



(12) **DEMANDE DE BREVET CANADIEN**
CANADIAN PATENT APPLICATION
(13) **A1**

(86) Date de dépôt PCT/PCT Filing Date: 2018/04/27
(87) Date publication PCT/PCT Publication Date: 2018/11/01
(85) Entrée phase nationale/National Entry: 2019/10/28
(86) N° demande PCT/PCT Application No.: US 2018/029668
(87) N° publication PCT/PCT Publication No.: 2018/200888
(30) Priorités/Priorities: 2017/04/27 (US62/491,114);
2017/10/07 (US62/569,532)

(51) Cl.Int./Int.Cl. *C12N 9/10* (2006.01),
C12N 15/52 (2006.01), *C12P 17/06* (2006.01),
C12P 7/22 (2006.01)
(71) Demandeur/Applicant:
REGENTS OF THE UNIVERSITY OF CALIFORNIA, US
(72) Inventeurs/Inventors:
KEASLING, JAY D., US;
D'ESPAUX, LEO, US;
WONG, JEFF, US;
LUO, XIAOZHOU, US;
REITER, MICHAEL, US;
DENBY, CHARLES, US;
LECHNER, ANNA, US
(74) Agent: ADE & COMPANY INC.

(54) Titre : MICRO-ORGANISMES ET PROCEDES DE PRODUCTION DE CANNABINOIDES ET DE DERIVES DE
CANNABINOIDES
(54) Title: MICROORGANISMS AND METHODS FOR PRODUCING CANNABINOIDS AND CANNABINOID
DERIVATIVES

(57) **Abrégé/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.

(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)



(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)

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).

(21) International Application Number:

PCT/US2018/029668

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(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,

(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.

WO 2018/200888 A1

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 2
CONTENANT LES PAGES 1 À 245

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 2
CONTAINING PAGES 1 TO 245

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

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

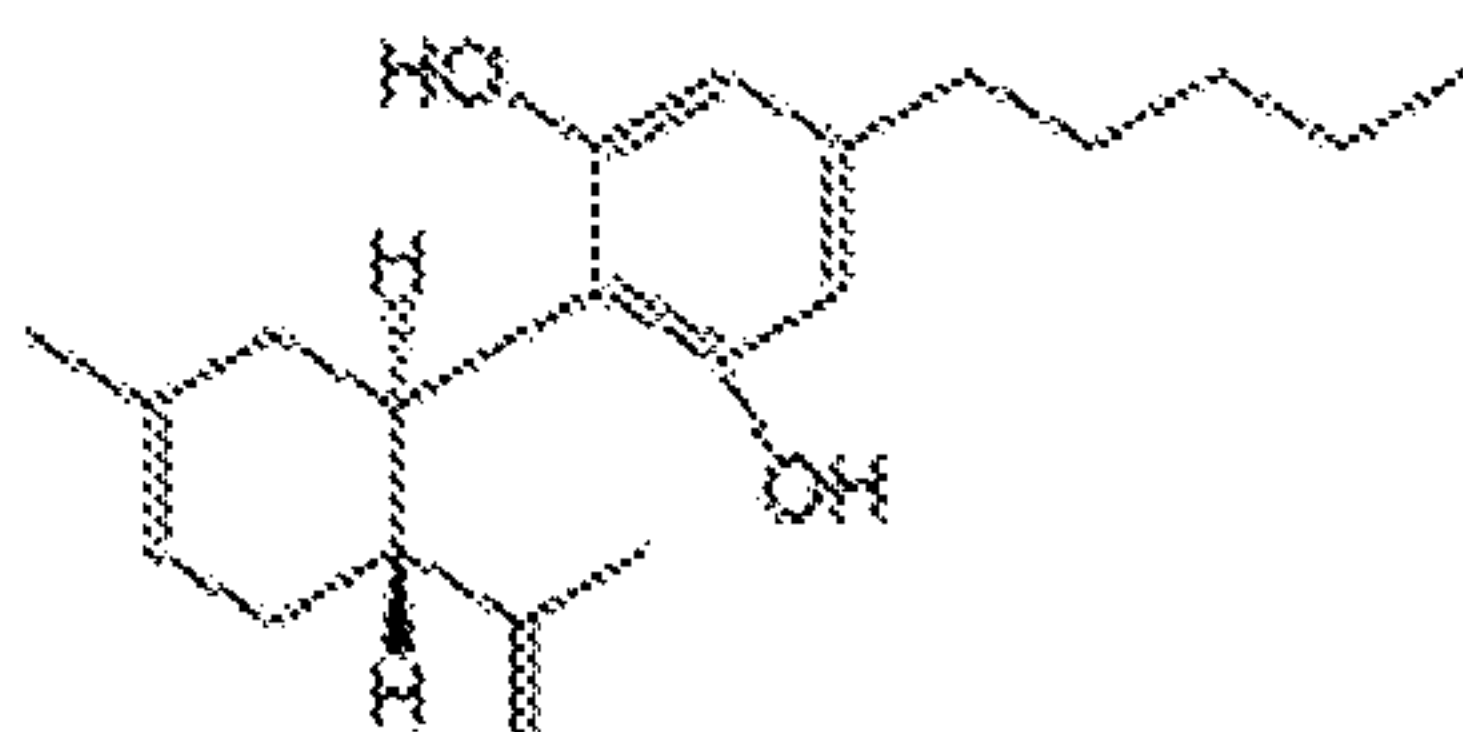
[0002] This invention was made with government support under Grant Numbers 1330914 and 1442724 awarded by the National Science Foundation. The government has certain rights in the invention.

REFERENCE TO A SEQUENCE LISTING

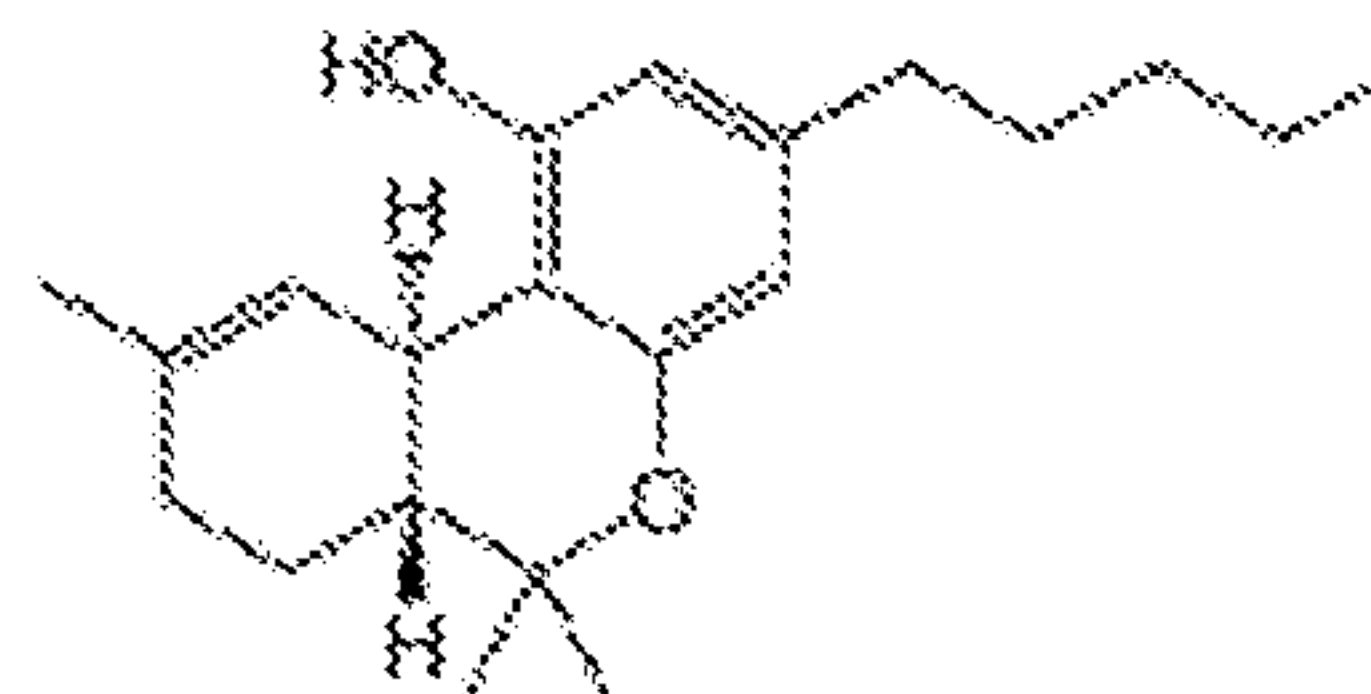
[0003] The Sequence Listing associated with this application is provided in text format in lieu of a paper copy, and is hereby incorporated by reference into the specification. The name of the text file containing the Sequence Listing is DENT_001_00WOSeqList_ST25_txt. The text file is about 673 KB, was 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

[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] 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.

[0009] 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.

[0010] 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.

[0011] 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 (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.

[0012] 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.

[0013] 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.

[0014] 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.

[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 pyruvate dehydrogenase complex (PDC) polypeptide. In some embodiments, the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117.

[0016] 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.

[0017] In certain embodiments of any of the foregoing or following, the genetically modified host cell is a prokaryotic cell.

[0018] 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.

[0019] In certain embodiments of any of the foregoing or following, at least one of the one or more heterologous nucleic acids is maintained extrachromosomally.

[0020] 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.

[0021] 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.

[0022] 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.

[0023] 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.

[0024] 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.

[0025] 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, cannabicyclic acid, cannabicyclol, cannabielsoinic acid, cannabielsoin, cannabicitranic acid, or cannabicitran.

[0026] 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.

[0027] 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.

[0028] 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.

[0029] 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.

[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 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.

[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 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.

[0032] 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.

[0033] 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.

[0034] 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.

[0035] 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.

[0036] 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.

[0037] 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.

[0038] 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.

[0039] 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.

[0040] 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.

[0041] 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.

[0042] 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.

[0043] 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.

[0044] 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.

[0045] 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.

[0046] 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

[0047] FIG. 1 provides a schematic diagram of biosynthetic pathways for generating cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives.

[0048] FIG. 2 depicts intracellular olivetolic acid production using pathway 1a and a tetraketide synthase (TKS) polypeptide/olivetolic acid cyclase (OAC) polypeptide.

- [0049] FIG. 3 depicts intracellular olivetolic acid production comparing pathway 1a and 1b.
- [0050] FIG. 4 provides schematic depictions of 3 expression constructs for olivetolic acid production.
- [0051] FIG. 5 provides schematic depictions of 2 expression constructs for olivetolic acid production.
- [0052] FIG. 6 provides schematic depictions of 3 expression constructs for geranyl pyrophosphate (GPP) production.
- [0053] FIG. 7 provides schematic depictions of 2 expression constructs for GPP production.
- [0054] FIG. 8 provides schematic depictions of 3 expression constructs for cannabinoid production.
- [0055] FIG. 9 depicts production of olivetolic acid using expression constructs 3 + 4 or expression constructs 3 + 5.
- [0056] FIG. 10 depicts production of olivetolic acid using Construct 3 and culturing the cells in medium comprising hexanoate; or using Construct 1.
- [0057] 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.
- [0058] 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.
- [0059] 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.
- [0060] FIG. 14 depicts cannabinoid biosynthetic pathways utilizing neryl pyrophosphate (NPP) or GPP.
- [0061] FIG. 15 depicts generation of cannabigerolic acid (CBGA) using a NphB polypeptide and the substrates olivetolic acid and GPP.
- [0062] FIG. 16 depicts an expression construct to produce GPP.
- [0063] FIG. 17 depicts an expression construct to produce hexanoyl-CoA and/or hexanoate.

- [0064] FIG. 18 depicts an expression construct to produce hexanoyl-CoA and/or hexanoate.
- [0065] FIG. 19 depicts an expression construct to produce olivetolic acid.
- [0066] FIG. 20 depicts an expression construct to produce CBGA.
- [0067] FIG. 21 depicts an expression construct to produce cannabidiolic acid (CBDA).
- [0068] FIG. 22 depicts an expression construct to produce CBDA.
- [0069] FIG. 23 depicts an expression construct to produce CBDA.
- [0070] FIG. 24 depicts an expression construct to produce tetrahydrocannabinolic acid (THCA).
- [0071] FIG. 25 depicts an expression construct to produce THCA.
- [0072] 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 μ M CBGA standard (FIG. 26B), and a mixture of ethyl acetate extraction of yL444 and 10 μ M CBGA standard (FIG. 26C). Peaks observed at 9.2 minutes indicated the presence of CBGA in the ethyl acetate extraction of yL444.
- [0073] 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.
- [0074] 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.
- [0075] 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.

[0076] 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.

[0077] FIG. 31 depicts expression constructs used in the production of the S35 strain.

[0078] FIG. 32 depicts expression constructs used in the production of the S37 strain.

[0079] FIG. 33 depicts expression constructs used in the production of the S38 strain.

[0080] FIG. 34 depicts expression constructs used in the production of the S39 strain.

[0081] FIG. 35 depicts expression constructs used in the production of the S41 strain.

[0082] FIG. 36 depicts expression constructs used in the production of the S42 strain.

[0083] FIG. 37 depicts expression constructs used in the production of the S43 strain.

[0084] FIG. 38 depicts expression constructs used in the production of the S44 strain.

[0085] FIG. 39 depicts expression constructs used in the production of the S45 strain.

[0086] FIG. 40 depicts expression constructs used in the production of the S46 strain.

[0087] FIG. 41 depicts expression constructs used in the production of the S47 strain.

[0088] FIGS. 42A, 42B, and 42C depict expression constructs used in the production of the S49 strain.

[0089] FIGS. 43A, 43B, and 43C depict expression constructs used in the production of the S50 strain.

[0090] 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.

[0091] FIG. 45 depicts expression constructs used in the production of the S78 strain.

[0092] FIG. 46 depicts expression constructs used in the production of the S80 strain.

[0093] FIG. 47 depicts expression constructs used in the production of the S81 strain.

[0094] FIG. 48 depicts expression constructs used in the production of the S82 strain.

[0095] FIG. 49 depicts expression constructs used in the production of the S83 strain.

[0096] FIG. 50 depicts expression constructs used in the production of the S84 strain.

[0097] FIG. 51 depicts expression constructs used in the production of the S85 strain.

[0098] FIG. 52 depicts expression constructs used in the production of the S86 strain.

[0099] FIG. 53 depicts expression constructs used in the production of the S87 strain.

[00100] FIG. 54 depicts expression constructs used in the production of the S88 strain.

- [00101] FIG. 55 depicts expression constructs used in the production of the S89 strain.
- [00102] FIGS. 56A, 56B, and 56C depict expression constructs used in the production of the S90 strain.
- [00103] FIGS. 57A, 57B, and 57C depict expression constructs used in the production of the S91 strain.
- [00104] FIG. 58 depicts expression constructs used in the production of the S94 strain.
- [00105] FIG. 59 depicts expression constructs used in the production of the S95 strain.
- [00106] FIG. 60 depicts expression constructs used in the production of the S97 strain.
- [00107] FIG. 61 depicts expression constructs used in the production of the S104 strain.
- [00108] FIG. 62 depicts expression constructs used in the production of the S108 strain.
- [00109] FIG. 63 depicts expression constructs used in the production of the S112 strain.
- [00110] FIG. 64 depicts expression constructs used in the production of the S114 strain.
- [00111] FIG. 65 depicts expression constructs used in the production of the S115 strain.
- [00112] FIG. 66 depicts expression constructs used in the production of the S116 strain.
- [00113] FIG. 67 depicts expression constructs used in the production of the S118 strain.
- [00114] FIG. 68 depicts expression constructs used in the production of the S123 strain.
- [00115] FIG. 69 depicts expression constructs used in the production of the S147 strain.
- [00116] FIG. 70 depicts expression constructs used in the production of the S164 strain.
- [00117] FIG. 71 depicts expression constructs used in the production of the S165 strain.
- [00118] FIG. 72 depicts expression constructs used in the production of the S166 strain.
- [00119] FIG. 73 depicts expression constructs used in the production of the S167 strain.

- [00120] FIG. 74 depicts expression constructs used in the production of the S168 strain.
- [00121] FIG. 75 depicts expression constructs used in the production of the S169 strain.
- [00122] FIG. 76 depicts expression constructs used in the production of the S170 strain.
- [00123] FIG. 77 depicts the MS/MS spectrum of the CBGA peak produced from a CsPT4 polypeptide expressing strain (S29).
- [00124] FIG. 78 depicts the MS/MS spectrum of an authentic CBGA standard.
- [00125] 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).
- [00126] FIG. 80 depicts CBGA produced by a CsGOT polypeptide at 1.06 min (scale $\times 10^2$ units).
- [00127] FIG. 81 depicts CBGA produced by a CsPT4 polypeptide at 1.06 min (scale $\times 10^4$ units).
- [00128] FIG. 82 depicts an authentic CBGA standard at 1.06 min (scale $\times 10^4$ units).
- [00129] FIG. 83 depicts CBDA produced by S34 at 1.02 min (top) and an authentic CBDA standard at 1.02 min (bottom).
- [00130] FIG. 84 depicts THCA produced from strain D123 at 1.29 min (top) and an authentic THCA standard at 1.29 min (bottom).
- [00131] FIG. 85 depicts expression constructs used in the production of the S34 strain.
- [00132] 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

- [00133] 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).

[00134] 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.

[00135] 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.

[00136] 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 dehydrogenase complex polypeptides (see FIGS. 1 and 11).

[00137] 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).

[00138] 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 CBDA, or derivatives thereof, via a CBDA or THCA synthase polypeptide.

[00139] 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.

[00140] 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.

[00141] 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 cannabinoids, cannabinoid derivatives, cannabinoid precursors, or cannabinoid precursor derivatives. In some embodiments, the *in vitro* methods are cell-free.

[00142] 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

[00143] 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.

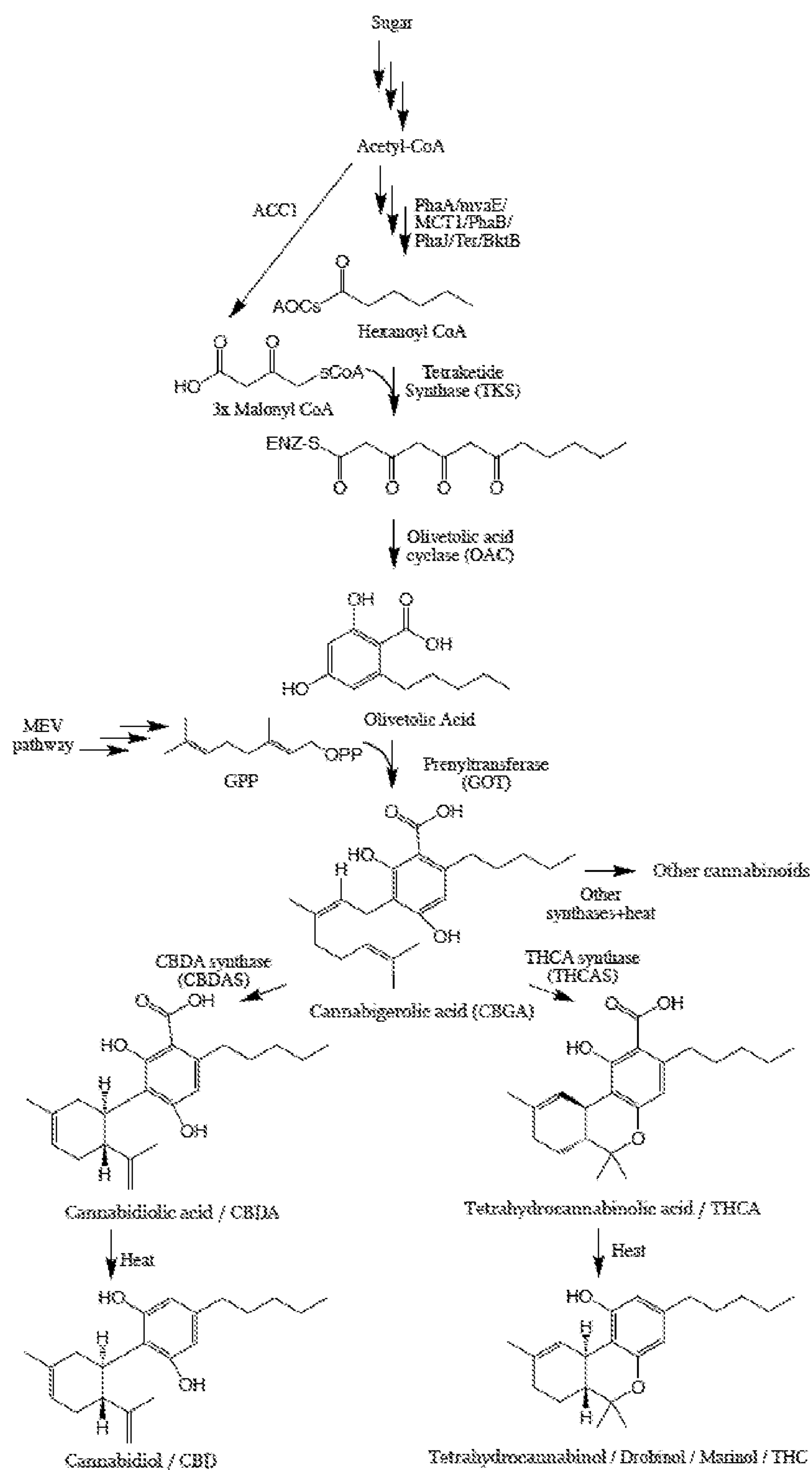
[00144] In *Cannabis*, cannabinoids are produced from the common metabolite precursors geranylpyrophosphate (GPP) and hexanoyl-CoA by the action of three polypeptides so far only identified in *Cannabis*. Hexanoyl-CoA and malonyl-CoA are combined to afford a 12-carbon tetraketide intermediate by a TKS 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.

[00145] 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.

[00146] 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).

[00147] 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

[00148] GPP and hexanoyl-CoA may also be generated through pathways comprising polypeptides that condense two molecules of acetyl-CoA to generate acetoacetyl-CoA and pyruvate dehydrogenase complex 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

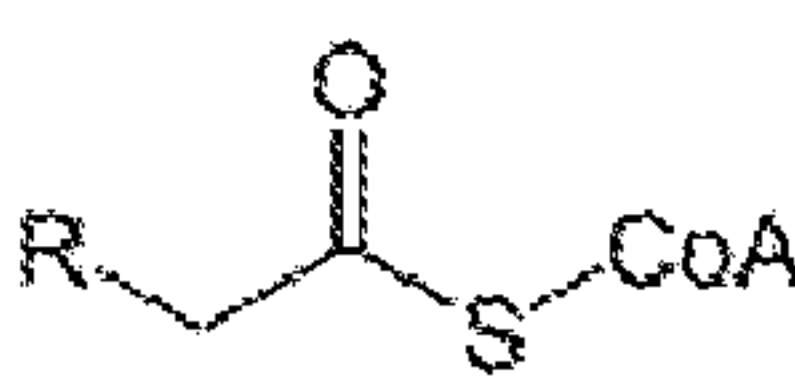
[00149] The practice of the present disclosure will employ, unless otherwise indicated, conventional techniques of molecular biology (including recombinant techniques), microbiology, cell biology, biochemistry, and immunology, which are within the skill of the art. Such techniques are explained fully 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.

[00150] “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, cannabitrilol (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), cannabicyclic 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-ethoxy-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).

[00151] “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.

[00152] 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).

[00153] 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, heterocyclalkenyl, heterocyclalkynyl, heteroarylalkenyl, heteroarylalkynyl, arylalkenyl, arylalkynyl, heterocycl, spirocycl, heterospirocycl, thioalkyl, sulfone, sulfonyl, sulfoxide, amido, alkylamino, dialkylamino, arylamino, alkylaryl amino, diarylamino, N-oxide, imide, enamine, imine, oxime, hydrazone, nitrile, aralkyl, cycloalkylalkyl, haloalkyl, heterocyclalkyl, 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.

[00154] 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, heterocyclalkenyl, heteroarylalkenyl, arylalkenyl, heterocycl, aralkyl, cycloalkylalkyl, heterocyclalkyl, 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.

[00155] 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.

[00156] 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.

[00157] 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.

[00158] “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.

[00159] “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.

[00160] 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.

[00161] 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.

[00162] 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.

[00163] 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.nih.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.

[00164] 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.

[00165] 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.

[00166] 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.

[00167] 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.

[00168] 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 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

[00169] 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).

[00170] 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.

[00171] 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.

[00172] 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

[00173] 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.

[00174] 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.

[00175] 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.

[00176] 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.

[00177] 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.

[00178] 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.

[00179] 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.

[00180] 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.

[00181] 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.

[00182] 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.

[00183] 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.

[00184] 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.

[00185] 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.

[00186] 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 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).

[00187] 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

[00188] 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.

[00189] 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.

[00190] 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.

[00191] 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.

[00192] 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

[00193] 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.

[00194] 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.

[00195] 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.

[00196] 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.

[00197] 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.

[00198] 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.

[00199] 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.

[00200] 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.

[00201] 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.

[00202] 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.

[00203] 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 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.

[00204] 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.

[00205] 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

[00206] 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.

[00207] 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.

[00208] 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.

[00209] 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.

[00210] 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.

[00211] 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.

[00212] 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.

[00213] 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.

[00214] 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.

[00215] 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.

[00216] 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.

[00217] 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.

[00218] 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

[00219] 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.

[00220] 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.

[00221] 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 dehydrogenase complex 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.

[00222] 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.

[00223] 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.

[00224] 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

[00225] 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.

[00226] 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.

[00227] 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.

[00228] 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.

[00229] 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.

[00230] 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.

[00231] 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

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.

[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: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.

[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: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.

[00234] 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.

[00235] 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.

[00236] 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.

[00237] 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.

[00238] 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.

[00239] 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.

[00240] 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.

[00241] 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.

[00242] 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.

[00243] 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.

[00244] 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.

[00245] 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.

[00246] 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.

[00247] 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.

[00248] 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.

[00249] 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.

[00250] 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.

[00251] 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.

[00252] 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.

[00253] 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.

[00254] 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.

[00255] 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.

[00256] 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.

[00257] 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.

[00258] 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.

[00259] 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.

[00260] 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.

[00261] 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.

[00262] 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.

[00263] 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.

[00264] 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.

[00265] 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.

[00266] 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.

[00267] 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.

[00268] 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.

[00269] 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.

[00270] 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

[00271] 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.

[00272] 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.

[00273] 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.

[00274] 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.

[00275] 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.

[00276] 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.

[00277] 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.

[00278] 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.

[00279] 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.

[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: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.

[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: 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.

[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: 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.

[00283] 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.

[00284] 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.

[00285] 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.

[00286] 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.

[00287] 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.

[00288] 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.

[00289] 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.

[00290] 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.

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 CBDAS polypeptides.

[00291] 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.

[00292] 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.

[00293] 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.

[00294] 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.

[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: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.

[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: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.

[00297] 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.

[00298] 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.

[00299] 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.

[00300] 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.

[00301] 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.

[00302] 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.

[00303] 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

[00304] 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.

[00305] 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.

[00306] 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.

[00307] 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.

[00308] 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.

[00309] 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.

[00310] 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.

[00311] 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.

[00312] 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 RRELIQKVRNM C-terminal amino acids.

[00313] 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.

[00314] 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.

[00315] 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.

[00316] 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.

[00317] 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.

[00318] 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.

[00319] 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.

[00320] 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.

[00321] 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.

[00322] 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.

[00323] 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.

[00324] 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.

[00325] 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.

[00326] 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

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.

[00327] 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.

[00328] 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.

[00329] 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.

[00330] 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.

[00331] 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.

[00332] 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.

[00333] 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.

[00334] 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.

[00335] 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.

[00336] 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

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.

[00337] 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.

[00338] 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.

[00339] 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.

[00340] 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%,

at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at 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.

[00341] 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.

[00342] 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

[00343] 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.

[00344] 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

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-

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

[00345] 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.

[00346] 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.

[00347] 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.

[00348] 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.

[00349] 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.

[00350] 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.

[00351] 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.

[00352] 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.

[00353] 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.

[00354] 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.

[00355] 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.

[00356] 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.

[00357] 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.

[00358] 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.

[00359] 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

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

[00360] 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.

[00361] 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.

[00362] 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.

[00363] 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.

[00364] 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.

[00365] 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.

[00366] 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.

[00367] 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.

[00368] 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

[00369] 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.

[00370] 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

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.

[00371] 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.

[00372] 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.

[00373] 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

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

[00374] 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.

[00375] 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.

[00376] 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.

[00377] 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.

[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: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

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.

[00379] 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.

[00380] 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.

[00381] 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.

[00382] 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.

[00383] 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.

[00384] 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.

[00385] 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.

[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: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.

[00387] 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

[00388] 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

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

[00389] 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.

[00390] 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.

[00391] 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.

[00392] 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.

[00393] 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.

[00394] 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

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.

[00395] 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.

[00396] 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.

[00397] 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.

[00398] 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.

[00399] 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.

[00400] 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.

[00401] 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.

[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: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.

[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: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.

[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: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.

[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: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.

[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: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.

[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: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.

[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: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.

[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: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.

[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: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.

[00411] 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.

[00412] 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.

[00413] 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.

[00414] 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.

[00415] 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.

[00416] 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

96%, at least 97%, at least 98%, at 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.

[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. 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.

[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: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.

[00419] 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

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.

[00420] 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.

[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: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.

[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: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.

[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: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.

[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: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.

[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: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.

[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: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.

[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:142. 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, 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.

[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: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.

[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: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

[00430] 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.

[00431] 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.

[00432] 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.

[00433] 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.

[00434] 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

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.

[00435] 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

[00436] 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

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

[00437] 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.

[00438] 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.

[00439] 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.

[00440] 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.

[00441] 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

[00442] 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 dehydrogenase complex (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.

[00443] 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.

[00444] 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.

[00445] 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.

[00446] 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.

[00447] 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

[00448] 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.

[00449] 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.

[00450] 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.

[00451] 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.

[00452] 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.

[00453] 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.

[00454] 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

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

[00455] 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.

[00456] 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).

[00457] 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.

[00458] 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.

[00459] 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

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

[00460] 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.

[00461] 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.

[00462] 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.

[00463] 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.

[00464] 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.

[00465] 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.

[00466] 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.

[00467] 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.

[00468] 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.

[00469] 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.

[00470] 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.

[00471] 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.

[00472] 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.

[00473] 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.

[00474] 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.

[00475] 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

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.

[00476] 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.

[00477] 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.

[00478] 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.

[00479] 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.

[00480] 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.

[00481] 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.

[00482] 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.

[00483] 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.

[00484] 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.

[00485] 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.

[00486] 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

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.

[00487] 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.

[00488] 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.

[00489] 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.

[00490] 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.

[00491] 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.

[00492] 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.

[00493] 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.

[00494] 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.

[00495] 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.

[00496] 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.

[00497] 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.

[00498] 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.

[00499] 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.

[00500] 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.

[00501] 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.

[00502] 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.

[00503] 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.

[00504] 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

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

[00505] 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.

[00506] 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.

[00507] 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,

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.

[00508] 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.

[00509] 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

[00510] 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.

[00511] 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.

[00512] 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 2C-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.

[00513] 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.

[00514] 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.

[00515] 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.

[00516] 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.

[00517] 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.

[00518] 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.

[00519] 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.

[00520] 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.

[00521] 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.

[00522] 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

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.

[00523] 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.

[00524] 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.

[00525] 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.

[00526] 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.

[00527] 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

[00528] 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.

[00529] 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

[00530] 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.

[00531] 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.

[00532] 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

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.

[00533] 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

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.

[00534] 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.

[00535] 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.

[00536] 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.

[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 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:110

[00538] 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.

[00539] 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.

[00540] 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.

[00541] 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.

[00542] 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.

[00543] 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.

[00544] 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.

[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 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

[00546] 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.

[00547] 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.

[00548] 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.

[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: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.

[00550] 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.

[00551] 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.

[00552] 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.

[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 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

[00554] 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.

[00555] 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.

[00556] 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.

[00557] 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.

[00558] 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.

[00559] 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.

[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, 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.

[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 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.

[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 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.

[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; 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.

[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 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.

[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 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.

[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; 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.

[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; 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.

[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 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.

[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; 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.

[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 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.

[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 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.

[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 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.

[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 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.

[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 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.

[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 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.

[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 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.

[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 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.

[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 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.

[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 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.

[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 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.

[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 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.

[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 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.

[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 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.

[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 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.

[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 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.

[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 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.

[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 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.

[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 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.

[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 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.

[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 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.

[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 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.

[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 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 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.

[00593] 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.

[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 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.

Exemplary Genetically Modified Host Cells Expressing NPPS Polypeptides

[00595] 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.

[00596] 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.

[00597] 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.

[00598] A GOT polypeptide, using NPP and olivetolic acid as substrates, can generate cannabimerolic 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. Cannabimerolic acid or derivatives thereof can serve as a substrate for a CBDAS or THCAS polypeptide to generate CBDA or THCA, or derivatives thereof, respectively.

[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 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.

[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 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.

[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 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.

[00602] 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.

[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 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.

[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; 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

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 2
CONTENANT LES PAGES 1 À 245

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 2
CONTAINING PAGES 1 TO 245

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

CLAIMS

What is claimed is:

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.
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.
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.
4. The genetically modified host cell of any one of claims 1-3, 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 (OAC) polypeptide, or one or more heterologous nucleic acids encoding a fusion TKS and OAC polypeptide.
5. The genetically modified host cell of claim 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.

6. The genetically modified host cell of claim 4 or 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.

7. The genetically modified host cell of any one of claims 1-6, 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.

8. The genetically modified host cell of claim 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 an acyl-activating enzyme (AAE) polypeptide.

9. The genetically modified host cell of claim 8, wherein the AAE polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:90.

10. The genetically modified host cell of claim 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.

11. The genetically modified host cell of claim 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.

12. The genetically modified host cell of claim 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.

13. The genetically modified host cell of claim 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.

14. The genetically modified host cell of claim 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.

15. The genetically modified host cell of any one of claims 7-14, 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.

16. The genetically modified host cell of claim 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.

17. The genetically modified host cell of any one of claims 7-16, 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.

18. The genetically modified host cell of claim 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.

19. The genetically modified host cell of any one of claims 1-18, 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.

20. The genetically modified host cell of claim 19, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an IDI polypeptide.

21. The genetically modified host cell of claim 20, wherein the IDI polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:58.

22. The genetically modified host cell of any one of claims 19-21, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an HMGR polypeptide.

23. The genetically modified host cell of claim 22, wherein the HMGR polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:22.

24. The genetically modified host cell of any one of claims 19-21, 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.

25. The genetically modified host cell of claim 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.

26. The genetically modified host cell of any one of claims 19-25, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an HMGS polypeptide.

27. The genetically modified host cell of claim 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.

28. The genetically modified host cell of any one of claims 19-27, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding an MK polypeptide.

29. The genetically modified host cell of claim 28, wherein the MK polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:64.

30. The genetically modified host cell of any one of claims 19-29, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a PMK polypeptide.

31. The genetically modified host cell of claim 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.

32. The genetically modified host cell of any one of claims 19-31, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a MVD polypeptide.

33. The genetically modified host cell of claim 32, wherein the MVD polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:66.

34. The genetically modified host cell of any one of claims 1-33, 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.

35. The genetically modified host cell of claim 34, wherein the polypeptide that condenses two molecules of acetyl-CoA to generate acetoacetyl-CoA is an acetoacetyl-CoA thiolase polypeptide.

36. The genetically modified host cell of claim 35, wherein the acetoacetyl-CoA thiolase polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:25.

37. The genetically modified host cell of any one of claims 1-36, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a pyruvate dehydrogenase complex (PDC) polypeptide.

38. The genetically modified host cell of claim 37, wherein the PDC polypeptide comprises an amino acid sequence having at least 50% sequence identity to SEQ ID NO:117.

39. The genetically modified host cell of any one of claims 1-38, wherein the genetically modified host cell is a eukaryotic cell.

40. The genetically modified host cell of claim 39, wherein the eukaryotic cell is a yeast cell.

41. The genetically modified host cell of claim 40, wherein the yeast cell is *Saccharomyces cerevisiae*.

42. The genetically modified host cell of claim 41, wherein the *Saccharomyces cerevisiae* is a protease-deficient strain of *Saccharomyces cerevisiae*.

43. The genetically modified host cell of any one of claims 1-39, wherein the genetically modified host cell is a plant cell.

44. The genetically modified host cell of any one of claims 1-38, wherein the genetically modified host cell is a prokaryotic cell.

45. The genetically modified host cell of any one of claims 1-44, wherein at least one of the one or more heterologous nucleic acids is integrated into the chromosome of the genetically modified host cell.

46. The genetically modified host cell of any one of claims 1-44, wherein at least one of the one or more heterologous nucleic acids is maintained extrachromosomally.

47. The genetically modified host cell of any one of claims 1-44, wherein two or more of the one or more heterologous nucleic acids are present in a single expression vector.

48. The genetically modified host cell of any one of claims 1-44, wherein at least one of the heterologous nucleic acids is operably linked to an inducible promoter.

49. The genetically modified host cell of any one of claims 1-44, wherein at least one of the heterologous nucleic acids is operably linked to a constitutive promoter.

50. The genetically modified host cell of any one of claims 1-49, 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.

51. The genetically modified host cell of any one of claims 1-50, wherein the genetically modified host cell further comprises one or more heterologous nucleic acids encoding a cannabinoid synthase polypeptide.

52. The genetically modified host cell of claim 51, wherein the cannabinoid synthase polypeptide is a tetrahydrocannabinolic acid (THCA) synthase polypeptide.

53. The genetically modified host cell of claim 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.

54. The genetically modified host cell of claim 51, wherein the cannabinoid synthase polypeptide is a cannabidiolic acid (CBDA) synthase polypeptide.

55. The genetically modified host cell of claim 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.

56. The genetically modified host cell of any one of claims 1-55, 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, cannabichromevarin, cannabigerovarinic acid, cannabigerovarin, cannabicyclolic acid, cannabicyclol, cannabielsoinic acid, cannabielsoin, cannabicitranic acid, or cannabicitran.

57. The genetically modified host cell of claim 56, wherein the cannabinoid is cannabigerolic acid.

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 claims 1-57 in a suitable medium; and
- b) recovering the produced cannabinoid or cannabinoid derivative.

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 claims 1-57 in a suitable medium comprising a carboxylic acid;
- b) recovering the produced cannabinoid or cannabinoid derivative.

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 claims 1-57 in a suitable medium comprising olivetolic acid or an olivetolic acid derivative;
- b) recovering the produced cannabinoid or cannabinoid derivative.

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.

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.

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 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.

64. The method of any one of claims 58-63, wherein the suitable medium comprises a fermentable sugar.

65. The method of any one of claims 58-63, wherein the suitable medium comprises a pretreated cellulosic feedstock.

66. The method of any one of claims 58-63, wherein the suitable medium comprises a non-fermentable carbon source.

67. The method of claim 66, wherein the non-fermentable carbon source comprises ethanol.

68. The method of any one of claims 58-67, 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, cannabicyclic acid, cannabicyclol, cannabielsoinic acid, cannabielsoin, cannabicitranic acid, or cannabicitran.

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.

70. An isolated or purified polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:110.

71. An isolated or purified polypeptide comprising an amino acid sequence having at least 65% sequence identity to SEQ ID NO:100.

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

times higher than a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:82.

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.

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.

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.

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.

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.

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 the genetically modified host cell.

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 the genetically modified host cell.

80. 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:100 into the genetically modified host cell.

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 claims 75-77 into the genetically modified host cell.

