                 360                 365

Ser Ala Met Ala Gly Ser Val Gly Gly Phe Asn Ala His Ala Ala Asn
    370                 375                 380
```

```
Leu Val Thr Ala Val Phe Leu Ala Leu Gly Gln Asp Pro Ala Gln Asn
385                 390                 395                 400


Val Glu Ser Ser Asn Cys Ile Thr Leu Met Lys Glu Val Asp Gly Asp
                405                 410                 415


Leu Arg Ile Ser Val Ser Met Pro Ser Ile Glu Val Gly Thr Ile Gly
            420                 425                 430


Gly Gly Thr Val Leu Glu Pro Gln Gly Ala Met Leu Asp Leu Leu Gly
        435                 440                 445


Val Arg Gly Pro His Ala Thr Ala Pro Gly Thr Asn Ala Arg Gln Leu
    450                 455                 460


Ala Arg Ile Val Ala Cys Ala Val Leu Ala Gly Glu Leu Ser Leu Cys
465                 470                 475                 480


Ala Ala Leu Ala Ala Gly His Leu Val Gln Ser His Met Thr His Asn
                485                 490                 495


Arg Lys Pro Ala Glu Pro Thr Lys Pro Asn Asn Leu Asp Ala Thr Asp
            500                 505                 510


Ile Asn Arg Leu Lys Asp Gly Ser Val Thr Cys Ile Lys Ser
        515                 520                 525


<210>  114
<211>  1581
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  Truncated 3-hydroxy-3-methyl-glutaryl-CoA reductase (tHMG1)

<400>  114
atgtccgatc aattagtcaa gaccgaagtc accaagaagt ccttcaccgc cccagttcaa       60

aaagcctcca ctccagtctt aactaacaag accgttattt ctggttccaa ggttaagtct      120

ttatcctccg ctcaatcctc ctcttccggt ccatcttctt cttctgaaga agatgattcc      180
```

cgtgacattg agtccttgga taagaaaatc agacctttgg aagaattaga agctttgtta 240

tcctctggta atactaagca attgaaaaac aaggaagttg ctgctttggt tattcacggt 300

aaattacctt tgtacgcttt agagaagaag ttaggtgaca ctacccgtgc tgtcgccgtt 360

agaagaaaag ctttgtctat tttagctgag gcccctgttt tggcttctga cagattacca 420

tacaagaatt acgattacga tagagttttc ggtgcctgtt gtgagaacgt tatcggttat 480

atgccattac cagtcggtgt tatcggtcca ttggttattg acggtacctc ttaccacatc 540

ccaatggcta ctactgaagg ttgtttagtc gcctccgcca tgagaggttg taaggctatc 600

aatgctggtg gtggtgctac taccgtcttg actaaggatg gtatgactag aggtccagtt 660

gtccgttttc caactttgaa aagatctggt gcttgtaaga tttggttgga ttctgaagaa 720

ggtcaaaatg ccattaagaa ggctttcaat tccacctcta gatttgccag attacaacat 780

attcaaacct gtttagccgg tgatttgttg ttcatgagat tcagaactac tactggtgat 840

gctatgggta tgaacatgat ctctaagggt gtcgaatatt ctttaaaaca aatggttgaa 900

gagtatggtt gggaagacat ggaggtcgtc tctgtctctg gtaactactg tactgataag 960

aaaccagctg ctatcaactg gatcgaaggt cgtggtaagt ctgttgttgc cgaagctact 1020

attccaggtg atgttgttag aaaggtttta aaatccgatg tctctgcctt ggttgagttg 1080

aacattgcta aaaacttggt tggttctgct atggctggtt ctgtcggtgg ttttaatgcc 1140

catgccgcca acttagtcac cgccgttttc ttagctttgg gtcaagatcc agctcaaaat 1200

gtcgaatcct ccaactgtat cactttgatg aaagaggtcg acggtgactt gcgtatctct 1260

gtttccatgc catctatcga agttggtact atcggtggtg gtactgtttt ggagccacaa 1320

ggtgctatgt tggacttatt gggtgttaga ggtccacacg ccactgctcc tggtaccaac 1380

gccagacaat tagctagaat cgttgcctgt gccgtcttag ctggtgagtt gtctttatgt 1440

gctgccttag ctgctggtca cttggtccaa tcccacatga ctcataacag aaagccagct 1500

gaacctacca agcctaacaa cttggatgcc accgatatta atcgtttaaa agatggttct 1560

gtcacttgca ttaagtccta a 1581

<210> 115
<211> 461

<212> PRT
<213> Saccharomyces sp.

<400> 115

```
Met Thr Glu Leu Lys Lys Gln Lys Thr Ala Glu Gln Lys Thr Arg Pro
1               5                  10                  15

Gln Asn Val Gly Ile Lys Gly Ile Gln Ile Tyr Ile Pro Thr Gln Cys
            20                  25                  30

Val Asn Gln Ser Glu Leu Glu Lys Phe Asp Gly Val Ser Gln Gly Lys
        35                  40                  45

Tyr Thr Ile Gly Leu Gly Gln Thr Asn Met Ser Phe Val Asn Asp Arg
    50                  55                  60

Glu Asp Ile Tyr Ser Met Ser Leu Thr Val Leu Ser Lys Leu Ile Lys
65                  70                  75                  80

Ser Tyr Asn Ile Asp Thr Asn Lys Ile Gly Arg Leu Glu Val Gly Thr
                85                  90                  95

Glu Thr Leu Ile Asp Lys Ser Lys Ser Val Lys Ser Val Leu Met Gln
            100                 105                 110

Leu Phe Gly Glu Asn Thr Asp Val Glu Gly Ile Asp Thr Leu Asn Ala
        115                 120                 125

Cys Tyr Gly Gly Thr Asn Ala Leu Phe Asn Ser Leu Asn Trp Ile Glu
    130                 135                 140

Ser Asn Ala Trp Asp Gly Arg Asp Ala Ile Val Val Cys Gly Asp Ile
145                 150                 155                 160

Ala Ile Tyr Asp Lys Gly Ala Ala Arg Pro Thr Gly Gly Ala Gly Thr
                165                 170                 175

Val Ala Met Trp Ile Gly Pro Asp Ala Pro Ile Val Phe Asp Ser Val
            180                 185                 190
```

```
Arg Ala Ser Tyr Met Glu His Ala Tyr Asp Phe Tyr Lys Pro Asp Phe
        195                 200                 205

Thr Ser Glu Tyr Pro Tyr Val Asp Gly His Phe Ser Leu Thr Cys Tyr
    210                 215                 220

Val Lys Ala Leu Asp Gln Val Tyr Lys Ser Tyr Ser Lys Lys Ala Ile
225                 230                 235                 240

Ser Lys Gly Leu Val Ser Asp Pro Ala Gly Ser Asp Ala Leu Asn Val
                245                 250                 255

Leu Lys Tyr Phe Asp Tyr Asn Val Phe His Val Pro Thr Cys Lys Leu
            260                 265                 270

Val Thr Lys Ser Tyr Gly Arg Leu Leu Tyr Asn Asp Phe Arg Ala Asn
        275                 280                 285

Pro Gln Leu Phe Pro Glu Val Asp Ala Glu Leu Ala Thr Arg Asp Tyr
    290                 295                 300

Asp Glu Ser Leu Thr Asp Lys Asn Ile Glu Lys Thr Phe Val Asn Val
305                 310                 315                 320

Ala Lys Pro Phe His Lys Glu Arg Val Ala Gln Ser Leu Ile Val Pro
                325                 330                 335

Thr Asn Thr Gly Asn Met Tyr Thr Ala Ser Val Tyr Ala Ala Phe Ala
            340                 345                 350

Ser Leu Leu Asn Tyr Val Gly Ser Asp Asp Leu Gln Gly Lys Arg Val
        355                 360                 365

Gly Leu Phe Ser Tyr Gly Ser Gly Leu Ala Ala Ser Leu Tyr Ser Cys
    370                 375                 380

Lys Ile Val Gly Asp Val Gln His Ile Ile Lys Glu Leu Asp Ile Thr
385                 390                 395                 400
```

Asn Lys Leu Ala Lys Arg Ile Thr Glu Thr Pro Lys Asp Tyr Glu Ala
405 410 415

Ala Ile Glu Leu Arg Glu Asn Ala His Leu Lys Lys Asn Phe Lys Pro
420 425 430

Gln Gly Ser Ile Glu His Leu Gln Ser Gly Val Tyr Tyr Leu Thr Asn
435 440 445

Ile Asp Asp Lys Phe Arg Arg Ser Tyr Asp Val Lys Lys
450 455 460

<210> 116
<211> 1386
<212> DNA
<213> Saccharomyces sp.

<400> 116
atgaccgaat tgaagaagca aaagactgct gaacaaaaaa cccgtccaca aaatgttggt 60

atcaagggta ttcaaatcta cattccaact caatgcgtca accaatctga attggaaaaa 120

tttgatggtg tttctcaagg taaatacact attggtttgg gtcaaactaa tatgtccttc 180

gttaacgaca gagaagatat ttactccatg tccttgaccg tcttgtccaa attgattaag 240

tcttataata ttgacaccaa caagatcggt agattggaag ttggtactga aactttgatt 300

gataagtcta agtctgttaa gtctgtttta atgcaattgt tcggtgaaaa tactgacgtt 360

gaaggtattg acactttgaa cgcttgttac ggtggtacta atgctttatt taactctttg 420

aactggattg aatccaacgc ttgggacggt agagatgcca ttgtcgtttg tggtgacatt 480

gctatctatg acaagggtgc cgctcgtcca actggtggtg ctggtaccgt tgctatgtgg 540

atcggtcctg acgccccaat cgtttttgat tctgttcgtg cttcttacat ggagcatgct 600

tacgactttt ataagccaga ctttacctcc gaatatccat acgtcgatgg tcatttctct 660

ttgacctgct acgttaaagc cttagatcaa gtctacaagt cttactccaa gaaggccatt 720

tccaagggtt tagtctccga tccagctggt tccgatgctt taaacgtttt aaagtacttc 780

gattacaacg tttttcatgt ccctacttgt aaattggtta ccaaatctta cggtagatta 840

ttgtacaacg atttcagagc taatccacaa ttatttccag aagtcgatgc tgagttggct 900

actagagatt acgacgaatc cttgaccgac aaaaatattg aaaagacttt tgttaacgtt 960

gctaagccat ttcacaaaga gagagttgcc caatctttga ttgtcccaac taatactggt 1020

aatatgtata ctgcttctgt ttacgctgcc tttgcttctt tgttgaacta tgtcggttct 1080

gacgacttac aaggtaagcg tgtcggtttg ttctcctacg gttccggttt ggctgcctct 1140

ttgtattctt gtaagattgt cggtgatgtt caacacatca tcaaggaatt ggatatcacc 1200

aataaattgg ccaagagaat cactgaaact cctaaagact atgaagctgc tatcgaattg 1260

agagaaaatg ctcatttaaa gaaaaacttt aaaccacaag gttctattga acacttgcaa 1320

tccggtgttt actacttaac taacatcgat gacaagttcc gtagatccta cgacgtcaag 1380

aagtaa 1386

<210> 117
<211> 568
<212> PRT
<213> Zymomonas mobilis

<400> 117

Met Ser Tyr Thr Val Gly Thr Tyr Leu Ala Glu Arg Leu Val Gln Ile
1 5 10 15

Gly Leu Lys His His Phe Ala Val Ala Gly Asp Tyr Asn Leu Val Leu
20 25 30

Leu Asp Asn Leu Leu Leu Asn Lys Asn Met Glu Gln Val Tyr Cys Cys
35 40 45

Asn Glu Leu Asn Cys Gly Phe Ser Ala Glu Gly Tyr Ala Arg Ala Lys
50 55 60

Gly Ala Ala Ala Ala Val Val Thr Tyr Ser Val Gly Ala Leu Ser Ala
65 70 75 80

Phe Asp Ala Ile Gly Gly Ala Tyr Ala Glu Asn Leu Pro Val Ile Leu
85 90 95

Ile Ser Gly Ala Pro Asn Asn Asn Asp His Ala Ala Gly His Val Leu

100 105 110

His His Ala Leu Gly Lys Thr Asp Tyr His Tyr Gln Leu Glu Met Ala
115 120 125

Lys Asn Ile Thr Ala Ala Ala Glu Ala Ile Tyr Thr Pro Glu Glu Ala
130 135 140

Pro Ala Lys Ile Asp His Val Ile Lys Thr Ala Leu Arg Glu Lys Lys
145 150 155 160

Pro Val Tyr Leu Glu Ile Ala Cys Asn Ile Ala Ser Met Pro Cys Ala
165 170 175

Ala Pro Gly Pro Ala Ser Ala Leu Phe Asn Asp Glu Ala Ser Asp Glu
180 185 190

Ala Ser Leu Asn Ala Ala Val Glu Glu Thr Leu Lys Phe Ile Ala Asn
195 200 205

Arg Asp Lys Val Ala Val Leu Val Gly Ser Lys Leu Arg Ala Ala Gly
210 215 220

Ala Glu Glu Ala Ala Val Lys Phe Ala Asp Ala Leu Gly Gly Ala Val
225 230 235 240

Ala Thr Met Ala Ala Ala Lys Ser Phe Phe Pro Glu Glu Asn Pro His
245 250 255

Tyr Ile Gly Thr Ser Trp Gly Glu Val Ser Tyr Pro Gly Val Glu Lys
260 265 270

Thr Met Lys Glu Ala Asp Ala Val Ile Ala Leu Ala Pro Val Phe Asn
275 280 285

Asp Tyr Ser Thr Thr Gly Trp Thr Asp Ile Pro Asp Pro Lys Lys Leu
290 295 300

Val Leu Ala Glu Pro Arg Ser Val Val Val Asn Gly Ile Arg Phe Pro

305 310 315 320

Ser Val His Leu Lys Asp Tyr Leu Thr Arg Leu Ala Gln Lys Val Ser
325 330 335

Lys Lys Thr Gly Ala Leu Asp Phe Phe Lys Ser Leu Asn Ala Gly Glu
340 345 350

Leu Lys Lys Ala Ala Pro Ala Asp Pro Ser Ala Pro Leu Val Asn Ala
355 360 365

Glu Ile Ala Arg Gln Val Glu Ala Leu Leu Thr Pro Asn Thr Thr Val
370 375 380

Ile Ala Glu Thr Gly Asp Ser Trp Phe Asn Ala Gln Arg Met Lys Leu
385 390 395 400

Pro Asn Gly Ala Arg Val Glu Tyr Glu Met Gln Trp Gly His Ile Gly
405 410 415

Trp Ser Val Pro Ala Ala Phe Gly Tyr Ala Val Gly Ala Pro Glu Arg
420 425 430

Arg Asn Ile Leu Met Val Gly Asp Gly Ser Phe Gln Leu Thr Ala Gln
435 440 445

Glu Val Ala Gln Met Val Arg Leu Lys Leu Pro Val Ile Ile Phe Leu
450 455 460

Ile Asn Asn Tyr Gly Tyr Thr Ile Glu Val Met Ile His Asp Gly Pro
465 470 475 480

Tyr Asn Asn Ile Lys Asn Trp Asp Tyr Ala Gly Leu Met Glu Val Phe
485 490 495

Asn Gly Asn Gly Gly Tyr Asp Ser Gly Ala Gly Lys Gly Leu Lys Ala
500 505 510

Lys Thr Gly Gly Glu Leu Ala Glu Ala Ile Lys Val Ala Leu Ala Asn

515 520 525

Thr Asp Gly Pro Thr Leu Ile Glu Cys Phe Ile Gly Arg Glu Asp Cys
530 535 540

Thr Glu Glu Leu Val Lys Trp Gly Lys Arg Val Ala Ala Ala Asn Ser
545 550 555 560

Arg Lys Pro Val Asn Lys Leu Leu
565

<210> 118
<211> 1707
<212> DNA
<213> Zymomonas mobilis

<400> 118
atgtcctaca ccgttggtac ctacttagct gagcgtttgg tccaaatcgg tttgaagcac 60

catttcgccg ttgctggtga ttacaacttg gtcttgttag ataatttatt attgaacaag 120

aacatggaac aagtctactg ctgtaatgaa ttgaactgtg gtttctctgc tgaaggttat 180

gctagagcta aaggtgccgc tgccgctgtt gtcacttact ctgttggtgc tttgtctgcc 240

ttcgacgcta ttggtggtgc ttacgccgag aatttacctg ttattttaat ttctggtgcc 300

cctaacaata acgatcatgc tgctggtcat gttttacacc acgctttggg taaaactgac 360

taccattatc aattagagat ggccaaaaac atcaccgccg ctgccgaggc catttacact 420

ccagaagaag ccccagccaa aattgatcac gtcatcaaaa ccgccttgag agagaaaaaa 480

cctgtttact tggaaatcgc ctgtaatatc gcctctatgc cttgcgccgc tcctggtcct 540

gcttccgcct tattcaacga tgaggcttct gatgaagctt ccttaaacgc tgctgttgag 600

gagactttaa agttcatcgc taatagagat aaggtcgctg ttttagtcgg ttctaagttg 660

cgtgctgccg gtgccgagga agctgctgtt aaattcgccg atgctttagg tggtgctgtc 720

gccaccatgg ccgccgccaa atccttttttc cctgaagaaa acccacacta catcggtact 780

tcttggggtg aagtctctta cccaggtgtc gaaaagacta tgaaggaagc cgatgccgtc 840

atcgccttgg ccccagtttt taatgattat tccaccactg gttggactga tatcccagat 900

cctaaaaagt tagttttagc cgagcctaga tccgttgttg ttaacggtat tagattccct 960

tccgttcact tgaaggatta cttaactaga ttggctcaaa aggtttccaa gaagaccggt 1020

gctttggact ttttcaaatc tttgaacgcc ggtgagttaa agaaggccgc ccctgctgac 1080

ccatctgctc cattggttaa cgctgagatt gctagacaag tcgaagcttt attgacccca 1140

aacactaccg ttatcgccga aactggtgac tcttggttta atgctcaaag aatgaagtta 1200

ccaaatggtg ccagagttga gtacgaaatg caatggggtc atatcggttg gtctgtccca 1260

gctgcttttg gttatgctgt tggtgcccct gagagaagaa acatcttgat ggttggtgac 1320

ggttccttcc aattgactgc tcaagaagtc gctcaaatgg ttagattaaa attaccagtc 1380

atcatcttct tgatcaataa ctacggttac actatcgaag tcatgattca cgatggtcct 1440

tacaataata ttaagaactg ggactatgct ggtttgatgg aagtctttaa tggtaacggt 1500

ggttacgatt ccggtgctgg taagggttta aaggctaaga ctggtggtga attagctgaa 1560

gccattaagg ttgccttggc taacaccgac ggtcctactt taatcgaatg tttcattggt 1620

agagaggatt gtaccgaaga gttagttaag tggggtaaga gagttgccgc tgctaattcc 1680

cgtaagcctg tcaataaatt gttataa 1707

<210> 119
<211> 1575
<212> DNA
<213> Artificial Sequence

<220>
<223> Truncated 3-hydroxy-3-methyl-glutaryl-CoA reductase (tHMG1)

<400> 119
atgcaattgg tgaagactga agtcaccaag aagtctttta ctgctcctgt acaaaaggct 60

tctacaccag ttttaaccaa taaaacagtc atttctggat cgaaagtcaa aagtttatca 120

tctgcgcaat cgagctcatc aggaccttca tcatctagtg aggaagatga ttcccgcgat 180

attgaaagct tggataagaa aatacgtcct ttagaagaat tagaagcatt attaagtagt 240

ggaaatacaa aacaattgaa gaacaaagag gtcgctgcct tggttattca cggtaagtta 300

cctttgtacg ctttggagaa aaaattaggt gatactacga gagcggttgc ggtacgtagg 360

aaggctcttt caattttggc agaagctcct gtattagcat ctgatcgttt accatataaa 420

aattatgact acgaccgcgt atttggcgct tgttgtgaaa atgttatagg ttacatgcct 480

ttgcccgttg gtgttatagg ccccttggtt atcgatggta catcttatca tataccaatg 540

gcaactacag agggttgttt ggtagcttct gccatgcgtg gctgtaaggc aatcaatgct 600

ggcggtggtg caacaactgt tttaactaag gatggtatga caagaggccc agtagtccgt 660

ttcccaactt tgaaaagatc tggtgcctgt aagatatggt tagactcaga agagggacaa 720

aacgcaatta aaaaagcttt taactctaca tcaagatttg cacgtctgca acatattcaa 780

acttgtctag caggagattt actcttcatg agatttagaa caactactgg tgacgcaatg 840

ggtatgaata tgatttctaa gggtgtcgaa tactcattaa agcaaatggt agaagagtat 900

ggctgggaag atatggaggt tgtctccgtt tctggtaact actgtaccga caaaaaacca 960

gctgccatca actggatcga aggtcgtggt aagagtgtcg tcgcagaagc tactattcct 1020

ggtgatgttg tcagaaaagt gttaaaaagt gatgtttccg cattggttga gttgaacatt 1080

gctaagaatt tggttggatc tgcaatggct gggtctgttg gtggatttaa cgcacatgca 1140

gctaatttag tgacagctgt tttcttggca ttaggacaag atcctgcaca aaatgtcgaa 1200

agttccaact gtataacatt gatgaaagaa gtggacggtg atttgagaat ttccgtatcc 1260

atgccatcca tcgaagtagg taccatcggt ggtggtactg ttctagaacc acaaggtgcc 1320

atgttggact tattaggtgt aagaggccca catgctaccg ctcctggtac caacgcacgt 1380

caattagcaa gaatagttgc ctgtgccgtc ttggcaggtg aattatcctt atgtgctgcc 1440

ctagcagccg gccatttggt tcaaagtcat atgacccaca acaggaaacc tgctgaacca 1500

acaaaaccta acaatttgga cgccactgat ataaatcgtt tgaaagatgg gtccgtcacc 1560

tgcattaaat cctaa 1575

<210> 120
<211> 1387
<212> DNA
<213> Saccharomyces sp.

<400> 120
atgactgaac taaaaaaaca aaagaccgct gaacaaaaaa ccagacctca aaatgtcggt 60

attaaaggta tccaaattta catcccaact caatgtgtca accaatctga gctagagaaa 120

tttgatggcg tttctcaagg taaatacaca attggtctgg gccaaaccaa catgtctttt 180

gtcaatgaca gagaagatat ctactcgatg tccctaactg ttttgtctaa gttgatcaag 240

agttacaaca tcgacaccaa caaaattggt agattagaag tcggtactga aactctgatt 300

gacaagtcca agtctgtcaa gtctgtcttg atgcaattgt ttggtgaaaa cactgacgtc 360

gaaggtattg acacgcttaa tgcctgttac ggtggtacca acgcgttgtt caactctttg 420

aactggattg aatctaacgc atgggatggt agagacgcca ttgtagtttg cggtgatatt 480

gccatctacg ataagggtgc cgcaagacca accggtggtg ccggtactgt tgctatgtgg 540

atcggtcctg atgctccaat tgtatttgac tctgtaagag cttcttacat ggaacacgcc 600

tacgattttt acaagccaga tttcaccagc gaatatcctt acgtcgatgg tcatttttca 660

ttaacttgtt acgtcaaggc tcttgatcaa gtttacaaga gttattccaa gaaggctatt 720

tctaaagggt tggttagcga tcccgctggt tcggatgctt tgaacgtttt gaaatatttc 780

gactacaacg ttttccatgt tccaacctgt aaattggtca caaaatcata cggtagatta 840

ctatataacg atttcagagc caatcctcaa ttgttcccag aagttgacgc cgaattagct 900

actcgcgatt atgacgaatc tttaaccgat aagaacattg aaaaaacttt tgttaatgtt 960

gctaagccat tccacaaaga gagagttgcc caatctttga ttgttccaac aaacacaggt 1020

aacatgtaca ccgcatctgt ttatgccgcc tttgcatctc tattaaacta tgttggatct 1080

gacgacttac aaggcaagcg tgttggttta ttttcttacg gttccggttt agctgcatct 1140

ctatattctt gcaaaattgt tggtgacgtc caacatatta tcaaggaatt agatattact 1200

aacaaattag ccaagagaat caccgaaact ccaaaggatt acgaagctgc catcgaattg 1260

agagaaaatg cccatttgaa gaagaacttc aaacctcaag gttccattga gcatttgcaa 1320

agtggtgttt actacttgac caacatcgat gacaaattta gaagatctta cgatgttaaa 1380

aaataat 1387

<210> 121
<211> 420
<212> PRT
<213> Catharanthus sp.

<400> 121

Met Leu Phe Ser Arg Gly Leu Tyr Arg Ile Ala Arg Thr Ser Leu Asn
1 5 10 15

```
Arg Ser Arg Leu Leu Tyr Pro Leu Gln Ser Gln Ser Pro Glu Leu Leu
            20                  25                  30

Gln Ser Phe Gln Phe Arg Ser Pro Ile Gly Ser Ser Gln Lys Val Ser
        35                  40                  45

Gly Phe Arg Val Ile Tyr Ser Cys Val Ser Ser Ala Leu Ala Asn Val
    50                  55                  60

Gly Gln Gln Val Gln Arg Gln Ser Asn Ser Val Ala Glu Glu Pro Leu
65                  70                  75                  80

Asp Pro Phe Ser Leu Val Ala Asp Glu Leu Ser Ile Leu Ala Asn Arg
                85                  90                  95

Leu Arg Ser Met Val Val Ala Glu Val Pro Lys Leu Ala Ser Ala Ala
            100                 105                 110

Glu Tyr Phe Phe Lys Leu Gly Val Glu Gly Lys Arg Phe Arg Pro Thr
        115                 120                 125

Val Leu Leu Leu Met Ala Thr Ala Ile Asp Ala Pro Ile Ser Arg Thr
    130                 135                 140

Pro Pro Asp Thr Ser Leu Asp Thr Leu Ser Thr Glu Leu Arg Leu Arg
145                 150                 155                 160

Gln Gln Ser Ile Ala Glu Ile Thr Glu Met Ile His Val Ala Ser Leu
                165                 170                 175

Leu His Asp Asp Val Leu Asp Asp Ala Glu Thr Arg Arg Gly Ile Gly
            180                 185                 190

Ser Leu Asn Phe Val Met Gly Asn Lys Leu Ala Val Leu Ala Gly Asp
        195                 200                 205

Phe Leu Leu Ser Arg Ala Cys Val Ala Leu Ala Ser Leu Lys Asn Thr
    210                 215                 220
```

```
Glu Val Val Ser Leu Leu Ala Thr Val Val Glu His Leu Val Thr Gly
225                 230                 235                 240

Glu Thr Met Gln Met Thr Thr Thr Ser Asp Gln Arg Cys Ser Met Glu
                245                 250                 255

Tyr Tyr Met Gln Lys Thr Tyr Tyr Lys Thr Ala Ser Leu Ile Ser Asn
            260                 265                 270

Ser Cys Lys Ala Ile Ala Leu Leu Ala Gly Gln Thr Ser Glu Val Ala
        275                 280                 285

Met Leu Ala Tyr Glu Tyr Gly Lys Asn Leu Gly Leu Ala Phe Gln Leu
    290                 295                 300

Ile Asp Asp Val Leu Asp Phe Thr Gly Thr Ser Ala Ser Leu Gly Lys
305                 310                 315                 320

Gly Ser Leu Ser Asp Ile Arg His Gly Ile Val Thr Ala Pro Ile Leu
                325                 330                 335

Phe Ala Ile Glu Glu Phe Pro Glu Leu Arg Ala Val Val Asp Glu Gly
            340                 345                 350

Phe Glu Asn Pro Tyr Asn Val Asp Leu Ala Leu His Tyr Leu Gly Lys
        355                 360                 365

Ser Arg Gly Ile Gln Arg Thr Arg Glu Leu Ala Ile Lys His Ala Asn
    370                 375                 380

Leu Ala Ser Asp Ala Ile Asp Ser Leu Pro Val Thr Asp Asp Glu His
385                 390                 395                 400

Val Leu Arg Ser Arg Arg Ala Leu Val Glu Leu Thr Gln Arg Val Ile
                405                 410                 415

Thr Arg Arg Lys
            420
```

<210> 122
<211> 1263
<212> DNA
<213> Catharanthus sp.

<400> 122
atgttattct ctcgtggttt atacagaatc gccagaactt ctttgaacag atcccgtttg 60

ttgtaccctt tacaatctca atctcctgaa ttgttacaat ccttccaatt cagatctcca 120

atcggttcct ctcaaaaggt ttccggtttc agagttatct actcctgcgt ttcctctgct 180

ttagctaacg ttggtcaaca agtccaaaga caatctaatt ccgttgctga agaacctttg 240

gacccattct ccttggttgc cgatgaatta tccattttag ctaacagatt gcgttctatg 300

gtcgtcgctg aagttccaaa gttagcctcc gccgccgaat atttcttcaa gttgggtgtc 360

gagggtaaaa gattcagacc aactgttttg ttgttaatgg ccaccgccat tgatgcccca 420

atctctagaa ccccacctga cacctcctta gatactttat ccaccgaatt gcgtttgaga 480

caacaatcta tcgccgaaat tactgaaatg attcatgtcg cttccttgtt gcacgatgat 540

gttttggatg atgctgaaac tagaagaggt attggttctt taaattttgt catgggtaac 600

aaattggctg ttttggccgg tgacttctta ttatctagag cttgtgttgc cttagcttct 660

ttgaaaaaca ctgaagtcgt ctccttgtta gccactgtcg ttgaacactt agttactggt 720

gagactatgc aaatgactac cacctccgat caaagatgtt ctatggaata ctacatgcaa 780

aagacctatt acaagactgc ctctttgatt tctaactcct gtaaagccat tgccttgtta 840

gctggtcaaa cttctgaagt tgccatgttg gcttacgaat acggtaaaaa cttgggtttg 900

gctttccaat tgattgatga tgttttggat ttcactggta cttctgcttc cttaggtaaa 960

ggttctttgt ctgatattcg tcacggtatc gttaccgccc caatcttgtt cgctattgaa 1020

gaattcccag agttaagagc tgttgttgac gaaggtttcg aaaaccctta caatgttgac 1080

ttagccttgc actacttggg taaatctaga ggtattcaac gtaccagaga attagccatt 1140

aaacatgcta acttagcctc tgacgccatt gactctttac cagtcactga tgatgagcac 1200

gtcttacgtt ccagacgtgc cttagttgaa ttgactcaaa gagttattac tagaagaaag 1260

taa 1263

<210> 123
<211> 421
<212> PRT
<213> Mangifera indica

<400> 123

```
Met Leu Phe Ser Tyr Gly Leu Ser Arg Ile Ser Ile Asn Pro Arg Ala
1               5                   10                  15


Ser Leu Leu Thr Cys Arg Trp Leu Leu Ser His Leu Thr Gly Ser Leu
            20                  25                  30


Ser Pro Ser Thr Ser Ser His Thr Ile Ser Asp Ser Val His Lys Val
        35                  40                  45


Trp Gly Cys Arg Glu Ala Tyr Thr Trp Ser Val Pro Ala Leu His Gly
    50                  55                  60


Phe Arg His Gln Ile His His Gln Ser Ser Ser Leu Ile Glu Asp Gln
65                  70                  75                  80


Leu Asp Pro Phe Ser Leu Val Ala Asp Glu Leu Ser Leu Val Ala Asn
                85                  90                  95


Arg Leu Arg Ser Met Val Val Thr Glu Val Pro Lys Leu Ala Ser Ala
            100                 105                 110


Ala Glu Tyr Phe Phe Lys Met Gly Val Glu Gly Lys Arg Phe Arg Pro
        115                 120                 125


Ala Val Leu Leu Leu Met Ala Thr Ala Leu Asn Val His Val Leu Glu
    130                 135                 140


Pro Leu Pro Glu Gly Ala Gly Asp Ala Leu Met Thr Glu Leu Arg Thr
145                 150                 155                 160


Arg Gln Gln Cys Ile Ala Glu Ile Thr Glu Met Ile His Val Ala Ser
                165                 170                 175
```

```
Leu Leu His Asp Asp Val Leu Asp Asp Ala Asp Thr Arg Arg Gly Ile
            180                 185                 190

Gly Ser Leu Asn Leu Val Met Gly Asn Lys Leu Ala Val Leu Ala Gly
        195                 200                 205

Asp Phe Leu Leu Ser Arg Ala Cys Val Ala Leu Ala Ser Leu Lys Asn
    210                 215                 220

Thr Glu Val Val Ser Leu Leu Ala Thr Val Val Glu His Leu Val Thr
225                 230                 235                 240

Gly Glu Thr Met Gln Met Thr Thr Ser Ser Asp Gln Arg Cys Ser Met
                245                 250                 255

Glu Tyr Tyr Met Gln Lys Thr Tyr Tyr Lys Thr Ala Ser Leu Ile Ser
            260                 265                 270

Asn Ser Cys Lys Ala Ile Ala Leu Leu Ala Gly Gln Ser Ala Glu Val
        275                 280                 285

Ala Met Leu Ala Phe Glu Phe Gly Lys Asn Leu Gly Leu Ala Tyr Gln
    290                 295                 300

Leu Ile Asp Asp Val Leu Asp Phe Thr Gly Thr Ser Ala Ser Leu Gly
305                 310                 315                 320

Lys Gly Ser Leu Ser Asp Ile Arg His Gly Ile Val Thr Ala Pro Ile
                325                 330                 335

Leu Phe Ala Met Glu Glu Phe Pro Gln Leu Arg Ala Val Ile Asp Gln
            340                 345                 350

Gly Phe Glu Asn Pro Ser Asn Val Asp Val Ala Leu Glu Tyr Leu Gly
        355                 360                 365

Lys Ser Arg Gly Ile Gln Arg Thr Arg Glu Leu Ala Thr Asn His Ala
    370                 375                 380
```

Asn Leu Ala Ala Ala Ala Ile Asp Ala Leu Pro Lys Thr Asp Asn Glu
385 390 395 400

Glu Val Arg Lys Ser Arg Arg Ala Leu Leu Asp Leu Thr Gln Arg Val
405 410 415

Ile Thr Arg Asn Lys
420

<210> 124
<211> 1266
<212> DNA
<213> Mangifera indica

<400> 124
atgttattct cttatggttt atctcgtatt tctattaacc ctcgtgcctc tttattgact 60

tgtagatggt tattatccca tttgactggt tctttatctc cttccacttc ttcccatact 120

atttctgact ccgtccataa agtctggggt tgcagagaag cctatacttg gtctgtccca 180

gctttacatg gttttagaca tcaaatccac catcaatcct cttccttgat tgaagatcaa 240

ttagacccat tctccttggt cgccgatgag ttgtccttgg ttgctaaccg tttaagatct 300

atggttgtca ctgaagtccc taaattagcc tctgccgccg aatacttttt caagatgggt 360

gtcgaaggta agcgtttcag accagctgtc ttgttgttaa tggccactgc cttaaacgtt 420

catgttttgg aacctttgcc tgaaggtgct ggtgacgctt taatgaccga gttgagaacc 480

cgtcaacaat gcattgctga aatcactgag atgatccacg tcgcctcttt attgcatgac 540

gatgttttag acgacgctga tactagaaga ggtattggtt ctttgaactt ggttatgggt 600

aacaaattgg ccgttttggc cggtgatttc ttgttatccc gtgcttgcgt tgctttagct 660

tctttgaaga acactgaagt tgtttctttg ttggccaccg tcgttgaaca cttagttact 720

ggtgagacta tgcaaatgac cacctcttct gaccaaagat gttccatgga atattacatg 780

caaaaaactt attacaaaac cgcctccttg atttctaact cctgtaaagc catcgcctta 840

ttagctggtc aatctgctga agttgccatg ttagccttcg agtttggtaa gaacttgggt 900

ttagcttacc aattgatcga tgatgtcttg gattttaccg gtacctctgc ttctttgggt 960

aagggttcct tgtccgacat tagacacggt attgttaccg ccccaatctt attcgctatg 1020

gaagagtttc cacaattgag agctgttatc gaccaaggtt tcgagaaccc atctaacgtt 1080

gacgtcgcct tagagtattt aggtaaatct agaggtatcc aacgtacccg tgaattagct 1140

actaaccatg ctaacttagc cgccgccgcc atcgatgcct tgcctaaaac cgataatgaa 1200

gaagtccgta agtccagacg tgctttatta gatttgactc aaagagtcat caccagaaac 1260

aaatag 1266

<210> 125
<211> 328
<212> PRT
<213> Mangifera indica

<400> 125

Met Pro Phe Val Val Pro Arg Arg Asn Arg Ser Leu Ser Val Ser Ala
1 5 10 15

Val Leu Thr Lys Glu Glu Thr Leu Arg Glu Glu Glu Glu Asp Pro Lys
20 25 30

Pro Val Phe Asp Phe Lys Ser Tyr Met Leu Gln Lys Gly Asn Ser Val
35 40 45

Asn Gln Ala Leu Asp Ala Val Val Ser Ile Arg Glu Pro Lys Lys Ile
50 55 60

His Glu Ala Met Arg Tyr Ser Leu Leu Ala Gly Gly Lys Arg Val Arg
65 70 75 80

Pro Val Leu Cys Ile Ala Ala Cys Glu Leu Val Gly Gly Asn Glu Ser
85 90 95

Met Ala Met Pro Ala Ala Cys Ala Val Glu Met Ile His Thr Met Ser
100 105 110

Leu Ile His Asp Asp Leu Pro Cys Met Asp Asn Asp Asp Leu Arg Arg
115 120 125

Gly Lys Pro Thr Asn His Lys Val Phe Gly Glu Asp Val Ala Val Leu
130 135 140

```
Ala Gly Asp Ala Leu Leu Ala Phe Ser Phe Glu Asn Met Ala Val Ser
145                 150                 155                 160

Thr Val Gly Val Leu Pro Ser Arg Val Val Lys Ala Val Gly Glu Leu
                165                 170                 175

Ala Lys Ser Ile Gly Ile Glu Gly Leu Val Ala Gly Gln Val Val Asp
            180                 185                 190

Ile Asn Ser Glu Gly Leu Lys Glu Val Gly Leu Asp His Leu Glu Phe
        195                 200                 205

Ile His Gln His Lys Thr Ala Ala Leu Leu Glu Gly Ser Val Val Leu
    210                 215                 220

Gly Ala Ile Leu Gly Gly Gly Ser Asp Asp Glu Val Glu Lys Leu Arg
225                 230                 235                 240

Thr Phe Ala Arg Cys Ile Gly Leu Leu Phe Gln Val Val Asp Asp Ile
                245                 250                 255

Leu Asp Val Thr Lys Ser Ser Arg Glu Leu Gly Lys Thr Ala Gly Lys
            260                 265                 270

Asp Leu Val Ala Asp Lys Val Thr Tyr Pro Lys Leu Leu Gly Ile Glu
        275                 280                 285

Lys Ser Arg Glu Leu Ala Asp Lys Leu Asn Lys Asp Ala Gln Gln Gln
    290                 295                 300

Leu Ser Gly Phe Asp Gln Glu Lys Ala Ala Pro Leu Ile Ala Leu Ser
305                 310                 315                 320

