Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 3069708 A1 2019/01/17

(21) 3 069 708

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2018/07/11
(87) Date publication PCT/PCT Publication Date: 2019/01/17
(85) Entrée phase nationale/National Entry: 2020/01/10
(86) N° demande PCT/PCT Application No.: US 2018/041579
(87) N° publication PCT/PCT Publication No.: 2019/014310
(30) Priorité/Priority: 2017/07/13 (US62/532,297)

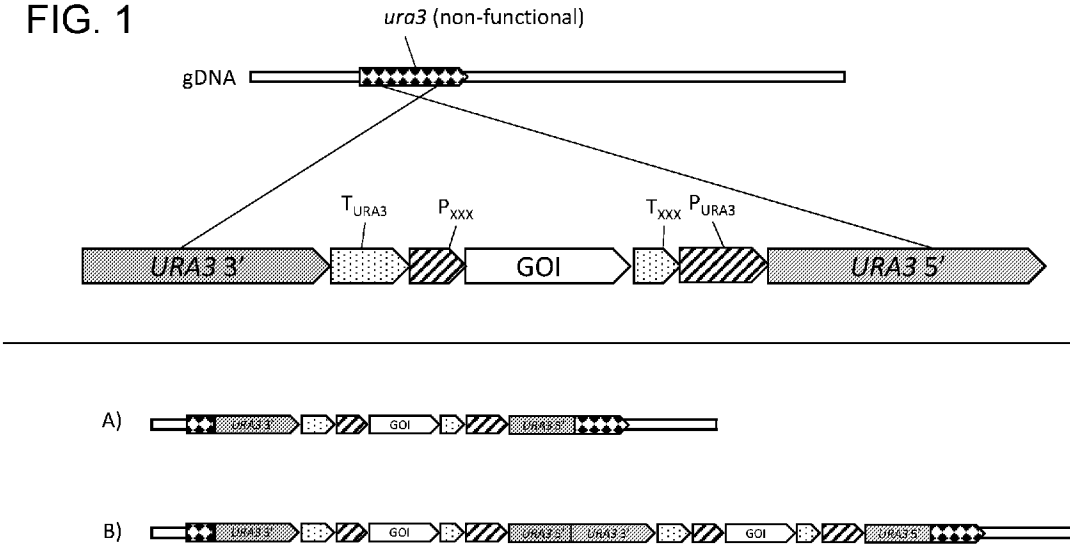
(51) Cl.Int./Int.Cl. *C07H 21/04* (2006.01),
C07K 14/415 (2006.01), *C12N 1/16* (2006.01),
C12N 1/19 (2006.01), *C12N 1/21* (2006.01),
C12N 15/09 (2006.01), *C12N 15/52* (2006.01),
C12N 5/10 (2006.01), *C12N 9/02* (2006.01),
C12N 9/88 (2006.01), *C12N 9/96* (2006.01)

(71) Demandeur/Applicant:
RADICI CHIMICA S.P.A., IT

(72) Inventeurs/Inventors:
AELING, KIMBELY ANN, US;
DE WALS, PIERRE-YVES, US;
SCOTT, ERIN, US;
KNIGHT, ERIC, US; ...

(54) Titre : PROCEDES BIOLOGIQUES POUR LA PREPARATION DE TERPENES
(54) Title: BIOLOGICAL METHODS FOR PREPARING TERPENES

FIG. 1



(57) Abrégé/Abstract:
The technology relates in part to biological methods for producing terpenes and to engineered cells and microorganisms capable of such production.

(21) 3 069 708

(13) A1

(72) Inventeurs(suite)/Inventors(continued): YI, JIAN, US; WALBRIDGE, MICHAEL, US; BEARDSLEE, THOMAS, US

(74) Agent: BERESKIN & PARR LLP/S.E.N.C.R.L.,S.R.L.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 January 2019 (17.01.2019)

WIPO | PCT

(10) International Publication Number
WO 2019/014310 A1

(51) International Patent Classification:
C07H 21/04 (2006.01) *C12N 9/02* (2006.01)
C07K 14/415 (2006.01) *C12N 9/88* (2006.01)
C12N 1/16 (2006.01) *C12N 9/96* (2006.01)
C12N 1/19 (2006.01) *C12N 15/09* (2006.01)
C12N 1/21 (2006.01) *C12N 15/52* (2006.01)
C12N 5/10 (2006.01)

(21) International Application Number:
PCT/US2018/041579

(22) International Filing Date:
11 July 2018 (11.07.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/532,297 13 July 2017 (13.07.2017) US

(71) Applicant: **VERDEZYNE (ABC), LLC** [US/US]; 2735 Sand Hill Road, Menlo Park, CA 94025 (US).

(72) Inventor: **AELING, Kimbely, Ann**; 324 Plaza Estival, San Clemente, CA 92672 (US).

(74) Agent: **DICKINSON, Kari A.** et al.; Grant IP, Inc., C/o CPA Global, 900 Second Avenue S., Suite 600, Minneapolis, MN 55402 (US).

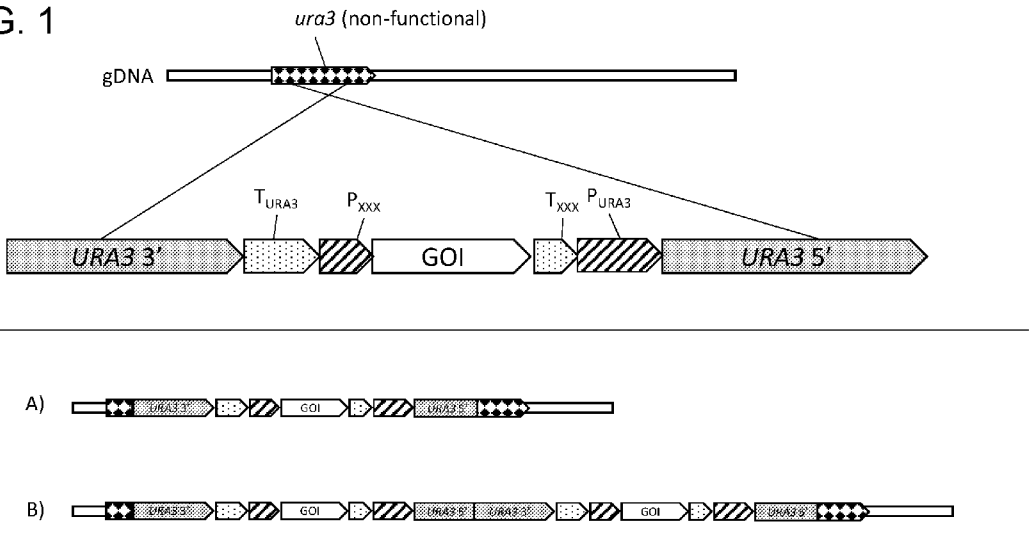
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS,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

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

Published:
— *with international search report (Art. 21(3))*

(54) Title: BIOLOGICAL METHODS FOR PREPARING TERPENES

FIG. 1



(57) Abstract: The technology relates in part to biological methods for producing terpenes and to engineered cells and microorganisms capable of such production.

WO 2019/014310 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 3
CONTENANT LES PAGES 1 À 261

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 3
CONTAINING PAGES 1 TO 261

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

NOM DU FICHIER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

BIOLOGICAL METHODS FOR PREPARING TERPENES

Related Patent Applications

This patent application claims the benefit of U.S. provisional patent application no. 62/532,297, filed on July 13, 2017, entitled BIOLOGICAL METHODS FOR PREPARING TERPENES, naming Kimberly Ann Aeling as inventor, and designated by Attorney Docket No. VRD-3002-PV. This patent application is related to an International patent application, filed simultaneously herewith, entitled BIOLOGICAL METHODS FOR MODIFYING CELLULAR CARBON FLUX, naming Tom Beardslee as inventor, and designated by Attorney Docket No. VAL-3001-PC, which claims the benefit of U.S. provisional patent application no. 62/532,292, filed on July 13, 2017, entitled BIOLOGICAL METHODS FOR MODIFYING CELLULAR CARBON FLUX, naming Tom Beardslee as inventor, and designated by Attorney Docket No. VRD-3001-PV. This patent application also is related to U.S. provisional patent application no. 61/222,902 filed on July 2, 2009, entitled BIOLOGICAL METHODS FOR PREPARING ADIPIC ACID, naming Stephen Picataggio as inventor, and designated by Attorney Docket No. VRD-1001-PV. This patent application also is related to International patent application no. PCT/US2010/040837 filed on July 1, 2010, entitled BIOLOGICAL METHODS FOR PREPARING ADIPIC ACID, naming Stephen Picataggio and Tom Beardslee as inventors, and designated by Attorney Docket No. VRD-1001-PC. This patent application also is related to U.S. provisional patent application no. 61/430,097 filed on January 5, 2011, entitled BIOLOGICAL METHODS FOR PREPARING ADIPIC ACID, naming Stephen Picataggio and Tom Beardslee as inventors, and designated by Attorney Docket No. VRD-1001-PV2. This patent application also is related to U.S. provisional patent application no. 61/482,160 filed on May 3, 2011, entitled BIOLOGICAL METHODS FOR PREPARING ADIPIC ACID, naming Stephen Picataggio and Tom Beardslee as inventors, and designated by Attorney Docket No. VRD-1001-PV3. This patent application also is related to U.S. patent application no. 13/245,777 filed on September 26, 2011, entitled BIOLOGICAL METHODS FOR PREPARING ADIPIC ACID, naming Stephen Picataggio and Tom Beardslee as inventors, and designated by Attorney Docket No. VRD-1001-CT. This patent application also is related to U.S. patent application no. 13/245,780 filed on September 26, 2011, entitled BIOLOGICAL METHODS FOR PREPARING ADIPIC ACID, naming Stephen Picataggio and Tom Beardslee as inventors, and designated by Attorney Docket No. VRD-1001-UT. This patent application also is related to U.S. patent application no. 13/245,782 filed on September 26, 2011, entitled BIOLOGICAL METHODS FOR PREPARING ADIPIC ACID,

naming Stephen Picataggio and Tom Beardslee as inventors, and designated by Attorney Docket No. VRD-1001-UT2. This patent application also is related to International patent application no. PCT/US2012/020230 filed on January 4, 2012, entitled BIOLOGICAL METHODS FOR PREPARING ADIPIC ACID, naming Stephen Picataggio and Tom Beardslee as inventors, and designated by Attorney Docket No. VRD-1001-PC2. This patent application also is related to International patent application no. PCT/US2012/056562 filed on September 21, 2012, entitled BIOLOGICAL METHODS FOR PREPARING ADIPIC ACID, naming Stephen Picataggio and Tom Beardslee as inventors, and designated by Attorney Docket No. VRD-1001-PC3. This patent application also is related to U.S. provisional patent application no. 62/136,350 filed on March 20, 2015, entitled BIOLOGICAL METHODS FOR PREPARING 3-HYDROXYPROPIONIC ACID, naming Eric Michael Knight as inventor, and designated by Attorney Docket No. VRD-2001-PV. This patent application also is related to International patent application no. PCT/US2016/023243 filed on March 18, 2016, entitled BIOLOGICAL METHODS FOR PREPARING 3-HYDROXYPROPIONIC ACID, naming Eric Michael Knight as inventor. This patent application also is related to U.S. provisional patent application no. 61/505,092 filed on July 6, 2011, entitled BIOLOGICAL METHODS FOR PREPARING SEBACIC ACID naming Stephen Picataggio and Tom Beardslee as inventors, and designated by Attorney Docket No. VRD-1005-PV. This patent application also is related to U.S. provisional patent application no. 61/523,216 filed August 12, 2011, entitled BIOLOGICAL METHODS FOR PREPARING DODECANEDIOIC ACID naming Stephen Picataggio and Tom Beardslee as inventors, and designated by Attorney Docket No. VRD-1006-PV. This patent application also is related to International patent application no. PCT/US2012/045615 filed on July 5, 2012, entitled BIOLOGICAL METHODS FOR PREPARING A FATTY DICARBOXYLIC ACID, naming Tom Beardslee, Stephen Picataggio, L. Dudley Eirich and Jose Miguel Laplaza as inventors, and designated by Attorney Docket No. VRD-1005-PC. This patent application also is related to International patent application no. PCT/US2012/045622 filed on July 5, 2012, entitled BIOLOGICAL METHODS FOR PREPARING A FATTY DICARBOXYLIC ACID, naming Tom Beardslee, Stephen Picataggio, Alex Hutagalung and Tom Fahland as inventors, and designated by Attorney Docket No. VRD-1006-PC. This patent application also is related to U.S. patent application no. 14/131,170 filed on April 14, 2014 entitled BIOLOGICAL METHODS FOR PREPARING A FATTY DICARBOXYLIC ACID, naming Tom Beardslee, Stephen Picataggio, L. Dudley Eirich and Jose Miguel Laplaza as inventors. This patent application also is related to U.S. patent application no. 14/131,174 filed on April 28, 2014, entitled BIOLOGICAL METHODS FOR PREPARING A FATTY DICARBOXYLIC ACID, naming Tom Beardslee,