1/101

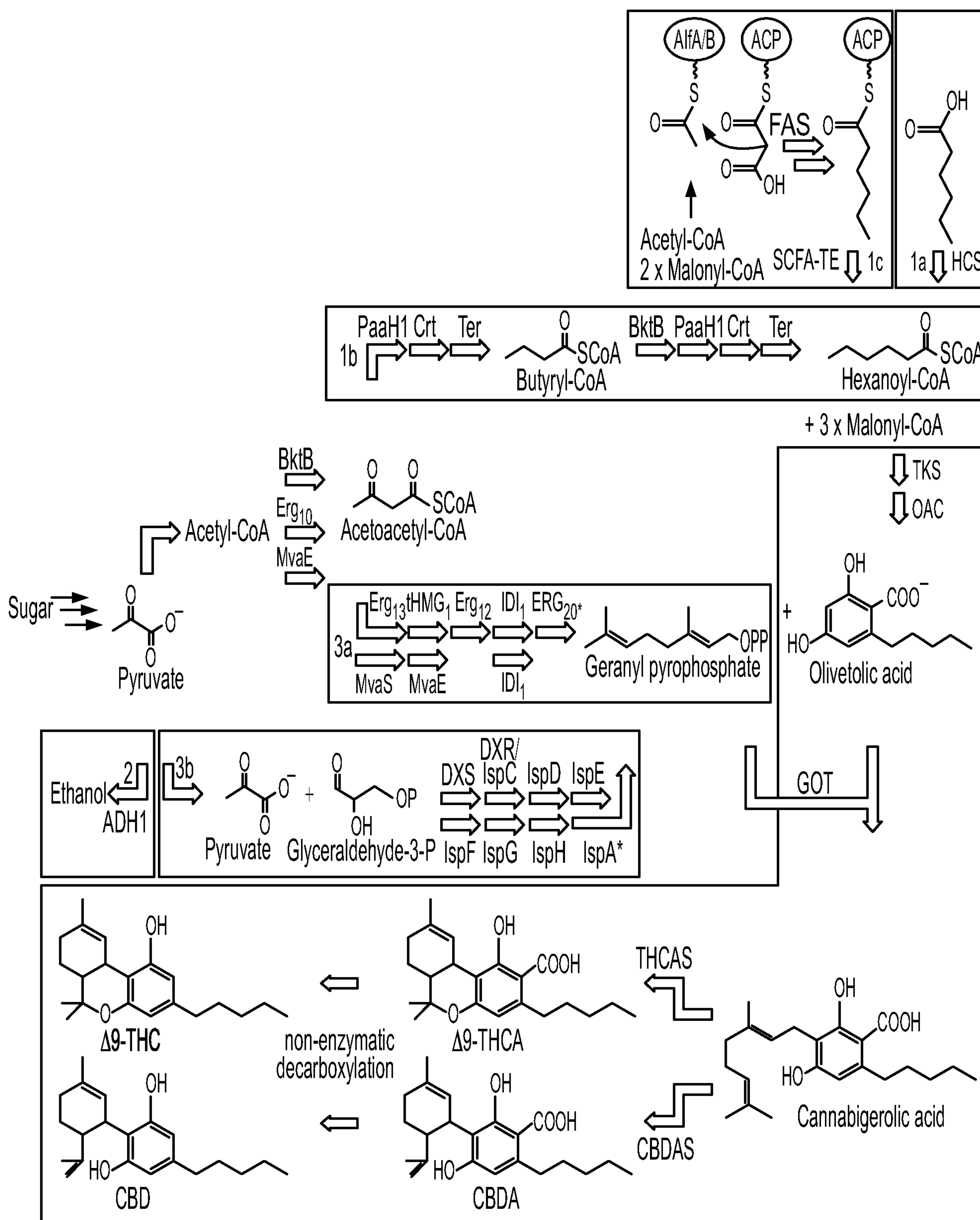


FIG. 1

2/101

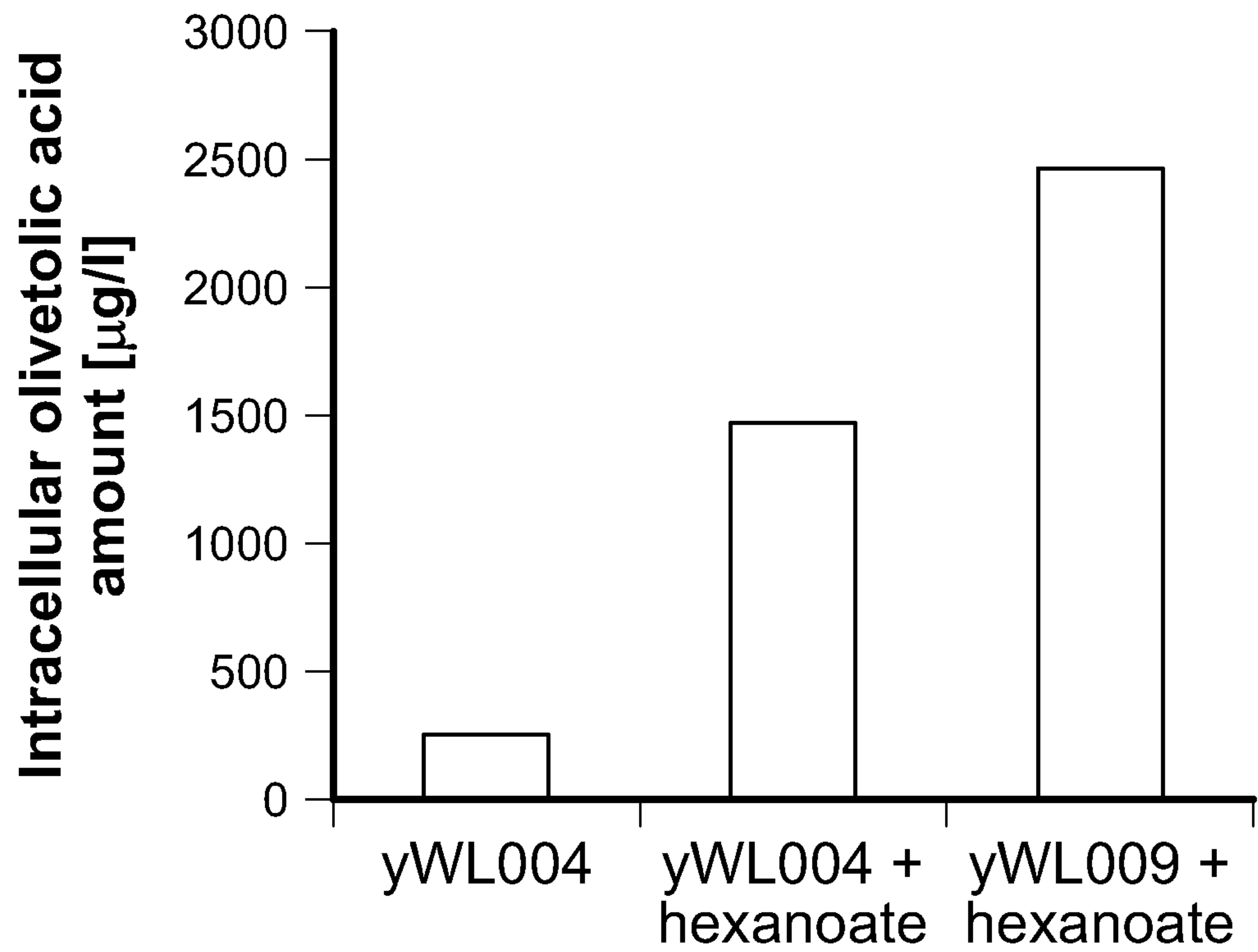


FIG. 2

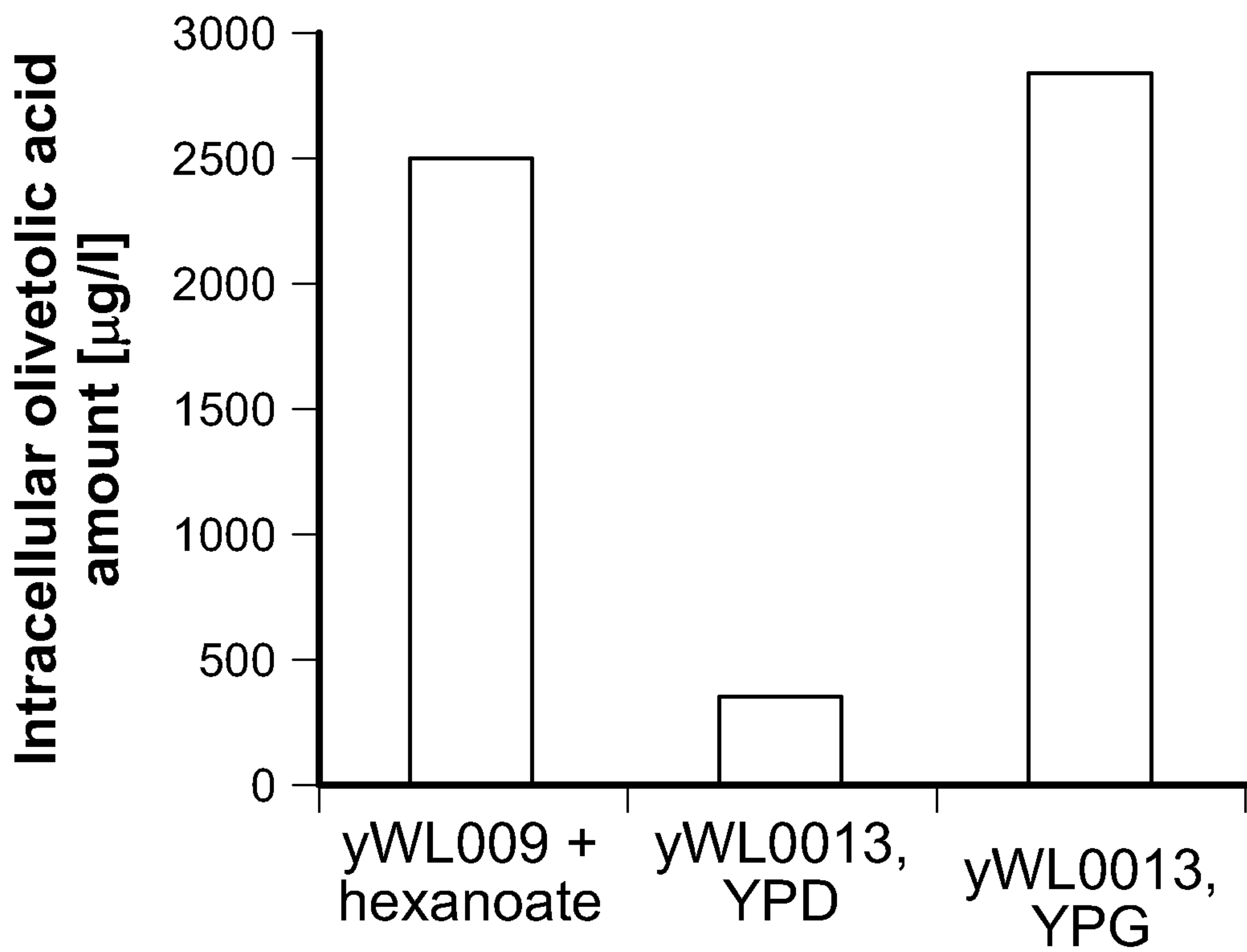


FIG. 3

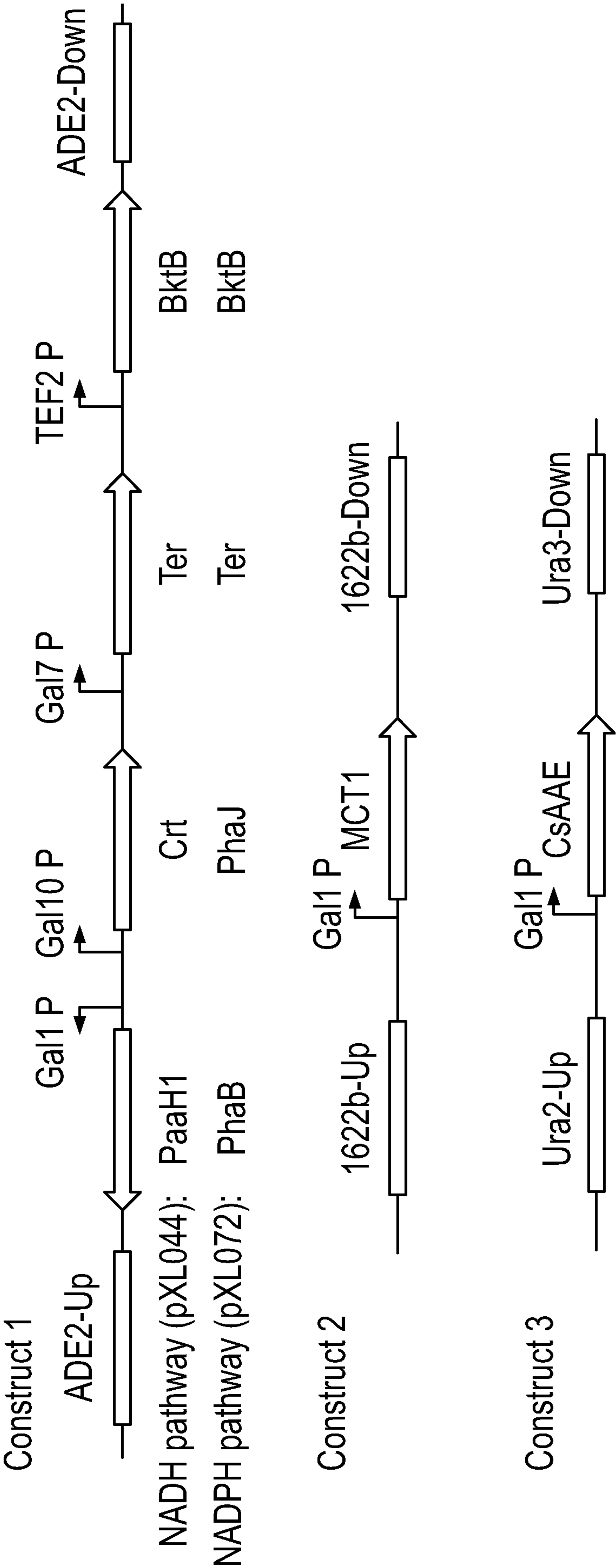
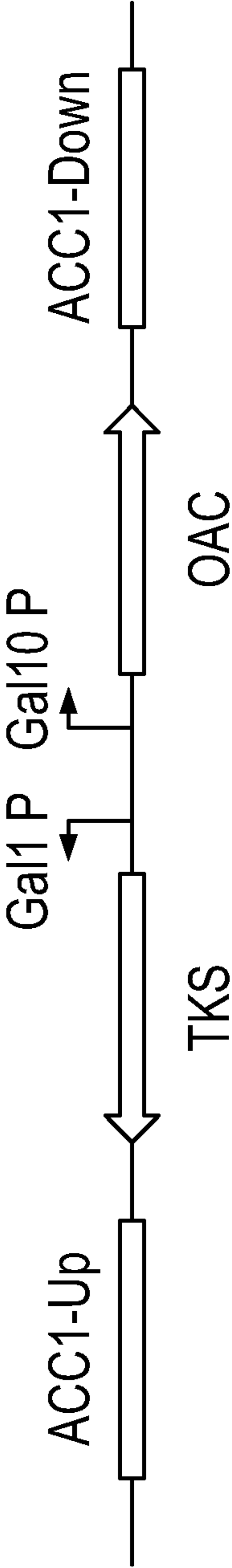


FIG. 4

Construct 4
Design 1:



Construct 5
Design 2:

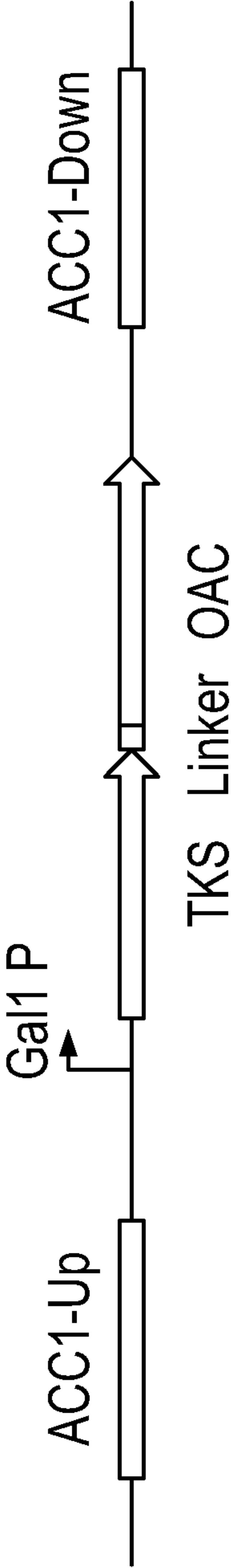


FIG. 5

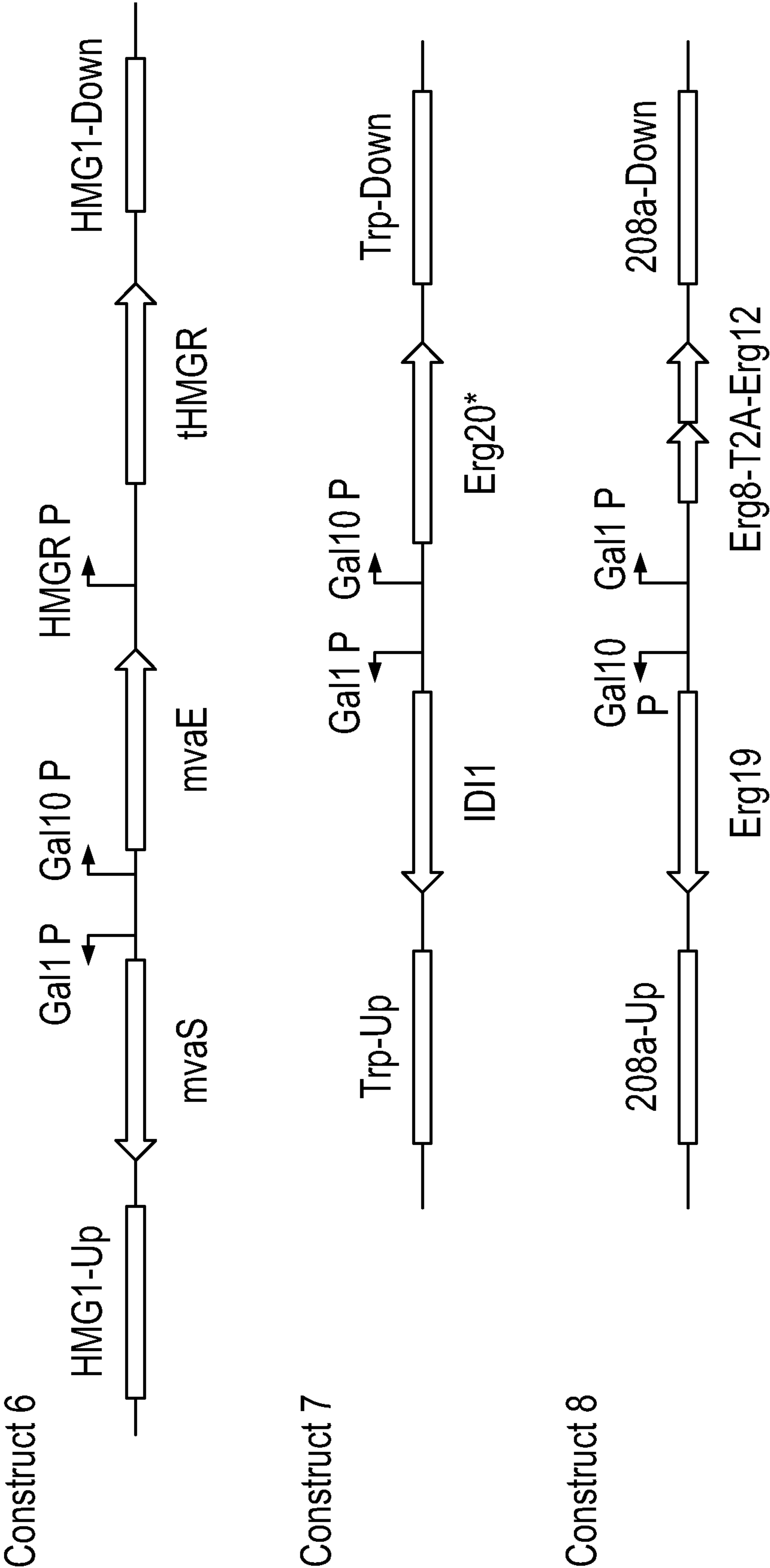


FIG. 6

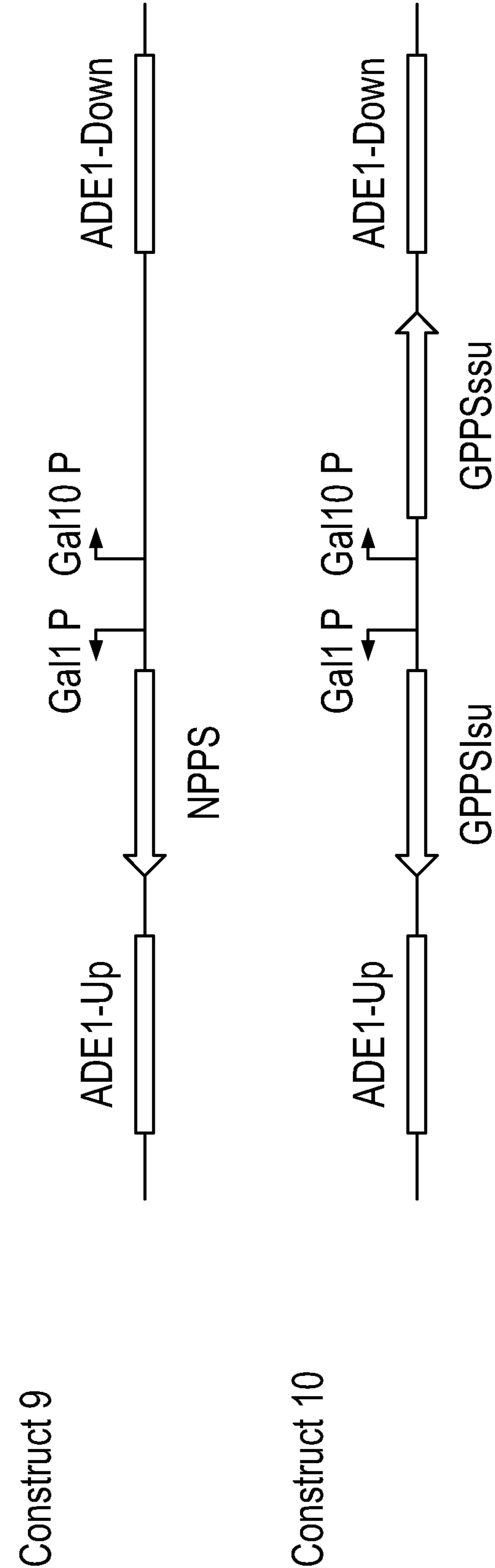


FIG. 7

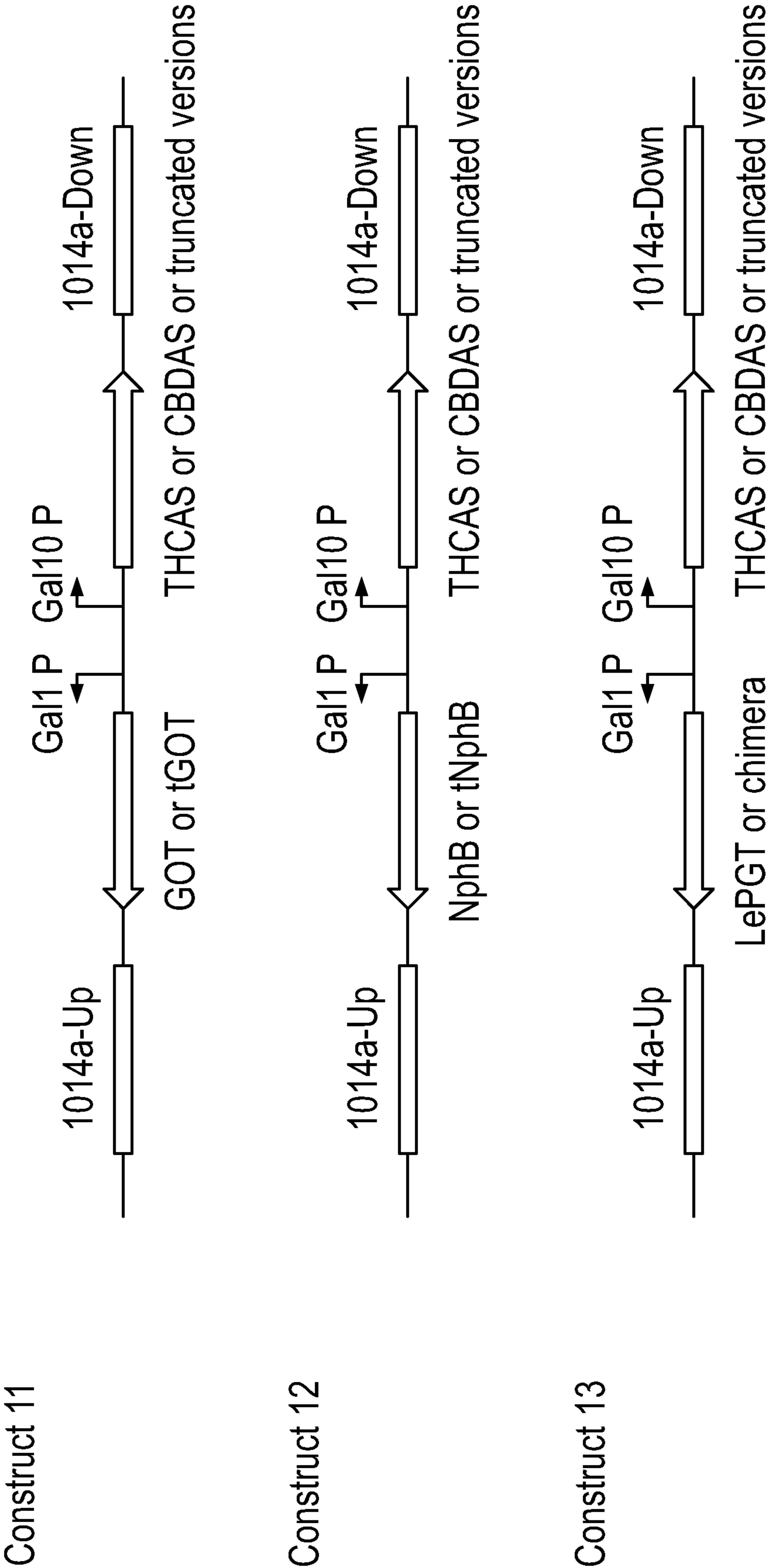
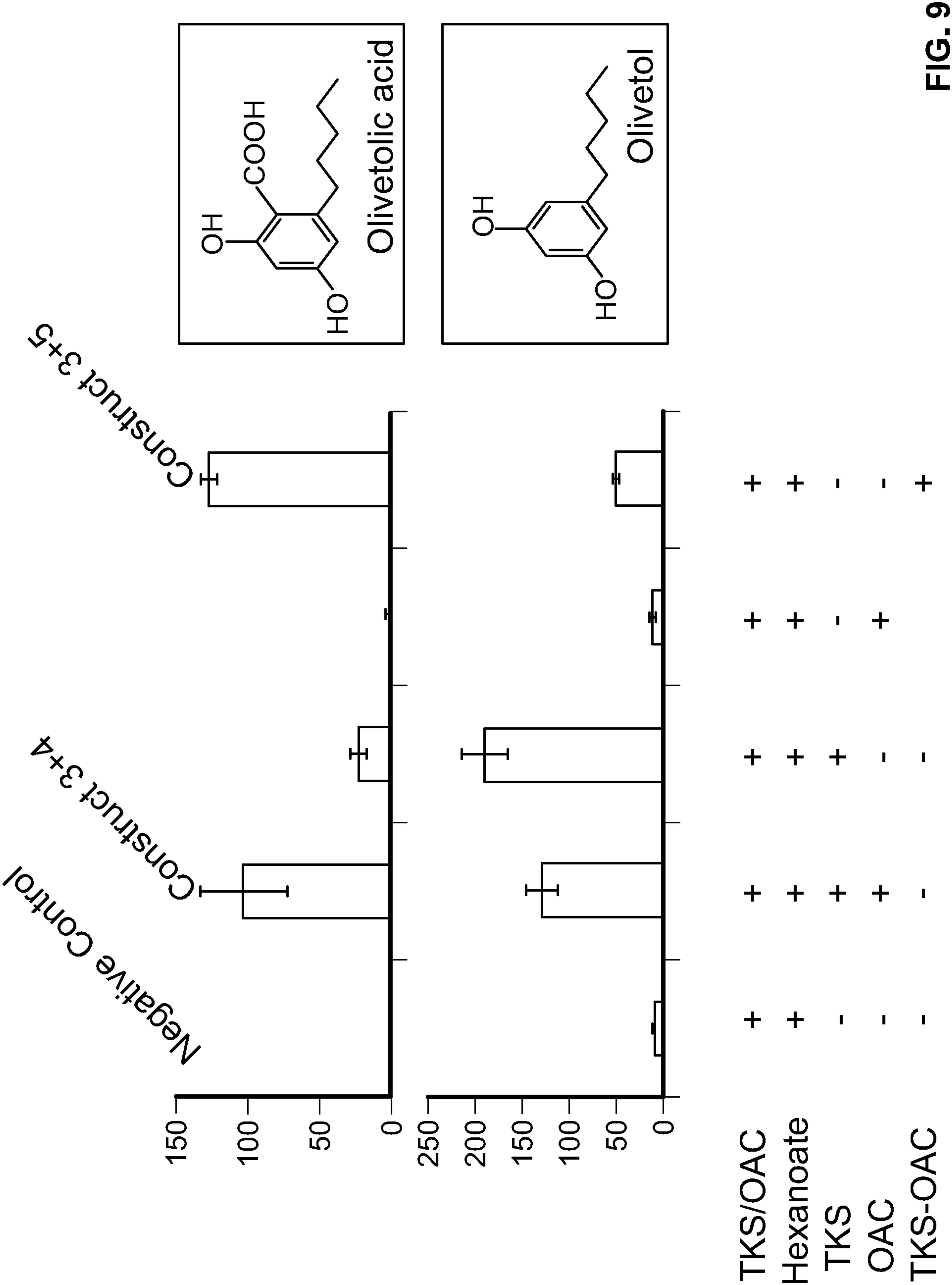


FIG. 8



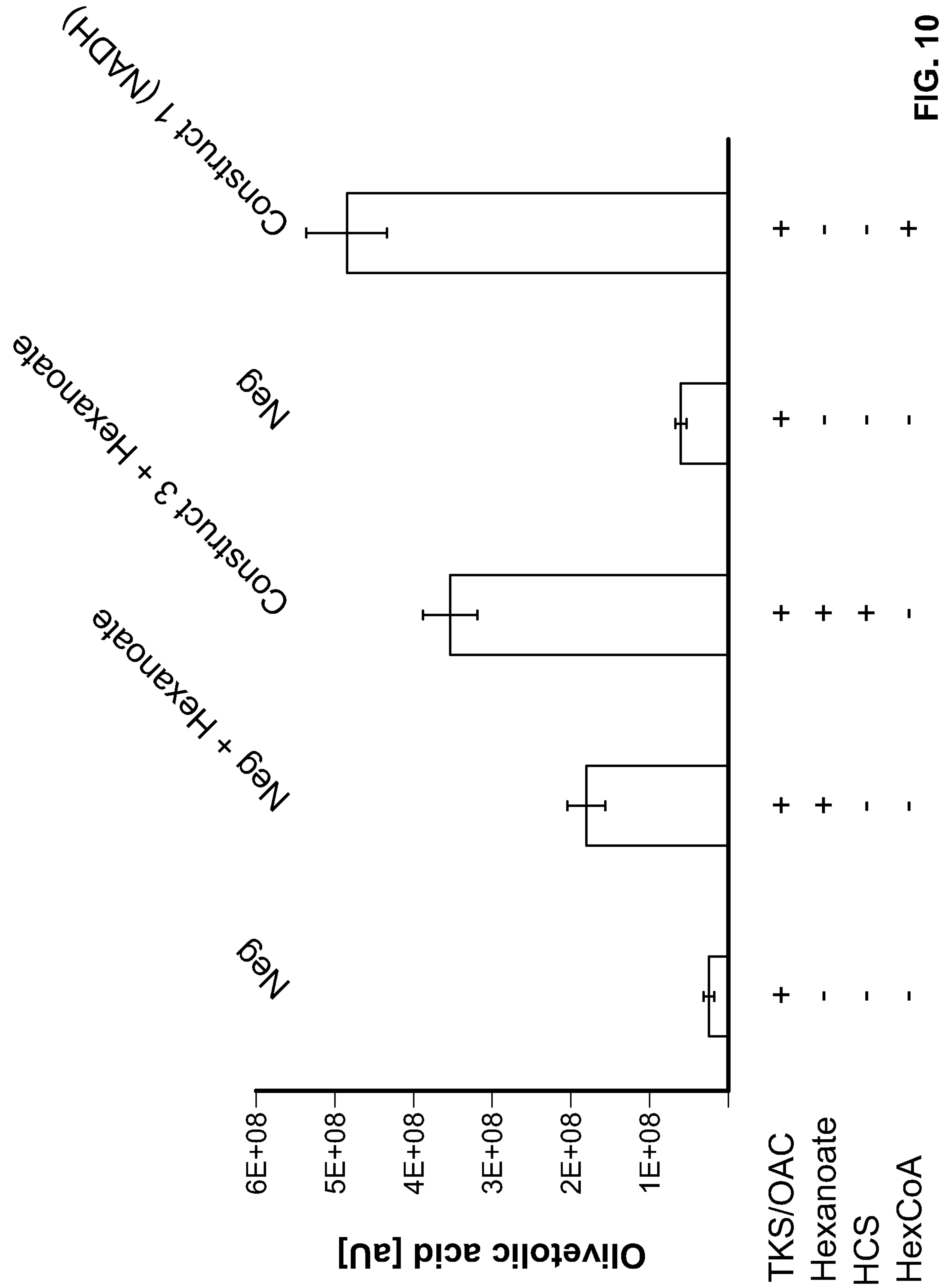


FIG. 10

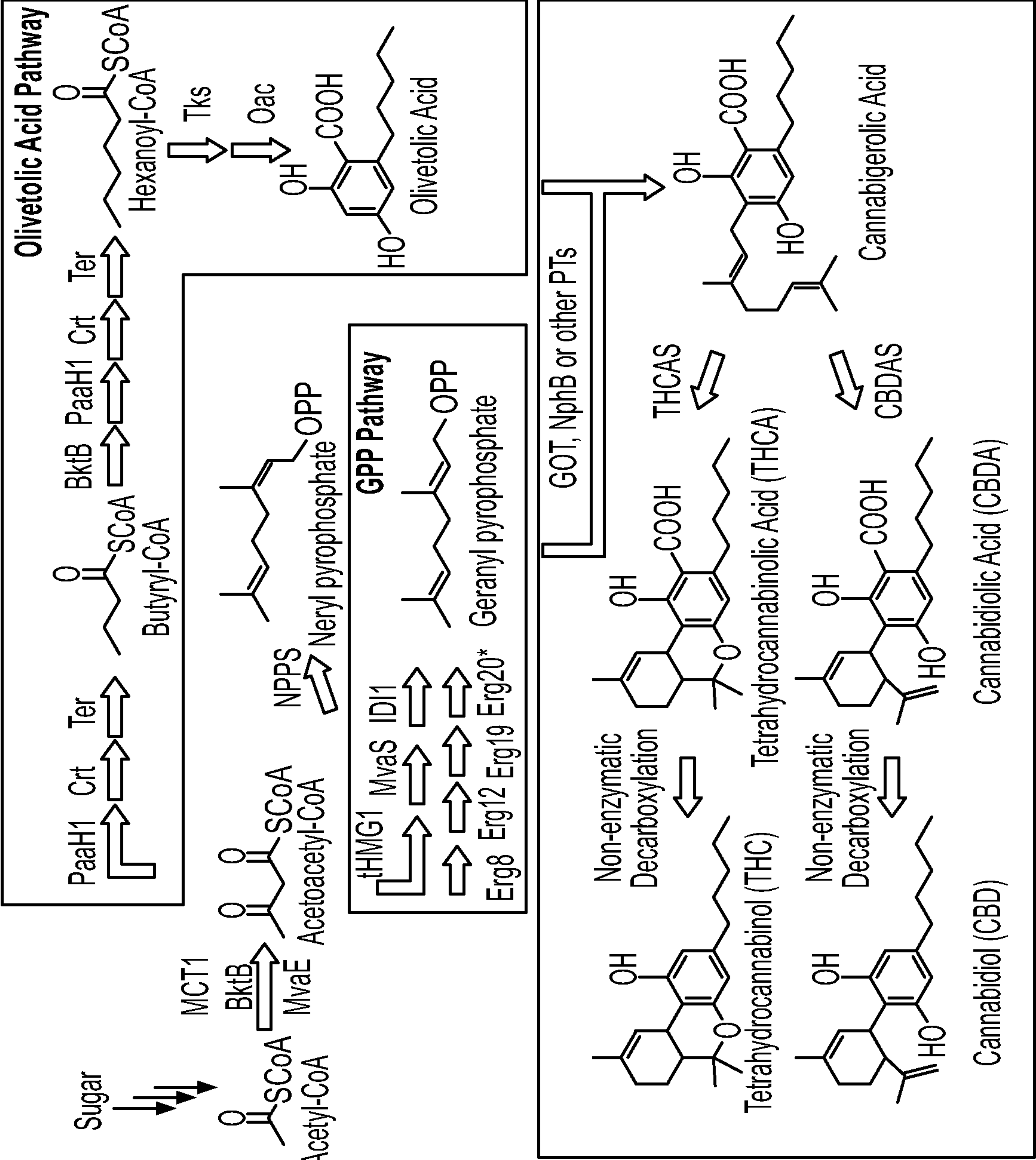


FIG. 11

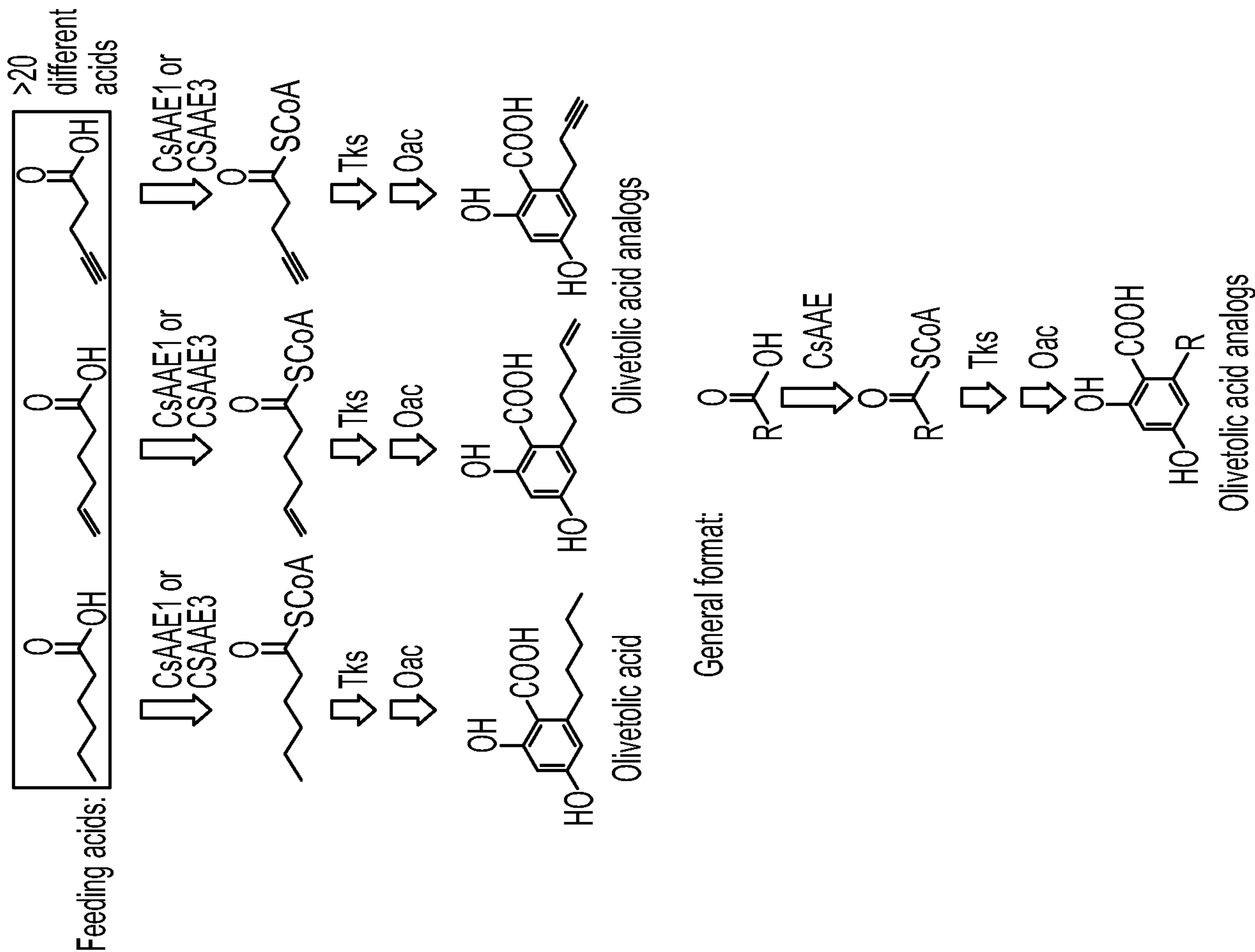


FIG. 11 (Cont.)

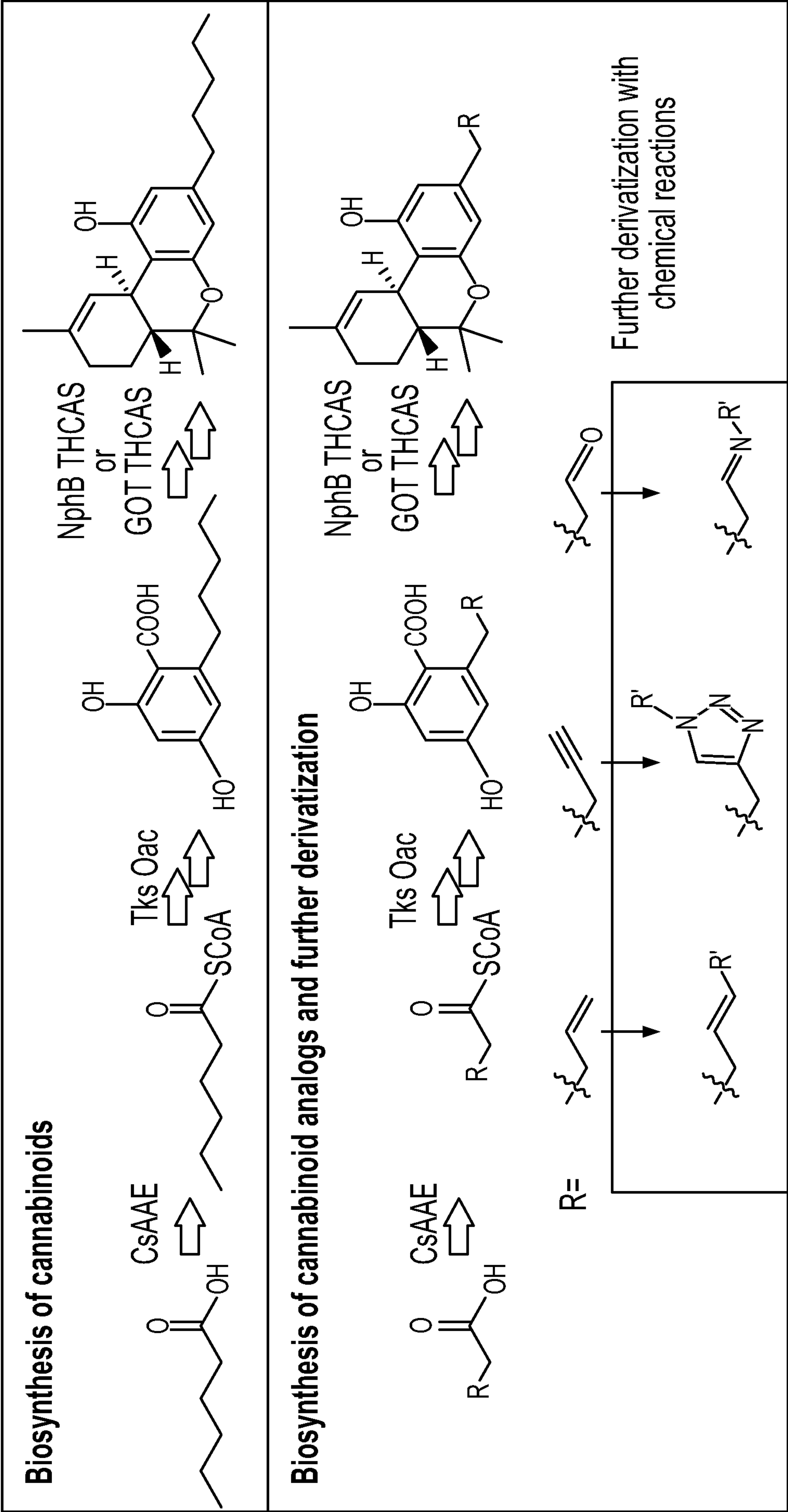


FIG. 13

14/101

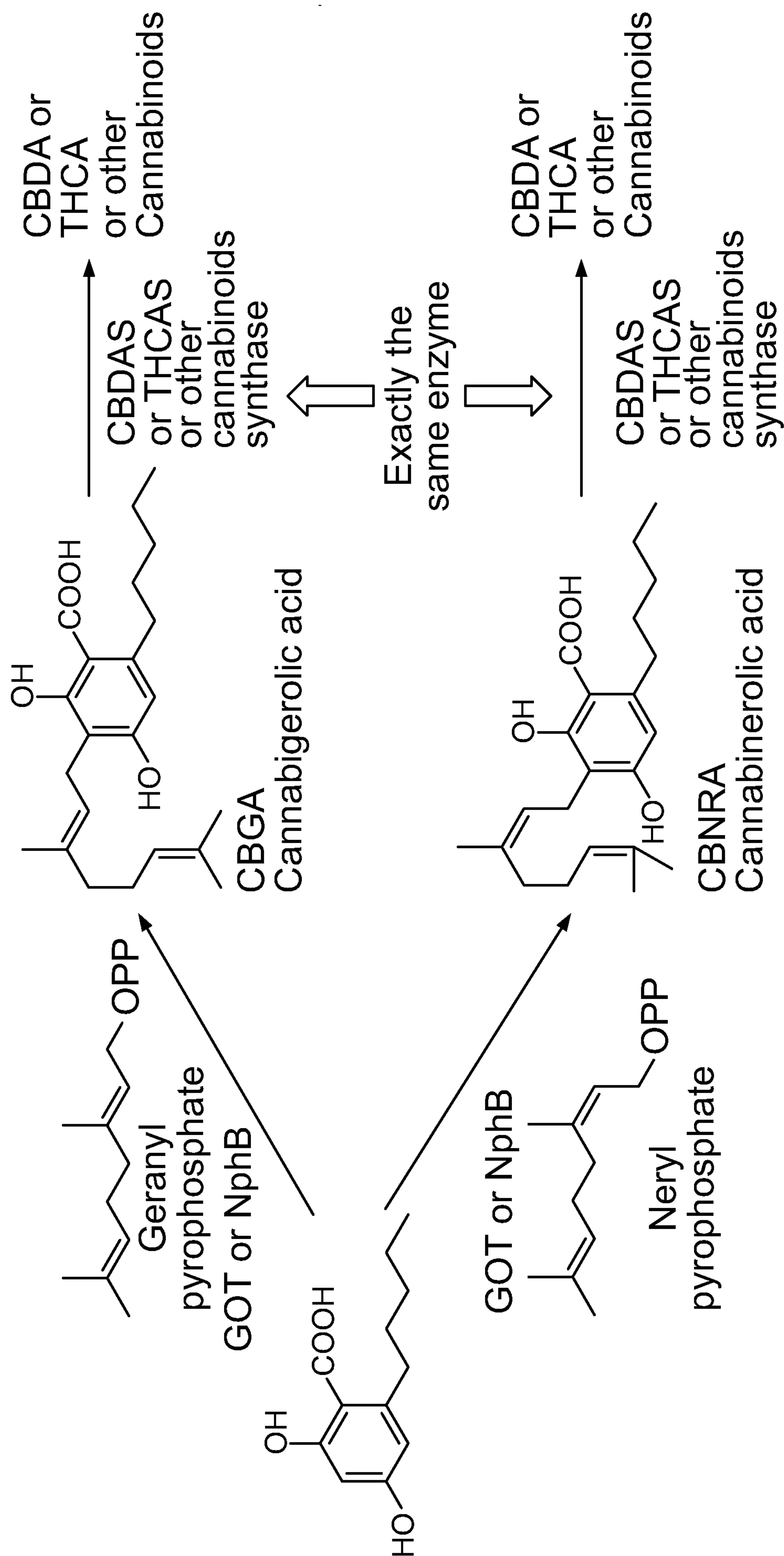
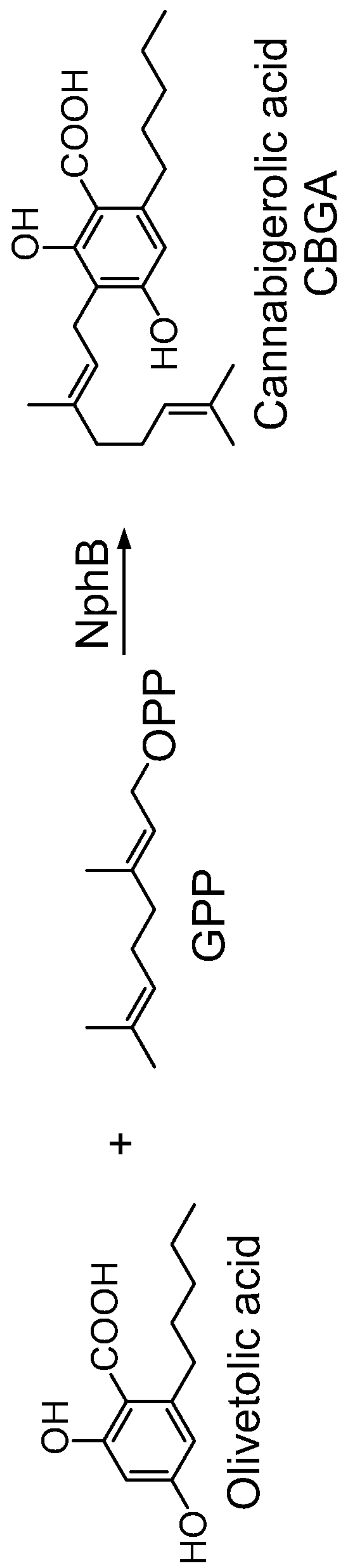