Asn Tyr Ile Ala Tyr Arg Gln Asn
                325
```

<210> 126
<211> 987

<212> DNA
<213> Mangifera indica

<400> 126
atgccattcg ttgttcctag aagaaaccgt tctttgtccg tttccgccgt tttgaccaag 60

gaagaaactt taagagagga agaagaagat ccaaagccag ttttcgactt caaatcttac 120

atgttacaaa agggtaattc tgttaatcaa gctttggatg ctgtcgtttc cattagagaa 180

cctaagaaaa tccatgaggc tatgcgttac tctttgttgg ctggtggtaa gagagttcgt 240

cctgttttgt gtattgccgc ctgtgaattg gtcggtggta acgaatctat ggctatgcca 300

gccgcctgtg ctgtcgaaat gatccacact atgtccttga ttcacgatga tttgccatgt 360

atggataatg acgatttgcg tcgtggtaaa cctaccaacc ataaagtttt cggtgaagac 420

gtcgccgttt tggctggtga cgctttatta gctttttcct tcgagaacat ggccgtttcc 480

actgttggtg tcttaccatc cagagttgtc aaggctgttg gtgaattggc caagtctatc 540

ggtattgaag gtttggttgc cggtcaagtc gtcgatatta attctgaggg tttaaaagag 600

gtcggtttag atcacttaga atttatccat caacacaaaa ccgctgcttt gttggagggt 660

tctgttgttt tgggtgctat tttaggtggt ggttctgatg atgaagtcga aaagttgcgt 720

acctttgcta gatgtatcgg tttgttgttt caagttgttg acgatatttt ggatgtcact 780

aagtcttcca gagaattggg taagactgcc ggtaaagatt tggttgctga taaagttact 840

tatccaaagt tgttaggtat tgaaaagtct cgtgaattgg ccgataagtt aaacaaggat 900

gctcaacaac aattatccgg ttttgatcaa gagaaggctg ccccttaat cgctttgtcc 960

aattacatcg cctacagaca aaactag 987

<210> 127
<211> 321
<212> PRT
<213> Artificial Sequence

<220>
<223> Truncated geranylgeranyl pyrophosphate synthase Cs2_GPPS_NTrunc

<400> 127

Met Val Ile Ala Glu Val Pro Lys Leu Ala Ser Ala Ala Glu Tyr Phe
1 5 10 15

```
Phe Lys Met Gly Val Glu Gly Lys Arg Phe Arg Pro Thr Val Leu Leu
            20                  25                  30

Leu Met Ala Thr Ala Leu Asn Val Arg Val Pro Glu Pro Leu His Asp
        35                  40                  45

Gly Val Glu Asp Ala Ser Ala Thr Glu Leu Arg Thr Arg Gln Gln Cys
    50                  55                  60

Ile Ala Glu Ile Thr Glu Met Ile His Val Ala Ser Leu Leu His Asp
65                  70                  75                  80

Asp Val Leu Asp Asp Ala Asp Thr Arg Arg Gly Ile Gly Ser Leu Asn
                85                  90                  95

Phe Val Met Gly Asn Lys Leu Ala Val Leu Ala Gly Asp Phe Leu Leu
            100                 105                 110

Ser Arg Ala Cys Val Ala Leu Ala Ser Leu Lys Asn Thr Glu Val Val
        115                 120                 125

Thr Leu Leu Ala Thr Val Val Glu His Leu Val Thr Gly Glu Thr Met
    130                 135                 140

Gln Met Thr Thr Ser Ser Asp Gln Arg Cys Ser Met Asp Tyr Tyr Met
145                 150                 155                 160

Gln Lys Thr Tyr Tyr Lys Thr Ala Ser Leu Ile Ser Asn Ser Cys Lys
                165                 170                 175

Ala Ile Ala Leu Leu Ala Gly Gln Thr Ala Glu Val Ala Ile Leu Ala
            180                 185                 190

Phe Asp Tyr Gly Lys Asn Leu Gly Leu Ala Tyr Gln Leu Ile Asp Asp
        195                 200                 205

Val Leu Asp Phe Thr Gly Thr Ser Ala Ser Leu Gly Lys Gly Ser Leu
    210                 215                 220
```

```
Ser Asp Ile Arg His Gly Ile Ile Thr Ala Pro Ile Leu Phe Ala Met
225                 230                 235                 240


Glu Glu Phe Pro Gln Leu Arg Thr Val Val Glu Gln Gly Phe Glu Asp
                245                 250                 255


Ser Ser Asn Val Asp Ile Ala Leu Glu Tyr Leu Gly Lys Ser Arg Gly
            260                 265                 270


Ile Gln Lys Thr Arg Glu Leu Ala Val Lys His Ala Asn Leu Ala Ala
        275                 280                 285


Ala Ala Ile Asp Ser Leu Pro Glu Asn Asn Asp Glu Asp Val Thr Lys
    290                 295                 300


Ser Arg Arg Ala Leu Leu Asp Leu Thr His Arg Val Ile Thr Arg Asn
305                 310                 315                 320


Lys
```

<210> 128
<211> 966
<212> DNA
<213> Artificial Sequence

<220>
<223> Truncated geranylgeranyl pyrophosphate synthase Cs2_GPPS_NTrunc

<400> 128

```
atggtcattg ctgaagttcc taaattagcc tctgccgccg aatacttctt caagatgggt       60

gtcgagggta agagatttcg tcctaccgtt ttgttgttaa tggccaccgc cttaaacgtc      120

agagtccctg aaccattaca tgatggtgtt gaagatgcct ctgccaccga gttgagaact      180

agacaacaat gtattgctga aatcaccgag atgattcacg ttgcctcttt gttgcacgat      240

gatgttttgg atgatgctga tacccgtcgt ggtatcggtt ctttgaactt tgtcatgggt      300

aacaagttgg ctgtcttggc tggtgatttc ttattgtctc gtgcctgcgt tgccttagcc      360

tctttaaaaa ataccgaagt tgttacttta ttggccactg ttgttgagca cttggttact      420
```

ggtgaaacta tgcaaatgac cacctcttcc gaccaacgtt gttccatgga ctattacatg 480

caaaagacct actacaagac cgcttctttg atttccaatt cttgtaaagc cattgcctta 540

ttagctggtc aaactgctga agttgccatc ttggccttcg actacggtaa aaacttgggt 600

ttagcttacc aattgattga tgacgtttta gattttactg gtacttctgc ttctttgggt 660

aaaggttctt tatccgatat tcgtcatggt atcattaccg ctccaatctt attcgctatg 720

gaagaatttc ctcaattgcg tactgtcgtt gaacaaggtt tcgaagactc ctccaacgtt 780

gacattgcct tagaatactt gggtaagtct cgtggtattc aaaagacccg tgaattagcc 840

gttaaacatg ccaacttagc cgccgccgcc atcgattcct tgcctgaaaa caacgatgag 900

gatgtcacca agtcccgtcg tgctttgtta gatttaactc acagagttat tacccgtaac 960

aagtaa 966

<210> 129
<211> 416
<212> PRT
<213> Quercus sp.

<400> 129

Met Leu Phe Ser Arg Ile Ser Arg Ile Arg Arg Pro Gly Ser Asn Gly
1 5 10 15

Phe Arg Trp Phe Leu Ser His Lys Thr His Leu Gln Phe Leu Asn Pro
20 25 30

Pro Ala Tyr Ser Tyr Ser Ser Thr His Lys Val Leu Gly Cys Arg Glu
35 40 45

Ile Phe Ser Trp Gly Leu Pro Ala Leu His Gly Phe Arg His Asn Ile
50 55 60

His His Gln Ser Ser Ser Ile Val Glu Glu Gln Asn Asp Pro Phe Ser
65 70 75 80

Leu Val Ala Asp Glu Leu Ser Met Val Ala Asn Arg Leu Arg Ser Met
85 90 95

```
Val Val Thr Glu Val Pro Lys Leu Ala Ser Ala Ala Glu Tyr Phe Phe
            100                 105                 110

Lys Met Gly Val Glu Gly Lys Arg Phe Arg Pro Thr Val Leu Leu Leu
        115                 120                 125

Met Ala Thr Ala Met Asn Ile Ser Ile Leu Glu Pro Ser Leu Arg Gly
    130                 135                 140

Pro Gly Asp Ala Leu Thr Thr Glu Leu Arg Ala Arg Gln Gln Arg Ile
145                 150                 155                 160

Ala Glu Ile Thr Glu Met Ile His Val Ala Ser Leu Leu His Asp Asp
                165                 170                 175

Val Leu Asp Asp Ala Asp Thr Arg Arg Gly Ile Gly Ser Leu Asn Phe
            180                 185                 190

Val Met Gly Asn Lys Leu Ala Val Leu Ala Gly Asp Phe Leu Leu Ser
        195                 200                 205

Arg Ala Cys Val Ala Leu Ala Ser Leu Lys Asn Thr Glu Val Val Ser
    210                 215                 220

Leu Leu Ala Lys Val Val Glu His Leu Val Thr Gly Glu Thr Met Gln
225                 230                 235                 240

Met Thr Thr Thr Cys Glu Gln Arg Cys Ser Met Glu Tyr Tyr Met Gln
                245                 250                 255

Lys Thr Tyr Tyr Lys Thr Ala Ser Leu Ile Ser Asn Ser Cys Lys Ala
            260                 265                 270

Ile Ala Leu Leu Gly Gly Gln Thr Ser Glu Val Ala Met Leu Ala Tyr
        275                 280                 285

Glu Tyr Gly Lys Asn Leu Gly Leu Ala Tyr Gln Leu Ile Asp Asp Val
    290                 295                 300
```

Leu Asp Phe Thr Gly Thr Ser Ala Ser Leu Gly Lys Gly Ser Leu Ser
305 310 315 320

Asp Ile Arg His Gly Ile Ile Thr Ala Pro Ile Leu Phe Ala Met Glu
325 330 335

Glu Phe Pro Gln Leu Arg Glu Val Val Asp Arg Gly Phe Asp Asp Pro
340 345 350

Ala Asn Val Asp Val Ala Leu Asp Tyr Leu Gly Lys Ser Arg Gly Ile
355 360 365

Gln Arg Ala Arg Glu Leu Ala Lys Lys His Ala Asn Ile Ala Ala Glu
370 375 380

Ala Ile Asp Ser Leu Pro Glu Ser Asn Asp Glu Asp Val Arg Lys Ser
385 390 395 400

Arg Arg Ala Leu Leu Asp Leu Thr Glu Arg Val Ile Thr Arg Thr Lys
405 410 415

<210> 130
<211> 1251
<212> DNA
<213> Quercus sp.

<400> 130
atgttgttct ctcgtatttc tcgtatccgt agaccaggtt ctaatggttt cagatggttc 60

ttgtcccata agactcattt acaattcttg aaccctccag cttattccta ctcttccact 120

cataaggtct tgggttgtag agaaattttt tcctggggtt tacctgcctt acatggtttc 180

agacacaaca ttcaccacca atcttcctct attgttgaag aacaaaatga ccctttctct 240

ttggtcgctg atgagttgtc catggttgct aacagattgc gttctatggt tgttactgaa 300

gttcctaaat tagcctccgc cgctgaatac ttttttaaaa tgggtgttga aggtaagaga 360

ttcagaccaa ctgttttatt gttgatggct accgccatga acatttccat cttagaacca 420

tctttgagag gtccaggtga cgctttgacc actgaattga gagccagaca acaaagaatt 480

gctgaaatta ccgagatgat ccacgttgct tccttgttgc acgatgacgt tttggatgac 540

gctgatacta gaagaggtat tggttcctta aactttgtca tgggtaataa attagctgtt 600

ttggctggtg atttttttgtt atctcgtgcc tgtgttgctt tagcttcttt gaagaacacc 660

gaagttgtct ccttgttagc caaggtcgtc gaacacttgg ttactggtga aactatgcaa 720

atgaccacta cttgtgaaca aagatgttcc atggaatact acatgcaaaa gacttactat 780

aagaccgctt ctttaatttc caactcctgt aaagccattg ctttattagg tggtcaaact 840

tctgaggtcg ctatgttagc ctacgaatat ggtaaaaact tgggtttagc ttaccaattg 900

attgatgatg tcttggattt cactggtact tctgcttcct tgggtaaggg ttccttgtct 960

gatattagac atggtatcat tactgctcca attttgtttg ctatggaaga attcccacaa 1020

ttacgtgaag ttgtcgatag aggtttcgac gatcctgcca acgtcgatgt tgccttggac 1080

tacttgggta agtctagagg tatccaaaga gccagagagt tagctaaaaa acacgctaac 1140

attgctgccg aagccatcga ctctttgcca gaatccaacg acgaggacgt cagaaagtcc 1200

cgtcgtgctt tgttggactt gaccgaaaga gtcattactc gtactaagta a 1251

<210> 131
<211> 427
<212> PRT
<213> Artificial Sequence

<220>
<223> Truncated geranylgeranyl pyrophosphate synthase Pa_GPPS_Ntrunc

<400> 131

Met Tyr Thr Arg Cys Ile Leu Arg Asp Lys Tyr Ser Arg Phe Asn Leu
1 5 10 15

Arg Arg Lys Phe Phe Thr Ser Ala Lys Ser Ile Asn Ala Leu Asn Gly
20 25 30

Leu Pro Asp Ser Gly Asn Pro Arg Gly Glu Ser Asn Gly Ile Ser Gln
35 40 45

Phe Glu Ile Gln Gln Val Phe Arg Cys Lys Glu Tyr Ile Trp Ile Asp
50 55 60

Arg His Lys Phe His Asp Val Gly Phe Gln Ala His His Lys Gly Ser

65 70 75 80

Ile Thr Asp Glu Glu Gln Val Asp Pro Phe Ser Leu Val Ala Asp Glu
85 90 95

Leu Ser Ile Leu Ala Asn Arg Leu Arg Ser Met Ile Leu Thr Glu Ile
100 105 110

Pro Lys Leu Gly Thr Ala Ala Glu Tyr Phe Phe Lys Leu Gly Val Glu
115 120 125

Gly Lys Arg Phe Arg Pro Met Val Leu Leu Leu Met Ala Ser Ser Leu
130 135 140

Thr Ile Gly Ile Pro Glu Val Ala Ala Asp Cys Leu Arg Lys Gly Leu
145 150 155 160

Asp Glu Glu Gln Arg Leu Arg Gln Gln Arg Ile Ala Glu Ile Thr Glu
165 170 175

Met Ile His Val Ala Ser Leu Leu His Asp Asp Val Leu Asp Asp Ala
180 185 190

Asp Thr Arg Arg Gly Val Gly Ser Leu Asn Phe Val Met Gly Asn Lys
195 200 205

Leu Ala Val Leu Ala Gly Asp Phe Leu Leu Ser Arg Ala Ser Val Ala
210 215 220

Leu Ala Ser Leu Lys Asn Thr Glu Val Val Glu Leu Leu Ser Lys Val
225 230 235 240

Leu Glu His Leu Val Thr Gly Glu Ile Met Gln Met Thr Asn Thr Asn
245 250 255

Glu Gln Arg Cys Ser Met Glu Tyr Tyr Met Gln Lys Thr Phe Tyr Lys
260 265 270

Thr Ala Ser Leu Met Ala Asn Ser Cys Lys Ala Ile Ala Leu Ile Ala

275 280 285

Gly Gln Pro Ala Glu Val Cys Met Leu Ala Tyr Asp Tyr Gly Arg Asn
290 295 300

Leu Gly Leu Ala Tyr Gln Leu Leu Asp Asp Val Leu Asp Phe Thr Gly
305 310 315 320

Thr Thr Ala Ser Leu Gly Lys Gly Ser Leu Ser Asp Ile Arg Gln Gly
325 330 335

Ile Val Thr Ala Pro Ile Leu Phe Ala Leu Glu Glu Phe Pro Gln Leu
340 345 350

His Asp Val Ile Asn Arg Lys Phe Lys Lys Pro Gly Asp Ile Asp Leu
355 360 365

Ala Leu Glu Phe Leu Gly Lys Ser Asp Gly Ile Arg Lys Ala Lys Gln
370 375 380

Leu Ala Ala Gln His Ala Gly Leu Ala Ala Phe Ser Val Glu Ser Phe
385 390 395 400

Pro Pro Ser Glu Ser Glu Tyr Val Lys Leu Cys Arg Lys Ala Leu Ile
405 410 415

Asp Leu Ser Glu Lys Val Ile Thr Arg Thr Arg
420 425

<210> 132
<211> 1284
<212> DNA
<213> Artificial Sequence

<220>
<223> Truncated geranylgeranyl pyrophosphate synthase Pa_GPPS_N_trunc

<400> 132
atgtataccc gttgcatttt aagagacaag tattctcgtt tcaacttgag acgtaaattc 60

ttcacttccg ctaaatccat caatgccttg aatggtttac ctgactctgg taaccctaga 120

ggtgaatcta acggtatctc ccaattcgaa attcaacaag ttttccgttg taaagaatac 180

atttggatcg atcgtcacaa gttccacgat gttggttttc aagctcatca caagggttcc 240

atcactgacg aggaacaagt tgacccttt tctttagtcg ctgatgaatt gtccatctta 300

gctaatcgtt taagatccat gatcttaacc gagattccaa agttaggtac cgctgccgaa 360

tacttttca agttgggtgt cgaaggtaag agatttagac caatggtttt gttgttgatg 420

gcctcctctt taactattgg tatccctgaa gttgccgctg attgtttgcg taagggtttg 480

gacgaagaac aaagattacg tcaacaacgt atcgctgaaa ttactgaaat gattcatgtc 540

gcctctttgt tgcacgatga tgttttggat gacgccgata ctagacgtgg tgttggttcc 600

ttgaactttg ttatgggtaa caagttggct gttttagccg gtgatttctt gttatctaga 660

gcttctgttg ccttagcttc tttaaagaac actgaggttg ttgagttatt gtctaaggtt 720

ttggagcact tagtcactgg tgagatcatg caaatgacta acactaatga acaaagatgt 780

tctatggaat attacatgca aaagactttc tacaagaccg cctctttgat ggctaattct 840

tgtaaagcca ttgccttgat cgctggtcaa cctgccgaag tctgcatgtt ggcctacgac 900

tacggtagaa acttgggttt agcttatcaa ttattggatg acgttttgga tttcactggt 960

accactgctt ctttaggtaa gggttcctta tccgacatca gacaaggtat tgttactgcc 1020

cctattttat tcgctttgga agaattccct caattacacg acgtcatcaa ccgtaagttc 1080

aaaaaaccag gtgacatcga tttggccttg gaatttttgg gtaagtctga tggtatccgt 1140

aaagccaaac aattggctgc tcaacatgct ggtttagctg ccttttctgt cgaatccttt 1200

ccaccatctg aatccgaata cgttaagtta tgtagaaagg ccttgatcga tttgtctgaa 1260

aaggtcatta ctcgtaccag ataa 1284

<210> 133
<211> 381
<212> PRT
<213> Abies grandis

<400> 133

Met Ala Tyr Ser Ala Met Ala Thr Met Gly Tyr Asn Gly Met Ala Ala
1 5 10 15

```
Ser Cys His Thr Leu His Pro Thr Ser Pro Leu Lys Pro Phe His Gly
            20                  25                  30

Ala Ser Thr Ser Leu Glu Ala Phe Asn Gly Glu His Met Gly Leu Leu
        35                  40                  45

Arg Gly Tyr Ser Lys Arg Lys Leu Ser Ser Tyr Lys Asn Pro Ala Ser
    50                  55                  60

Arg Ser Ser Asn Ala Thr Val Ala Gln Leu Leu Asn Pro Pro Gln Lys
65                  70                  75                  80

Gly Lys Lys Ala Val Glu Phe Asp Phe Asn Lys Tyr Met Asp Ser Lys
                85                  90                  95

Ala Met Thr Val Asn Glu Ala Leu Asn Lys Ala Ile Pro Leu Arg Tyr
            100                 105                 110

Pro Gln Lys Ile Tyr Glu Ser Met Arg Tyr Ser Leu Leu Ala Gly Gly
        115                 120                 125

Lys Arg Val Arg Pro Val Leu Cys Ile Ala Ala Cys Glu Leu Val Gly
    130                 135                 140

Gly Thr Glu Glu Leu Ala Ile Pro Thr Ala Cys Ala Ile Glu Met Ile
145                 150                 155                 160

His Thr Met Ser Leu Met His Asp Asp Leu Pro Cys Ile Asp Asn Asp
                165                 170                 175

Asp Leu Arg Arg Gly Lys Pro Thr Asn His Lys Ile Phe Gly Glu Asp
            180                 185                 190

Thr Ala Val Thr Ala Gly Asn Ala Leu His Ser Tyr Ala Phe Glu His
        195                 200                 205

Ile Ala Val Ser Thr Ser Lys Thr Val Gly Ala Asp Arg Ile Leu Arg
    210                 215                 220
```

```
Met Val Ser Glu Leu Gly Arg Ala Thr Gly Ser Glu Gly Val Met Gly
225                 230                 235                 240


Gly Gln Met Val Asp Ile Ala Ser Glu Gly Asp Pro Ser Ile Asp Leu
                245                 250                 255


Gln Thr Leu Glu Trp Ile His Ile His Lys Thr Ala Met Leu Leu Glu
            260                 265                 270


Cys Ser Val Val Cys Gly Ala Ile Ile Gly Gly Ala Ser Glu Ile Val
        275                 280                 285


Ile Glu Arg Ala Arg Arg Tyr Ala Arg Cys Val Gly Leu Leu Phe Gln
    290                 295                 300


Val Val Asp Asp Ile Leu Asp Val Thr Lys Ser Ser Asp Glu Leu Gly
305                 310                 315                 320


Lys Thr Ala Gly Lys Asp Leu Ile Ser Asp Lys Ala Thr Tyr Pro Lys
                325                 330                 335


Leu Met Gly Leu Glu Lys Ala Lys Glu Phe Ser Asp Glu Leu Leu Asn
            340                 345                 350


Arg Ala Lys Gly Glu Leu Ser Cys Phe Asp Pro Val Lys Ala Ala Pro
        355                 360                 365


Leu Leu Gly Leu Ala Asp Tyr Val Ala Phe Arg Gln Asn
    370                 375                 380


<210>  134
<211>  1146
<212>  DNA
<213>  Abies grandis

<400>  134
atggcttatt ctgctatggc tactatgggt tacaacggta tggctgcttc ttgtcacact       60

ttacacccaa cttctccatt gaaacctttt cacggtgctt ctacttcctt ggaagccttc      120

aatggtgaac acatgggttt gttaagaggt tattctaagc gtaagttgtc ctcttacaaa      180
```

aatccagctt ctcgttcctc caatgctacc gtcgctcaat tattgaaccc accacaaaag 240

ggtaagaagg ctgttgaatt tgacttcaat aagtatatgg attctaaggc tatgaccgtc 300

aacgaggctt tgaataaagc catcccattg cgttacccac aaaagatcta cgaatctatg 360

agatattctt tgttagctgg tggtaagaga gtccgtccag ttttgtgtat cgccgcttgt 420

gaattagtcg gtggtactga ggagttagct attccaaccg cctgtgccat cgaaatgatc 480

cacaccatgt ctttgatgca cgatgatttg ccatgtatcg acaacgatga cttgagacgt 540

ggtaaaccta ccaatcataa gattttcggt gaagatactg ctgttactgc cggtaacgct 600

ttacactctt acgccttcga acatattgct gtttctactt ccaagactgt tggtgctgat 660

agaattttga gaatggtttc tgaattaggt cgtgctactg gttccgaagg tgttatgggt 720

ggtcaaatgg tcgatattgc ttctgaaggt gacccttcca ttgatttgca aactttagaa 780

tggatccaca tccacaagac tgctatgtta ttagaatgtt ctgttgtctg tggtgccatc 840

atcggtggtg cttctgaaat tgttattgag agagccagac gttatgctcg ttgtgtcggt 900

ttattgtttc aagttgttga cgacatttta gatgttacca aatcttctga cgaattgggt 960

aaaactgctg gtaaagattt aatctccgat aaagccacct accctaagtt gatgggtttg 1020

gagaaggcca aagagttttc cgatgaatta ttaaacagag ctaaaggtga attgtcttgc 1080

ttcgatccag ttaaggctgc cccattgtta ggtttggctg actacgttgc cttcagacaa 1140

aactaa 1146

<210> 135
<211> 257
<212> PRT
<213> Artificial Sequence

<220>
<223> Truncated geranylgeranyl pyrophosphate synthase Pb_GPPS_NTrunc

<400> 135

Met Ala Ala Ile Phe Pro Ser Ile Pro Ser Asn Phe Lys Pro Pro Gln
1 5 10 15

Ile Ser Gln Thr Leu Thr Arg Arg Arg Arg Pro Asn Arg Thr Leu Cys
20 25 30

```
Thr Ala Thr Ser Asp Gln Ser Tyr Leu Ser Ala Ser Ser Ala Asp Ile
        35                  40                  45

Tyr Ser His Leu Leu Arg Ser Leu Pro Ala Thr Ile His Pro Ser Val
    50                  55                  60

Lys Ala Pro Ile His Ser Leu Leu Ser Ser Pro Ile Pro Pro Thr Ile
65                  70                  75                  80

Ala Pro Pro Leu Cys Leu Ala Ala Thr Glu Leu Val Gly Gly Asn Pro
                85                  90                  95

Asn Ser Ala Ile Asn Ala Ala Cys Ala Ile His Leu Ile His Ala Val
            100                 105                 110

Thr His Thr Arg Thr Ala Pro Pro Leu Ala Glu Phe Ser Pro Gly Val
        115                 120                 125

Leu Leu Met Thr Gly Asp Gly Leu Leu Val Leu Ala Tyr Glu Met Leu
    130                 135                 140

Ala Arg Ser Pro Ala Val Asp Ala Asp Thr Ser Val Arg Val Leu Lys
145                 150                 155                 160

Glu Val Ala Arg Thr Ala Ala Ala Val Ala Ala Ala Tyr Glu Gly Gly
                165                 170                 175

Arg Glu Gly Glu Leu Ala Ala Gly Ala Ala Ala Cys Gly Val Ile Leu
            180                 185                 190

Gly Gly Gly Asn Glu Glu Glu Val Glu Arg Gly Arg Arg Val Gly Met
        195                 200                 205

Phe Ala Gly Lys Met Glu Leu Val Glu Ala Glu Val Glu Leu Arg Leu
    210                 215                 220

Gly Phe Glu Asp Ala Lys Ala Gly Ala Val Arg Arg Leu Leu Glu Glu
225                 230                 235                 240
```

Met Arg Phe Thr Gln Ser Phe Val Asn Val Arg Asn Pro Phe Tyr Gly
245 250 255

Lys

<210> 136
<211> 774
<212> DNA
<213> Artificial Sequence

<220>
<223> Truncated geranylgeranyl pyrophosphate synthase Pb_GPPS_NTrunc

<400> 136
atggctgcta tctttccatc cattccatcc aacttcaaac cacctcaaat ctctcaaact 60

ttgaccagac gtagaagacc aaaccgtact ttatgtactg ccacctctga ccaatcttac 120

ttgtccgctt cttctgccga catttattct catttgttaa gatctttacc agctactatt 180

catccatctg ttaaagcccc aatccattct ttattgtcct ctccaattcc tccaaccatc 240

gctccacctt tgtgtttagc tgctaccgaa ttggttggtg gtaaccctaa ctctgccatt 300

aacgccgcct gtgccattca tttgattcat gctgttactc atactagaac cgctccacca 360

ttagctgaat tttctcctgg tgttttgttg atgactggtg atggtttatt agttttggct 420

tacgagatgt tggccagatc cccagctgtt gatgccgata cttctgtccg tgttttgaag 480

gaagtcgcta gaaccgccgc cgccgtcgcc gctgcttatg aaggtggtag agaaggtgaa 540

ttagctgccg gtgccgctgc ttgtggtgtc attttgggtg gtggtaacga agaagaggtc 600

gaaagaggtc gtagagtcgg tatgttcgct ggtaaaatgg aattagttga agctgaagtc 660

gaattgagat tgggtttcga agatgctaaa gccggtgccg ttagaagatt gttggaagaa 720

atgcgtttca cccaatcttt tgtcaacgtt agaaaccctt tttatggtaa gtaa 774

<210> 137
<211> 420
<212> PRT
<213> Azadirachta indica

<400> 137

```
Met Leu Phe Ser Arg Gly Leu Ser Arg Ile Ser Arg Ile Pro Arg Asn
1               5                  10                  15

Ser Leu Ile Gly Cys Arg Trp Leu Val Ser Tyr Arg Pro Asp Thr Ile
            20                  25                  30

Leu Ser Gly Ser Ser His Ser Val Gly Asp Ser Thr Gln Lys Val Leu
        35                  40                  45

Gly Cys Arg Glu Ala Tyr Leu Trp Ser Leu Pro Ala Leu His Gly Ile
    50                  55                  60

Arg His Gln Ile His Gln Gln Ser Ser Ser Leu Ile Glu Glu Glu Leu
65                  70                  75                  80

Asp Pro Phe Ser Leu Val Ala Asp Glu Leu Ser Leu Val Ala Asn Arg
                85                  90                  95

Leu Arg Ser Met Val Val Ala Glu Val Pro Lys Leu Ala Ser Ala Ala
            100                 105                 110

Glu Tyr Phe Phe Lys Met Gly Val Glu Gly Lys Arg Phe Arg Pro Thr
        115                 120                 125

Val Leu Leu Leu Met Ala Ser Ala Leu Asn Val Gln Val Pro Gln Pro
    130                 135                 140

Leu Ser Asp Gly Val Gly Asp Ala Leu Thr Thr Glu Leu Arg Thr Arg
145                 150                 155                 160

Gln Gln Cys Ile Ala Glu Ile Thr Glu Met Ile His Val Ala Ser Leu
                165                 170                 175

Leu His Asp Asp Val Leu Asp Asp Ala Asp Thr Arg Arg Gly Ile Gly
            180                 185                 190

Ser Leu Asn Phe Val Met Gly Asn Lys Leu Ala Val Leu Ala Gly Asp
        195                 200                 205
```

```
Phe Leu Leu Ser Arg Ala Cys Val Ala Leu Ala Ser Leu Lys Asn Thr
    210                 215                 220

Glu Val Val Ser Leu Leu Ala Thr Val Val Glu His Leu Val Thr Gly
225                 230                 235                 240

Glu Thr Met Gln Met Thr Thr Thr Ala Glu Gln Arg Arg Ser Met Asp
                245                 250                 255

Tyr Tyr Met Gln Lys Thr Tyr Tyr Lys Thr Ala Ser Leu Ile Ser Asn
            260                 265                 270

Ser Cys Lys Ala Ile Ala Leu Leu Ala Gly Gln Thr Thr Glu Val Ala
        275                 280                 285

Met Leu Ala Phe Asp Tyr Gly Lys Asn Leu Gly Leu Ala Phe Gln Leu
    290                 295                 300

Ile Asp Asp Val Leu Asp Phe Thr Gly Thr Ser Ala Ser Leu Gly Lys
305                 310                 315                 320

Gly Ser Leu Ser Asp Ile Arg His Gly Ile Val Thr Ala Pro Ile Leu
                325                 330                 335

Phe Ala Met Glu Glu Phe Pro Glu Leu Arg Lys Val Val Asp Lys Gly
            340                 345                 350

Phe Asp Asp Pro Ser Asn Val Asp Ile Ala Leu Glu Tyr Leu Gly Lys
        355                 360                 365

Ser Arg Gly Ile Gln Arg Thr Arg Glu Leu Ala Gln Lys His Ala Asn
    370                 375                 380

Leu Ala Thr Val Ala Leu Asp Ser Leu Pro Glu Ser Asn Asp Asp Asp
385                 390                 395                 400

Val Lys Lys Ser Arg Arg Ala Leu Leu Asp Leu Ala Gln Arg Val Ile
                405                 410                 415
```

Thr Arg Asn Lys
420

<210> 138
<211> 1263
<212> DNA
<213> Azadirachta indica

<400> 138
atgttgtttt ccagaggttt atctcgtatt tccagaatcc cacgtaactc tttgatcggt 60

tgtagatggt tagtttctta ccgtcctgat accattttat ctggttcctc tcactccgtt 120

ggtgactcta ctcaaaaggt tttaggttgt cgtgaagctt acttgtggtc tttaccagcc 180

ttgcacggta ttagacacca aattcatcaa caatcctctt ctttgattga agaagaattg 240

gatccattct ctttagttgc tgatgaattg tctttagtcg ctaaccgttt gagatccatg 300

gtcgtcgctg aagtcccaaa attagcctcc gccgccgagt acttcttcaa gatgggtgtt 360

gagggtaaga gattccgtcc aactgtctta ttgttgatgg cctccgcctt aaacgttcaa 420

gtcccacaac ctttgtctga cggtgttggt gatgctttga ctaccgagtt gagaactaga 480

caacaatgca ttgctgagat tactgaaatg atccatgttg cttctttgtt gcatgacgac 540

gttttggatg atgctgacac tagacgtggt atcggttctt tgaacttcgt tatgggtaac 600

aagttggctg tcttggctgg tgatttcttg ttgtccagag cctgtgttgc tttagcttcc 660

ttgaagaata ctgaggttgt ctctttgttg gccaccgttg ttgaacactt ggtcaccggt 720

gaaactatgc aaatgactac tactgctgaa caaagacgtt ccatggatta ttacatgcaa 780

aagacttact ataagaccgc ctctttgatt tccaactctt gtaaagccat tgccttgtta 840

gctggtcaaa ctaccgaagt tgctatgttg gctttcgatt acggtaagaa tttgggttta 900

gcttttcaat tgatcgatga cgtcttggat tttactggta cctctgcttc tttaggtaaa 960

ggttccttgt ctgatattag acacggtatc gttaccgctc caattttatt cgctatggaa 1020

gaattcccag aattaagaaa ggttgttgat aagggttttg acgacccttc caacgttgac 1080

attgctttgg agtatttggg taagtctaga ggtattcaaa gaaccagaga attggctcaa 1140

aaacatgcca atttggccac cgtcgccttg gattctttac cagaatccaa cgacgacgat 1200

gttaagaagt ctcgtagagc tttattggac ttggctcaaa gagttattac tagaaacaag 1260

taa 1263

<210> 139
<211> 513
<212> PRT
<213> Artificial Sequence

<220>
<223> Truncated geranylgeranyl pyrophosphate synthase Es_GPPS_NTrunc

<400> 139

Met Arg Arg Ser Gly Ser Ala Thr Ala Ala Ala Ala Ala Thr Leu Ala
1 5 10 15

Arg His Ala Asn Ala Cys Cys Arg Ala Arg Ser Pro Ala Leu Gly Leu
20 25 30

Leu Pro Gly Ala Ala Ala Ser Ser Ser Thr His Arg Ala Ala Leu Ser
35 40 45

Ser Asn Ser Gly His Gly Gly Asp Gly Ser Gly His Tyr Asp Ala Ala
50 55 60

Met Arg Arg Arg Glu Ser Cys Ala Ser Arg Ser Arg His Arg Trp Ser
65 70 75 80

Gly Gln Glu Ala Ala Ala Ala Ser Ala Thr Thr Thr Thr Ala Arg Arg
85 90 95

Ala Pro Gly Gly Val Ala Gly Ala Ser Gly Gln Gly Ser Ala Ala Gly
100 105 110

Ser Val Arg Ala Leu Ser Ser Ser Phe Leu Ala Asp Ala Val Arg Glu
115 120 125

Thr Ala Thr Asn His Cys Ile Asp Arg Val Val Asn Gly Gly Leu Asp
130 135 140

Gly Ser Val Pro Val Asp Lys Asp Thr Pro Thr Val Glu Val Gln Asp
145 150 155 160

```
Phe Val Tyr Asp Ile Asp Phe Ala Gln Arg Pro Ser Gly Ala Ser Gln
                165                 170                 175

Ser Leu Ala Asp Gly Pro Asp Pro Phe Glu Leu Val Ser Ala Glu Leu
            180                 185                 190

Ala Gly Leu Ser Asp Gly Ile Lys Ser Leu Ile Gly Thr Glu His Ala
        195                 200                 205

Val Leu Asn Ala Ala Ala Lys Tyr Phe Phe Glu Leu Asp Gly Gly Lys
    210                 215                 220

Lys Ile Arg Pro Thr Met Val Ile Leu Met Ser Gln Ala Cys Asn Ser
225                 230                 235                 240

Asn Ser Gln Gln Val Arg Pro Asp Val Gln Pro Gly Thr Glu Leu Val
                245                 250                 255

Asn Pro Leu Gln Leu Arg Leu Ala Glu Ile Thr Glu Met Ile His Ala
            260                 265                 270

Ala Ser Leu Phe His Asp Asp Val Ile Asp Glu Ala Asp Thr Arg Arg
        275                 280                 285

Gly Val Pro Ser Val Asn Lys Val Phe Gly Asn Lys Leu Ala Ile Leu
    290                 295                 300

Ala Gly Asp Phe Leu Leu Ala Arg Ser Ser Met Ser Leu Ala Arg Leu
305                 310                 315                 320

Arg Ser Leu Glu Ser Val Glu Leu Met Ser Ala Ala Ile Glu His Leu
                325                 330                 335

Val Lys Gly Glu Val Leu Gln Met Arg Pro Thr Glu Asp Gly Gly Gly
            340                 345                 350

Ala Phe Glu Tyr Tyr Val Arg Lys Asn Tyr Tyr Lys Thr Gly Ser Leu
        355                 360                 365
```

```
Met Ala Asn Ser Cys Lys Ala Ser Ala Val Leu Gly Gln His Asp Leu
    370                 375                 380


Glu Val Gln Glu Val Ala Phe Glu Tyr Gly Lys Arg Val Gly Leu Ala
385                 390                 395                 400


Phe Gln Leu Val Asp Asp Ile Leu Asp Phe Glu Gly Asn Thr Phe Thr
                405                 410                 415


Leu Gly Lys Pro Ala Leu Asn Asp Leu Arg Gln Gly Leu Ala Thr Ala
            420                 425                 430


Pro Val Leu Leu Ala Ala Glu Gln Gln Pro Gly Leu Ala Lys Leu Ile
        435                 440                 445


Ser Arg Lys Phe Arg Gly Pro Gly Asp Val Asp Glu Ala Leu Glu Leu
    450                 455                 460


Val His Arg Ser Asp Gly Ile Ala Arg Ala Lys Glu Val Ala Val Val
465                 470                 475                 480


Gln Ala Glu Lys Ala Met Ser Ala Ile Leu Thr Leu His Asp Ser Pro
                485                 490                 495


Ala Gln Asn Ala Leu Val Gln Leu Ala His Lys Ile Val Asn Arg Asn
            500                 505                 510


His
```

<210> 140
<211> 1542
<212> DNA
<213> Artificial Sequence

<220>
<223> Truncated geranylgeranyl pyrophosphate synthase Es_GPPS_NTrunc

<400> 140
atgcgtagat ccggttccgc taccgccgct gccgctgcca ccttagccag acacgccaac 60

gcctgttgta gagcccgttc cccagcttta ggtttgttgc ctggtgccgc cgcttcttcc 120

tctactcaca gagccgcctt gtcttctaat tctggtcatg gtggtgatgg ttccggtcat 180

tacgacgctg ctatgagaag aagagaatct tgcgcttcca gatctcgtca cagatggtcc 240

ggtcaagaag ctgccgccgc ctccgccact accaccaccg ctcgtcgtgc tccaggtggt 300

gtcgccggtg cttctggtca aggttctgct gccggttctg ttagagcctt atcctcttct 360

tttttagccg atgccgttcg tgaaaccgct actaaccact gtatcgaccg tgttgtcaac 420

ggtggtttgg acggttctgt cccagtcgat aaagataccc caactgtcga agttcaagac 480

tttgtttatg atattgactt tgctcaacgt ccatccggtg cctctcaatc tttagctgac 540

ggtccagatc cattcgagtt agtttccgct gagttggccg gtttgtctga tggtattaag 600

tctttgattg gtaccgaaca tgctgtcttg aacgccgccg ccaaatattt cttcgaatta 660

gatggtggta aaaagatcag acctactatg gttatcttaa tgtcccaagc ttgtaactct 720

aattcccaac aagttcgtcc tgacgttcaa ccaggtactg aattagtcaa tcctttgcaa 780

ttaagattgg ctgaaatcac cgagatgatt catgctgctt ctttattcca cgacgatgtt 840

attgatgagg ctgatactag acgtggtgtc ccttctgtta ataaagtttt cggtaacaaa 900

ttagccatct tggccggtga cttcttattg gctagatcct ctatgtcctt ggcccgttta 960

agatccttgg agtccgtcga attgatgtcc gccgctatcg aacacttggt caaaggtgaa 1020

gttttacaaa tgcgtccaac tgaggacggt ggtggtgctt tcgagtacta cgtcagaaaa 1080

aattactaca agactggttc tttgatggct aactcctgta aggcctccgc cgttttaggt 1140

caacacgact tagaagtcca agaggtcgct tttgaatacg gtaagagagt cggtttggct 1200

ttccaattgg ttgacgatat tttagatttt gaaggtaata ctttcacttt gggtaagcca 1260

gctttaaacg acttgagaca aggtttagcc actgcccctg tcttgttagc tgctgaacaa 1320

caacctggtt tagctaaatt gatctccaga aagtttagag gtcctggtga tgtcgatgaa 1380

gctttggaat tggtccacag atccgacggt attgctagag ctaaggaggt tgctgttgtc 1440

caagccgaaa aagctatgtc tgccattttg accttgcatg actccccagc tcaaaatgct 1500

ttggttcaat tggctcacaa aatcgtcaat cgtaaccatt ag 1542

<210> 141

<211> 415
<212> PRT
<213> Solanum sp.

<400> 141