Stephen Picataggio, Alex Hutagalung and Tom Fahland as inventors. This patent application also is related to U.S. provisional patent application no. 61/739,656 filed December 19, 2012, entitled BIOLOGICAL METHODS FOR PREPARING A FATTY DICARBOXYLIC ACID naming Jose Laplaza, Tom Beardslee, Dudley Eirich and Stephen Picataggio as inventors, and designated by Attorney Docket No. VRD-1007-PV. This patent application also is related to U.S. provisional patent application no. 61/739,661 filed December 19, 2012, entitled BIOLOGICAL METHODS FOR PREPARING A FATTY DICARBOXYLIC ACID naming Tom Beardslee, Alex Hutagalung and Stephen Picataggio as inventors, and designated by Attorney Docket No. VRD-1008-PV. This patent application also is related to International patent application no. PCT/US2013/076664 filed on December 19, 2013, entitled BIOLOGICAL METHODS FOR PREPARING A FATTY DICARBOXYLIC ACID, naming Jose Laplaza, Tom Beardslee, Dudley Eirich and Stephen Picataggio as inventors. This patent application also is related to International patent application no. PCT/US2013/076739 filed on December 19, 2013, entitled BIOLOGICAL METHODS FOR PREPARING A FATTY DICARBOXYLIC ACID, naming Tom Beardslee, Alex Hutagalung and Stephen Picataggio as inventors. This patent application also is related to U.S. patent application no. 14/654,442 filed on June 19, 2015 entitled BIOLOGICAL METHODS FOR PREPARING A FATTY DICARBOXYLIC ACID, naming Jose Laplaza, Tom Beardslee, Dudley Eirich and Stephen Picataggio as inventors. This patent application also is related to U.S. patent application no. 14/654,458 filed on June 19, 2015, entitled BIOLOGICAL METHODS FOR PREPARING A FATTY DICARBOXYLIC ACID, naming Tom Beardslee, Alex Hutagalung and Stephen Picataggio as inventors. This patent application is also related to U.S. provisional patent application no. 62/011,500 filed on June 12, 2014 entitled PURIFICATION OF POLYCARBOXYLIC ACIDS, naming Jose Laplaza as inventor. This patent application is also related to U.S. patent application no. 14/738,600 filed on June 12, 2015 entitled PURIFICATION OF POLYCARBOXYLIC ACIDS, naming Jose Laplaza, William Andrew Evanko and Jason H. Radany as inventors. This patent application also is related to International patent application no. PCT/US2015/035634 filed June 12, 2015 entitled PURIFICATION OF POLYCARBOXYLIC ACIDS, naming Jose Laplaza as inventor.

The entire content of each of the foregoing patent applications is incorporated herein by reference, including, without limitation, all text, tables and drawings.

Field

The technology relates in part to biological methods for producing terpenes and to engineered cells and microorganisms capable of such production.

Background

Cells and microorganisms employ various enzyme-driven biological pathways to support metabolism and growth. A cell synthesizes native proteins, including enzymes, in vivo based on the sequence of deoxyribonucleic acid (DNA) encoding the protein. DNA first is transcribed into a complementary ribonucleic acid (RNA) that contains a ribonucleotide sequence encoding the protein. RNA then directs translation of the encoded protein by interaction with various cellular components, such as ribosomes. When the resulting protein is an enzyme, it can participate as a biological catalyst in biochemical pathways involved in producing a variety of organic molecules by the cell or organism.

These pathways can be exploited for the harvesting of naturally produced organic molecules. The pathways also can be altered to increase production or to produce specific molecules that may be commercially valuable. Advances in recombinant molecular biology methodology allow researchers to isolate DNA from one cell or organism and insert it into another cell or organism, thus altering the cellular synthesis of enzymes or other proteins. Advances in recombinant molecular biology methodology also allow endogenous genes, carried in the genomic DNA of a cell or microorganism, to be increased in copy number, thus altering the cellular synthesis of enzymes or other proteins. Such genetic engineering can change the biological pathways within the host cell or organism, causing it to produce a desired product. Microorganic industrial production can minimize the use of caustic chemicals and the production of toxic byproducts, thus providing a “clean” source for certain compounds. The use of appropriate plant-derived feedstocks allows production of “green” compounds while further minimizing the need for and use of petroleum-derived compounds.

Summary

Provided herein in certain aspects are genetically modified microorganisms comprising one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides,

where expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide. Also provided herein in certain aspects are genetically modified microorganisms comprising one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides, and a genetic modification that alters the expression of a polypeptide providing for transport of acetyl-carnitine in the microorganisms. Also provided herein in certain aspects are methods for producing a terpene comprising contacting a genetically modified microorganism provided herein with a feedstock comprising a carbon source, and culturing the microorganism under conditions in which the terpenes are produced from the feedstock.

Also provided herein in certain aspects is a genetically modified *Candida viswanathii* yeast, comprising one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides. Also provided herein in certain aspects are methods for producing a terpene comprising contacting a genetically modified *Candida viswanathii* yeast provided herein with a feedstock comprising a carbon source, and culturing the microorganism under conditions in which the terpenes are produced from the feedstock.

Also provided in certain aspects is a genetically modified yeast comprising one or more heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase. Also provided in certain aspects is a genetically modified yeast comprising one or more heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase. Also provided in certain aspects is a genetically modified yeast comprising one or more heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, and cytochrome p450 reductase.

Certain embodiments are described further in the following description, examples, claims and drawings.

Incorporation by Reference

All publications, patents and patent applications, GENBANK sequences (e.g., available at the World Wide Web Uniform Resource Locator (URL) ncbi.nlm.nih.gov of the National Center for

Biotechnology Information (NCBI), sequences available through other databases, and websites and other published materials referred to throughout the entire disclosure herein, unless noted otherwise, are incorporated by reference in their entirety. Citation of any publications, patents and patent applications, GENBANK (and other database) sequences, websites and other published materials herein is not an admission that any of the foregoing is pertinent prior art, nor does it constitute any admission as to the contents or date of these publications or documents.

Brief Description of the Drawings

The drawings illustrate certain embodiments of the technology and are not limiting. For clarity and ease of illustration, the drawings are not made to scale and, in some instances, various aspects may be shown exaggerated or enlarged to facilitate an understanding of particular embodiments.

Fig. 1 is a diagrammatic representation of a cassette for the addition of a gene of interest (GOI) into a host non-functional ura3 locus using the single crossover integration method. The core of the cassette contains the GOI gene with a promoter (P_{XXX}) and terminator (T_{XXX}) for controlling transcription of the GOI gene. The *URA3* gene selectable marker is split with a 3' portion of the gene at one end of the cassette and a 5' portion positioned at the other end of the cassette. The segment of the expression cassette containing the gene of interest (GOI) is positioned between the *URA3* promoter (P_{URA3}) and terminator (T_{URA3}). Parts (A) and (B) of Fig. 1 show results of integration of one copy (A) and two copies (B) of the cassette. Integration of one cassette generates an added, functional GOI expression unit and may or may not provide for expression of a functional Ura3p, depending on the nature of the ura3 locus and the location of the split in the *URA3* selectable marker. Integration of two copies of the cassette generates a complete, functional *URA3* sequence by combining the 5' end of *URA3* from one copy of the cassette and the 3' end of *URA3* from the second copy of the cassette. Additional copies may also be integrated. Transformants are selected for by growth on uracil-free media. This integration method thus favors selection of transformants containing multiple copies of the GOI.

Fig. 2 is a diagrammatic illustration of an exemplary gene cassette for use in a double crossover homologous recombination integration referred to as "knock out" mutagenesis. Two slightly different cassettes are depicted for use in separately disrupting each of the two *FAT1* alleles in a diploid yeast. The two cassette-containing nucleic acid segments are referred to as "Deletion

1" and "Deletion 2," respectively. Each cassette contains a *URA3* gene including a URA3 promoter (P_{URA3}) and terminator (T_{URA3}). Additionally, each cassette contains a repeat of the P_{URA3} sequence immediately downstream of the terminator sequence. The two separate deletion cassette-containing fragments differ in the sequences of the target gene that they contain on each side of the *URA3* cassette.

Fig. 3A, Fig. 3B, and Fig. 3C show diagrammatic illustrations of a "knock in" gene disruption method which disrupts one target gene ("GOI1") and also adds a desired gene of interest ("GOI2") at the disrupted locus. The basic URA3 disruption cassette is the same as that described in Fig. 2, except for an additional expression cassette immediately downstream of the second P_{URA3} repeat sequence. This expression cassette contains the gene of interest, GOI2, for adding to the endogenous GOI1 locus and includes a promoter (P_{XXX}) and terminator (T_{XXX}) for controlling transcription of GOI2. Immediately upstream of the first P_{URA3} sequence is a sequence of nucleotides of the GOI1 gene, and immediately downstream of the terminator (T_{XXX}) for GOI2 is another sequence of the GOI1 gene. These sequences are for use in integration of the cassettes into the GOI1 locus. Fig. 3B shows the locus after the integration of the knock-in cassette which incorporates the functional URA3 selection gene. To remove the *URA3* gene, transformants are grown in the presence of 5-FOA to facilitate a "loop-out" event that is driven by the direct repeat sequences on either side of the *URA3* gene (in this case P_{URA3}). The result of that event is shown in Fig. 3C which depicts the P_{URA3} sequence that remains followed by the functional GOI2 cassette.

Fig. 4 is a schematic illustration of the general, unmodified flow of carbon from a fatty acid carbon source in a wild-type eukaryotic cell, such as, for example, a yeast cell. "FA": fatty acid; "Co-A": coenzyme A; "PL": phospholipid; "TAG": triacylglyceride; FA-CoA": fatty acyl-CoA "*FAA1*" and "*FAT1*": acyl-CoA synthetase genes; "*PEX11*": peroxisomal biogenesis factor gene; "*PXA1*": peroxisomal transport protein gene; "β-Ox": β-oxidation; "Ac-CoA": acetyl-CoA; "*CAT2*": carnitine acetyltransferase gene; "Carn": carnitine; "Ac-Carn": acetyl-carnitine; "*CRC*1": mitochondrial acetyl-carnitine transport protein; "Cit": citrate; "TCA": tricarboxylic acid cycle; "Iso-Cit": isocitrate; "*ICL1*": isocitrate lyase gene; "Succ": succinate; "*MLS1*": malate synthase gene; "Glx": glyoxylate; "Mal": malate.