FIG. 14

15/101



Titer of Cannabigerolic acid

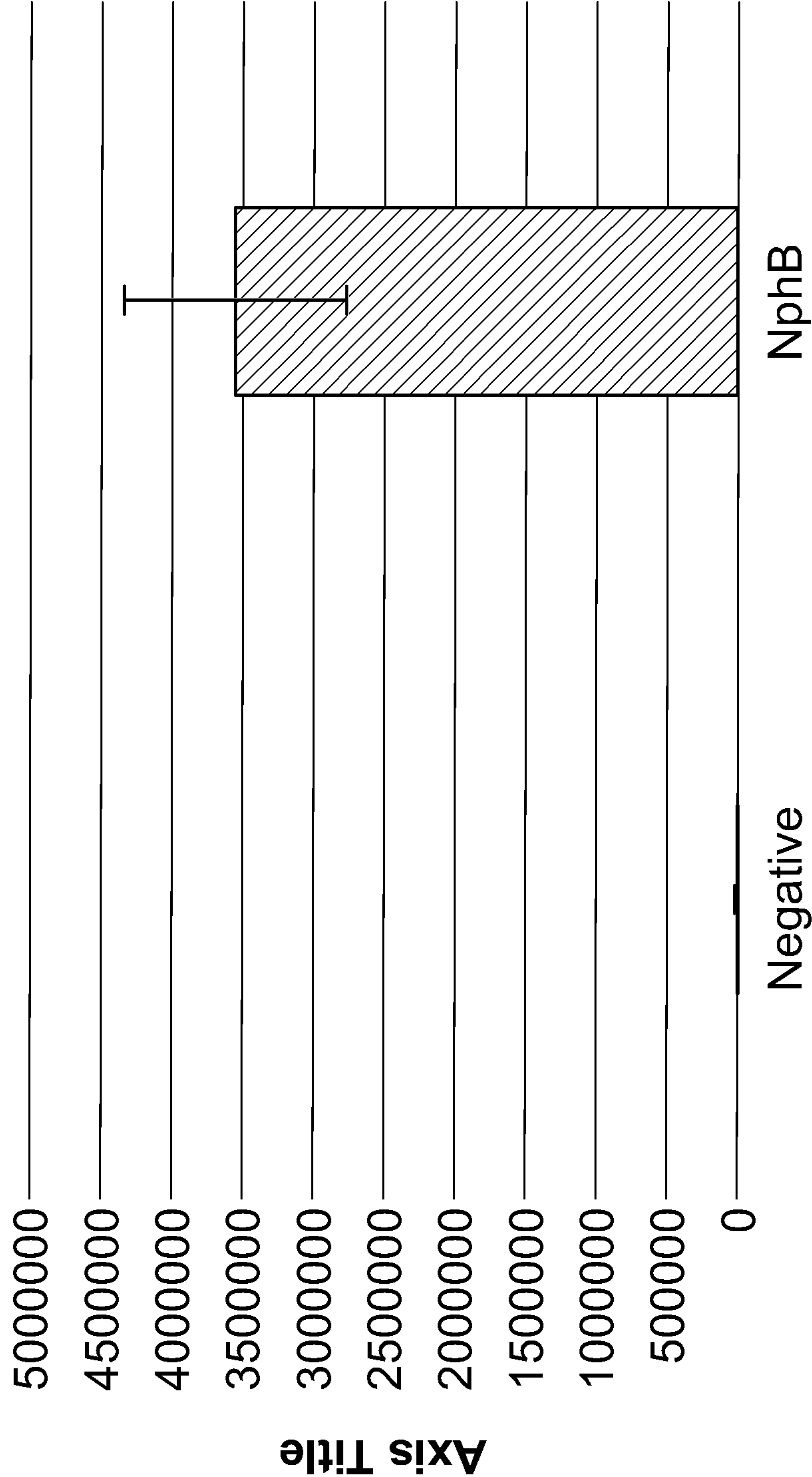


FIG. 15

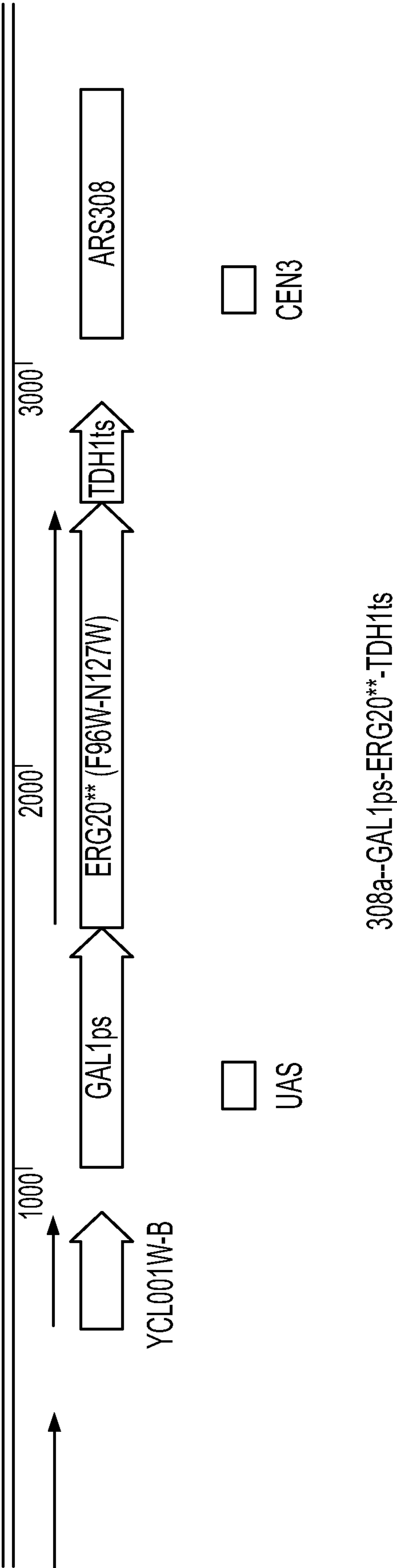
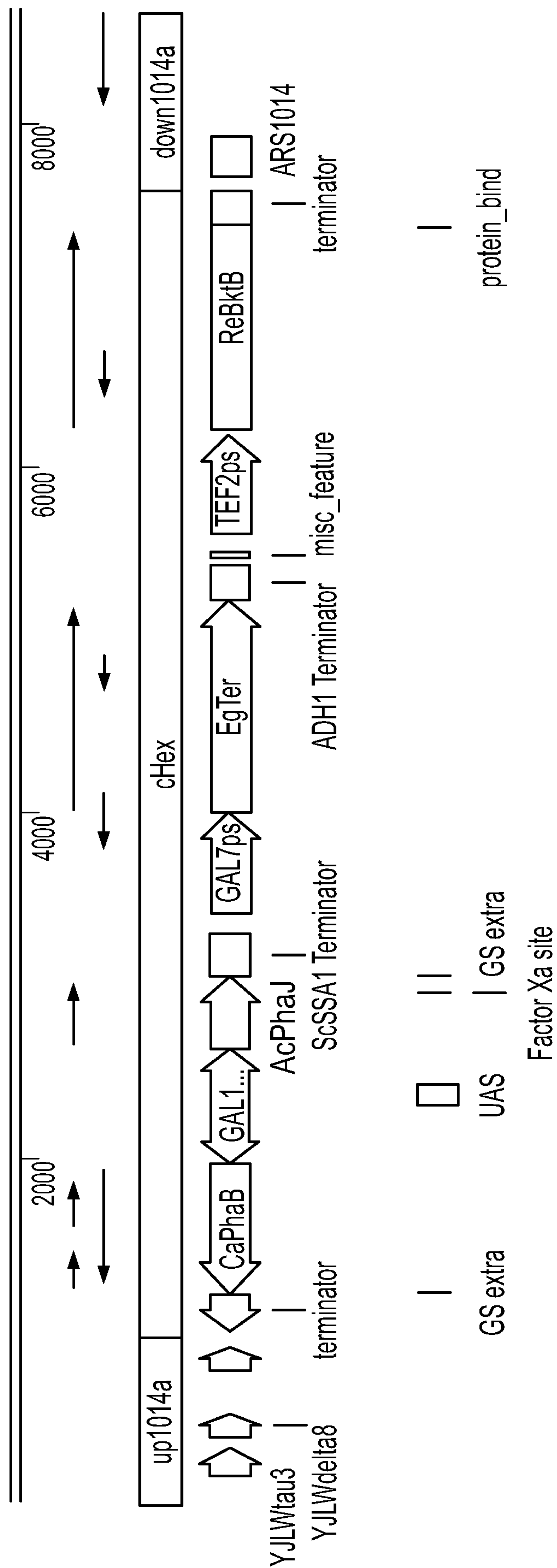


FIG. 16

17/101



1014a--cHexCoA

FIG. 17

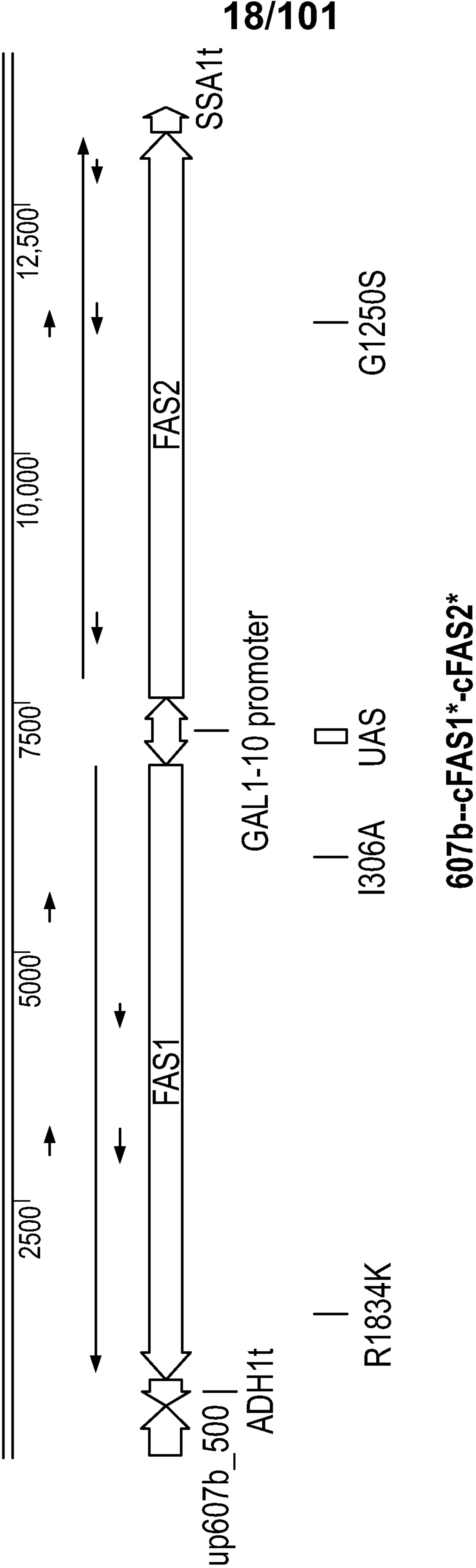


FIG. 18

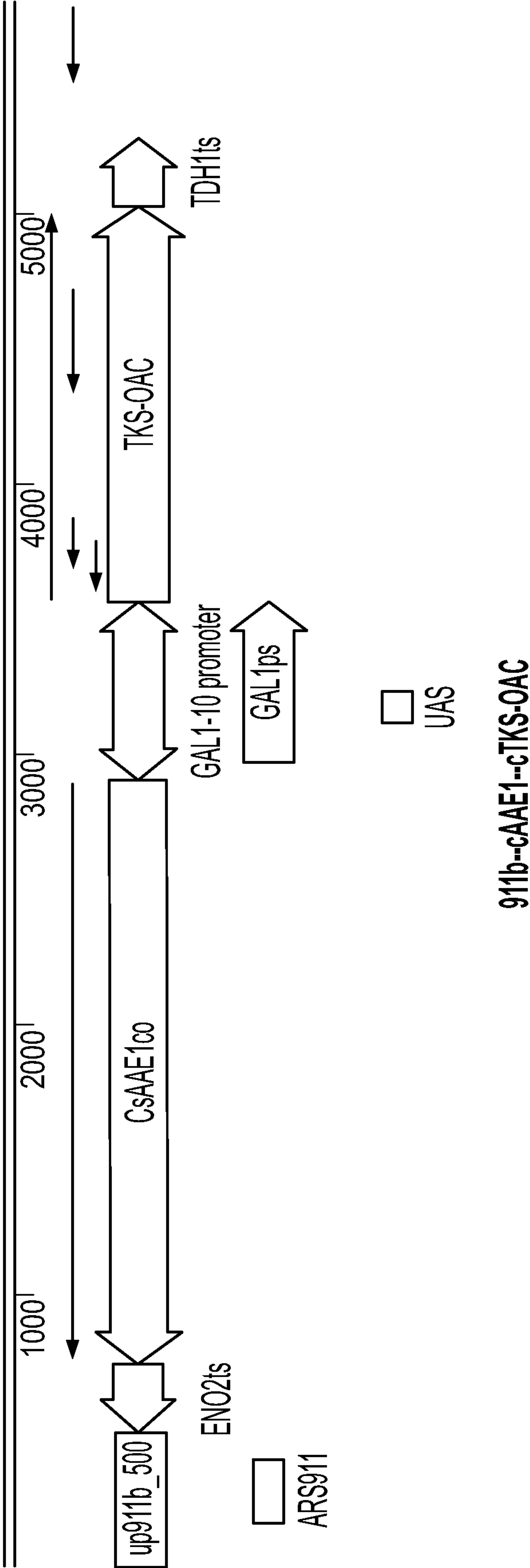
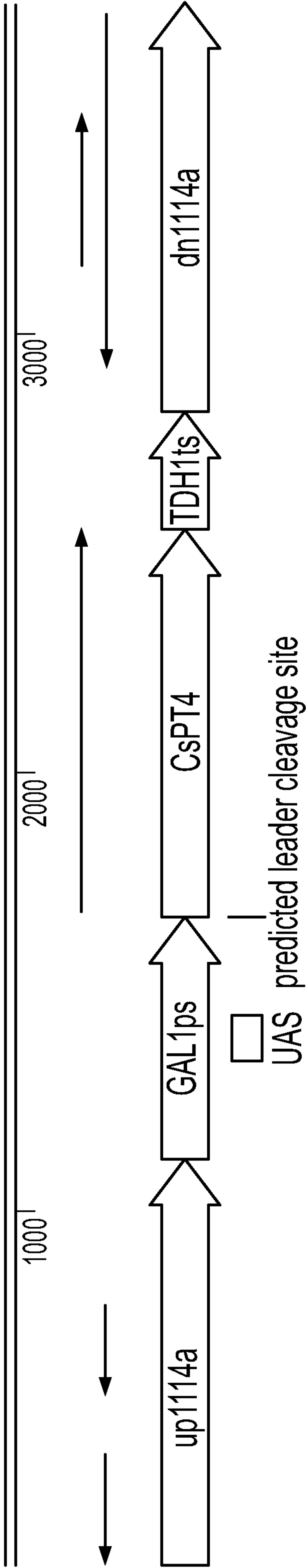


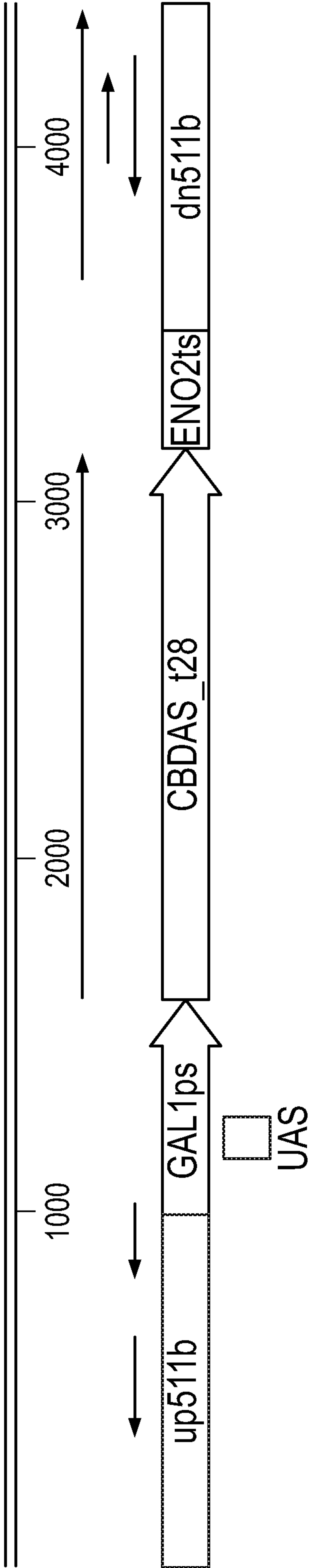
FIG. 19

20/101



1114a--tCsPT4

FIG. 20



511b--GAL1ps-CBDAS t28-ENO2ts

FIG. 21

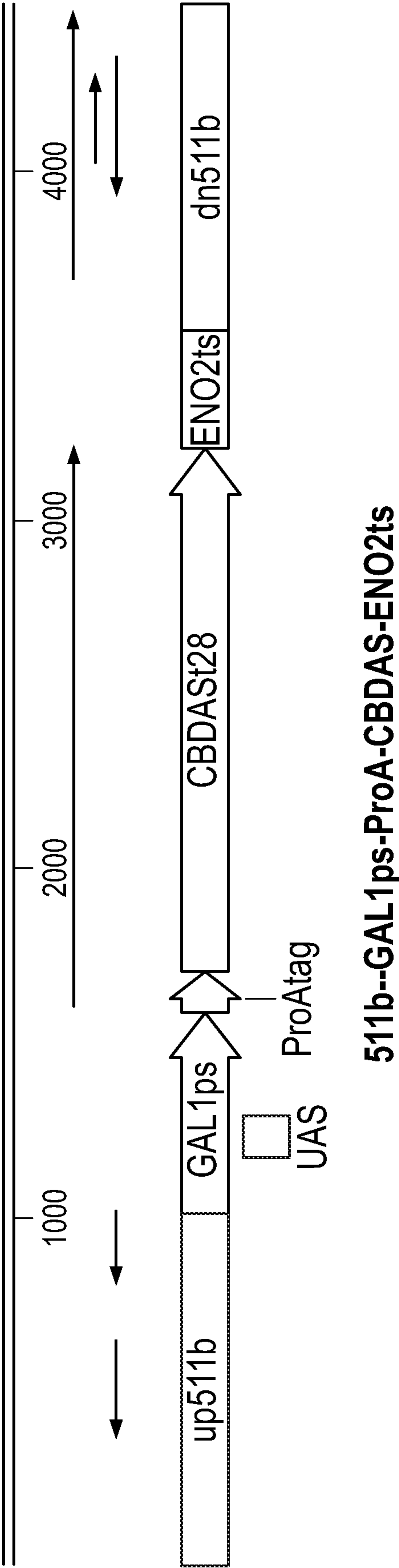


FIG. 22

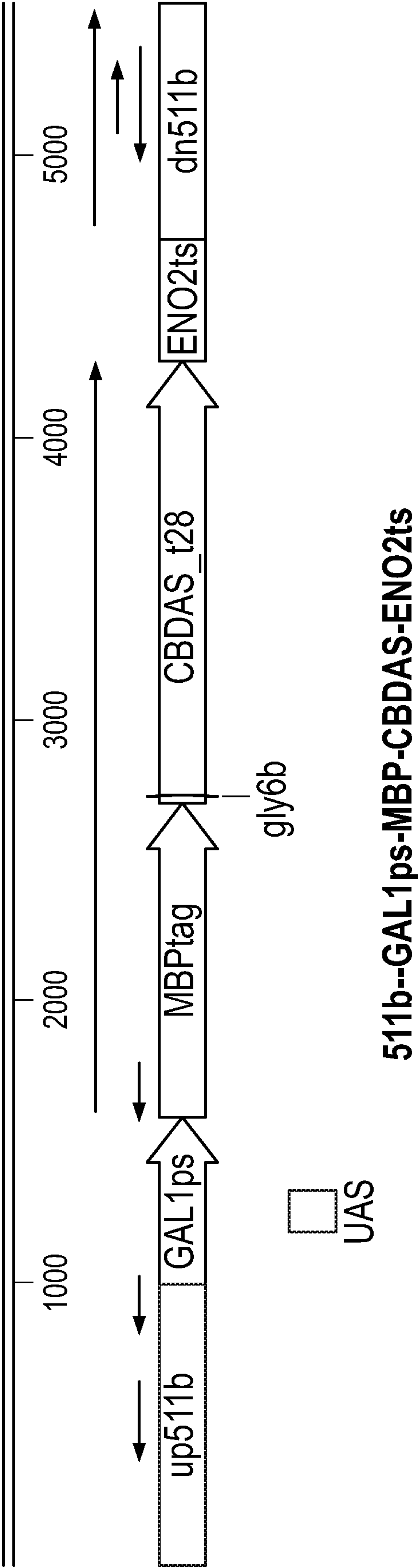


FIG. 23

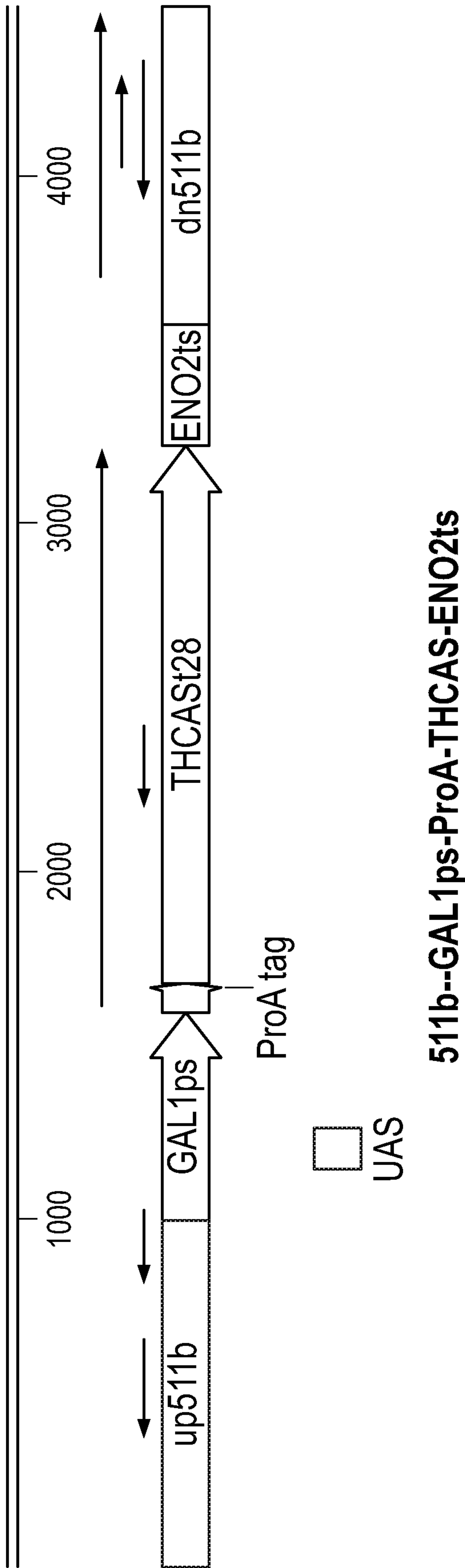
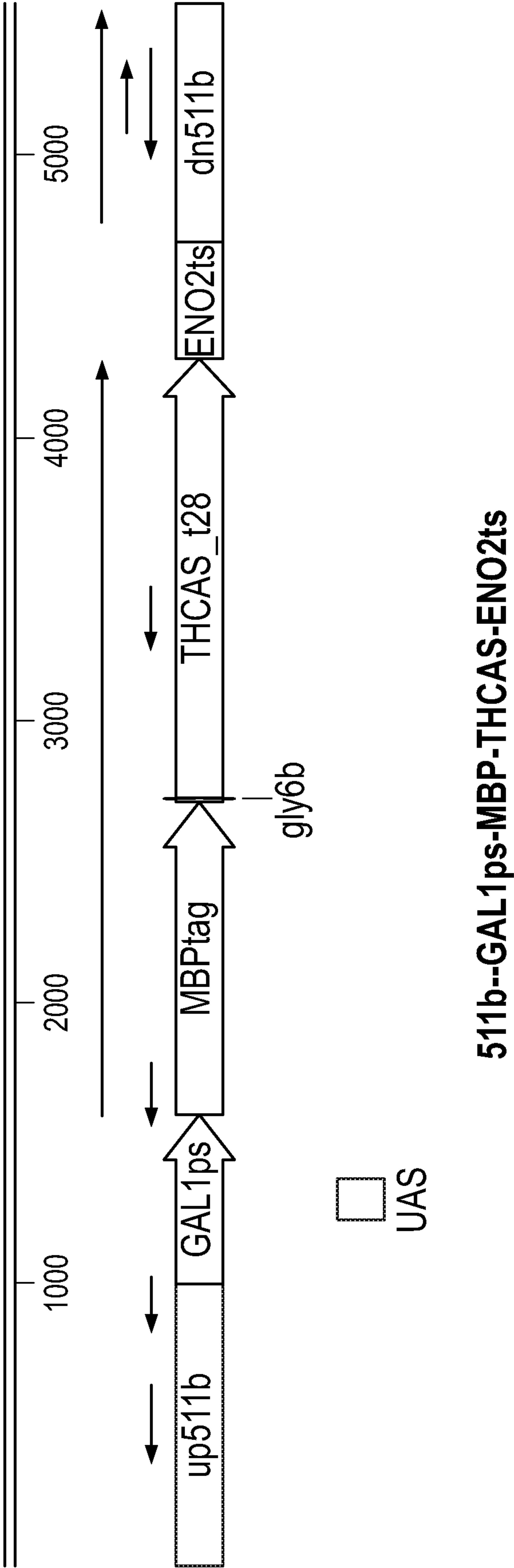


FIG. 24



511b--GAL1ps-MBP-THCAS-ENO2ts

FIG. 25

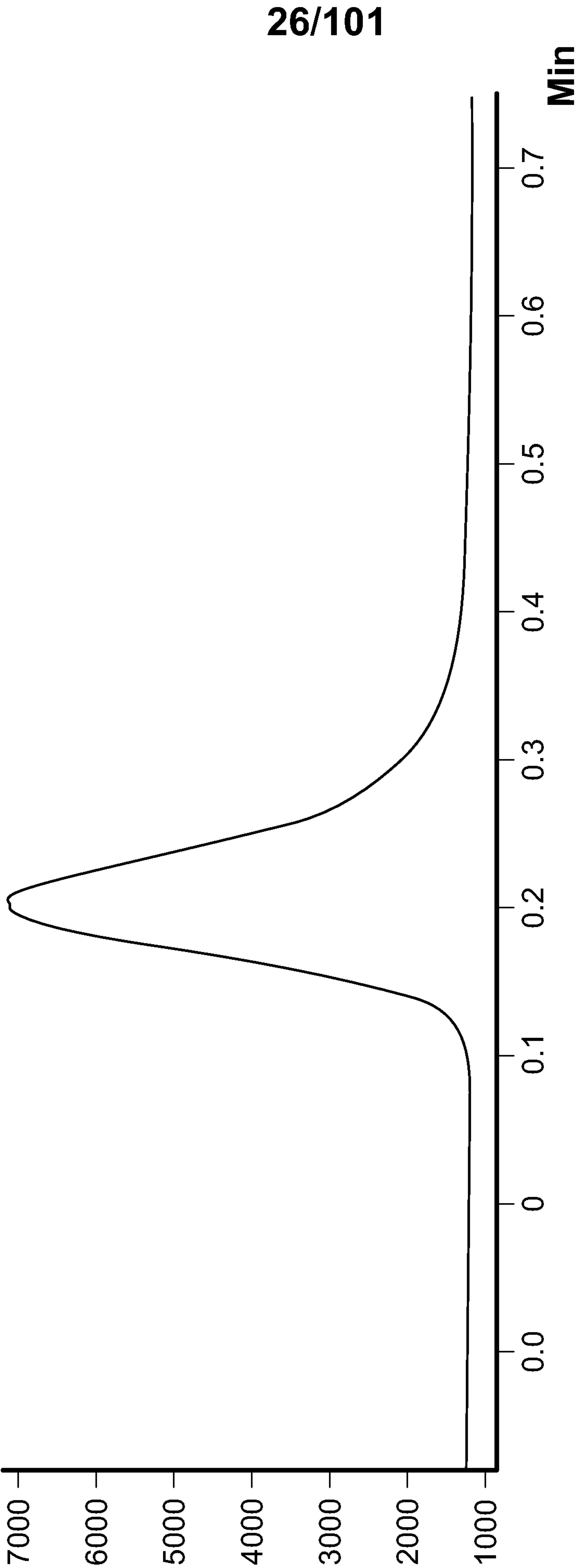


FIG. 26A

27/101

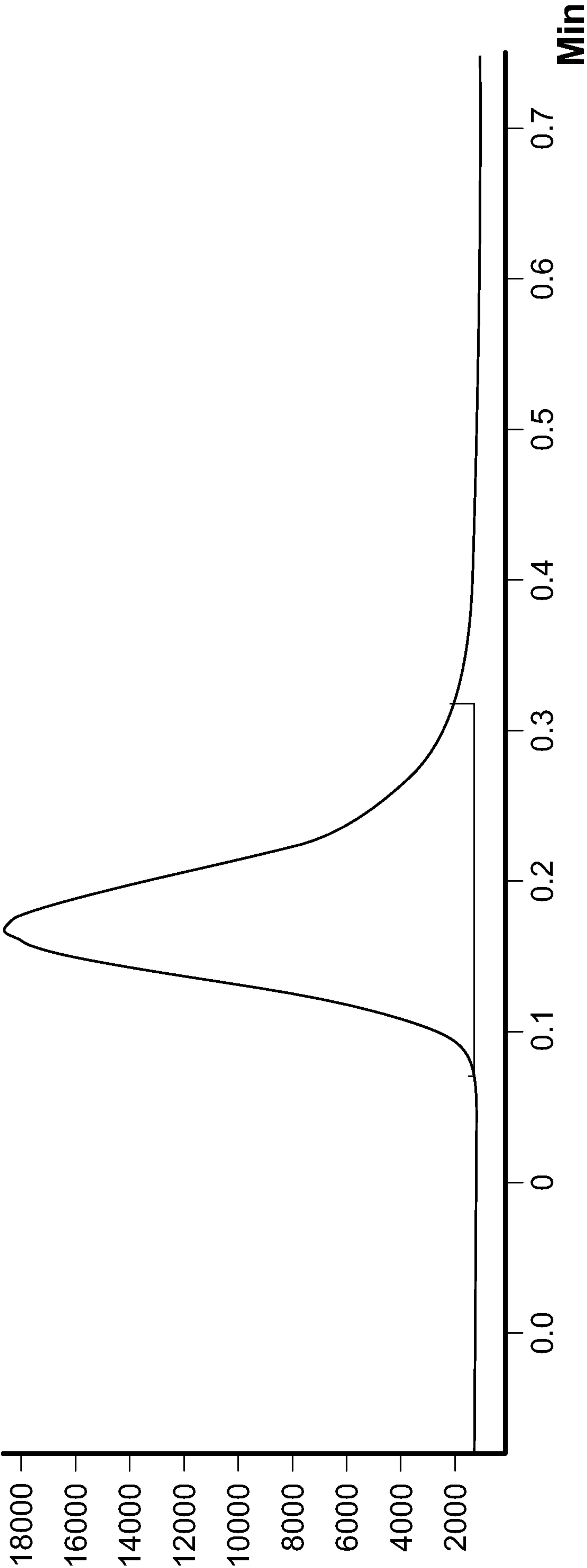


FIG. 26B

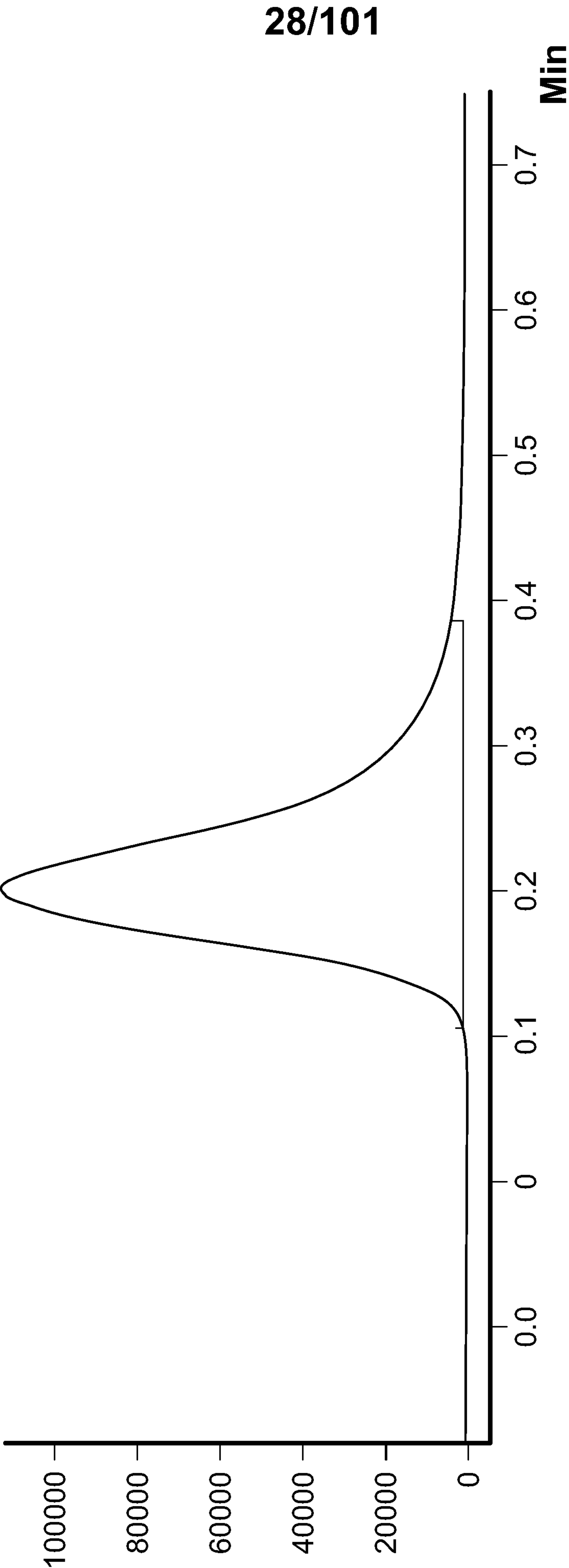


FIG. 26C

THCA synthase with N-terminal truncation and ProA signal sequence

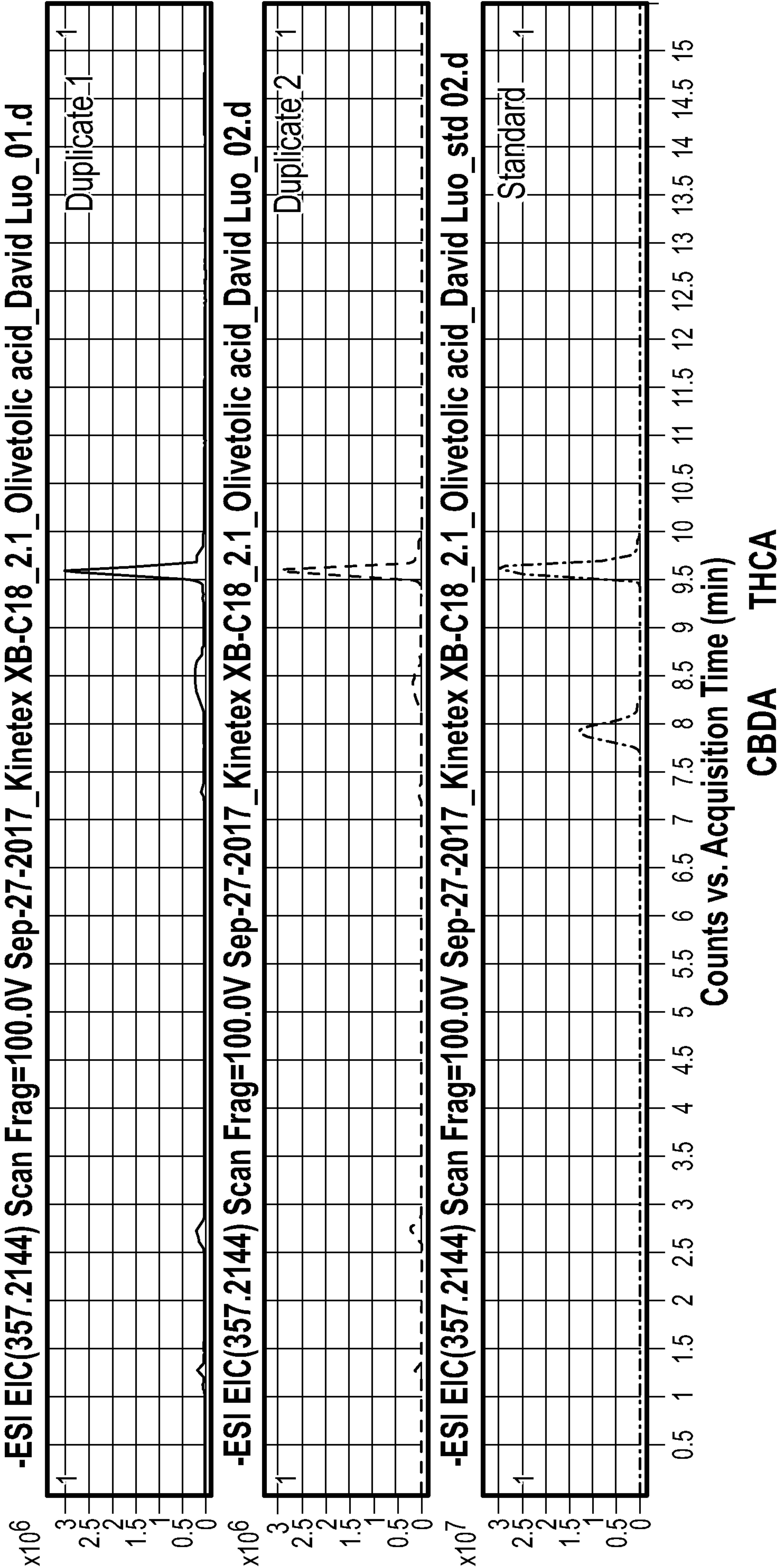


FIG. 27

CBDA synthase with N-terminal truncation and ProA signal sequence

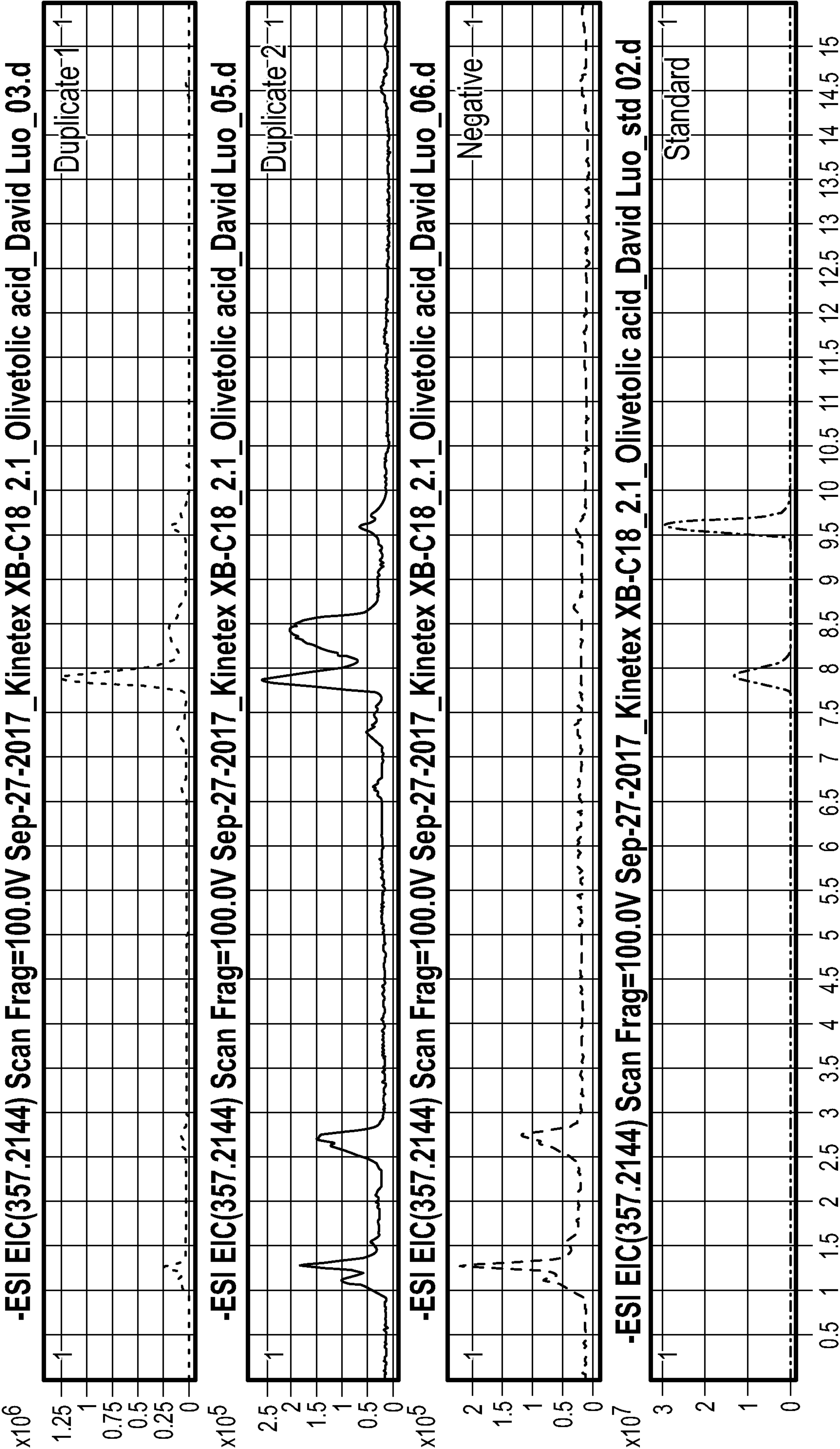
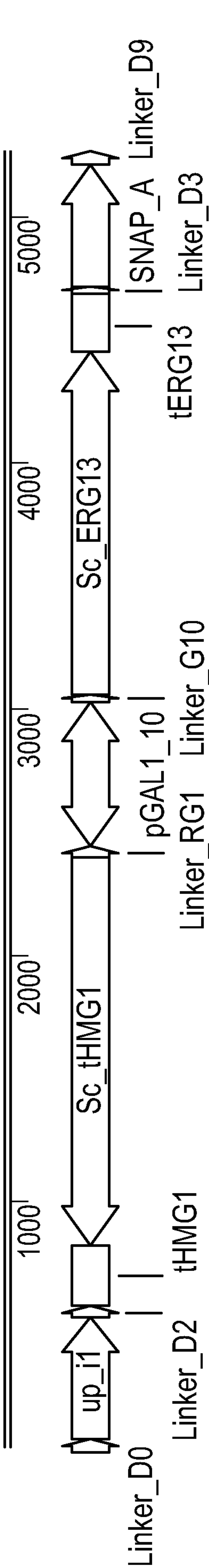
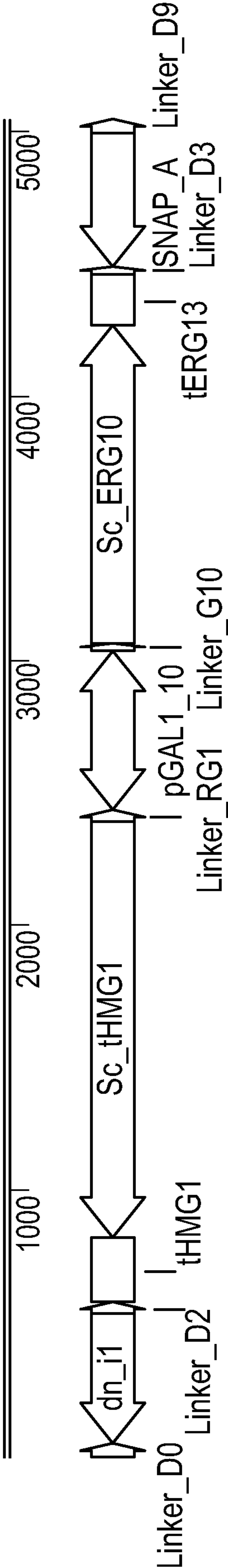