```
Met Ile Phe Ser Lys Gly Leu Ala Gln Ile Ser Arg Asn Arg Phe Ser
1               5                   10                  15


Arg Cys Arg Trp Leu Phe Ser Leu Arg Pro Ile Pro Gln Leu His Gln
            20                  25                  30


Ser Asn His Ile His Asp Pro Pro Lys Val Leu Gly Cys Arg Val Ile
        35                  40                  45


His Ser Trp Val Ser Asn Ala Leu Ser Gly Ile Gly Gln Gln Ile His
    50                  55                  60


Gln Gln Ser Thr Ala Val Ala Glu Glu Gln Val Asp Pro Phe Ser Leu
65                  70                  75                  80


Val Ala Asp Glu Leu Ser Leu Leu Thr Asn Arg Leu Arg Ser Met Val
                85                  90                  95


Val Ala Glu Val Pro Lys Leu Ala Ser Ala Ala Glu Tyr Phe Phe Lys
            100                 105                 110


Leu Gly Val Glu Gly Lys Arg Phe Arg Pro Thr Val Leu Leu Leu Met
        115                 120                 125


Ala Thr Ala Leu Asn Val Gln Ile Pro Arg Ser Ala Pro Gln Val Asp
    130                 135                 140


Val Asp Ser Phe Ser Gly Asp Leu Arg Thr Arg Gln Gln Cys Ile Ala
145                 150                 155                 160


Glu Ile Thr Glu Met Ile His Val Ala Ser Leu Leu His Asp Asp Val
                165                 170                 175


Leu Asp Asp Ala Asp Thr Arg Arg Gly Ile Gly Ser Leu Asn Phe Val
            180                 185                 190
```

```
Met Gly Asn Lys Leu Ala Val Leu Ala Gly Asp Phe Leu Leu Ser Arg
        195                 200                 205

Ala Cys Val Ala Leu Ala Ser Leu Lys Asn Thr Glu Val Val Cys Leu
    210                 215                 220

Leu Ala Thr Val Val Glu His Leu Val Thr Gly Glu Thr Met Gln Met
225                 230                 235                 240

Thr Thr Ser Ser Asp Glu Arg Cys Ser Met Glu Tyr Tyr Met Gln Lys
                245                 250                 255

Thr Tyr Tyr Lys Thr Ala Ser Leu Ile Ser Asn Ser Cys Lys Ala Ile
            260                 265                 270

Ala Leu Leu Ala Gly His Ser Ala Glu Val Ser Val Leu Ala Phe Asp
        275                 280                 285

Tyr Gly Lys Asn Leu Gly Leu Ala Phe Gln Leu Ile Asp Asp Val Leu
    290                 295                 300

Asp Phe Thr Gly Thr Ser Ala Thr Leu Gly Lys Gly Ser Leu Ser Asp
305                 310                 315                 320

Ile Arg His Gly Ile Val Thr Ala Pro Ile Leu Tyr Ala Met Glu Glu
                325                 330                 335

Phe Pro Gln Leu Arg Thr Leu Val Asp Arg Gly Phe Asp Asp Pro Val
            340                 345                 350

Asn Val Glu Ile Ala Leu Asp Tyr Leu Gly Lys Ser Arg Gly Ile Gln
        355                 360                 365

Arg Thr Arg Glu Leu Ala Arg Lys His Ala Ser Leu Ala Ser Ala Ala
    370                 375                 380

Ile Asp Ser Leu Pro Glu Ser Asp Asp Glu Glu Val Gln Arg Ser Arg
385                 390                 395                 400
```

Arg Ala Leu Val Glu Leu Thr His Arg Val Ile Thr Arg Thr Lys
405 410 415

<210> 142
<211> 1248
<212> DNA
<213> Solanum sp.

<400> 142
atgatctttt ccaagggttt agctcaaatc tctcgtaata gattctctcg ttgcagatgg 60

ttattctctt tgcgtccaat tcctcaatta caccaatcca atcacatcca cgacccacca 120

aaagttttgg gttgtcgtgt cattcactct tgggtttcta atgccttgtc tggtatcggt 180

caacaaatcc atcaacaatc tactgccgtt gccgaggaac aagtcgaccc tttttctttg 240

gttgctgatg agttatcctt gttaaccaac agattgagat ccatggttgt cgctgaagtc 300

cctaagttag cctccgccgc tgagtatttc tttaagttag gtgtcgaagg taaacgtttc 360

cgtccaactg tcttgttgtt gatggccact gccttaaacg tccaaattcc tcgttctgct 420

ccacaagttg acgttgactc tttttctggt gacttgagaa ctagacaaca atgtatcgct 480

gaaattactg aaatgattca cgtcgcctct ttgttgcatg atgacgtctt agatgatgct 540

gatactagaa gaggtattgg ttccttaaat tttgttatgg gtaataagtt ggctgttttg 600

gctggtgatt tcttgttatc cagagcctgc gtcgccttag cctccttgaa gaacaccgaa 660

gttgtctgtt tattggccac cgttgtcgaa catttggtta ccggtgaaac tatgcaaatg 720

actacctcct ccgatgaaag atgttccatg gaatactaca tgcaaaagac ctactataag 780

actgcctctt tgatttctaa ctcttgtaaa gccattgcct tgttagccgg tcactctgct 840

gaagtttctg tcttggcctt cgattacggt aagaacttag gtttggcttt tcaattgatc 900

gacgatgttt tggacttcac cggtacctct gctactttgg gtaaaggttc cttgtccgat 960

atcagacatg gtatcgttac tgctcctatt ttgtatgcta tggaagaatt ccctcaatta 1020

cgtactttgg ttgacagagg tttcgatgat ccagttaatg ttgagatcgc tttggattac 1080

ttgggtaaat cccgtggtat tcaaagaact agagaattag ccagaaagca tgcctcttta 1140

gcctctgccg ccatcgattc cttgcctgaa tccgacgatg aggaagttca aagatctcgt 1200

agagctttgg tcgaattgac ccatagagtc attactcgta ctaagtaa 1248

<210> 143
<211> 330
<212> PRT
<213> Hevea brasiliensis

<400> 143

Met Gln Phe Leu Arg Gly Leu Ser Pro Ile Ser Arg Ser Gly Leu Arg
1 5 10 15

Leu Phe Leu Ser Arg Gln Leu Tyr Pro Phe Pro Val Ala Asn Ser Ser
20 25 30

Gln Leu Leu Gly Asp Ser Thr Gln Lys Val Phe Asn Arg Arg Glu Thr
35 40 45

Tyr Ser Trp Ser Leu Val Asp Ser His Gly Phe Lys Gln Gln Ile His
50 55 60

His Gln Ser Ser Phe Leu Ser Glu Glu Pro Leu Asp Pro Phe Ser Leu
65 70 75 80

Val Ala Asp Glu Leu Ser Leu Val Ala Asn Arg Leu Arg Ala Met Leu
85 90 95

Val Ser Glu Val Pro Lys Leu Ala Ser Ala Ala Glu Tyr Phe Phe Lys
100 105 110

Met Gly Val Glu Gly Lys Arg Leu Arg Pro Thr Val Leu Leu Leu Met
115 120 125

Ala Thr Ala Leu Asn Val His Ile His Glu Pro Met Pro Asn Gly Val
130 135 140

Gly Asp Thr Leu Gly Ala Glu Leu Arg Thr Arg Gln Gln Cys Ile Ala
145 150 155 160

Glu Ile Thr Glu Met Ile His Val Ala Ser Leu Leu His Asp Asp Val
165 170 175

```
Leu Asp Asp Ala Asp Thr Arg Arg Gly Ile Gly Ser Leu Asn Phe Val
            180                 185                 190

Met Gly Asn Lys Val Ala Val Leu Ala Gly Asp Phe Leu Leu Ser Arg
        195                 200                 205

Ala Cys Val Ala Leu Ala Ser Leu Lys Asn Thr Glu Val Val Ser Leu
    210                 215                 220

Leu Ala Thr Val Val Glu His Leu Val Thr Gly Glu Thr Met Gln Met
225                 230                 235                 240

Thr Ser Thr Ser Glu Gln Arg Cys Ser Met Asp His Tyr Met Gln Lys
                245                 250                 255

Thr Tyr Tyr Lys Thr Ala Ser Leu Ile Ser Asp Ser Cys Lys Ala Ile
            260                 265                 270

Ala Leu Leu Ala Gly Gln Thr Thr Glu Val Ala Met Leu Ala Phe Glu
        275                 280                 285

Tyr Gly Lys Ser Leu Gly Leu Ala Phe Gln Leu Ile Asp Asp Val Leu
    290                 295                 300

Asp Phe Thr Gly Thr Ser Ala Ser Leu Gly Lys Gly Ser Leu Ser Asp
305                 310                 315                 320

Ile Arg His Val Ile Arg Leu Ser Leu Ile
                325                 330
```

<210> 144
<211> 993
<212> DNA
<213> Hevea brasiliensis

<400> 144
atgcaatttt tgagaggttt gtcccctatt tccagatccg gtttgcgttt attcttatct 60

cgtcaattat atccattccc agtcgccaac tcctcccaat tattaggtga ctctactcaa 120

aaggttttta acagacgtga gacttactct tggtctttgg tcgactctca cggttttaag 180

caacaaattc atcaccaatc ctctttttg tctgaagaac cattggatcc attctctttg 240

gttgctgatg aattatcctt ggtcgctaac agattgcgtg ctatgttggt ttctgaagtc 300

ccaaaattag cctccgccgc tgaatatttt ttcaagatgg gtgttgaagg taagagattg 360

cgtccaaccg tcttgttatt aatggccact gctttaaacg ttcatatcca tgaacctatg 420

cctaacggtg ttggtgacac tttgggtgcc gaattgagaa ctagacaaca atgcatcgct 480

gaaatcaccg aaatgatcca tgttgcttct ttattacatg acgacgtttt agacgatgcc 540

gataccagaa gaggtattgg ttctttgaac ttcgttatgg gtaacaaggt tgctgttttg 600

gccggtgact ttttgttgtc cagagcttgt gttgccttag cttctttgaa gaataccgaa 660

gtcgtttctt tattggccac cgtcgtcgaa cacttggtta ctggtgagac tatgcaaatg 720

acctccactt ctgagcaacg ttgttccatg gatcattata tgcaaaagac ttactataag 780

accgcttcct taatctctga ttcctgtaaa gccatcgcct tgttagctgg tcaaactacc 840

gaggtcgcca tgttggcctt cgaatatggt aagtctttgg gtttagcttt tcaattaatc 900

gacgatgttt tagacttcac cggtacttct gcttccttgg gtaagggttc tttgtccgac 960

attagacacg ttattagatt atccttaatt taa 993

<210> 145
<211> 566
<212> PRT
<213> Artificial Sequence

<220>
<223> Mutant medium-chain fatty acid CoA ligase Ec_FADK_v1

<400> 145

Met His Pro Thr Gly Pro His Leu Gly Pro Asp Val Leu Phe Arg Glu
1 5 10 15

Ser Asn Met Lys Val Thr Leu Thr Phe Asn Glu Gln Arg Arg Ala Ala
20 25 30

Tyr Arg Gln Gln Gly Leu Trp Gly Asp Ala Ser Leu Ala Asp Tyr Trp
35 40 45

```
Gln Gln Thr Ala Arg Ala Met Pro Asp Lys Ile Ala Val Val Asp Asn
    50                  55                  60

His Gly Ala Ser Tyr Thr Tyr Ser Ala Leu Asp His Ala Ala Ser Cys
65                  70                  75                  80

Leu Ala Asn Trp Met Leu Ala Lys Gly Ile Glu Ser Gly Asp Arg Ile
                85                  90                  95

Ala Phe Gln Leu Pro Gly Trp Cys Glu Phe Thr Val Ile Tyr Leu Ala
            100                 105                 110

Cys Leu Lys Ile Gly Ala Val Ser Val Pro Leu Leu Pro Ser Trp Arg
        115                 120                 125

Glu Ala Glu Leu Val Trp Val Leu Asn Lys Cys Gln Ala Lys Met Phe
    130                 135                 140

Phe Ala Pro Thr Leu Phe Lys Gln Thr Arg Pro Val Asp Leu Ile Leu
145                 150                 155                 160

Pro Leu Gln Asn Gln Leu Pro Gln Leu Gln Gln Ile Val Gly Val Asp
                165                 170                 175

Lys Leu Ala Pro Ala Thr Ser Ser Leu Ser Leu Ser Gln Ile Ile Ala
            180                 185                 190

Asp Asn Thr Ser Leu Thr Thr Ala Ile Thr Thr His Gly Asp Glu Leu
        195                 200                 205

Ala Ala Val Leu Phe Thr Ser Gly Thr Glu Gly Leu Pro Lys Gly Val
    210                 215                 220

Met Leu Thr His Asn Asn Ile Leu Ala Ser Glu Arg Ala Tyr Cys Ala
225                 230                 235                 240

Arg Leu Asn Leu Thr Trp Gln Asp Val Phe Met Met Pro Ala Pro Leu
                245                 250                 255
```

```
Gly His Ala Thr Gly Phe Leu His Gly Val Thr Ala Pro Phe Leu Ile
            260                 265                 270

Gly Ala Arg Ser Val Leu Leu Asp Ile Phe Thr Pro Asp Ala Cys Leu
        275                 280                 285

Ala Leu Leu Glu Gln Gln Arg Cys Thr Cys Met Leu Gly Ala Thr Pro
    290                 295                 300

Phe Val Tyr Asp Leu Leu Asn Val Leu Glu Lys Gln Pro Ala Asp Leu
305                 310                 315                 320

Ser Ala Leu Arg Phe Phe Leu Cys Gly Gly Thr Thr Ile Pro Lys Lys
                325                 330                 335

Val Ala Arg Glu Cys Gln Gln Arg Gly Ile Lys Leu Leu Ser Val Tyr
            340                 345                 350

Gly Ser Thr Glu Ser Ser Pro His Ala Val Val Asn Leu Asp Asp Pro
        355                 360                 365

Leu Ser Arg Phe Met His Thr Asp Gly Tyr Ala Ala Ala Gly Val Glu
    370                 375                 380

Ile Lys Val Val Asp Asp Ala Arg Lys Thr Leu Pro Pro Gly Cys Glu
385                 390                 395                 400

Gly Glu Glu Ala Ser Arg Gly Pro Asn Val Phe Met Gly Tyr Phe Asp
                405                 410                 415

Glu Pro Glu Leu Thr Ala Arg Ala Leu Asp Glu Glu Gly Trp Tyr Tyr
            420                 425                 430

Ser Gly Asp Leu Cys Arg Met Asp Glu Ala Gly Tyr Ile Lys Ile Thr
        435                 440                 445

Gly Arg Lys Lys Asp Ile Ile Val Arg Gly Gly Glu Asn Ile Ser Ser
    450                 455                 460
```

```
Arg Glu Val Glu Asp Ile Leu Leu Gln His Pro Lys Ile His Asp Ala
465                 470                 475                 480


Cys Val Val Ala Met Ser Asp Glu Arg Leu Gly Glu Arg Ser Cys Ala
                485                 490                 495


Tyr Val Val Leu Lys Ala Pro His His Ser Leu Ser Leu Glu Glu Val
            500                 505                 510


Val Ala Phe Phe Ser Arg Lys Arg Val Ala Lys Tyr Lys Tyr Pro Glu
        515                 520                 525


His Ile Val Val Ile Glu Lys Leu Pro Arg Thr Thr Ser Gly Lys Ile
    530                 535                 540


Gln Lys Phe Leu Leu Arg Lys Asp Ile Met Arg Arg Leu Thr Gln Asp
545                 550                 555                 560


Val Cys Glu Glu Ile Glu
                565


<210>  146
<211>  1701
<212>  DNA
<213>  Escherichia coli

<400>  146
atgcatccaa ctggtccaca cttaggtcct gatgtcttat ttagagaatc taatatgaaa       60

gtcactttga cctttaatga acaaagacgt gccgcttaca gacaacaagg tttgtggggt      120

gacgcttctt tggctgacta ctggcaacaa actgctagag ctatgccaga caagatcgcc      180

gttgtcgata accacggtgc ttcttatacc tactctgctt tggatcatgc cgcttcttgt      240

ttggctaatt ggatgttggc taagggtatc gaatctggtg atcgtattgc ttttcaattg      300

ccaggttggt gtgaatttac cgttatctac ttggcttgtt tgaagattgg tgctgtttct      360

gtcccattgt tgccatcttg gagagaagcc gaattggttt gggttttgaa caaatgccaa      420

gctaagatgt tctttgctcc aaccttgttc aagcaaacta gaccagttga cttgatttta      480

cctttacaaa atcaattacc acaattgcaa caaatcgttg gtgttgacaa gttagctcca      540
```

gccacctcct ctttgtcctt gtcccaaatt atcgctgata atacttcttt aaccaccgct 600

atcactactc acggtgatga gttggctgct gttttgttca cttccggtac tgagggtttg 660

ccaaagggtg ttatgttgac ccacaataac attttggctt ccgaaagagc ttattgtgct 720

cgtttgaact tgacctggca agatgttttc atgatgccag ctccattggg tcatgctact 780

ggtttcttgc acggtgttac tgccccattc ttgattggtg ctagatctgt cttgttggat 840

atctttaccc cagacgcttg cttagcttta ttggaacaac aaagatgtac ctgtatgtta 900

ggtgctactc catttgttta cgatttattg aacgttttgg aaaaacaacc agctgatttg 960

tctgccttga gattcttttt gtgtggtggt actactattc caaagaaagt tgctagagaa 1020

tgccaacaaa gaggtatcaa gttgttgtcc gtctatggtt ccactgaatc ttctcctcat 1080

gctgttgtca atttagatga cccattgtct agattcatgc acaccgatgg ttacgccgct 1140

gctggtgttg agattaaggt tgtcgacgat gctagaaaga ccttacctcc aggttgtgaa 1200

ggtgaagaag cctctagagg tccaaatgtc tttatgggtt acttcgacga gccagaattg 1260

actgctagag ctttagatga ggaaggttgg tattactctg gtgatttgtg tagaatggat 1320

gaagctggtt acattaaaat cactggtaga aagaaggaca ttattgttag aggtggtgaa 1380

aatatctcct ccagagaagt tgaagatatt ttattgcaac acccaaagat tcatgatgct 1440

tgtgttgttg ctatgtccga tgagagatta ggtgaaagat cttgtgctta cgttgttttg 1500

aaggctccac atcactcttt gtctttagaa gaagtcgttg ctttcttctc tagaaagaga 1560

gtcgccaagt acaagtaccc agaacacatt gttgttatcg aaaaattgcc tagaactact 1620

tctggtaaaa ttcaaaaatt cttgttgaga aaggatatca tgagacgttt gacccaagat 1680

gtctgtgaag aaattgaata a 1701

<210> 147
<211> 566
<212> PRT
<213> Escherichia coli

<400> 147

Met His Pro Thr Gly Pro His Leu Gly Pro Asp Val Leu Phe Arg Glu
1 5 10 15

```
Ser Asn Met Lys Val Thr Leu Thr Phe Asn Glu Gln Arg Arg Ala Ala
            20                  25                  30

Tyr Arg Gln Gln Gly Leu Trp Gly Asp Ala Ser Leu Ala Asp Tyr Trp
        35                  40                  45

Gln Gln Thr Ala Arg Ala Met Pro Asp Lys Ile Ala Val Val Asp Asn
    50                  55                  60

His Gly Ala Ser Tyr Thr Tyr Ser Ala Leu Asp His Ala Ala Ser Cys
65                  70                  75                  80

Leu Ala Asn Trp Met Leu Ala Lys Gly Ile Glu Ser Gly Asp Arg Ile
                85                  90                  95

Ala Phe Gln Leu Pro Gly Trp Cys Glu Phe Thr Val Ile Tyr Leu Ala
            100                 105                 110

Cys Leu Lys Ile Gly Ala Val Ser Val Pro Leu Leu Pro Ser Trp Arg
        115                 120                 125

Glu Ala Glu Leu Val Trp Val Leu Asn Lys Cys Gln Ala Lys Met Phe
    130                 135                 140

Phe Ala Pro Thr Leu Phe Lys Gln Thr Arg Pro Val Asp Leu Ile Leu
145                 150                 155                 160

Pro Leu Gln Asn Gln Leu Pro Gln Leu Gln Gln Ile Val Gly Val Asp
                165                 170                 175

Lys Leu Ala Pro Ala Thr Ser Ser Leu Ser Leu Ser Gln Ile Ile Ala
            180                 185                 190

Asp Asn Thr Ser Leu Thr Thr Ala Ile Thr Thr His Gly Asp Glu Leu
        195                 200                 205

Ala Ala Val Leu Phe Thr Ser Gly Thr Glu Gly Leu Pro Lys Gly Val
    210                 215                 220
```

```
Met Leu Thr His Asn Asn Ile Leu Ala Ser Glu Arg Ala Tyr Cys Ala
225                 230                 235                 240

Arg Leu Asn Leu Thr Trp Gln Asp Val Phe Met Met Pro Ala Pro Leu
                245                 250                 255

Gly His Ala Thr Gly Phe Leu His Gly Val Thr Ala Pro Phe Leu Ile
            260                 265                 270

Gly Ala Arg Ser Val Leu Leu Asp Ile Phe Thr Pro Asp Ala Cys Leu
        275                 280                 285

Ala Leu Leu Glu Gln Gln Arg Cys Thr Cys Met Leu Gly Ala Thr Pro
    290                 295                 300

Phe Val Tyr Asp Leu Leu Asn Val Leu Glu Lys Gln Pro Ala Asp Leu
305                 310                 315                 320

Ser Ala Leu Arg Phe Phe Leu Cys Gly Gly Thr Thr Ile Pro Lys Lys
                325                 330                 335

Val Ala Arg Glu Cys Gln Gln Arg Gly Ile Lys Leu Leu Ser Val Tyr
            340                 345                 350

Gly Ser Thr Glu Ser Ser Pro His Ala Val Val Asn Leu Asp Asp Pro
        355                 360                 365

Leu Ser Arg Phe Met His Thr Asp Gly Tyr Ala Ala Ala Gly Val Glu
    370                 375                 380

Ile Lys Val Val Asp Asp Ala Arg Lys Thr Leu Pro Pro Gly Cys Glu
385                 390                 395                 400

Gly Glu Glu Ala Ser Arg Gly Pro Asn Val Phe Met Gly Tyr Phe Asp
                405                 410                 415

Glu Pro Glu Leu Thr Ala Arg Ala Leu Asp Glu Glu Gly Trp Tyr Tyr
            420                 425                 430
```

Ser Gly Asp Leu Cys Arg Met Asp Glu Ala Gly Tyr Ile Lys Ile Thr
435 440 445

Gly Arg Lys Lys Asp Ile Ile Val Arg Gly Gly Glu Asn Ile Ser Ser
450 455 460

Arg Glu Val Glu Asp Ile Leu Leu Gln His Pro Lys Ile His Asp Ala
465 470 475 480

Cys Val Val Ala Met Ser Asp Glu Arg Leu Gly Glu Arg Ser Cys Ala
485 490 495

Tyr Val Val Leu Lys Ala Pro His His Ser Leu Ser Leu Glu Glu Val
500 505 510

Val Ala Phe Phe Ser Arg Lys Arg Val Ala Lys Tyr Lys Tyr Pro Glu
515 520 525

His Ile Val Val Ile Glu Lys Leu Pro Arg Thr Thr Ser Gly Lys Ile
530 535 540

Gln Lys Phe Leu Leu Arg Lys Asp Ile Met Arg Arg Leu Thr Gln Asp
545 550 555 560

Val Cys Glu Glu Ile Glu
565

<210> 148
<211> 1701
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial medium-chain fatty acid CoA ligase Ec_FADK_v2 nucleotide sequence

<400> 148
atgcatccaa ctggtcctca cttaggtcca gatgtcttat tcagagaatc taacatgaaa 60

gtcactttaa cttttaacga acaacgtaga gctgcttata gacaacaagg tttgtggggt 120

gatgcttcct tggctgacta ctggcaacaa actgctagag ccatgccaga taaaattgcc 180

gttgttgaca atcacggtgc ttcttacact tattctgcct tagatcacgc tgcttcctgt 240

ttagctaact ggatgttagc taagggtatt gaatccggtg atagaattgc tttccaattg 300

ccaggttggt gcgaatttac tgtcatttat ttagcttgtt taaagattgg tgccgtctcc 360

gtccctttgt tgccatcctg gagagaggcc gagttggttt gggttttaaa caagtgtcaa 420

gctaaaatgt tctttgctcc taccttgttc aagcaaacca gaccagttga cttaattttg 480

ccattacaaa accaattacc acaattgcaa caaatcgtcg gtgttgataa attagctcca 540

gccacttctt ctttgtcctt atcccaaatt attgctgata acacttcttt aactactgct 600

attactactc acggtgatga attggccgct gttttgttca cttccggtac tgaaggtttg 660

cctaaaggtg tcatgttgac tcacaacaac attttggcct ctgaaagagc ttactgtgcc 720

cgtttaaatt tgacctggca agatgtcttc atgatgcctg ctccattggg tcacgctacc 780

ggtttcttac acggtgtcac tgccccattc ttgatcggtg ctcgttctgt tttattggat 840

atctttactc cagatgcttg cttggcttta ttggaacaac aaagatgtac ctgcatgtta 900

ggtgctactc ctttcgtcta tgatttattg aacgtcttag aaaaacaacc agctgattta 960

tccgctttaa gattcttttt gtgtggtggt actactatcc caaaaaaggt cgccagagaa 1020

tgtcaacaaa gaggtattaa attattgtcc gtttatggtt ccactgaatc ttcccctcat 1080

gctgttgtca atttagacga ccctttgtcc agattcatgc acactgatgg ttacgccgct 1140

gctggtgtcg aaatcaaggt tgttgatgac gctagaaaaa ctttaccacc tggttgcgaa 1200

ggtgaagagg cttccagagg tccaaacgtc tttatgggtt actttgatga accagaattg 1260

actgccagag ctttggatga ggaaggttgg tattattctg gtgatttgtg tagaatggat 1320

gaagccggtt acatcaagat caccggtaga aagaaagaca tcatcgttag aggtggtgaa 1380

aacatttctt ctagagaagt tgaagacatt ttgttgcaac acccaaagat ccacgacgct 1440

tgtgtcgtcg ccatgtctga cgaaagattg ggtgaacgtt cttgtgctta cgtcgtcttg 1500

aaagccccac accactcttt gtctttggaa gaagtcgttg ctttttttctc tcgtaagcgt 1560

gttgccaagt acaagtaccc agagcacatc gttgttattg aaaaattgcc tcgtactact 1620

tccggtaaga ttcaaaagtt cttattacgt aaggacatca tgagaagatt gactcaagac 1680

gtctgcgaag aaattgaata a 1701

<210> 149
<211> 521
<212> PRT
<213> Artificial Sequence

<220>
<223> Truncated acyl activating enzyme (Cs_AAE3_Ctrunc)

<400> 149

```
Met Glu Lys Ser Gly Tyr Gly Arg Asp Gly Ile Tyr Arg Ser Leu Arg
1               5                   10                  15


Pro Pro Leu His Leu Pro Asn Asn Asn Asn Leu Ser Met Val Ser Phe
            20                  25                  30


Leu Phe Arg Asn Ser Ser Ser Tyr Pro Gln Lys Pro Ala Leu Ile Asp
        35                  40                  45


Ser Glu Thr Asn Gln Ile Leu Ser Phe Ser His Phe Lys Ser Thr Val
    50                  55                  60


Ile Lys Val Ser His Gly Phe Leu Asn Leu Gly Ile Lys Lys Asn Asp
65                  70                  75                  80


Trp Leu Ile Tyr Ala Pro Asn Ser Ile His Phe Pro Val Cys Phe Leu
                85                  90                  95


Gly Ile Ile Ala Ser Gly Ala Ile Ala Thr Thr Ser Asn Pro Leu Tyr
            100                 105                 110


Thr Val Ser Glu Leu Ser Lys Gln Val Lys Asp Ser Asn Pro Lys Leu
        115                 120                 125


Ile Ile Thr Val Pro Gln Leu Leu Glu Lys Val Lys Gly Phe Asn Leu
    130                 135                 140


Pro Thr Ile Leu Ile Gly Pro Asp Ser Glu Gln Glu Ser Ser Ser Asp
145                 150                 155                 160
```

```
Lys Val Met Thr Phe Asn Asp Leu Val Asn Leu Gly Gly Ser Ser Gly
                165                 170                 175

Ser Glu Phe Pro Ile Val Asp Asp Phe Lys Gln Ser Asp Thr Ala Ala
            180                 185                 190

Leu Leu Tyr Ser Ser Gly Thr Thr Gly Met Ser Lys Gly Trp Leu Thr
        195                 200                 205

His Lys Asn Phe Ile Ala Ser Ser Leu Met Val Thr Met Glu Gln Asp
    210                 215                 220

Leu Val Gly Glu Met Asp Asn Val Phe Leu Cys Phe Leu Pro Met Phe
225                 230                 235                 240

His Val Phe Gly Leu Ala Ile Ile Thr Tyr Ala Gln Leu Gln Arg Gly
                245                 250                 255

Asn Thr Val Ile Ser Ala Arg Phe Asp Leu Glu Lys Met Leu Lys Asp
            260                 265                 270

Val Glu Lys Tyr Val Thr His Leu Trp Trp Pro Pro Val Ile Leu Ala
        275                 280                 285

Leu Ser Lys Asn Ser Met Val Lys Phe Asn Leu Ser Ser Ile Lys Tyr
    290                 295                 300

Ile Gly Ser Gly Ala Ala Pro Leu Gly Lys Asp Leu Met Glu Glu Cys
305                 310                 315                 320

Ser Lys Trp Pro Tyr Gly Ile Val Ala Gln Gly Tyr Gly Met Thr Glu
                325                 330                 335

Thr Cys Gly Ile Val Ser Met Glu Asp Ile Arg Gly Gly Lys Arg Asn
            340                 345                 350

Ser Gly Ser Ala Gly Met Leu Ala Ser Gly Val Glu Ala Gln Ile Val
        355                 360                 365
```

Ser Val Asp Thr Leu Lys Pro Leu Pro Pro Asn Gln Leu Gly Glu Ile
370 375 380

Trp Val Lys Gly Pro Asn Met Met Gln Gly Tyr Phe Asn Asn Pro Gln
385 390 395 400

Ala Thr Lys Leu Thr Ile Asp Lys Lys Gly Trp Val His Thr Gly Asp
405 410 415

Leu Gly Tyr Phe Asp Glu Asp Gly His Leu Tyr Trp Asp Arg Ile Lys
420 425 430

Glu Leu Ile Lys Tyr Lys Gly Phe Gln Val Ala Pro Ala Glu Leu Glu
435 440 445

Gly Leu Leu Val Ser His Pro Glu Ile Leu Asp Ala Trp Ile Pro Phe
450 455 460

Pro Asp Ala Glu Ala Gly Glu Val Pro Val Ala Tyr Trp Arg Ser Pro
465 470 475 480

Asn Ser Ser Leu Thr Glu Asn Asp Val Lys Lys Phe Ile Ala Gly Gln
485 490 495

Val Ala Ser Phe Lys Arg Leu Arg Lys Val Thr Phe Ile Asn Ser Val
500 505 510

Pro Lys Ser Ala Ser Gly Lys Ile Leu
515 520

<210> 150
<211> 1563
<212> DNA
<213> Artificial Sequence

<220>
<223> Truncated acyl activating enzyme (Cs_AAE3_Ctrunc)

<400> 150
atggaaaaat ctggttatgg tagagacggt atctacagat ccttgcgtcc tccattacac 60

ttgccaaaca ataataactt atctatggtt tccttttttgt tccgtaactc ttcctcttac 120

ccacaaaaac ctgctttgat tgactccgaa accaatcaaa tcttgtcctt ttcccacttc 180

aaatctactg tcattaaagt ctctcacggt ttcttgaact taggtattaa gaagaacgac 240

tggttgatct acgctcctaa ttccatccac tttccagttt gtttcttggg tatcattgct 300

tctggtgcca ttgctaccac ttctaaccct ttatacactg tttctgagtt atctaagcaa 360

gttaaagatt ctaacccaaa attgattatc actgtcccac aattattaga aaaggtcaag 420

ggtttcaatt taccaaccat tttaatcggt ccagactccg aacaagagtc ttcttccgat 480

aaagttatga cttttaacga cttagttaac ttgggtggtt cttctggttc tgagttccca 540

atcgtcgatg atttcaagca atctgacacc gccgctttat tgtattcctc tggtactact 600

ggtatgtcta agggttggtt gactcacaaa aactttatcg cttcctcttt gatggttacc 660

atggaacaag acttggttgg tgaaatggat aacgtcttct tgtgtttttt accaatgttc 720

catgttttcg gtttagctat cattacttac gctcaattac aaagaggtaa cactgtcatc 780

tctgctcgtt ttgacttaga aaagatgttg aaagacgttg aaaagtacgt tactcacttg 840

tggtggcctc ctgttatttt agctttgtct aagaattcta tggttaaatt caacttgtcc 900

tctatcaagt acattggttc tggtgccgct ccattaggta aggacttgat ggaagaatgt 960

tctaaatggc cttacggtat cgtcgctcaa ggttacggta tgactgaaac ttgtggtatc 1020

gtttctatgg aagacatcag aggtggtaag cgtaactccg gttctgctgg tatgttggct 1080

tccggtgttg aagcccaaat tgtttctgtc gatactttga aacctttgcc acctaaccaa 1140

ttaggtgaaa tttgggttaa aggtcctaac atgatgcaag gttacttcaa taaccctcaa 1200

gctactaagt taactattga taagaagggt tgggttcata ctggtgattt gggttacttc 1260

gatgaagatg gtcatttgta ctgggataga atcaaagaat taattaagta taaaggtttc 1320

caagttgccc cagctgaatt ggaaggtttg ttggtttctc atcctgaaat tttagatgct 1380

tggattcctt tcccagacgc tgaagccggt gaagttccag ttgcttactg gagatcccct 1440

aactcttcct tgactgaaaa cgacgtcaag aagttcatcg ctggtcaagt tgcttccttt 1500

aagagattaa gaaaagtcac cttcatcaac tccgttccaa agtctgcttc cggtaagatt 1560

ttg 1563

<210> 151
<211> 518
<212> PRT
<213> Artificial Sequence

<220>
<223> Truncated cannabidiolic acid synthase Cs_CBDASt28

<400> 151

```
Met Ser Asn Pro Arg Glu Asn Phe Leu Lys Cys Phe Ser Gln Tyr Ile
1               5                  10                  15


Pro Asn Asn Ala Thr Asn Leu Lys Leu Val Tyr Thr Gln Asn Asn Pro
            20                  25                  30


Leu Tyr Met Ser Val Leu Asn Ser Thr Ile His Asn Leu Arg Phe Thr
        35                  40                  45


Ser Asp Thr Thr Pro Lys Pro Leu Val Ile Val Thr Pro Ser His Val
    50                  55                  60


Ser His Ile Gln Gly Thr Ile Leu Cys Ser Lys Lys Val Gly Leu Gln
65                  70                  75                  80


Ile Arg Thr Arg Ser Gly Gly His Asp Ser Glu Gly Met Ser Tyr Ile
                85                  90                  95


Ser Gln Val Pro Phe Val Ile Val Asp Leu Arg Asn Met Arg Ser Ile
            100                 105                 110


Lys Ile Asp Val His Ser Gln Thr Ala Trp Val Glu Ala Gly Ala Thr
        115                 120                 125


Leu Gly Glu Val Tyr Tyr Trp Val Asn Glu Lys Asn Glu Asn Leu Ser
    130                 135                 140


Leu Ala Ala Gly Tyr Cys Pro Thr Val Cys Ala Gly Gly His Phe Gly
145                 150                 155                 160


Gly Gly Gly Tyr Gly Pro Leu Met Arg Asn Tyr Gly Leu Ala Ala Asp
                165                 170                 175
```

```
Asn Ile Ile Asp Ala His Leu Val Asn Val His Gly Lys Val Leu Asp
            180                 185                 190

Arg Lys Ser Met Gly Glu Asp Leu Phe Trp Ala Leu Arg Gly Gly Gly
        195                 200                 205

Ala Glu Ser Phe Gly Ile Ile Val Ala Trp Lys Ile Arg Leu Val Ala
    210                 215                 220

Val Pro Lys Ser Thr Met Phe Ser Val Lys Lys Ile Met Glu Ile His
225                 230                 235                 240

Glu Leu Val Lys Leu Val Asn Lys Trp Gln Asn Ile Ala Tyr Lys Tyr
                245                 250                 255

Asp Lys Asp Leu Leu Leu Met Thr His Phe Ile Thr Arg Asn Ile Thr
            260                 265                 270

Asp Asn Gln Gly Lys Asn Lys Thr Ala Ile His Thr Tyr Phe Ser Ser
        275                 280                 285

Val Phe Leu Gly Gly Val Asp Ser Leu Val Asp Leu Met Asn Lys Ser
    290                 295                 300

Phe Pro Glu Leu Gly Ile Lys Lys Thr Asp Cys Arg Gln Leu Ser Trp
305                 310                 315                 320

Ile Asp Thr Ile Ile Phe Tyr Ser Gly Val Val Asn Tyr Asp Thr Asp
                325                 330                 335

Asn Phe Asn Lys Glu Ile Leu Leu Asp Arg Ser Ala Gly Gln Asn Gly
            340                 345                 350

Ala Phe Lys Ile Lys Leu Asp Tyr Val Lys Lys Pro Ile Pro Glu Ser
        355                 360                 365