Fig. 5 is a schematic illustration of an engineered carbon flux pathway of a modified cell for use in producing a target molecule. The figure depicts cellular modifications in some embodiments

of a eukaryotic (i.e., yeast in this example) platform system for developing particular target molecule production systems. The platform system contains an acetyl group carbon recycle loop that diverts acetyl moieties generated in the breakdown of fatty acids in peroxisomal β-oxidation("β-ox") into cytosolic fatty acid synthesis to regenerate a fatty acid that can be subjected to another cycle of peroxisomal β-oxidation. The recycle loop is depicted by the dark, solid reaction arrows beginning with extracellular fatty acid ("FA") internalization in the upper left corner of the figure. Free fatty acids that have entered the cell can undergo oxidation to dicarboxylic acids (DCA) through ω-oxidation ("ω-ox"). Multiple modifications introduced via genetic manipulation, as well as unmodified activities of the cell, are indicated as follows: acyl-CoA synthetase gene deletions shown as "faa1Δ" and "fat1Δ" and resulting disruption of cytosolic activation of fatty acids (indicated as a lightly shaded dotted line reaction arrow below the gene deletion symbols and extending from "FA" to "FA-CoA") and diminished entry of FA-CoA into lipid (triacylglycerides ("TAG") and phospholipids ("PL")) biosynthesis; endogenous, unmodified peroxisomal enzymes acyl-CoA synthetase ("FAA2") and thioesterase ("TES"); unmodified glyoxylate cycle ("GlyOx") showing endogenous isocitrate lyase enzyme ("ICL1"); unmodified endogenous peroxisomal carnitine acetyltransferase ("CAT2") for conversion of acetyl-CoA ("Ac-CoA") to acetyl-carnitine ("AC-Carn"); modified (indicated by diagonal hatch lines) cytosolic carnitine acetyltransferase ("$CAT2^{cyt}$") and acetyl-CoA carboxylase ("ACC1") enzymes; unmodified endogenous fatty acid synthase enzyme complex ("FAS"); modified and added cytosolic thioesterase enzyme ("TES^{cyt}") showing added activity as solid, dark reaction arrow extending from FA-CoA to FA (which represents the final segment of the recycle loop); modified (gene deletion) peroxisomal transport protein ("pxa1Δ") showing disrupted (lightly shaded dotted line reaction arrow) acyl-CoA ("FA-CoA") import activity; modified peroxisomal biogenesis factor ("PEX11") activity; modified (promoter replacement) mitochondrial acetyl-carnitine transport protein ("CRC1") showing diminished (lightly shaded dotted line reaction arrow) acetyl-carnitine import activity; modified mitochondrial carnitine acetyltransferase ("CAT2") activity showing decreased conversion (lightly shaded dotted line reaction arrow) of AC-Carn to AC-CoA; unmodified mitochondrial tricarboxylic acid cycle ("TCA"); lightly shaded dashed lines reflect unmodified cellular activities that are not part of the carbon recycle loop shown in dark, solid lines. The details of the modifications in this exemplary engineered platform system are provided in the Detailed Description that follows.

Fig. 6 is a schematic illustration of an engineered carbon flux pathway of a modified cell for use in producing a target molecule. The figure depicts cellular modifications in some embodiments

of a eukaryotic (i.e., yeast in this example) platform system for developing particular target molecule production systems. The platform system is similar to that shown in Fig. 5 except for the following: modified (gene deletion) peroxisomal carnitine acetyltransferase ("cat2Δ") showing disrupted (lightly shaded dotted line reaction arrow) generation of peroxisomal acetyl-carnitine (AC-Carn shown with an "X" over it); modified and added peroxisomal acetyl-CoA hydrolase for converting acetyl-CoA (AC-CoA) to acetate ("Ac"); modified (promoter replacement) cytosolic acetyl-CoA synthetase ("ACS") activity. The details of the modifications in this exemplary engineered platform system are provided in the Detailed Description that follows.

Fig. 7 is a schematic illustration of an engineered carbon flux pathway of a modified cell for use in producing a target molecule. The figure depicts cellular modifications in some embodiments of a eukaryotic (i.e., yeast in this example) platform system for developing particular target molecule production systems. The carbon recycle loop in this platform system, depicted by the dark, solid reaction arrows, extends through mitochondrial metabolism and differs from that shown in Figs. 5 and 6. The mitochondrial acetyl-carnitine transporter ("CRC1") and carnitine acetytransferase ("CAT2") are unmodified in this exemplary platform system. A cytosolic ATP citrate lyase ("ACL1/2") activity is added to the system. The details of the modifications in this exemplary engineered platform system are provided in the Detailed Description that follows.

Fig. 8 is a schematic illustration of an engineered carbon flux pathway of a modified cell for use in producing a target molecule. The figure depicts cellular modifications in some embodiments of a eukaryotic (i.e., yeast in this example) platform system for for the enhanced production of malonyl-CoA and various target molecules that can be synthesized using malonyl-CoA as a precursor. The details of the modifications in this exemplary engineered platform system are provided in the Detailed Description that follows.

Fig. 9 is a schematic illustration of an example of an engineered production pathway for cell- or microbial-based synthesis of 3-hydroxypropionic acid ("3HP"). Added cytosolic malonyl-CoA reductase ("MCR") activity and modified 3-hydroxy-propionate-dehydrogenase ("HPD1") activities for 3HP synthesis are shown as well as modified (gene deleted) endogenous semialdehyde dehydrogenase ("ald6Δ") activity. The details of the modifications in this exemplary engineered platform system are provided in the Detailed Description that follows.

Fig. 10 is a schematic illustration of an example of an engineered production pathway for cell- or microbial-based synthesis of triacetic acid lactone ("TAL"). Added 2-pyrone synthase ("2PS") activity for TAL synthesis is shown. The details of the modifications in this exemplary engineered platform system are provided in the Detailed Description that follows.

Fig. 11 is a schematic illustration depicting cellular modifications in some embodiments of a eukaryotic (i.e., yeast in this example) platform system for the enhanced generation of acetyl-CoA and the production of a diverse array of target molecules (e.g., terpenes). In one aspect, Fig. 11 differs from Fig. 8 in that it shows an embodiment of the platform system in which target molecule production pathways extend from acetyl-CoA, instead of malonyl-CoA , as a precursor molecule. ("Mev" refers to the mevalonate pathway; "IPP" refers to isopentenyl diphosphate; "DMAPP" refers to dimethylallyl diphosphate.) The details of the modifications in this exemplary engineered platform system are provided in the Detailed Description that follows.

Fig. 12 is a restriction endonuclease site map of plasmid pAA061 showing the relative placement of the following nucleic acid sequences: *Candida* strain ATCC 20336 orotidine-5'-phosphate decarboxylase (*URA3*) gene promoter (Prom), open-reading frame and terminator (Term); β-lactamase (ampicillin-resistance) gene promoter (P(BLA)) and ORF (AP^r); and the *Escherichia coli* origin of replication (ORI). Also shown are the *Candida* strain ATCC 20336 phosphoglycerate kinase (*PGK*) gene promoter and terminator that were added to pAA061 to form pAA105.

Fig. 13 is a restriction endonuclease site map of plasmid pAA105 which was constructed by ligating the *Candida* strain ATCC 20336 phosphoglycerate kinase (*PGK*) gene promoter and terminator with the PstI/NdeI fragment of pAA061 (Fig. 12).

Fig. 14 is a restriction endonuclease site map of plasmid pAA219 which was constructed by inserting the *Candida* strain ATCC 20336 cytochrome P450 reductase (*CPRB*) ORF between the *PGK* gene promoter and terminator in pAA105 (Fig. 13).

Fig. 15 is a restriction endonuclease site map of a PstI/NdeI fragment of plasmid pAA073 which contains the *Candida* strain ATCC 20336 acyl-CoA oxidase (*POX4*) gene promoter and terminator with restriction sites between them for incorporating ORFs to be controlled by the inducible *POX4* promoter.

Fig. 16 is a restriction endonuclease site map of a fragment of plasmid pAA073. Plasmid pAA073 was constructed by ligating the PstI/NdeI fragment shown in Fig. 15 with the PstI/NdeI fragment of pAA061 (Fig. 12).

Fig. 17 is a restriction endonuclease site map of plasmid pAA153 which was constructed by inserting the *Candida* strain ATCC 20336 cytochrome P450 monooxygenase (*CYP52A14*) ORF between the *POX4* gene promoter and terminator in pAA073.

Fig. 18 is a diagrammatic representation of pAA153 (Fig. 17) linearized by endonuclease cutting of the plasmid at the *Cla*I site to yield a cassette for use in the addition of a *Candida* strain ATCC 20336 *CYP52A14* gene into a host non-functional ura3 locus using the single crossover integration method. The core of the cassette contains the *CYP52A14* gene with a *POX4* promoter and terminator for controlling transcription of the gene. Cutting of the plasmid at the *Cla*I site splits the *URA3* selectable marker and yields a linear DNA fragment with the *CYP52A14* gene expression cassette positioned between the *URA3* promoter (URA3 Prom) and terminator (URA3 Term).

Fig. 19 is a diagrammatic representation of plasmid pAA367 generated by (1) PCR amplification of two separate fragments of pAA153 (Fig. 17), one fragment containing a 3' URA3 sequence and the URA3 terminator and another fragment containing a *CYP52A14* gene expression cassette with a *POX4* promoter and terminator followed by the URA3 promoter and a 5' URA3 sequence, (2) joining of the two amplicons by overlap extension PCR to generate a single amplified fragment and (3) cloning of the single fragment into pCR-BluntII-TOPO.

Fig. 20 is a diagrammatic representation of a linear DNA expression cassette obtained by amplification from pAA367 (Fig. 19) that does not contain nucleic acid encoding an antibiotic selection marker (i.e., antibiotic-free).

Fig. 21A, Fig. 21B, and Fig. 21C show diagrammatic illustrations of a "knock out" gene disruption method which disrupts a target gene ("GOI"). Fig. 21A shows a double-crossover gene knock-out cassette for knocking out the function of a GOI containing a *URA3* selectable marker gene (including the gene promoter (P_{URA3}) and terminator (T_{URA3})) between 5' and 3' homologous sequences for the GOI. The *URA3* selectable marker also has DNA sequence

direct repeats (P_{URA3}) at the beginning and at the end of the gene sequence. After transformation of the double-crossover gene knock-out cassette into a Ura$^-$ mutant, the *URA3* marker allows selection on SC-URA plates for colonies that have integrated the construct (Fig. 21B) disrupting the GOI and generating a Ura$^+$ phenotype. Subsequent growth of Ura$^+$ transformants on 5-fluoroorotic acid (5-FOA) yields Ura$^-$ cells resulting from removal of the *URA3* selectable marker from the genome by a second crossover homologous recombination between the DNA sequence direct repeats (P_{URA3}) (Fig. 21C). A DNA sequence direct repeat remains in the genome as a "scar" left behind at the gene knock out site. The *URA3* selection marker may now be used again for further genetic modifications.

Fig. 22 is a diagrammatic depiction of how the knock out gene disruption method illustrated in Figs. 21A-21C, which regenerates an auxotrophic (Ura$^-$) cell after the second homologous recombination event, enables the same *URA3*-based selection method to be used repeatedly on the same cell, for example, such as in the disruption of the second allele ("Deletion 2") of a gene of interest (GOI) following the disruption of the first allele.

Fig. 23 is a restriction endonuclease site map of plasmid pAA298 containing a double-crossover gene knock-out cassette for knocking out the function of a *FAT1* gene. As shown in the figure, the double-crossover gene knock-out cassette includes a *URA3* selectable marker gene (including the gene promoter and terminator) between 5' and 3' homologous sequences ("N-Fat1" and "C-Fat1," respectively) for the *FAT1* gene. The plasmid also contains elements from pCR-BluntII-TOPO.

Fig. 24 shows restriction endonuclease site maps of plasmids pAA1519 and pAA1520 each containing a double-crossover gene knock-out cassette for knocking out the function of a *CAT2* gene. Each *CAT2* gene deletion cassette includes a 5' *Candida viswanathii* strain ATCC 20336 *CAT2* DNA fragment ("Cv CAT2 5' homology"), a 3' *Candida viswanathii* strain ATCC 20336 *CAT2* DNA fragment ("Cv CAT2 3' homology") and a *Candida viswanathii* strain ATCC 20336 *URA3* gene fragment containing a *URA3* ORF ("Cv URA3"), *URA3* promoter and a *URA3* terminator followed by a repeat of the promoter. The gene knock-out cassettes are contained within pCR-BluntII-TOPO.

Fig. 25A and Fig. 25B show a comparison of the N-terminal (Fig. 25A) and C-terminal (Fig. 25B) amino acid sequences of an unmodified *Candida* strain ATCC 20336 carnitine acetyltransferase

("Cv-CAT2 from pAA426") protein and of modified carnitine acetyltransferase proteins lacking one or both of the N-terminal mitochondrial targeting sequence (mts) and the C-terminal peroxisomal targeting sequence (pts). ("Cv-CAT2(-mts)" refers to the protein lacking only an N-terminal mitochondrial targeting sequence; "Cv-CAT2(-pts)" refers to the protein lacking only a C-terminal peroxisomal targeting sequence; "Cv-CAT2(-mts-pts)" refers to the protein lacking the N-terminal mitochondrial targeting sequence and the C-terminal peroxisomal targeting sequence.)