FIG. 28

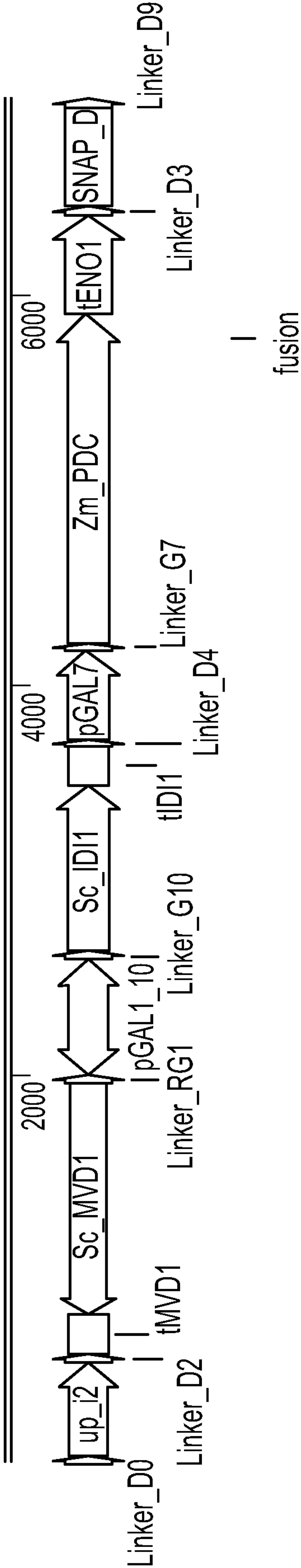


S22_i1_GPP_native_1A
5259 bp

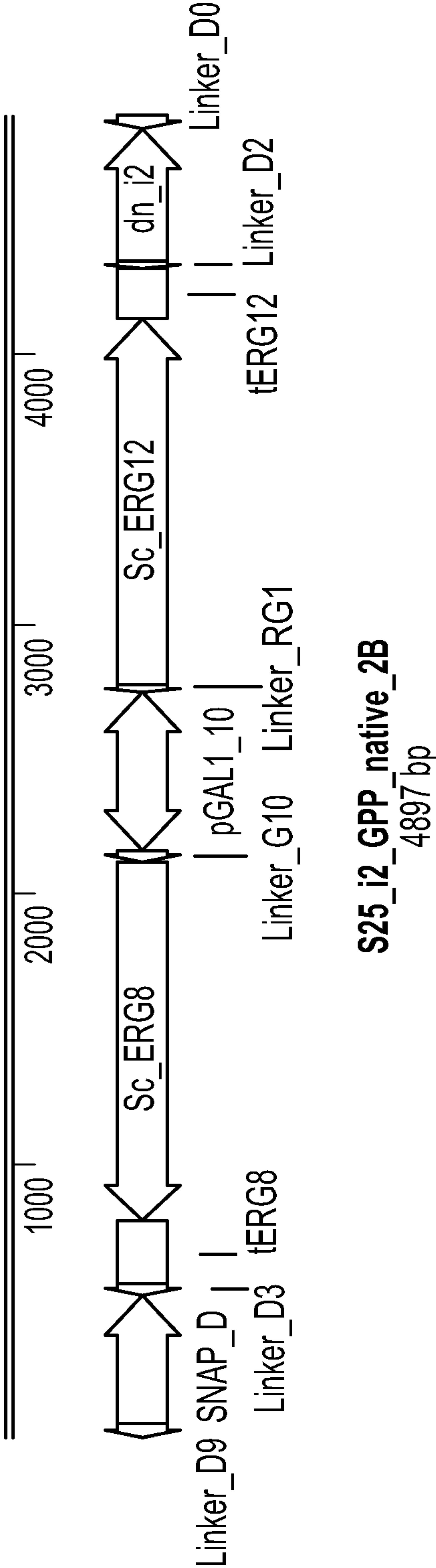


S23_i1_GPP_native_1B
5028 bp

FIG. 29A

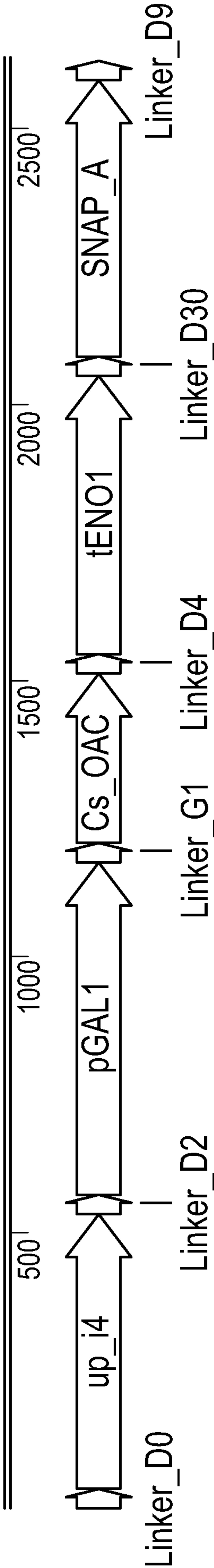


S24_i2_GPP_native_2A
7002 bp

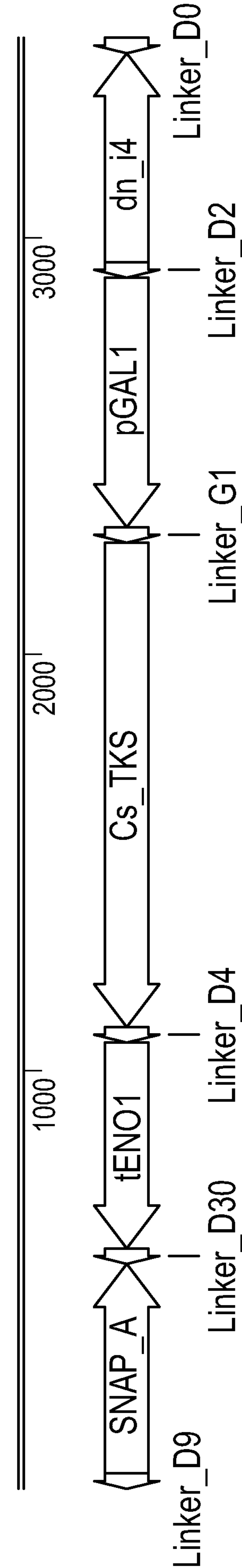


S25_i2_GPP_native_2B
4897 bp

FIG. 29B

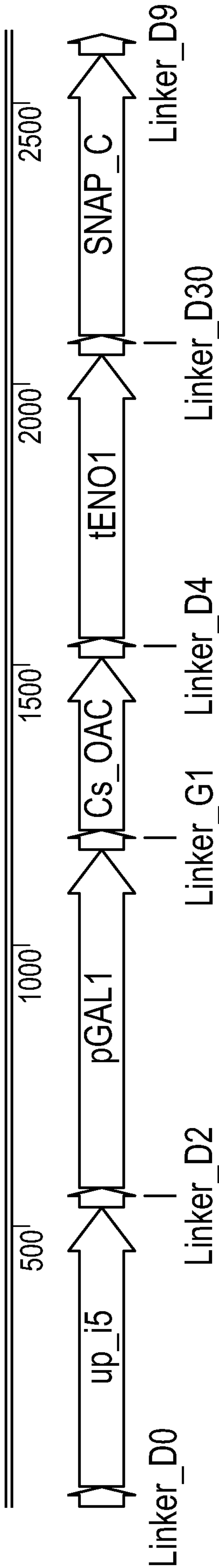


S244_i4A_OAC
2625 bp

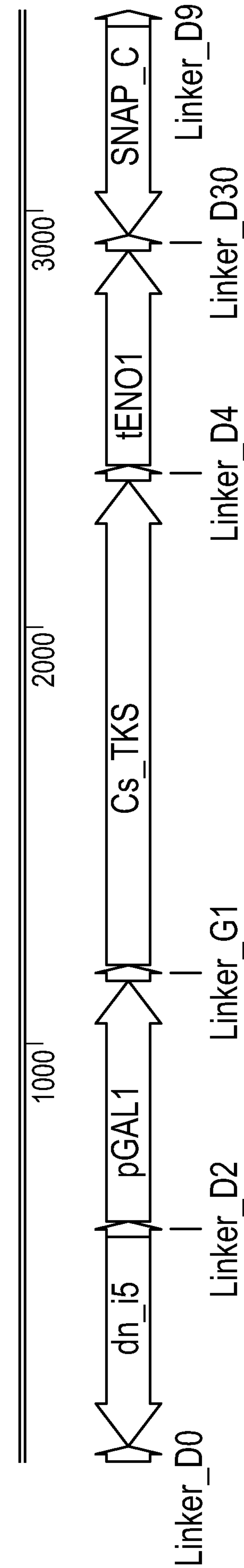


S247_i4B_TKS
3477 bp

FIG. 30A



S248 i5A_OAC
2625 bp



S251 i5B_TKS
3477 bp

FIG. 30B

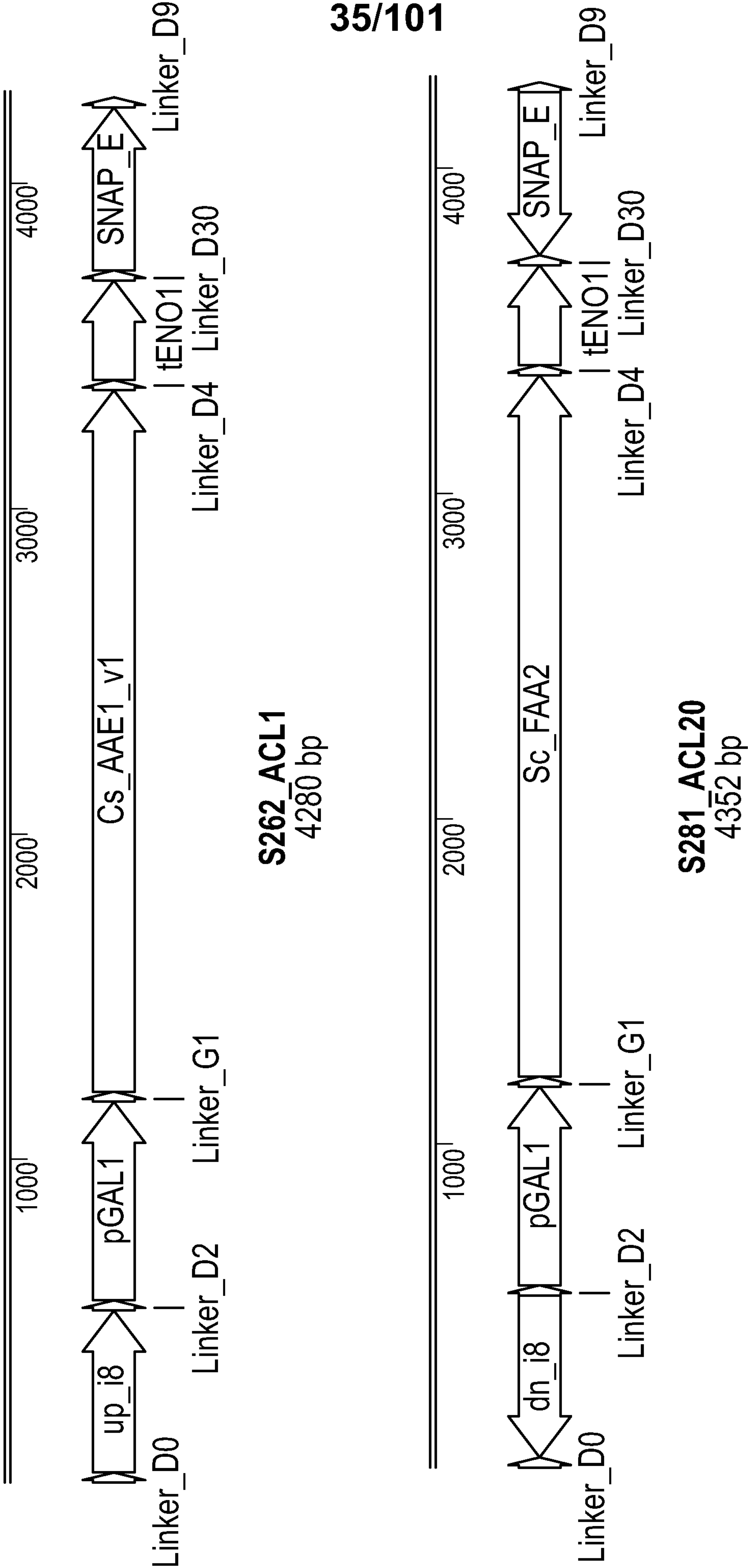


FIG. 30C

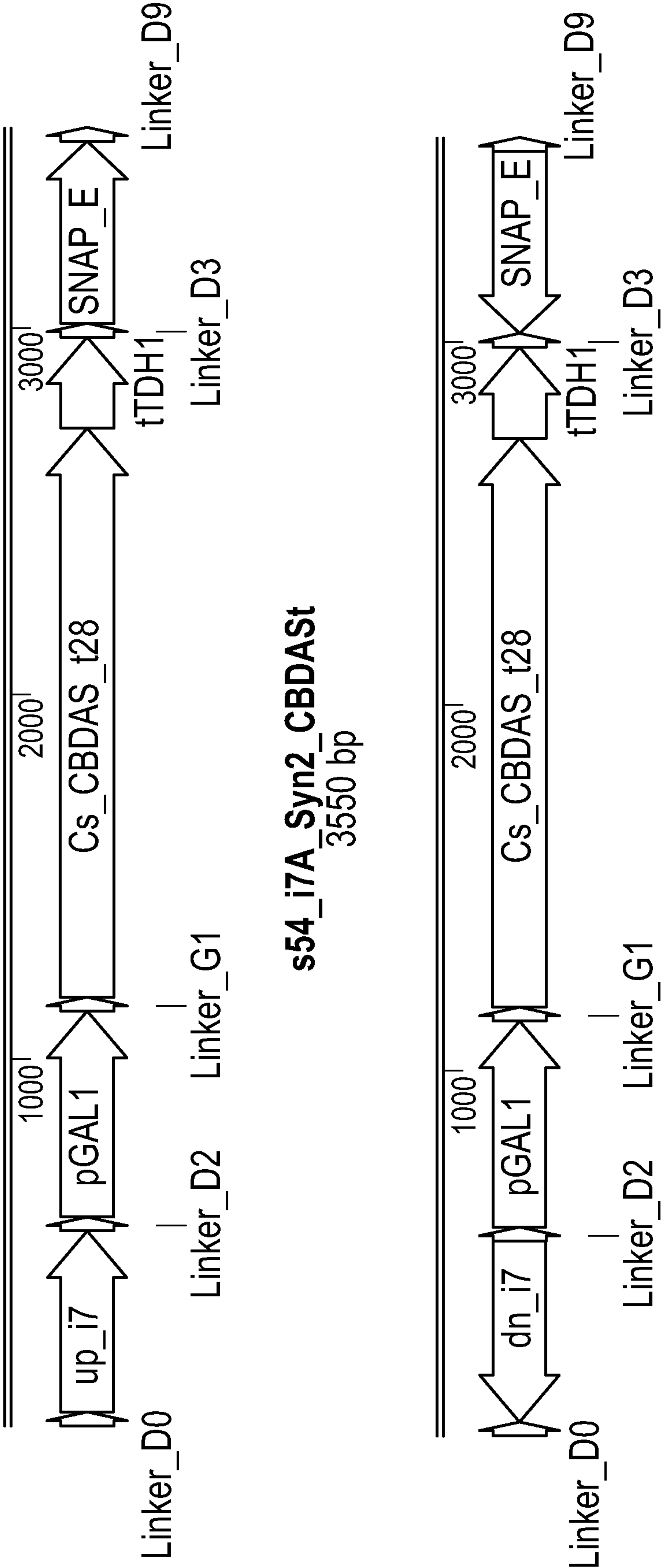


FIG. 31

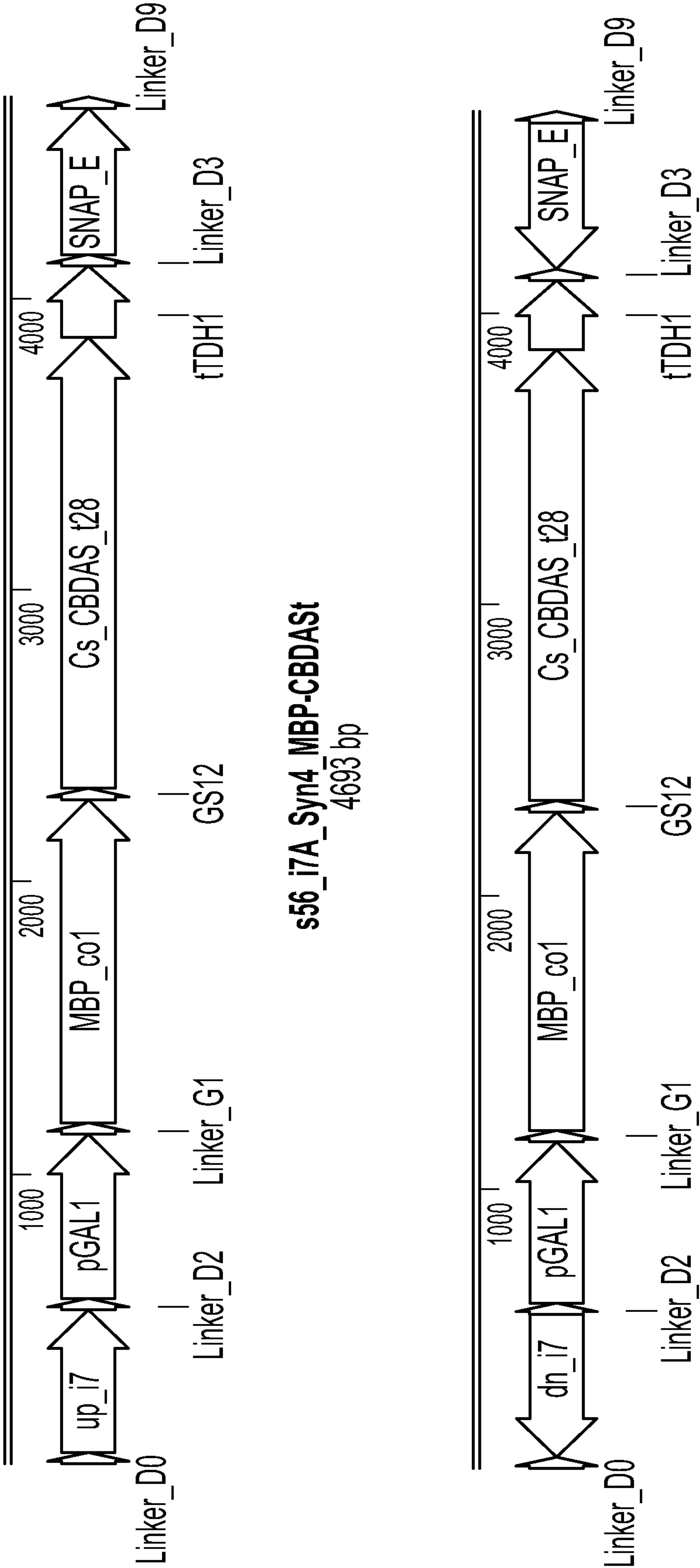


FIG. 32

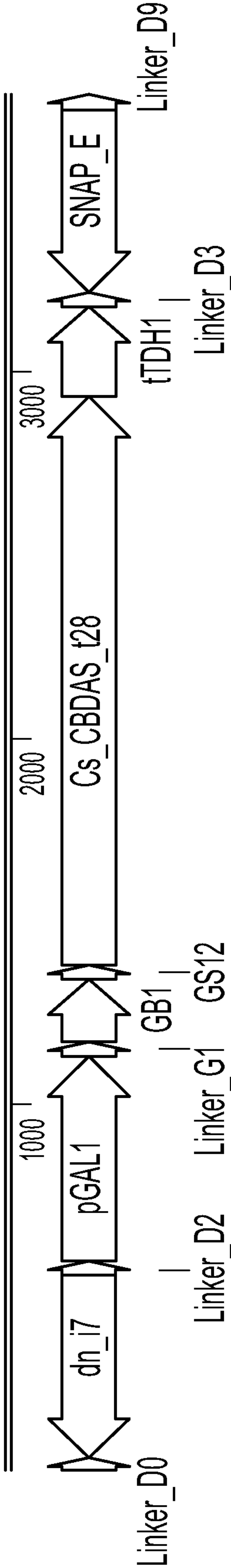
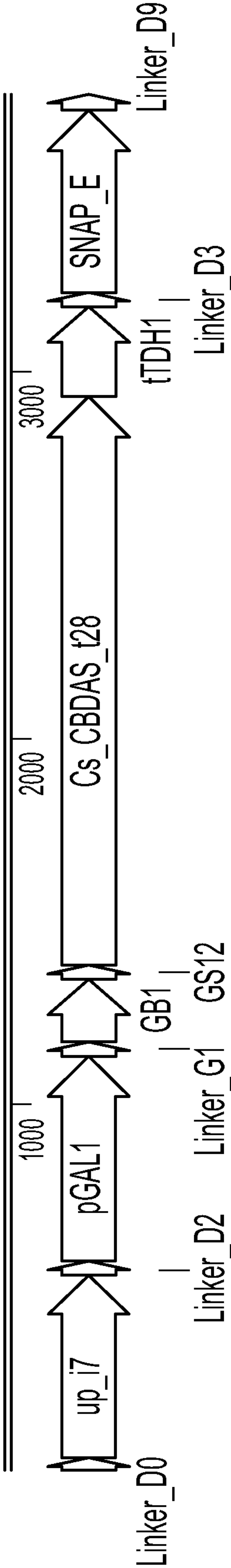
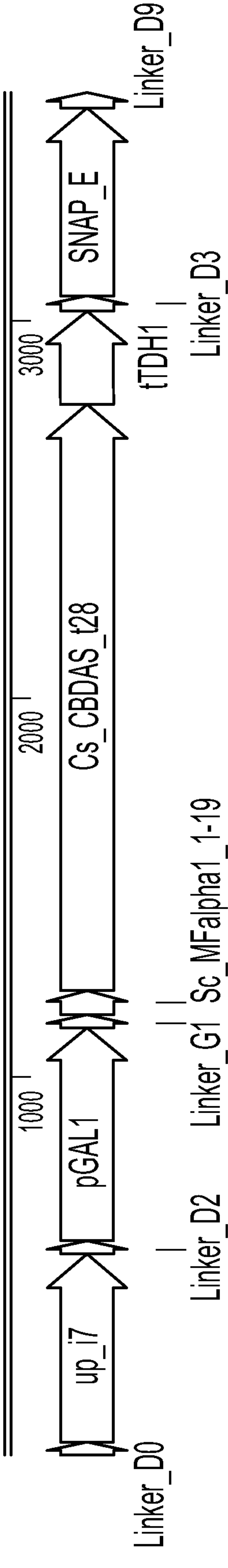
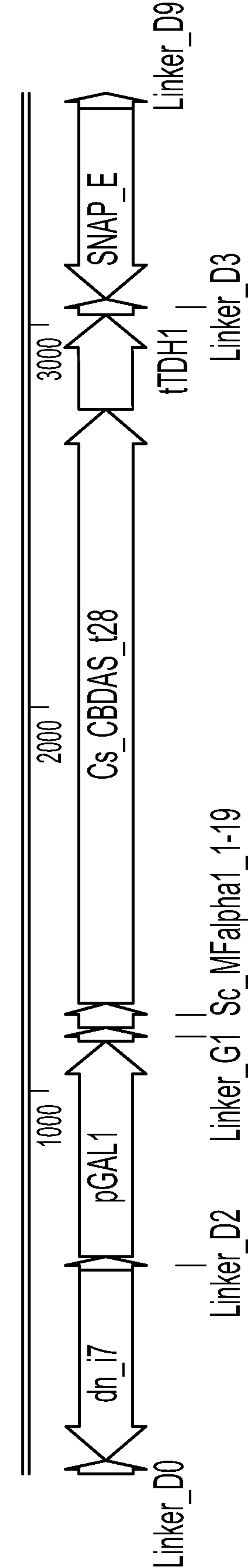


FIG. 33



s58_i7A_Syn7_MFer-CBDAST



s72_i7B_Syn7_MFer-CBDAST

FIG. 34

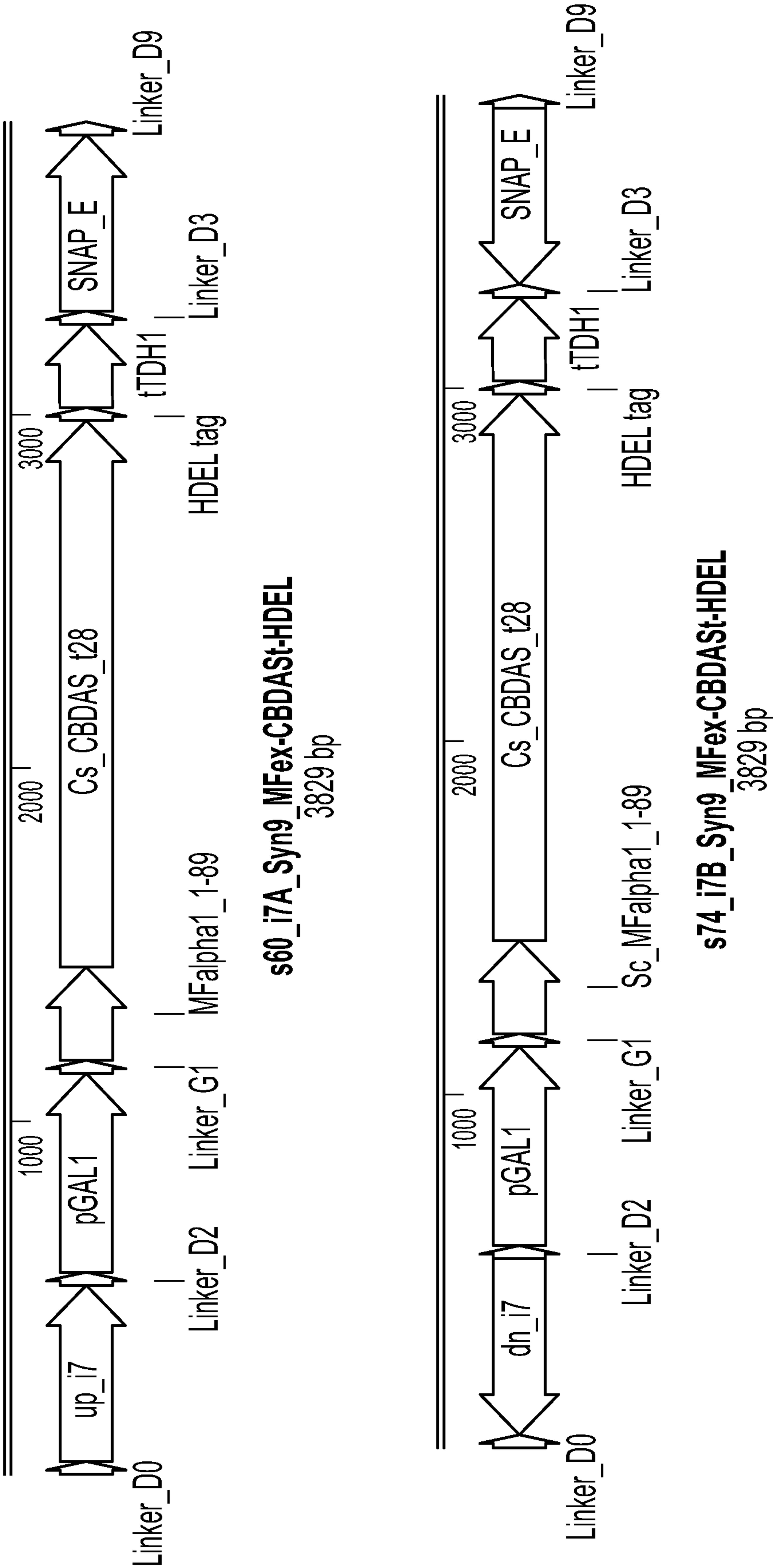


FIG. 35

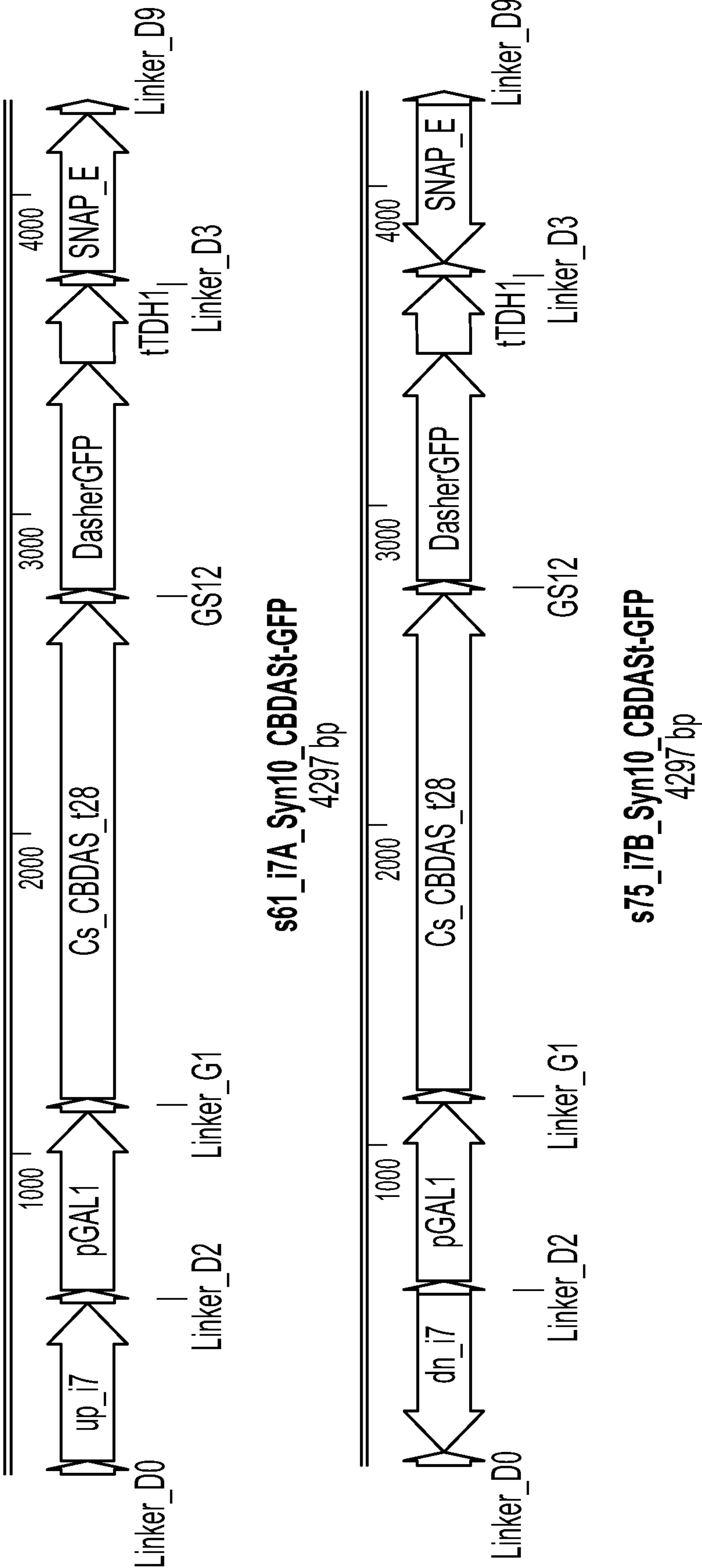


FIG. 36

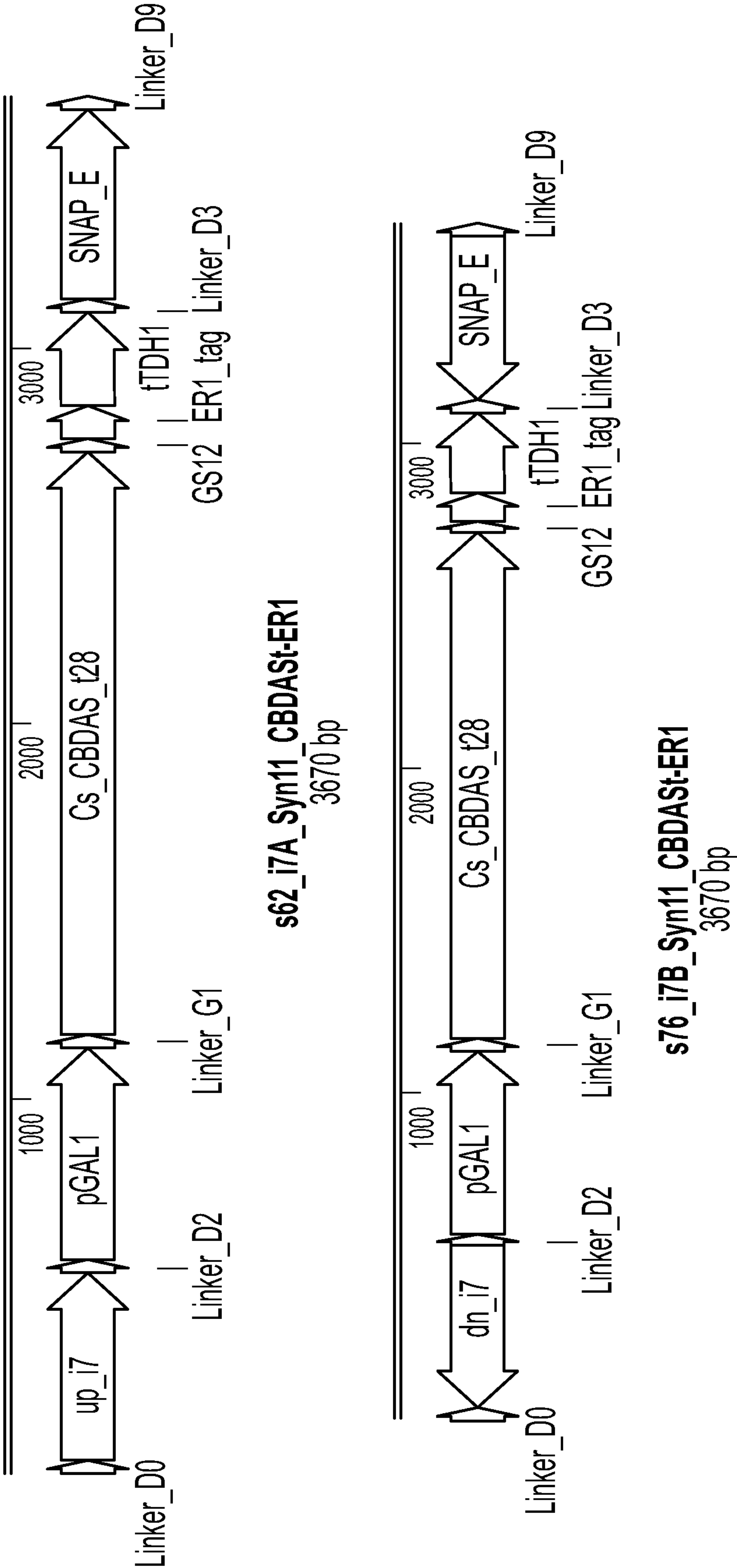


FIG. 37

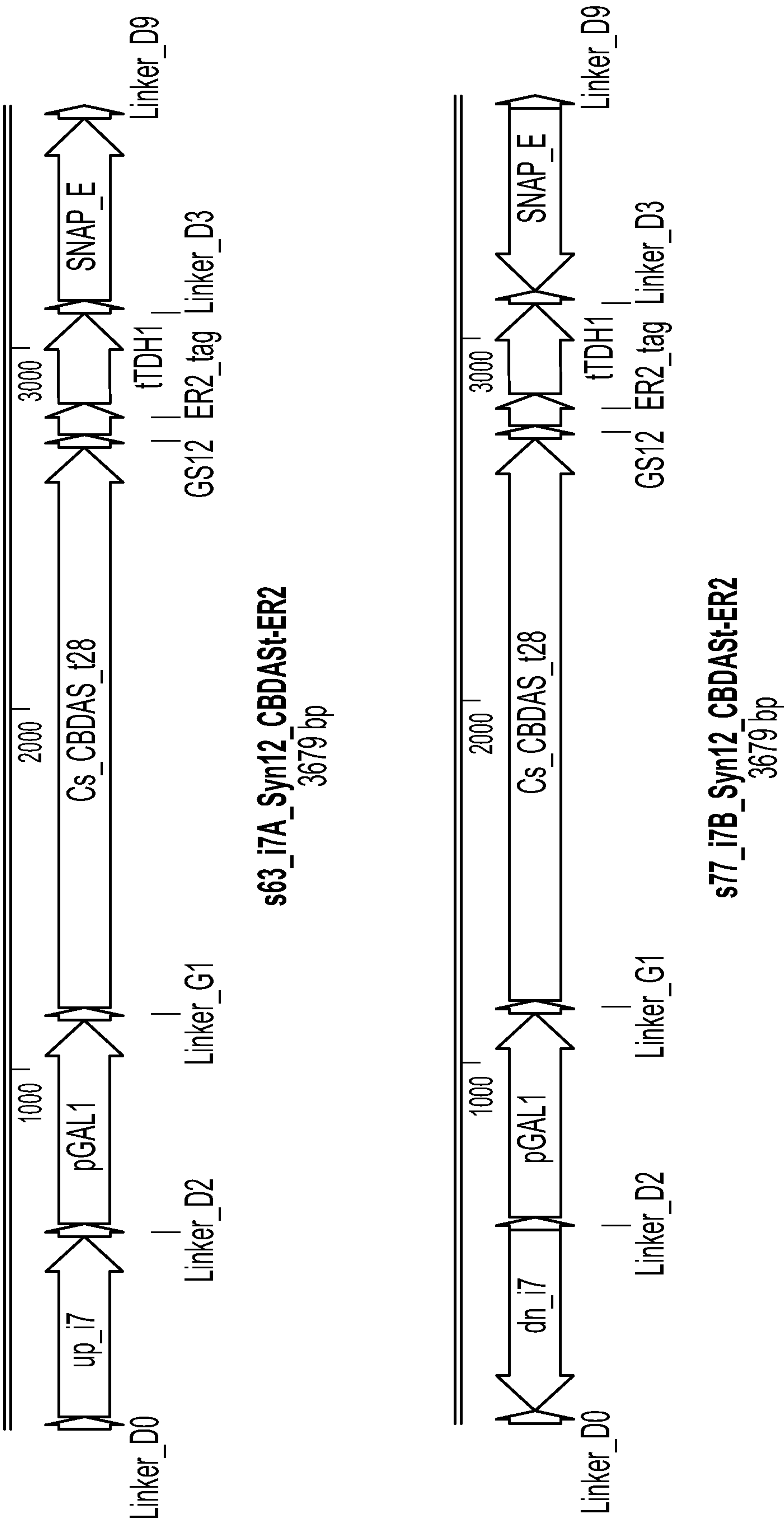


FIG. 38

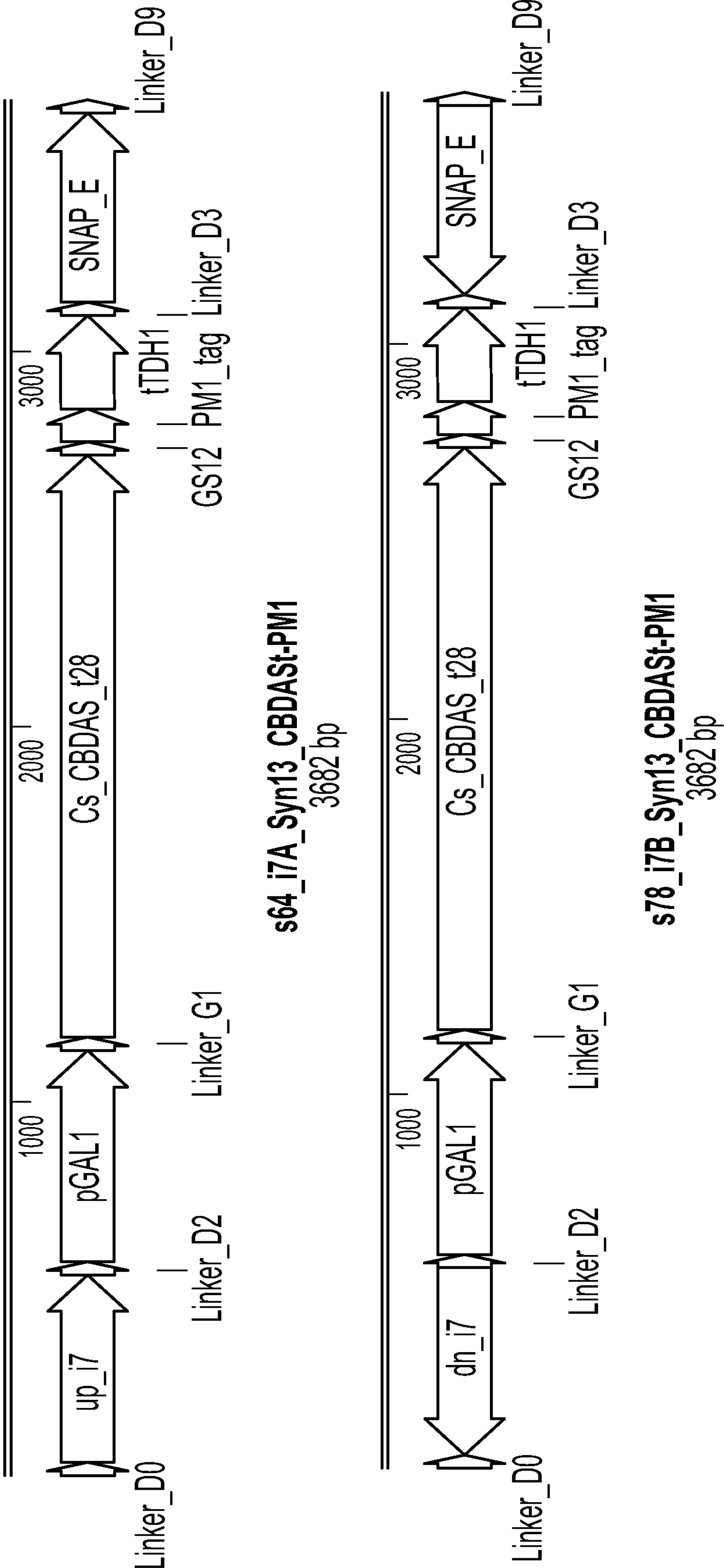


FIG. 39

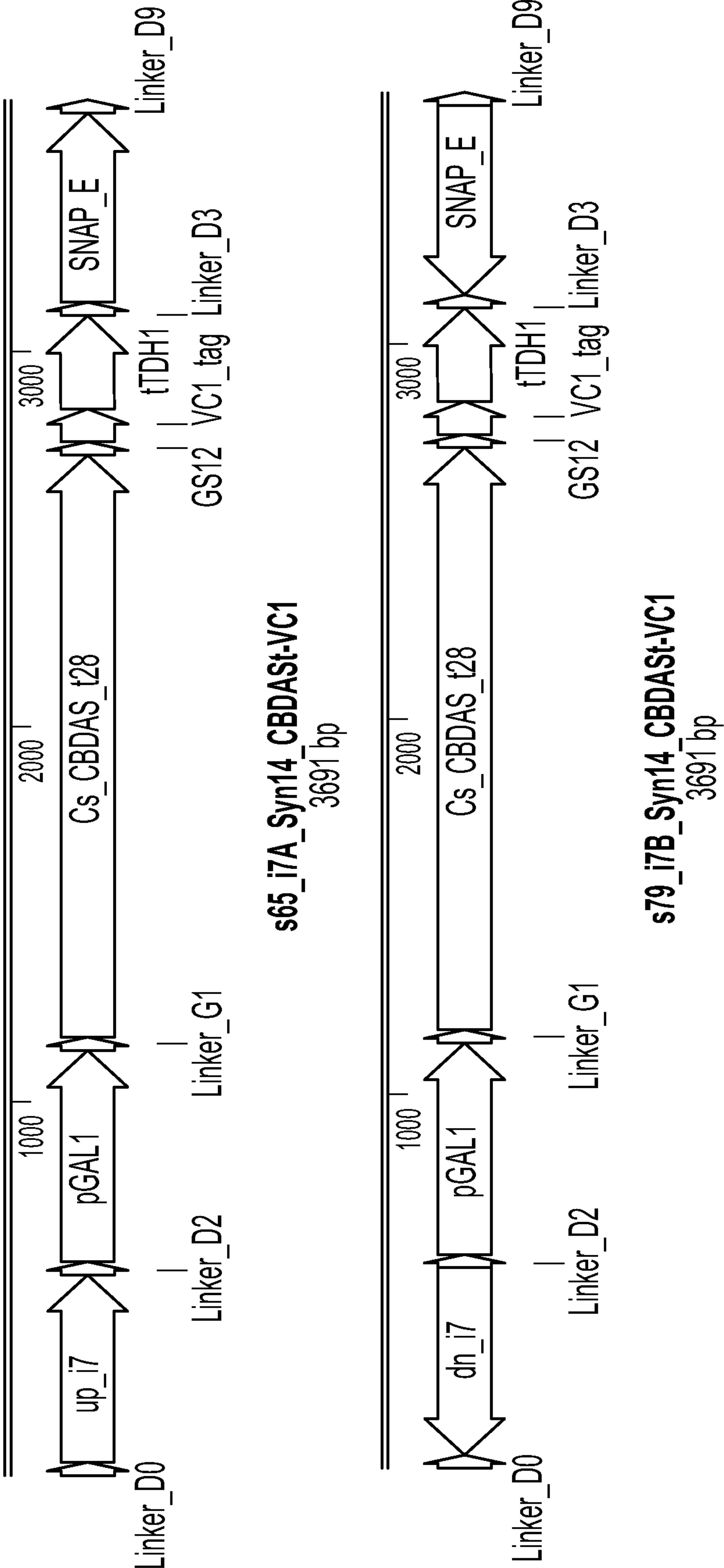


FIG. 40

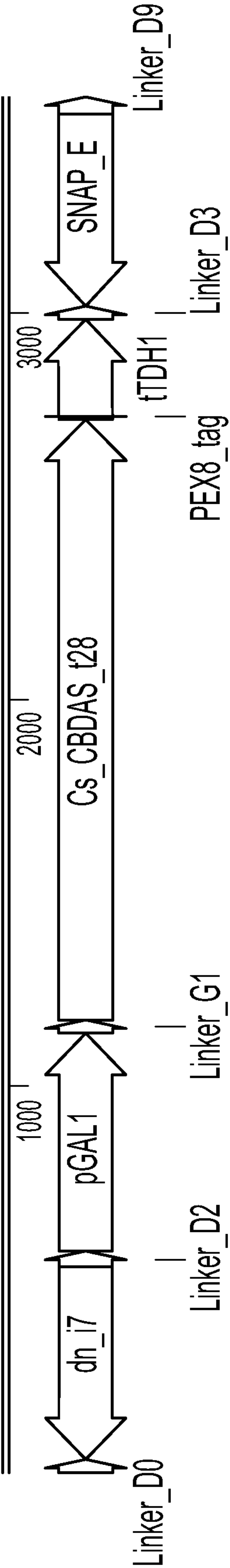
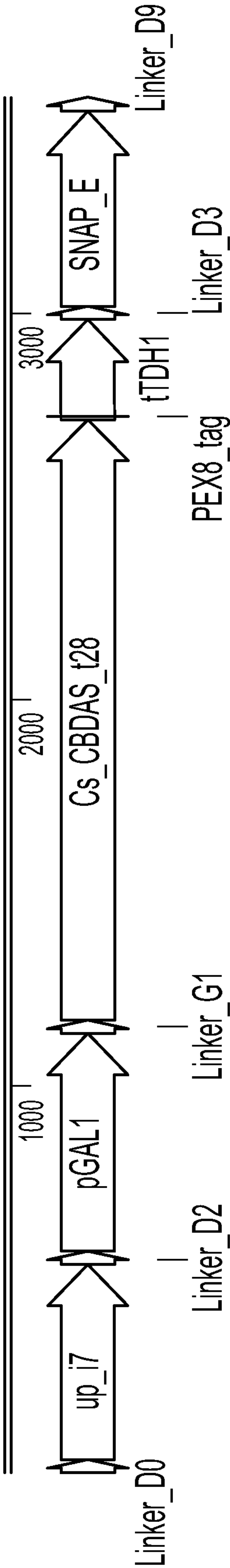
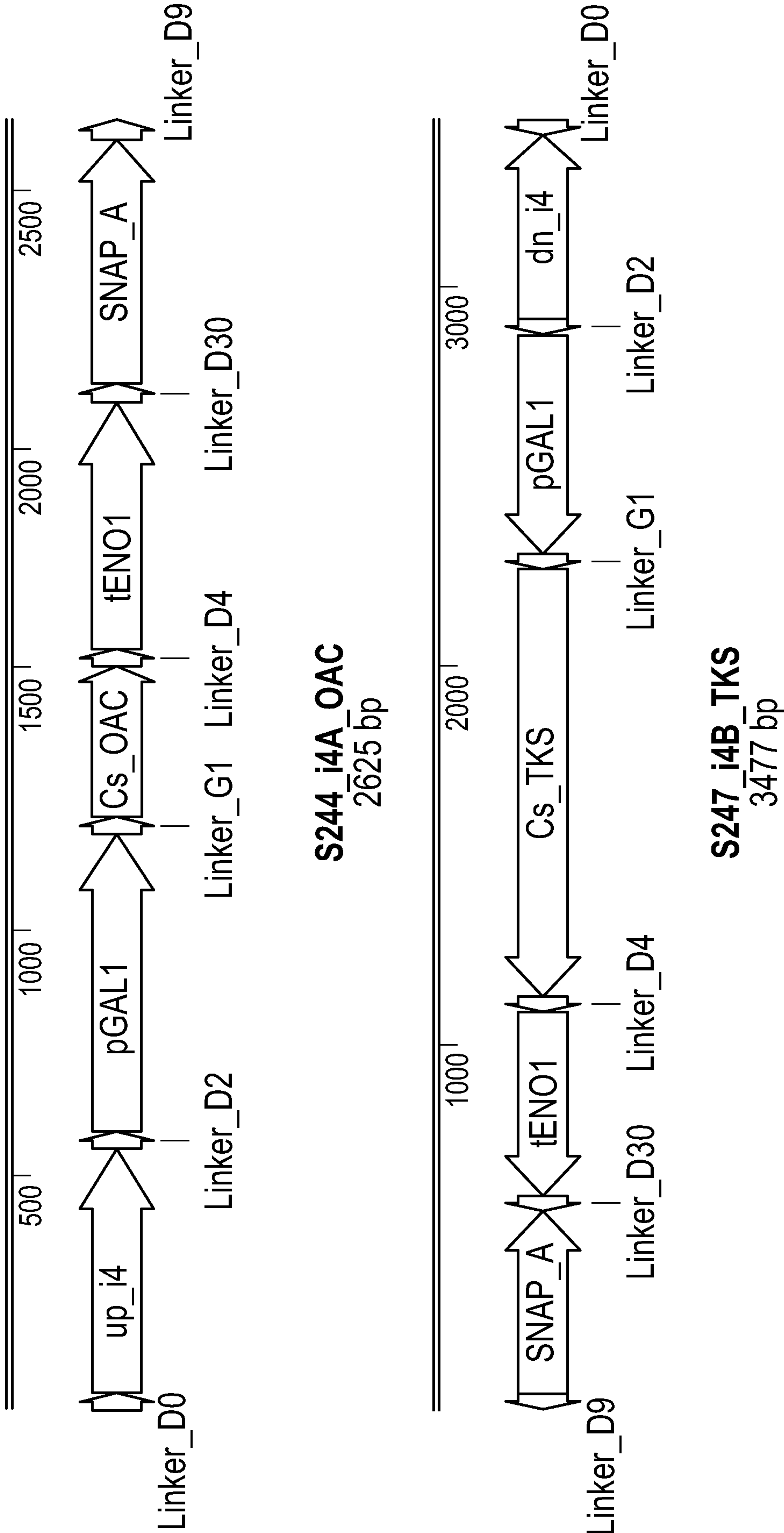