Val Phe Val Gln Ile Leu Glu Lys Leu Tyr Glu Glu Asp Ile Gly Ala
    370                 375                 380
```

Gly Met Tyr Ala Leu Tyr Pro Tyr Gly Gly Ile Met Asp Glu Ile Ser
385 390 395 400

Glu Ser Ala Ile Pro Phe Pro His Arg Ala Gly Ile Leu Tyr Glu Leu
405 410 415

Trp Tyr Ile Cys Ser Trp Glu Lys Gln Glu Asp Asn Glu Lys His Leu
420 425 430

Asn Trp Ile Arg Asn Ile Tyr Asn Phe Met Thr Pro Tyr Val Ser Lys
435 440 445

Asn Pro Arg Leu Ala Tyr Leu Asn Tyr Arg Asp Leu Asp Ile Gly Ile
450 455 460

Asn Asp Pro Lys Asn Pro Asn Asn Tyr Thr Gln Ala Arg Ile Trp Gly
465 470 475 480

Glu Lys Tyr Phe Gly Lys Asn Phe Asp Arg Leu Val Lys Val Lys Thr
485 490 495

Leu Val Asp Pro Asn Asn Phe Phe Arg Asn Glu Gln Ser Ile Pro Pro
500 505 510

Leu Pro Arg His Arg His
515

<210> 152
<211> 1557
<212> DNA
<213> Artificial Sequence

<220>
<223> Truncated cannabidiolic acid synthase Cs_CBDASt28

<400> 152
atgtctaatc caagagagaa tttcttaaag tgtttttctc aatacatccc aaacaatgct 60

actaacttaa agttggttta cactcaaaat aacccattgt acatgtctgt cttgaactct 120

accattcaca atttgcgttt tacttctgac accaccccta agccattagt tattgttacc 180

ccatcccacg tctctcacat ccaaggtact attttgtgtt ctaaaaaggt tggtttgcaa 240

attagaacta gatctggtgg tcacgactcc gagggtatgt cttacatctc tcaagttcca 300

ttcgttattg tcgacttgcg taacatgcgt tccatcaaaa tcgatgttca ctcccaaact 360

gcttgggtcg aagccggtgc cactttaggt gaggtttatt actgggtcaa tgagaagaat 420

gagaatttgt ccttggctgc tggttattgt ccaaccgtct gtgctggtgg tcattttggt 480

ggtggtggtt acggtccatt aatgagaaac tatggtttgg ctgccgataa cattatcgac 540

gctcacttgg ttaatgtcca cggtaaggtc ttagatagaa aatccatggg tgaggacttg 600

ttctgggctt tgagaggtgg tggtgctgag tcctttggta tcatcgttgc ttggaaaatt 660

cgtttagttg ctgtcccaaa atctactatg ttttctgtta agaagatcat ggaaattcac 720

gagttggtta agttggttaa taagtggcaa aatattgcct acaagtatga caaagacttg 780

ttattgatga ctcacttcat cactagaaac atcaccgata accaaggtaa aaataaaact 840

gctatccata cctacttctc ctccgttttc ttgggtggtg tcgactcctt agttgatttg 900

atgaacaaat cttttcctga attaggtatc aagaagactg attgtcgtca attgtcctgg 960

attgatacca ttatctttta ctctggtgtc gtcaattacg acaccgataa tttcaataag 1020

gaaattttat tggacagatc tgccggtcaa aacggtgctt tcaagatcaa gttggactac 1080

gttaaaaaac caatcccaga atccgtcttt gtccaaattt tggagaagtt atacgaggaa 1140

gacatcggtg ctggtatgta tgccttatat ccatacggtg gtattatgga tgaaatttcc 1200

gaatctgcta tcccatttcc acatcgtgct ggtattttgt atgaattatg gtacatttgt 1260

tcctgggaaa agcaagaaga taacgagaag cacttgaatt ggatcagaaa tatctacaat 1320

ttcatgactc cttacgtttc taagaatcct cgtttggctt acttgaacta cagagatttg 1380

gacatcggta ttaatgaccc aaagaaccca ataactata ctcaagctag aatttggggt 1440

gaaaagtact tcggtaaaaa ctttgacaga ttggttaagg ttaagacttt agttgatcca 1500

aataacttct tcagaaatga acaatccatc ccaccattgc ctagacacag acactaa 1557

<210> 153
<211> 519
<212> PRT
<213> Artificial Sequence

<220>
<223> Truncated tetrahydrocannabinolic acid synthase Cs_THCASt28

<400> 153

```
Met Ser Asn Pro Arg Glu Asn Phe Leu Lys Cys Phe Ser Lys His Ile
1               5                  10                  15


Pro Asn Asn Val Ala Asn Pro Lys Leu Val Tyr Thr Gln His Asp Gln
            20                  25                  30


Leu Tyr Met Ser Ile Leu Asn Ser Thr Ile Gln Asn Leu Arg Phe Ile
        35                  40                  45


Ser Asp Thr Thr Pro Lys Pro Leu Val Ile Val Thr Pro Ser Asn Asn
    50                  55                  60


Ser His Ile Gln Ala Thr Ile Leu Cys Ser Lys Lys Val Gly Leu Gln
65                  70                  75                  80


Ile Arg Thr Arg Ser Gly Gly His Asp Ala Glu Gly Met Ser Tyr Ile
                85                  90                  95


Ser Gln Val Pro Phe Val Val Val Asp Leu Arg Asn Met His Ser Ile
            100                 105                 110


Lys Ile Asp Val His Ser Gln Thr Ala Trp Val Glu Ala Gly Ala Thr
        115                 120                 125


Leu Gly Glu Val Tyr Tyr Trp Ile Asn Glu Lys Asn Glu Asn Leu Ser
    130                 135                 140


Phe Pro Gly Gly Tyr Cys Pro Thr Val Gly Val Gly Gly His Phe Ser
145                 150                 155                 160


Gly Gly Gly Tyr Gly Ala Leu Met Arg Asn Tyr Gly Leu Ala Ala Asp
                165                 170                 175


Asn Ile Ile Asp Ala His Leu Val Asn Val Asp Gly Lys Val Leu Asp
            180                 185                 190
```

```
Arg Lys Ser Met Gly Glu Asp Leu Phe Trp Ala Ile Arg Gly Gly Gly
        195                 200                 205

Gly Glu Asn Phe Gly Ile Ile Ala Ala Trp Lys Ile Lys Leu Val Ala
    210                 215                 220

Val Pro Ser Lys Ser Thr Ile Phe Ser Val Lys Lys Asn Met Glu Ile
225                 230                 235                 240

His Gly Leu Val Lys Leu Phe Asn Lys Trp Gln Asn Ile Ala Tyr Lys
                245                 250                 255

Tyr Asp Lys Asp Leu Val Leu Met Thr His Phe Ile Thr Lys Asn Ile
            260                 265                 270

Thr Asp Asn His Gly Lys Asn Lys Thr Thr Val His Gly Tyr Phe Ser
        275                 280                 285

Ser Ile Phe His Gly Gly Val Asp Ser Leu Val Asp Leu Met Asn Lys
    290                 295                 300

Ser Phe Pro Glu Leu Gly Ile Lys Lys Thr Asp Cys Lys Glu Phe Ser
305                 310                 315                 320

Trp Ile Asp Thr Thr Ile Phe Tyr Ser Gly Val Val Asn Phe Asn Thr
                325                 330                 335

Ala Asn Phe Lys Lys Glu Ile Leu Leu Asp Arg Ser Ala Gly Lys Lys
            340                 345                 350

Thr Ala Phe Ser Ile Lys Leu Asp Tyr Val Lys Lys Pro Ile Pro Glu
        355                 360                 365

Thr Ala Met Val Lys Ile Leu Glu Lys Leu Tyr Glu Glu Asp Val Gly
    370                 375                 380

Ala Gly Met Tyr Val Leu Tyr Pro Tyr Gly Gly Ile Met Glu Glu Ile
385                 390                 395                 400
```

Ser Glu Ser Ala Ile Pro Phe Pro His Arg Ala Gly Ile Met Tyr Glu
405 410 415

Leu Trp Tyr Thr Ala Ser Trp Glu Lys Gln Glu Asp Asn Glu Lys His
420 425 430

Ile Asn Trp Val Arg Ser Val Tyr Asn Phe Thr Thr Pro Tyr Val Ser
435 440 445

Gln Asn Pro Arg Leu Ala Tyr Leu Asn Tyr Arg Asp Leu Asp Leu Gly
450 455 460

Lys Thr Asn His Ala Ser Pro Asn Asn Tyr Thr Gln Ala Arg Ile Trp
465 470 475 480

Gly Glu Lys Tyr Phe Gly Lys Asn Phe Asn Arg Leu Val Lys Val Lys
485 490 495

Thr Lys Val Asp Pro Asn Asn Phe Phe Arg Asn Glu Gln Ser Ile Pro
500 505 510

Pro Leu Pro Pro His His His
515

<210> 154
<211> 1560
<212> DNA
<213> Artificial Sequence

<220>
<223> Truncated tetrahydrocannabinolic acid synthase Cs_THCASt28

<400> 154
atgtctaacc ctcgtgagaa cttcttgaaa tgtttctcca aacatatccc aaacaatgtc 60

gctaacccta agttagttta cactcaacat gatcaattat atatgtctat cttgaactct 120

accatccaaa acttgagatt catctccgat accaccccaa aaccattggt tattgttacc 180

ccatccaaca attctcatat tcaagctacc attttgtgct ccaaaaaggt cggtttgcaa 240

atccgtacta gatctggtgg tcacgatgct gaaggtatgt cttacatttc ccaagtccca 300

ttcgttgttg tcgatttaag aaatatgcac tctatcaaaa tcgacgttca ctctcaaact 360

gcttgggttg aagccggtgc cactttaggt gaggtttact actggattaa cgaaaagaat 420

gaaaacttat cctttccagg tggttactgt ccaactgttg gtgttggtgg tcacttctct 480

ggtggtggtt atggtgcctt gatgagaaac tacggtttag ctgctgataa tattatcgac 540

gctcacttgg ttaatgtcga cggtaaggtt ttggacagaa aatccatggg tgaagattta 600

ttctgggcca ttagaggtgg tggtggtgaa aacttcggta tcattgctgc ttggaaaatt 660

aaattggtcg ctgtcccatc caagtctact attttctccg tcaagaaaaa catggaaatt 720

catggtttgg ttaaattatt caacaagtgg caaaacattg cttacaaata cgacaaagac 780

ttagttttga tgacccactt cattactaaa aacattaccg acaaccatgg taaaaataaa 840

actactgttc acggttactt ctcttccatt tttcatggtg gtgtcgactc cttggtcgat 900

ttaatgaaca aatctttccc tgagttgggt atcaagaaga ccgactgtaa agaattctct 960

tggatcgaca ctactatttt ctactctggt gtcgttaact tcaacaccgc taatttcaag 1020

aaggaaattt tattagatag atccgctggt aaaaagaccg ctttctctat caaattagac 1080

tacgttaaaa aaccaatccc agaaaccgct atggtcaaaa tcttggaaaa attatatgaa 1140

gaagacgttg gtgccggtat gtacgtctta tatccatatg gtggtattat ggaagagatc 1200

tctgaatccg ctatcccttt tccacacaga gccggtatta tgtacgaatt atggtacact 1260

gcttcctggg agaaacaaga agataatgaa aagcacatta actgggttag atctgtttac 1320

aacttcacta ctccatacgt ctctcaaaac ccaagattag cctacttaaa ctaccgtgat 1380

ttggatttag gtaaaactaa tcacgcttcc ccaaacaact acacccaagc tagaatttgg 1440

ggtgagaagt actttggtaa gaacttcaac cgtttagtca aggtcaagac taaagttgat 1500

ccaaacaatt ttttcagaaa cgaacaatct atcccacctt taccaccaca ccaccattag 1560

<210> 155
<211> 545
<212> PRT
<213> Cannabis sativa

<400> 155

Met Asn Cys Ser Ala Phe Ser Phe Trp Phe Val Cys Lys Ile Ile Phe

```
1               5                   10                  15

Phe Phe Leu Ser Phe His Ile Gln Ile Ser Ile Ala Asn Pro Arg Glu
            20                  25                  30

Asn Phe Leu Lys Cys Phe Ser Lys His Ile Pro Asn Asn Val Ala Asn
        35                  40                  45

Pro Lys Leu Val Tyr Thr Gln His Asp Gln Leu Tyr Met Ser Ile Leu
    50                  55                  60

Asn Ser Thr Ile Gln Asn Leu Arg Phe Ile Ser Asp Thr Thr Pro Lys
65                  70                  75                  80

Pro Leu Val Ile Val Thr Pro Ser Asn Asn Ser His Ile Gln Ala Thr
                85                  90                  95

Ile Leu Cys Ser Lys Lys Val Gly Leu Gln Ile Arg Thr Arg Ser Gly
            100                 105                 110

Gly His Asp Ala Glu Gly Met Ser Tyr Ile Ser Gln Val Pro Phe Val
        115                 120                 125

Val Val Asp Leu Arg Asn Met His Ser Ile Lys Ile Asp Val His Ser
    130                 135                 140

Gln Thr Ala Trp Val Glu Ala Gly Ala Thr Leu Gly Glu Val Tyr Tyr
145                 150                 155                 160

Trp Ile Asn Glu Lys Asn Glu Asn Leu Ser Phe Pro Gly Gly Tyr Cys
                165                 170                 175

Pro Thr Val Gly Val Gly Gly His Phe Ser Gly Gly Gly Tyr Gly Ala
            180                 185                 190

Leu Met Arg Asn Tyr Gly Leu Ala Ala Asp Asn Ile Ile Asp Ala His
        195                 200                 205