Fig. 26 is a map of plasmid pAA1164 containing all the elements of the pCR-BluntII-TOPO vector, two separate portions of a URA3 gene selectable marker originally cloned from *Candida* strain ATCC 20336, nucleic acid encoding a modified *Candida* strain ATCC 20336 Pox5p (i.e., Pox5(F98G)) and the *HDE* gene promoter and *POX4* gene terminator, both from *Candida* strain ATCC 20336. Also shown are the locations of sequences corresponding to oligonucleotides oAA4722 and oAA4723 which were used as primers in PCR amplification of the larger fragment sequence between these two sites. The amplified DNA fragment was used in the construction of pAA1610 (see details provided in the Examples herein). The "CvURA3" segment positioned following the 3' *URA3* fragment ("CvURA3 F23 to Stop") corresponds to the URA3 terminator, whereas the "CvURA3" segment positioned in front of the 5' *URA3* fragment ("CvURA3 M1 to L22") corresponds to the URA3 promoter.

Fig. 27 shows a comparison of the N-terminal amino acid sequences of an unmodified *Candida* strain ATCC 20336 cytosolic carnitine acetyltransferase ("Cv-Yat1p") protein and of modified Yat1p carnitine acetyltransferase proteins containing an added N-terminal mitochondrial targeting sequence (mts) in place of the initiating methionine of the unmodified Yat1p. ("Cv-Yat1p+CAT2mts" refers to the protein with an added N-terminal mitochondrial targeting sequence from the *Candida* strain ATCC 20336 mitochondrial carnitine acetyltransferase; "Cv-Yat1p+CIT1mts" refers to the protein with an added N-terminal mitochondrial targeting sequence from the *Candida* strain ATCC 20336 citrate synthase; "Cv-Yat1p+COX4mts" refers to the protein with an added N-terminal mitochondrial targeting sequence from the *Candida* strain ATCC 20336 cytochrome c oxidase.)

Fig. 28 is a restriction endonuclease site map of plasmid pAA245 which contains all the elements of the pCR-BluntII-TOPO vector, DNA encoding a *Candida* strain ATCC 20336 acetyl-CoA carboxylase ("CvACC1") enzyme and a 5' partial intron ("CvACC1 5' Intron partial") at the

5' end of the ACC1 ORF. Also shown are the locations of sequences corresponding to oligonucleotides oAA0784 and oAA0785 which can be used as primers in PCR amplification of the Acc1-encoding DNA fragment from genomic DNA.

Fig. 29 is a restriction endonuclease site map of plasmid pAA326 which was generated by cutting pAA245 (Fig. 28) with BspQI and ligating the resulting *ACC1* gene fragment including 5' partial intron into BspQI-cut plasmid pAA105 (Fig.13) to put the gene under the control of the *PGK* promoter and terminator from *Candida* strain ATCC 20336.

Fig. 30 is a restriction endonuclease site map of plasmid pAA1634 generated by ligating a SpeI/XbaI fragment of pAA326 (Fig. 29) containing DNA encoding amino acids R643 to the STOP codon of Candida strain ATCC 20336 *ACC1* with SpeI/XbaI-digested pAA601 (Fig. 12). A series of site-directed mutagenesis reactions was performed on pAA1634 to introduce mutations into the truncated *ACC1* coding sequence as described in the Examples herein. Also shown in the figure is an illustration of two DNA fragments, one containing an *HDE* gene promoter and one containing DNA encoding amino acids M1 – S642 of the Acc1p (both from *Candida* strain ATCC 20336) that were subsequently ligated with pAA1634 to generate plasmids containing DNA encoding full-length mutant Acc1p.

Fig. 31 is a map of plasmid pAA2247 which was generated by ligating a SbfI/MluI fragment of plasmid pAA1908 containing DNA encoding a *Candida* strain ATCC 20336 Acc1p mutant (S1158A) surrounded by the *HDE* gene promoter ("Prom") and *PGK* gene terminator ("Term") from *Candida* strain ATCC 20336 with SbfI/MluI-digested plasmid pAA2153. Plasmid pAA2153 contains DNA encoding a *Candida* strain ATCC 20336 *URA3* selectable marker with a direct repeat of the T_{URA3} sequence located just upstream of the *URA3* gene promoter sequence (P_{URA3}) to yield T_{URA3}-P_{URA3}-*URA3*-T_{URA3}. The *URA3* selectable marker in pAA2153 is placed between genomic DNA sequence elements ("IGR5 5' homology" and "IGR5 3' homology") from *Candida* strain ATCC 20336 which are named IGR5. The IGR5 homology regions target integration of the intervening DNA into genomic DNA by homologous recombination. Also shown are the locations of DNA corresponding to primers oAA7259 and oAA7260 which can be used to amplify pAA2247 to generate a linear DNA for transformation of host cells.

Fig. 32 shows maps of plasmids pAA1613 and pAA1701 which contain double-crossover *Candida* strain ATCC 20336 *CRC1* gene knock-out cassettes that can be used to disrupt

Candida CRC1 genes. The cassette in each plasmid contains a 5' *Candida* strain ATCC 20336 *CRC1* DNA fragment ("CRC1 5' homology"), a 3' *Candida* strain ATCC 20336 *CRC1* DNA fragment ("CRC1 3' homology") and a *Candida* strain ATCC 20336 *URA3* gene fragment containing a *URA3* promoter, *URA3* ORF, and a *URA3* terminator followed by a repeat of the *URA3* promoter. The cassettes in the two plasmids differ in the sizes of the *CRC1* gene homology regions and the orientation of the *URA3* selectable marker between the homology regions. The gene knock-out cassettes are contained within pCR-BluntII-TOPO. Also shown are the locations of DNA corresponding to primers oAA5511 and oAA5512 which can be used to PCR amplify linear cassettes from each plasmid to generate a linear DNA for transformation of host cells.

Fig. 33 is a map of plasmid pAA2214 which was generated by ligating a plasmid backbone containing the IGR5 homology regions and *URA3* selectable marker (with T_{URA3} repeat) amplified from plasmid pAA2247 (Fig. 31) with a 1,816-bp DNA fragment containing a *Candida* strain ATCC 20336 modified *CRC1* gene expression cassette. The expression cassette contains the glucose-6-phosphate isomerase (*G6PI*) promoter and *POX4* gene terminator from ATCC 20336 for controlling expression of the Crc1p. Also shown are the locations of DNA corresponding to primers oAA7259 and oAA7260 which can be used to PCR amplify a linear cassette from pAA2214 to generate a linear DNA for transformation of host cells.

Fig. 34 is a map of plasmid pAA2311 which contains a single-crossover cassette with nucleic acid encoding a *Candida* strain ATCC 20336 *CRC1* gene protein linked to a *Candida* strain ATCC 20336 *G6PI* low-expression promoter. The *CRC1* expression elements, P_{G6PI}-*CRC1*-T_{POX4}, were obtained as a DNA fragment amplified from plasmid pAA2214 (Fig. 33). This fragment was ligated with a fragment amplified from pAA1164 (Fig. 26) containing all the elements of the pCR-BluntII-TOPO vector, two separate portions of a URA3 selectable marker and a *POX4* gene terminator to yield pAA2311. Also shown are the locations of sequences corresponding to oligonucleotides oAA2206 and oAA2209 which can be used as primers in PCR amplification of a 3,307-bp linear, antibiotic-free DNA fragment for use in transforming host cells for expression of the Crc1p.

Fig. 35 is a map of plasmid pAA879 which contains a double-crossover knock-in cassette with nucleic acid encoding a *Candida* strain ATCC 20336 cytochrome P450 reductase (*CPRB*) gene protein. The plasmid contains all the elements of the pCR-BluntII-TOPO vector, two separate

portions ("FAA1 5' homology" and "CvFAA1 3' homology") of a *Candida* strain ATCC 20336 *FAA1* gene, elements for the expression of CprB protein ("POX4 Promoter," "CvCPRB" and "POX4 term") and a *URA3* selectable marker with P_{URA3} repeat (URA3 Prom-URA3-URA3 Term-URA3 Prom). Also shown are the locations of sequences corresponding to oligonucleotides oAA3557 and oAA3564 which can be used as primers in PCR amplification of a linear DNA fragment for use in transforming host cells for disruption of the *FAA1* gene and expression of CPRB protein.

Fig. 36 is a map of plasmid pAA208 which includes two separate portions ("POX4 5' homology" and "POX4 3' homology") of a *Candida* strain ATCC 20336 *POX4* gene and a *Candida* strain ATCC 20336 *URA3* gene selectable marker with P_{URA3} repeat (URA3 Prom-URA3-URA3 Term-URA3 Prom).

Fig. 37 is a map of plasmid pAA850 which contains a double-crossover knock-in cassette with nucleic acid encoding a *Candida* strain ATCC 20336 *PEX11* gene protein. The plasmid contains all the elements of the pCR-BluntII-TOPO vector, two separate portions ("POX4 5' homology" and "POX4 3' homology") of a *Candida* strain ATCC 20336 *POX4* gene, elements for the expression of Pex11 protein ("POX4 Prom," "PEX11" and "POX4 term") and a *URA3* selectable marker with P_{URA3} repeat (URA3 Prom-URA3-URA3 Term-URA3 Prom). Also shown are the locations of sequences corresponding to oligonucleotides oAA3355 and oAA3357 which can be used as primers in PCR amplification of a linear DNA fragment for use in transforming host cells for disruption of the *POX4* gene and expression of Pex11 protein.

Fig. 38 shows a partial amino acid sequence of *Candida* strain ATCC 20336 Pox5p acyl-CoA oxidase and the results of analysis of the amino acid sequence using HotSpot Wizard (a software program tool for identifying sites for engineering of substrate specificity and/or activity of enzymes using a combination of structural, functional and sequence analysis). HotSpot Wizard identified several amino acid positions, or "hotspots," of Pox5p to mutate, with each position given a score from 1 (cold) to 9 (hot). The different HotSpot residues identified are highlighted in the figure and shaded according to the score assigned to the residue.

Fig. 39 is an illustration of an overlap extension PCR-based method for generating nucleic acids encoding mutants (F98G and W429F) of a *Candida* strain ATCC 20336 Pox5 acyl-CoA oxidase. The oligonucleotides ("Oligos") used in the PCR amplifications are listed in the table shown in

the figure. Oligos B and C contain the desired point mutations. Sequences for each of the oligonucleotides are provided in the Examples herein.

Fig. 40 shows maps of plasmids pAA1117 and pAA1155 which contain *Candida* strain ATCC 20336 *PXA1* gene knock-out cassettes that can be used to disrupt alleles of *PXA1* genes. Both plasmids contain all the elements of the pCR-BluntII-TOPO vector, two separate portions ("PXA1 5' homology" and "PXA1 3' homology") of a *Candida* strain ATCC 20336 *PXA1* gene, and a *URA3* selectable marker with P_{URA3} repeat (URA3 Prom-URA3-URA3 Term-URA3 Prom). Also shown are the locations of sequences corresponding to oligonucleotides oAA2679/oAA2684 and oAA2914/oAA2919 which can be used as primers in PCR amplification of pAA1117 and pAA1155, respectively, to obtain a linear DNA fragment for use in transforming host cells for disruption of the *PXA1* gene. Plasmids pAA1117 and pAA1155 differ in the sizes of *PXA1* gene 5' and 3' homology sequences and the orientation of the *URA3* selectable marker nucleic acid sequence contained in the plasmids.

Fig. 41 is a flow diagram showing the parent-daughter relationship for exemplary engineered yeast strains that can be used in generating cells and organisms for use in target molecule platform and production systems described herein. Strains in bold type are Crc^- strains.

Fig. 42 shows a restriction endonuclease site map of plasmid pAA276 which contains a *Candida* strain ATCC 20336 *FAA1* gene knock-out cassette that can be used to disrupt alleles of *FAA1* genes. The plasmid contains all the elements of the pCR-BluntII-TOPO vector, two separate portions ("FAA1 N terminal" and "FAA1 C terminal") of a *Candida* strain ATCC 20336 *FAA1* gene, and a *URA3* selectable marker with URA3 Promoter repeat (Promoter URA3 -URA3 - Terminator URA3 – Promoter URA3).