FIG. 41



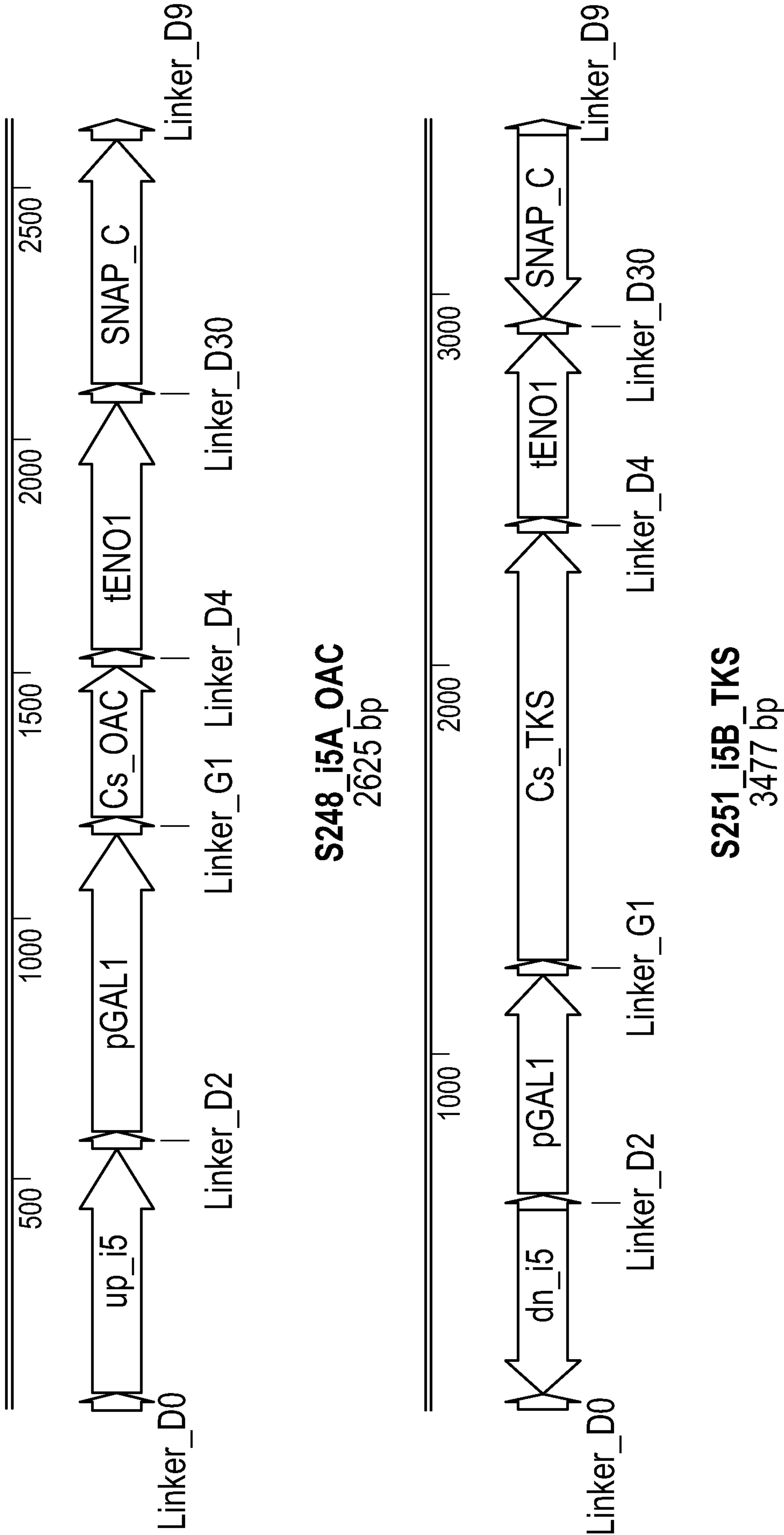
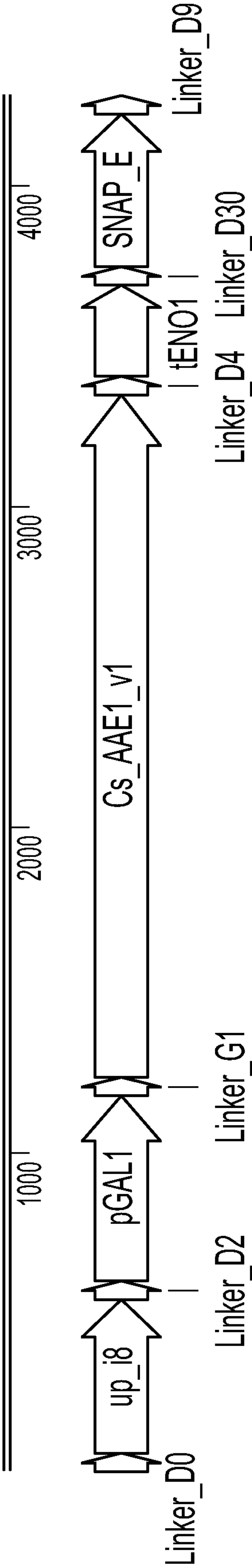
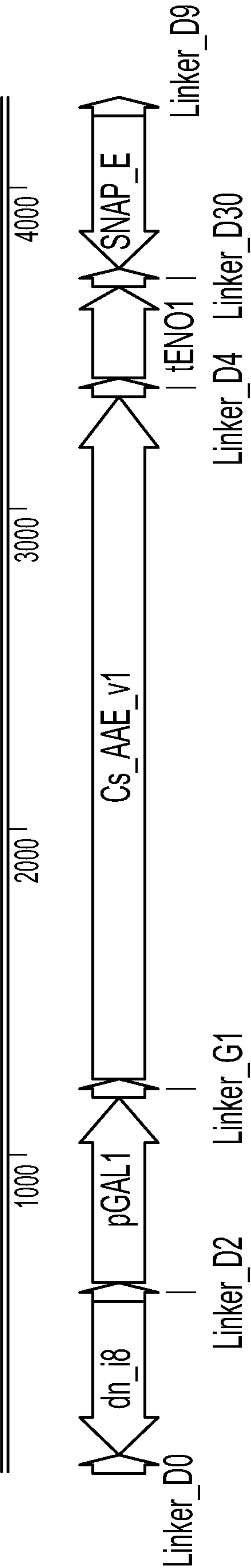


FIG. 42B

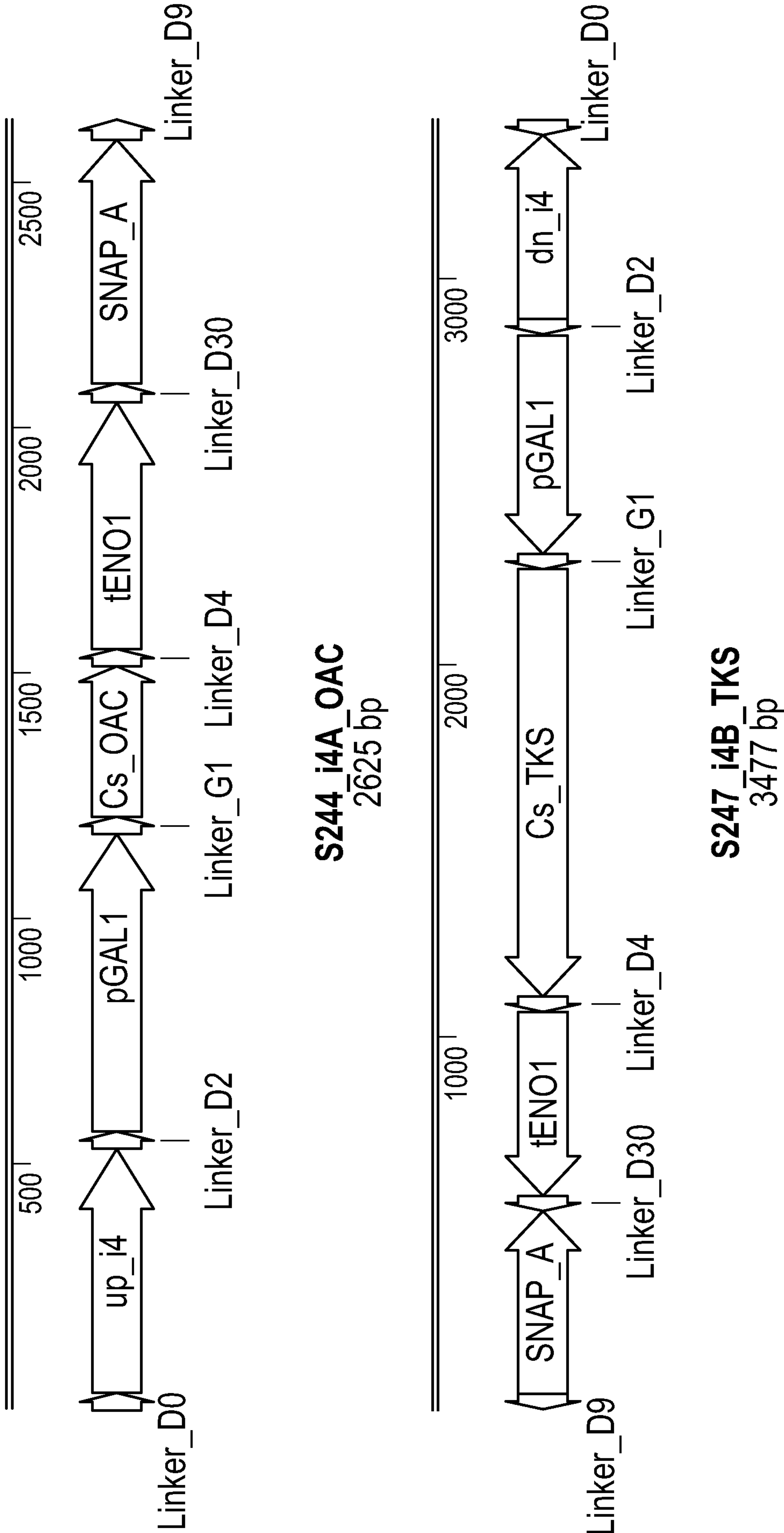


S262_ACL1
4280 bp



S274_ACL13
4280 bp

FIG. 42C



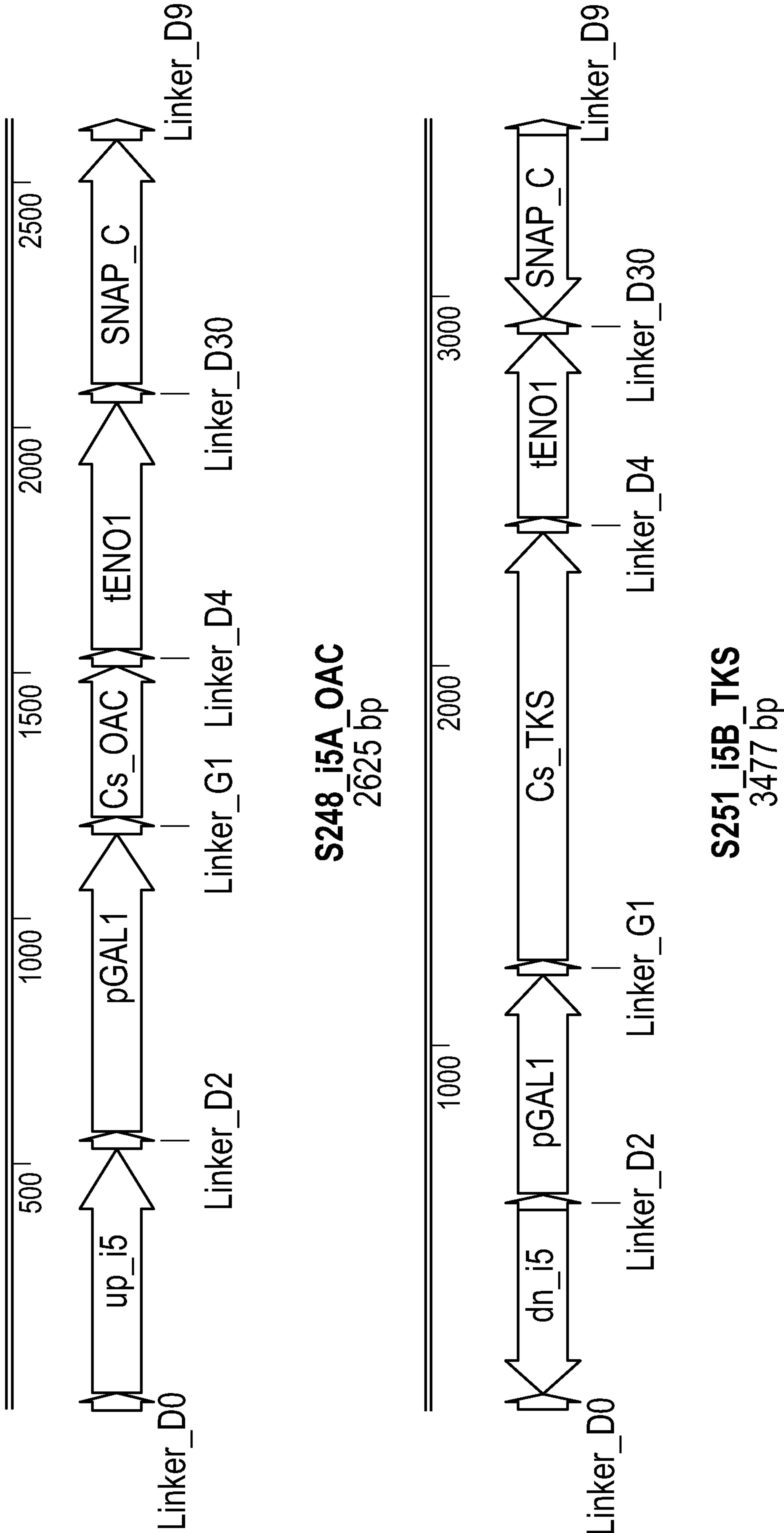
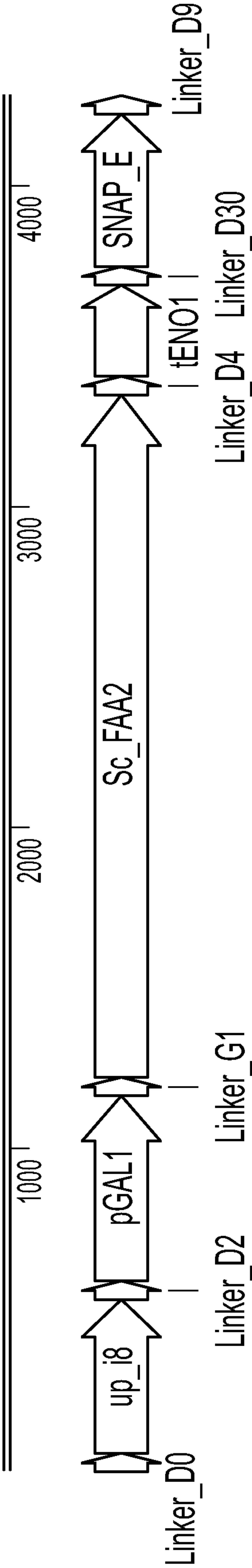
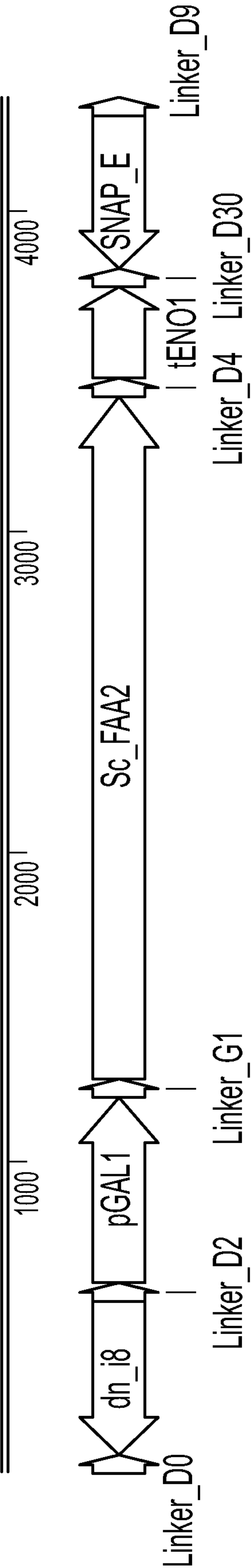


FIG. 43B



S269_ACL8
4352 bp



S281_ACL20
4352 bp

FIG. 43C

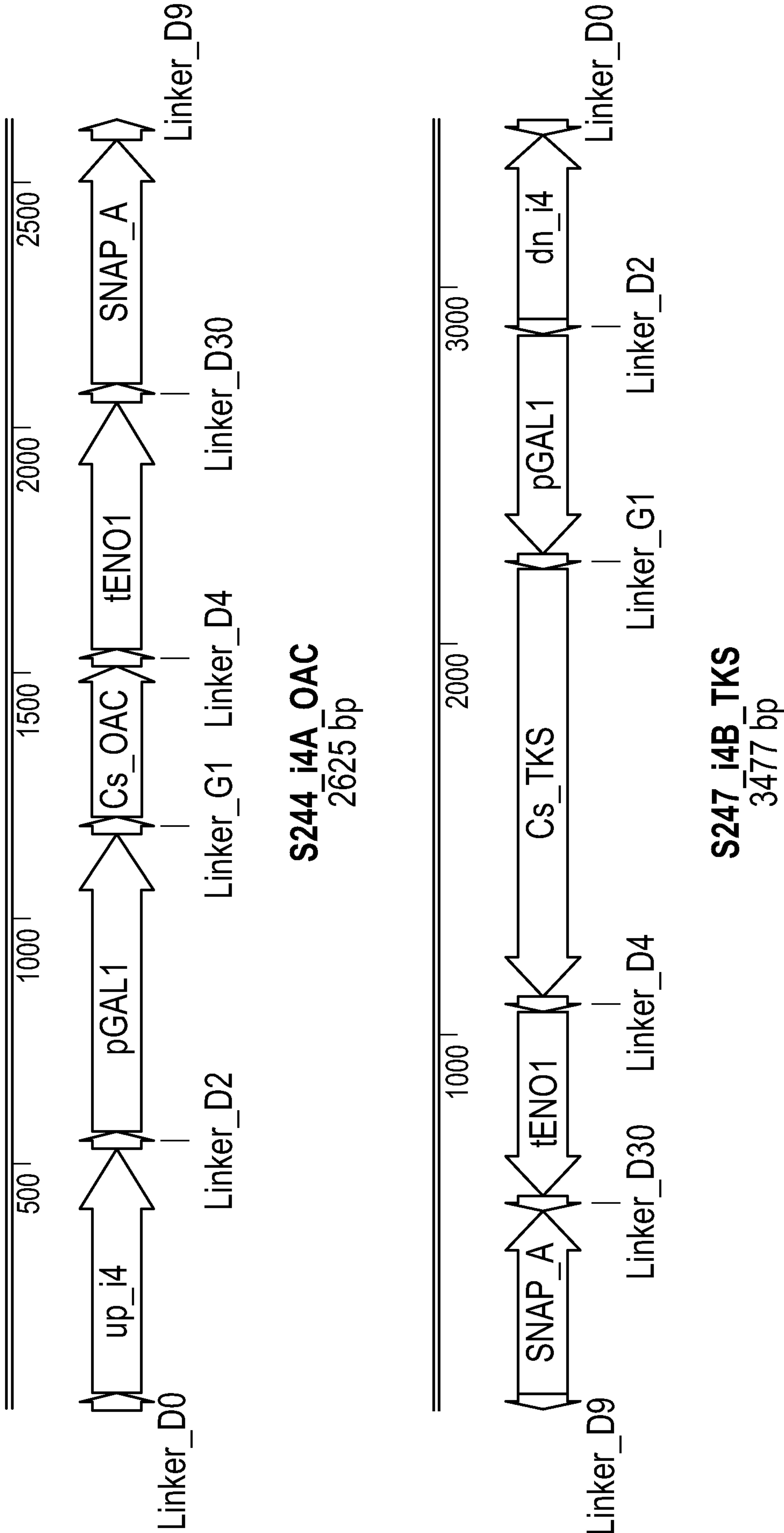


FIG. 44A

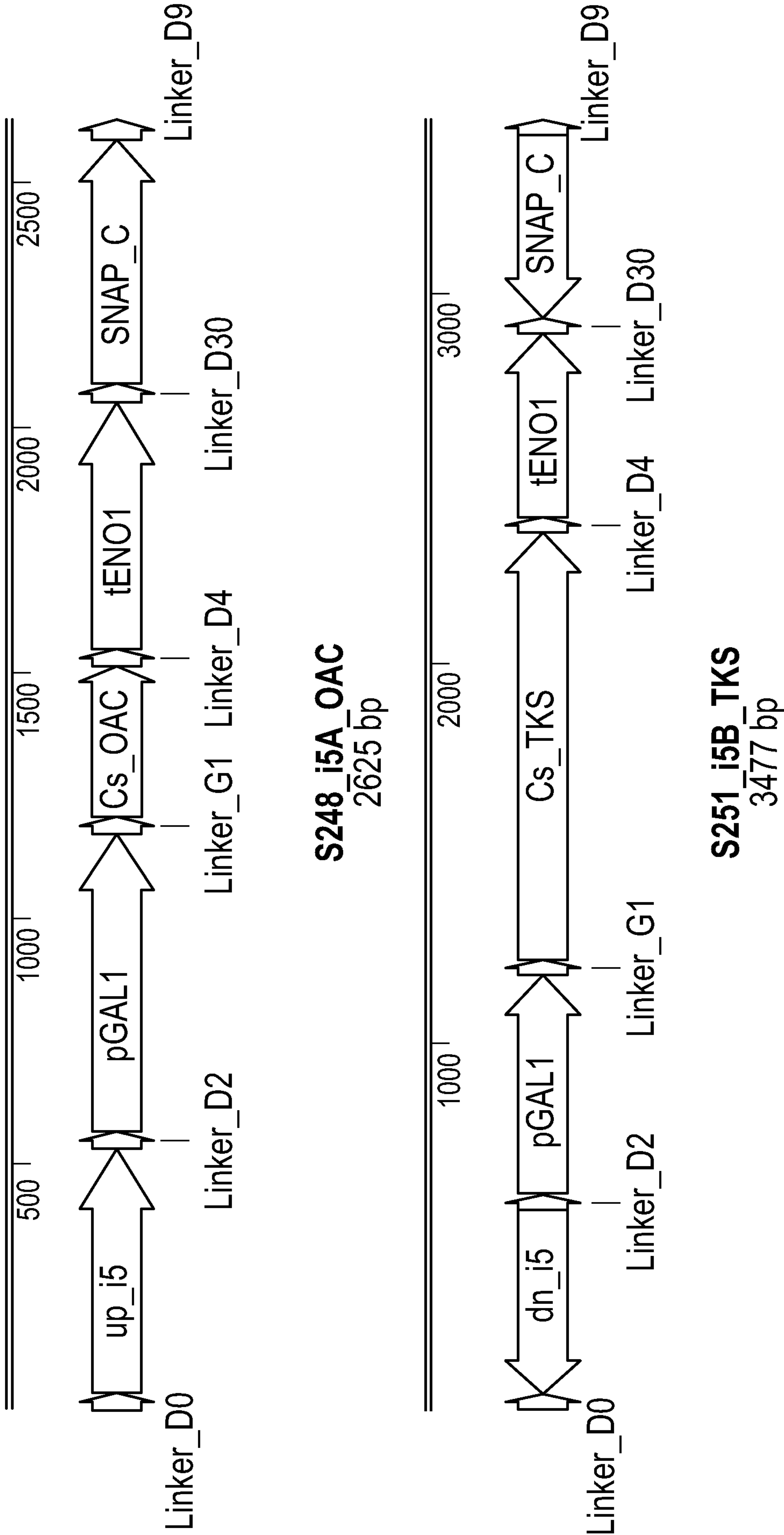
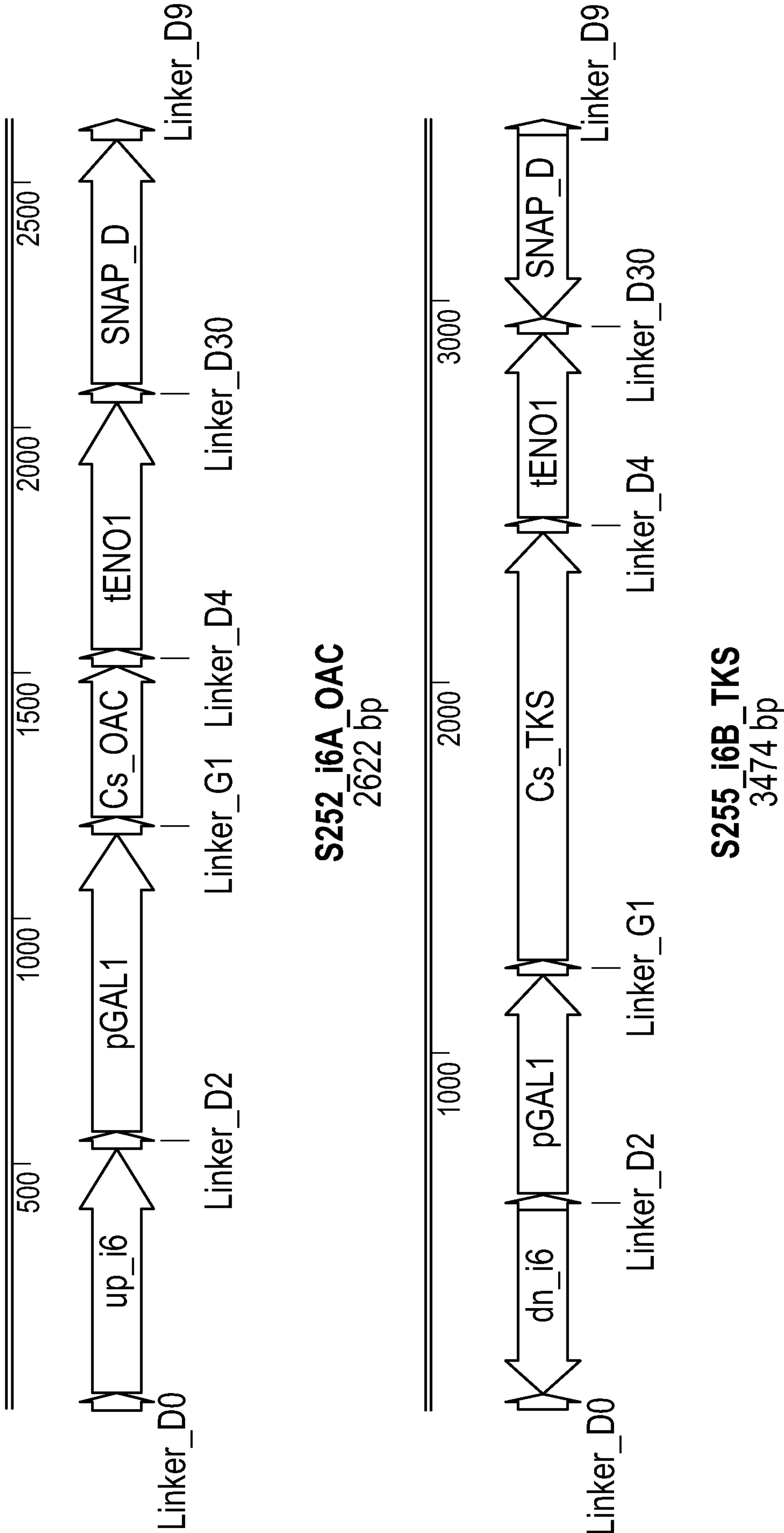
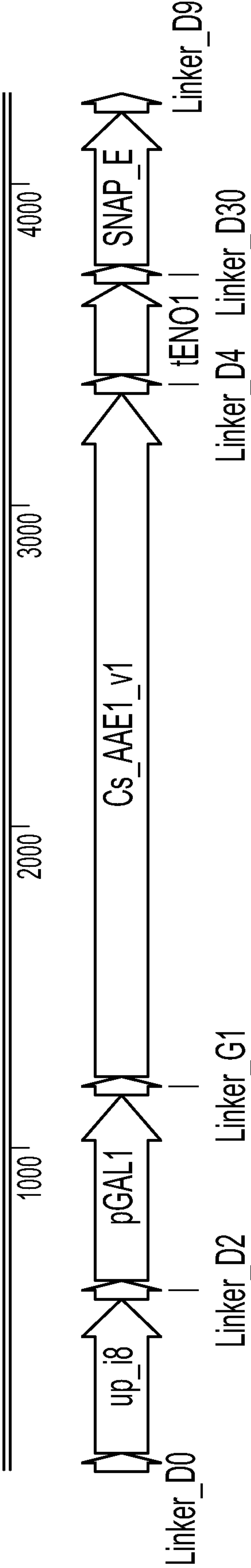
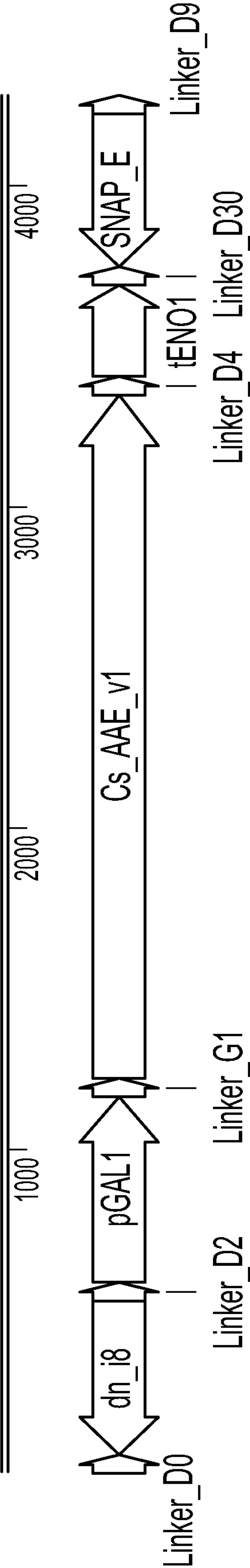


FIG. 44B



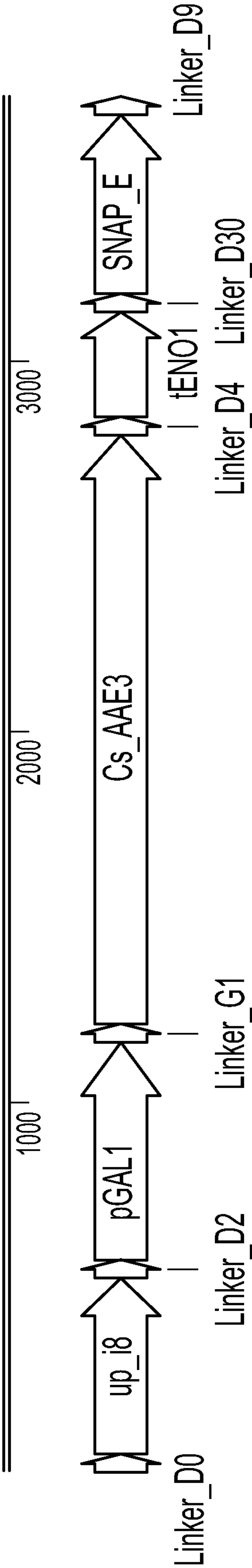


S262_ACL1
4280 bp

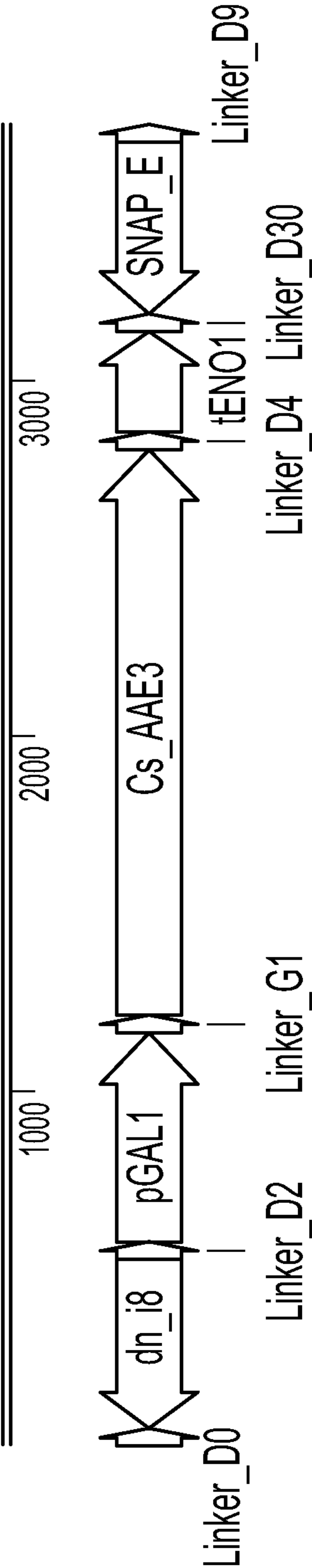


S274_ACL13
4280 bp

FIG. 45



S266_ACL5
3719 bp



S278_ACL17
3719 bp

FIG. 46

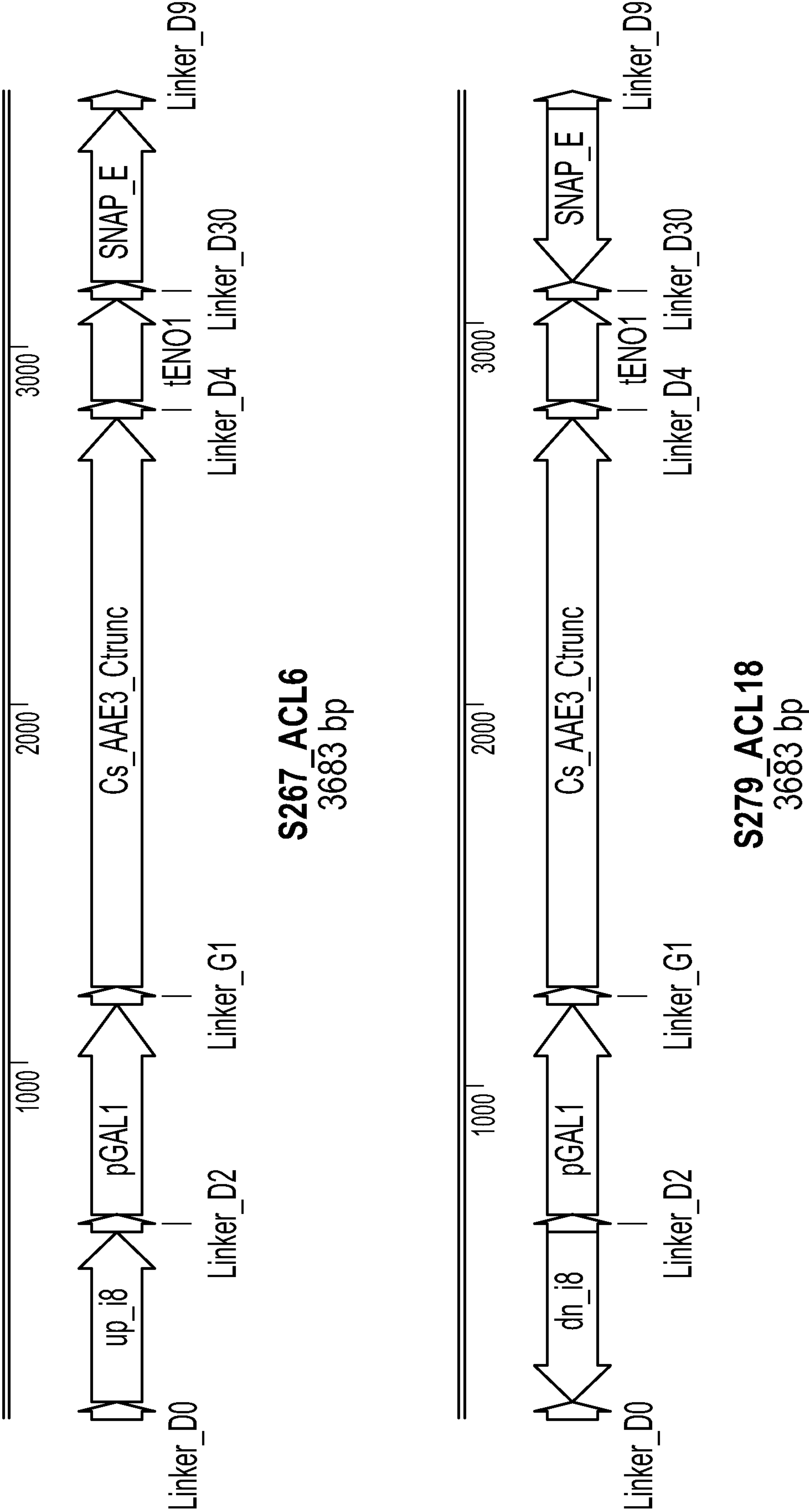
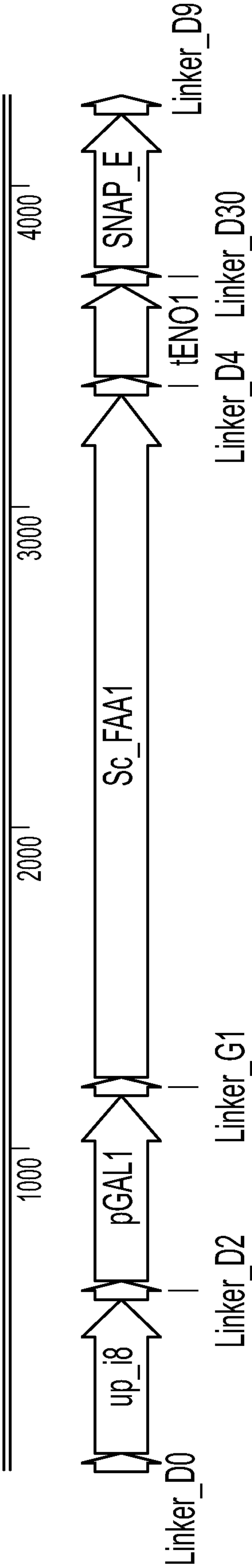
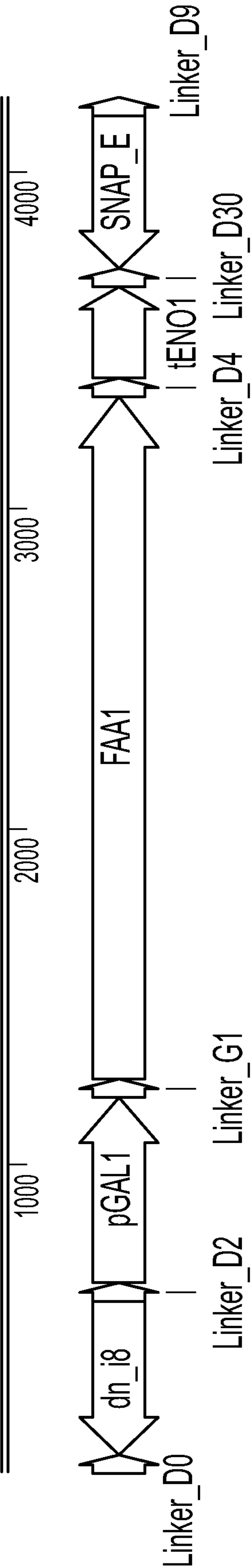