Leu Val Asn Val Asp Gly Lys Val Leu Asp Arg Lys Ser Met Gly Glu
```

210 215 220

Asp Leu Phe Trp Ala Ile Arg Gly Gly Gly Gly Glu Asn Phe Gly Ile
225 230 235 240

Ile Ala Ala Trp Lys Ile Lys Leu Val Ala Val Pro Ser Lys Ser Thr
245 250 255

Ile Phe Ser Val Lys Lys Asn Met Glu Ile His Gly Leu Val Lys Leu
260 265 270

Phe Asn Lys Trp Gln Asn Ile Ala Tyr Lys Tyr Asp Lys Asp Leu Val
275 280 285

Leu Met Thr His Phe Ile Thr Lys Asn Ile Thr Asp Asn His Gly Lys
290 295 300

Asn Lys Thr Thr Val His Gly Tyr Phe Ser Ser Ile Phe His Gly Gly
305 310 315 320

Val Asp Ser Leu Val Asp Leu Met Asn Lys Ser Phe Pro Glu Leu Gly
325 330 335

Ile Lys Lys Thr Asp Cys Lys Glu Phe Ser Trp Ile Asp Thr Thr Ile
340 345 350

Phe Tyr Ser Gly Val Val Asn Phe Asn Thr Ala Asn Phe Lys Lys Glu
355 360 365

Ile Leu Leu Asp Arg Ser Ala Gly Lys Lys Thr Ala Phe Ser Ile Lys
370 375 380

Leu Asp Tyr Val Lys Lys Pro Ile Pro Glu Thr Ala Met Val Lys Ile
385 390 395 400

Leu Glu Lys Leu Tyr Glu Glu Asp Val Gly Ala Gly Met Tyr Val Leu
405 410 415

Tyr Pro Tyr Gly Gly Ile Met Glu Glu Ile Ser Glu Ser Ala Ile Pro

420 425 430

Phe Pro His Arg Ala Gly Ile Met Tyr Glu Leu Trp Tyr Thr Ala Ser
435 440 445

Trp Glu Lys Gln Glu Asp Asn Glu Lys His Ile Asn Trp Val Arg Ser
450 455 460

Val Tyr Asn Phe Thr Thr Pro Tyr Val Ser Gln Asn Pro Arg Leu Ala
465 470 475 480

Tyr Leu Asn Tyr Arg Asp Leu Asp Leu Gly Lys Thr Asn His Ala Ser
485 490 495

Pro Asn Asn Tyr Thr Gln Ala Arg Ile Trp Gly Glu Lys Tyr Phe Gly
500 505 510

Lys Asn Phe Asn Arg Leu Val Lys Val Lys Thr Lys Val Asp Pro Asn
515 520 525

Asn Phe Phe Arg Asn Glu Gln Ser Ile Pro Pro Leu Pro Pro His His
530 535 540

His
545

<210> 156
<211> 1638
<212> DNA
<213> Cannabis sativa

<400> 156
atgaattgtt ctgctttctc tttctggttc gtttgtaaga tcatcttttt cttcttatct 60

ttccatattc aaatctctat cgctaaccct cgtgagaact tcttgaaatg tttctccaaa 120

catatcccaa acaatgtcgc taaccctaag ttagtttaca ctcaacatga tcaattatat 180

atgtctatct tgaactctac catccaaaac ttgagattca tctccgatac caccccaaaa 240

ccattggtta ttgttacccc atccaacaat tctcatattc aagctaccat tttgtgctcc 300

aaaaaggtcg gtttgcaaat ccgtactaga tctggtggtc acgatgctga aggtatgtct 360

tacatttccc aagtcccatt cgttgttgtc gatttaagaa atatgcactc tatcaaaatc 420

gacgttcact ctcaaactgc ttgggttgaa gccggtgcca ctttaggtga ggtttactac 480

tggattaacg aaaagaatga aaacttatcc tttccaggtg gttactgtcc aactgttggt 540

gttggtggtc acttctctgg tggtggttat ggtgccttga tgagaaacta cggtttagct 600

gctgataata ttatcgacgc tcacttggtt aatgtcgacg gtaaggtttt ggacagaaaa 660

tccatgggtg aagatttatt ctgggccatt agaggtggtg gtggtgaaaa cttcggtatc 720

attgctgctt ggaaaattaa attggtcgct gtcccatcca agtctactat tttctccgtc 780

aagaaaaaca tggaaattca tggtttggtt aaattattca acaagtggca aaacattgct 840

tacaaatacg acaaagactt agttttgatg acccacttca ttactaaaaa cattaccgac 900

aaccatggta aaaataaaac tactgttcac ggttacttct cttccatttt tcatggtggt 960

gtcgactcct tggtcgattt aatgaacaaa tctttccctg agttgggtat caagaagacc 1020

gactgtaaag aattctcttg gatcgacact actattttct actctggtgt cgttaacttc 1080

aacaccgcta atttcaagaa ggaaatttta ttagatagat ccgctggtaa aaagaccgct 1140

ttctctatca aattagacta cgttaaaaaa ccaatcccag aaaccgctat ggtcaaaatc 1200

ttggaaaaat tatatgaaga agacgttggt gccggtatgt acgtcttata tccatatggt 1260

ggtattatgg aagagatctc tgaatccgct atccctttc cacacagagc cggtattatg 1320

tacgaattat ggtacactgc ttcctgggag aaacaagaag ataatgaaaa gcacattaac 1380

tgggttagat ctgtttacaa cttcactact ccatacgtct ctcaaaaccc aagattagcc 1440

tacttaaact accgtgattt ggatttaggt aaaactaatc acgcttcccc aaacaactac 1500

acccaagcta gaatttgggg tgagaagtac tttggtaaga acttcaaccg tttagtcaag 1560

gtcaagacta aagttgatcc aaacaatttt ttcagaaacg aacaatctat cccaccttta 1620

ccaccacacc accattag 1638

<210> 157
<211> 1197
<212> DNA
<213> Artificial Sequence

<220>

<223> Artificial Erg10p: acetoacetyl CoA thiolase nucleotide sequence

<400> 157
atgtcccaaa atgtttacat tgtttctact gctagaactc ctatcggttc cttccaaggt 60

tccttatctt ccaaaactgc cgtcgaattg ggtgccgttg ccttgaaagg tgctttagct 120

aaagttccag agttagacgc ttccaaagat ttcgatgaaa ttatcttcgg taacgtttta 180

tccgctaact tgggtcaagc tccagccaga caagttgcct tggctgccgg tttgtctaat 240

cacatcgttg cttctactgt caacaaagtt tgtgcctctg ctatgaaagc tatcatttta 300

ggtgcccaat ctattaaatg tggtaatgct gacgttgttg tcgctggtgg ttgtgagtcc 360

atgaccaacg ccccttacta catgccagcc gccagagccg gtgccaaatt cggtcaaact 420

gttttggttg acggtgttga aagagatggt ttgaacgatg cctatgacgg tttggctatg 480

ggtgttcacg ctgaaaagtg tgctagagac tgggacatta ccagagaaca acaagataat 540

ttcgctattg aatcttacca aaagtcccaa aaatctcaaa aggaaggtaa gtttgacaat 600

gaaatcgttc cagttactat caagggtttt cgtggtaagc ctgatactca agtcaccaag 660

gatgaagaac cagcccgttt acacgtcgaa aagttgagat ctgccagaac cgttttccaa 720

aaagaaaacg gtaccgttac tgctgccaat gcttctccaa tcaacgatgg tgccgctgct 780

gttattttag tctctgagaa ggttttgaag gagaaaaatt tgaagccttt agccatcatt 840

aagggttggg gtgaagctgc tcaccaacca gctgatttca cttgggcccc ttctttagct 900

gtcccaaagg ctttaaaaca cgctggtatt gaagatatca actctgttga ctacttcgaa 960

ttcaatgaag ctttctctgt cgtcggtttg gtcaatacca aaatcttgaa gttggatcct 1020

tctaaggtta acgtttacgg tggtgctgtc gccttaggtc acccttttagg ttgttctggt 1080

gctagagttg ttgtcacctt gttgtccatt ttacaacaag aaggtggtaa gatcggtgtt 1140

gctgctatct gtaacggtgg tggtggtgct tcttccattg tcatcgaaaa gatctag 1197

<210> 158
<211> 1191
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial mevalonate pyrophosphate decarboxylase (Sc_ERG19)
nucleotide sequence

<400> 158
atgactgtct acactgcctc cgttactgcc cctgtcaaca ttgccacctt gaagtattgg 60

ggtaaaagag atactaaatt gaacttacca actaactcct ccatttctgt cactttgtct 120

caagatgatt tgagaacctt gacttccgct gccaccgccc ctgaatttga gagagatact 180

ttgtggttaa atggtgaacc tcattctatt gacaacgaaa gaacccaaaa ctgtttacgt 240

gacttgagac aattgcgtaa ggaaatggaa tctaaagacg cttctttacc taccttgtct 300

caatggaaat tgcatatcgt ttctgaaaat aacttcccta ctgctgccgg tttggcttcc 360

tccgctgctg gttttgctgc tttagtttct gccatcgcca aattatatca attgccacaa 420

tccacttccg aaatctctag aatcgctaga aaaggttccg gttctgcttg tagatccttg 480

ttcggtggtt acgttgcttg ggaaatgggt aaagctgaag acggtcatga ttctatggcc 540

gttcaaattg ccgactcctc cgattggcct caaatgaaag cttgtgtctt ggttgtctcc 600

gatatcaaaa aggatgtctc ttctactcaa ggtatgcaat taactgttgc cacttccgaa 660

ttgttcaaag agcgtatcga acacgttgtt ccaaagagat ttgaagttat gagaaaagct 720

atcgtcgaaa aggacttcgc tacctttgcc aaggagacta tgatggattc taactccttc 780

cacgctactt gtttggattc ctttccacct attttctaca tgaatgacac ctccaaacgt 840

attatctctt ggtgtcacac cattaaccaa ttttatggtg aaactatcgt cgcttacact 900

ttcgatgccg gtccaaacgc tgtcttgtac tatttggctg aaaacgaatc caagttattt 960

gcttttatct ataagttgtt cggttccgtc cctggttggg acaagaaatt caccactgaa 1020

caattggaag ctttcaacca ccaattcgaa tcttccaatt tcactgctag agaattagat 1080

ttggaattac aaaaggatgt cgctagagtc atcttaactc aagttggttc cggtccacaa 1140

gaaactaacg aatctttgat tgatgctaaa actggtttgc ctaaagaata a 1191

<210> 159
<211> 867
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial isopentenyl pyrophosphate isomerase Sc_IDI1 nucleotide sequence

<400> 159
atgaccgctg acaacaactc catgccacat ggtgctgtct cctcctacgc taaattagtc 60

caaaaccaaa cccctgaaga cattttagaa gagttccctg aaatcattcc attgcaacaa 120

agaccaaaca ctagatcctc cgagacttct aacgatgaat ctggtgaaac ttgtttttct 180

ggtcatgatg aagaacaaat caagttgatg aacgagaatt gtattgtttt ggactgggat 240

gacaacgcta tcggtgctgg taccaaaaag gtctgtcact tgatggaaaa catcgaaaag 300

ggtttgttgc atagagcctt ttccgtcttc atcttcaacg aacaaggtga gttattattg 360

caacaaagag ccactgaaaa aatcaccttt ccagatttat ggaccaacac ctgttgctcc 420

catccattgt gtattgatga tgaattgggt ttgaaaggta agttggacga caagattaaa 480

ggtgccatca ccgccgctgt tcgtaagtta gaccatgaat tgggtatccc tgaagacgaa 540

actaagacta gaggtaaatt ccatttcttg aatcgtattc actacatggc tccttccaat 600

gaaccatggg gtgaacacga aatcgactac attttgtttt acaaaattaa tgctaaagaa 660

aatttaaccg ttaacccaaa cgtcaacgag gttagagatt tcaagtgggt ctctccaaac 720

gatttgaaga ctatgttcgc tgacccatcc tacaagttca ctccatggtt taagatcatc 780

tgtgaaaact atttgtttaa ctggtgggag caattggacg acttatctga agttgaaaat 840

gatcgtcaaa ttcaccgtat gttgtaa 867

<210> 160
<211> 1356
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial phosphomevalonate kinase Sc_ERG8 nucleotide sequence

<400> 160
atgtccgagt taagagcctt ctccgctcct ggtaaagcct tattagctgg tggttactta 60

gtcttggata ctaaatatga agccttcgtc gtcggtttat ctgccagaat gcatgccgtc 120

gcccatccat acggttcctt gcaaggttct gacaagtttg aggtccgtgt caagtctaaa 180

caattcaaag atggtgaatg gttgtatcat atttctccaa aatccggttt cattccagtt 240

tctatcggtg gttctaagaa cccattcatc gaaaaagtca tcgctaacgt tttctcttac 300

ttcaagccta atatggatga ttattgcaat agaaatttat tcgttattga tatcttctcc 360

gatgacgcct atcattccca agaagactct gttaccgagc atagaggtaa cagaagatta 420

tctttccact ctcacagaat tgaagaagtt ccaaaaactg gtttaggttc ttctgctggt 480

ttagtcaccg ttttaaccac tgccttggct tctttctttg tttccgactt agaaaataac 540

gtcgacaagt atcgtgaagt catccacaac ttggcccaag ttgctcattg tcaagctcaa 600

ggtaagattg gttccggttt cgatgttgct gccgccgcct acggttccat cagatataga 660

agattccctc cagctttgat ttctaactta ccagatattg gttctgctac ttatggttcc 720

aagttggctc acttggttga cgaagaagat tggaacatta ccatcaagtc caatcacttg 780

ccatctggtt taactttgtg gatgggtgat atcaagaacg gttctgaaac tgtcaaattg 840

gtccaaaagg tcaaaaattg gtacgattcc catatgccag agtctttgaa gatctatact 900

gaattggacc acgctaactc tcgtttcatg gatggtttgt ctaagttgga cagattgcat 960

gaaactcacg acgactactc tgaccaaatt ttcgagtcct tggaaagaaa cgactgcact 1020

tgtcaaaagt atccagaaat caccgaggtt agagatgccg ttgctactat tagaagatcc 1080

ttcagaaaga ttaccaagga atctggtgct gatattgagc ctccagttca aacttctttg 1140

ttggatgatt gccaaacttt aaaaggtgtt ttaacttgtt taattcctgg tgctggtggt 1200

tacgacgcca tcgccgttat caccaaacaa gacgtcgact taagagccca aactgccaac 1260

gacaaaagat tctccaaggt tcaatggttg gacgtcactc aagctgattg gggtgttaga 1320

aaagaaaagg acccagagac ttacttggat aaatag 1356

<210> 161
<211> 1059
<212> DNA
<213> Artificial Sequence

<220>
<223> Mutant farnesyl pyrophosphate synthase (Erg20mut, F96W, N127W)

<400> 161
atggcttctg agaaggagat tcgtcgtgag agattcttga atgtttttcc taaattagtc 60

gaggaattga acgcttcttt gttggcttat ggtatgccta aggaagcttg tgattggtat 120

gctcactcct tgaattataa tactccaggt ggtaaattga accgtggttt gtctgttgtt 180

gacacttacg ctattttatc taacaagacc gtcgagcaat tgggtcaaga agagtatgaa 240

aaggtcgcta ttttaggttg gtgtattgaa ttgttgcaag cttactggtt ggttgccgat 300

gacatgatgg acaagtctat tactcgtcgt ggtcaacctt gctggtataa ggtcccagag 360

gttggtgaaa ttgctatctg ggacgctttc atgttggaag ctgctatcta taaattgttg 420

aaatcccact tcagaaacga gaaatactac attgacatca ccgagttgtt ccacgaagtc 480

actttccaaa ctgagttagg tcaattaatg gacttgatca ccgctccaga agacaaagtt 540

gacttgtcca agttttcctt gaaaaagcac tctttcatcg ttactttcaa gactgcttat 600

tactctttct acttaccagt tgccttggct atgtacgtcg ccggtatcac tgacgaaaag 660

gacttgaagc aagctcgtga cgttttgatt ccattaggtg aatatttcca aatccaagat 720

gactacttag actgttttgg taccctgaa caaatcggta agatcggtac tgatattcaa 780

gataacaagt gctcttgggt tatcaacaag gctttagagt tagcctccgc cgaacaacgt 840

aaaactttag atgaaaacta cggtaaaaaa gactctgttg ctgaggccaa gtgtaagaag 900

atttttaacg atttaaaaat cgaacaattg tatcacgaat atgaagagtc cattgctaag 960

gatttgaagg ctaaaatttc tcaagttgac gaatcccgtg gtttcaaagc tgacgttttg 1020

actgcttttt taaacaaggt ttacaagcgt tccaaataa 1059

<210> 162
<211> 1158
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial tetraketide synthase (TKS) nucleotide sequence

<400> 162
atgaaccatt taagagctga gggtccagct tccgtcttgg ctatcggtac tgctaatcca 60

gagaacattt tattacaaga tgagtttcca gattactatt tccgtgttac taagtccgag 120

catatgaccc aattgaaaga aaagttccgt aaaatctgtg ataaatctat gattagaaaa 180

agaaactgct ttttaaacga agaacacttg aagcaaaacc caagattagt tgaacacgag 240

atgcaaacct tggacgctag acaagatatg ttggttgtcg aggttcctaa attgggtaaa 300

gacgcctgtg ctaaagctat caaagagtgg ggtcaaccta agtccaagat cactcactta 360

atcttcactt ccgcttccac cactgacatg cctggtgctg attaccactg tgccaagttg 420

ttgggtttgt ctccttctgt caagagagtt atgatgtacc aattaggttg ttacggtggt 480

ggtactgtct taagaattgc taaggacatc gctgaaaaca acaaaggtgc tagagtttta 540

gccgtttgtt gtgacatcat ggcttgttta tttcgtggtc catctgaatc tgacttggag 600

ttgttggttg gtcaagctat ttttggtgat ggtgccgctg ccgtcatcgt tggtgctgag 660

ccagatgaat ccgttggtga aagaccaatt ttcgaattag tctctactgg tcaaactatt 720

ttgccaaact ccgagggtac tatcggtggt catattcgtg aagccggttt aatctttgat 780

ttgcacaaag acgttccaat gttgatctct aacaacatcg aaaagtgttt aattgaggct 840

tttactccaa ttggtatctc tgactggaac tctatcttct ggatcactca tccaggtggt 900

aaggctatct tggacaaggt tgaagaaaaa ttacatttaa agtccgataa attcgtcgat 960

tctcgtcatg ttttgtctga acacggtaac atgtcttcct ccactgtctt gtttgttatg 1020

gatgaattac gtaagagatc tttggaggag ggtaagtcta ctactggtga tggtttcgaa 1080

tggggtgttt tgttcggttt cggtcctggt ttgactgttg aacgtgttgt tgttagatct 1140

gttccaatta agtactag 1158

<210> 163
<211> 306
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial olivetolic acid cyclase (OAC) nucleotide sequence

<400> 163
atggccgtca aacacttgat cgtcttaaaa ttcaaggatg aaattactga agctcaaaaa 60

gaagagttct tcaaaaccta tgtcaattta gtcaacatta ttcctgctat gaaggacgtt 120

tactggggta aggatgtcac ccaaaagaac aaggaagaag gttacactca cattgttgaa 180

gtcactttcg aatctgttga aactatccaa gattatatta tccacccagc tcatgtcggt 240

tttggtgatg tttacagatc tttttgggaa aaattgttga tctttgacta tactccaaga 300

aaataa 306

<210> 164
<211> 2163

<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial acyl-activating enzyme Cs_AAE1_v1 nucleotide sequence

<400> 164
atgggtaaga attacaagtc cttagactct gttgttgctt ctgactttat tgctttaggt 60

attacttccg aagttgctga aaccttacac ggtagattgg ctgaaattgt ttgcaactac 120

ggtgctgcta cccctcaaac ttggattaac attgctaatc atattttgtc tccagatttg 180

ccattttctt tacaccaaat gttgttctac ggttgttaca aggatttcgg tcctgctcct 240

ccagcttgga ttcctgatcc agaaaaagtc aaatctacta acttgggtgc tttgttggaa 300

aagagaggta aggagttttt gggtgttaag tacaaggacc caatttcttc tttctctcac 360

ttccaagaat tctctgttag aaaccctgaa gtttactgga gaactgtttt gatggatgag 420

atgaagattt ctttttctaa ggacccagag tgtatcttaa gaagagacga cattaacaat 480

ccaggtggtt ctgagtggtt accaggtggt tacttgaact ctgccaaaaa ttgcttgaac 540

gttaactcta acaagaaatt gaatgacact atgattgtct ggagagatga gggtaacgat 600

gatttgcctt tgaataaatt gactttggat caattgagaa aaagagtctg gttggttggt 660

tacgctttgg aagaaatggg tttagaaaaa ggttgtgcta tcgccatcga tatgcctatg 720

cacgttgatg ctgttgttat ttatttggct attgttttag ctggttatgt tgttgtttcc 780

atcgccgact ccttctctgc tccagaaatc tccaccagat tgagattgtc taaagccaaa 840

gccattttca cccaagacca catcattaga ggtaagaagc gtattccatt gtattctcgt 900

gttgttgaag ctaaatctcc tatggctatc gtcatcccat gctctggttc taacatcggt 960

gctgaattaa gagacggtga tatttcttgg gactactttt tagaaagagc taaagaattc 1020

aaaaactgcg agtttactgc tagagaacaa cctgtcgacg cttatactaa tattttattc 1080

tcttctggta ctactggtga acctaaggct attccatgga cccaagctac tcctttgaaa 1140

gccgctgctg atggttggtc ccatttagac atcagaaaag gtgatgtcat cgtctggcca 1200

actaacttag gttggatgat gggtccatgg ttagtctacg cttctttgtt gaatggtgcc 1260

tctatcgcct tatataatgg ttccccttta gtctctggtt ttgctaaatt cgttcaagat 1320

gctaaggtta ccatgttagg tgttgtccct tctatcgtta gatcttggaa atctactaac 1380

tgtgtttctg gttacgactg gtccactatt cgttgtttct cttcttctgg tgaagcttcc 1440

aatgtcgatg agtacttatg gttaatgggt cgtgctaact acaagccagt catcgaaatg 1500

tgcggtggta ctgaaattgg tggtgctttt tccgctggtt ctttttttaca agcccaatcc 1560

ttgtcttcct tctcctctca atgtatgggt tgtactttat atatcttaga taagaatggt 1620

taccctatgc ctaaaaacaa gccaggtatt ggtgaattag ctttgggtcc tgttatgttt 1680

ggtgcttcta aaaccttgtt aaatggtaat catcacgacg tttacttcaa aggtatgcct 1740

actttgaacg gtgaggtttt gagacgtcat ggtgatattt tcgaattaac ttccaacggt 1800

tattatcacg ctcacggtag agctgatgat actatgaaca ttggtggtat taagatctct 1860

tccatcgaaa ttgagagagt ttgtaacgag gttgacgatc gtgttttcga aactactgct 1920

attggtgtcc ctcctttagg tggtggtcca gaacaattgg ttatcttttt cgtcttgaag 1980

gactccaacg acaccactat cgacttaaac caattaagat tgtctttcaa cttgggtttg 2040

caaaagaagt tgaatccatt atttaaggtt actcgtgtcg ttccattgtc ctccttgcca 2100

agaactgcta ccaacaagat tatgcgtaga gtcttgagac aacaattctc tcactttgag 2160

taa 2163

<210> 165
<211> 2163
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial acyl-activating enzyme Cs_AAE1_v2 nucleotide sequence

<400> 165
atgggtaaga actacaaatc cttagattcc gtcgtcgctt ctgatttcat cgctttgggt 60

attacttctg aagttgctga aaccttgcat ggtagattgg ctgaaattgt ctgtaactac 120

ggtgctgcta ccccacaaac ttggatcaac attgctaacc acatcttatc ccctgacttg 180

ccattctcct tacaccaaat gttgttctac ggttgttata aagatttcgg tccagctcct 240

cctgcttgga ttcctgaccc agagaaggtt aagtctacta atttaggtgc tttgttagag 300

aagagaggta aggaattttt aggtgttaag tataaagatc caatttcttc cttctctcac 360

ttccaagaat tttctgttag aaacccagaa gtttactgga gaactgtttt gatggatgaa 420

atgaagatct ctttttccaa ggacccagag tgtattttga gacgtgatga catcaacaat 480

ccaggtggtt ctgagtggtt accaggtggt tacttgaact ctgccaagaa ttgtttgaac 540

gttaactcta acaaaaagtt gaacgatacc atgattgttt ggagagacga aggtaacgat 600

gatttgccat tgaataagtt aaccttggat caattgagaa aaagagtctg gttagtcggt 660

tacgctttgg aagagatggg tttggaaaag ggttgtgcta tcgccatcga tatgccaatg 720

catgttgatg ctgttgttat ctatttggcc attgttttgg ctggttacgt tgttgtttcc 780

atcgctgact ccttctctgc tccagaaatt tctactagat taagattgtc taaagccaaa 840

gccattttca ctcaagacca tattattaga ggtaagaaaa gaattccatt gtattccaga 900

gttgttgaag ctaaatcccc aatggccatc gtcatcccat gctctggttc taatattggt 960

gccgaattga gagacggtga tatctcttgg gactactttt tggagcgtgc taaagaattt 1020

aaaaactgcg aattcaccgc cagagaacaa ccagttgacg cctacactaa cattttgttt 1080

tcttctggta ctactggtga acctaaggct attccatgga ctcaagctac tccattgaaa 1140

gccgccgccg atggttggtc ccacttagat attagaaagg gtgatgtcat cgtctggcct 1200

actaacttgg gttggatgat gggtccttgg ttggtttacg cttccttatt gaacggtgcc 1260

tctatcgctt tatataatgg ttccccttta gtttctggtt ttgctaaatt cgttcaagat 1320

gctaaggtta ctatgttggg tgtcgtccca tccattgtcc gttcctggaa gtctaccaat 1380

tgtgtttctg gttatgattg gtctactatt cgttgttttt cttcctctgg tgaagcttct 1440

aatgtcgatg aatatttgtg gttaatgggt agagctaact acaagccagt tattgaaatg 1500

tgtggtggta ctgaaattgg tggtgctttc tctgctggtt ccttttgca agctcaatcc 1560

ttgtcttctt tctcctccca atgtatgggt tgcactttat acatcttgga caagaatggt 1620

taccctatgc caaagaataa accaggtatt ggtgaattgg ctttgggtcc agtcatgttc 1680

ggtgcttcta agactttgtt gaacggtaac catcatgacg tctacttcaa gggtatgcct 1740

accttgaacg gtgaagtttt aagacgtcac ggtgacattt tcgaattgac ttccaacggt 1800

tattatcatg ctcacggtag agctgacgac actatgaaca tcggtggtat taagatctct 1860

tctatcgaaa ttgaaagagt ttgcaacgag gttgatgatc gtgtcttcga aaccactgct 1920

attggtgtcc ctcctttagg tggtggtcct gagcaattgg ttattttctt tgtcttaaag 1980

gattctaacg acaccactat tgacttaaat caattgagat tgtccttcaa tttgggtttg 2040

caaaagaagt tgaacccatt attcaaggtt actcgtgtcg ttcctttgtc ctctttgcca 2100

agaaccgcta ccaataaaat tatgagacgt gttttgcgtc aacaattctc tcactttgaa 2160

taa 2163

<210> 166
<211> 1602
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial acyl-activating enzyme Cs_AAE3 nucleotide sequence

<400> 166
atggaaaaat ctggttatgg tagagacggt atctacagat ccttgcgtcc tccattacac 60

ttgccaaaca ataataactt atctatggtt tccttttttgt tccgtaactc ttcctcttac 120

ccacaaaaac ctgctttgat tgactccgaa accaatcaaa tcttgtcctt ttcccacttc 180

aaatctactg tcattaaagt ctctcacggt ttcttgaact taggtattaa gaagaacgac 240

tggttgatct acgctcctaa ttccatccac tttccagttt gtttcttggg tatcattgct 300

tctggtgcca ttgctaccac ttctaaccct ttatacactg tttctgagtt atctaagcaa 360

gttaaagatt ctaacccaaa attgattatc actgtcccac aattattaga aaaggtcaag 420

ggtttcaatt taccaaccat tttaatcggt ccagactccg aacaagagtc ttcttccgat 480

aaagttatga cttttaacga cttagttaac ttgggtggtt cttctggttc tgagttccca 540

atcgtcgatg atttcaagca atctgacacc gccgctttat tgtattcctc tggtactact 600

ggtatgtcta agggttggtt gactcacaaa aactttatcg cttcctcttt gatggttacc 660

atggaacaag acttggttgg tgaaatggat aacgtcttct tgtgtttttt accaatgttc 720

catgttttcg gtttagctat cattacttac gctcaattac aaagaggtaa cactgtcatc 780

tctgctcgtt ttgacttaga aaagatgttg aaagacgttg aaaagtacgt tactcacttg 840

tggtggcctc ctgttatttt agctttgtct aagaattcta tggttaaatt caacttgtcc 900

tctatcaagt acattggttc tggtgccgct ccattaggta aggacttgat ggaagaatgt 960

tctaaatggc cttacggtat cgtcgctcaa ggttacggta tgactgaaac ttgtggtatc 1020

gtttctatgg aagacatcag aggtggtaag cgtaactccg gttctgctgg tatgttggct 1080

tccggtgttg aagcccaaat tgtttctgtc gatactttga aacctttgcc acctaaccaa 1140

ttaggtgaaa tttgggttaa aggtcctaac atgatgcaag gttacttcaa taaccctcaa 1200

gctactaagt taactattga taagaagggt tgggttcata ctggtgattt gggttacttc 1260

gatgaagatg gtcatttgta ctgggataga atcaaagaat taattaagta taaaggtttc 1320

caagttgccc cagctgaatt ggaaggtttg ttggtttctc atcctgaaat tttagatgct 1380

tggattcctt tcccagacgc tgaagccggt gaagttccag ttgcttactg gagatcccct 1440

aactcttcct tgactgaaaa cgacgtcaag aagttcatcg ctggtcaagt tgcttccttt 1500

aagagattaa gaaaagtcac cttcatcaac tccgttccaa agtctgcttc cggtaagatt 1560

ttgagaagag aattaatcca aaaggttcgt tccaacatgt ag 1602

<210> 167
<211> 1635
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial cannabidiolic acid synthase (CBDAS) nucleotide sequence

<400> 167
atgaaatgtt ccacctttc tttctggttt gtttgtaaga tcatcttctt cttcttctcc 60

ttcaacatcc aaacttccat cgctaatcca agagagaatt tcttaaagtg tttttctcaa 120

tacatcccaa acaatgctac taacttaaag ttggtttaca ctcaaaataa cccattgtac 180

atgtctgtct tgaactctac cattcacaat ttgcgtttta cttctgacac caccctaag 240

ccattagtta ttgttacccc atcccacgtc tctcacatcc aaggtactat tttgtgttct 300

aaaaaggttg gtttgcaaat tagaactaga tctggtggtc acgactccga gggtatgtct 360

tacatctctc aagttccatt cgttattgtc gacttgcgta acatgcgttc catcaaaatc 420

gatgttcact cccaaactgc ttgggtcgaa gccggtgcca ctttaggtga ggtttattac 480

tgggtcaatg agaagaatga gaatttgtcc ttggctgctg gttattgtcc aaccgtctgt 540

gctggtggtc attttggtgg tggtggttac ggtccattaa tgagaaacta tggtttggct 600

gccgataaca ttatcgacgc tcacttggtt aatgtccacg gtaaggtctt agatagaaaa 660

tccatgggtg aggacttgtt ctgggctttg agaggtggtg gtgctgagtc ctttggtatc 720

atcgttgctt ggaaaattcg tttagttgct gtcccaaaat ctactatgtt ttctgttaag 780

aagatcatgg aaattcacga gttggttaag ttggttaata agtggcaaaa tattgcctac 840

aagtatgaca aagacttgtt attgatgact cacttcatca ctagaaacat caccgataac 900

caaggtaaaa ataaaactgc tatccatacc tacttctcct ccgttttctt gggtggtgtc 960

gactccttag ttgatttgat gaacaaatct tttcctgaat taggtatcaa gaagactgat 1020

tgtcgtcaat tgtcctggat tgataccatt atcttttact ctggtgtcgt caattacgac 1080

accgataatt tcaataagga aattttattg gacagatctg ccggtcaaaa cggtgctttc 1140

aagatcaagt tggactacgt taaaaaacca atcccagaat ccgtctttgt ccaaattttg 1200

gagaagttat acgaggaaga catcggtgct ggtatgtatg ccttatatcc atacggtggt 1260

attatggatg aaatttccga atctgctatc ccatttccac atcgtgctgg tattttgtat 1320

gaattatggt acatttgttc ctgggaaaag caagaagata acgagaagca cttgaattgg 1380

atcagaaata tctacaattt catgactcct tacgtttcta agaatcctcg tttggcttac 1440

ttgaactaca gagatttgga catcggtatt aatgacccaa agaacccaaa taactatact 1500

caagctagaa tttggggtga aaagtacttc ggtaaaaact ttgacagatt ggttaaggtt 1560

aagactttag ttgatccaaa taacttcttc agaaatgaac aatccatccc accattgcct 1620

agacacagac actaa 1635

<210> 168
<211> 2235
<212> DNA
<213> Saccharomyces sp.

<400> 168
atggccgctc cagattatgc acttaccgat ttaattgaat cggatcctcg tttcgaaagt 60

ttgaagacaa gattagccgg ttacaccaaa ggctctgatg aatatattga agagctatac 120

tctcaattac cactgaccag ctaccccagg tacaaaacat ttttaaagaa acaggcggtt 180

gccatttcga atccggataa tgaagctggt tttagctcga tttataggag ttctctttct 240

tctgaaaatc tagtgagctg tgtggataaa aacttaagaa ctgcatacga tcacttcatg 300

ttttctgcaa ggagatggcc tcaacgtgac tgtttaggtt caaggccaat tgataaagcc 360

acaggcacct gggaggaaac attccgtttc gagtcgtact ccacggtatc taaaagatgt 420

cataatatcg gaagtggtat attgtctttg gtaaacacga aaaggaaacg tcctttggaa 480

gccaatgatt ttgttgttgc tatcttatca cacaacaacc ctgaatggat cctaacagat 540

ttggcctgtc aggcctattc tctaactaac acggctttgt acgaaacatt aggtccaaac 600

acctccgagt acatattgaa tttaaccgag gcccccattc tgatttttgc aaaatcaaat 660

atgtatcatg tattgaagat ggtgcctgat atgaaatttg ttaatacttt ggtttgtatg 720

gatgaattaa ctcatgacga gctccgtatg ctaaatgaat cgttgctacc cgttaagtgc 780

aactctctca atgaaaaaat cacatttttt tcattggagc aggtagaaca agttggttgc 840

tttaacaaaa ttcctgcaat tccacctacc ccagattcct tgtatactat ttcgtttact 900

tctggtacta caggtttacc taaaggtgtg gaaatgtctc acagaaacat tgcgtctggg 960

atagcatttg ctttttctac cttcagaata ccgccagata aaagaaacca acagttatat 1020

gatatgtgtt ttttgccatt ggctcatatt tttgaaagaa tggttattgc gtatgatcta 1080

gccatcgggt ttggaatagg cttcttacat aaaccagacc caactgtatt ggtagaggat 1140

ttgaagattt tgaaacctta cgcggttgcc ctggttccta gaatattaac acggtttgaa 1200

gccggtataa aaaatgcttt ggataaatcg actgtccaga ggaacgtagc aaatactata 1260

ttggattcta aatcggccag atttaccgca agaggtggtc cagataaatc gattatgaat 1320

tttctagttt atcatcgcgt attgattgat aaaatcagag actctttagg tttgtccaat 1380

aactcgttta taattaccgg atcagctccc atatctaaag ataccttact atttttaaga 1440

agcgccttgg atattggtat aagacagggc tacggcttaa ctgaaacttt tgctggtgtc 1500

tgtttaagcg aaccgtttga aaaagatgtc ggatcttgtg gtgccatagg tatttctgca 1560

gaatgtagat tgaagtctgt tccagaaatg ggttaccatg ccgacaagga tttaaaaggt 1620

gaactgcaaa ttcgtggccc acaggttttt gaaagatatt ttaaaaatcc gaatgaaact 1680

tcaaaagccg ttgaccaaga tggttggttt tccacgggag atgttgcatt tatcgatgca 1740

aaaggtcgca tcagcgtcat tgatcgagtc aagaactttt tcaagctagc acatggtgaa 1800

tatattgctc cagagaaaat cgaaaatatt tatttatcat catgccccta tatcacgcaa 1860

atatttgtct ttggagatcc tttgaagaca tttttagttg gcatcgttgg tgttgatgtt 1920

gatgcagcgc aaccgatttt agctgcaaag cacccagagg tgaaaacgtg gactaaggaa 1980

gtgctagtag aaaacttaaa tcgtaataaa aagctaagga aggaattttt aaacaaaatt 2040

aataaatgca tcgatgggct acaaggattt gaaaaattgc acaacatcaa agtcggactt 2100

gagcctttga ctctcgagga tgatgttgtg acgccaactt ttaaaataaa gcgtgccaaa 2160

gcatcaaaat tcttcaaaga tacattagac caactatacg ccgaaggttc actagtcaag 2220

acagaaaagc tttag 2235

<210> 169
<211> 744
<212> PRT
<213> Saccharomyces sp.

<400> 169

Met Ala Ala Pro Asp Tyr Ala Leu Thr Asp Leu Ile Glu Ser Asp Pro
1 5 10 15

Arg Phe Glu Ser Leu Lys Thr Arg Leu Ala Gly Tyr Thr Lys Gly Ser
20 25 30

Asp Glu Tyr Ile Glu Glu Leu Tyr Ser Gln Leu Pro Leu Thr Ser Tyr
35 40 45

Pro Arg Tyr Lys Thr Phe Leu Lys Lys Gln Ala Val Ala Ile Ser Asn
50 55 60

Pro Asp Asn Glu Ala Gly Phe Ser Ser Ile Tyr Arg Ser Ser Leu Ser
65 70 75 80

Ser Glu Asn Leu Val Ser Cys Val Asp Lys Asn Leu Arg Thr Ala Tyr
85 90 95

Asp His Phe Met Phe Ser Ala Arg Arg Trp Pro Gln Arg Asp Cys Leu
100 105 110

Gly Ser Arg Pro Ile Asp Lys Ala Thr Gly Thr Trp Glu Glu Thr Phe

115 120 125

Arg Phe Glu Ser Tyr Ser Thr Val Ser Lys Arg Cys His Asn Ile Gly
130 135 140

Ser Gly Ile Leu Ser Leu Val Asn Thr Lys Arg Lys Arg Pro Leu Glu
145 150 155 160

Ala Asn Asp Phe Val Val Ala Ile Leu Ser His Asn Asn Pro Glu Trp
165 170 175

Ile Leu Thr Asp Leu Ala Cys Gln Ala Tyr Ser Leu Thr Asn Thr Ala
180 185 190

Leu Tyr Glu Thr Leu Gly Pro Asn Thr Ser Glu Tyr Ile Leu Asn Leu
195 200 205

Thr Glu Ala Pro Ile Leu Ile Phe Ala Lys Ser Asn Met Tyr His Val
210 215 220

Leu Lys Met Val Pro Asp Met Lys Phe Val Asn Thr Leu Val Cys Met
225 230 235 240

Asp Glu Leu Thr His Asp Glu Leu Arg Met Leu Asn Glu Ser Leu Leu
245 250 255

Pro Val Lys Cys Asn Ser Leu Asn Glu Lys Ile Thr Phe Phe Ser Leu
260 265 270

Glu Gln Val Glu Gln Val Gly Cys Phe Asn Lys Ile Pro Ala Ile Pro
275 280 285

Pro Thr Pro Asp Ser Leu Tyr Thr Ile Ser Phe Thr Ser Gly Thr Thr
290 295 300

Gly Leu Pro Lys Gly Val Glu Met Ser His Arg Asn Ile Ala Ser Gly
305 310 315 320

Ile Ala Phe Ala Phe Ser Thr Phe Arg Ile Pro Pro Asp Lys Arg Asn

325 330 335

Gln Gln Leu Tyr Asp Met Cys Phe Leu Pro Leu Ala His Ile Phe Glu
340 345 350

Arg Met Val Ile Ala Tyr Asp Leu Ala Ile Gly Phe Gly Ile Gly Phe
355 360 365

Leu His Lys Pro Asp Pro Thr Val Leu Val Glu Asp Leu Lys Ile Leu
370 375 380

Lys Pro Tyr Ala Val Ala Leu Val Pro Arg Ile Leu Thr Arg Phe Glu
385 390 395 400

Ala Gly Ile Lys Asn Ala Leu Asp Lys Ser Thr Val Gln Arg Asn Val
405 410 415

Ala Asn Thr Ile Leu Asp Ser Lys Ser Ala Arg Phe Thr Ala Arg Gly
420 425 430

Gly Pro Asp Lys Ser Ile Met Asn Phe Leu Val Tyr His Arg Val Leu
435 440 445

Ile Asp Lys Ile Arg Asp Ser Leu Gly Leu Ser Asn Asn Ser Phe Ile
450 455 460

Ile Thr Gly Ser Ala Pro Ile Ser Lys Asp Thr Leu Leu Phe Leu Arg
465 470 475 480

Ser Ala Leu Asp Ile Gly Ile Arg Gln Gly Tyr Gly Leu Thr Glu Thr
485 490 495

Phe Ala Gly Val Cys Leu Ser Glu Pro Phe Glu Lys Asp Val Gly Ser
500 505 510

Cys Gly Ala Ile Gly Ile Ser Ala Glu Cys Arg Leu Lys Ser Val Pro
515 520 525

Glu Met Gly Tyr His Ala Asp Lys Asp Leu Lys Gly Glu Leu Gln Ile

530 535 540

Arg Gly Pro Gln Val Phe Glu Arg Tyr Phe Lys Asn Pro Asn Glu Thr
545 550 555 560

Ser Lys Ala Val Asp Gln Asp Gly Trp Phe Ser Thr Gly Asp Val Ala
565 570 575

Phe Ile Asp Ala Lys Gly Arg Ile Ser Val Ile Asp Arg Val Lys Asn
580 585 590

Phe Phe Lys Leu Ala His Gly Glu Tyr Ile Ala Pro Glu Lys Ile Glu
595 600 605

Asn Ile Tyr Leu Ser Ser Cys Pro Tyr Ile Thr Gln Ile Phe Val Phe
610 615 620

Gly Asp Pro Leu Lys Thr Phe Leu Val Gly Ile Val Gly Val Asp Val
625 630 635 640

Asp Ala Ala Gln Pro Ile Leu Ala Ala Lys His Pro Glu Val Lys Thr
645 650 655

Trp Thr Lys Glu Val Leu Val Glu Asn Leu Asn Arg Asn Lys Lys Leu
660 665 670

Arg Lys Glu Phe Leu Asn Lys Ile Asn Lys Cys Ile Asp Gly Leu Gln
675 680 685

Gly Phe Glu Lys Leu His Asn Ile Lys Val Gly Leu Glu Pro Leu Thr
690 695 700

Leu Glu Asp Asp Val Val Thr Pro Thr Phe Lys Ile Lys Arg Ala Lys
705 710 715 720

Ala Ser Lys Phe Phe Lys Asp Thr Leu Asp Gln Leu Tyr Ala Glu Gly
725 730 735

Ser Leu Val Lys Thr Glu Lys Leu

740

<210> 170
<211> 1113
<212> DNA
<213> Artificial Sequence

<220>
<223> MBPtag

<400> 170
atgaaaatcg aagagggtaa attggtcatc tggatcaatg gtgacaaagg ttacaacggt 60

ttggctgaag tcggtaaaaa attcgagaaa gacactggta ttaaggttac cgtcgaacac 120

ccagataagt tggaagaaaa atttccacaa gttgccgcta ctggtgatgg tccagacatc 180

attttctggg cccacgacag atttggtggt tatgctcaat ctggtttgtt agccgagatc 240

accccagaca aagcctttca agataaatta tacccattta cctgggatgc tgtccgttac 300

aacggtaagt tgatcgctta cccaatcgcc gttgaagctt tgtctttaat ctacaataaa 360

gacttattgc caaaccctcc aaagacctgg gaagaaattc ctgccttgga taaggaatta 420

aaggctaaag gtaaatctgc cttaatgttc aacttacaag agccttactt tacttggcca 480

ttgattgctg ctgatggtgg ttatgctttt aagtacgaaa atggtaaata cgacattaaa 540

gatgttggtg ttgacaatgc cggtgctaaa gccggtttaa ctttcttagt cgacttgatc 600

aagaacaagc acatgaatgc tgacactgat tattctatcg ctgaagccgc cttcaacaag 660

ggtgaaactg ctatgactat caatggtcct tgggcctggt ctaatattga cacctccaaa 720

gtcaactacg gtgttactgt cttaccaact ttcaaaggtc aaccttccaa gccatttgtc 780

ggtgttttgt ctgctggtat taacgctgcc tctccaaaca aagaattggc caaggaattt 840

ttggaaaact acttgttgac tgacgaaggt ttagaggctg ttaacaaaga caaaccattg 900

ggtgctgtcg ccttgaaatc ctacgaagaa gaattagcca aggatccaag aatcgccgct 960

accatggaaa atgctcaaaa aggtgaaatt atgccaaaca ttccacaaat gtccgctttt 1020

tggtacgctg ttagaactgc tgttattaat gctgcttctg gtagacaaac tgtcgatgaa 1080

gctttgaagg acgctcaaac cagaatcact aag 1113

<210> 171

<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> GS12 Linker

<400> 171
ggaggtggag gaggtggttc cggaggaggt ggttct 36

<210> 172
<211> 12
<212> PRT
<213> Artificial Sequence

<220>
<223> GS12 Linker

<400> 172

Gly Gly Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser
1 5 10

<210> 173
<211> 174
<212> DNA
<213> Artificial Sequence

<220>
<223> GB1 tag

<400> 173
atgtctgaca cttacaagtt gatcttgaac ggtaagactt tgaaaggtga aactactacc 60

gaagctgttg atgctgccac tgctgaaaag gtttttaagc aatacgccaa tgataacggt 120

gtcgacggtg aatggactta cgatgatgcc actaagactt ttaccgttac tgaa 174

<210> 174
<211> 58
<212> PRT
<213> Artificial Sequence

<220>
<223> GB1 tag

<400> 174

Met Ser Asp Thr Tyr Lys Leu Ile Leu Asn Gly Lys Thr Leu Lys Gly

1 5 10 15

Glu Thr Thr Thr Glu Ala Val Asp Ala Ala Thr Ala Glu Lys Val Phe
20 25 30

Lys Gln Tyr Ala Asn Asp Asn Gly Val Asp Gly Glu Trp Thr Tyr Asp
35 40 45

Asp Ala Thr Lys Thr Phe Thr Val Thr Glu
50 55

<210> 175
<211> 57
<212> DNA
<213> Artificial Sequence

<220>
<223> MFalpha1_1-19

<400> 175
atgagatttc cttcaatttt tactgcagtt ttattcgcag catcctccgc attagct 57

<210> 176
<211> 19
<212> PRT
<213> Artificial Sequence

<220>
<223> MFalpha1_1-19

<400> 176

Met Arg Phe Pro Ser Ile Phe Thr Ala Val Leu Phe Ala Ala Ser Ser
1 5 10 15

Ala Leu Ala

<210> 177
<211> 267
<212> DNA
<213> Artificial Sequence

<220>
<223> MFalpha1_1-89

<400> 177
atgagatttc cttcaatttt tactgcagtt ttattcgcag catcctccgc attagctgct 60

ccagtcaaca ctacaacaga agatgaaacg gcacaaattc cggctgaagc tgtcatcggt 120

tacttagatt tagaagggga tttcgatgtt gctgttttgc cattttccaa cagcacaaat 180

aacgggttat tgtttataaa tactactatt gccagcattg ctgctaaaga agaaggggta 240

tctttggata aaagagaggc tgaagct 267

<210> 178
<211> 89
<212> PRT
<213> Artificial Sequence

<220>
<223> MFalpha1_1-89

<400> 178

Met Arg Phe Pro Ser Ile Phe Thr Ala Val Leu Phe Ala Ala Ser Ser
1 5 10 15

Ala Leu Ala Ala Pro Val Asn Thr Thr Thr Glu Asp Glu Thr Ala Gln
20 25 30

Ile Pro Ala Glu Ala Val Ile Gly Tyr Leu Asp Leu Glu Gly Asp Phe
35 40 45

Asp Val Ala Val Leu Pro Phe Ser Asn Ser Thr Asn Asn Gly Leu Leu
50 55 60

Phe Ile Asn Thr Thr Ile Ala Ser Ile Ala Ala Lys Glu Glu Gly Val
65 70 75 80

Ser Leu Asp Lys Arg Glu Ala Glu Ala
85

<210> 179
<211> 714
<212> DNA
<213> Artificial Sequence

<220>
<223> DasherGFP

<400> 179
atgaccgcac taacagaagg agctaaacta ttcgaaaagg agattcctta cattacagaa 60

ttagagggtg atgtcgaagg aatgaaattc attatcaagg gcgagggtac tggtgacgct 120

actaccggta cgattaaagc aaagtacatc tgtacaacag gtgaccttcc tgttccgtgg 180

gctactctgg tgagcacttt gtcttatgga gttcaatgtt ttgctaaata cccttcgcac 240

attaaagact ttttcaaaag tgcaatgcct gagggctata ctcaggagag aacaatatct 300

ttcgaaggag atggtgtgta taagactagg gctatggtca cgtatgaaag aggatccatc 360

tacaatagag taactttaac tggtgaaaac ttcaaaaagg acggtcacat ccttagaaag 420

aatgttgcct ttcaatgccc accatccatc ttgtacattt tgccagacac agttaacaat 480

ggtatcagag ttgagtttaa ccaagcttat gacatagagg gtgtcaccga aaagttggtt 540

acaaaatgtt cacagatgaa tcgtcccctg gcaggatcag ctgccgtcca tatcccacgt 600

taccatcata tcacttatca taccaagctg tccaaagatc gtgatgagag aagggatcac 660

atgtgtttgg ttgaagtggt aaaggccgtg gatttggata cttaccaagg ttga 714

<210> 180
<211> 237
<212> PRT
<213> Artificial Sequence

<220>
<223> DasherGFP

<400> 180

Met Thr Ala Leu Thr Glu Gly Ala Lys Leu Phe Glu Lys Glu Ile Pro
1 5 10 15

Tyr Ile Thr Glu Leu Glu Gly Asp Val Glu Gly Met Lys Phe Ile Ile
20 25 30

Lys Gly Glu Gly Thr Gly Asp Ala Thr Thr Gly Thr Ile Lys Ala Lys
35 40 45

Tyr Ile Cys Thr Thr Gly Asp Leu Pro Val Pro Trp Ala Thr Leu Val

50 55 60

Ser Thr Leu Ser Tyr Gly Val Gln Cys Phe Ala Lys Tyr Pro Ser His
65 70 75 80

Ile Lys Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Thr Gln Glu
85 90 95

Arg Thr Ile Ser Phe Glu Gly Asp Gly Val Tyr Lys Thr Arg Ala Met
100 105 110

Val Thr Tyr Glu Arg Gly Ser Ile Tyr Asn Arg Val Thr Leu Thr Gly
115 120 125

Glu Asn Phe Lys Lys Asp Gly His Ile Leu Arg Lys Asn Val Ala Phe
130 135 140

Gln Cys Pro Pro Ser Ile Leu Tyr Ile Leu Pro Asp Thr Val Asn Asn
145 150 155 160

Gly Ile Arg Val Glu Phe Asn Gln Ala Tyr Asp Ile Glu Gly Val Thr
165 170 175

Glu Lys Leu Val Thr Lys Cys Ser Gln Met Asn Arg Pro Leu Ala Gly
180 185 190

Ser Ala Ala Val His Ile Pro Arg Tyr His His Ile Thr Tyr His Thr
195 200 205

Lys Leu Ser Lys Asp Arg Asp Glu Arg Arg Asp His Met Cys Leu Val
210 215 220

Glu Val Val Lys Ala Val Asp Leu Asp Thr Tyr Gln Gly
225 230 235

<210> 181
<211> 87
<212> DNA
<213> Artificial Sequence

<220>
<223> ER1 tag

<400> 181
atattagagc aacctctgaa atttgtgctt actgcggccg tcgtgctctt gacgacgtcg 60

gttctttgtt gtgtagtatt tacataa 87

<210> 182
<211> 28
<212> PRT
<213> Artificial Sequence

<220>
<223> ER1 tag

<400> 182

Ile Leu Glu Gln Pro Leu Lys Phe Val Leu Thr Ala Ala Val Val Leu
1 5 10 15

Leu Thr Thr Ser Val Leu Cys Cys Val Val Phe Thr
20 25

<210> 183
<211> 96
<212> DNA
<213> Artificial Sequence

<220>
<223> ER2 tag

<400> 183
tctacctctg aaaaccaaag taaaggtagt ggtacattgg ttgtcatatt ggccatttta 60

atgctaggtg ttgcttatta tttgttgaac gaataa 96

<210> 184
<211> 31
<212> PRT
<213> Artificial Sequence

<220>
<223> ER2 tag

<400> 184

Ser Thr Ser Glu Asn Gln Ser Lys Gly Ser Gly Thr Leu Val Val Ile

1 5 10 15

Leu Ala Ile Leu Met Leu Gly Val Ala Tyr Tyr Leu Leu Asn Glu
20 25 30

<210> 185
<211> 99
<212> DNA
<213> Artificial Sequence

<220>
<223> PM1 tag

<400> 185
tggtacaagg atctaaaaat gaagatgtgt ctggctttag taatcatcat attgcttgtt 60

gtaatcatcg tccccattgc tgttcacttt agtcgataa 99

<210> 186
<211> 32
<212> PRT
<213> Artificial Sequence

<220>
<223> PM1 tag

<400> 186

Trp Tyr Lys Asp Leu Lys Met Lys Met Cys Leu Ala Leu Val Ile Ile
1 5 10 15

Ile Leu Leu Val Val Ile Ile Val Pro Ile Ala Val His Phe Ser Arg
20 25 30

<210> 187
<211> 108
<212> DNA
<213> Artificial Sequence

<220>
<223> VC1 tag

<400> 187
aatataaaag aaataatgtg gtggcagaag gtcaaaaata ttacgttatt aactttcact 60

attatactat ttgtaagtgc tgctttcatg tttttctatc tgtggtaa 108

<210> 188
<211> 35
<212> PRT
<213> Artificial Sequence

<220>
<223> VC1 tag

<400> 188

Asn Ile Lys Glu Ile Met Trp Trp Gln Lys Val Lys Asn Ile Thr Leu
1 5 10 15

Leu Thr Phe Thr Ile Ile Leu Phe Val Ser Ala Ala Phe Met Phe Phe
20 25 30

Tyr Leu Trp
35

<210> 189
<211> 12
<212> DNA
<213> Artificial Sequence

<220>
<223> PEX8 tag

<400> 189
tctaaattat aa 12

<210> 190
<211> 3
<212> PRT
<213> Artificial Sequence

<220>
<223> PEX8 tag

<400> 190

Ser Lys Leu
1

<210> 191
<211> 2103
<212> DNA

<213> Saccharomyces cerevisiae

<400> 191
atggttgctc aatataccgt tccagttggg aaagccgcca atgagcatga aactgctcca 60

agaagaaatt atcaatgccg cgagaagccg ctcgtcagac cgcctaacac aaagtgttcc 120

actgtttatg agtttgttct agagtgcttt cagaagaaca aaaattcaaa tgctatgggt 180

tggagggatg ttaaggaaat tcatgaagaa tccaaatcgg ttatgaaaaa agttgatggc 240

aaggagactt cagtggaaaa gaaatggatg tattatgaac tatcgcatta tcattataat 300

tcatttgacc aattgaccga tatcatgcat gaaattggtc gtgggttggt gaaaatagga 360

ttaaagccta atgatgatga caaattacat ctttacgcag ccacttctca caagtggatg 420

aagatgttct taggagcgca gtctcaaggt attcctgtcg tcactgccta cgatactttg 480

ggagagaaag ggctaattca ttctttggtg caaacggggt ctaaggccat ttttaccgat 540

aactctttat taccatcctt gatcaaacca gtgcaagccg ctcaagacgt aaaatacata 600

attcatttcg attccatcag ttctgaggac aggaggcaaa gtggtaagat ctatcaatct 660

gctcatgatg ccatcaacag aattaaagaa gttagacctg atatcaagac ctttagcttt 720

gacgacatct tgaagctagg taaagaatcc tgtaacgaaa tcgatgttca tccacctggc 780

aaggatgatc tttgttgcat catgtatacg tctggttcta caggtgagcc aaagggtgtt 840

gtcttgaaac attcaaatgt tgtcgcaggt gttggtggtg caagtttgaa tgttttgaag 900

tttgtgggca ataccgaccg tgttatctgt tttttgccac tagctcatat ttttgaattg 960

gttttcgaac tattgtcctt ttattggggg gcctgcattg gttatgccac cgtaaaaact 1020

ttaactagca gctctgtgag aaattgtcaa ggtgatttgc aagaattcaa gcccacaatc 1080

atggttggtg tcgccgctgt ttgggaaaca gtgagaaaag ggatcttaaa ccaaattgat 1140

aatttgccct tcctcaccaa gaaaatcttc tggaccgcgt ataataccaa gttgaacatg 1200

caacgtctcc acatccctgg tggcggcgcc ttaggaaact tggttttcaa aaaaatcaga 1260

actgccacag gtggccaatt aagatatttg ttaaacggtg gttctccaat cagtcgggat 1320

gctcaggaat tcatcacaaa tttaatctgc cctatgctta ttggttacgg tttaaccgag 1380

acatgcgcta gtaccaccat cttggatcct gctaattttg aactcggcgt cgctggtgac 1440

ctaacaggtt gtgttaccgt caaactagtt gatgttgaag aattaggtta ttttgctaaa 1500

aacaaccaag gtgaagtttg gatcacaggt gccaatgtca cgcctgaata ttataagaat 1560

gaggaagaaa cttctcaagc tttaacaagc gatggttggt tcaagaccgg tgacatcggt 1620

gaatgggaag caaatggcca tttgaaaata attgacagga agaaaaactt ggtcaaaaca 1680

atgaacggtg aatatatcgc actcgagaaa ttagagtccg tttacagatc taacgaatat 1740

gttgctaaca tttgtgttta tgccgaccaa tctaagacta agccagttgg tattattgta 1800

ccaaatcatg ctccattaac gaagcttgct aaaaagttgg gaattatgga acaaaaagac 1860

agttcaatta atatcgaaaa ttatttggag gatgcaaaat tgattaaagc tgtttattct 1920

gatcttttga agacaggtaa agaccaaggt ttggttggca ttgaattact agcaggcata 1980

gtgttctttg acggcgaatg gactccacaa aacggttttg ttacgtccgc tcagaaattg 2040

aaaagaaaag acattttgaa tgctgtcaaa gataaagttg acgccgttta tagttcgtct 2100

taa 2103

<210> 192
<211> 700
<212> PRT
<213> Saccharomyces cerevisiae

<400> 192

Met Val Ala Gln Tyr Thr Val Pro Val Gly Lys Ala Ala Asn Glu His
1 5 10 15

Glu Thr Ala Pro Arg Arg Asn Tyr Gln Cys Arg Glu Lys Pro Leu Val
20 25 30

Arg Pro Pro Asn Thr Lys Cys Ser Thr Val Tyr Glu Phe Val Leu Glu
35 40 45

Cys Phe Gln Lys Asn Lys Asn Ser Asn Ala Met Gly Trp Arg Asp Val
50 55 60

Lys Glu Ile His Glu Glu Ser Lys Ser Val Met Lys Lys Val Asp Gly
65 70 75 80

Lys Glu Thr Ser Val Glu Lys Lys Trp Met Tyr Tyr Glu Leu Ser His

85 90 95

Tyr His Tyr Asn Ser Phe Asp Gln Leu Thr Asp Ile Met His Glu Ile
100 105 110

Gly Arg Gly Leu Val Lys Ile Gly Leu Lys Pro Asn Asp Asp Asp Lys
115 120 125

Leu His Leu Tyr Ala Ala Thr Ser His Lys Trp Met Lys Met Phe Leu
130 135 140

Gly Ala Gln Ser Gln Gly Ile Pro Val Val Thr Ala Tyr Asp Thr Leu
145 150 155 160

Gly Glu Lys Gly Leu Ile His Ser Leu Val Gln Thr Gly Ser Lys Ala
165 170 175

Ile Phe Thr Asp Asn Ser Leu Leu Pro Ser Leu Ile Lys Pro Val Gln
180 185 190

Ala Ala Gln Asp Val Lys Tyr Ile Ile His Phe Asp Ser Ile Ser Ser
195 200 205

Glu Asp Arg Arg Gln Ser Gly Lys Ile Tyr Gln Ser Ala His Asp Ala
210 215 220

Ile Asn Arg Ile Lys Glu Val Arg Pro Asp Ile Lys Thr Phe Ser Phe
225 230 235 240

Asp Asp Ile Leu Lys Leu Gly Lys Glu Ser Cys Asn Glu Ile Asp Val
245 250 255

His Pro Pro Gly Lys Asp Asp Leu Cys Cys Ile Met Tyr Thr Ser Gly
260 265 270

Ser Thr Gly Glu Pro Lys Gly Val Val Leu Lys His Ser Asn Val Val
275 280 285

Ala Gly Val Gly Gly Ala Ser Leu Asn Val Leu Lys Phe Val Gly Asn

290 295 300

Thr Asp Arg Val Ile Cys Phe Leu Pro Leu Ala His Ile Phe Glu Leu
305 310 315 320

Val Phe Glu Leu Leu Ser Phe Tyr Trp Gly Ala Cys Ile Gly Tyr Ala
325 330 335

Thr Val Lys Thr Leu Thr Ser Ser Ser Val Arg Asn Cys Gln Gly Asp
340 345 350

Leu Gln Glu Phe Lys Pro Thr Ile Met Val Gly Val Ala Ala Val Trp
355 360 365

Glu Thr Val Arg Lys Gly Ile Leu Asn Gln Ile Asp Asn Leu Pro Phe
370 375 380

Leu Thr Lys Lys Ile Phe Trp Thr Ala Tyr Asn Thr Lys Leu Asn Met
385 390 395 400

Gln Arg Leu His Ile Pro Gly Gly Gly Ala Leu Gly Asn Leu Val Phe
405 410 415

Lys Lys Ile Arg Thr Ala Thr Gly Gly Gln Leu Arg Tyr Leu Leu Asn
420 425 430

Gly Gly Ser Pro Ile Ser Arg Asp Ala Gln Glu Phe Ile Thr Asn Leu
435 440 445

Ile Cys Pro Met Leu Ile Gly Tyr Gly Leu Thr Glu Thr Cys Ala Ser
450 455 460

Thr Thr Ile Leu Asp Pro Ala Asn Phe Glu Leu Gly Val Ala Gly Asp
465 470 475 480

Leu Thr Gly Cys Val Thr Val Lys Leu Val Asp Val Glu Glu Leu Gly
485 490 495

Tyr Phe Ala Lys Asn Asn Gln Gly Glu Val Trp Ile Thr Gly Ala Asn

```
            500                 505                 510

Val Thr Pro Glu Tyr Tyr Lys Asn Glu Glu Glu Thr Ser Gln Ala Leu
        515                 520                 525

Thr Ser Asp Gly Trp Phe Lys Thr Gly Asp Ile Gly Glu Trp Glu Ala
    530                 535                 540

Asn Gly His Leu Lys Ile Ile Asp Arg Lys Lys Asn Leu Val Lys Thr
545                 550                 555                 560

Met Asn Gly Glu Tyr Ile Ala Leu Glu Lys Leu Glu Ser Val Tyr Arg
                565                 570                 575

Ser Asn Glu Tyr Val Ala Asn Ile Cys Val Tyr Ala Asp Gln Ser Lys
            580                 585                 590

Thr Lys Pro Val Gly Ile Ile Val Pro Asn His Ala Pro Leu Thr Lys
        595                 600                 605

Leu Ala Lys Lys Leu Gly Ile Met Glu Gln Lys Asp Ser Ser Ile Asn
    610                 615                 620

Ile Glu Asn Tyr Leu Glu Asp Ala Lys Leu Ile Lys Ala Val Tyr Ser
625                 630                 635                 640

Asp Leu Leu Lys Thr Gly Lys Asp Gln Gly Leu Val Gly Ile Glu Leu
                645                 650                 655

Leu Ala Gly Ile Val Phe Phe Asp Gly Glu Trp Thr Pro Gln Asn Gly
            660                 665                 670

Phe Val Thr Ser Ala Gln Lys Leu Lys Arg Lys Asp Ile Leu Asn Ala
        675                 680                 685

Val Lys Asp Lys Val Asp Ala Val Tyr Ser Ser Ser
    690                 695                 700
```

<210> 193

<211> 2226
<212> DNA
<213> Artificial Sequence

<220>
<223> Truncated medium chain fatty acyl-CoA synthetase Sc_FAA2_Ctrunc

<400> 193
atggccgctc cagattatgc acttaccgat ttaattgaat cggatcctcg tttcgaaagt 60

ttgaagacaa gattagccgg ttacaccaaa ggctctgatg aatatattga agagctatac 120

tctcaattac cactgaccag ctaccccagg tacaaaacat ttttaaagaa acaggcggtt 180

gccatttcga atccggataa tgaagctggt tttagctcga tttataggag ttctctttct 240

tctgaaaatc tagtgagctg tgtggataaa aacttaagaa ctgcatacga tcacttcatg 300

ttttctgcaa ggagatggcc tcaacgtgac tgtttaggtt caaggccaat tgataaagcc 360

acaggcacct gggaggaaac attccgtttc gagtcgtact ccacggtatc taaaagatgt 420

cataatatcg gaagtggtat attgtctttg gtaaacacga aaaggaaacg tcctttggaa 480

gccaatgatt ttgttgttgc tatcttatca cacaacaacc ctgaatggat cctaacagat 540

ttggcctgtc aggcctattc tctaactaac acggctttgt acgaaacatt aggtccaaac 600

acctccgagt acatattgaa tttaaccgag gcccccattc tgatttttgc aaaatcaaat 660

atgtatcatg tattgaagat ggtgcctgat atgaaatttg ttaatacttt ggtttgtatg 720

gatgaattaa ctcatgacga gctccgtatg ctaaatgaat cgttgctacc cgttaagtgc 780

aactctctca atgaaaaaat cacatttttt tcattggagc aggtagaaca agttggttgc 840

tttaacaaaa ttcctgcaat tccacctacc ccagattcct tgtatactat ttcgtttact 900

tctggtacta caggtttacc taaaggtgtg gaaatgtctc acagaaacat tgcgtctggg 960

atagcatttg ctttttctac cttcagaata ccgccagata aaagaaacca acagttatat 1020

gatatgtgtt ttttgccatt ggctcatatt tttgaaagaa tggttattgc gtatgatcta 1080

gccatcgggt ttggaatagg cttcttacat aaaccagacc caactgtatt ggtagaggat 1140

ttgaagattt tgaaacctta cgcggttgcc ctggttccta gaatattaac acggtttgaa 1200

gccggtataa aaaatgcttt ggataaatcg actgtccaga ggaacgtagc aaatactata 1260

ttggattcta aatcggccag atttaccgca agaggtggtc cagataaatc gattatgaat 1320

tttctagttt atcatcgcgt attgattgat aaaatcagag actctttagg tttgtccaat 1380

aactcgttta taattaccgg atcagctccc atatctaaag ataccttact atttttaaga 1440

agcgccttgg atattggtat aagacagggc tacggcttaa ctgaaacttt tgctggtgtc 1500

tgtttaagcg aaccgtttga aaaagatgtc ggatcttgtg gtgccatagg tatttctgca 1560

gaatgtagat tgaagtctgt tccagaaatg ggttaccatg ccgacaagga tttaaaaggt 1620

gaactgcaaa ttcgtggccc acaggttttt gaaagatatt ttaaaaatcc gaatgaaact 1680

tcaaaagccg ttgaccaaga tggttggttt tccacgggag atgttgcatt tatcgatgca 1740

aaaggtcgca tcagcgtcat tgatcgagtc aagaactttt tcaagctagc acatggtgaa 1800

tatattgctc cagagaaaat cgaaaatatt tatttatcat catgccccta tatcacgcaa 1860

atatttgtct ttggagatcc tttgaagaca tttttagttg gcatcgttgg tgttgatgtt 1920

gatgcagcgc aaccgatttt agctgcaaag cacccagagg tgaaaacgtg gactaaggaa 1980

gtgctagtag aaaacttaaa tcgtaataaa aagctaagga aggaattttt aaacaaaatt 2040

aataaatgca tcgatgggct acaaggattt gaaaaattgc acaacatcaa agtcggactt 2100

gagcctttga ctctcgagga tgatgttgtg acgccaactt ttaaaataaa gcgtgccaaa 2160

gcatcaaaat tcttcaaaga tacattagac caactatacg ccgaaggttc actagtcaag 2220

acatag 2226

<210> 194
<211> 741
<212> PRT
<213> Artificial Sequence

<220>
<223> Truncated medium chain fatty acyl-CoA synthetase Sc_FAA2_Ctrunc

<400> 194

Met Ala Ala Pro Asp Tyr Ala Leu Thr Asp Leu Ile Glu Ser Asp Pro
1 5 10 15

Arg Phe Glu Ser Leu Lys Thr Arg Leu Ala Gly Tyr Thr Lys Gly Ser
20 25 30

Asp Glu Tyr Ile Glu Glu Leu Tyr Ser Gln Leu Pro Leu Thr Ser Tyr