Fig. 43 shows a restriction endonuclease site map of plasmid pAA918 which contains a *Candida* strain ATCC 20336 *POX5* gene knock-out cassette that can be used to disrupt alleles of *POX5* genes. The plasmid contains all the elements of the pCR-BluntII-TOPO vector, two separate portions ("5' FR POX5" and "3' FR POX5") of a *Candida* strain ATCC 20336 *POX5* gene, and a *URA3* selectable marker with URA3 terminator repeat (Tura3 – Pura3 - URA3 -Tura3).

Fig. 44 shows photographs of the agar plates obtained from spot growth assays of wild-type *Candida* strain ATCC 20336 ("CAT2/CAT2") and mutant strains as follows: *cat2-Δ1::*P_{URA3}*/cat2-*

Δ2::URA3 (sAA4594) and a Cat2$^-$ *Candida* strain (*cat2-Δ1::*P_{URA3}*/cat2-Δ2::*P_{URA3}) that had been transformed with either pAA1610 or pAA1876. The upper and lower photographs on the left side of the figure are control agar plates containing synthetic complete media with dextrose minus uracil ("SCD-URA"), and the upper and lower photographs on the right side of the figure are plates containing yeast nitrogen base without amino acids, plus phosphate and 2% oleic acid ("YNBP + 2% oleic acid"). Each row of "spots" corresponds to serial dilutions of cells of the strain designated to the right of each row (increasing dilutions from right-to-left for the control agar plates and from left-to-right for the plates containing YNBP + 2% oleic acid).

Fig. 45 shows photographs of the agar plates obtained from spot growth assays of wild-type *Candida* strain ATCC 20336 ("CAT2/CAT2") and mutant strains as follows: *cat2-Δ1::*P_{URA3}*/cat2-Δ2::URA3* (sAA4594) and a Cat2$^-$ *Candida* strain (*cat2-Δ1::*P_{URA3}*/cat2-Δ2::*P_{URA3}) that had been transformed with one or more of plasmids pAA1610, pAA1967, pAA1968 and pAA1969. The photograph on the left side of the figure is of control agar plates containing synthetic complete media with dextrose minus uracil ("SCD-URA"), and the photograph on the right side of the figure is of plates containing yeast nitrogen base without amino acids, plus phosphate and 2% oleic acid ("YNBP + 2% oleic acid"). Each row of "spots" corresponds to serial dilutions of cells of the strain designated to the right of each row (increasing dilutions from right-to-left for the control agar plates and from left-to-right for the plates containing YNBP + 2% oleic acid).

Fig. 46 shows photographs of the agar plates obtained from spot growth assays of wild-type *Candida* strain ATCC 20336 ("CAT2/CAT2") and mutant strains as follows: *cat2-Δ1::*P_{URA3}*/cat2-Δ2::URA3* (sAA4594) and a Cat2$^-$ *Candida* strain (*cat2-Δ1::*P_{URA3}*/cat2-Δ2::*P_{URA3}) that had been transformed with pAA1846, pAA1875 and one or more of plasmids pAA1967, pAA1968, pAA1969 and pAA1847. The photograph on the left side of the figure is of control agar plates containing synthetic complete media with dextrose minus uracil ("SCD-URA"), and the photograph on the right side of the figure is of plates containing yeast nitrogen base without amino acids, plus phosphate and 2% oleic acid ("YNBP + 2% oleic acid"). Each row of "spots" corresponds to serial dilutions of cells of the strain designated to the right of each row (increasing dilutions from right-to-left for the control agar plates and from left-to-right for the plates containing YNBP + 2% oleic acid).

Fig. 47 shows photographs of the agar plates obtained from spot growth assays of wild-type *Candida* strain ATCC 20336 and mutant strains as follows: *ura3/ura3 pox4a::ura3/pox4b::ura3*

faa1:: P_{URA3}/*faa1::*P_{URA3} *fat1-Δ1::*P_{URA3}/*fat1-Δ2::URA3* (sAA875), *ura3/ura3 pox4a::ura3/pox4b::ura3 faa1::*P_{URA3}/*faa1::*P_{URA3} *fat1-Δ1::*P_{URA3}/*fat1-Δ2::*P_{URA3} *crc1-Δ1::URA3/CRC1* (sAA4057), *ura3/ura3 pox4a::ura3/pox4b::ura3 faa1::*P_{URA3}/*faa1::*P_{URA3} *fat1-Δ1::*P_{URA3}/*fat1-Δ2::*P_{URA3} *crc1-Δ1::*P_{URA3}/*crc1-Δ2::URA3* (sAA4281), *ura3/ura3 crc1-Δ1::URA3/CRC1* (sAA4368), and *ura3/ura3 crc1-Δ1::*P_{URA3}/*crc1-Δ2::URA3* (sAA9398). Also shown are the agar plates obtained from spot growth assays of strains sAA5916, sAA5917 and sAA5918 generated by transforming strain sAA4377 (*ura3/ura3 pox4a::ura3/pox4b::ura3 faa1::*P_{URA3}/*faa1::*P_{URA3} *fat1-Δ1::*P_{URA3}/*fat1-Δ2::*P_{URA3} *crc1-Δ1::*P_{URA3}/*crc1-Δ2::*P_{URA3}) with a double-crossover integration cassette containing DNA encoding a *Candida* strain ATCC 20336 Crc1p linked to the *G6PI*, *PIC2* or *SUL2* promoter, respectively. The photographs on the left side of the figure are of control agar plates containing synthetic complete media with dextrose minus uracil ("SCD-URA"), and the photographs on the right side of the figure are of plates containing yeast nitrogen base without amino acids, plus phosphate and 2% oleic acid ("YNBP + 2% oleic acid"). Each row of "spots" corresponds to serial dilutions of cells of the strain designated to the right of each row (increasing dilutions from right-to-left for the control agar plates and from left-to-right for the plates containing YNBP + 2% oleic acid).

Fig. 48A shows GC-MS of valencene production from sAA7449 with its library match shown in Fig. 48B.

Fig. 49A shows GC-MS of valencene production from sAA7453 with its library match shown in Fig. 49B.

Fig. 50 shows the yeast upper mevalonate pathway. The enzymes acetyl-CoA C-acetyltransferase (*ERG10*), HMG-CoA synthase (*ERG13*), and HMG-CoA reductase (*HMG1*), are native to yeast described herein. The bracketed number under each enzyme refers to the IUBMB enzyme nomenclature for the enzyme.

Fig. 51 shows the yeast lower mevalonate pathway. The enzymes Mevalonate kinase (*ERG12*), Phosphomevalonate kinase (*ERG8*), Diphosphomevalonate decarboxylase (*MVD1*; *ERG19*), Isopentenyl diphosphate delta isomerase (*IDI1*), are native to yeast described herein. The bracketed number under each enzyme refers to the IUBMB enzyme nomenclature for the enzyme.

Fig. 52 shows a biosynthetic pathway to ergosterol or geranylgeranyl diphosphate. *ERG20*, *ERG9*, and *BTS1* are native to yeast described herein.

Fig. 53 shows carotenoid production from farnesyl pyrophosphate. *CrtE*, *CrtB*, *CrtI, CrtY, CrtZ, and CrtW* are heterologous genes.

Fig. 54 shows an alternative route to lutein, canthaxanthin, and astaxanthin from β-carotene by the cytochrome P450 hydroxylase *CrtS*. *CrtS* is a heterologous enzyme which requires the activity of *CrtR*, a heterologous cytochrome p450 reductase.

Fig. 55 shows a route from farnesyl pyrophosphate to valencene by TPS1, a heterologous valencene synthase.

Fig. 56 show an absorption spectra of extracted carotenoids from sAA001, sAA7443, sAA7444, and sAA7445. Strains sAA7443, aAA7444 and sAA7445 are generated from sAA001 plus CsCrtE, CsCrtB and CsCrtI each under a glyceraldehyde-3-phosphate dehydrogenase (GPD) promoter. Lycopene absorption maxima are expected at 443, 472, and 502 nm.

Fig. 57 shows an HPLC chromatogram of lycopene production in sAA001 and sAA8503.

Detailed Description

There are multiple cellular metabolic pathways that utilize carbon-containing molecules for varying purposes, such as, for example, biomass production, energy generation and growth. Different metabolic pathways can occur in distinct areas of the cell. For example, in eukaryotic microorganisms metabolic processes such as glycolysis, the pentose phosphate pathway and gluconeogenesis occur in the cytoplasm, whereas β-oxidation, the tricarboxylic acid (TCA) cycle and glyoxylate cycle are carried out, in whole or in part, in cellular organelles. The different metabolic pathways can be differentially utilized to maintain the basic functions required for survival of a microorganism under a variety of conditions (e.g., varying carbon source, nutrient, and oxygen availabilities). These features contribute to the ability of microorganisms to readily adapt to a variety of environmental conditions. The adaptability of microorganisms facilitates manipulation of microbial metabolic processes for the production of commercially useful materials (e.g., proteins, lipids, and organic acids).

There are a number of molecules that are valuable as final products, and/or as raw materials in generating a desired product, that incorporate carbon atoms resulting from catabolic and anabolic carbon-metabolism pathways in microorganisms. A desired molecule that can be produced in cells and microorganisms is referred to herein as a "target" molecule or product. Some of these molecules are generated in wild-type microorganisms, whereas others that are not produced in a native microorganism can be generated through modification of a wild-type organism. In either case, the goal in the development of cellular and microbial production systems is to maximize yield and efficiency and minimize loss for process economy. Biological cell- or organism-based systems for production of carbon-containing molecules, such as, for example, polymers of substituted or unsubstituted hydrocarbons, may not be optimally efficient due to loss of carbon atoms that are transferred into other, non-target-producing, metabolic processes in the cell. In order for a bioproduction system to be cost-efficient and economically viable, it generally should meet certain metrics of titer, rate and yield with minimal by-product formation.

In order to minimize carbon loss and increase process efficiency of biological cell-based production systems, provided herein are modified cells and organisms (e.g., microorganisms) in which the flux of cellular carbon has been altered relative to an unmodified cell or organism. In some aspects, the cells or organisms are modified to redirect carbon from entering one or more growth and/or energy production metabolic pathways so that it is available for use in other inherent and/or engineered production processes. Alteration of carbon flux facilitates engineering of the cells or organisms for enhanced production of desired target molecules, including, for example, organic acids, terpenes and precursor molecules that can be used in the production of industrial chemicals. As such, modified cells and microorganisms provided herein are useful as platform systems that can be used for enhanced production of many different desired target molecules (e.g., terpenes) either singly or multiply in co-production microbial systems. Also provided herein are methods of modifying cellular carbon flux and methods of generating cells or microorganisms in which carbon flux has been optimized for production of target molecules, e.g., carbon-containing compounds, terpenes. Further provided herein are methods of producing target molecules (e.g., terpenes) using cell-based or microbial biosynthesis systems, including, for example, modified cells or microorganisms in which the flux of cellular carbon has been altered relative to an unmodified microorganism. Also provided are

compositions, including, but not limited to, nucleic acids and chemical media and combinations, that can be used in the methods provided herein.