FIG. 47

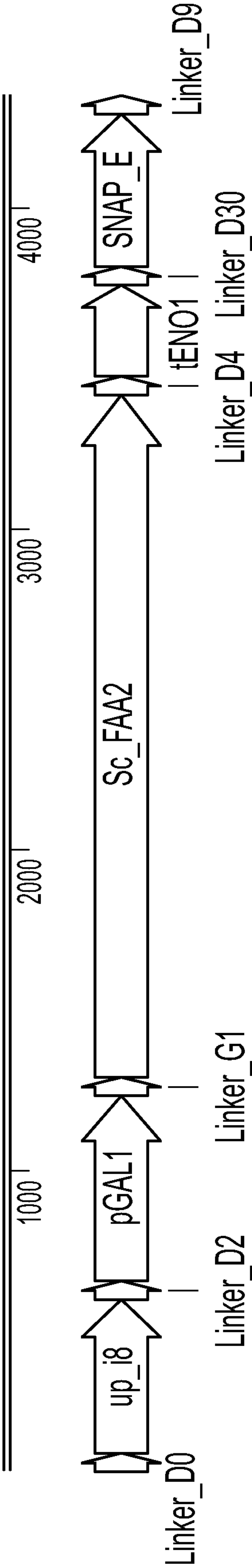


S268_ACL7
4220 bp

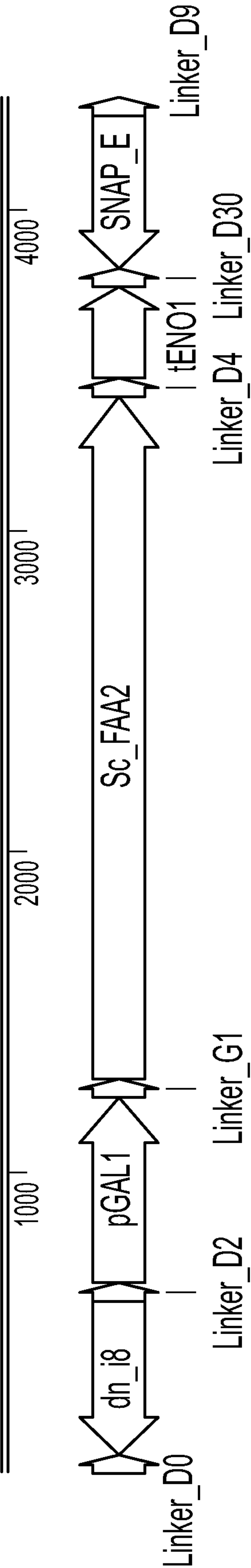


S280_ACL19
4220 bp

FIG. 48

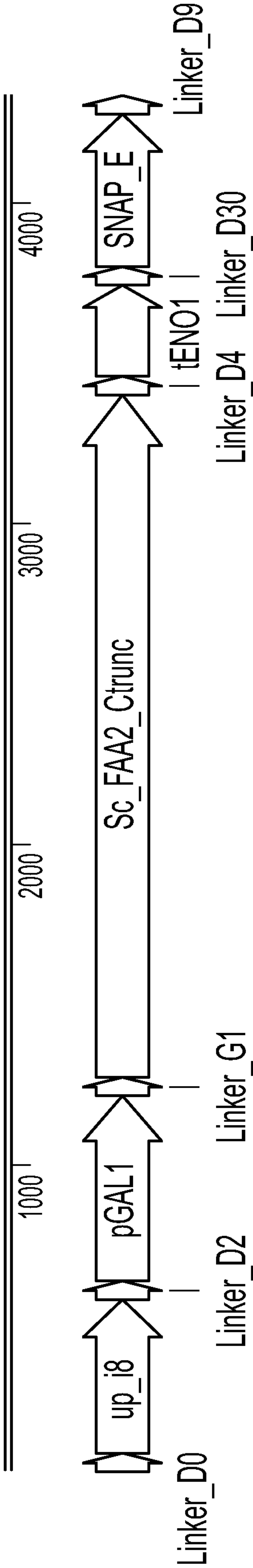


S269_ACL8
4352 bp

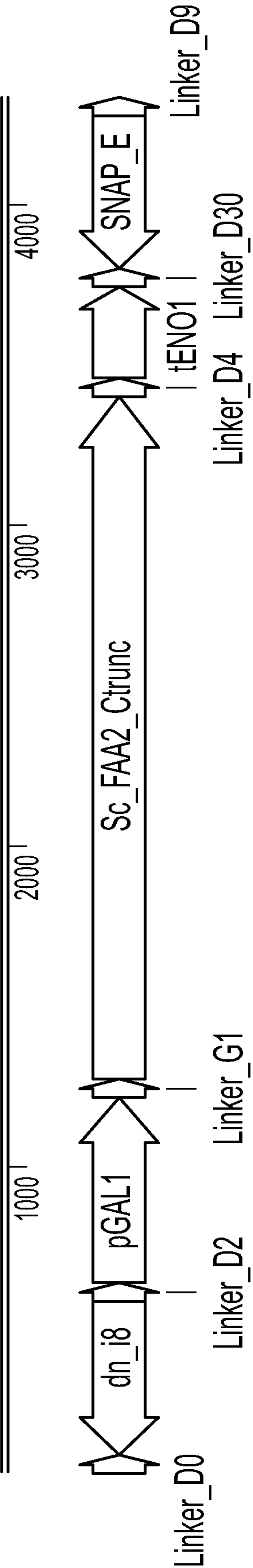


S281_ACL20
4352 bp

FIG. 49

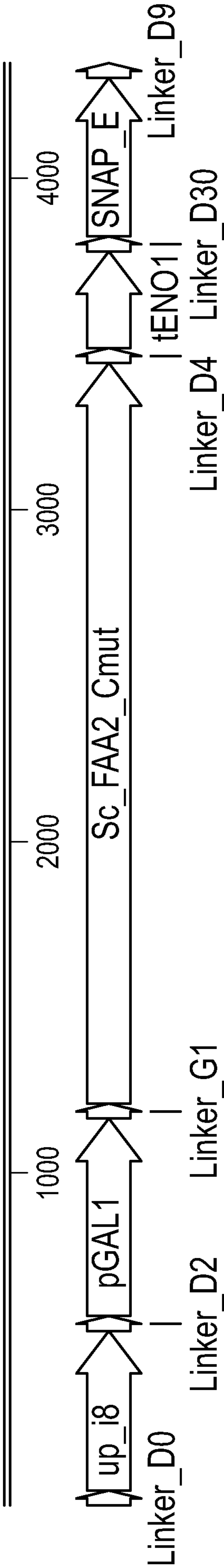


S270 ACL9
4343 bp



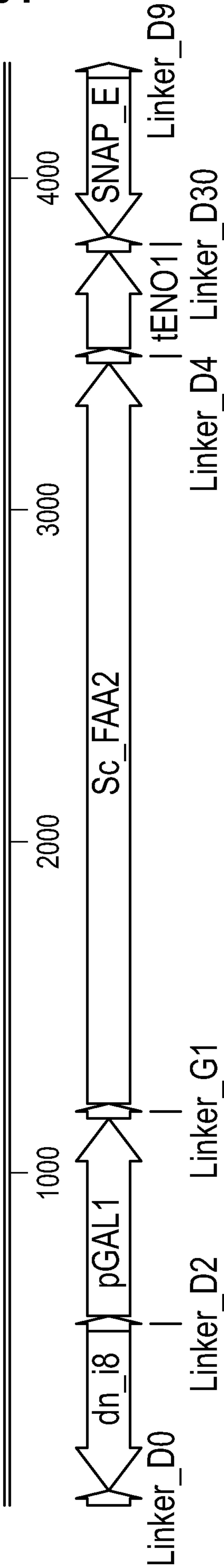
S282 ACL21
4343 bp

FIG. 50



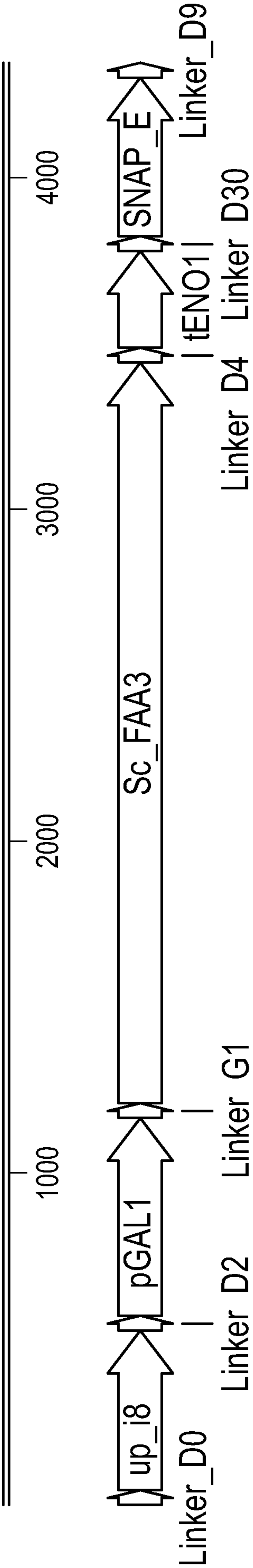
S271_ACL10
4355 bp

62/101



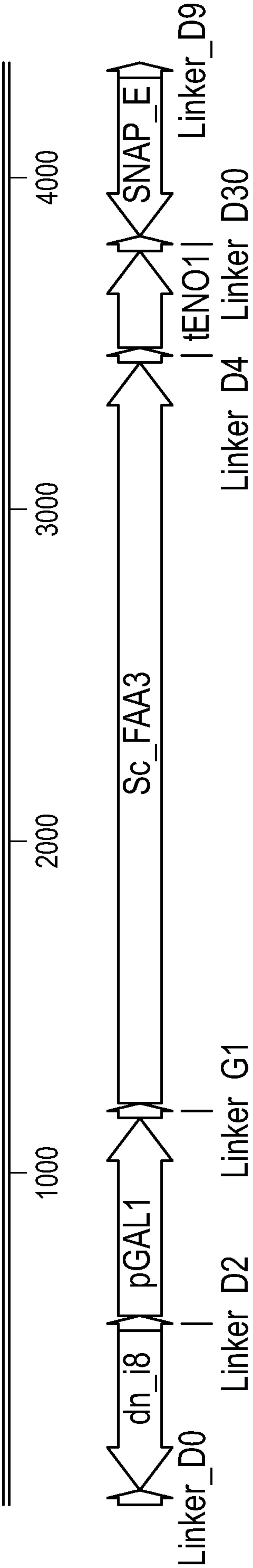
S281_ACL20
4352 bp

FIG. 51



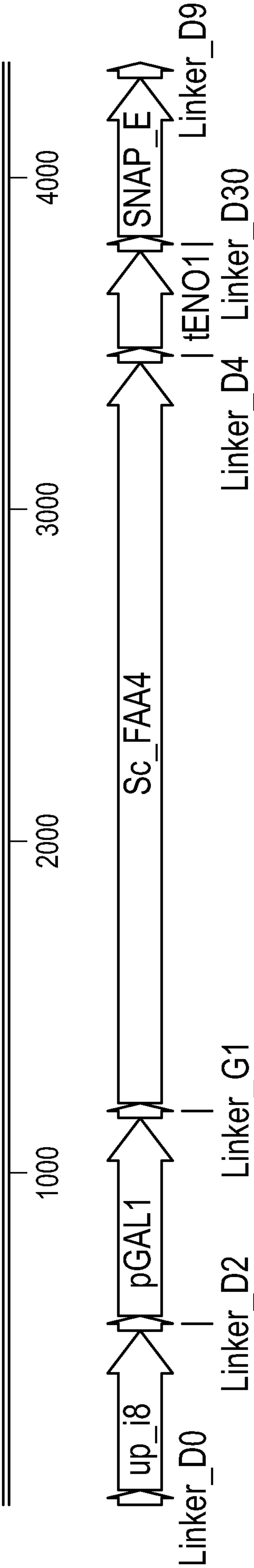
S272_ACL11
4202 bp

63/101



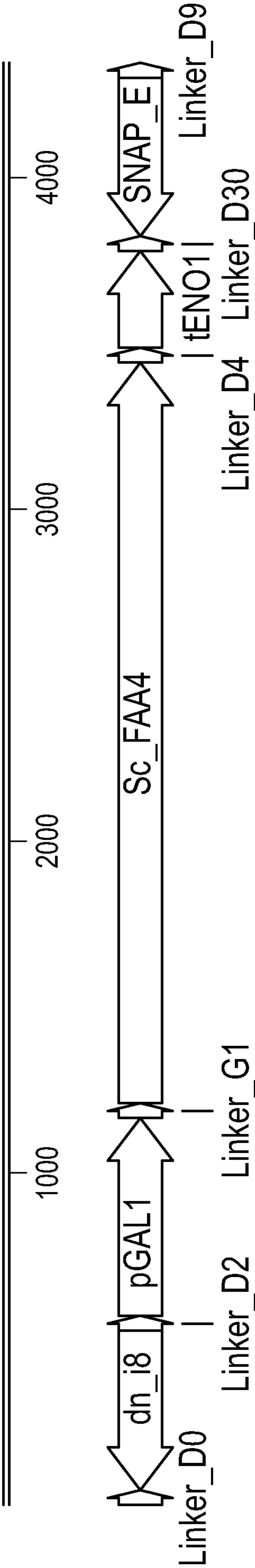
S284_ACL23
4202 bp

FIG. 52



S273_ACL12
4202 bp

64/101



S285_ACL24
4202 bp

FIG. 53

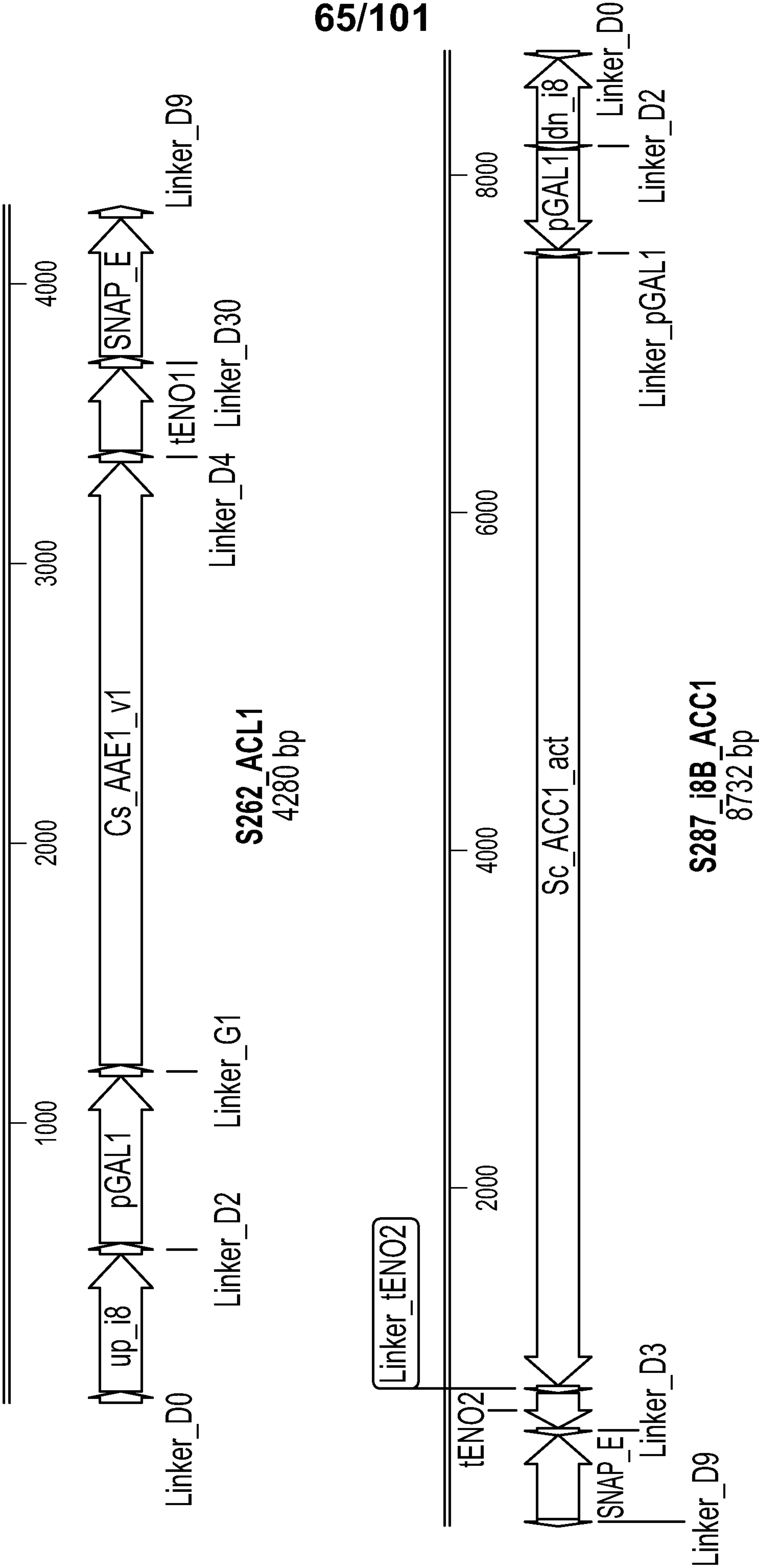


FIG. 54

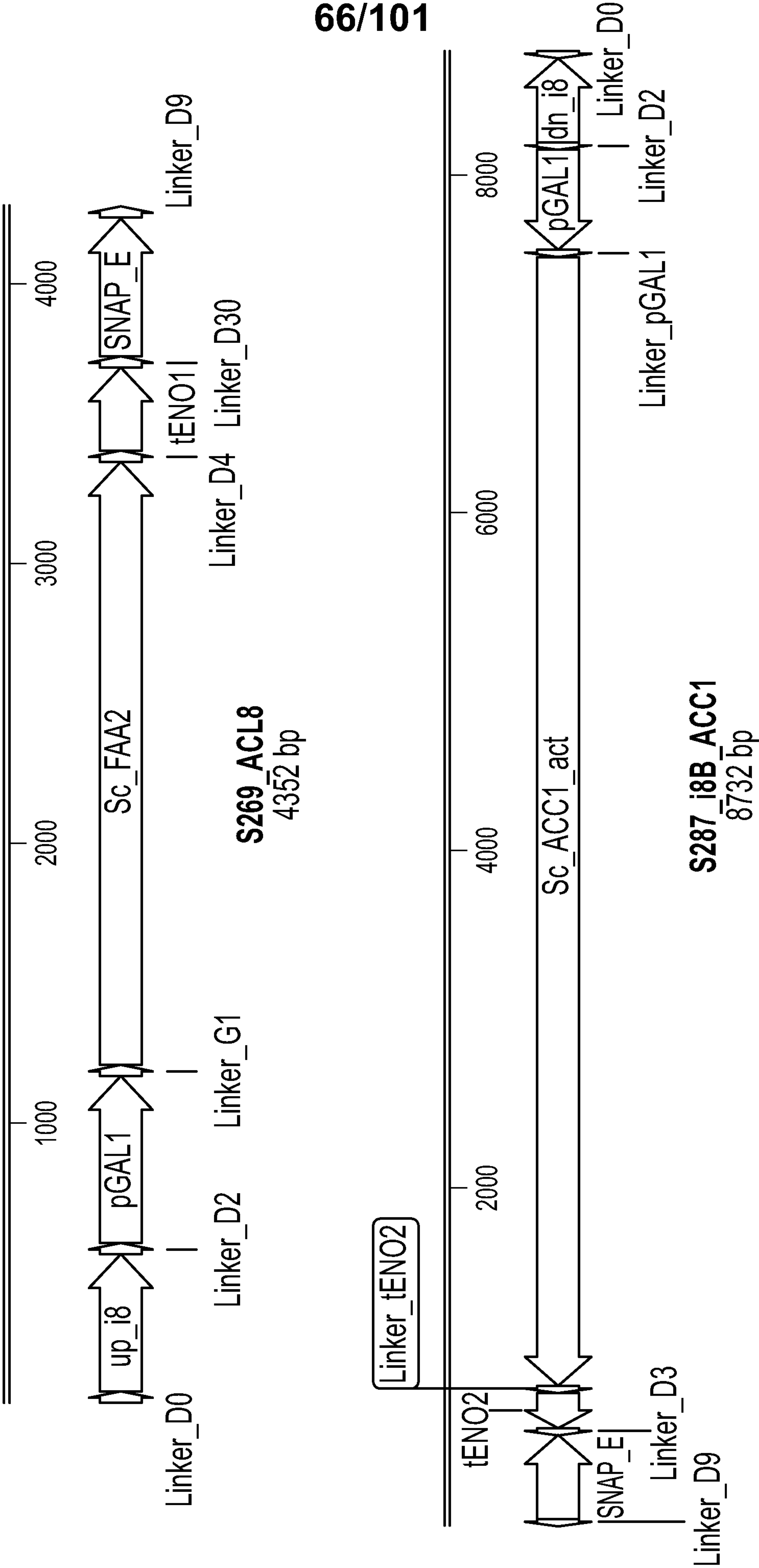


FIG. 55

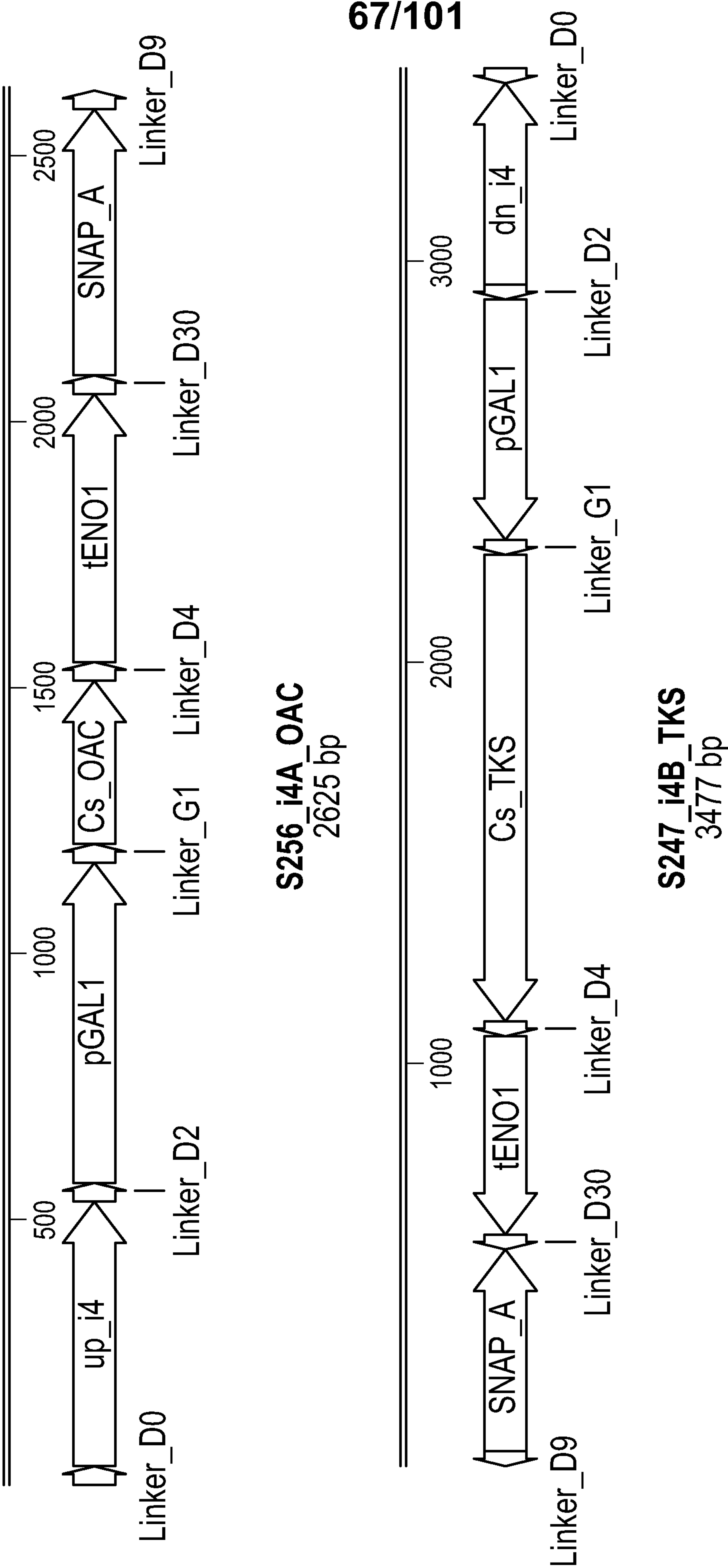


FIG. 56A

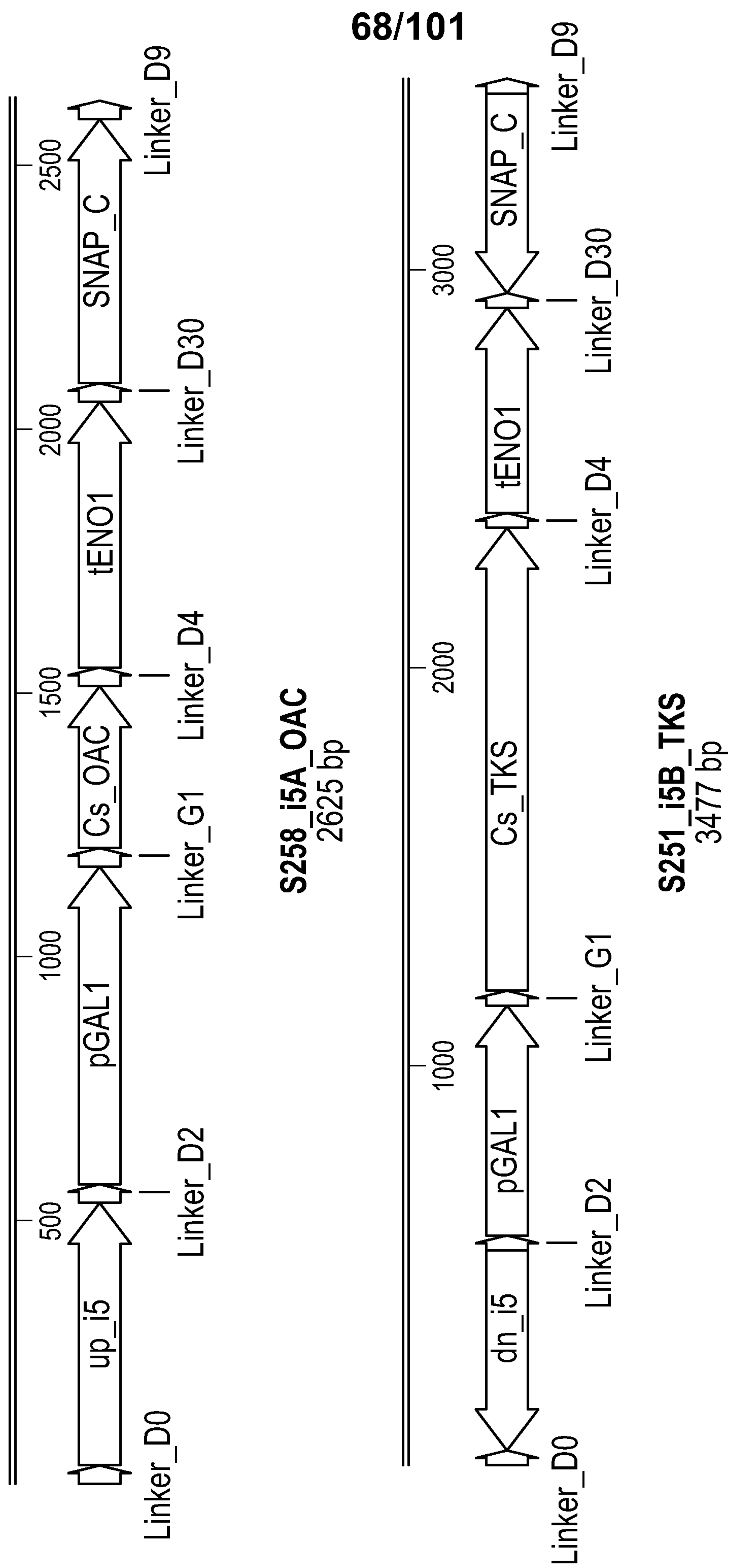


FIG. 56B

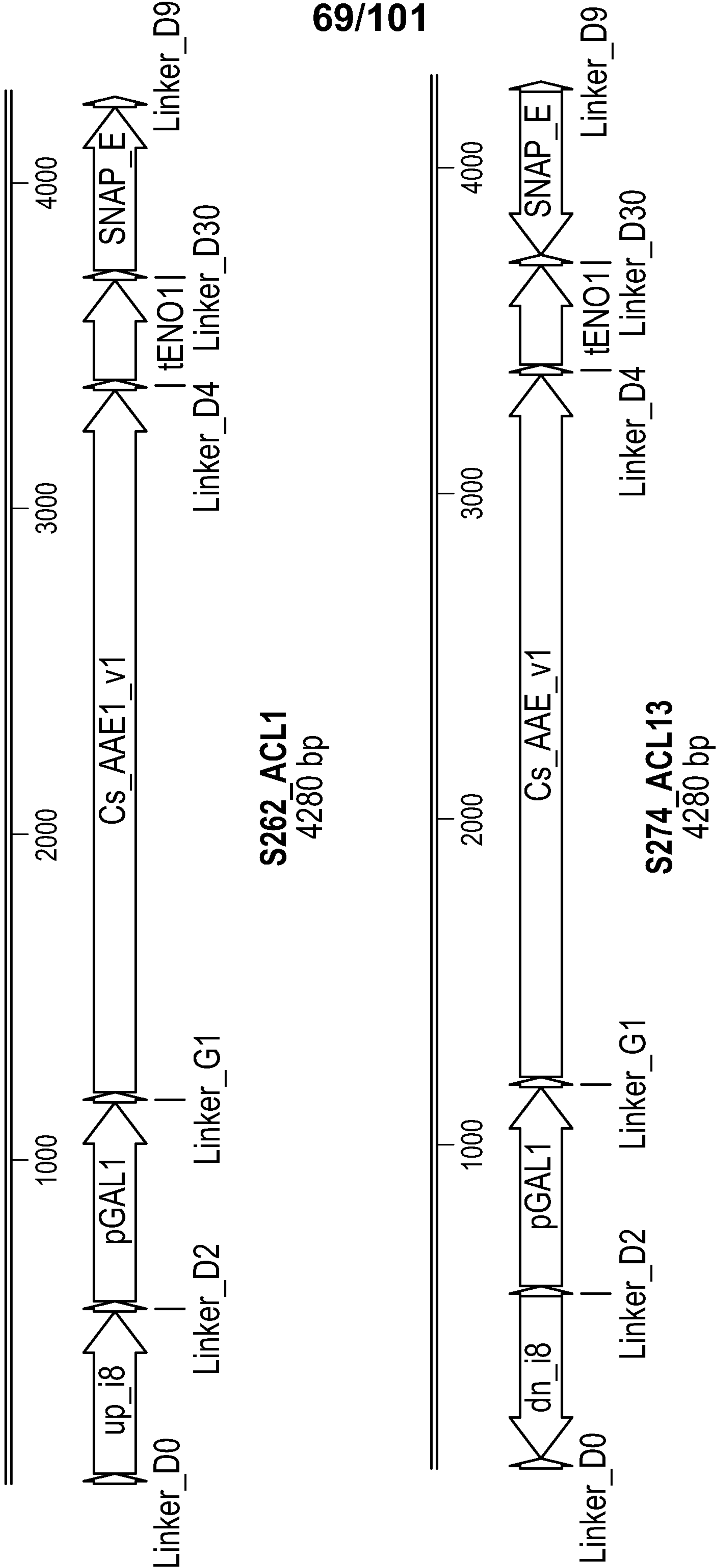


FIG. 56C

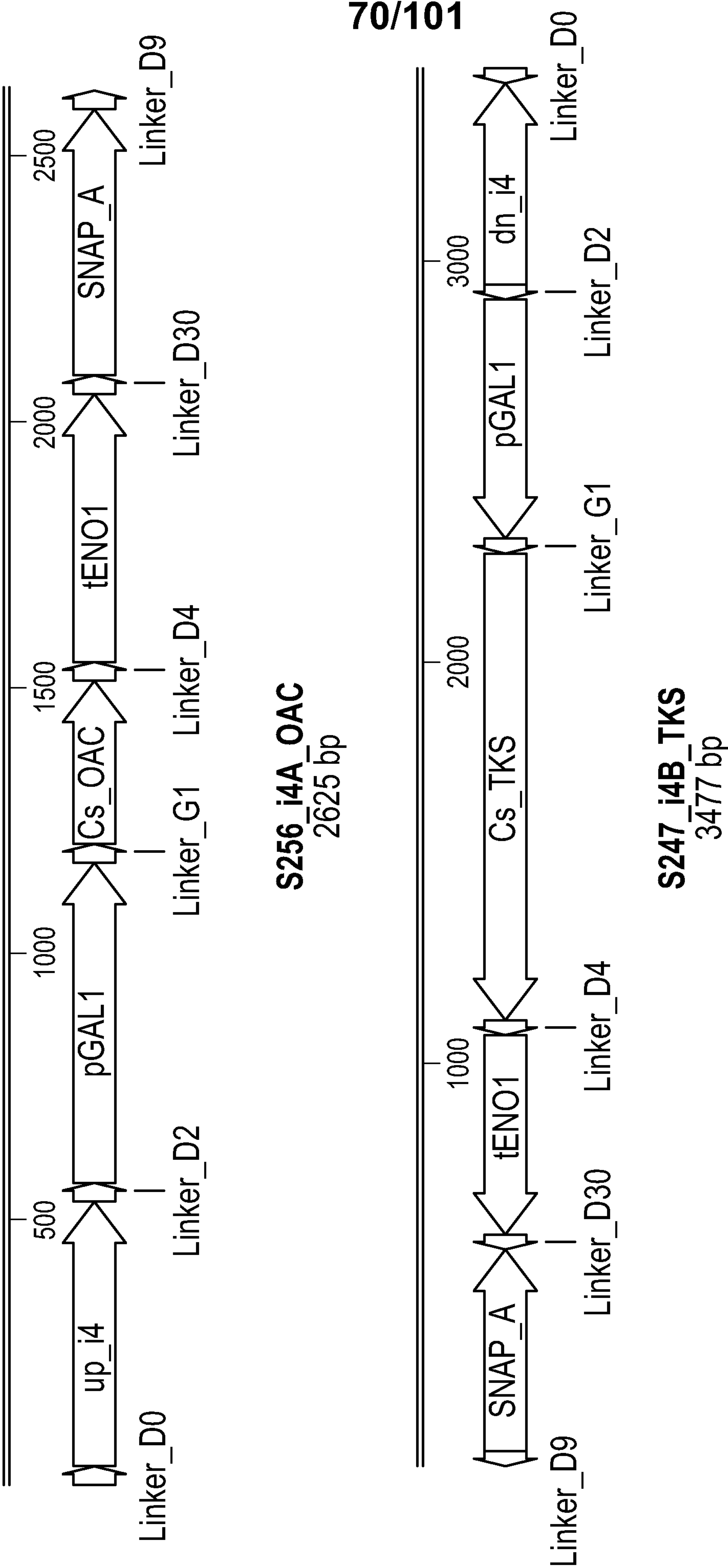


FIG. 57A

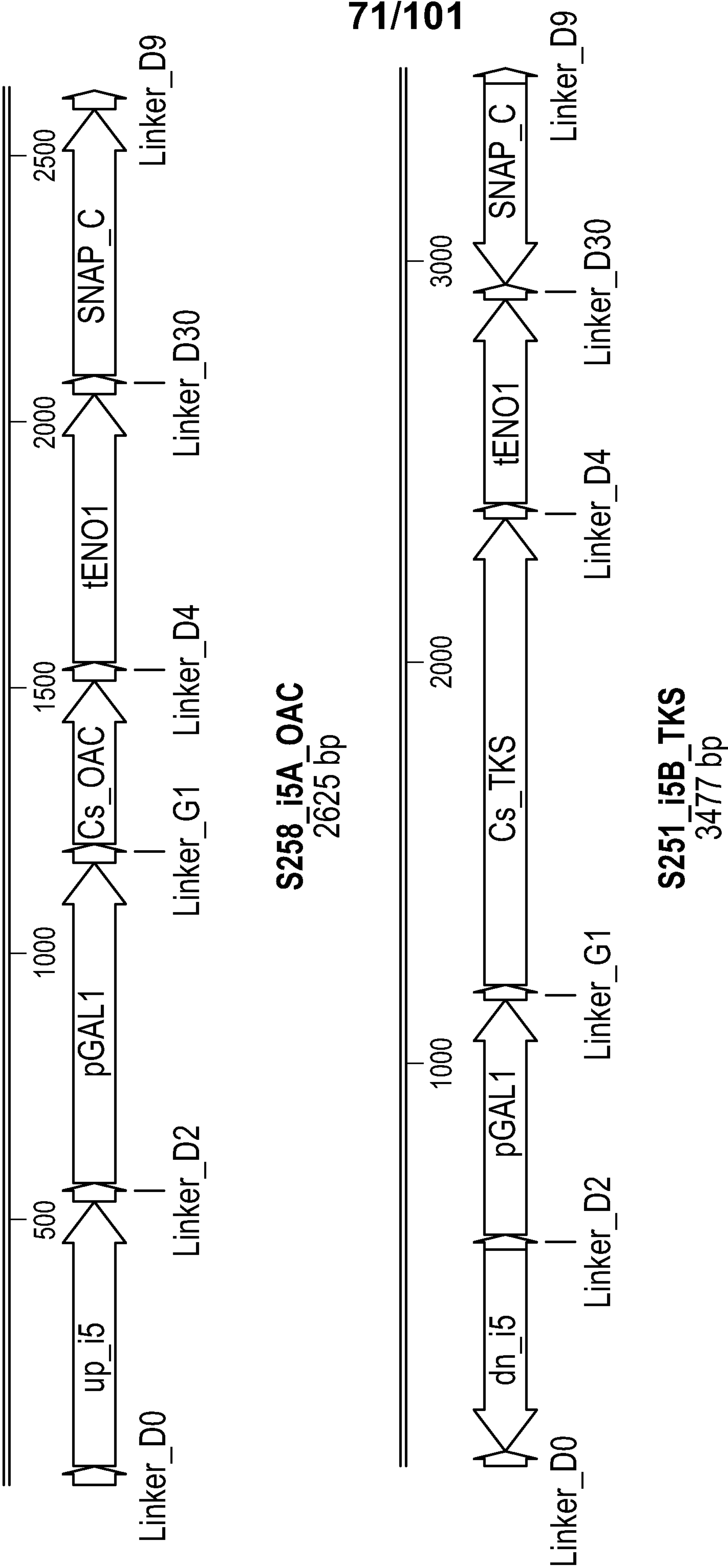
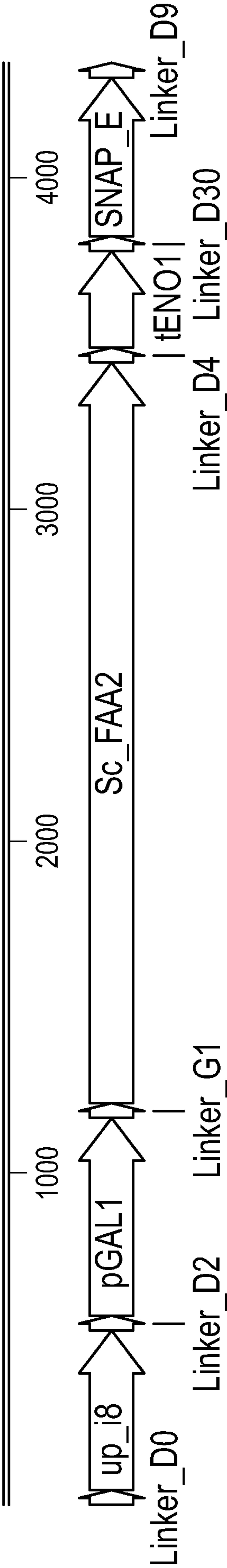
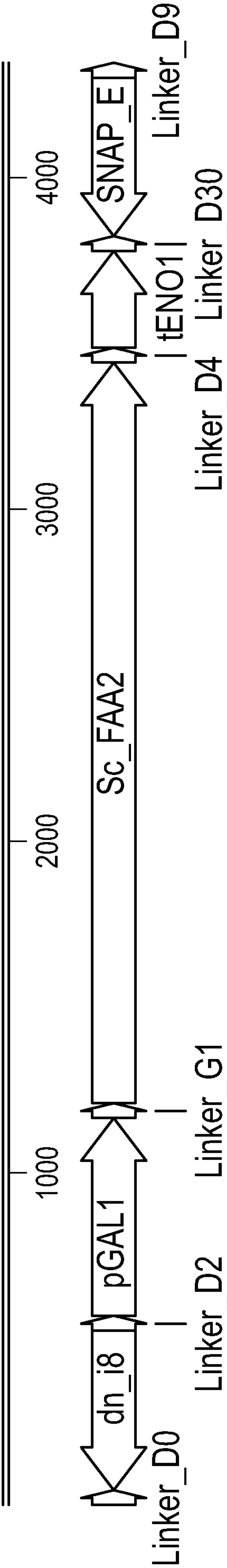


FIG. 57B

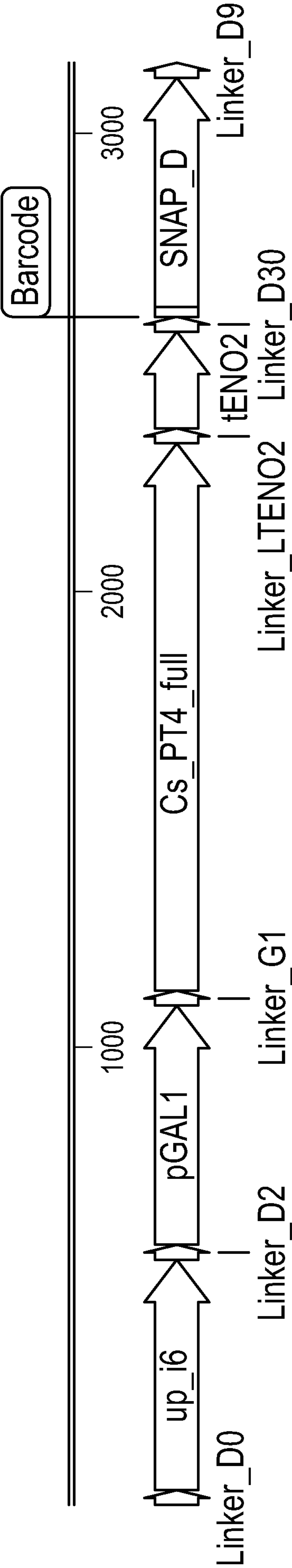


S269_ACL8
4352 bp

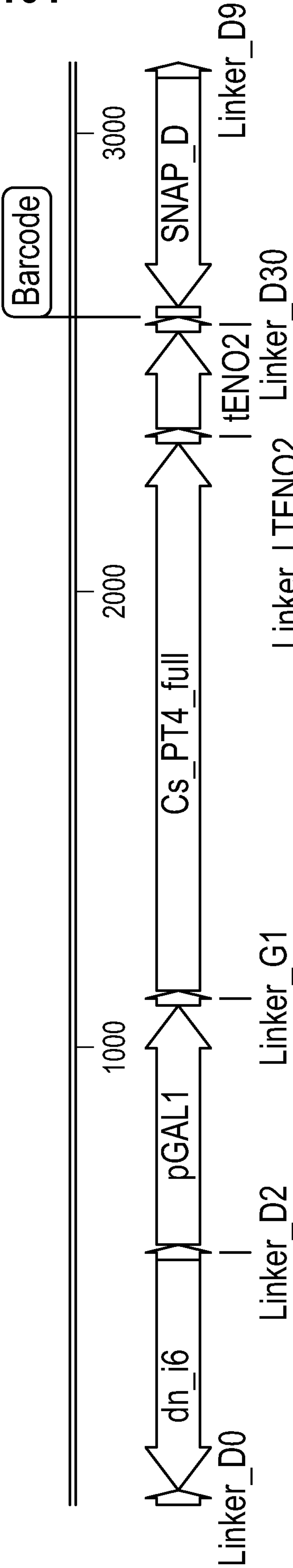


S281_ACL20
4352 bp

FIG. 57C



S169
3153 bp



S170
3153 bp

FIG. 58

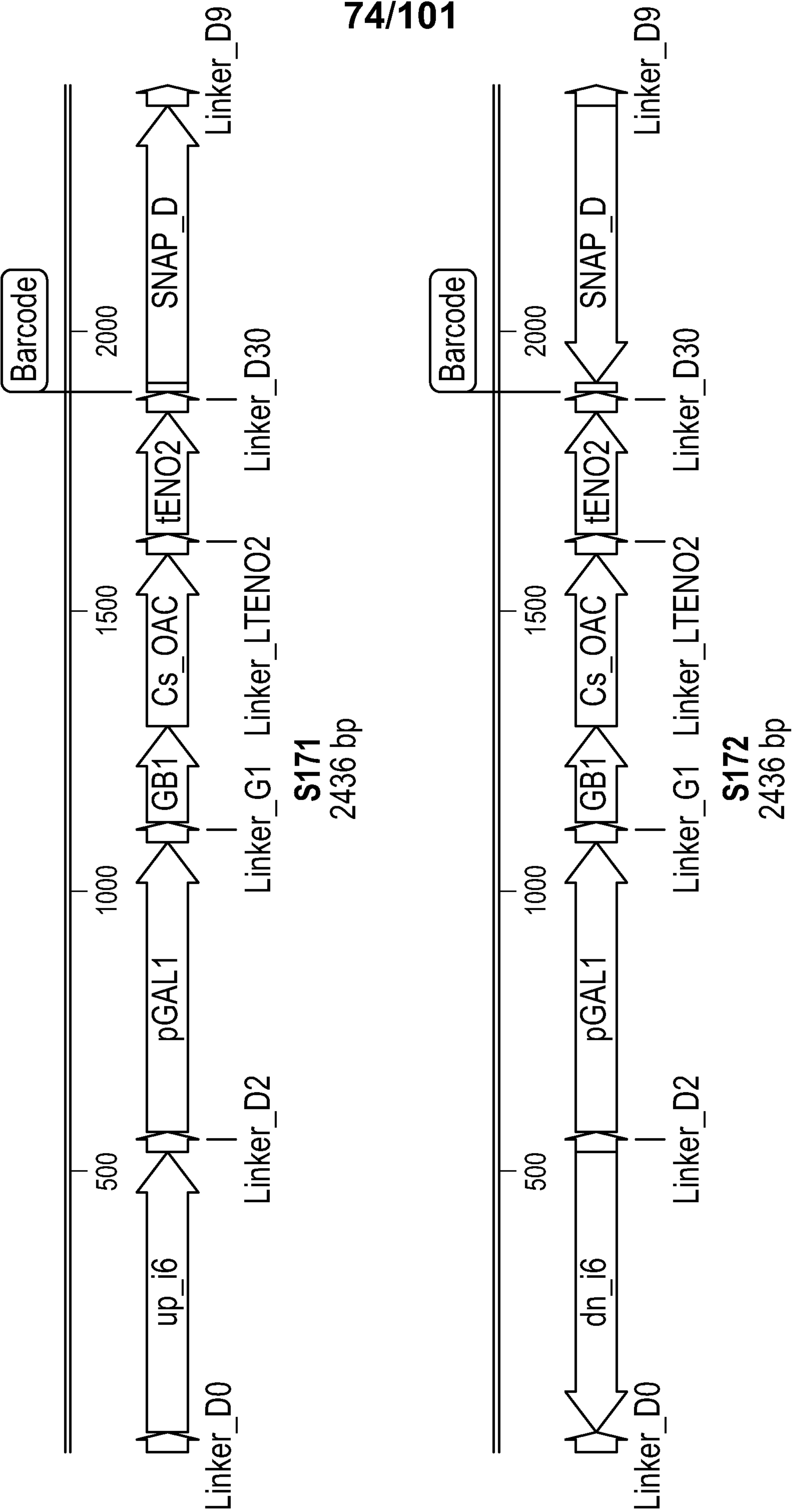
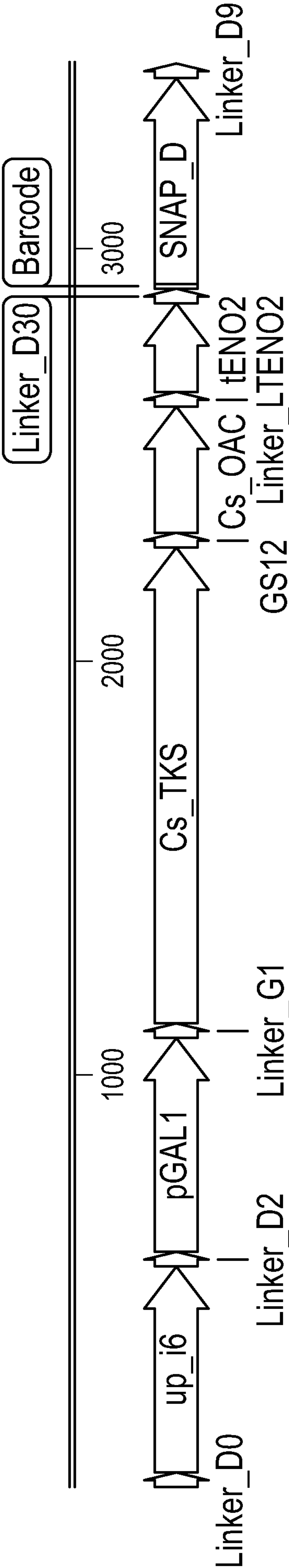
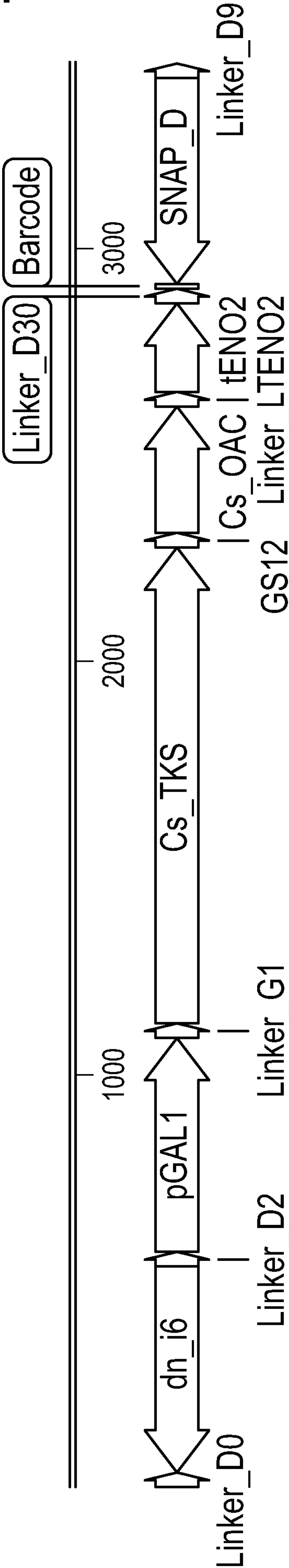