```
        35                  40                  45

Pro Arg Tyr Lys Thr Phe Leu Lys Lys Gln Ala Val Ala Ile Ser Asn
    50                  55                  60

Pro Asp Asn Glu Ala Gly Phe Ser Ser Ile Tyr Arg Ser Ser Leu Ser
65                  70                  75                  80

Ser Glu Asn Leu Val Ser Cys Val Asp Lys Asn Leu Arg Thr Ala Tyr
                85                  90                  95

Asp His Phe Met Phe Ser Ala Arg Arg Trp Pro Gln Arg Asp Cys Leu
            100                 105                 110

Gly Ser Arg Pro Ile Asp Lys Ala Thr Gly Thr Trp Glu Glu Thr Phe
        115                 120                 125

Arg Phe Glu Ser Tyr Ser Thr Val Ser Lys Arg Cys His Asn Ile Gly
    130                 135                 140

Ser Gly Ile Leu Ser Leu Val Asn Thr Lys Arg Lys Arg Pro Leu Glu
145                 150                 155                 160

Ala Asn Asp Phe Val Val Ala Ile Leu Ser His Asn Asn Pro Glu Trp
                165                 170                 175

Ile Leu Thr Asp Leu Ala Cys Gln Ala Tyr Ser Leu Thr Asn Thr Ala
            180                 185                 190

Leu Tyr Glu Thr Leu Gly Pro Asn Thr Ser Glu Tyr Ile Leu Asn Leu
        195                 200                 205

Thr Glu Ala Pro Ile Leu Ile Phe Ala Lys Ser Asn Met Tyr His Val
    210                 215                 220

Leu Lys Met Val Pro Asp Met Lys Phe Val Asn Thr Leu Val Cys Met
225                 230                 235                 240

Asp Glu Leu Thr His Asp Glu Leu Arg Met Leu Asn Glu Ser Leu Leu
```

245 250 255

Pro Val Lys Cys Asn Ser Leu Asn Glu Lys Ile Thr Phe Phe Ser Leu
260 265 270

Glu Gln Val Glu Gln Val Gly Cys Phe Asn Lys Ile Pro Ala Ile Pro
275 280 285

Pro Thr Pro Asp Ser Leu Tyr Thr Ile Ser Phe Thr Ser Gly Thr Thr
290 295 300

Gly Leu Pro Lys Gly Val Glu Met Ser His Arg Asn Ile Ala Ser Gly
305 310 315 320

Ile Ala Phe Ala Phe Ser Thr Phe Arg Ile Pro Pro Asp Lys Arg Asn
325 330 335

Gln Gln Leu Tyr Asp Met Cys Phe Leu Pro Leu Ala His Ile Phe Glu
340 345 350

Arg Met Val Ile Ala Tyr Asp Leu Ala Ile Gly Phe Gly Ile Gly Phe
355 360 365

Leu His Lys Pro Asp Pro Thr Val Leu Val Glu Asp Leu Lys Ile Leu
370 375 380

Lys Pro Tyr Ala Val Ala Leu Val Pro Arg Ile Leu Thr Arg Phe Glu
385 390 395 400

Ala Gly Ile Lys Asn Ala Leu Asp Lys Ser Thr Val Gln Arg Asn Val
405 410 415

Ala Asn Thr Ile Leu Asp Ser Lys Ser Ala Arg Phe Thr Ala Arg Gly
420 425 430

Gly Pro Asp Lys Ser Ile Met Asn Phe Leu Val Tyr His Arg Val Leu
435 440 445

Ile Asp Lys Ile Arg Asp Ser Leu Gly Leu Ser Asn Asn Ser Phe Ile

450 455 460

Ile Thr Gly Ser Ala Pro Ile Ser Lys Asp Thr Leu Leu Phe Leu Arg
465 470 475 480

Ser Ala Leu Asp Ile Gly Ile Arg Gln Gly Tyr Gly Leu Thr Glu Thr
485 490 495

Phe Ala Gly Val Cys Leu Ser Glu Pro Phe Glu Lys Asp Val Gly Ser
500 505 510

Cys Gly Ala Ile Gly Ile Ser Ala Glu Cys Arg Leu Lys Ser Val Pro
515 520 525

Glu Met Gly Tyr His Ala Asp Lys Asp Leu Lys Gly Glu Leu Gln Ile
530 535 540

Arg Gly Pro Gln Val Phe Glu Arg Tyr Phe Lys Asn Pro Asn Glu Thr
545 550 555 560

Ser Lys Ala Val Asp Gln Asp Gly Trp Phe Ser Thr Gly Asp Val Ala
565 570 575

Phe Ile Asp Ala Lys Gly Arg Ile Ser Val Ile Asp Arg Val Lys Asn
580 585 590

Phe Phe Lys Leu Ala His Gly Glu Tyr Ile Ala Pro Glu Lys Ile Glu
595 600 605

Asn Ile Tyr Leu Ser Ser Cys Pro Tyr Ile Thr Gln Ile Phe Val Phe
610 615 620

Gly Asp Pro Leu Lys Thr Phe Leu Val Gly Ile Val Gly Val Asp Val
625 630 635 640

Asp Ala Ala Gln Pro Ile Leu Ala Ala Lys His Pro Glu Val Lys Thr
645 650 655

Trp Thr Lys Glu Val Leu Val Glu Asn Leu Asn Arg Asn Lys Lys Leu

660 665 670

Arg Lys Glu Phe Leu Asn Lys Ile Asn Lys Cys Ile Asp Gly Leu Gln
675 680 685

Gly Phe Glu Lys Leu His Asn Ile Lys Val Gly Leu Glu Pro Leu Thr
690 695 700

Leu Glu Asp Asp Val Val Thr Pro Thr Phe Lys Ile Lys Arg Ala Lys
705 710 715 720

Ala Ser Lys Phe Phe Lys Asp Thr Leu Asp Gln Leu Tyr Ala Glu Gly
725 730 735

Ser Leu Val Lys Thr
740

<210> 195
<211> 2238
<212> DNA
<213> Artificial Sequence

<220>
<223> Mutated medium chain fatty acyl-CoA synthetase Sc_FAA2_Cmut

<400> 195
atggccgctc cagattatgc acttaccgat ttaattgaat cggatcctcg tttcgaaagt 60

ttgaagacaa gattagccgg ttacaccaaa ggctctgatg aatatattga agagctatac 120

tctcaattac cactgaccag ctaccccagg tacaaaacat ttttaaagaa acaggcggtt 180

gccatttcga atccggataa tgaagctggt tttagctcga tttataggag ttctctttct 240

tctgaaaatc tagtgagctg tgtggataaa aacttaagaa ctgcatacga tcacttcatg 300

ttttctgcaa ggagatggcc tcaacgtgac tgtttaggtt caaggccaat tgataaagcc 360

acaggcacct gggaggaaac attccgtttc gagtcgtact ccacggtatc taaaagatgt 420

cataatatcg gaagtggtat attgtctttg gtaaacacga aaaggaaacg tcctttggaa 480

gccaatgatt ttgttgttgc tatcttatca cacaacaacc ctgaatggat cctaacagat 540

ttggcctgtc aggcctattc tctaactaac acggctttgt acgaaacatt aggtccaaac 600

acctccgagt acatattgaa tttaaccgag gcccccattc tgatttttgc aaaatcaaat 660

atgtatcatg tattgaagat ggtgcctgat atgaaatttg ttaatacttt ggtttgtatg 720

gatgaattaa ctcatgacga gctccgtatg ctaaatgaat cgttgctacc cgttaagtgc 780

aactctctca atgaaaaaat cacatttttt tcattggagc aggtagaaca agttggttgc 840

tttaacaaaa ttcctgcaat tccacctacc ccagattcct tgtatactat ttcgtttact 900

tctggtacta caggtttacc taaaggtgtg gaaatgtctc acagaaacat tgcgtctggg 960

atagcatttg ctttttctac cttcagaata ccgccagata aaagaaacca acagttatat 1020

gatatgtgtt ttttgccatt ggctcatatt tttgaaagaa tggttattgc gtatgatcta 1080

gccatcgggt ttggaatagg cttcttacat aaaccagacc caactgtatt ggtagaggat 1140

ttgaagattt tgaaacctta cgcggttgcc ctggttccta gaatattaac acggtttgaa 1200

gccggtataa aaaatgcttt ggataaatcg actgtccaga ggaacgtagc aaatactata 1260

ttggattcta aatcggccag atttaccgca agaggtggtc cagataaatc gattatgaat 1320

tttctagttt atcatcgcgt attgattgat aaaatcagag actctttagg tttgtccaat 1380

aactcgttta taattaccgg atcagctccc atatctaaag ataccttact atttttaaga 1440

agcgccttgg atattggtat aagacagggc tacggcttaa ctgaaacttt tgctggtgtc 1500

tgtttaagcg aaccgtttga aaaagatgtc ggatcttgtg gtgccatagg tatttctgca 1560

gaatgtagat tgaagtctgt tccagaaatg ggttaccatg ccgacaagga tttaaaaggt 1620

gaactgcaaa ttcgtggccc acaggttttt gaaagatatt ttaaaaatcc gaatgaaact 1680

tcaaaagccg ttgaccaaga tggttggttt tccacgggag atgttgcatt tatcgatgca 1740

aaaggtcgca tcagcgtcat tgatcgagtc aagaactttt tcaagctagc acatggtgaa 1800

tatattgctc cagagaaaat cgaaaatatt tatttatcat catgccccta tatcacgcaa 1860

atatttgtct ttggagatcc tttgaagaca tttttagttg gcatcgttgg tgttgatgtt 1920

gatgcagcgc aaccgatttt agctgcaaag cacccagagg tgaaaacgtg gactaaggaa 1980

gtgctagtag aaaacttaaa tcgtaataaa aagctaagga aggaattttt aaacaaaatt 2040

aataaatgca tcgatgggct acaaggattt gaaaaattgc acaacatcaa agtcggactt 2100

gagcctttga ctctcgagga tgatgttgtg acgccaactt ttaaaataaa gcgtgccaaa 2160

gcatcaaaat tcttcaaaga tacattagac caactatacg ccgaaggttc actagtcaag 2220

acagaaaagc ttaaatag 2238

<210> 196
<211> 745
<212> PRT
<213> Artificial Sequence

<220>
<223> Mutated medium chain fatty acyl-CoA synthetase Sc_FAA2_Cmut

<400> 196

Met Ala Ala Pro Asp Tyr Ala Leu Thr Asp Leu Ile Glu Ser Asp Pro
1 5 10 15

Arg Phe Glu Ser Leu Lys Thr Arg Leu Ala Gly Tyr Thr Lys Gly Ser
20 25 30

Asp Glu Tyr Ile Glu Glu Leu Tyr Ser Gln Leu Pro Leu Thr Ser Tyr
35 40 45

Pro Arg Tyr Lys Thr Phe Leu Lys Lys Gln Ala Val Ala Ile Ser Asn
50 55 60

Pro Asp Asn Glu Ala Gly Phe Ser Ser Ile Tyr Arg Ser Ser Leu Ser
65 70 75 80

Ser Glu Asn Leu Val Ser Cys Val Asp Lys Asn Leu Arg Thr Ala Tyr
85 90 95

Asp His Phe Met Phe Ser Ala Arg Arg Trp Pro Gln Arg Asp Cys Leu
100 105 110

Gly Ser Arg Pro Ile Asp Lys Ala Thr Gly Thr Trp Glu Glu Thr Phe
115 120 125

Arg Phe Glu Ser Tyr Ser Thr Val Ser Lys Arg Cys His Asn Ile Gly
130 135 140

Ser Gly Ile Leu Ser Leu Val Asn Thr Lys Arg Lys Arg Pro Leu Glu

145 150 155 160

Ala Asn Asp Phe Val Val Ala Ile Leu Ser His Asn Asn Pro Glu Trp
165 170 175

Ile Leu Thr Asp Leu Ala Cys Gln Ala Tyr Ser Leu Thr Asn Thr Ala
180 185 190

Leu Tyr Glu Thr Leu Gly Pro Asn Thr Ser Glu Tyr Ile Leu Asn Leu
195 200 205

Thr Glu Ala Pro Ile Leu Ile Phe Ala Lys Ser Asn Met Tyr His Val
210 215 220

Leu Lys Met Val Pro Asp Met Lys Phe Val Asn Thr Leu Val Cys Met
225 230 235 240

Asp Glu Leu Thr His Asp Glu Leu Arg Met Leu Asn Glu Ser Leu Leu
245 250 255

Pro Val Lys Cys Asn Ser Leu Asn Glu Lys Ile Thr Phe Phe Ser Leu
260 265 270

Glu Gln Val Glu Gln Val Gly Cys Phe Asn Lys Ile Pro Ala Ile Pro
275 280 285

Pro Thr Pro Asp Ser Leu Tyr Thr Ile Ser Phe Thr Ser Gly Thr Thr
290 295 300

Gly Leu Pro Lys Gly Val Glu Met Ser His Arg Asn Ile Ala Ser Gly
305 310 315 320

Ile Ala Phe Ala Phe Ser Thr Phe Arg Ile Pro Pro Asp Lys Arg Asn
325 330 335

Gln Gln Leu Tyr Asp Met Cys Phe Leu Pro Leu Ala His Ile Phe Glu
340 345 350

Arg Met Val Ile Ala Tyr Asp Leu Ala Ile Gly Phe Gly Ile Gly Phe

355 360 365

Leu His Lys Pro Asp Pro Thr Val Leu Val Glu Asp Leu Lys Ile Leu
370 375 380

Lys Pro Tyr Ala Val Ala Leu Val Pro Arg Ile Leu Thr Arg Phe Glu
385 390 395 400

Ala Gly Ile Lys Asn Ala Leu Asp Lys Ser Thr Val Gln Arg Asn Val
405 410 415

Ala Asn Thr Ile Leu Asp Ser Lys Ser Ala Arg Phe Thr Ala Arg Gly
420 425 430

Gly Pro Asp Lys Ser Ile Met Asn Phe Leu Val Tyr His Arg Val Leu
435 440 445

Ile Asp Lys Ile Arg Asp Ser Leu Gly Leu Ser Asn Asn Ser Phe Ile
450 455 460

Ile Thr Gly Ser Ala Pro Ile Ser Lys Asp Thr Leu Leu Phe Leu Arg
465 470 475 480

Ser Ala Leu Asp Ile Gly Ile Arg Gln Gly Tyr Gly Leu Thr Glu Thr
485 490 495

Phe Ala Gly Val Cys Leu Ser Glu Pro Phe Glu Lys Asp Val Gly Ser
500 505 510

Cys Gly Ala Ile Gly Ile Ser Ala Glu Cys Arg Leu Lys Ser Val Pro
515 520 525

Glu Met Gly Tyr His Ala Asp Lys Asp Leu Lys Gly Glu Leu Gln Ile
530 535 540

Arg Gly Pro Gln Val Phe Glu Arg Tyr Phe Lys Asn Pro Asn Glu Thr
545 550 555 560

Ser Lys Ala Val Asp Gln Asp Gly Trp Phe Ser Thr Gly Asp Val Ala

565 570 575

Phe Ile Asp Ala Lys Gly Arg Ile Ser Val Ile Asp Arg Val Lys Asn
580 585 590

Phe Phe Lys Leu Ala His Gly Glu Tyr Ile Ala Pro Glu Lys Ile Glu
595 600 605

Asn Ile Tyr Leu Ser Ser Cys Pro Tyr Ile Thr Gln Ile Phe Val Phe
610 615 620

Gly Asp Pro Leu Lys Thr Phe Leu Val Gly Ile Val Gly Val Asp Val
625 630 635 640

Asp Ala Ala Gln Pro Ile Leu Ala Ala Lys His Pro Glu Val Lys Thr
645 650 655

Trp Thr Lys Glu Val Leu Val Glu Asn Leu Asn Arg Asn Lys Lys Leu
660 665 670

Arg Lys Glu Phe Leu Asn Lys Ile Asn Lys Cys Ile Asp Gly Leu Gln
675 680 685

Gly Phe Glu Lys Leu His Asn Ile Lys Val Gly Leu Glu Pro Leu Thr
690 695 700

Leu Glu Asp Asp Val Val Thr Pro Thr Phe Lys Ile Lys Arg Ala Lys
705 710 715 720

Ala Ser Lys Phe Phe Lys Asp Thr Leu Asp Gln Leu Tyr Ala Glu Gly
725 730 735

Ser Leu Val Lys Thr Glu Lys Leu Lys
740 745

<210> 197
<211> 2085
<212> DNA
<213> Saccharomyces cerevisiae

<400> 197
atgtccgaac aacactctgt cgcagtcggt aaagctgcta atgagcacga gactgcccct 60

aggagaaatg ttagagtcaa gaagcggccc ttaattagac cattgaactc gtcagcatct 120

acgctgtatg aatttgccct agagtgtttc aacaagggtg gaaaacgaga tggtatggct 180

tggagagatg tcatcgagat tcatgagaca aagaaaacca ttgtgagaaa ggtagacggc 240

aaggataaat ctatagaaaa gacatggctg tattatgaaa tgtcaccata taaaatgatg 300

acctaccagg aactgatctg ggtgatgcac gatatgggcc gtgggctggc aaaaataggc 360

atcaagccca atggagaaca caaattccac atcttcgcat ctacttccca taaatggatg 420

aagattttcc ttggttgcat atcccagggt atccccgtag taaccgcgta tgatactttg 480

ggtgagagcg gtttgattca ctccatggtt gaaaccgagt ctgctgctat tttcactgat 540

aatcaattat tggctaaaat gatagtgcct ttgcaatctg ctaaagatat caaatttctt 600

atccataacg aacctatcga ccccaatgac agaagacaaa acggcaaact ttacaaggct 660

gctaaggatg ccattaataa gatcagagaa gttaggccag acataaaaat ttatagtttt 720

gaagaagttg tcaagatagg taaaaaaagt aaagatgagg tcaaacttca tccacctgag 780

ccaaaagatt tggcttgtat catgtacacc tcgggctcga tcagtgcacc aaaaggtgta 840

gtattgactc attataatat tgtttcgggt atcgctggtg taggtcacaa cgtctttgga 900

tggatcggct ctacagaccg tgttttgtcg ttcttgccat tggctcatat ttttgaactg 960

gtctttgaat tcgaagcctt ttactggaac ggtattcttg ggtacggtag tgttaagact 1020

ttgactaata cttcgactcg taattgtaag ggtgacctgg ttgagtttaa gcctactatt 1080

atgatcggtg tggctgccgt ttgggaaact gtgagaaaag ctattttgga aaagatcagc 1140

gatttaactc ccgtactcca aaagattttt tggtctgcct atagtatgaa agaaaagagt 1200

gtaccatgca ccgggttttt aagtcgtatg gtcttcaaga aagtcagaca agccaccggt 1260

ggtcatctta agtatattat gaacggtggg tctgcgatca gtattgatgc tcagaaattc 1320

ttttctatcg tcctgtgtcc tatgattatc ggttacggcc ttactgaaac agttgcgaat 1380

gcttgtgttt tggagcctga tcatttcgaa tatggtatag ttggtgatct tgttggatcg 1440

gtcactgcca aattggtgga tgttaaggac ctaggttatt atgcaaaaaa caatcaaggt 1500

gaattgcttc taaagggtgc gccggtctgt tctgaatatt ataagaatcc aatagaaacg 1560

gcggtctctt tcacttacga tggatggttt cgtactggtg atattgttga atggactccc 1620

aagggacaac ttaaaattat tgatagaaga aagaatttgg ttaaaaccct aaatggtgaa 1680

tatattgcat tagaaaagtt agaatctgtt tacaggtcaa actcctatgt gaaaaatatc 1740

tgtgtttatg ccgatgaaag tagggttaaa ccggtgggta ttgtggtacc caacccagga 1800

cccctatcta aatttgctgt caaattgcgt attatgaaaa agggtgaaga catcgaaaac 1860

tatatccatg acaaagcatt acgaaatgct gttttcaaag agatgatcgc aacagccaaa 1920

tctcaaggtt tggttggtat tgaactatta tgtggtattg ttttctttga tgaagaatgg 1980

acacctgaaa atggctttgt cacatctgct caaaaattaa agagaagaga aatcttagcc 2040

gctgttaaat cagaagtcga aagggtttac aaagaaaatt cttag 2085

<210> 198
<211> 694
<212> PRT
<213> Saccharomyces cerevisiae

<400> 198

Met Ser Glu Gln His Ser Val Ala Val Gly Lys Ala Ala Asn Glu His
1 5 10 15

Glu Thr Ala Pro Arg Arg Asn Val Arg Val Lys Lys Arg Pro Leu Ile
20 25 30

Arg Pro Leu Asn Ser Ser Ala Ser Thr Leu Tyr Glu Phe Ala Leu Glu
35 40 45

Cys Phe Asn Lys Gly Gly Lys Arg Asp Gly Met Ala Trp Arg Asp Val
50 55 60

Ile Glu Ile His Glu Thr Lys Lys Thr Ile Val Arg Lys Val Asp Gly
65 70 75 80

Lys Asp Lys Ser Ile Glu Lys Thr Trp Leu Tyr Tyr Glu Met Ser Pro
85 90 95

Tyr Lys Met Met Thr Tyr Gln Glu Leu Ile Trp Val Met His Asp Met

100 105 110

Gly Arg Gly Leu Ala Lys Ile Gly Ile Lys Pro Asn Gly Glu His Lys
115 120 125

Phe His Ile Phe Ala Ser Thr Ser His Lys Trp Met Lys Ile Phe Leu
130 135 140

Gly Cys Ile Ser Gln Gly Ile Pro Val Val Thr Ala Tyr Asp Thr Leu
145 150 155 160

Gly Glu Ser Gly Leu Ile His Ser Met Val Glu Thr Glu Ser Ala Ala
165 170 175

Ile Phe Thr Asp Asn Gln Leu Leu Ala Lys Met Ile Val Pro Leu Gln
180 185 190

Ser Ala Lys Asp Ile Lys Phe Leu Ile His Asn Glu Pro Ile Asp Pro
195 200 205

Asn Asp Arg Arg Gln Asn Gly Lys Leu Tyr Lys Ala Ala Lys Asp Ala
210 215 220

Ile Asn Lys Ile Arg Glu Val Arg Pro Asp Ile Lys Ile Tyr Ser Phe
225 230 235 240

Glu Glu Val Val Lys Ile Gly Lys Lys Ser Lys Asp Glu Val Lys Leu
245 250 255

His Pro Pro Glu Pro Lys Asp Leu Ala Cys Ile Met Tyr Thr Ser Gly
260 265 270

Ser Ile Ser Ala Pro Lys Gly Val Val Leu Thr His Tyr Asn Ile Val
275 280 285

Ser Gly Ile Ala Gly Val Gly His Asn Val Phe Gly Trp Ile Gly Ser
290 295 300

Thr Asp Arg Val Leu Ser Phe Leu Pro Leu Ala His Ile Phe Glu Leu

305 310 315 320

Val Phe Glu Phe Glu Ala Phe Tyr Trp Asn Gly Ile Leu Gly Tyr Gly
325 330 335

Ser Val Lys Thr Leu Thr Asn Thr Ser Thr Arg Asn Cys Lys Gly Asp
340 345 350

Leu Val Glu Phe Lys Pro Thr Ile Met Ile Gly Val Ala Ala Val Trp
355 360 365

Glu Thr Val Arg Lys Ala Ile Leu Glu Lys Ile Ser Asp Leu Thr Pro
370 375 380

Val Leu Gln Lys Ile Phe Trp Ser Ala Tyr Ser Met Lys Glu Lys Ser
385 390 395 400

Val Pro Cys Thr Gly Phe Leu Ser Arg Met Val Phe Lys Lys Val Arg
405 410 415

Gln Ala Thr Gly Gly His Leu Lys Tyr Ile Met Asn Gly Gly Ser Ala
420 425 430

Ile Ser Ile Asp Ala Gln Lys Phe Phe Ser Ile Val Leu Cys Pro Met
435 440 445

Ile Ile Gly Tyr Gly Leu Thr Glu Thr Val Ala Asn Ala Cys Val Leu
450 455 460

Glu Pro Asp His Phe Glu Tyr Gly Ile Val Gly Asp Leu Val Gly Ser
465 470 475 480

Val Thr Ala Lys Leu Val Asp Val Lys Asp Leu Gly Tyr Tyr Ala Lys
485 490 495

Asn Asn Gln Gly Glu Leu Leu Leu Lys Gly Ala Pro Val Cys Ser Glu
500 505 510

Tyr Tyr Lys Asn Pro Ile Glu Thr Ala Val Ser Phe Thr Tyr Asp Gly

515 520 525

Trp Phe Arg Thr Gly Asp Ile Val Glu Trp Thr Pro Lys Gly Gln Leu
530 535 540

Lys Ile Ile Asp Arg Arg Lys Asn Leu Val Lys Thr Leu Asn Gly Glu
545 550 555 560

Tyr Ile Ala Leu Glu Lys Leu Glu Ser Val Tyr Arg Ser Asn Ser Tyr
565 570 575

Val Lys Asn Ile Cys Val Tyr Ala Asp Glu Ser Arg Val Lys Pro Val
580 585 590

Gly Ile Val Val Pro Asn Pro Gly Pro Leu Ser Lys Phe Ala Val Lys
595 600 605

Leu Arg Ile Met Lys Lys Gly Glu Asp Ile Glu Asn Tyr Ile His Asp
610 615 620

Lys Ala Leu Arg Asn Ala Val Phe Lys Glu Met Ile Ala Thr Ala Lys
625 630 635 640

Ser Gln Gly Leu Val Gly Ile Glu Leu Leu Cys Gly Ile Val Phe Phe
645 650 655

Asp Glu Glu Trp Thr Pro Glu Asn Gly Phe Val Thr Ser Ala Gln Lys
660 665 670

Leu Lys Arg Arg Glu Ile Leu Ala Ala Val Lys Ser Glu Val Glu Arg
675 680 685

Val Tyr Lys Glu Asn Ser
690

<210> 199
<211> 2085
<212> DNA
<213> Saccharomyces cerevisiae

<400> 199

atgaccgaac aatattccgt tgcagttggc gaagccgaca atgagcatga aaccgctcca 60

agaagaaata tcagggttaa agacaagcct ttgattagac ccataaactc ctcagcatct 120

acactgtacg aattcgccct ggaatgtttt accaaaggtg gtaagagaga cggtatggca 180

tggagagata ttatagatat acatgagacg aaaaaaacca tagtcaagag ggtggatggt 240

aaggataagc ccatcgaaaa aacatggttg tactacgaac tgactcccta cataaccatg 300

acatacgagg agatgatctg cgtaatgcac gacattggac gtgggctgat aaagattggt 360

gttaaaccta acggtgagaa caagttccac atctttgcct ctacatctca caagtggatg 420

aaaacttttc ttggttgcat gtcacaaggt attcctgtgg tcaccgcgta cgacactttg 480

ggtgagagcg gtttgattca ctccatggtg gaaacggatt ccgtcgccat tttcacggac 540

aaccagctgt tgtccaaatt agcagttcct ttgaaaaccg ccaagaacgt aaaattcgtc 600

attcacaacg aacccatcga tccaagtgac aaaagacaaa atggtaagct ttacaaggct 660

gccaaggatg ctgttgacaa aatcaaggaa gttagaccgg acataaaaat ctacagtttc 720

gatgaaatta ttgagatagg taaaaaggcc aaggacgagg ttgaattgca tttccccaag 780

cctgaagatc cagcttgtat catgtacact tctggttcca ctggtacacc aaagggtgtg 840

gtattgacac attacaacat tgtagctggt attggtggtg tgggccataa cgttatcgga 900

tggattggcc caacagaccg tattatcgca ttcttgccat tggctcatat ttttgaatta 960

atctttgaat tcgaagcgtt ctactggaat ggtatcctag ggtacgccac tgtcaagact 1020

ttaaccccaa cttctacacg taattgccaa ggtgacctga tggagtttaa acctaccgta 1080

atggtaggtg ttgccgcagt ttgggaaaca gtgagaaaag gtatcctggc caagatcaac 1140

gaattgcccg gttggtctca aacgcttttc tggactgtct atgctttgaa agagagaaat 1200

ataccatgca gcggcttgct gagtgggttg atcttcaaga gaatcagaga agcaaccggt 1260

ggaaacttaa ggtttattct gaacggtggg tctgcaatca gcatagacgc ccaaaaattc 1320

ctctccaacc ttctatgtcc tatgctcatt ggatatgggc taactgaggg tgtggctaat 1380

gcctgtgtcc tggagcctga acattttgat tacggtattg ctggtgacct tgtcggaact 1440

attacagcta aattggtgga tgtcgaagat ttgggctatt ttgccaagaa taaccaaggt 1500

gaattgctgt taaagggtgc acccatctgt tctgaatact ataagaatcc tgaagaaact 1560

gctgcggcct ttaccgatga tggctggttc cgtaccggtg atatcgctga atggaccccc 1620

aagggacaaa ttaagatcat tgatagaaag aaaaatttgg tcaagacctt aaacggtgag 1680

tacattgcat tggaaaaatt agaatccatt tacagatcaa atccttacgt ccaaaacatc 1740

tgtgtctacg ctgatgaaaa caaagttaag cctgtcggta ttgtggtccc taacttagga 1800

cacttgtcta agctggctat cgaattaggt ataatggtac caggtgaaga tgtcgaaagc 1860

tatatccatg aaaagaagct acaggatgcc gtttgcaaag atatgctgtc aactgccaaa 1920

tctcaaggct tgaatggtat tgaattatta tgtggcattg ttttctttga agaagaatgg 1980

actccagaaa acggtcttgt tacatccgcc caaaaattaa agagaagaga tattctagcg 2040

gctgtcaagc cagatgtgga aagagtttat aaagaaaaca cttaa 2085

<210> 200
<211> 694
<212> PRT
<213> Saccharomyces cerevisiae

<400> 200

Met Thr Glu Gln Tyr Ser Val Ala Val Gly Glu Ala Asp Asn Glu His
1 5 10 15

Glu Thr Ala Pro Arg Arg Asn Ile Arg Val Lys Asp Lys Pro Leu Ile
20 25 30

Arg Pro Ile Asn Ser Ser Ala Ser Thr Leu Tyr Glu Phe Ala Leu Glu
35 40 45

Cys Phe Thr Lys Gly Gly Lys Arg Asp Gly Met Ala Trp Arg Asp Ile
50 55 60

Ile Asp Ile His Glu Thr Lys Lys Thr Ile Val Lys Arg Val Asp Gly
65 70 75 80

Lys Asp Lys Pro Ile Glu Lys Thr Trp Leu Tyr Tyr Glu Leu Thr Pro
85 90 95

Tyr Ile Thr Met Thr Tyr Glu Glu Met Ile Cys Val Met His Asp Ile

100 105 110

Gly Arg Gly Leu Ile Lys Ile Gly Val Lys Pro Asn Gly Glu Asn Lys
115 120 125

Phe His Ile Phe Ala Ser Thr Ser His Lys Trp Met Lys Thr Phe Leu
130 135 140

Gly Cys Met Ser Gln Gly Ile Pro Val Val Thr Ala Tyr Asp Thr Leu
145 150 155 160

Gly Glu Ser Gly Leu Ile His Ser Met Val Glu Thr Asp Ser Val Ala
165 170 175

Ile Phe Thr Asp Asn Gln Leu Leu Ser Lys Leu Ala Val Pro Leu Lys
180 185 190

Thr Ala Lys Asn Val Lys Phe Val Ile His Asn Glu Pro Ile Asp Pro
195 200 205

Ser Asp Lys Arg Gln Asn Gly Lys Leu Tyr Lys Ala Ala Lys Asp Ala
210 215 220

Val Asp Lys Ile Lys Glu Val Arg Pro Asp Ile Lys Ile Tyr Ser Phe
225 230 235 240

Asp Glu Ile Ile Glu Ile Gly Lys Lys Ala Lys Asp Glu Val Glu Leu
245 250 255

His Phe Pro Lys Pro Glu Asp Pro Ala Cys Ile Met Tyr Thr Ser Gly
260 265 270

Ser Thr Gly Thr Pro Lys Gly Val Val Leu Thr His Tyr Asn Ile Val
275 280 285

Ala Gly Ile Gly Gly Val Gly His Asn Val Ile Gly Trp Ile Gly Pro
290 295 300

Thr Asp Arg Ile Ile Ala Phe Leu Pro Leu Ala His Ile Phe Glu Leu

305 310 315 320

Ile Phe Glu Phe Glu Ala Phe Tyr Trp Asn Gly Ile Leu Gly Tyr Ala
325 330 335

Thr Val Lys Thr Leu Thr Pro Thr Ser Thr Arg Asn Cys Gln Gly Asp
340 345 350

Leu Met Glu Phe Lys Pro Thr Val Met Val Gly Val Ala Ala Val Trp
355 360 365

Glu Thr Val Arg Lys Gly Ile Leu Ala Lys Ile Asn Glu Leu Pro Gly
370 375 380

Trp Ser Gln Thr Leu Phe Trp Thr Val Tyr Ala Leu Lys Glu Arg Asn
385 390 395 400

Ile Pro Cys Ser Gly Leu Leu Ser Gly Leu Ile Phe Lys Arg Ile Arg
405 410 415

Glu Ala Thr Gly Gly Asn Leu Arg Phe Ile Leu Asn Gly Gly Ser Ala
420 425 430

Ile Ser Ile Asp Ala Gln Lys Phe Leu Ser Asn Leu Leu Cys Pro Met
435 440 445

Leu Ile Gly Tyr Gly Leu Thr Glu Gly Val Ala Asn Ala Cys Val Leu
450 455 460

Glu Pro Glu His Phe Asp Tyr Gly Ile Ala Gly Asp Leu Val Gly Thr
465 470 475 480

Ile Thr Ala Lys Leu Val Asp Val Glu Asp Leu Gly Tyr Phe Ala Lys
485 490 495

Asn Asn Gln Gly Glu Leu Leu Leu Lys Gly Ala Pro Ile Cys Ser Glu
500 505 510

Tyr Tyr Lys Asn Pro Glu Glu Thr Ala Ala Ala Phe Thr Asp Asp Gly

515 520 525

Trp Phe Arg Thr Gly Asp Ile Ala Glu Trp Thr Pro Lys Gly Gln Ile
530 535 540

Lys Ile Ile Asp Arg Lys Lys Asn Leu Val Lys Thr Leu Asn Gly Glu
545 550 555 560

Tyr Ile Ala Leu Glu Lys Leu Glu Ser Ile Tyr Arg Ser Asn Pro Tyr
565 570 575

Val Gln Asn Ile Cys Val Tyr Ala Asp Glu Asn Lys Val Lys Pro Val
580 585 590

Gly Ile Val Val Pro Asn Leu Gly His Leu Ser Lys Leu Ala Ile Glu
595 600 605

Leu Gly Ile Met Val Pro Gly Glu Asp Val Glu Ser Tyr Ile His Glu
610 615 620

Lys Lys Leu Gln Asp Ala Val Cys Lys Asp Met Leu Ser Thr Ala Lys
625 630 635 640

Ser Gln Gly Leu Asn Gly Ile Glu Leu Leu Cys Gly Ile Val Phe Phe
645 650 655

Glu Glu Glu Trp Thr Pro Glu Asn Gly Leu Val Thr Ser Ala Gln Lys
660 665 670

Leu Lys Arg Arg Asp Ile Leu Ala Ala Val Lys Pro Asp Val Glu Arg
675 680 685

Val Tyr Lys Glu Asn Thr
690

<210> 201
<211> 6702
<212> DNA
<213> Artificial Sequence

<220>
<223> Mutated acetyl-CoA carboxylase (ACC1) (S659A, S1157A)

<400> 201
atgagcgaag aaagcttatt cgagtcttct ccacagaaga tggagtacga aattacaaac 60

tactcagaaa gacatacaga acttccaggt catttcattg gcctcaatac agtagataaa 120

ctagaggagt ccccgttaag ggactttgtt aagagtcacg gtggtcacac ggtcatatcc 180

aagatcctga tagcaaataa tggtattgcc gccgtgaaag aaattagatc cgtcagaaaa 240

tgggcatacg agacgttcgg cgatgacaga accgtccaat tcgtcgccat ggccacccca 300

gaagatctgg aggccaacgc agaatatatc cgtatggccg atcaatacat tgaagtgcca 360

ggtggtacta ataataacaa ctacgctaac gtagacttga tcgtagacat cgccgaaaga 420

gcagacgtag acgccgtatg ggctggctgg ggtcacgcct ccgagaatcc actattgcct 480

gaaaaattgt cccagtctaa gaggaaagtc atctttattg ggcctccagg taacgccatg 540

aggtctttag gtgataaaat ctcctctacc attgtcgctc aaagtgctaa agtcccatgt 600

attccatggt ctggtaccgg tgttgacacc gttcacgtgg acgagaaaac cggtctggtc 660

tctgtcgacg atgacatcta tcaaaagggt tgttgtacct ctcctgaaga tggtttacaa 720

aaggccaagc gtattggttt tcctgtcatg attaaggcat ccgaaggtgg tggtggtaaa 780

ggtatcagac aagttgaacg tgaagaagat ttcatcgctt tataccacca ggcagccaac 840

gaaattccag gctcccccat tttcatcatg aagttggccg gtagagcgcg tcacttggaa 900

gttcaactgc tagcagatca gtacggtaca aatatttcct tgttcggtag agactgttcc 960

gttcagagac gtcatcaaaa aattatcgaa gaagcaccag ttacaattgc caaggctgaa 1020

acatttcacg agatggaaaa ggctgccgtc agactgggga aactagtcgg ttatgtctct 1080

gccggtaccg tggagtatct atattctcat gatgatggaa aattctactt tttagaattg 1140

aacccaagat tacaagtcga gcatccaaca acggaaatgg tctccggtgt taacttacct 1200

gcagctcaat tacaaatcgc tatgggtatc cctatgcata gaataagtga cattagaact 1260

ttatatggta tgaatcctca ttctgcctca gaaatcgatt tcgaattcaa aactcaagat 1320

gccaccaaga aacaaagaag acctattcca aagggtcatt gtaccgcttg tcgtatcaca 1380

tcagaagatc caaacgatgg attcaagcca tcgggtggta ctttgcatga actaaacttc 1440

cgttcttcct ctaatgtttg gggttacttc tccgtgggta acaatggtaa tattcactcc 1500

ttttcggact ctcagttcgg ccatattttt gcttttggtg aaaatagaca agcttccagg 1560

aaacacatgg ttgttgccct gaaggaattg tccattaggg gtgatttcag aactactgtg 1620

gaatacttga tcaaactttt ggaaactgaa gatttcgagg ataacactat taccaccggt 1680

tggttggacg atttgattac tcataaaatg accgctgaaa agcctgatcc aactcttgcc 1740

gtcatttgcg gtgccgctac aaaggctttc ttagcatctg aagaagcccg ccacaagtat 1800

atcgaatcct tacaaaaggg acaagttcta tctaaagacc tactgcaaac tatgttccct 1860

gtagatttta tccatgaggg taaaagatac aagttcaccg tagctaaatc cggtaatgac 1920

cgttacacat tatttatcaa tggttctaaa tgtgatatca tactgcgtca actatctgat 1980

ggtggtcttt tgattgccat aggcggtaaa tcgcatacca tctattggaa agaagaagtt 2040

gctgctacaa gattatccgt tgactctatg actactttgt tggaagttga aaacgatcca 2100

acccagttgc gtactccatc ccctggtaaa ttggttaaat tcttggtgga aaatggtgaa 2160

cacattatca agggccaacc atatgcagaa attgaagtta tgaaaatgca aatgcctttg 2220

gtttctcaag aaaatggtat cgtccagtta ttaaagcaac ctggttctac cattgttgca 2280

ggtgatatca tggctattat gactcttgac gatccatcca aggtcaagca cgctctacca 2340

tttgaaggta tgctgccaga ttttggttct ccagttatcg aaggaaccaa acctgcctat 2400

aaattcaagt cattagtgtc tactttggaa aacattttga agggttatga caaccaagtt 2460

attatgaacg cttccttgca acaattgata gaggttttga gaaatccaaa actgccttac 2520

tcagaatgga aactacacat ctctgcttta cattcaagat tgcctgctaa gctagatgaa 2580

caaatggaag agttagttgc acgttctttg agacgtggtg ctgttttccc agctagacaa 2640

ttaagtaaat tgattgatat ggccgtgaag aatcctgaat acaaccccga caaattgctg 2700

ggcgccgtcg tggaaccatt ggcggatatt gctcataagt actctaacgg gttagaagcc 2760

catgaacatt ctatatttgt ccatttcttg gaagaatatt acgaagttga aaagttattc 2820

aatggtccaa atgttcgtga ggaaaatatc attctgaaat tgcgtgatga aaaccctaaa 2880

gatctagata aagttgcgct aactgttttg tctcattcga aagtttcagc gaagaataac 2940

ctgatcctag ctatcttgaa acattatcaa ccattgtgca agttatcttc taaagtttct 3000

gccattttct ctactcctct acaacatatt gttgaactag aatctaaggc taccgctaag 3060

gtcgctctac aagcaagaga aattttgatt caaggcgctt taccttcggt caaggaaaga 3120

actgaacaaa ttgaacatat cttaaaatcc tctgttgtga aggttgccta tggctcatcc 3180

aatccaaagc gctctgaacc agatttgaat atcttgaagg acttgatcga ttctaattac 3240

gttgtgttcg atgttttact tcaattccta acccatcaag acccagttgt gactgctgca 3300

gctgctcaag tctatattcg tcgtgcttat cgtgcttaca ccataggaga tattagagtt 3360

cacgaaggtg tcacagttcc aattgttgaa tggaaattcc aactaccttc agctgcgttc 3420

tccacctttc caactgttaa atctaaaatg ggtatgaaca gggctgtttc tgtttcagat 3480

ttgtcatatg ttgcaaacag tcagtcatct ccgttaagag aaggtatttt gatggctgtg 3540

gatcatttag atgatgttga tgaaattttg tcacaaagtt tggaagttat tcctcgtcac 3600

caatcttctt ctaacggacc tgctcctgat cgttctggta gctccgcatc gttgagtaat 3660

gttgctaatg tttgtgttgc ttctacagaa ggtttcgaat ctgaagagga aattttggta 3720

aggttgagag aaattttgga tttgaataag caggaattaa tcaatgcttc tatccgtcgt 3780

atcacattta tgttcggttt taaagatggg tcttatccaa agtattatac ttttaacggt 3840

ccaaattata acgaaaatga aacaattcgt cacattgagc cggctttggc cttccaactg 3900

gaattaggaa gattgtccaa cttcaacatt aaaccaattt tcactgataa tagaaacatc 3960

catgtctacg aagctgttag taagacttct ccattggata agagattctt tacaagaggt 4020

attattagaa cgggtcatat ccgtgatgac atttctattc aagaatatct gacttctgaa 4080

gctaacagat tgatgagtga tatattggat aatttagaag tcaccgacac ttcaaattct 4140

gatttgaatc atatcttcat caacttcatt gcggtgtttg atatctctcc agaagatgtc 4200

gaagccgcct tcggtggttt cttagaaaga tttggtaaga gattgttgag attgcgtgtt 4260

tcttctgccg aaattagaat catcatcaaa gatcctcaaa caggtgcccc agtaccattg 4320

cgtgccttga tcaataacgt ttctggttat gttatcaaaa cagaaatgta caccgaagtc 4380

aagaacgcaa aaggtgaatg ggtatttaag tctttgggta aacctggatc catgcattta 4440

agacctattg ctactcctta ccctgttaag gaatggttgc aaccaaaacg ttataaggca 4500

cacttgatgg gtaccacata tgtctatgac ttcccagaat tattccgcca agcatcgtca 4560

tcccaatgga aaaatttctc tgcagatgtt aagttaacag atgatttctt tatttccaac 4620

gagttgattg aagatgaaaa cggcgaatta actgaggtgg aaagagaacc tggtgccaac 4680

gctattggta tggttgcctt taagattact gtaaagactc ctgaatatcc aagaggccgt 4740

caatttgttg ttgttgctaa cgatatcaca ttcaagatcg gttcctttgg tccacaagaa 4800

gacgaattct tcaataaggt tactgaatat gctagaaagc gtggtatccc aagaatttac 4860

ttggctgcaa actcaggtgc cagaattggt atggctgaag agattgttcc actatttcaa 4920

gttgcatgga atgatgctgc caatccggac aagggcttcc aatacttata cttaacaagt 4980

gaaggtatgg aaactttaaa gaaatttgac aaagaaaatt ctgttctcac tgaacgtact 5040

gttataaacg gtgaagaaag atttgtcatc aagacaatta ttggttctga agatgggtta 5100

ggtgtcgaat gtctacgtgg atctggttta attgctggtg caacgtcaag ggcttaccac 5160

gatatcttca ctatcacctt agtcacttgt agatccgtcg gtatcggtgc ttatttggtt 5220

cgtttgggtc aaagagctat tcaggtcgaa ggccagccaa ttattttaac tggtgctcct 5280

gcaatcaaca aaatgctggg tagagaagtt tatacttcta acttacaatt gggtggtact 5340

caaatcatgt ataacaacgg tgtttcacat ttgactgctg ttgacgattt agctggtgta 5400

gagaagattg ttgaatggat gtcttatgtt ccagccaagc gtaatatgcc agttcctatc 5460

ttggaaacta aagacacatg ggatagacca gttgatttca ctccaactaa tgatgaaact 5520

tacgatgtaa gatggatgat tgaaggtcgt gagactgaaa gtggatttga atatggtttg 5580

tttgataaag ggtctttctt tgaaactttg tcaggatggg ccaaaggtgt tgtcgttggt 5640

agagcccgtc ttggtggtat tccactgggt gttattggtg ttgaaacaag aactgtcgag 5700

aacttgattc ctgctgatcc agctaatcca aatagtgctg aaacattaat tcaagaacct 5760

ggtcaagttt ggcatccaaa ctccgccttc aagactgctc aagctatcaa tgactttaac 5820

aacggtgaac aattgccaat gatgattttg gccaactgga gaggtttctc tggtggtcaa 5880

cgtgatatgt tcaacgaagt cttgaagtat ggttcgttta ttgttgacgc attggtggat 5940

tacaaacaac caattattat ctatatccca cctaccggtg aactaagagg tggttcatgg 6000

gttgttgtcg atccaactat caacgctgac caaatggaaa tgtatgccga cgtcaacgct 6060

agagctggtg ttttggaacc acaaggtatg gttggtatca agttccgtag agaaaaattg 6120

ctggacacca tgaacagatt ggatgacaag tacagagaat tgagatctca attatccaac 6180

aagagtttgg ctccagaagt acatcagcaa atatccaagc aattagctga tcgtgagaga 6240

gaactattgc caatttacgg acaaatcagt cttcaatttg ctgatttgca cgataggtct 6300

tcacgtatgg tggccaaggg tgttatttct aaggaactgg aatggaccga ggcacgtcgt 6360

ttcttcttct ggagattgag aagaagattg aacgaagaat atttgattaa aaggttgagc 6420

catcaggtag gcgaagcatc aagattagaa aagatcgcaa gaattagatc gtggtaccct 6480

gcttcagtgg accatgaaga tgataggcaa gtcgcaacat ggattgaaga aaactacaaa 6540

actttggacg ataaactaaa gggtttgaaa ttagagtcat tcgctcaaga cttagctaaa 6600

aagatcagaa gcgaccatga caatgctatt gatggattat ctgaagttat caagatgtta 6660

tctaccgatg ataaagaaaa attgttgaag actttgaaat aa 6702

<210> 202
<211> 894
<212> DNA
<213> Artificial Sequence

<220>
<223> Truncated geranylgeranyl pyrophosphate synthase Ag_GPPS_Ntrunc

<400> 202
atgtcttttg acttcaataa gtatatggat tctaaggcta tgaccgtcaa cgaggctttg 60

aataaagcca tcccattgcg ttacccacaa aagatctacg aatctatgag atattctttg 120

ttagctggtg gtaagagagt ccgtccagtt ttgtgtatcg ccgcttgtga attagtcggt 180

ggtactgagg agttagctat tccaaccgcc tgtgccatcg aaatgatcca caccatgtct 240

ttgatgcacg atgatttgcc atgtatcgac aacgatgact tgagacgtgg taaacctacc 300

aatcataaga ttttcggtga agatactgct gttactgccg gtaacgcttt acactcttac 360

gccttcgaac atattgctgt ttctacttcc aagactgttg gtgctgatag aattttgaga 420

atggtttctg aattaggtcg tgctactggt tccgaaggtg ttatgggtgg tcaaatggtc 480

gatattgctt ctgaaggtga cccttccatt gatttgcaaa ctttagaatg gatccacatc 540

cacaagactg ctatgttatt agaatgttct gttgtctgtg gtgccatcat cggtggtgct 600

tctgaaattg ttattgagag agccagacgt tatgctcgtt gtgtcggttt attgtttcaa 660

gttgttgacg acattttaga tgttaccaaa tcttctgacg aattgggtaa aactgctggt 720

aaagatttaa tctccgataa agccacctac cctaagttga tgggtttgga gaaggccaaa 780

gagttttccg atgaattatt aaacagagct aaaggtgaat tgtcttgctt cgatccagtt 840

aaggctgccc cattgttagg tttggctgac tacgttgcct tcagacaaaa ctaa 894

<210> 203
<211> 297
<212> PRT
<213> Artificial Sequence

<220>
<223> Truncated geranylgeranyl pyrophosphate synthase Ag_GPPS_Ntrunc

<400> 203

Met Ser Phe Asp Phe Asn Lys Tyr Met Asp Ser Lys Ala Met Thr Val
1 5 10 15

Asn Glu Ala Leu Asn Lys Ala Ile Pro Leu Arg Tyr Pro Gln Lys Ile
20 25 30

Tyr Glu Ser Met Arg Tyr Ser Leu Leu Ala Gly Gly Lys Arg Val Arg
35 40 45

Pro Val Leu Cys Ile Ala Ala Cys Glu Leu Val Gly Gly Thr Glu Glu
50 55 60

Leu Ala Ile Pro Thr Ala Cys Ala Ile Glu Met Ile His Thr Met Ser
65 70 75 80

Leu Met His Asp Asp Leu Pro Cys Ile Asp Asn Asp Asp Leu Arg Arg
85 90 95

Gly Lys Pro Thr Asn His Lys Ile Phe Gly Glu Asp Thr Ala Val Thr
100 105 110

Ala Gly Asn Ala Leu His Ser Tyr Ala Phe Glu His Ile Ala Val Ser
115 120 125

Thr Ser Lys Thr Val Gly Ala Asp Arg Ile Leu Arg Met Val Ser Glu

130 135 140

Leu Gly Arg Ala Thr Gly Ser Glu Gly Val Met Gly Gly Gln Met Val
145 150 155 160

Asp Ile Ala Ser Glu Gly Asp Pro Ser Ile Asp Leu Gln Thr Leu Glu
165 170 175

Trp Ile His Ile His Lys Thr Ala Met Leu Leu Glu Cys Ser Val Val
180 185 190

Cys Gly Ala Ile Ile Gly Gly Ala Ser Glu Ile Val Ile Glu Arg Ala
195 200 205

Arg Arg Tyr Ala Arg Cys Val Gly Leu Leu Phe Gln Val Val Asp Asp
210 215 220

Ile Leu Asp Val Thr Lys Ser Ser Asp Glu Leu Gly Lys Thr Ala Gly
225 230 235 240

Lys Asp Leu Ile Ser Asp Lys Ala Thr Tyr Pro Lys Leu Met Gly Leu
245 250 255

Glu Lys Ala Lys Glu Phe Ser Asp Glu Leu Leu Asn Arg Ala Lys Gly
260 265 270

Glu Leu Ser Cys Phe Asp Pro Val Lys Ala Ala Pro Leu Leu Gly Leu
275 280 285

Ala Asp Tyr Val Ala Phe Arg Gln Asn
290 295

<210> 204
<211> 1356
<212> DNA
<213> Saccharomyces cerevisiae

<400> 204
ttatttatca agataagttt ccggatcttt ttctttccta acaccccagt cagcctgagt 60

tacatccagc cattgaacct tagaaaatct tttgtcatca gcggtttgag ccctaagatc 120

aacatcttgc ttagcaatca ctgcaatggc gtcataacca ccagcaccag gtattaagca 180

agtaagaact ccttttaagg tctggcaatc atccaataag ctagtttgta cgggaggttc 240

gatatcggca ccagattctt tagttatttt tctaaaggaa cgtctaattg tggcaactgc 300

atctctaact tctgtgatct caggatactt ttgacaggta cagtcattcc tctcaagaga 360

ctcaaatatc tgatcgctgt aatcgtcatg agtctcgtgt aagcgatcta gtttagatag 420

tccatccata aatctagaat ttgcatgatc gagttctgta tatattttca agctttccgg 480

catatgcgaa tcataccaat tttttacctt ctggaccagt tttactgttt ctgaaccatt 540

cttaatatcg cccatccata aagttaatcc cgaaggtaaa tggttacttt taatcgttat 600

attccagtct tcttcattaa ccaaatgcgc cagtttactg ccgtaagtag cacttccaat 660

atctggcaaa ttagagatta atgcgggtgg gaatcttcta tatctgatag atccatatgc 720

tgccgccgct acatcaaacc cgcttccaat tttaccctga gcttgacaat gagcaacttg 780

tgataaatta tgaataactt ctctatattt gtctacatta ttttccaggt ccgatacaaa 840

aaaggaggcc aaagctgtag ttaaaactgt gactaaacct gccgaggagc ccagccctgt 900

tttgggaact tcttcaattc tgtgcgaatg aaaactcaat cttctgttgc cacgatgttc 960

ggtaacgctg tcctcctgag aatggtaggc atcatcagag aaaatatcaa taacgaacaa 1020

gtttctattg cagtagtcgt ccatgttagg cttaaagtag ctaaatacgt tagcgataac 1080

tttttcaatg aaagggttct tagatccgcc tatcgaaaca ggaatgaagc cagttttagg 1140

acttatatgg tacagccact ccccatcttt aaattgttta cttttcacac gcacttcaaa 1200

cttatcagac tcttgcaatg aaccgtaagg atgggctaca gcatgcattc ttgccgataa 1260

tccgactaca aatgcttcat atttcggatc taaaactaaa tatccaccag ctagtaacgc 1320

tttccctggg gcactgaagg ctctcaactc tgacat 1356

<210> 205
<211> 451
<212> PRT
<213> Saccharomyces cerevisiae

<400> 205

Met Ser Glu Leu Arg Ala Phe Ser Ala Pro Gly Lys Ala Leu Leu Ala

1 5 10 15

Gly Gly Tyr Leu Val Leu Asp Pro Lys Tyr Glu Ala Phe Val Val Gly
20 25 30

Leu Ser Ala Arg Met His Ala Val Ala His Pro Tyr Gly Ser Leu Gln
35 40 45

Glu Ser Asp Lys Phe Glu Val Arg Val Lys Ser Lys Gln Phe Lys Asp
50 55 60

Gly Glu Trp Leu Tyr His Ile Ser Pro Lys Thr Gly Phe Ile Pro Val
65 70 75 80

Ser Ile Gly Gly Ser Lys Asn Pro Phe Ile Glu Lys Val Ile Ala Asn
85 90 95

Val Phe Ser Tyr Phe Lys Pro Asn Met Asp Asp Tyr Cys Asn Arg Asn
100 105 110

Leu Phe Val Ile Asp Ile Phe Ser Asp Asp Ala Tyr His Ser Gln Glu
115 120 125

Asp Ser Val Thr Glu His Arg Gly Asn Arg Arg Leu Ser Phe His Ser
130 135 140

His Arg Ile Glu Glu Val Pro Lys Thr Gly Leu Gly Ser Ser Ala Gly
145 150 155 160

Leu Val Thr Val Leu Thr Thr Ala Leu Ala Ser Phe Phe Val Ser Asp
165 170 175

Leu Glu Asn Asn Val Asp Lys Tyr Arg Glu Val Ile His Asn Leu Ser
180 185 190

Gln Val Ala His Cys Gln Ala Gln Gly Lys Ile Gly Ser Gly Phe Asp
195 200 205

Val Ala Ala Ala Ala Tyr Gly Ser Ile Arg Tyr Arg Arg Phe Pro Pro

210 215 220

Ala Leu Ile Ser Asn Leu Pro Asp Ile Gly Ser Ala Thr Tyr Gly Ser
225 230 235 240

Lys Leu Ala His Leu Val Asn Glu Glu Asp Trp Asn Ile Thr Ile Lys
245 250 255

Ser Asn His Leu Pro Ser Gly Leu Thr Leu Trp Met Gly Asp Ile Lys
260 265 270

Asn Gly Ser Glu Thr Val Lys Leu Val Gln Lys Val Lys Asn Trp Tyr
275 280 285

Asp Ser His Met Pro Glu Ser Leu Lys Ile Tyr Thr Glu Leu Asp His
290 295 300

Ala Asn Ser Arg Phe Met Asp Gly Leu Ser Lys Leu Asp Arg Leu His
305 310 315 320

Glu Thr His Asp Asp Tyr Ser Asp Gln Ile Phe Glu Ser Leu Glu Arg
325 330 335

Asn Asp Cys Thr Cys Gln Lys Tyr Pro Glu Ile Thr Glu Val Arg Asp
340 345 350

Ala Val Ala Thr Ile Arg Arg Ser Phe Arg Lys Ile Thr Lys Glu Ser
355 360 365

Gly Ala Asp Ile Glu Pro Pro Val Gln Thr Ser Leu Leu Asp Asp Cys
370 375 380

Gln Thr Leu Lys Gly Val Leu Thr Cys Leu Ile Pro Gly Ala Gly Gly
385 390 395 400

Tyr Asp Ala Ile Ala Val Ile Ala Lys Gln Asp Val Asp Leu Arg Ala
405 410 415

Gln Thr Ala Asp Asp Lys Arg Phe Ser Lys Val Gln Trp Leu Asp Val

420 425 430

Thr Gln Ala Asp Trp Gly Val Arg Lys Glu Lys Asp Pro Glu Thr Tyr
435 440 445

Leu Asp Lys
450

<210> 206
<211> 1332
<212> DNA
<213> Saccharomyces cerevisiae

<400> 206
atgtcattac cgttcttaac ttctgcaccg ggaaaggtta ttatttttgg tgaacactct 60

gctgtgtaca acaagcctgc cgtcgctgct agtgtgtctg cgttgagaac ctacctgcta 120

ataagcgagt catctgcacc agatactatt gaattggact tcccggacat tagctttaat 180

cataagtggt ccatcaatga tttcaatgcc atcaccgagg atcaagtaaa ctcccaaaaa 240

ttggccaagg ctcaacaagc caccgatggc ttgtctcagg aactcgttag tcttttggat 300

ccgttgttag ctcaactatc cgaatccttc cactaccatg cagcgttttg tttcctgtat 360

atgtttgttt gcctatgccc ccatgccaag aatattaagt tttctttaaa gtctacttta 420

cccatcggtg ctgggttggg ctcaagcgcc tctatttctg tatcactggc cttagctatg 480

gcctacttgg gggggttaat aggatctaat gacttggaaa agctgtcaga aaacgataag 540

catatagtga atcaatgggc cttcataggt gaaaagtgta ttcacggtac cccttcagga 600

atagataacg ctgtggccac ttatggtaat gccctgctat ttgaaaaaga ctcacataat 660

ggaacaataa acacaaacaa ttttaagttc ttagatgatt tcccagccat tccaatgatc 720

ctaacctata ctagaattcc aaggtctaca aaagatcttg ttgctcgcgt tcgtgtgttg 780

gtcaccgaga aatttcctga agttatgaag ccaattctag atgccatggg tgaatgtgcc 840

ctacaaggct tagagatcat gactaagtta agtaaatgta aaggcaccga tgacgaggct 900

gtagaaacta ataatgaact gtatgaacaa ctattggaat tgataagaat aaatcatgga 960

ctgcttgtct caatcggtgt ttctcatcct ggattagaac ttattaaaaa tctgagcgat 1020

gatttgagaa ttggctccac aaaacttacc ggtgctggtg gcggcggttg ctctttgact 1080

ttgttacgaa gagacattac tcaagagcaa attgacagtt tcaaaaagaa attgcaagat 1140

gattttagtt acgagacatt tgaaacagac ttgggtggga ctggctgctg tttgttaagc 1200

gcaaaaaatt tgaataaaga tcttaaaatc aaatccctag tattccaatt atttgaaaat 1260

aaaactacca caaagcaaca aattgacgat ctattattgc caggaaacac gaatttacca 1320

tggacttcat aa 1332

<210> 207
<211> 2233
<212> PRT
<213> Artificial Sequence

<220>
<223> Mutated acetyl-CoA carboxylase (ACC1) (S659A, S1157A)

<400> 207

Met Ser Glu Glu Ser Leu Phe Glu Ser Ser Pro Gln Lys Met Glu Tyr
1 5 10 15

Glu Ile Thr Asn Tyr Ser Glu Arg His Thr Glu Leu Pro Gly His Phe
20 25 30

Ile Gly Leu Asn Thr Val Asp Lys Leu Glu Glu Ser Pro Leu Arg Asp
35 40 45

Phe Val Lys Ser His Gly Gly His Thr Val Ile Ser Lys Ile Leu Ile
50 55 60

Ala Asn Asn Gly Ile Ala Ala Val Lys Glu Ile Arg Ser Val Arg Lys
65 70 75 80

Trp Ala Tyr Glu Thr Phe Gly Asp Asp Arg Thr Val Gln Phe Val Ala
85 90 95

Met Ala Thr Pro Glu Asp Leu Glu Ala Asn Ala Glu Tyr Ile Arg Met
100 105 110

Ala Asp Gln Tyr Ile Glu Val Pro Gly Gly Thr Asn Asn Asn Asn Tyr
115 120 125