Terpenes

Provided herein are methods for producing terpenes and engineered cells and microorganisms capable of producing terpenes. Terpenes are compounds generally made from isoprene units $(C_5H_8)_n$. Isoprene, also known as 2-methyl-1,3-butadiene or 2-Methylbuta-1,3-diene, is an organic compound having the formula $CH_2=C(CH_3)-CH=CH_2$. Isoprene units may be assembled in various combinations to produce thousands of possible terpene compounds. Terpenes may include hemiterpenes (one isoprene unit), monoterpenes (two isoprene units), sesquiterpenes (three isoprene units), diterpenes (four isoprene units), sesterterpenes (five isoprene units), triterpenes (six isoprene units), sesquarterpenes (seven isoprene units), tetraterpenes (eight isoprene units), and polyterpenes (many isoprene units; more than eight isoprene units). Hemiterpenes may include, for example, isoprene, prenol, isovaleric acid, benzoate, and prenyl acetate. Monoterpenes may include, for example, α-ocimene, β-ocimene, α-myrcene, β-myrcene, geraniol, citronellal, citronellol, linalool, citral A, halomon, S-limonene, R-limonene, phellandrene, α-terpinene, menthol, S-carvone, R-carvone, safranal, terpineol, thymol, carvacrol, umbellulone, piperitone, pulegone, rose oxide, lactisole, sabinene, camphene, eucalyptol, thujene, thujone, pinene, nepetalactone, ascaridole, borneol, verbenone, camphor, ethyl fenchol, and cyclosantene. Sesquiterpenes may include, for example, farnesene, farnesol, nerolidol, zingiberene, humulene, bisabolane, bisabolene, elemol, caryophyllene, guaiol, vetivazulene, cadinene, caryophyllene, cuparane, laurene, laurane, oppositane, mutisianthol, thapsane, lepidozanes, chiloscyphane, pinguisanes, herbertanes, botrydial, ngaione, longifolene, copaene, patchoulol, norpatchoulenol, santalol, cortisol, progesterone, oestrogen, testosterone, and 11-ketotestosterone. Diterpenes may include, for example, phytol, cembrene A, retinol, retinal, labdane, abietic acid, ferruginol, tetrahydrocannabinol, forskolin, cafestol, and gibberellins. Sesterterpenes may include, for example, geranylfarnesol and amyrin. Triterpenes may include, for example, squalene, sapelenins, tangshenoside I, cholecalciferol, dihydrotachysterol, lanosterol, cholesterol, cycloartenol, lovastatin, cucurbitacin, 20-hydroxyecdysone, campesterol, stigmasterol, ergosterol, bile acids, betulinic acid, withaferin A, ginsenosides, eleutherosides, astragalosides, bacoside A, araloside A, ziziphin, and gymnemic acids. Tetraterpenes may include, for example, phytoene, phytofluene, lycopene, torulene, γ-carotene, δ-carotene, citranaxanthin, rubixanthin, α-carotene, β-carotene, astaxanthin,

cryptoxanthins, canthaxanthin, violaxanthin, antheraxanthin, zeaxanthin, lutein, diadinoxanthin, diatoxanthin, neoxanthin, fucoxanthins and flavoxanthins. Polyterpenes may include, for example, polypropylene, rubber and latex.

Terpenes may include molecules containing only carbon (C) and hydrogen (H) atoms and also may include molecules containing carbon (C) and hydrogen (H) atoms in addition to other atoms (e.g. oxygen (O)). Terpene molecules containing atoms in addition to carbon (C) and hydrogen (H) may be referred to as terpenoids. In certain instances, compounds referred to as terpenoids contain only carbon (C) and hydrogen (H). Use of the term terpene herein includes terpenoids. Terpene structures may be acyclic (no carbon rings), monocyclic (one carbon ring), bicyclic (two carbon rings), tricyclic (three carbon rings), tetracyclic (four carbon rings), pentacyclic (five carbon rings), hexacyclic (six carbon rings), heptacyclic (seven carbon rings), or octacyclic (eight carbon rings). Terpene structures may contain more than eight carbon rings.

Terpenes may include carotenoids. Carotenoids (also referred to as tetraterpenoids) generally are pigments that provide red, orange, and yellow pigmentation to certain organisms, and include certain 40-carbon molecules. Over 700 carotenoids are known, which include, for example, carotenes (e.g., γ-carotene, β-carotene, α-carotene, ε-carotene, δ-carotene), xanthophylls (e.g., lutein, zeaxanthin, neoxanthin, violaxanthin, flavoxanthin, and α- and β-cryptoxanthin), torulene, lycopene, lutein epoxide, hydroxyechinenone, phoenicoxanthin, astaxanthin, fucoxanthin, 19′-butanoyloxyfucoxanthin, 19'-hexanoyloxyfucoxanthin, diatoxanthin, diadinoxanthin 19-hexanoyloxyparacentrone 3-acetate, gyroxanthin, canthaxanthin, citranaxanthin, and apo-caroten-ester. About 40 to 50 carotenoids can be found in food, and about 6 carotenoids can be found in human serum. Carotenoids may be synthesized de novo by organisms including, for example, archae, bacteria, fungi, algae, plants, aphids & spider mites. Carotenoids have various uses including, for example, antioxidants, color attractants, protection of plants from sun damage, precursors to vitamins, inflammatory disease protection, eye health, bone growth, and immune function.

Provided herein are microorganisms genetically modified to produce terpenes. For example, a genetically modified microorganism may comprise one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides. Terpene biosynthesis polypeptides may include any polypeptide capable of producing a terpene molecule or a terpene precursor. Terpene biosynthesis polypeptides may include one or more enzymes. For example, terpene

biosynthesis polypeptides may include one or more of terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase, valencene synthase, and cytochrome p450 reductase. In some embodiments, a genetically modified microorganism comprises heterologous nucleic acids encoding one or more of phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase. In some embodiments, a genetically modified microorganism comprises heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase. In some embodiments, a genetically modified microorganism comprises heterologous nucleic acids encoding one or more of geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase. In some embodiments, a genetically modified microorganism comprises heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase. In some embodiments, a genetically modified microorganism comprises heterologous nucleic acids encoding one or more of geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, and phytoene synthase. In some embodiments, a genetically modified microorganism comprises heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, and phytoene synthase. In some embodiments, a genetically modified microorganism comprises a heterologous nucleic acid encoding valencene synthase.

In some embodiments, a genetically modified microorganism comprises heterologous nucleic acids encoding one or more of geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, and cytochrome p450 reductase. In some embodiments, a genetically modified microorganism comprises heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, and cytochrome p450 reductase. In some embodiments, a genetically modified microorganism comprises heterologous nucleic acids encoding one or more of geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, and cytochrome p450 reductase. In some embodiments, a genetically modified microorganism comprises heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, phytoene

synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, and cytochrome p450 reductase.

Terpene biosynthesis polypeptides may include any polypeptide (e.g., enzyme) capable of producing a terpene precursor. In certain embodiments, terpene biosynthesis polypeptides include one or more polypeptides (e.g., enzymes) in a mevalonate pathway (e.g., upper mevalonate pathway, lower mevalonate pathway). Typically, a mevalonate pathway begins with cytoplasmic acetyl-CoA, and the end product of a mevalonate pathway is isopentenyl diphosphate (IPP), which can serve as a monomer unit for terpene production. Polypeptides (e.g., enzymes) in a mevalonate pathway may include, for example, one or more of acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, isopentyl diphosphate delta isomerase, dimethylallyltranstransferase, and farnesyl diphosphate synthetase. In some embodiments, amounts and/or activities of one or more proteins chosen from acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, isopentyl diphosphate delta isomerase, dimethylallyltranstransferase, and farnesyl diphosphate synthetase are increased in a genetically modified microorganism provided herein. In some embodiments, a genetically modified microorganism comprises heterologous nucleic acids encoding one or more of acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, isopentyl diphosphate delta isomerase, dimethylallyltranstransferase, and farnesyl diphosphate synthetase.

In some embodiments, a heterologous nucleic acid encoding a terpene biosynthesis polypeptide is endogenously expressed in a microorganism that is different from the microorganism genetically modified to produce terpenes. In some embodiments, a heterologous nucleic acid encoding a terpene biosynthesis polypeptide is endogenously expressed in a microorganism that is of a different species, genus, family, order, class, phylum or kingdom than the microorganism genetically modified to produce terpenes. For example, a heterologous nucleic acid encoding a terpene biosynthesis polypeptide may be endogenously expressed in a microorganism chosen from *Cronobacter* spp, *Callitropsis* spp, *Xanthophyllomyces* spp, *Agrobacterium* spp, and *Pantoea* spp.

A heterologous nucleic acid encoding a terpene biosynthesis polypeptide may include an endogenous nucleic acid linked to a promoter or other regulatory element that is not normally functionally linked to the endogenous nucleic acid sequence. Promoters may include, for example, a mutated form of an endogenous promoter, an endogenous promoter replaced with a modified version of the endogenous promoter, and an endogenous promoter replaced with a promoter located elsewhere in the same genome or from a different genome. A modification may be performed by directed mutagenesis, and the mutated promoter may be referred to as "heterologous" because it is no longer native. In certain instances, a modification may be performed by replacing an endogenous promoter with a modified promoter, the latter of which would be considered heterologous because the modified promoter is not normally functionally linked to an endogenous nucleic acid.

In some embodiments, a heterologous nucleic acid encoding a terpene biosynthesis polypeptide is regulated according to an inducible system. For example, a heterologous nucleic acid encoding a terpene biosynthesis polypeptide may be regulated by a nucleic acid that provides for fatty acid induction of expression of the terpene biosynthesis polypeptide. The fatty acid that can induce expression of a terpene biosynthesis polypeptide may be a saturated fatty acid or an unsaturated fatty acid. For example, a fatty acid may be chosen from one or more of oleic acid, palmitoleic acid, erucic acid, linoleic acid, palmitic acid, caproic acid, enanthic acid, caprylic acid pelargonic acid, capric acid, undecylic acid, lauric acid, myristic acid, pentadecanoic acid, margaric acid, stearic acid arachidic acid, behenic acid, tridecylic acid, and linolenic acid.

In some embodiments, a nucleic acid that provides for fatty acid induction of expression of a terpene biosynthesis polypeptide is a promoter or a portion of a promoter. In some embodiments, a nucleic acid that provides for fatty acid induction of expression of a terpene biosynthesis polypeptide comprises a fatty acid response element. In some embodiments, a nucleic acid that provides for fatty acid induction of expression of a terpene biosynthesis polypeptide is a promoter, or a portion of a promoter, comprising a fatty acid response element. In some embodiments, a fatty acid response element comprises an oleic acid response element. In some embodiments, a nucleic acid that provides for fatty acid induction of expression of a terpene biosynthesis polypeptide comprises a promoter region of a heterologous gene. In some embodiments, a nucleic acid that provides for fatty acid induction of expression of a terpene biosynthesis polypeptide comprises a promoter region of an endogenous gene. For example, a promoter region may be a promoter region of a gene

encoding hydratase-dehydrogenase-epimerase (HDE), acyl co-A oxidase (POX; e.g., POX4), acyl co-A thiolase (POT), peroxin (PEX), peroxisomal protein POX18, or peroxisomal adenine nucleotide transporter protein (ANT1). In some embodiments, a nucleic acid that provides for fatty acid induction of expression of a terpene biosynthesis polypeptide comprises a promoter region of a gene encoding hydratase-dehydrogenase-epimerase (HDE).

In some embodiments, a nucleic acid that provides for fatty acid induction of expression of a terpene biosynthesis polypeptide comprises a promoter region of a gene from *Candida*. For example, a promoter region may be a promoter region of a gene encoding *Candida* hydratase-dehydrogenase-epimerase (HDE), *Candida* acyl co-A oxidase (POX; e.g., POX4), *Candida* acyl co-A thiolase (POT), *Candida* peroxin (PEX), peroxisomal protein POX18, or *Candida* peroxisomal adenine nucleotide transporter protein (ANT1). In some embodiments, a nucleic acid that provides for fatty acid induction of expression of a terpene biosynthesis polypeptide comprises a promoter region of a gene encoding *Candida* hydratase-dehydrogenase-epimerase (HDE).

In some embodiments, a heterologous nucleic acid encoding a terpene biosynthesis polypeptide is regulated by a nucleic acid that provides for alkane induction of expression of the terpene biosynthesis polypeptide. An alkane may be chosen from one or more of hexane, heptane, nonane, decane, undecane, dodecane, tridecane, tetradecane, pentadecane, hexadecane, heptadecane, and octadecane, for example. In some embodiments, a nucleic acid that provides for alkane induction of expression of a terpene biosynthesis polypeptide is a promoter or a portion of a promoter. In some embodiments, a nucleic acid that provides for alkane induction of expression of a terpene biosynthesis polypeptide comprises an alkane response element. In some embodiments, a nucleic acid that provides for alkane induction of expression of a terpene biosynthesis polypeptide is a promoter, or a portion of a promoter, comprising an alkane response element. In some embodiments, an alkane response element comprises an alkane response element 1 (ARE1) sequence or an ARE1-like sequence as described herein. In some embodiments, a nucleic acid that provides for alkane induction of expression of a terpene biosynthesis polypeptide comprises a promoter region, or a portion of a promoter region, of a heterologous gene.