FIG. 59

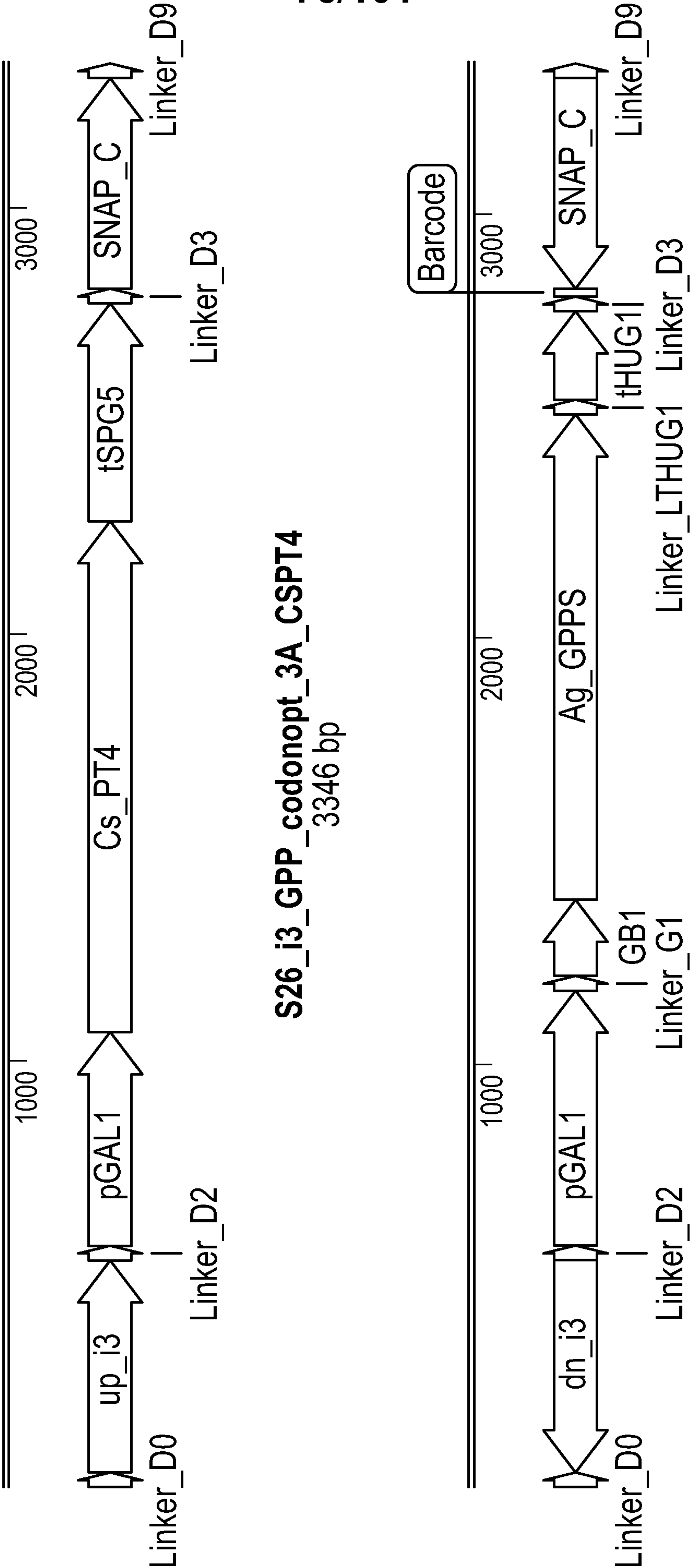


S175
3453 bp



S176
3453 bp

FIG. 60



S151

3363 bp

FIG. 61

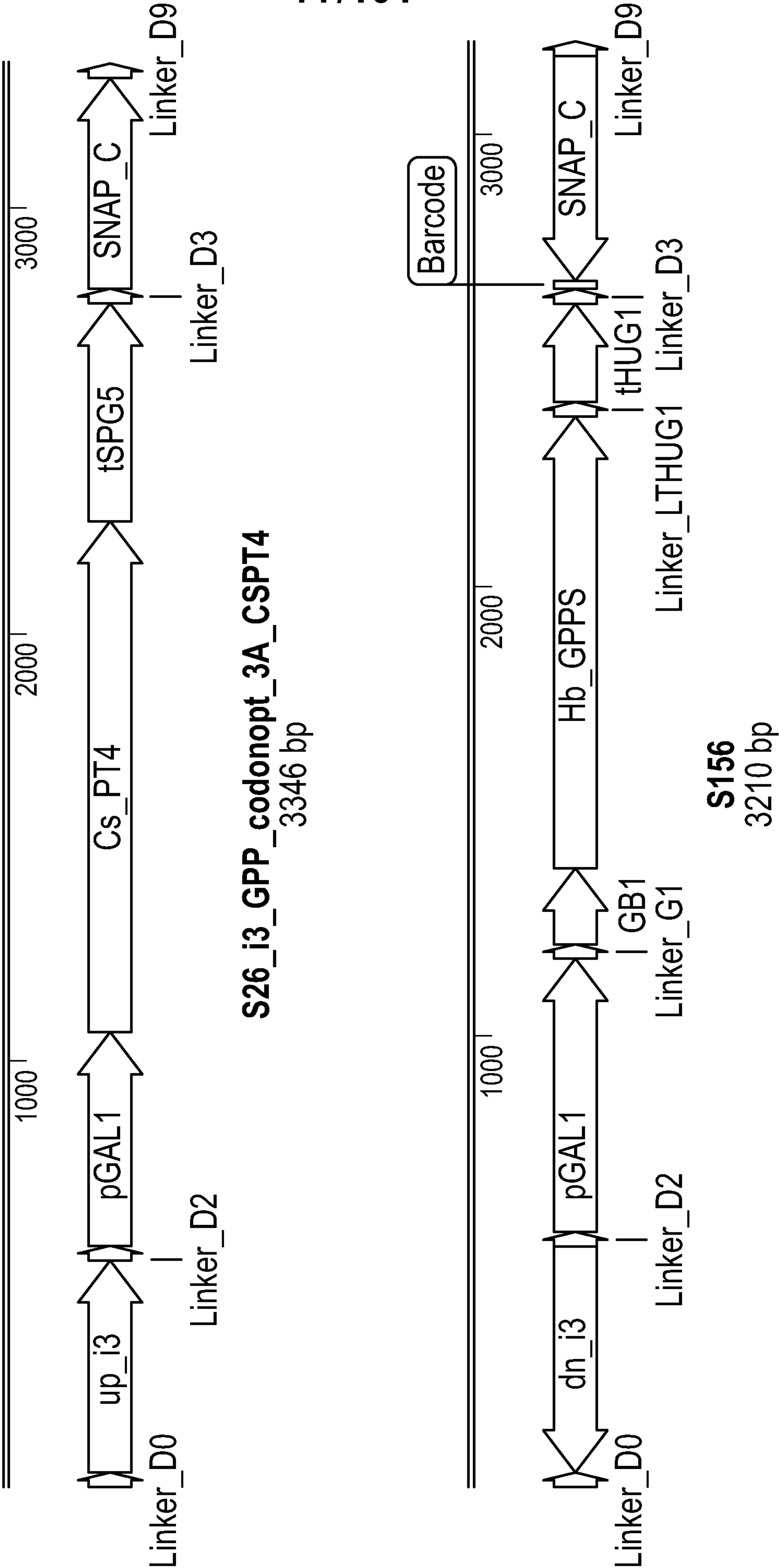
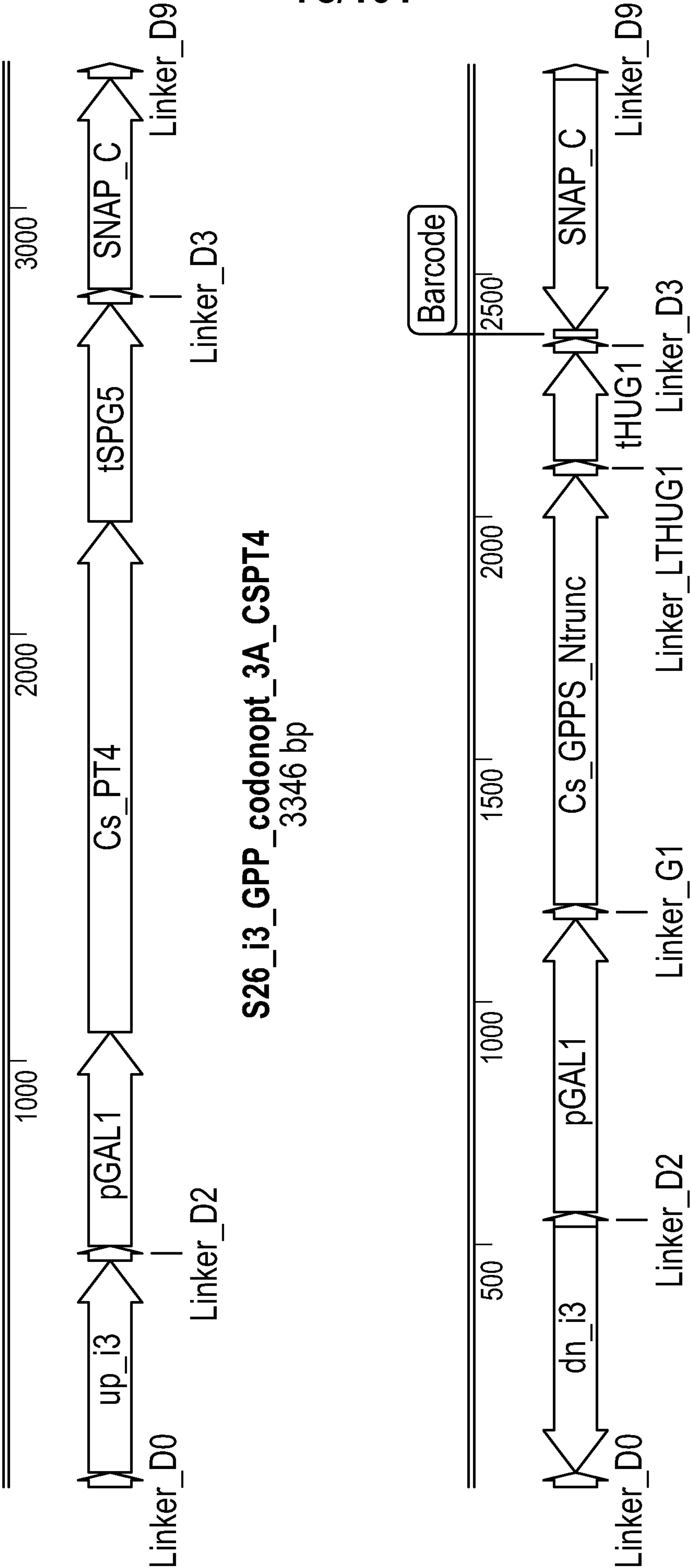