```
Ala Asn Val Asp Leu Ile Val Asp Ile Ala Glu Arg Ala Asp Val Asp
    130                 135                 140

Ala Val Trp Ala Gly Trp Gly His Ala Ser Glu Asn Pro Leu Leu Pro
145                 150                 155                 160

Glu Lys Leu Ser Gln Ser Lys Arg Lys Val Ile Phe Ile Gly Pro Pro
                165                 170                 175

Gly Asn Ala Met Arg Ser Leu Gly Asp Lys Ile Ser Ser Thr Ile Val
            180                 185                 190

Ala Gln Ser Ala Lys Val Pro Cys Ile Pro Trp Ser Gly Thr Gly Val
        195                 200                 205

Asp Thr Val His Val Asp Glu Lys Thr Gly Leu Val Ser Val Asp Asp
    210                 215                 220

Asp Ile Tyr Gln Lys Gly Cys Cys Thr Ser Pro Glu Asp Gly Leu Gln
225                 230                 235                 240

Lys Ala Lys Arg Ile Gly Phe Pro Val Met Ile Lys Ala Ser Glu Gly
                245                 250                 255

Gly Gly Gly Lys Gly Ile Arg Gln Val Glu Arg Glu Glu Asp Phe Ile
            260                 265                 270

Ala Leu Tyr His Gln Ala Ala Asn Glu Ile Pro Gly Ser Pro Ile Phe
        275                 280                 285

Ile Met Lys Leu Ala Gly Arg Ala Arg His Leu Glu Val Gln Leu Leu
    290                 295                 300

Ala Asp Gln Tyr Gly Thr Asn Ile Ser Leu Phe Gly Arg Asp Cys Ser
305                 310                 315                 320

Val Gln Arg Arg His Gln Lys Ile Ile Glu Glu Ala Pro Val Thr Ile
                325                 330                 335
```

```
Ala Lys Ala Glu Thr Phe His Glu Met Glu Lys Ala Ala Val Arg Leu
            340                 345                 350

Gly Lys Leu Val Gly Tyr Val Ser Ala Gly Thr Val Glu Tyr Leu Tyr
        355                 360                 365

Ser His Asp Asp Gly Lys Phe Tyr Phe Leu Glu Leu Asn Pro Arg Leu
    370                 375                 380

Gln Val Glu His Pro Thr Thr Glu Met Val Ser Gly Val Asn Leu Pro
385                 390                 395                 400

Ala Ala Gln Leu Gln Ile Ala Met Gly Ile Pro Met His Arg Ile Ser
                405                 410                 415

Asp Ile Arg Thr Leu Tyr Gly Met Asn Pro His Ser Ala Ser Glu Ile
            420                 425                 430

Asp Phe Glu Phe Lys Thr Gln Asp Ala Thr Lys Lys Gln Arg Arg Pro
        435                 440                 445

Ile Pro Lys Gly His Cys Thr Ala Cys Arg Ile Thr Ser Glu Asp Pro
    450                 455                 460

Asn Asp Gly Phe Lys Pro Ser Gly Gly Thr Leu His Glu Leu Asn Phe
465                 470                 475                 480

Arg Ser Ser Ser Asn Val Trp Gly Tyr Phe Ser Val Gly Asn Asn Gly
                485                 490                 495

Asn Ile His Ser Phe Ser Asp Ser Gln Phe Gly His Ile Phe Ala Phe
            500                 505                 510

Gly Glu Asn Arg Gln Ala Ser Arg Lys His Met Val Val Ala Leu Lys
        515                 520                 525

Glu Leu Ser Ile Arg Gly Asp Phe Arg Thr Thr Val Glu Tyr Leu Ile
    530                 535                 540
```

```
Lys Leu Leu Glu Thr Glu Asp Phe Glu Asp Asn Thr Ile Thr Thr Gly
545                 550                 555                 560

Trp Leu Asp Asp Leu Ile Thr His Lys Met Thr Ala Glu Lys Pro Asp
                565                 570                 575

Pro Thr Leu Ala Val Ile Cys Gly Ala Ala Thr Lys Ala Phe Leu Ala
            580                 585                 590

Ser Glu Glu Ala Arg His Lys Tyr Ile Glu Ser Leu Gln Lys Gly Gln
        595                 600                 605

Val Leu Ser Lys Asp Leu Leu Gln Thr Met Phe Pro Val Asp Phe Ile
    610                 615                 620

His Glu Gly Lys Arg Tyr Lys Phe Thr Val Ala Lys Ser Gly Asn Asp
625                 630                 635                 640

Arg Tyr Thr Leu Phe Ile Asn Gly Ser Lys Cys Asp Ile Ile Leu Arg
                645                 650                 655

Gln Leu Ser Asp Gly Gly Leu Leu Ile Ala Ile Gly Gly Lys Ser His
            660                 665                 670

Thr Ile Tyr Trp Lys Glu Glu Val Ala Ala Thr Arg Leu Ser Val Asp
        675                 680                 685

Ser Met Thr Thr Leu Leu Glu Val Glu Asn Asp Pro Thr Gln Leu Arg
    690                 695                 700

Thr Pro Ser Pro Gly Lys Leu Val Lys Phe Leu Val Glu Asn Gly Glu
705                 710                 715                 720

His Ile Ile Lys Gly Gln Pro Tyr Ala Glu Ile Glu Val Met Lys Met
                725                 730                 735

Gln Met Pro Leu Val Ser Gln Glu Asn Gly Ile Val Gln Leu Leu Lys
            740                 745                 750
```

```
Gln Pro Gly Ser Thr Ile Val Ala Gly Asp Ile Met Ala Ile Met Thr
        755                 760                 765

Leu Asp Asp Pro Ser Lys Val Lys His Ala Leu Pro Phe Glu Gly Met
    770                 775                 780

Leu Pro Asp Phe Gly Ser Pro Val Ile Glu Gly Thr Lys Pro Ala Tyr
785                 790                 795                 800

Lys Phe Lys Ser Leu Val Ser Thr Leu Glu Asn Ile Leu Lys Gly Tyr
                805                 810                 815

Asp Asn Gln Val Ile Met Asn Ala Ser Leu Gln Gln Leu Ile Glu Val
            820                 825                 830

Leu Arg Asn Pro Lys Leu Pro Tyr Ser Glu Trp Lys Leu His Ile Ser
        835                 840                 845

Ala Leu His Ser Arg Leu Pro Ala Lys Leu Asp Glu Gln Met Glu Glu
    850                 855                 860

Leu Val Ala Arg Ser Leu Arg Arg Gly Ala Val Phe Pro Ala Arg Gln
865                 870                 875                 880

Leu Ser Lys Leu Ile Asp Met Ala Val Lys Asn Pro Glu Tyr Asn Pro
                885                 890                 895

Asp Lys Leu Leu Gly Ala Val Val Glu Pro Leu Ala Asp Ile Ala His
            900                 905                 910

Lys Tyr Ser Asn Gly Leu Glu Ala His Glu His Ser Ile Phe Val His
        915                 920                 925

Phe Leu Glu Glu Tyr Tyr Glu Val Glu Lys Leu Phe Asn Gly Pro Asn
    930                 935                 940

Val Arg Glu Glu Asn Ile Ile Leu Lys Leu Arg Asp Glu Asn Pro Lys
945                 950                 955                 960
```

```
Asp Leu Asp Lys Val Ala Leu Thr Val Leu Ser His Ser Lys Val Ser
               965                 970                 975

Ala Lys Asn Asn Leu Ile Leu Ala Ile Leu Lys His Tyr Gln Pro Leu
           980                 985                 990

Cys Lys Leu Ser Ser Lys Val Ser  Ala Ile Phe Ser Thr  Pro Leu Gln
        995                 1000                1005

His Ile  Val Glu Leu Glu Ser  Lys Ala Thr Ala Lys  Val Ala Leu
    1010                1015                1020

Gln Ala  Arg Glu Ile Leu Ile  Gln Gly Ala Leu Pro  Ser Val Lys
    1025                1030                1035

Glu Arg  Thr Glu Gln Ile Glu  His Ile Leu Lys Ser  Ser Val Val
    1040                1045                1050

Lys Val  Ala Tyr Gly Ser Ser  Asn Pro Lys Arg Ser  Glu Pro Asp
    1055                1060                1065

Leu Asn  Ile Leu Lys Asp Leu  Ile Asp Ser Asn Tyr  Val Val Phe
    1070                1075                1080

Asp Val  Leu Leu Gln Phe Leu  Thr His Gln Asp Pro  Val Val Thr
    1085                1090                1095

Ala Ala  Ala Ala Gln Val Tyr  Ile Arg Arg Ala Tyr  Arg Ala Tyr
    1100                1105                1110

Thr Ile  Gly Asp Ile Arg Val  His Glu Gly Val Thr  Val Pro Ile
    1115                1120                1125

Val Glu  Trp Lys Phe Gln Leu  Pro Ser Ala Ala Phe  Ser Thr Phe
    1130                1135                1140

Pro Thr  Val Lys Ser Lys Met  Gly Met Asn Arg Ala  Val Ser Val
    1145                1150                1155
```

```
Ser Asp  Leu Ser Tyr Val Ala  Asn Ser Gln Ser Ser  Pro Leu Arg
    1160                1165                1170

Glu Gly  Ile Leu Met Ala Val  Asp His Leu Asp Asp  Val Asp Glu
    1175                1180                1185

Ile Leu  Ser Gln Ser Leu Glu  Val Ile Pro Arg His  Gln Ser Ser
    1190                1195                1200

Ser Asn  Gly Pro Ala Pro Asp  Arg Ser Gly Ser Ser  Ala Ser Leu
    1205                1210                1215

Ser Asn  Val Ala Asn Val Cys  Val Ala Ser Thr Glu  Gly Phe Glu
    1220                1225                1230

Ser Glu  Glu Glu Ile Leu Val  Arg Leu Arg Glu Ile  Leu Asp Leu
    1235                1240                1245

Asn Lys  Gln Glu Leu Ile Asn  Ala Ser Ile Arg Arg  Ile Thr Phe
    1250                1255                1260

Met Phe  Gly Phe Lys Asp Gly  Ser Tyr Pro Lys Tyr  Tyr Thr Phe
    1265                1270                1275

Asn Gly  Pro Asn Tyr Asn Glu  Asn Glu Thr Ile Arg  His Ile Glu
    1280                1285                1290

Pro Ala  Leu Ala Phe Gln Leu  Glu Leu Gly Arg Leu  Ser Asn Phe
    1295                1300                1305

Asn Ile  Lys Pro Ile Phe Thr  Asp Asn Arg Asn Ile  His Val Tyr
    1310                1315                1320

Glu Ala  Val Ser Lys Thr Ser  Pro Leu Asp Lys Arg  Phe Phe Thr
    1325                1330                1335

Arg Gly  Ile Ile Arg Thr Gly  His Ile Arg Asp Asp  Ile Ser Ile
    1340                1345                1350
```

```
Gln Glu  Tyr Leu Thr Ser Glu  Ala Asn Arg Leu Met  Ser Asp Ile
    1355                1360                1365

Leu Asp  Asn Leu Glu Val Thr  Asp Thr Ser Asn Ser  Asp Leu Asn
    1370                1375                1380

His Ile  Phe Ile Asn Phe Ile  Ala Val Phe Asp Ile  Ser Pro Glu
    1385                1390                1395

Asp Val  Glu Ala Ala Phe Gly  Gly Phe Leu Glu Arg  Phe Gly Lys
    1400                1405                1410

Arg Leu  Leu Arg Leu Arg Val  Ser Ser Ala Glu Ile  Arg Ile Ile
    1415                1420                1425

Ile Lys  Asp Pro Gln Thr Gly  Ala Pro Val Pro Leu  Arg Ala Leu
    1430                1435                1440

Ile Asn  Asn Val Ser Gly Tyr  Val Ile Lys Thr Glu  Met Tyr Thr
    1445                1450                1455

Glu Val  Lys Asn Ala Lys Gly  Glu Trp Val Phe Lys  Ser Leu Gly
    1460                1465                1470

Lys Pro  Gly Ser Met His Leu  Arg Pro Ile Ala Thr  Pro Tyr Pro
    1475                1480                1485

Val Lys  Glu Trp Leu Gln Pro  Lys Arg Tyr Lys Ala  His Leu Met
    1490                1495                1500

Gly Thr  Thr Tyr Val Tyr Asp  Phe Pro Glu Leu Phe  Arg Gln Ala
    1505                1510                1515

Ser Ser  Ser Gln Trp Lys Asn  Phe Ser Ala Asp Val  Lys Leu Thr
    1520                1525                1530

Asp Asp  Phe Phe Ile Ser Asn  Glu Leu Ile Glu Asp  Glu Asn Gly
    1535                1540                1545
```

```
Glu Leu  Thr Glu Val Glu Arg  Glu Pro Gly Ala Asn  Ala Ile Gly
    1550                1555                1560

Met Val  Ala Phe Lys Ile Thr  Val Lys Thr Pro Glu  Tyr Pro Arg
    1565                1570                1575

Gly Arg  Gln Phe Val Val Val  Ala Asn Asp Ile Thr  Phe Lys Ile
    1580                1585                1590

Gly Ser  Phe Gly Pro Gln Glu  Asp Glu Phe Phe Asn  Lys Val Thr
    1595                1600                1605

Glu Tyr  Ala Arg Lys Arg Gly  Ile Pro Arg Ile Tyr  Leu Ala Ala
    1610                1615                1620

Asn Ser  Gly Ala Arg Ile Gly  Met Ala Glu Glu Ile  Val Pro Leu
    1625                1630                1635

Phe Gln  Val Ala Trp Asn Asp  Ala Ala Asn Pro Asp  Lys Gly Phe
    1640                1645                1650

Gln Tyr  Leu Tyr Leu Thr Ser  Glu Gly Met Glu Thr  Leu Lys Lys
    1655                1660                1665

Phe Asp  Lys Glu Asn Ser Val  Leu Thr Glu Arg Thr  Val Ile Asn
    1670                1675                1680

Gly Glu  Glu Arg Phe Val Ile  Lys Thr Ile Ile Gly  Ser Glu Asp
    1685                1690                1695

Gly Leu  Gly Val Glu Cys Leu  Arg Gly Ser Gly Leu  Ile Ala Gly
    1700                1705                1710

Ala Thr  Ser Arg Ala Tyr His  Asp Ile Phe Thr Ile  Thr Leu Val
    1715                1720                1725

Thr Cys  Arg Ser Val Gly Ile  Gly Ala Tyr Leu Val  Arg Leu Gly
    1730                1735                1740
```

```
Gln Arg  Ala Ile Gln Val Glu  Gly Gln Pro Ile Ile  Leu Thr Gly
    1745                1750                1755

Ala Pro  Ala Ile Asn Lys Met  Leu Gly Arg Glu Val  Tyr Thr Ser
    1760                1765                1770

Asn Leu  Gln Leu Gly Gly Thr  Gln Ile Met Tyr Asn  Asn Gly Val
    1775                1780                1785

Ser His  Leu Thr Ala Val Asp  Asp Leu Ala Gly Val  Glu Lys Ile
    1790                1795                1800

Val Glu  Trp Met Ser Tyr Val  Pro Ala Lys Arg Asn  Met Pro Val
    1805                1810                1815

Pro Ile  Leu Glu Thr Lys Asp  Thr Trp Asp Arg Pro  Val Asp Phe
    1820                1825                1830

Thr Pro  Thr Asn Asp Glu Thr  Tyr Asp Val Arg Trp  Met Ile Glu
    1835                1840                1845

Gly Arg  Glu Thr Glu Ser Gly  Phe Glu Tyr Gly Leu  Phe Asp Lys
    1850                1855                1860

Gly Ser  Phe Phe Glu Thr Leu  Ser Gly Trp Ala Lys  Gly Val Val
    1865                1870                1875

Val Gly  Arg Ala Arg Leu Gly  Gly Ile Pro Leu Gly  Val Ile Gly
    1880                1885                1890

Val Glu  Thr Arg Thr Val Glu  Asn Leu Ile Pro Ala  Asp Pro Ala
    1895                1900                1905

Asn Pro  Asn Ser Ala Glu Thr  Leu Ile Gln Glu Pro  Gly Gln Val
    1910                1915                1920

Trp His  Pro Asn Ser Ala Phe  Lys Thr Ala Gln Ala  Ile Asn Asp
    1925                1930                1935
```

```
Phe Asn  Asn Gly Glu Gln Leu  Pro Met Met Ile Leu  Ala Asn Trp
    1940                1945                1950

Arg Gly  Phe Ser Gly Gly Gln  Arg Asp Met Phe Asn  Glu Val Leu
    1955                1960                1965

Lys Tyr  Gly Ser Phe Ile Val  Asp Ala Leu Val Asp  Tyr Lys Gln
    1970                1975                1980

Pro Ile  Ile Ile Tyr Ile Pro  Pro Thr Gly Glu Leu  Arg Gly Gly
    1985                1990                1995