In some embodiments, a heterologous nucleic acid encoding a terpene biosynthesis polypeptide is regulated by a nucleic acid that provides for glucose induction of expression of the terpene

biosynthesis polypeptide. In some embodiments, a nucleic acid that provides for glucose induction of expression of a terpene biosynthesis polypeptide comprises a promoter region of a gene encoding glyceraldehyde-3-phosphate dehydrogenase (GPD). In some embodiments, a nucleic acid that provides for glucose induction of expression of a terpene biosynthesis polypeptide comprises a promoter region of a gene from *Candida*. In some embodiments, a nucleic acid that provides for glucose induction of expression of a terpene biosynthesis polypeptide comprises a promoter region of a gene encoding *Candida* glyceraldehyde-3-phosphate dehydrogenase (GPD).

Modification of an acetyl-CoA C-acetyltransferase activity

An initial step in the upper mevalonate pathway is the conversion of 2 acetyl-CoA molecules to acetoacetyl-CoA and CoA. An acetyl-CoA C-acetyltransferase enzyme (also referred to as ERG10, acetoacetyl-CoA thiolase, acetyl-CoA acetyltransferase, ergosterol biosynthesis protein 10, beta-acetoacetyl coenzyme A thiolase, 2-methylacetoacetyl-CoA thiolase, 3-oxothiolase, acetyl coenzyme A thiolase, acetyl-CoA:N-acetyltransferase, thiolase II, LPB3, TSM0115; EC 2.3.1.9) is a cytosolic and/or peroxisomal enzyme that can transfer an acetyl group from one acetyl-CoA molecule to another, forming acetoacetyl-CoA. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of an acetyl-CoA C-acetyltransferase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of an acetyl-CoA C-acetyltransferase, may be modified to decrease the amount and/or activity of an acetyl-CoA C-acetyltransferase, or may be modified to alternately increase and decrease the amount and/or activity of an acetyl-CoA C-acetyltransferase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of an acetyl-CoA C-acetyltransferase in a cell is increased. Increasing the amount and/or activity of an acetyl-CoA C-acetyltransferase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., mevalonate production, isopentenyl diphosphate production, geranyl diphosphate production, farnesyl diphosphate production, geranylgeranyl diphosphate production, terpene production, carotenoid production, beta carotene production, lycopene production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, acetyl-CoA C-acetyltransferase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host acetyl-CoA C-acetyltransferase can be increased by increasing the number of copies of a nucleic acid encoding an acetyl-CoA C-acetyltransferase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an acetyl-CoA C-acetyltransferase; or by increasing the number of copies of a nucleic acid encoding an acetyl-CoA C-acetyltransferase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an acetyl-CoA C-acetyltransferase. In some embodiments, an acetyl-CoA C-acetyltransferase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of acetyl-CoA C-acetyltransferase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding an acetyl-CoA C-acetyltransferase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding an acetyl-CoA C-acetyltransferase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

In some embodiments, an acetyl-CoA C-acetyltransferase enzyme can be a fungal or bacterial protein. In a particular embodiment, the acetyl-CoA C-acetyltransferase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein. An example of a *Candida viswanathii* nucleotide sequence (SEQ ID NO: 371) encoding a polypeptide (*ERG10*; SEQ ID NO: 337) having an acetyl-CoA C-acetyltransferase activity is provided herein. Additional nonlimiting examples of nucleotide sequences encoding polypeptides having acetyl-CoA C-acetyltransferase activity include: *Candida tropicalis CTPACTB* gene for acetoacetyl-CoA thiolase A (Genbank accession no. D13471.1), *Candida tropicalis CTPACTA* gene for acetoacetyl-CoA thiolase A (Genbank accession no. D13470.1), and *Candida tropicalis MYA-3404* acetyl-CoA acetyltransferase IB (NCBI Reference Sequence: XM_002547232.1).

Presence, absence or amount of acetyl-CoA C-acetyltransferase activity can be detected by any suitable method known in the art. For example, detection can be performed by using an

acetyltransferase activity kit (e.g., ENZO, ADI-907-026). Nucleic acid sequences encoding native and/or modified acetyl-CoA C-acetyltransferase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding an acetyl-CoA C-acetyltransferase can be modified. For example, the amount of an acetyl-CoA C-acetyltransferase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of acetyl-CoA C-acetyltransferase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing acetyl-CoA C-acetyltransferase activity in a cell can be accomplished by modifying the amount of acetyl-CoA C-acetyltransferase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous acetyl-CoA C-acetyltransferase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type acetyl-CoA C-acetyltransferase such that the encoded modified or substituted acetyl-CoA C-acetyltransferase protein has a reduced enzyme activity.

Modification of an HMG-CoA synthase activity

A further step in the upper mevalonate pathway is the reaction in which acetyl-CoA condenses with acetoacetyl-CoA to form 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA). An HMG-CoA synthase enzyme (also referred to as ERG13, hydroxymethylglutaryl-CoA synthase, (S)-3-hydroxy-3-methylglutaryl-CoA acetoacetyl-CoA-lyase, (CoA-acetylating), 3-hydroxy-3-methylglutaryl CoA synthetase, 3-hydroxy-3-methylglutaryl coenzyme A synthase, 3-hydroxy-3-methylglutaryl coenzyme A synthetase, 3-hydroxy-3-methylglutaryl-CoA synthase, 3-hydroxy-3-methylglutaryl-coenzyme A synthase, beta-hydroxy-beta-methylglutaryl-CoA synthase, acetoacetyl coenzyme A transacetase, hydroxymethylglutaryl coenzyme A synthase, and hydroxymethylglutaryl coenzyme A-condensing enzyme; EC 2.3.3.10) contains a catalytic cysteine residue that acts as a nucleophile in the acetylation of the enzyme by acetyl-CoA (first substrate) to produce an acetyl-enzyme thioester, releasing the reduced coenzyme A. A subsequent nucleophilic attack on acetoacetyl-CoA (second substrate) forms HMG-CoA. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of an HMG-CoA synthase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of an HMG-CoA synthase, may be modified to decrease the amount and/or activity of an HMG-CoA synthase, or may be modified to alternately increase and decrease the amount and/or activity of an HMG-CoA synthase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of an HMG-CoA synthase in a cell is increased. Increasing the amount and/or activity of an HMG-CoA synthase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., mevalonate production, isopentenyl diphosphate production, geranyl diphosphate production, farnesyl diphosphate production, geranylgeranyl diphosphate production, terpene production, carotenoid production, beta carotene production, lycopene production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, HMG-CoA synthase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host HMG-CoA synthase can be increased by increasing the number of copies of a nucleic acid encoding an HMG-CoA

synthase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an HMG-CoA synthase; or by increasing the number of copies of a nucleic acid encoding an HMG-CoA synthase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an HMG-CoA synthase. In some embodiments, an HMG-CoA synthase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of HMG-CoA synthase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding an HMG-CoA synthase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding an HMG-CoA synthase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

In some embodiments, an HMG-CoA synthase enzyme can be a fungal or bacterial protein. In a particular embodiment, the HMG-CoA synthase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein. An example of a *Candida viswanathii* nucleotide sequence (SEQ ID NO: 372) encoding a polypeptide (*ERG13*; SEQ ID NO: 338) having an HMG-CoA synthase is provided herein. Additional nonlimiting examples of nucleotide sequences encoding polypeptides having HMG-CoA synthase activity include: *Candida tropicalis MYA-3404* hydroxymethylglutaryl-CoA synthase (NCBI Reference Sequence: XM_002546412.1), and *Candida tanzawaensis NRRL Y-17324* hydroxymethylglutaryl-CoA synthase (NCBI Reference Sequence: XM_020207398.1).

Presence, absence or amount of HMG-CoA synthase activity can be detected by any suitable method known in the art. For example, detection can be performed by using a visible wavelength spectrophotometric assay for HMG-CoA synthase (see e.g., Skaff et al. (2010) Anal. Biochem. 396(1):96-102). Nucleic acid sequences encoding native and/or modified HMG-CoA synthase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding an HMG-CoA synthase can be modified. For example, the amount of an HMG-CoA synthase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of HMG-CoA synthase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing HMG-CoA synthase activity in a cell can be accomplished by modifying the amount of HMG-CoA synthase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous HMG-CoA synthase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type HMG-CoA synthase such that the encoded modified or substituted HMG-CoA synthase protein has a reduced enzyme activity.

Modification of an HMG-CoA reductase activity

A further (rate controlling) step in the upper mevalonate pathway is the conversion of HMG-CoA to mevalonate by HMG-CoA reductase. HMG-CoA reductase (EC 1.1.1.34) may also be referred to as HMG1, 3-hydroxy-3-methyl-glutaryl-coenzyme A reductase, HMGCR, LDLCQ3, 3-hydroxy-3-methylglutaryl-CoA reductase, and hydroxymethylglutaryl-CoA reductase. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of an HMG-CoA reductase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of an HMG-CoA reductase, may be modified to decrease the amount and/or activity of an HMG-CoA reductase, or may be modified to alternately increase and decrease the amount and/or activity of an HMG-

CoA reductase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of an HMG-CoA reductase in a cell is increased. Increasing the amount and/or activity of an HMG-CoA reductase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., mevalonate production, isopentenyl diphosphate production, geranyl diphosphate production, farnesyl diphosphate production, geranylgeranyl diphosphate production, terpene production, carotenoid production, beta carotene production, lycopene production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, HMG-CoA reductase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host HMG-CoA reductase can be increased by increasing the number of copies of a nucleic acid encoding an HMG-CoA reductase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an HMG-CoA reductase; or by increasing the number of copies of a nucleic acid encoding an HMG-CoA reductase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an HMG-CoA reductase. In some embodiments, an HMG-CoA reductase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of HMG-CoA reductase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding an HMG-CoA reductase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding an HMG-CoA reductase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

In some embodiments, an HMG-CoA reductase enzyme can be a fungal or bacterial protein. In a particular embodiment, the HMG-CoA reductase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein. An example of a *Candida viswanathii* nucleotide sequence (SEQ ID NO: 373) encoding a polypeptide (*HMG1*; SEQ ID NO: 339) having an HMG-CoA reductase activity is provided herein. An

additional nonlimiting example of a nucleotide sequence encoding a polypeptide having HMG-CoA reductase activity includes: *Candida tropicalis MYA-3404* 3-hydroxy-3-methylglutaryl-coenzyme A reductase (NCBI Reference Sequence: XM_002550004.1).

Presence, absence or amount of HMG-CoA reductase activity can be detected by any suitable method known in the art. For example, detection can be performed by using an HMG-CoA Reductase Assay Kit (e.g., SIGMA-ALDRICH, CS1090 Sigma); and/or a Colorimetric HMG-CoA Reductase Activity Assay Kit (ABCAM, ab204701). Nucleic acid sequences encoding native and/or modified HMG-CoA reductase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding an HMG-CoA reductase can be modified. For example, the amount of an HMG-CoA reductase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of HMG-CoA reductase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing HMG-CoA reductase activity in a cell can be accomplished by modifying the amount of HMG-CoA reductase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous HMG-CoA reductase gene in a cell or organism with a weaker

heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type HMG-CoA reductase such that the encoded modified or substituted HMG-CoA reductase protein has a reduced enzyme activity.

Modification of a mevalonate kinase activity

The initial step in the lower mevalonate pathway is the conversion of mevalonate to phosphomevalonate by mevalonate kinase. Mevalonate kinase (EC 2.7.1.36) may also be referred to as ERG12, MVK, LRBP, MK, MVLK, and POROK3. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a mevalonate kinase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a mevalonate kinase, may be modified to decrease the amount and/or activity of a mevalonate kinase, or may be modified to alternately increase and decrease the amount and/or activity of a mevalonate kinase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a mevalonate kinase in a cell is increased. Increasing the amount and/or activity of a mevalonate kinase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., isopentenyl diphosphate production, geranyl diphosphate production, farnesyl diphosphate production, geranylgeranyl diphosphate production, terpene production, carotenoid production, beta carotene production, lycopene production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, mevalonate kinase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host mevalonate kinase can be increased by increasing the number of copies of a nucleic acid encoding a mevalonate kinase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a mevalonate kinase; or by increasing the number of copies of a nucleic acid encoding a mevalonate kinase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a mevalonate kinase. In some embodiments, a mevalonate kinase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems

and methods provided herein, the amount of mevalonate kinase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a mevalonate kinase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a mevalonate kinase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

In some embodiments, a mevalonate kinase enzyme can be a fungal or bacterial protein. In a particular embodiment, the mevalonate kinase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein. An example of a *Candida viswanathii* nucleotide sequence (SEQ ID NO: 374) encoding a polypeptide (*ERG12*; SEQ ID NO: 340) having a mevalonate kinase activity is provided herein.