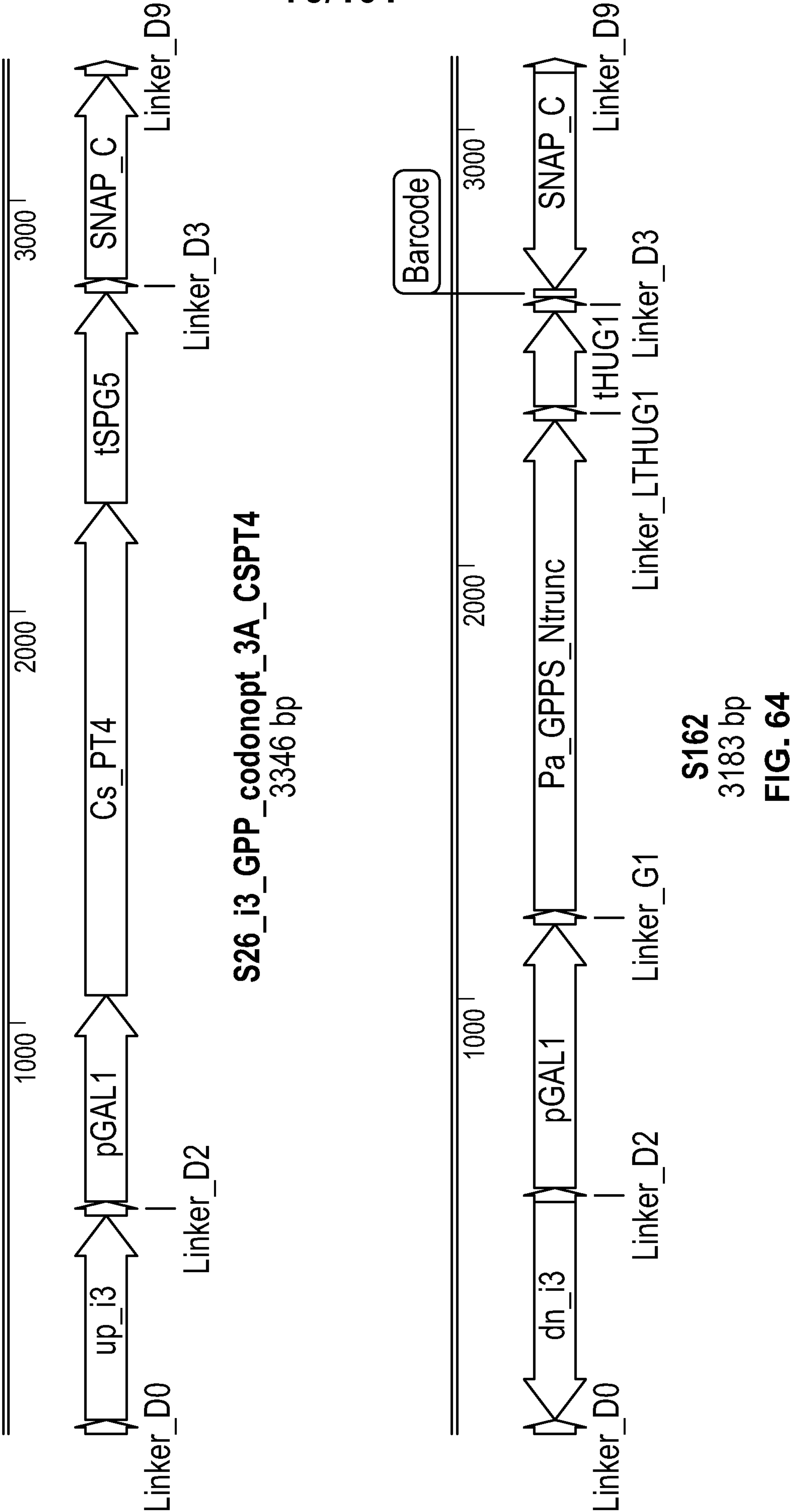
FIG. 62

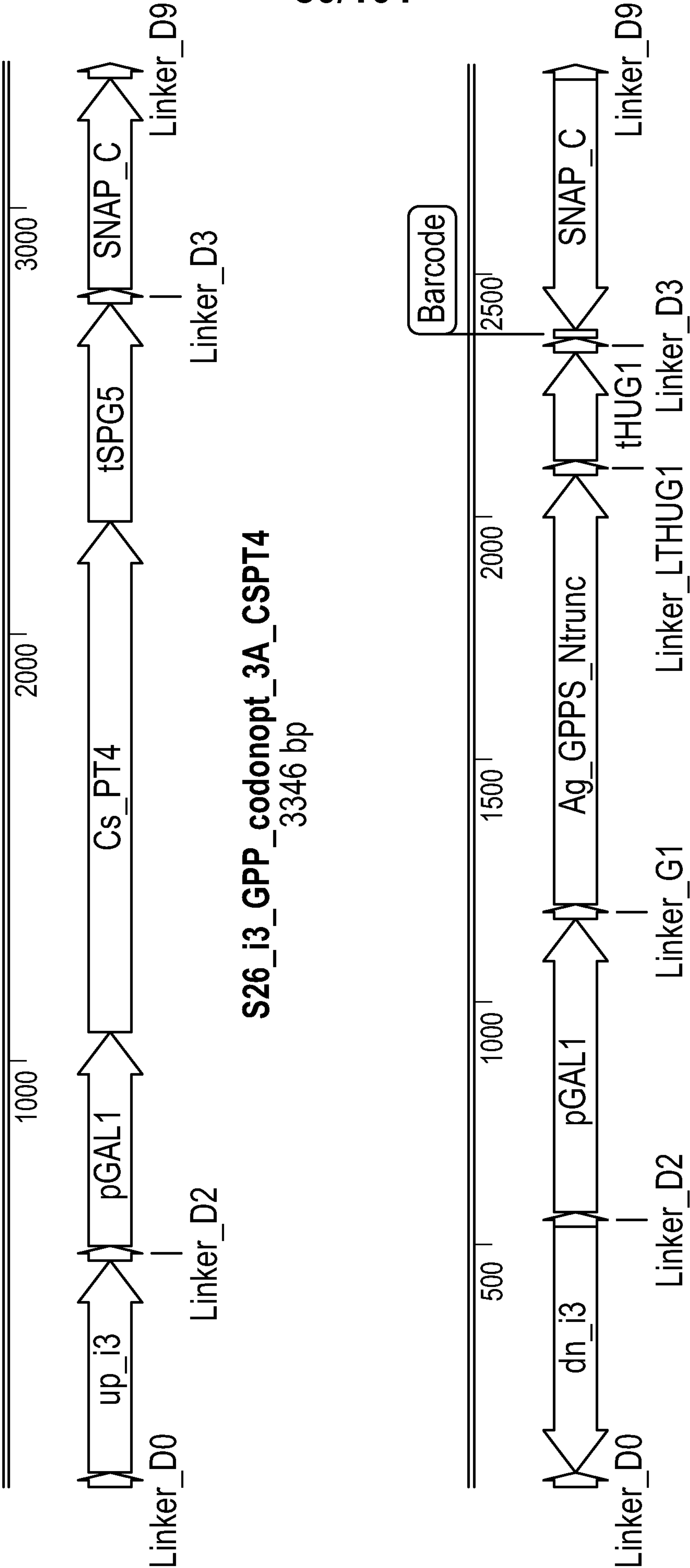


S160

2922 bp

FIG. 63

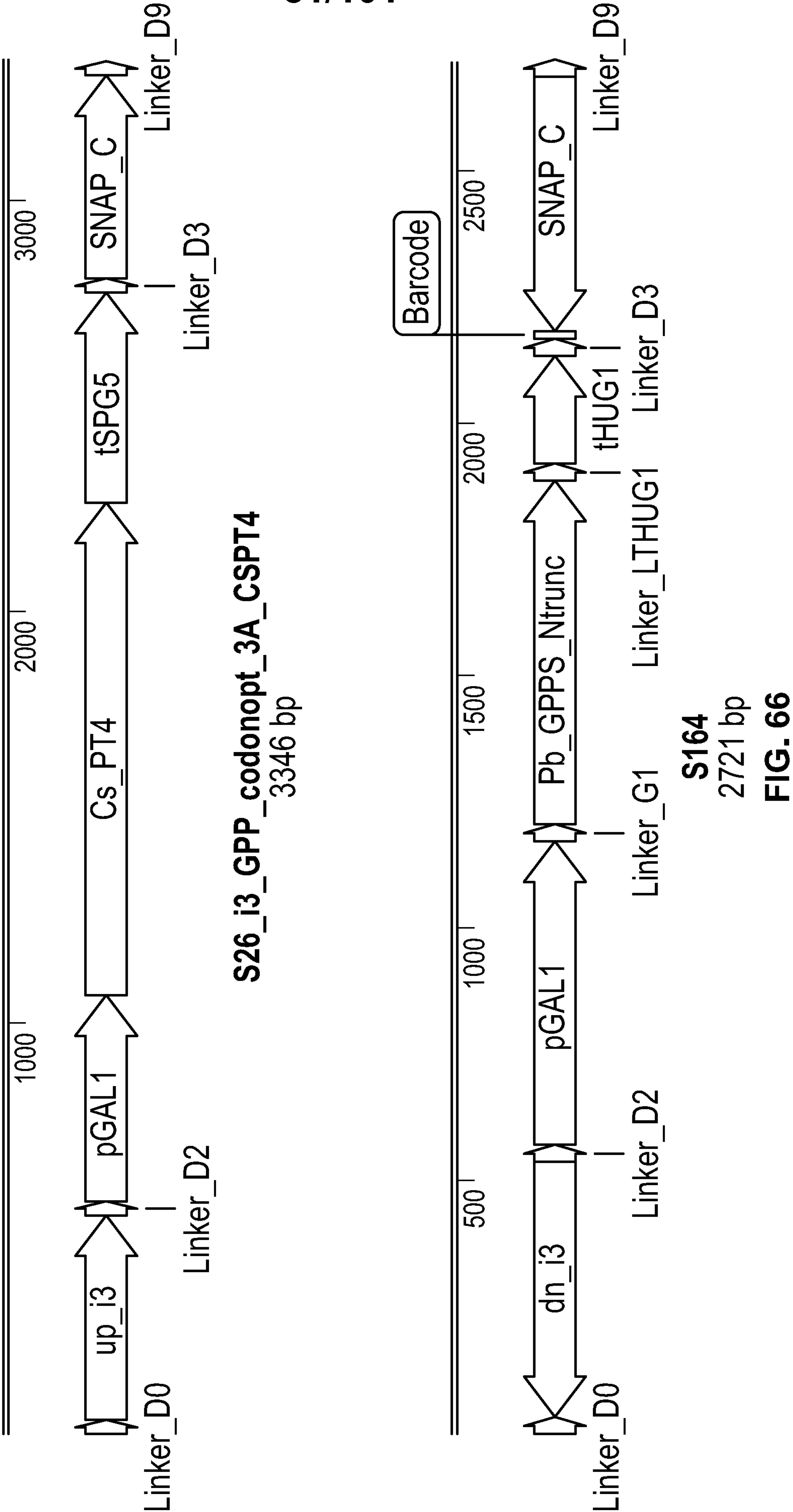




S163

2937 bp

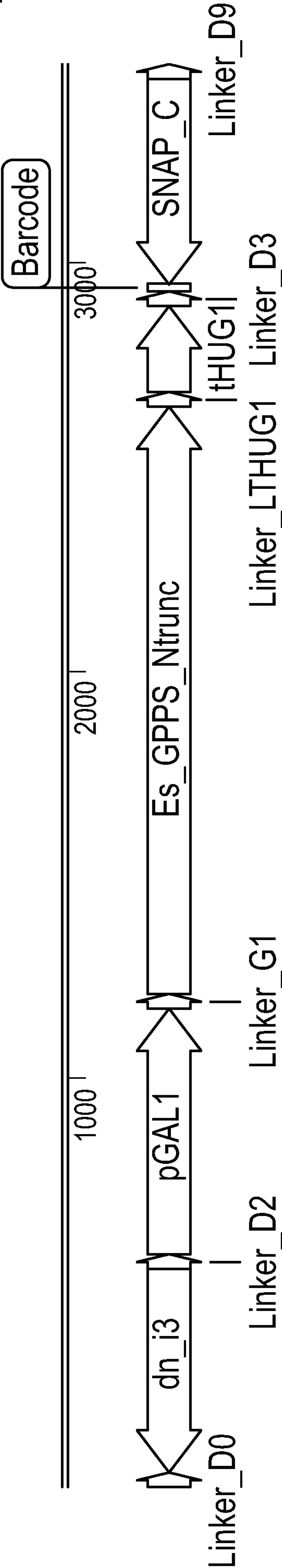
FIG. 65





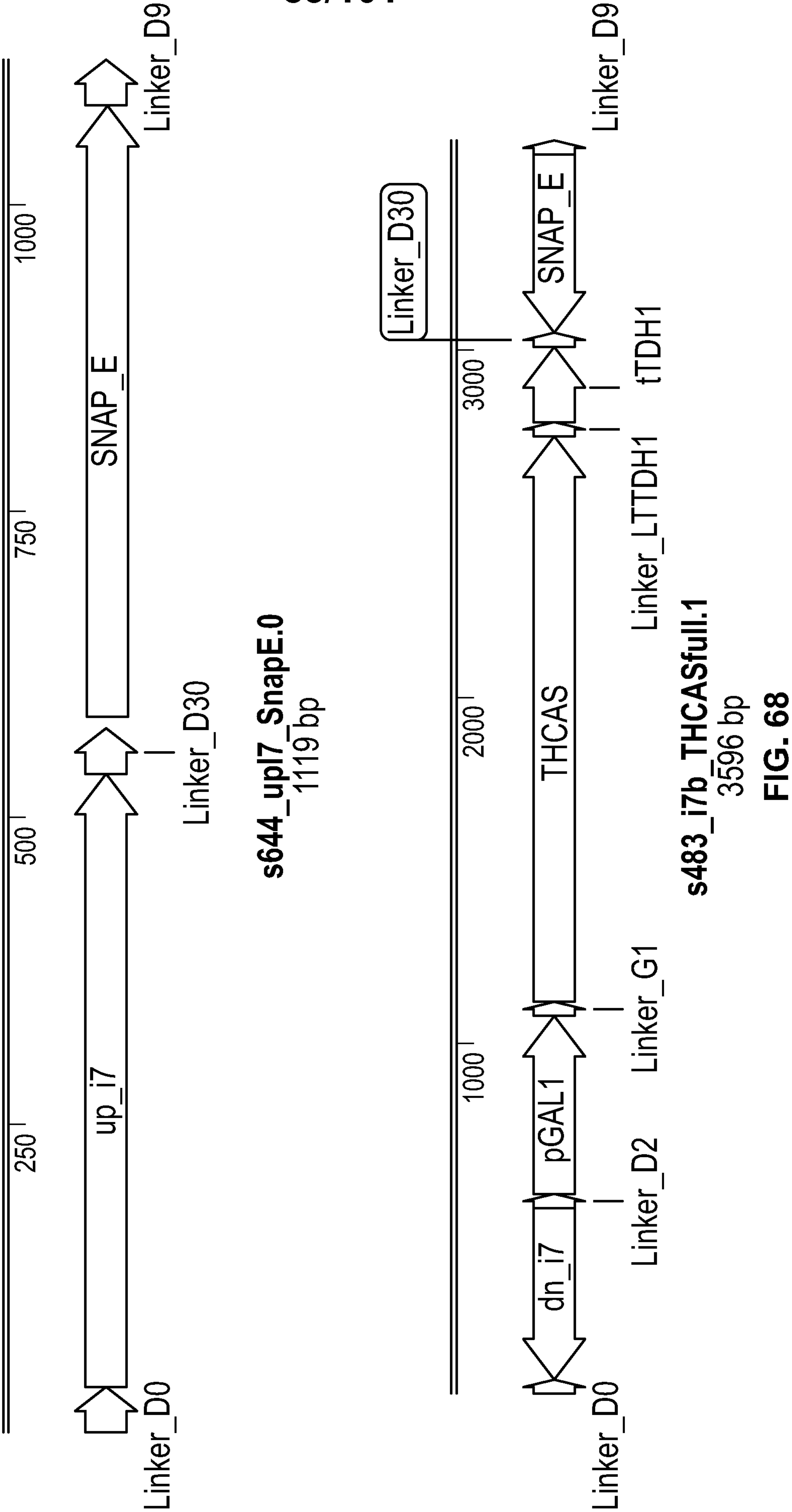
S26_i3_GPP_codonopt_3A_CSPT4
3346 bp

82/101



S166
3483 bp

FIG. 67



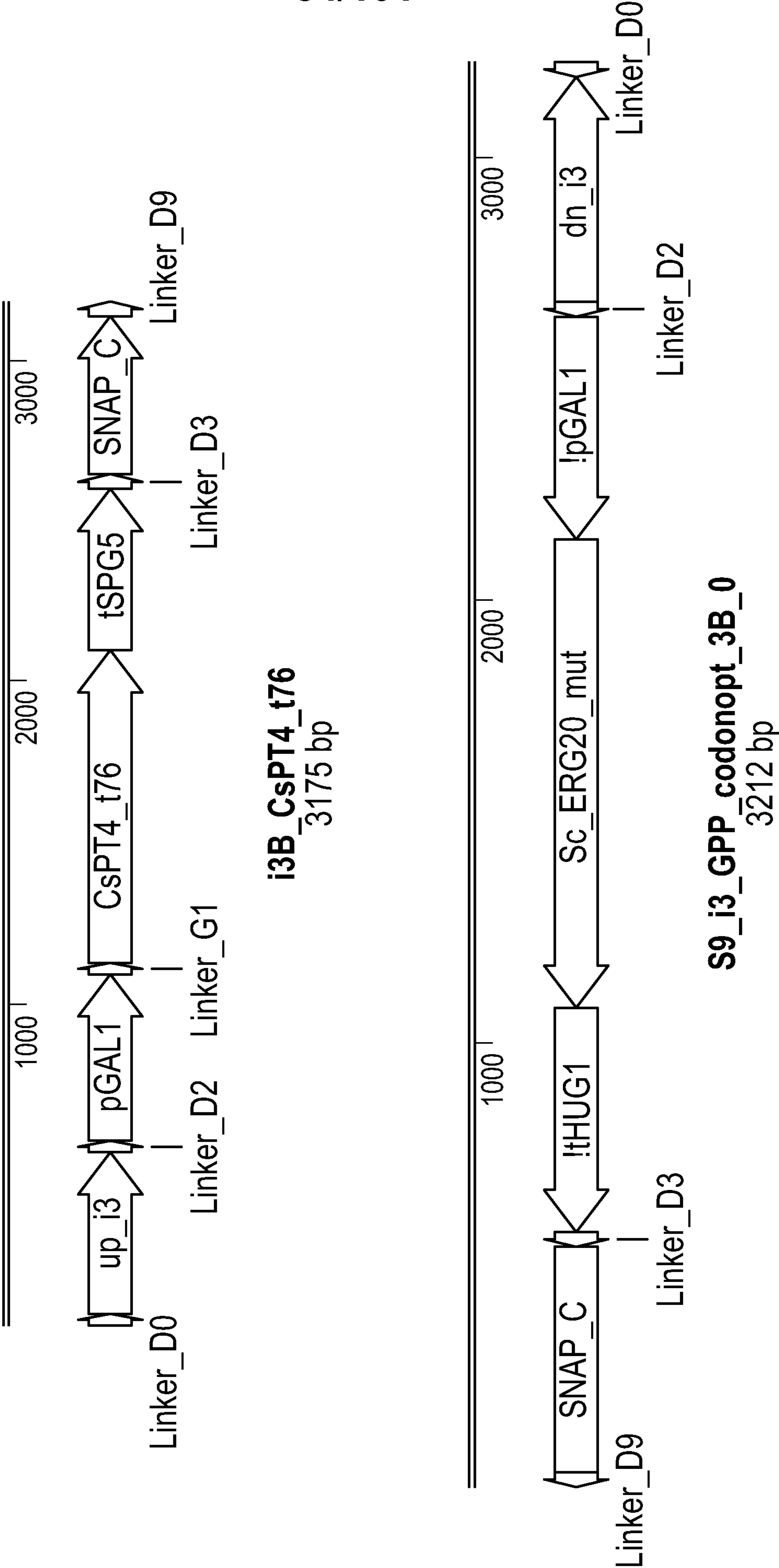


FIG. 69

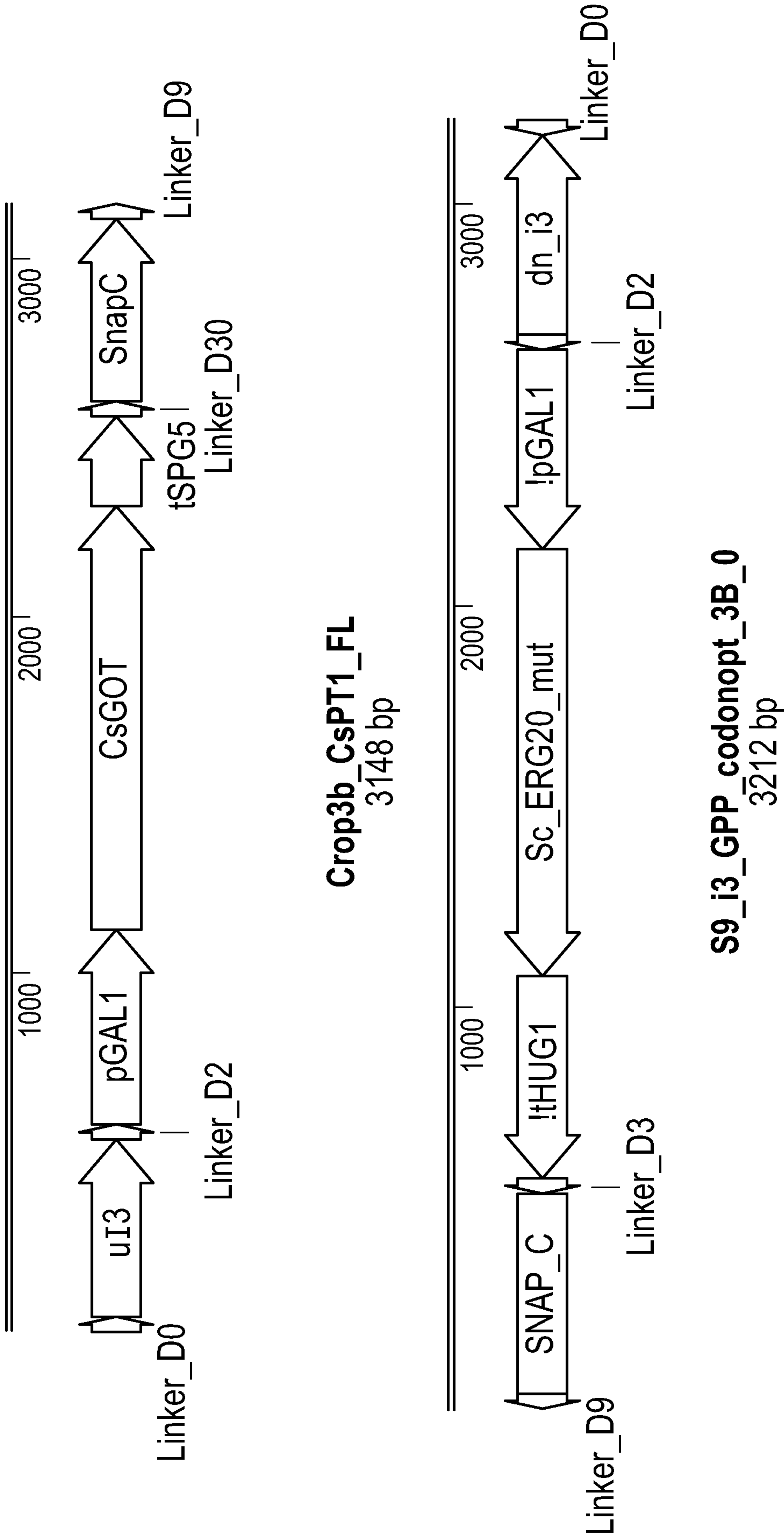


FIG. 70

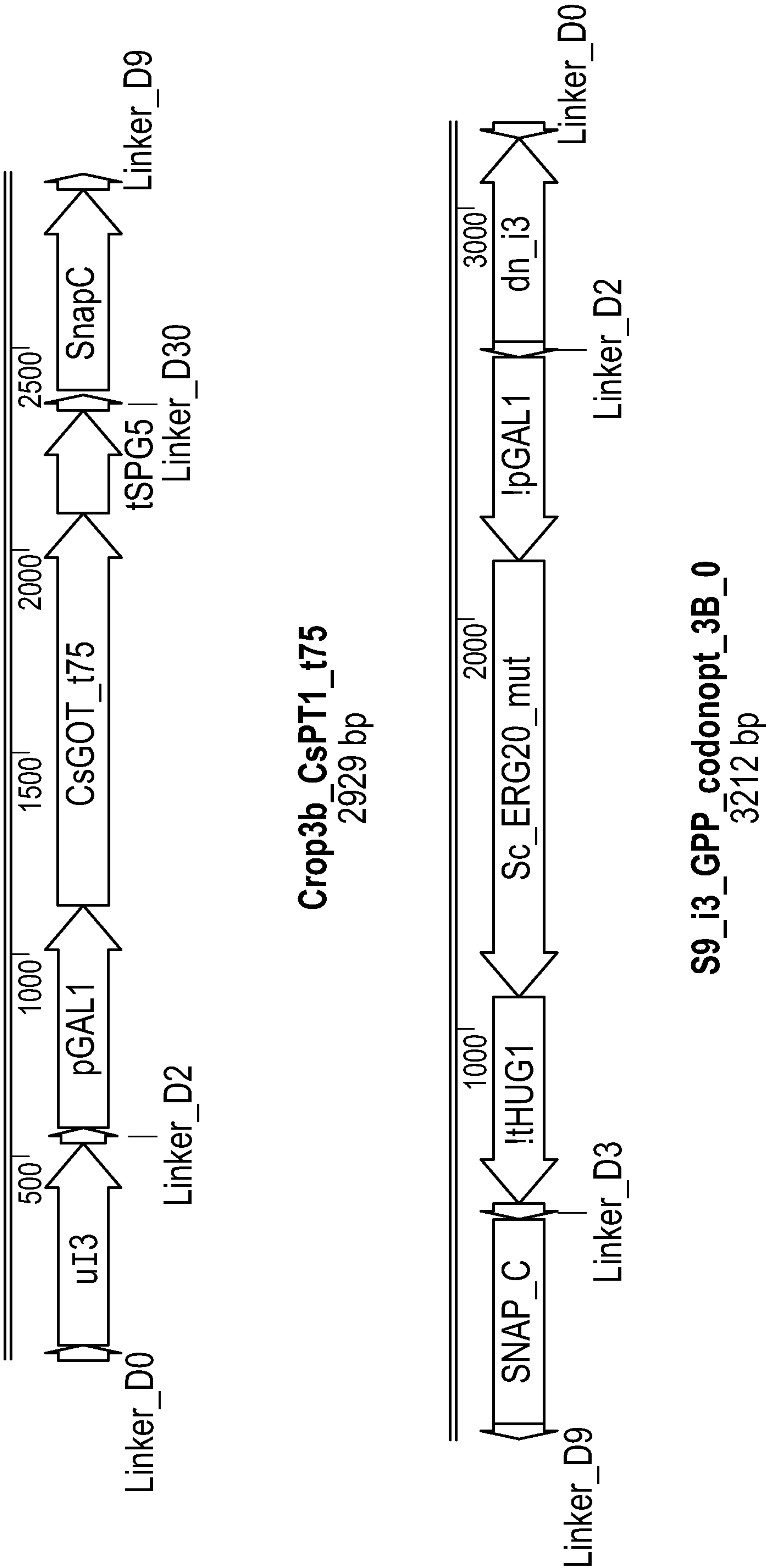


FIG. 71

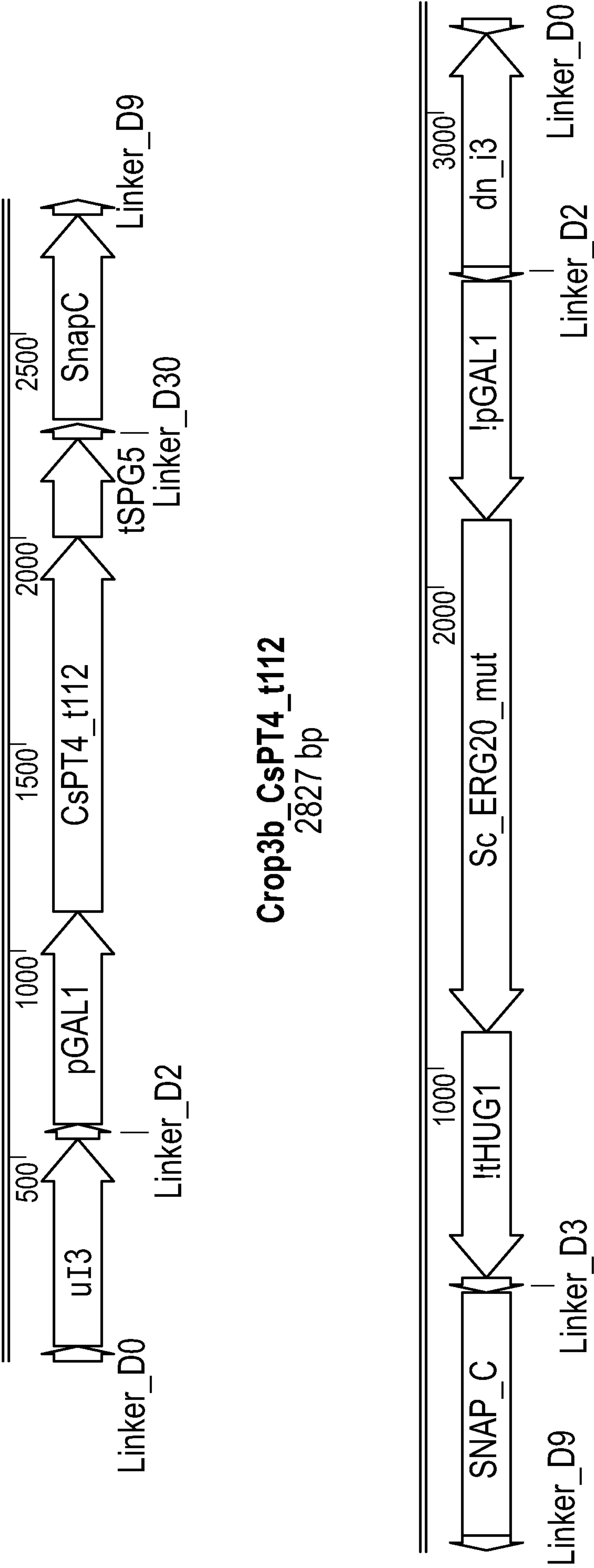


FIG. 72

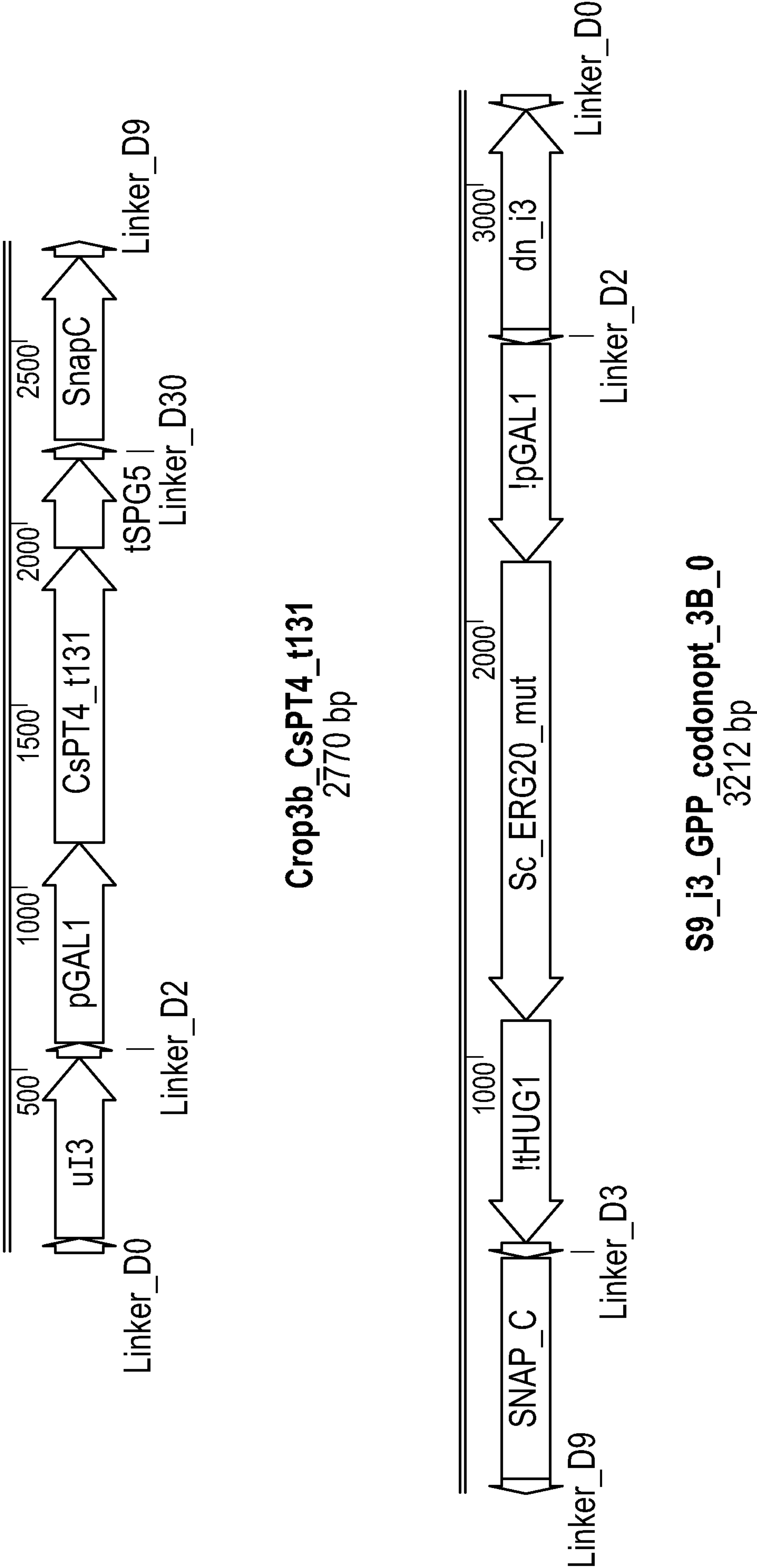


FIG. 73

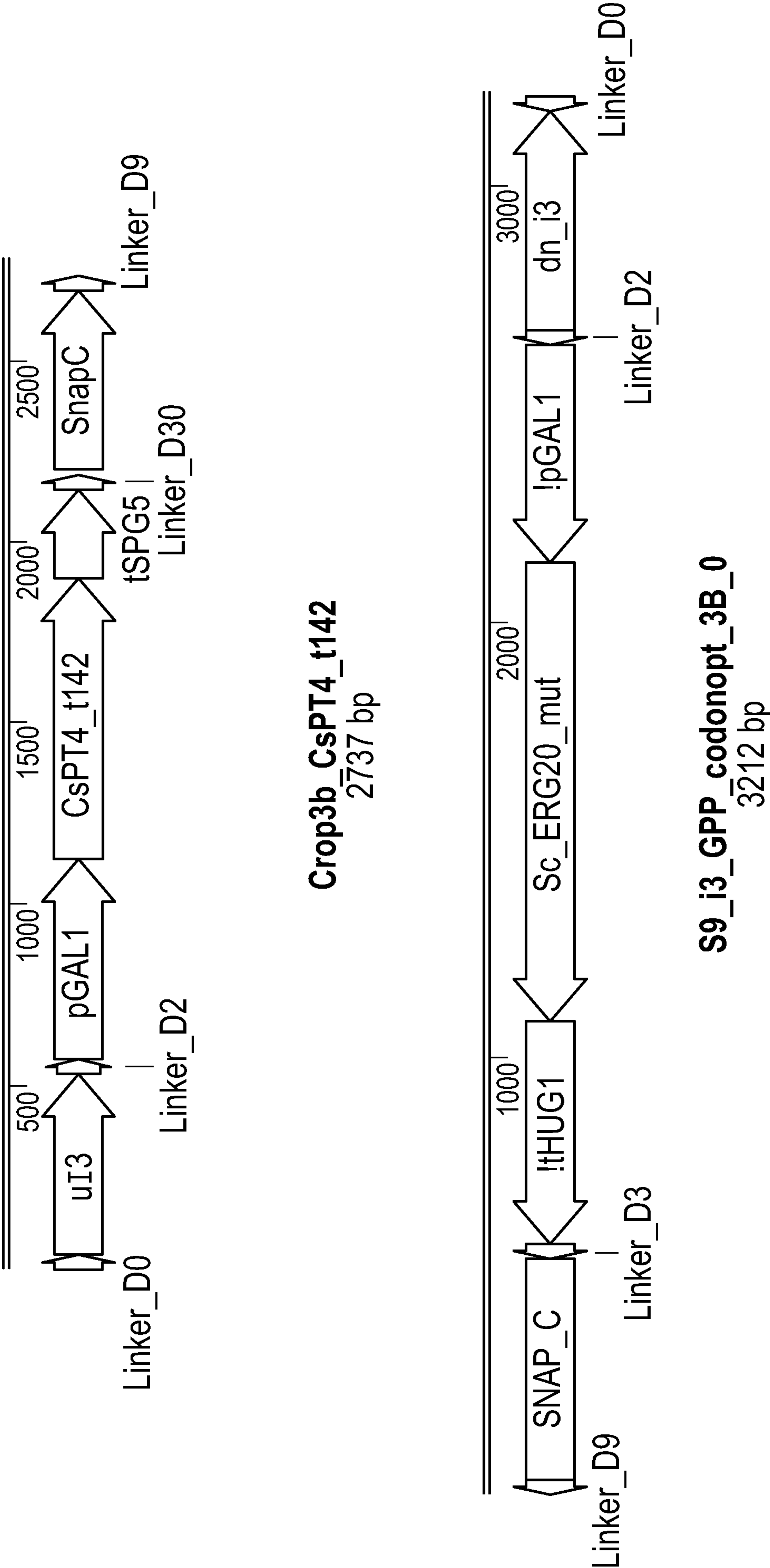


FIG. 74

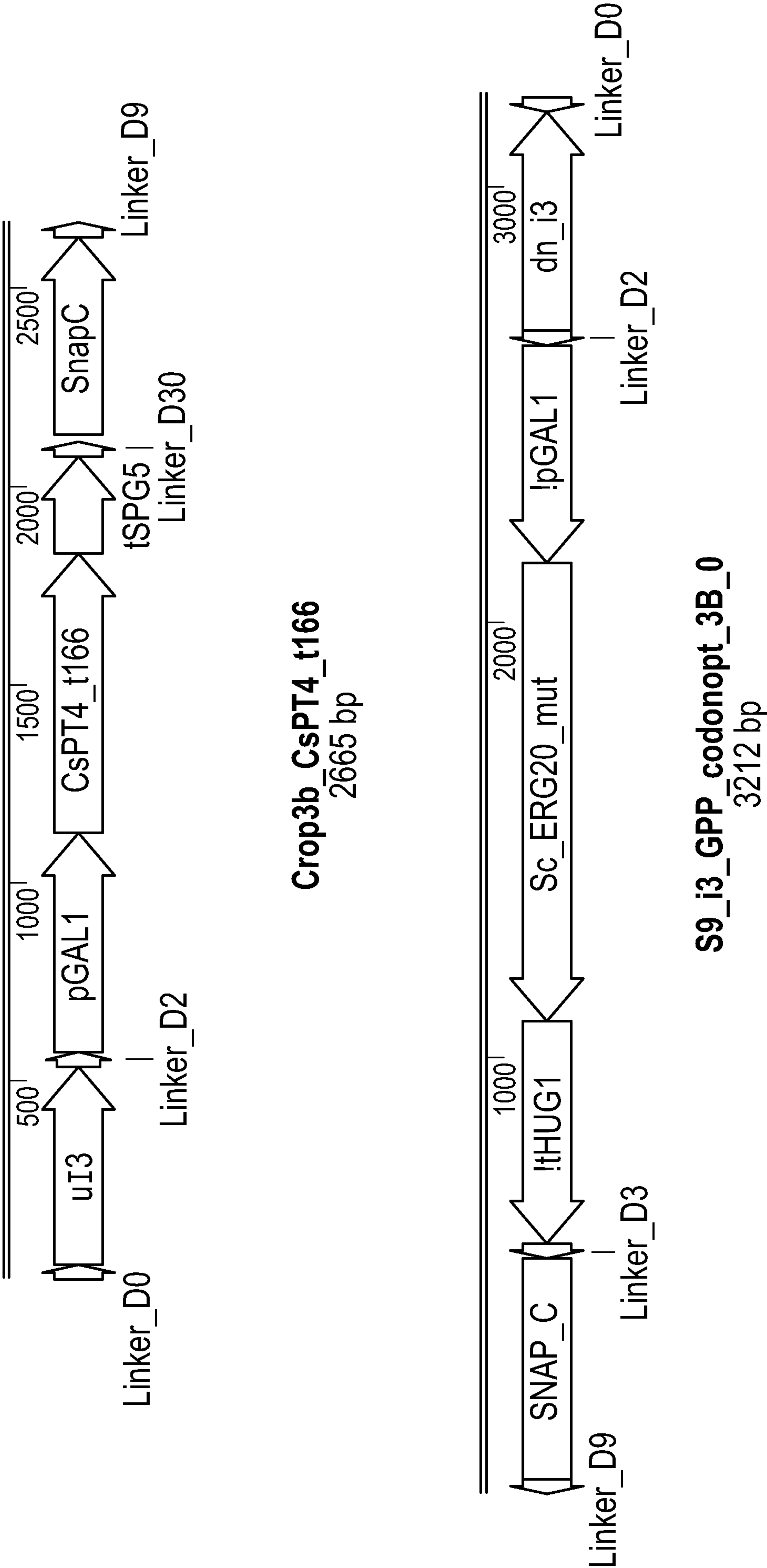


FIG. 75

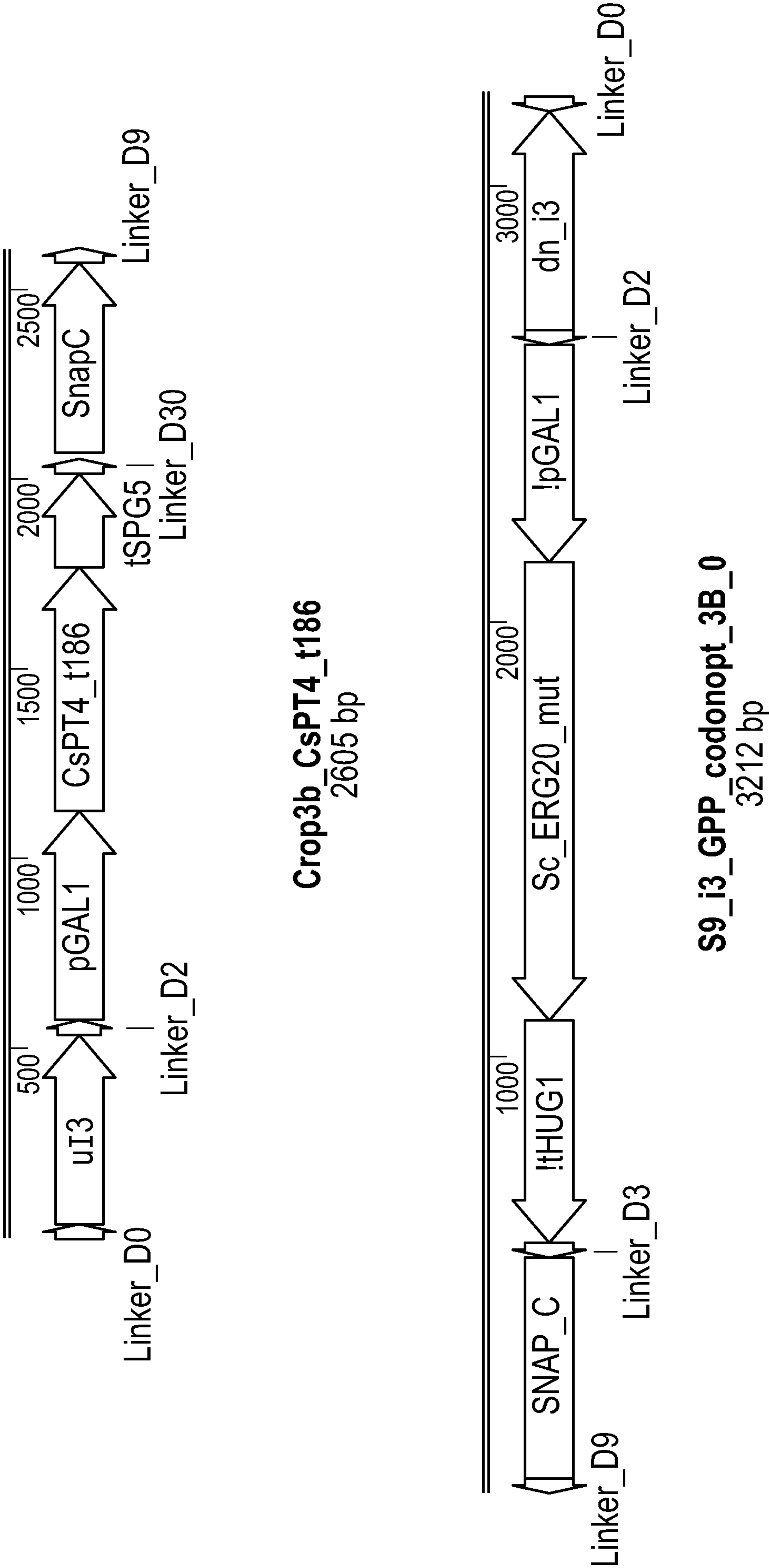


FIG. 76

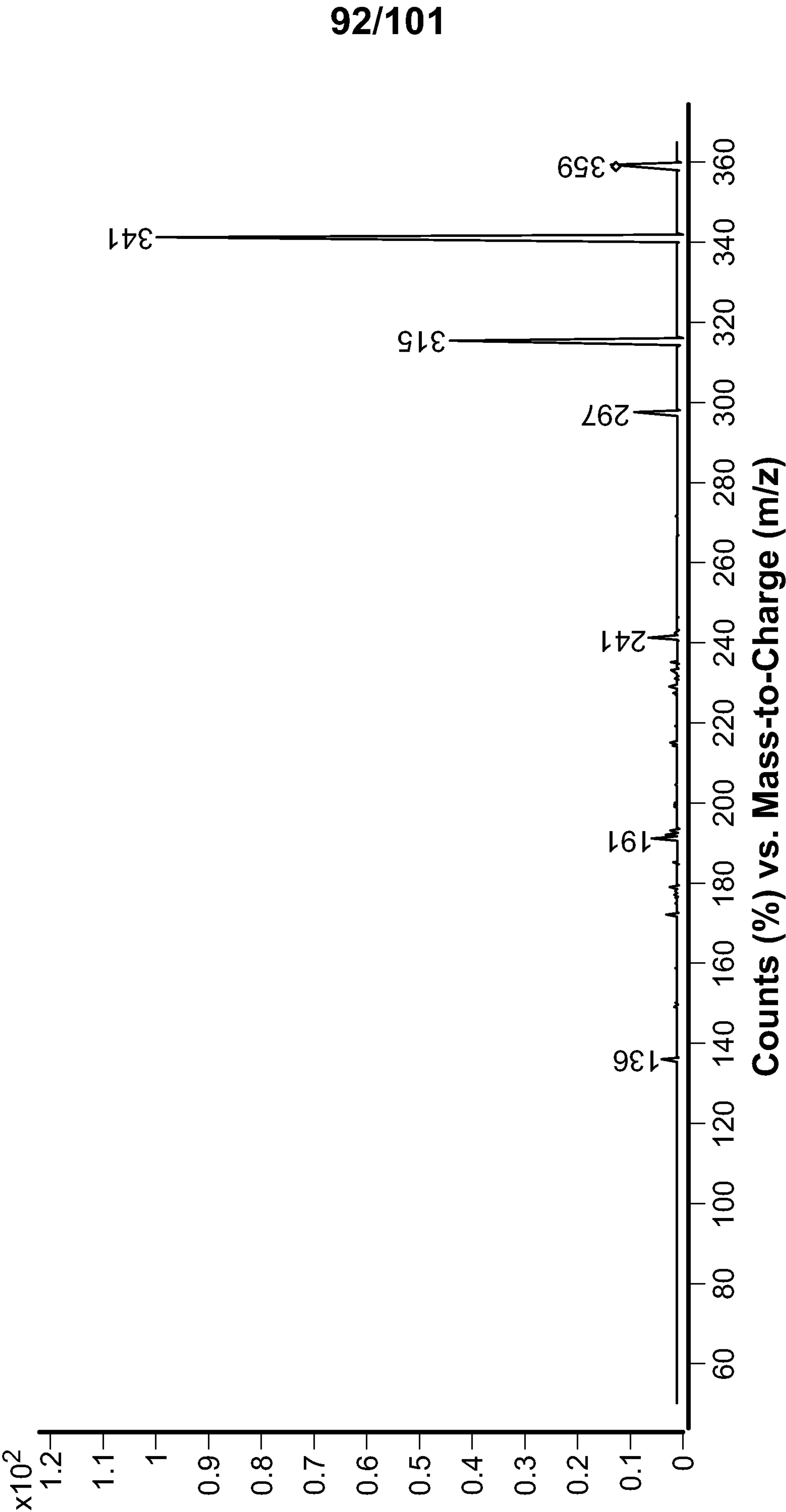


FIG. 77

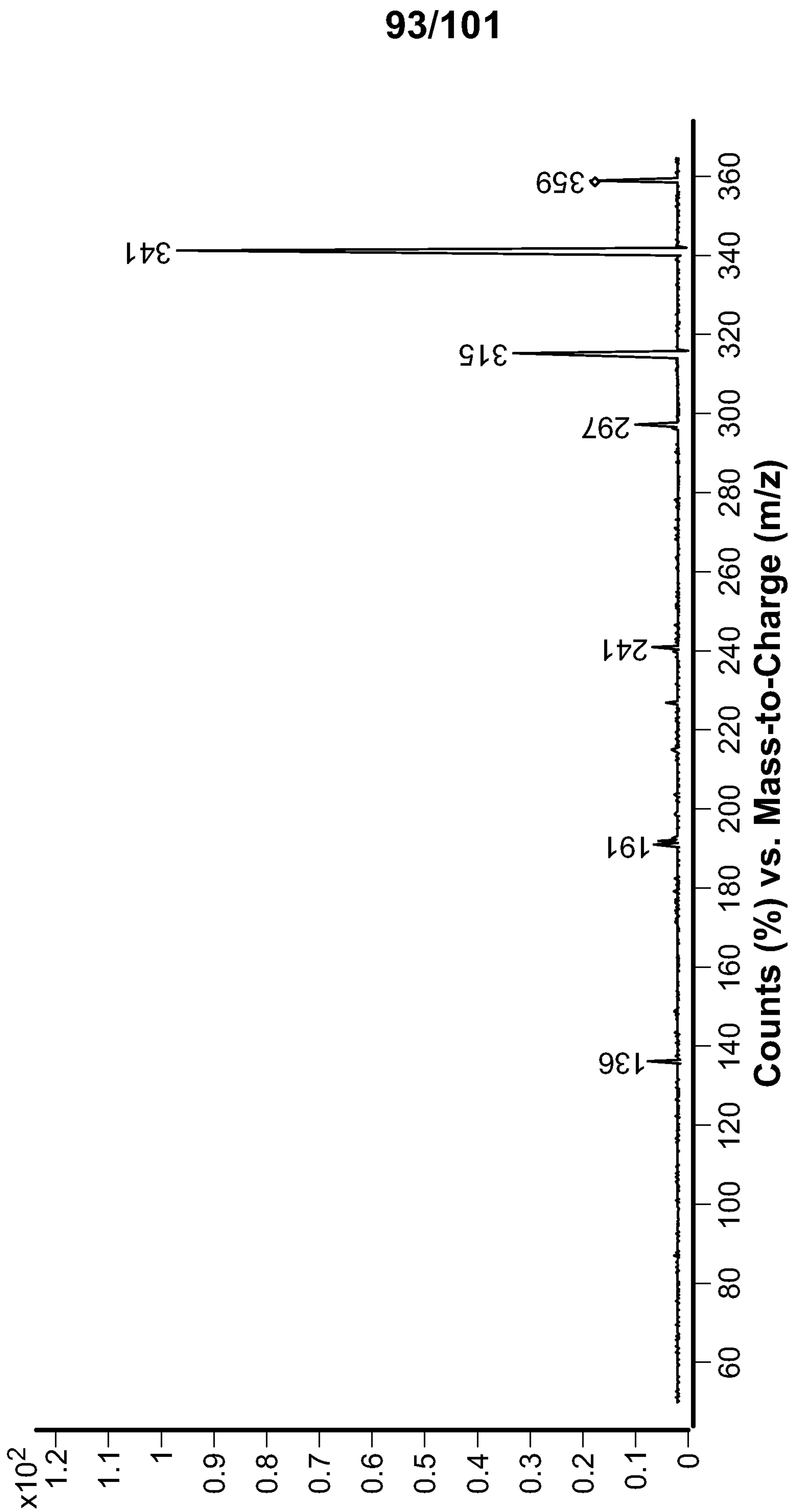
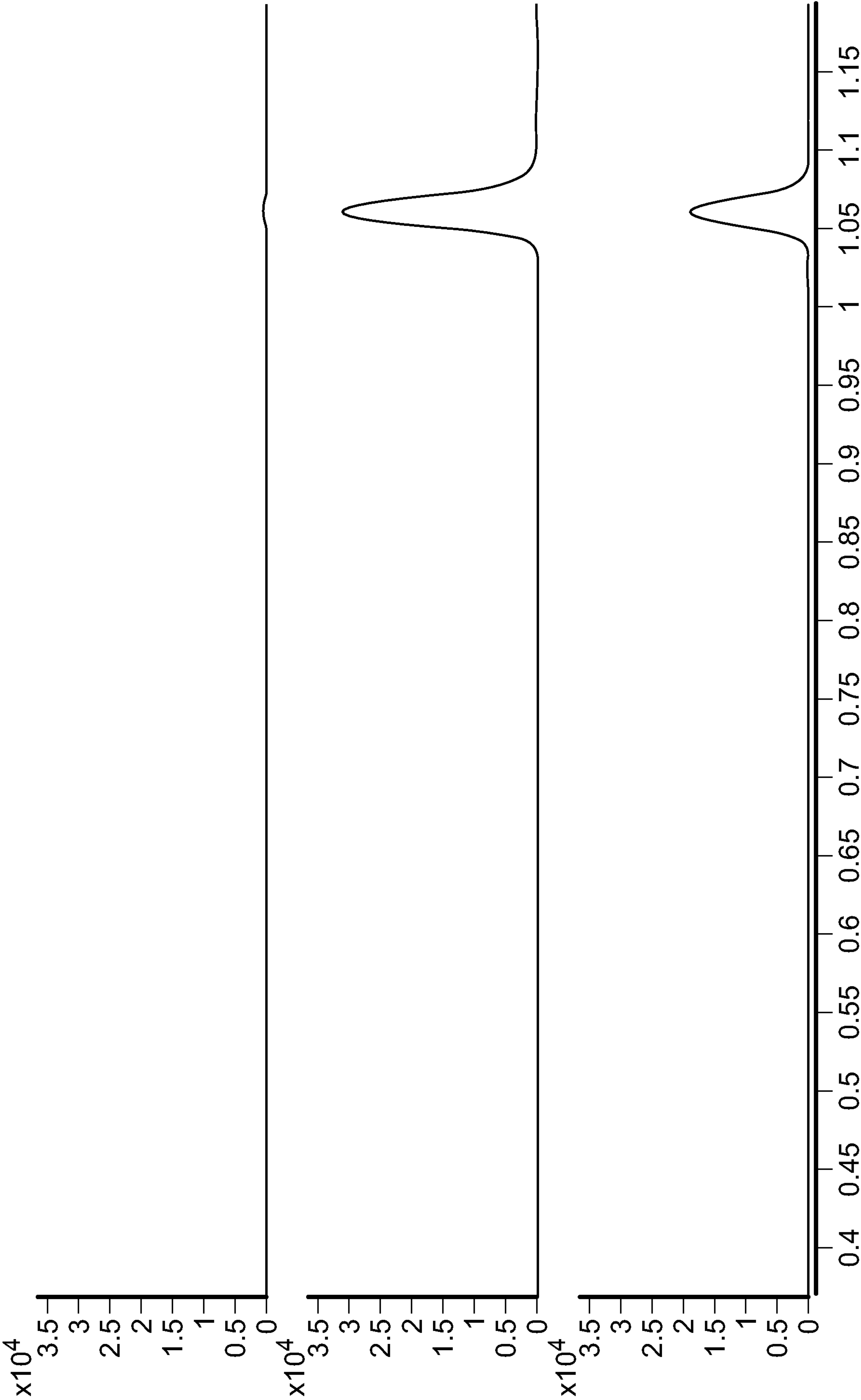


FIG. 78

94/101



Counts vs. Acquisition Time (min)

FIG. 79

95/101

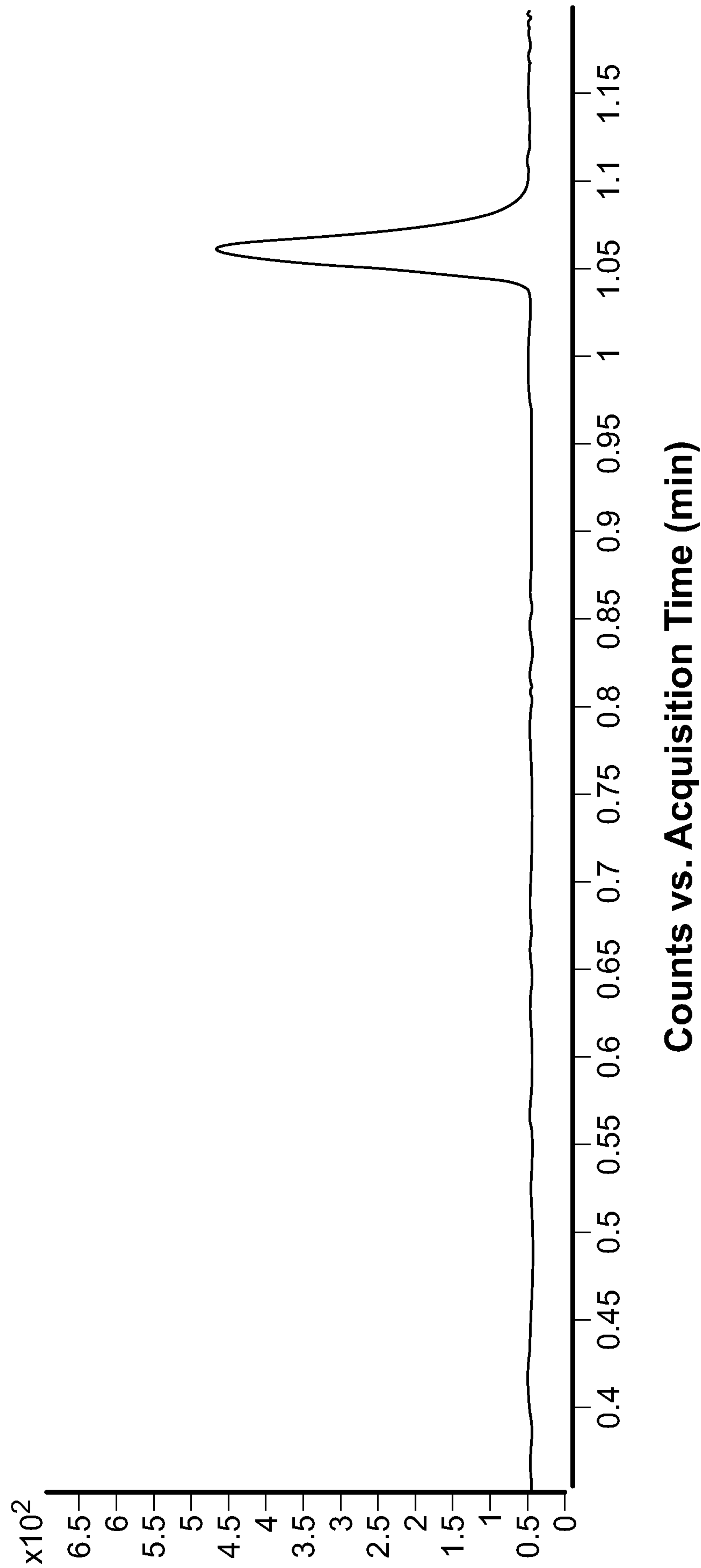
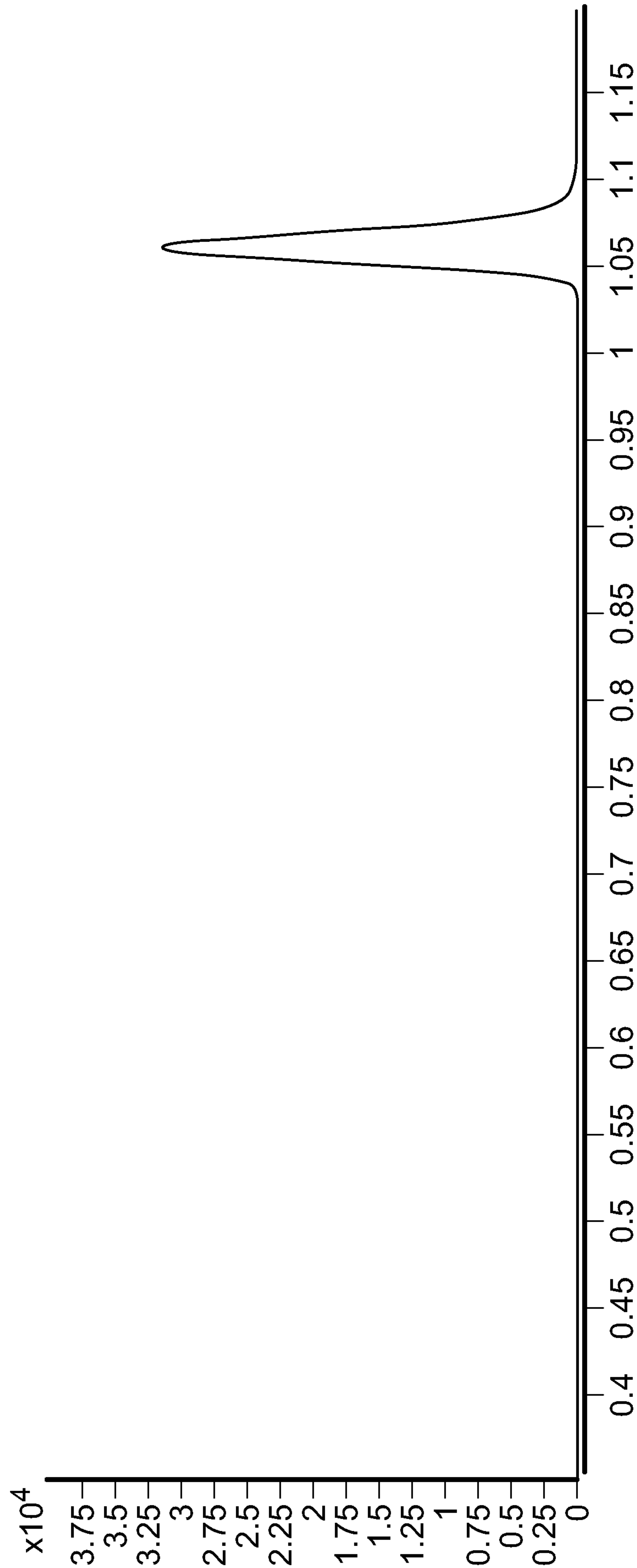


FIG. 80

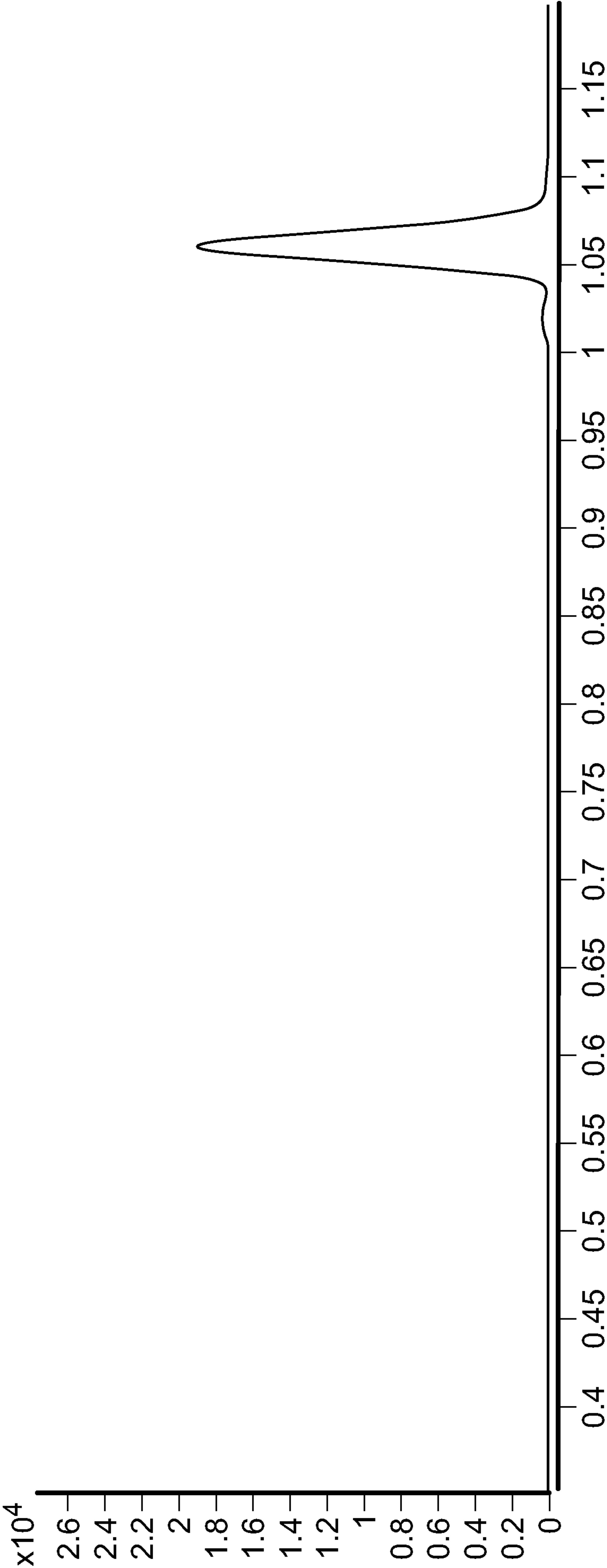
96/101



Counts vs. Acquisition Time (min)

FIG. 81

97/101



Counts vs. Acquisition Time (min)

FIG. 82

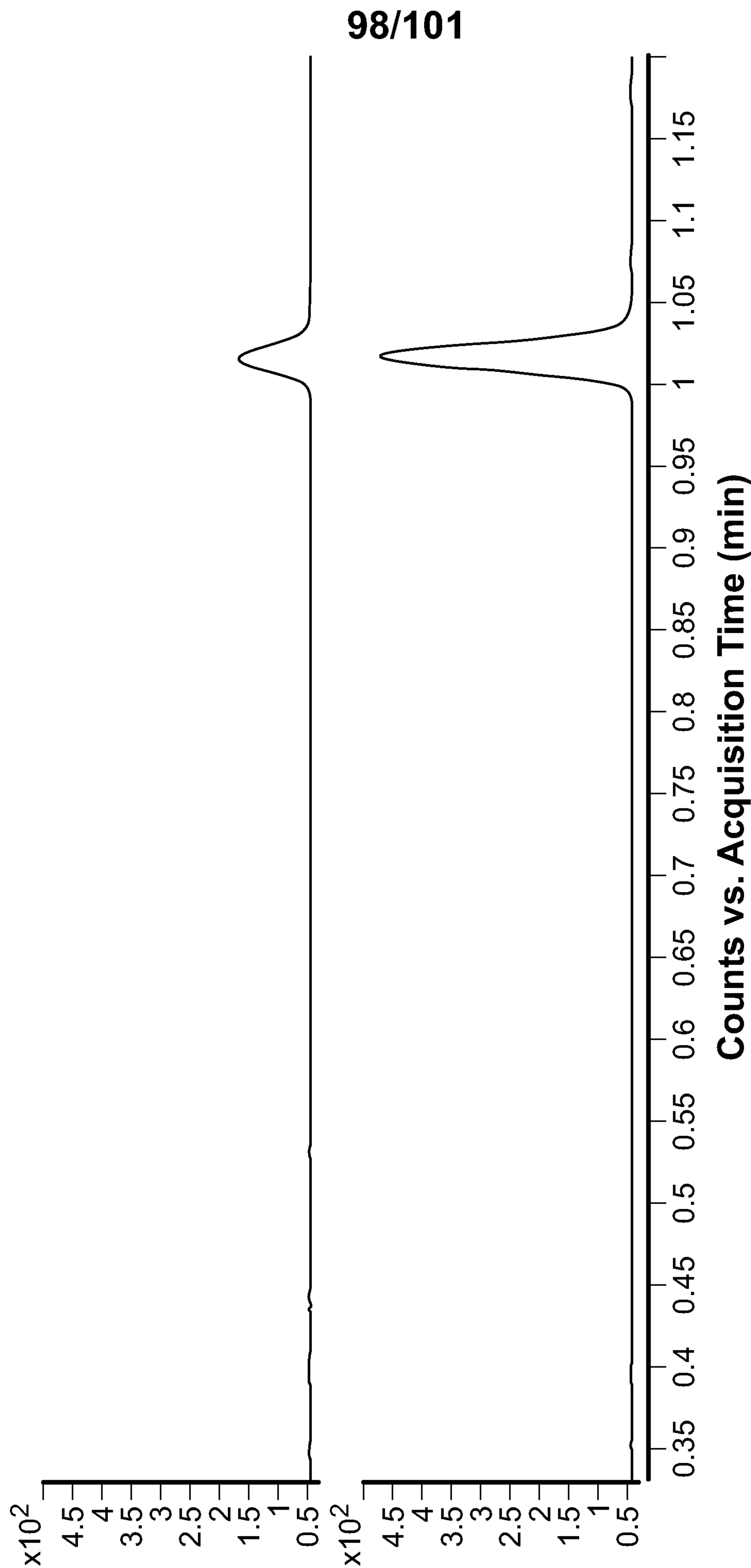


FIG. 83

99/101

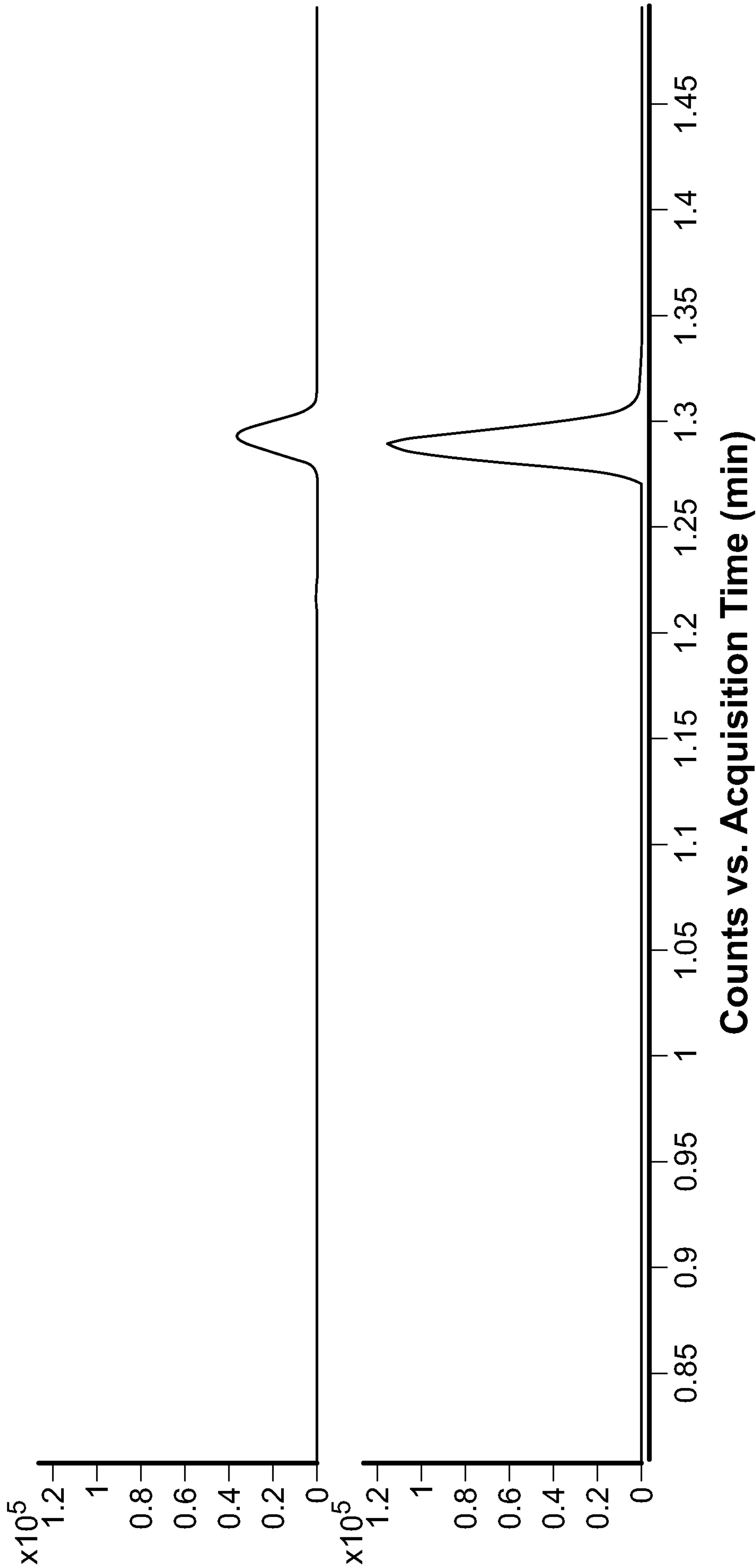
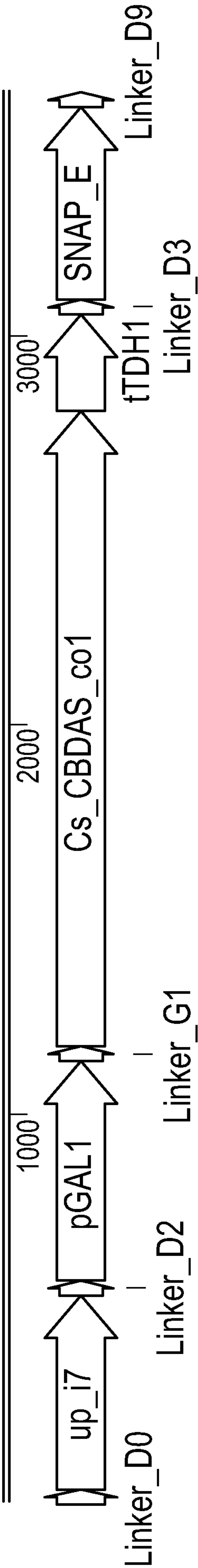
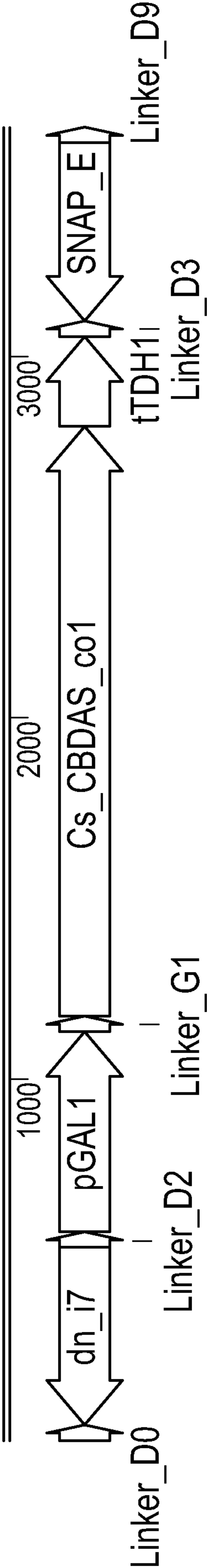


FIG. 84

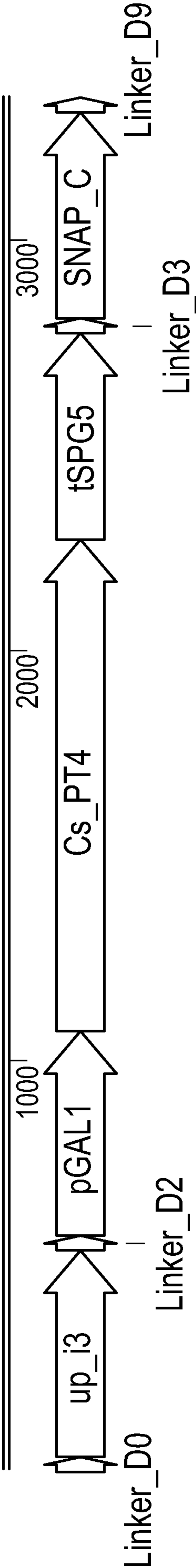


s53_i7A_Syn1_CBDASfull
3628 bp



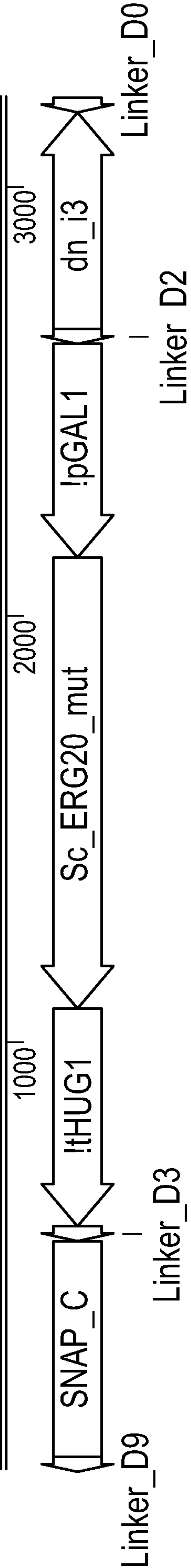
s67_i7B_Syn1_CBDASfull
3628 bp

FIG. 85



S26 i3_GPP_codonopt_3A_CSPT4
3346 bp

101/101



S9 i3_GPP_codonopt_3B_0
3212 bp

FIG. 86