Ser Trp  Val Val Val Asp Pro  Thr Ile Asn Ala Asp  Gln Met Glu
    2000                2005                2010

Met Tyr  Ala Asp Val Asn Ala  Arg Ala Gly Val Leu  Glu Pro Gln
    2015                2020                2025

Gly Met  Val Gly Ile Lys Phe  Arg Arg Glu Lys Leu  Leu Asp Thr
    2030                2035                2040

Met Asn  Arg Leu Asp Asp Lys  Tyr Arg Glu Leu Arg  Ser Gln Leu
    2045                2050                2055

Ser Asn  Lys Ser Leu Ala Pro  Glu Val His Gln Gln  Ile Ser Lys
    2060                2065                2070

Gln Leu  Ala Asp Arg Glu Arg  Glu Leu Leu Pro Ile  Tyr Gly Gln
    2075                2080                2085

Ile Ser  Leu Gln Phe Ala Asp  Leu His Asp Arg Ser  Ser Arg Met
    2090                2095                2100

Val Ala  Lys Gly Val Ile Ser  Lys Glu Leu Glu Trp  Thr Glu Ala
    2105                2110                2115

Arg Arg  Phe Phe Phe Trp Arg  Leu Arg Arg Arg Leu  Asn Glu Glu
    2120                2125                2130
```

```
Tyr Leu  Ile Lys Arg Leu Ser  His Gln Val Gly Glu  Ala Ser Arg
    2135                2140                2145


Leu Glu  Lys Ile Ala Arg Ile  Arg Ser Trp Tyr Pro  Ala Ser Val
    2150                2155                2160


Asp His  Glu Asp Asp Arg Gln  Val Ala Thr Trp Ile  Glu Glu Asn
    2165                2170                2175


Tyr Lys  Thr Leu Asp Asp Lys  Leu Lys Gly Leu Lys  Leu Glu Ser
    2180                2185                2190


Phe Ala  Gln Asp Leu Ala Lys  Lys Ile Arg Ser Asp  His Asp Asn
    2195                2200                2205


Ala Ile  Asp Gly Leu Ser Glu  Val Ile Lys Met Leu  Ser Thr Asp
    2210                2215                2220


Asp Lys  Glu Lys Leu Leu Lys  Thr Leu Lys
    2225                2230


<210>  208
<211>  524
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Truncated 3-hydroxy-3-methyl-glutaryl-CoA reductase (Sc_tHMG1)

<400>  208

Met Gln Leu Val Lys Thr Glu Val Thr Lys Lys Ser Phe Thr Ala Pro
1               5                  10                  15


Val Gln Lys Ala Ser Thr Pro Val Leu Thr Asn Lys Thr Val Ile Ser
            20                  25                  30


Gly Ser Lys Val Lys Ser Leu Ser Ser Ala Gln Ser Ser Ser Ser Gly
        35                  40                  45
```

```
Pro Ser Ser Ser Ser Glu Glu Asp Asp Ser Arg Asp Ile Glu Ser Leu
    50                  55                  60

Asp Lys Lys Ile Arg Pro Leu Glu Glu Leu Glu Ala Leu Leu Ser Ser
65                  70                  75                  80

Gly Asn Thr Lys Gln Leu Lys Asn Lys Glu Val Ala Ala Leu Val Ile
                85                  90                  95

His Gly Lys Leu Pro Leu Tyr Ala Leu Glu Lys Lys Leu Gly Asp Thr
            100                 105                 110

Thr Arg Ala Val Ala Val Arg Arg Lys Ala Leu Ser Ile Leu Ala Glu
        115                 120                 125

Ala Pro Val Leu Ala Ser Asp Arg Leu Pro Tyr Lys Asn Tyr Asp Tyr
    130                 135                 140

Asp Arg Val Phe Gly Ala Cys Cys Glu Asn Val Ile Gly Tyr Met Pro
145                 150                 155                 160

Leu Pro Val Gly Val Ile Gly Pro Leu Val Ile Asp Gly Thr Ser Tyr
                165                 170                 175

His Ile Pro Met Ala Thr Thr Glu Gly Cys Leu Val Ala Ser Ala Met
            180                 185                 190

Arg Gly Cys Lys Ala Ile Asn Ala Gly Gly Gly Ala Thr Thr Val Leu
        195                 200                 205

Thr Lys Asp Gly Met Thr Arg Gly Pro Val Val Arg Phe Pro Thr Leu
    210                 215                 220

Lys Arg Ser Gly Ala Cys Lys Ile Trp Leu Asp Ser Glu Glu Gly Gln
225                 230                 235                 240

Asn Ala Ile Lys Lys Ala Phe Asn Ser Thr Ser Arg Phe Ala Arg Leu
                245                 250                 255
```

```
Gln His Ile Gln Thr Cys Leu Ala Gly Asp Leu Leu Phe Met Arg Phe
            260                 265                 270

Arg Thr Thr Thr Gly Asp Ala Met Gly Met Asn Met Ile Ser Lys Gly
        275                 280                 285

Val Glu Tyr Ser Leu Lys Gln Met Val Glu Glu Tyr Gly Trp Glu Asp
    290                 295                 300

Met Glu Val Val Ser Val Ser Gly Asn Tyr Cys Thr Asp Lys Lys Pro
305                 310                 315                 320

Ala Ala Ile Asn Trp Ile Glu Gly Arg Gly Lys Ser Val Val Ala Glu
                325                 330                 335

Ala Thr Ile Pro Gly Asp Val Val Arg Lys Val Leu Lys Ser Asp Val
            340                 345                 350

Ser Ala Leu Val Glu Leu Asn Ile Ala Lys Asn Leu Val Gly Ser Ala
        355                 360                 365

Met Ala Gly Ser Val Gly Gly Phe Asn Ala His Ala Ala Asn Leu Val
    370                 375                 380

Thr Ala Val Phe Leu Ala Leu Gly Gln Asp Pro Ala Gln Asn Val Glu
385                 390                 395                 400

Ser Ser Asn Cys Ile Thr Leu Met Lys Glu Val Asp Gly Asp Leu Arg
                405                 410                 415

Ile Ser Val Ser Met Pro Ser Ile Glu Val Gly Thr Ile Gly Gly Gly
            420                 425                 430

Thr Val Leu Glu Pro Gln Gly Ala Met Leu Asp Leu Leu Gly Val Arg
        435                 440                 445

Gly Pro His Ala Thr Ala Pro Gly Thr Asn Ala Arg Gln Leu Ala Arg
    450                 455                 460
```

Ile Val Ala Cys Ala Val Leu Ala Gly Glu Leu Ser Leu Cys Ala Ala
465 470 475 480

Leu Ala Ala Gly His Leu Val Gln Ser His Met Thr His Asn Arg Lys
485 490 495

Pro Ala Glu Pro Thr Lys Pro Asn Asn Leu Asp Ala Thr Asp Ile Asn
500 505 510

Arg Leu Lys Asp Gly Ser Val Thr Cys Ile Lys Ser
515 520

<210> 209
<211> 1197
<212> DNA
<213> Saccharomyces cerevisiae

<400> 209
atgtctcaga acgtttacat tgtatcgact gccagaaccc caattggttc attccagggt 60

tctctatcct ccaagacagc agtggaattg ggtgctgttg ctttaaaagg cgccttggct 120

aaggttccag aattggatgc atccaaggat tttgacgaaa ttatttttgg taacgttctt 180

tctgccaatt tgggccaagc tccggccaga caagttgctt tggctgccgg tttgagtaat 240

catatcgttg caagcacagt taacaaggtc tgtgcatccg ctatgaaggc aatcattttg 300

ggtgctcaat ccatcaaatg tggtaatgct gatgttgtcg tagctggtgg ttgtgaatct 360

atgactaacg caccatacta catgccagca gcccgtgcgg gtgccaaatt tggccaaact 420

gttcttgttg atggtgtcga aagagatggg ttgaacgatg cgtacgatgg tctagccatg 480

ggtgtacacg cagaaaagtg tgcccgtgat tgggatatta ctagagaaca acaagacaat 540

tttgccatcg aatcctacca aaaatctcaa aaatctcaaa aggaaggtaa attcgacaat 600

gaaattgtac ctgttaccat taagggattt agaggtaagc ctgatactca agtcacgaag 660

gacgaggaac ctgctagatt acacgttgaa aaattgagat ctgcaaggac tgttttccaa 720

aaagaaaacg gtactgttac tgccgctaac gcttctccaa tcaacgatgg tgctgcagcc 780

gtcatcttgg tttccgaaaa agttttgaag gaaaagaatt tgaagccttt ggctattatc 840

aaaggttggg gtgaggccgc tcatcaacca gctgatttta catgggctcc atctcttgca 900

gttccaaagg ctttgaaaca tgctggcatc gaagacatca attctgttga ttactttgaa 960

ttcaatgaag ccttttcggt tgtcggtttg gtgaacacta agattttgaa gctagaccca 1020

tctaaggtta atgtatatgg tggtgctgtt gctctaggtc acccattggg ttgttctggt 1080

gctagagtgg ttgttacact gctatccatc ttacagcaag aaggaggtaa gatcggtgtt 1140

gccgccattt gtaatggtgg tggtggtgct tcctctattg tcattgaaaa gatatga 1197

<210> 210
<211> 867
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t112

<400> 210
atgtcttacg ttgtcaaggg tatgatctct attgcttgtg gtttgttcgg tagagaattg 60

tttaacaaca gacacttgtt ctcttggggt ttgatgtgga aagctttctt cgctttggtc 120

ccaattttgt ctttcaattt cttcgccgcc atcatgaacc aaatctacga tgttgatatc 180

gaccgtatca acaagccaga cttaccttta gtttccggtg aaatgtccat tgaaactgct 240

tggatcttgt ctatcattgt tgccttgact ggtttaattg ttactattaa gttgaagtcc 300

gctccattgt ttgtcttcat ctacatcttc ggtatcttcg ctggtttcgc ttactccgtc 360

ccacctatta gatggaaaca atatcctttt accaatttct tgatcactat ttcctctcat 420

gttggtttgg ctttcacttc ttactctgcc accacttctg ctttaggttt gcctttcgtt 480

tggcgtcctg ccttctcttt cattattgct ttcatgactg tcatgggtat gactattgcc 540

tttgctaaag acatttctga tatcgaaggt gatgctaagt acggtgtctc taccgttgct 600

accaagttag gtgctagaaa tatgactttt gttgtttctg gtgtcttatt gttgaactac 660

ttggtttcta tctctattgg tatcatttgg ccacaagttt tcaagtctaa cattatgatc 720

ttgtctcatg ctattttggc tttctgtttg atctttcaaa ctcgtgaatt agccttagcc 780

aattatgcct ctgccccatc ccgtcaattt ttcgaattca tctggttgtt atactatgcc 840

gaatacttcg tttacgtctt catttaa 867

<210> 211
<211> 288
<212> PRT
<213> Artificial Sequence

<220>
<223> Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t112

<400> 211

Met Ser Tyr Val Val Lys Gly Met Ile Ser Ile Ala Cys Gly Leu Phe
1 5 10 15

Gly Arg Glu Leu Phe Asn Asn Arg His Leu Phe Ser Trp Gly Leu Met
20 25 30

Trp Lys Ala Phe Phe Ala Leu Val Pro Ile Leu Ser Phe Asn Phe Phe
35 40 45

Ala Ala Ile Met Asn Gln Ile Tyr Asp Val Asp Ile Asp Arg Ile Asn
50 55 60

Lys Pro Asp Leu Pro Leu Val Ser Gly Glu Met Ser Ile Glu Thr Ala
65 70 75 80

Trp Ile Leu Ser Ile Ile Val Ala Leu Thr Gly Leu Ile Val Thr Ile
85 90 95

Lys Leu Lys Ser Ala Pro Leu Phe Val Phe Ile Tyr Ile Phe Gly Ile
100 105 110

Phe Ala Gly Phe Ala Tyr Ser Val Pro Pro Ile Arg Trp Lys Gln Tyr
115 120 125

Pro Phe Thr Asn Phe Leu Ile Thr Ile Ser Ser His Val Gly Leu Ala
130 135 140

Phe Thr Ser Tyr Ser Ala Thr Thr Ser Ala Leu Gly Leu Pro Phe Val
145 150 155 160

Trp Arg Pro Ala Phe Ser Phe Ile Ile Ala Phe Met Thr Val Met Gly

165 170 175

Met Thr Ile Ala Phe Ala Lys Asp Ile Ser Asp Ile Glu Gly Asp Ala
180 185 190

Lys Tyr Gly Val Ser Thr Val Ala Thr Lys Leu Gly Ala Arg Asn Met
195 200 205

Thr Phe Val Val Ser Gly Val Leu Leu Leu Asn Tyr Leu Val Ser Ile
210 215 220

Ser Ile Gly Ile Ile Trp Pro Gln Val Phe Lys Ser Asn Ile Met Ile
225 230 235 240

Leu Ser His Ala Ile Leu Ala Phe Cys Leu Ile Phe Gln Thr Arg Glu
245 250 255

Leu Ala Leu Ala Asn Tyr Ala Ser Ala Pro Ser Arg Gln Phe Phe Glu
260 265 270

Phe Ile Trp Leu Leu Tyr Tyr Ala Glu Tyr Phe Val Tyr Val Phe Ile
275 280 285

<210> 212
<211> 810
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t131 nucleotide sequence

<400> 212
atgtctaaca acagacactt gttctcttgg ggtttgatgt ggaaagcttt cttcgctttg 60

gtcccaattt tgtctttcaa tttcttcgcc gccatcatga accaaatcta cgatgttgat 120

atcgaccgta tcaacaagcc agacttacct ttagtttccg gtgaaatgtc cattgaaact 180

gcttggatct tgtctatcat tgttgccttg actggtttaa ttgttactat taagttgaag 240

tccgctccat tgtttgtctt catctacatc ttcggtatct tcgctggttt cgcttactcc 300

gtcccaccta ttagatggaa acaatatcct tttaccaatt tcttgatcac tatttcctct 360

catgttggtt tggctttcac ttcttactct gccaccactt ctgctttagg tttgcctttc 420

gtttggcgtc ctgccttctc tttcattatt gctttcatga ctgtcatggg tatgactatt 480

gcctttgcta aagacatttc tgatatcgaa ggtgatgcta agtacggtgt ctctaccgtt 540

gctaccaagt taggtgctag aaatatgact tttgttgttt ctggtgtctt attgttgaac 600

tacttggttt ctatctctat tggtatcatt tggccacaag ttttcaagtc taacattatg 660

atcttgtctc atgctatttt ggctttctgt ttgatctttc aaactcgtga attagcctta 720

gccaattatg cctctgcccc atcccgtcaa tttttcgaat tcatctggtt gttatactat 780

gccgaatact tcgtttacgt cttcatttaa 810

<210> 213
<211> 269
<212> PRT
<213> Artificial Sequence

<220>
<223> Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t131

<400> 213

Met Ser Asn Asn Arg His Leu Phe Ser Trp Gly Leu Met Trp Lys Ala
1 5 10 15

Phe Phe Ala Leu Val Pro Ile Leu Ser Phe Asn Phe Phe Ala Ala Ile
20 25 30

Met Asn Gln Ile Tyr Asp Val Asp Ile Asp Arg Ile Asn Lys Pro Asp
35 40 45

Leu Pro Leu Val Ser Gly Glu Met Ser Ile Glu Thr Ala Trp Ile Leu
50 55 60

Ser Ile Ile Val Ala Leu Thr Gly Leu Ile Val Thr Ile Lys Leu Lys
65 70 75 80

Ser Ala Pro Leu Phe Val Phe Ile Tyr Ile Phe Gly Ile Phe Ala Gly
85 90 95

```
Phe Ala Tyr Ser Val Pro Pro Ile Arg Trp Lys Gln Tyr Pro Phe Thr
            100                 105                 110

Asn Phe Leu Ile Thr Ile Ser Ser His Val Gly Leu Ala Phe Thr Ser
        115                 120                 125

Tyr Ser Ala Thr Thr Ser Ala Leu Gly Leu Pro Phe Val Trp Arg Pro
    130                 135                 140

Ala Phe Ser Phe Ile Ile Ala Phe Met Thr Val Met Gly Met Thr Ile
145                 150                 155                 160

Ala Phe Ala Lys Asp Ile Ser Asp Ile Glu Gly Asp Ala Lys Tyr Gly
                165                 170                 175

Val Ser Thr Val Ala Thr Lys Leu Gly Ala Arg Asn Met Thr Phe Val
            180                 185                 190

Val Ser Gly Val Leu Leu Leu Asn Tyr Leu Val Ser Ile Ser Ile Gly
        195                 200                 205

Ile Ile Trp Pro Gln Val Phe Lys Ser Asn Ile Met Ile Leu Ser His
    210                 215                 220

Ala Ile Leu Ala Phe Cys Leu Ile Phe Gln Thr Arg Glu Leu Ala Leu
225                 230                 235                 240

Ala Asn Tyr Ala Ser Ala Pro Ser Arg Gln Phe Phe Glu Phe Ile Trp
                245                 250                 255

Leu Leu Tyr Tyr Ala Glu Tyr Phe Val Tyr Val Phe Ile
            260                 265
```

<210> 214
<211> 777
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial truncated geranyl pyrophosphate olivetolic acid

geranyltransferase CsPT4_t142 nucleotide sequence

<400> 214
atgtcttgga aagctttctt cgctttggtc ccaattttgt ctttcaattt cttcgccgcc 60

atcatgaacc aaatctacga tgttgatatc gaccgtatca acaagccaga cttaccttta 120

gtttccggtg aaatgtccat tgaaactgct tggatcttgt ctatcattgt tgccttgact 180

ggtttaattg ttactattaa gttgaagtcc gctccattgt ttgtcttcat ctacatcttc 240

ggtatcttcg ctggtttcgc ttactccgtc ccacctatta gatggaaaca atatcctttt 300

accaatttct tgatcactat ttcctctcat gttggtttgg ctttcacttc ttactctgcc 360

accacttctg ctttaggttt gcctttcgtt tggcgtcctg ccttctcttt cattattgct 420

ttcatgactg tcatgggtat gactattgcc tttgctaaag acatttctga tatcgaaggt 480

gatgctaagt acggtgtctc taccgttgct accaagttag gtgctagaaa tatgactttt 540

gttgtttctg gtgtcttatt gttgaactac ttggtttcta tctctattgg tatcatttgg 600

ccacaagttt tcaagtctaa cattatgatc ttgtctcatg ctattttggc tttctgtttg 660

atctttcaaa ctcgtgaatt agccttagcc aattatgcct ctgccccatc ccgtcaattt 720

ttcgaattca tctggttgtt atactatgcc gaatacttcg tttacgtctt catttaa 777

<210> 215
<211> 258
<212> PRT
<213> Artificial Sequence

<220>
<223> Truncated geranyl pyrophosphate olivetolic acid
geranyltransferase CsPT4_t142

<400> 215

Met Ser Trp Lys Ala Phe Phe Ala Leu Val Pro Ile Leu Ser Phe Asn
1 5 10 15

Phe Phe Ala Ala Ile Met Asn Gln Ile Tyr Asp Val Asp Ile Asp Arg
20 25 30

Ile Asn Lys Pro Asp Leu Pro Leu Val Ser Gly Glu Met Ser Ile Glu
35 40 45

```
Thr Ala Trp Ile Leu Ser Ile Ile Val Ala Leu Thr Gly Leu Ile Val
    50                  55                  60

Thr Ile Lys Leu Lys Ser Ala Pro Leu Phe Val Phe Ile Tyr Ile Phe
65                  70                  75                  80

Gly Ile Phe Ala Gly Phe Ala Tyr Ser Val Pro Pro Ile Arg Trp Lys
                85                  90                  95

Gln Tyr Pro Phe Thr Asn Phe Leu Ile Thr Ile Ser Ser His Val Gly
            100                 105                 110

Leu Ala Phe Thr Ser Tyr Ser Ala Thr Thr Ser Ala Leu Gly Leu Pro
        115                 120                 125

Phe Val Trp Arg Pro Ala Phe Ser Phe Ile Ile Ala Phe Met Thr Val
    130                 135                 140

Met Gly Met Thr Ile Ala Phe Ala Lys Asp Ile Ser Asp Ile Glu Gly
145                 150                 155                 160

Asp Ala Lys Tyr Gly Val Ser Thr Val Ala Thr Lys Leu Gly Ala Arg
                165                 170                 175

Asn Met Thr Phe Val Val Ser Gly Val Leu Leu Leu Asn Tyr Leu Val
            180                 185                 190

Ser Ile Ser Ile Gly Ile Ile Trp Pro Gln Val Phe Lys Ser Asn Ile
        195                 200                 205

Met Ile Leu Ser His Ala Ile Leu Ala Phe Cys Leu Ile Phe Gln Thr
    210                 215                 220

Arg Glu Leu Ala Leu Ala Asn Tyr Ala Ser Ala Pro Ser Arg Gln Phe
225                 230                 235                 240

Phe Glu Phe Ile Trp Leu Leu Tyr Tyr Ala Glu Tyr Phe Val Tyr Val
                245                 250                 255
```

Phe Ile

<210> 216
<211> 705
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t166 nucleotide sequence

<400> 216
atgtctgatg ttgatatcga ccgtatcaac aagccagact tacctttagt ttccggtgaa 60

atgtccattg aaactgcttg gatcttgtct atcattgttg ccttgactgg tttaattgtt 120

actattaagt tgaagtccgc tccattgttt gtcttcatct acatcttcgg tatcttcgct 180

ggtttcgctt actccgtccc acctattaga tggaaacaat atccttttac caatttcttg 240

atcactattt cctctcatgt tggtttggct ttcacttctt actctgccac cacttctgct 300

ttaggtttgc ctttcgtttg gcgtcctgcc ttctctttca ttattgcttt catgactgtc 360

atgggtatga ctattgcctt tgctaaagac atttctgata tcgaaggtga tgctaagtac 420

ggtgtctcta ccgttgctac caagttaggt gctagaaata tgacttttgt tgtttctggt 480

gtcttattgt tgaactactt ggtttctatc tctattggta tcatttggcc acaagttttc 540

aagtctaaca ttatgatctt gtctcatgct attttggctt tctgtttgat ctttcaaact 600

cgtgaattag ccttagccaa ttatgcctct gccccatccc gtcaattttt cgaattcatc 660

tggttgttat actatgccga atacttcgtt tacgtcttca tttaa 705

<210> 217
<211> 234
<212> PRT
<213> Artificial Sequence

<220>
<223> Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t166

<400> 217

Met Ser Asp Val Asp Ile Asp Arg Ile Asn Lys Pro Asp Leu Pro Leu

```
1               5                   10                  15

Val Ser Gly Glu Met Ser Ile Glu Thr Ala Trp Ile Leu Ser Ile Ile
            20                  25                  30

Val Ala Leu Thr Gly Leu Ile Val Thr Ile Lys Leu Lys Ser Ala Pro
        35                  40                  45

Leu Phe Val Phe Ile Tyr Ile Phe Gly Ile Phe Ala Gly Phe Ala Tyr
    50                  55                  60

Ser Val Pro Pro Ile Arg Trp Lys Gln Tyr Pro Phe Thr Asn Phe Leu
65                  70                  75                  80

Ile Thr Ile Ser Ser His Val Gly Leu Ala Phe Thr Ser Tyr Ser Ala
                85                  90                  95

Thr Thr Ser Ala Leu Gly Leu Pro Phe Val Trp Arg Pro Ala Phe Ser
            100                 105                 110

Phe Ile Ile Ala Phe Met Thr Val Met Gly Met Thr Ile Ala Phe Ala
        115                 120                 125

Lys Asp Ile Ser Asp Ile Glu Gly Asp Ala Lys Tyr Gly Val Ser Thr
    130                 135                 140

Val Ala Thr Lys Leu Gly Ala Arg Asn Met Thr Phe Val Val Ser Gly
145                 150                 155                 160

Val Leu Leu Leu Asn Tyr Leu Val Ser Ile Ser Ile Gly Ile Ile Trp
                165                 170                 175

Pro Gln Val Phe Lys Ser Asn Ile Met Ile Leu Ser His Ala Ile Leu
            180                 185                 190

Ala Phe Cys Leu Ile Phe Gln Thr Arg Glu Leu Ala Leu Ala Asn Tyr
        195                 200                 205

Ala Ser Ala Pro Ser Arg Gln Phe Phe Glu Phe Ile Trp Leu Leu Tyr
```

210 215 220

Tyr Ala Glu Tyr Phe Val Tyr Val Phe Ile
225 230

<210> 218
<211> 645
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t186 nucleotide sequence

<400> 218
atgtctattg aaactgcttg gatcttgtct atcattgttg ccttgactgg tttaattgtt 60

actattaagt tgaagtccgc tccattgttt gtcttcatct acatcttcgg tatcttcgct 120

ggtttcgctt actccgtccc acctattaga tggaaacaat atccttttac caatttcttg 180

atcactattt cctctcatgt tggtttggct ttcacttctt actctgccac cacttctgct 240

ttaggtttgc ctttcgtttg gcgtcctgcc ttctctttca ttattgcttt catgactgtc 300

atgggtatga ctattgcctt tgctaaagac atttctgata tcgaaggtga tgctaagtac 360

ggtgtctcta ccgttgctac caagttaggt gctagaaata tgacttttgt tgtttctggt 420

gtcttattgt tgaactactt ggtttctatc tctattggta tcatttggcc acaagttttc 480

aagtctaaca ttatgatctt gtctcatgct attttggctt tctgtttgat ctttcaaact 540

cgtgaattag ccttagccaa ttatgcctct gccccatccc gtcaattttt cgaattcatc 600

tggttgttat actatgccga atacttcgtt tacgtcttca tttaa 645

<210> 219
<211> 214
<212> PRT
<213> Artificial Sequence

<220>
<223> Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t186

<400> 219

Met Ser Ile Glu Thr Ala Trp Ile Leu Ser Ile Ile Val Ala Leu Thr

1 5 10 15

Gly Leu Ile Val Thr Ile Lys Leu Lys Ser Ala Pro Leu Phe Val Phe
20 25 30

Ile Tyr Ile Phe Gly Ile Phe Ala Gly Phe Ala Tyr Ser Val Pro Pro
35 40 45

Ile Arg Trp Lys Gln Tyr Pro Phe Thr Asn Phe Leu Ile Thr Ile Ser
50 55 60

Ser His Val Gly Leu Ala Phe Thr Ser Tyr Ser Ala Thr Thr Ser Ala
65 70 75 80

Leu Gly Leu Pro Phe Val Trp Arg Pro Ala Phe Ser Phe Ile Ile Ala
85 90 95

Phe Met Thr Val Met Gly Met Thr Ile Ala Phe Ala Lys Asp Ile Ser
100 105 110

Asp Ile Glu Gly Asp Ala Lys Tyr Gly Val Ser Thr Val Ala Thr Lys
115 120 125

Leu Gly Ala Arg Asn Met Thr Phe Val Val Ser Gly Val Leu Leu Leu
130 135 140

Asn Tyr Leu Val Ser Ile Ser Ile Gly Ile Ile Trp Pro Gln Val Phe
145 150 155 160

Lys Ser Asn Ile Met Ile Leu Ser His Ala Ile Leu Ala Phe Cys Leu
165 170 175

Ile Phe Gln Thr Arg Glu Leu Ala Leu Ala Asn Tyr Ala Ser Ala Pro
180 185 190

Ser Arg Gln Phe Phe Glu Phe Ile Trp Leu Leu Tyr Tyr Ala Glu Tyr
195 200 205

Phe Val Tyr Val Phe Ile

210

<210> 220
<211> 1188
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial geranyl pyrophosphate olivetolic acid geranyltransferase CsGOT (CsPT1) nucleotide sequence

<400> 220
atgggtttat cttccgtttg tacttttttct ttccaaacta actaccacac tttgttaaat 60

ccacacaaca acaaccctaa aacctccttg ttatgttaca gacacccaaa gacccctatt 120

aaatactcct acaacaactt cccatccaaa cactgctcca ctaagtcctt tcacttgcaa 180

aacaagtgtt ctgaatcctt gtccattgcc aagaactcta ttcgtgccgc tactactaac 240

caaactgagc cacctgaatc cgataaccac tccgtcgcca ccaagatctt gaattttggt 300

aaagcttgct ggaaattgca aagaccatac actattattg ctttcacttc ctgtgcttgt 360

ggtttattcg gtaaggaatt attgcataac accaacttga tttcttggtc cttaatgttc 420

aaagccttct tctttttagt tgccatttta tgtattgctt ctttcactac tactattaat 480

caaatttacg atttgcacat tgacagaatc aataagcctg acttgccatt agcttccggt 540

gaaatttctg ttaacactgc ttggatcatg tccatcattg tcgctttgtt cggtttaatt 600

atcaccatca aaatgaaggg tggtcctttg tacatcttcg gttattgctt cggtattttc 660

ggtggtattg tctactctgt cccaccattc agatggaagc aaaacccatc cactgccttt 720

ttgttgaatt tcttggctca catcattacc aattttactt tctactatgc ctcccgtgct 780

gctttaggtt tgccttttga gttacgtcca tccttcactt ttttattggc ttttatgaag 840

tccatgggtt ctgctttagc cttaattaag gacgcctctg acgttgaagg tgatactaag 900

ttcggtatct ctactttagc ctctaagtac ggttctcgta acttgacctt gttctgttct 960

ggtattgtct tgttgtctta cgtcgccgct attttggccg gtatcatctg gccacaagct 1020

ttcaactcta acgttatgtt gttgtctcat gctatcttag ctttctggtt gatcttacaa 1080

accagagact tcgctttgac taactacgac ccagaagccg gtcgtagatt ctacgaattc 1140

atgtggaaat tgtactacgc cgagtacttg gtctacgttt tcatttag 1188

<210> 221
<211> 972
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4t nucleotide sequence

<400> 221
atgtctgctg gctctgacca aattgaaggt tccccgcatc acgaatcaga taatagtatt 60

gccacaaaga tcttaaactt tgggcataca tgttggaaat tacaaaggcc ctacgtcgtc 120

aaaggaatga taagcatcgc ttgcggtctg ttcggaaggg aattatttaa caataggcat 180

ctattcagct gggggttaat gtggaaagct ttcttcgcgt tagtgccaat cctaagcttt 240

aacttttcg ccgccatcat gaaccagatt tatgatgttg atatcgacag gataaataag 300

ccagatcttc cattggtatc cggtgaaatg tcaatagaaa ctgcatggat attatctatt 360

atcgttgcgc tgaccggact gatagtaaca atcaaattga aatctgcacc cctgtttgtt 420

tttatatata tatttggtat tttcgctgga ttcgcttact cagtgccacc tatcaggtgg 480

aagcagtacc cattcacgaa ttttctgatc acgatctcta gccacgtcgg gttagcgttc 540

acatcttact ctgcaaccac gagtgccttg gggcttcctt tcgtctggcg tccagctttt 600

agttttatca ttgcctttat gaccgtaatg ggaatgacga tcgcattcgc aaaggacatt 660

tctgacatag agggggatgc aaaatacggt gtctccactg tggcgacaaa attaggagct 720

aggaatatga ctttcgtggt gtccggtgta ttattactaa attatctggt atctataagt 780

atcggcatca tatggccgca agtgtttaaa tccaacatta tgatactgag tcatgctatt 840

ttggcttttt gtctgatttt tcagacgcgt gagttggcgc ttgcaaacta tgcctctgcg 900

cccagcaggc agttttttga attcatatgg ttattgtact atgccgagta tttcgtctac 960

gtatttattt aa 972

<210> 222
<211> 969
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsGOT_t75 (CsPT1_t75) nucleotide sequence

<400> 222
atgtctgccg ctactactaa ccaaactgag ccacctgaat ccgataacca ctccgtcgcc 60

accaagatct tgaattttgg taaagcttgc tggaaattgc aaagaccata cactattatt 120

gctttcactt cctgtgcttg tggtttattc ggtaaggaat tattgcataa caccaacttg 180

atttcttggt ccttaatgtt caaagccttc ttctttttag ttgccatttt atgtattgct 240

tctttcacta ctactattaa tcaaatttac gatttgcaca ttgacagaat caataagcct 300

gacttgccat tagcttccgg tgaaatttct gttaacactg cttggatcat gtccatcatt 360

gtcgctttgt tcggtttaat tatcaccatc aaaatgaagg gtggtccttt gtacatcttc 420

ggttattgct tcggtatttt cggtggtatt gtctactctg tcccaccatt cagatggaag 480

caaaacccat ccactgcctt tttgttgaat ttcttggctc acatcattac caattttact 540

ttctactatg cctcccgtgc tgctttaggt ttgccttttg agttacgtcc atccttcact 600

tttttattgg cttttatgaa gtccatgggt tctgctttag ccttaattaa ggacgcctct 660

gacgttgaag gtgatactaa gttcggtatc tctactttag cctctaagta cggttctcgt 720

aacttgacct tgttctgttc tggtattgtc ttgttgtctt acgtcgccgc tattttggcc 780

ggtatcatct ggccacaagc tttcaactct aacgttatgt tgttgtctca tgctatctta 840

gctttctggt tgatcttaca aaccagagac ttcgctttga ctaactacga cccagaagcc 900

ggtcgtagat tctacgaatt catgtggaaa ttgtactacg ccgagtactt ggtctacgtt 960

ttcatttag 969

<210> 223
<211> 322
<212> PRT
<213> Artificial Sequence

<220>
<223> Truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsGOT_t75 (CsPT1_t75)

<400> 223

Met Ser Ala Ala Thr Thr Asn Gln Thr Glu Pro Pro Glu Ser Asp Asn

```
1               5                   10                  15

His Ser Val Ala Thr Lys Ile Leu Asn Phe Gly Lys Ala Cys Trp Lys
            20                  25                  30

Leu Gln Arg Pro Tyr Thr Ile Ile Ala Phe Thr Ser Cys Ala Cys Gly
        35                  40                  45

Leu Phe Gly Lys Glu Leu Leu His Asn Thr Asn Leu Ile Ser Trp Ser
    50                  55                  60

Leu Met Phe Lys Ala Phe Phe Phe Leu Val Ala Ile Leu Cys Ile Ala
65                  70                  75                  80

Ser Phe Thr Thr Thr Ile Asn Gln Ile Tyr Asp Leu His Ile Asp Arg
                85                  90                  95

Ile Asn Lys Pro Asp Leu Pro Leu Ala Ser Gly Glu Ile Ser Val Asn
            100                 105                 110

Thr Ala Trp Ile Met Ser Ile Ile Val Ala Leu Phe Gly Leu Ile Ile
        115                 120                 125

Thr Ile Lys Met Lys Gly Gly Pro Leu Tyr Ile Phe Gly Tyr Cys Phe
    130                 135                 140

Gly Ile Phe Gly Gly Ile Val Tyr Ser Val Pro Pro Phe Arg Trp Lys
145                 150                 155                 160

Gln Asn Pro Ser Thr Ala Phe Leu Leu Asn Phe Leu Ala His Ile Ile
                165                 170                 175

Thr Asn Phe Thr Phe Tyr Tyr Ala Ser Arg Ala Ala Leu Gly Leu Pro
            180                 185                 190

Phe Glu Leu Arg Pro Ser Phe Thr Phe Leu Leu Ala Phe Met Lys Ser
        195                 200                 205

Met Gly Ser Ala Leu Ala Leu Ile Lys Asp Ala Ser Asp Val Glu Gly
```

210 215 220

Asp Thr Lys Phe Gly Ile Ser Thr Leu Ala Ser Lys Tyr Gly Ser Arg
225 230 235 240

Asn Leu Thr Leu Phe Cys Ser Gly Ile Val Leu Leu Ser Tyr Val Ala
245 250 255

Ala Ile Leu Ala Gly Ile Ile Trp Pro Gln Ala Phe Asn Ser Asn Val
260 265 270

Met Leu Leu Ser His Ala Ile Leu Ala Phe Trp Leu Ile Leu Gln Thr
275 280 285

Arg Asp Phe Ala Leu Thr Asn Tyr Asp Pro Glu Ala Gly Arg Arg Phe
290 295 300

Tyr Glu Phe Met Trp Lys Leu Tyr Tyr Ala Glu Tyr Leu Val Tyr Val
305 310 315 320

Phe Ile

<210> 224
<211> 972
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial truncated geranyl pyrophosphate olivetolic acid geranyltransferase CsPT4_t76 (CsPT4t) nucleotide sequence

<400> 224
atgtccgccg gttctgatca aatcgaaggt tcccctcatc atgagtccga taactccatt 60

gctactaaaa ttttaaattt cggtcatact tgttggaagt tgcaacgtcc ttacgttgtc 120

aagggtatga tctctattgc ttgtggtttg ttcggtagag aattgtttaa caacagacac 180

ttgttctctt ggggtttgat gtggaaagct ttcttcgctt tggtcccaat tttgtctttc 240

aatttcttcg ccgccatcat gaaccaaatc tacgatgttg atatcgaccg tatcaacaag 300

ccagacttac ctttagtttc cggtgaaatg tccattgaaa ctgcttggat cttgtctatc 360

attgttgcct tgactggttt aattgttact attaagttga agtccgctcc attgtttgtc 420

ttcatctaca tcttcggtat cttcgctggt ttcgcttact ccgtcccacc tattagatgg 480

aaacaatatc cttttaccaa tttcttgatc actatttcct ctcatgttgg tttggctttc 540

acttcttact ctgccaccac ttctgcttta ggtttgcctt tcgtttggcg tcctgccttc 600

tctttcatta ttgctttcat gactgtcatg ggtatgacta ttgcctttgc taaagacatt 660

tctgatatcg aaggtgatgc taagtacggt gtctctaccg ttgctaccaa gttaggtgct 720

agaaatatga cttttgttgt ttctggtgtc ttattgttga actacttggt ttctatctct 780

attggtatca tttggccaca agttttcaag tctaacatta tgatcttgtc tcatgctatt 840

ttggctttct gtttgatctt tcaaactcgt gaattagcct tagccaatta tgcctctgcc 900

ccatcccgtc aatttttcga attcatctgg ttgttatact atgccgaata cttcgtttac 960

gtcttcattt aa 972

<210> 225
<211> 1197
<212> DNA
<213> Cannabis sativa

<400> 225
atgggactct cattagtttg tacctttca tttcaaacta attatcatac tttattaaac 60

cctcataata agaatcccaa aaactcatta ttatcttatc aacaccccaa aacaccaata 120

attaaatcct cttatgataa ttttccctct aaatattgct taaccaagaa ctttcattta 180

cttggactca attcacacaa cagaataagc tcacaatcaa ggtccattag ggcaggtagc 240

gatcaaattg aaggttctcc tcatcatgaa tctgataatt caatagcaac taaaatttta 300

aattttggac atacttgttg gaaacttcaa agaccatatg tagtaaaagg gatgatttca 360

atcgcttgtg gtttgtttgg gagagagttg ttcaataaca gacatttatt cagttggggt 420

ttgatgtgga aggcattctt tgctttggtg cctatattgt ccttcaattt ctttgcagca 480

atcatgaatc aaatttacga tgtggacatc gacaggataa acaagcctga tctaccacta 540

gtttcagggg aaatgtcaat tgaaacagct tggattttga gcataattgt ggcactaact 600

gggttgatag taactataaa attgaaatct gcaccacttt ttgttttcat ttacattttt 660

ggtatatttg ctgggtttgc ctattctgtt ccaccaatta gatggaagca atatcctttt 720

accaattttc taattaccat atcgagtcat gtgggcttag ctttcacatc atattctgca 780

accacatcag ctcttggttt accatttgtg tggaggcctg cttttagttt catcatagca 840

ttcatgacag ttatgggtat gactattgct tttgccaaag atatttcaga tattgaaggc 900

gacgccaaat atggggtatc aactgttgca accaaattag gtgctaggaa catgacattt 960

gttgtttctg gagttcttct tctaaactac ttggtttcta tatctattgg gataatttgg 1020

cctcaggttt tcaagagtaa cataatgata ctttctcatg caatcttagc attttgctta 1080

atcttccaga ctcgtgagct tgctctagca aattacgcct cggcgccaag cagacaattc 1140

ttcgagttta tctggttgct atattatgct gaatactttg tatatgtatt tatataa 1197