Presence, absence or amount of mevalonate kinase activity can be detected by any suitable method known in the art. For example, detection can be performed by using a mevalonate kinase assay (see e.g., Green et al. (1970) Anal. Biochem. 38(1):130-138; Gibson et al. (1989) Enzyme 41(1):47-55). Nucleic acid sequences encoding native and/or modified mevalonate kinase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding a mevalonate kinase can be modified. For example, the amount of a mevalonate kinase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some

embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of mevalonate kinase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing mevalonate kinase activity in a cell can be accomplished by modifying the amount of mevalonate kinase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous mevalonate kinase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type mevalonate kinase such that the encoded modified or substituted mevalonate kinase protein has a reduced enzyme activity.

Modification of a phosphomevalonate kinase activity

A further step in the lower mevalonate pathway is the conversion of phosphomevalonate to diphosphomevalonate (mevalonate-5-phosphate to mevalonate-5-pyrophosphate; 5-phosphomevalonate to 5-diphosphomevalonate) by phosphomevalonate kinase. Phosphomevalonate kinase (ERG8; EC 2.7.4.2) may also be referred to as 5-phosphomevalonate phosphotransferase and PMVK. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a phosphomevalonate kinase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a phosphomevalonate kinase, may be modified to decrease the amount and/or activity of a phosphomevalonate kinase, or may be modified to alternately increase and decrease the amount and/or activity of a phosphomevalonate kinase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a phosphomevalonate kinase in a cell is increased. Increasing the amount and/or activity of a phosphomevalonate kinase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., isopentenyl diphosphate production, geranyl diphosphate production, farnesyl diphosphate production, geranylgeranyl diphosphate production, terpene production, carotenoid production, beta carotene production, lycopene

production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, phosphomevalonate kinase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host phosphomevalonate kinase can be increased by increasing the number of copies of a nucleic acid encoding a phosphomevalonate kinase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a phosphomevalonate kinase; or by increasing the number of copies of a nucleic acid encoding a phosphomevalonate kinase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a phosphomevalonate kinase. In some embodiments, a phosphomevalonate kinase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of phosphomevalonate kinase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a phosphomevalonate kinase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding an phosphomevalonate kinase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

In some embodiments, a phosphomevalonate kinase enzyme can be a fungal or bacterial protein. In a particular embodiment, the phosphomevalonate kinase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein. An example of a *Candida viswanathii* nucleotide sequence (SEQ ID NO: 375) encoding a polypeptide (*ERG8*; SEQ ID NO: 341) having a phosphomevalonate kinase activity is provided herein.

Presence, absence or amount of phosphomevalonate kinase activity can be detected by any suitable method known in the art. For example, detection can be performed by using a microplate assay for phosphomevalonate kinase activity (see e.g., Schulte et al. (1999) Anal. Biochem. 269(2):245-54). Nucleic acid sequences encoding native and/or modified phosphomevalonate kinase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations

thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding a phosphomevalonate kinase can be modified. For example, the amount of a phosphomevalonate kinase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of phosphomevalonate kinase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing phosphomevalonate kinase activity in a cell can be accomplished by modifying the amount of phosphomevalonate kinase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous phosphomevalonate kinase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type phosphomevalonate kinase such that the encoded modified or substituted phosphomevalonate kinase protein has a reduced enzyme activity.

Modification of a diphosphomevalonate decarboxylase activity

A further step in the lower mevalonate pathway is the conversion of mevalonate 5-diphosphate (diphosphomevalonate; mevalonate-5-pyrophosphate; 5-diphosphomevalonate) to isopentenyl diphosphate. A diphosphomevalonate decarboxylase enzyme (also referred to as MVD1,

ERG19, pyrophosphomevalonate decarboxylase, mevalonate-5-pyrophosphate decarboxylase, pyrophosphomevalonic acid decarboxylase, 5-pyrophosphomevalonate decarboxylase, mevalonate 5-diphosphate decarboxylase, and ATP:(R)-5-diphosphomevalonate carboxy-lyase (dehydrating); EC 4.1.1.33) can convert mevalonate 5-diphosphate to isopentenyl diphosphate through ATP dependent decarboxylation. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a diphosphomevalonate decarboxylase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a diphosphomevalonate decarboxylase, may be modified to decrease the amount and/or activity of a diphosphomevalonate decarboxylase, or may be modified to alternately increase and decrease the amount and/or activity of a diphosphomevalonate decarboxylase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a diphosphomevalonate decarboxylase in a cell is increased. Increasing the amount and/or activity of a diphosphomevalonate decarboxylase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., isopentenyl diphosphate production, geranyl diphosphate production, farnesyl diphosphate production, geranylgeranyl diphosphate production, terpene production, carotenoid production, beta carotene production, lycopene production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, diphosphomevalonate decarboxylase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host diphosphomevalonate decarboxylase can be increased by increasing the number of copies of a nucleic acid encoding a diphosphomevalonate decarboxylase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a diphosphomevalonate decarboxylase; or by increasing the number of copies of a nucleic acid encoding a diphosphomevalonate decarboxylase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a diphosphomevalonate decarboxylase. In some embodiments, a diphosphomevalonate decarboxylase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of

diphosphomevalonate decarboxylase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a diphosphomevalonate decarboxylase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a diphosphomevalonate decarboxylase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

In some embodiments, a diphosphomevalonate decarboxylase enzyme can be a fungal or bacterial protein. In a particular embodiment, the diphosphomevalonate decarboxylase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein. An example of a *Candida viswanathii* nucleotide sequence (SEQ ID NO: 376) encoding a polypeptide (*MVD1* (*ERG19)*; SEQ ID NO: 342) having a diphosphomevalonate decarboxylase activity is provided herein. Additional nonlimiting examples of nucleotide sequences encoding polypeptides having diphosphomevalonate decarboxylase activity include: *Candida tropicalis MYA-3404* diphosphomevalonate decarboxylase (NCBI Reference Sequence: XM_002549937.1) and *Candida tanzawaensis NRRL Y-17324* Diphosphomevalonate decarboxylase (NCBI Reference Sequence: XM_020210701.1).

Presence, absence or amount of diphosphomevalonate decarboxylase activity can be detected by any suitable method known in the art. Nucleic acid sequences encoding native and/or modified diphosphomevalonate decarboxylase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding a diphosphomevalonate decarboxylase can be modified. For example, the amount of a diphosphomevalonate decarboxylase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of

expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of diphosphomevalonate decarboxylase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing diphosphomevalonate decarboxylase activity in a cell can be accomplished by modifying the amount of diphosphomevalonate decarboxylase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous diphosphomevalonate decarboxylase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type diphosphomevalonate decarboxylase such that the encoded modified or substituted diphosphomevalonate decarboxylase protein has a reduced enzyme activity.

Modification of an isopentenyl diphosphate delta isomerase activity

A further step in the lower mevalonate pathway is the conversion of isopentenyl diphosphate (isopentenyl pyrophosphate, IPP) to dimethylallyl diphosphate (dimethylallyl pyrophosphate, DMAPP). An isopentenyl diphosphate delta isomerase enzyme (also referred to as IDI1, isopentenyl pyrophosphate isomerase, IPP isomerase; EC 5.3.3.2) can catalyze the isomerization of isopentenyl diphosphate to dimethylallyl diphosphate by an antarafacial transposition of hydrogen. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of an isopentenyl diphosphate delta isomerase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of an isopentenyl diphosphate delta isomerase, may be modified to decrease the amount and/or activity of an isopentenyl diphosphate delta isomerase, or may be modified to alternately increase and decrease the amount and/or activity of an isopentenyl diphosphate delta isomerase depending, for example, on the cellular

location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of an isopentenyl diphosphate delta isomerase in a cell is increased. Increasing the amount and/or activity of an isopentenyl diphosphate delta isomerase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., geranyl diphosphate production, farnesyl diphosphate production, geranylgeranyl diphosphate production, terpene production, carotenoid production, beta carotene production, lycopene production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, isopentenyl diphosphate delta isomerase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host isopentenyl diphosphate delta isomerase can be increased by increasing the number of copies of a nucleic acid encoding an isopentenyl diphosphate delta isomerase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an isopentenyl diphosphate delta isomerase; or by increasing the number of copies of a nucleic acid encoding an isopentenyl diphosphate delta isomerase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an isopentenyl diphosphate delta isomerase. In some embodiments, an isopentenyl diphosphate delta isomerase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of isopentenyl diphosphate delta isomerase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding an isopentenyl diphosphate delta isomerase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding an isopentenyl diphosphate delta isomerase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

In some embodiments, an isopentenyl diphosphate delta isomerase enzyme can be a fungal or bacterial protein. In a particular embodiment, the isopentenyl diphosphate delta isomerase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*),

Chronobacter (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein. An example of a *Candida viswanathii* nucleotide sequence (SEQ ID NO: 377) encoding a polypeptide (*IDI1*; SEQ ID NO: 343) having an isopentenyl diphosphate delta isomerase activity is provided herein. An additional nonlimiting example of a nucleotide sequence encoding a polypeptide having isopentenyl diphosphate delta isomerase activity is a *Candida tropicalis MYA-3404* isopentenyl-diphosphate delta-isomerase (NCBI Reference Sequence: XM_002545339.1).

Presence, absence or amount of isopentenyl diphosphate delta isomerase activity can be detected by any suitable method known in the art. For example, detection can be performed by using an isopentenyl diphosphate delta isomerase activity assay (see e.g., Diaz et al. (2012) Insect Biochem. Mol. Bio. 42(10):751-757). Nucleic acid sequences encoding native and/or modified isopentenyl diphosphate delta isomerase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding an isopentenyl diphosphate delta isomerase can be modified. For example, the amount of an isopentenyl diphosphate delta isomerase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of isopentenyl diphosphate delta isomerase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the

host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing isopentenyl diphosphate delta isomerase activity in a cell can be accomplished by modifying the amount of isopentenyl diphosphate delta isomerase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous isopentenyl diphosphate delta isomerase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type isopentenyl diphosphate delta isomerase such that the encoded modified or substituted isopentenyl diphosphate delta isomerase protein has a reduced enzyme activity.

Modification of a terpene synthase activity

Provided herein are one or more modifications to one or more terpene synthase activities. Terpene synthases may include one or more enzymes in a family of enzymes involved in terpene biosynthesis, such as, for example, terpene synthases known in the art and described in Yamada et al. (2015) PNAS 112(3):857-862. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a terpene synthase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a terpene synthase, may be modified to decrease the amount and/or activity of a terpene synthase, or may be modified to alternately increase and decrease the amount and/or activity of a terpene synthase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a terpene synthase in a cell is increased. Increasing the amount and/or activity of a terpene synthase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., mevalonate production, isopentenyl diphosphate production, geranyl diphosphate production, farnesyl diphosphate production, geranylgeranyl diphosphate production, terpene production, carotenoid production, beta carotene production, lycopene production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, terpene synthase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host terpene synthase can be increased by increasing the number of copies of a nucleic acid encoding a terpene synthase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a terpene synthase; or by increasing the number of copies of a nucleic acid a terpene synthase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a terpene synthase. In some embodiments, a terpene synthase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of terpene synthase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a terpene synthase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a terpene synthase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed. In some embodiments, a terpene synthase is not endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, a terpene synthase protein can be expressed in a cell by introducing heterologous nucleic acid encoding a terpene synthase into a cell or microorganism.

In some embodiments, a terpene synthase enzyme can be a fungal or bacterial protein. In a particular embodiment, the terpene synthase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein.

Presence, absence or amount of terpene synthase activity can be detected by any suitable method known in the art. For example, detection can be performed by using an appropriate terpene synthase activity assay. Nucleic acid sequences encoding native and/or modified terpene synthase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding a terpene synthase can be modified. For example, the amount of a terpene synthase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of terpene synthase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing terpene synthase activity in a cell can be accomplished by modifying the amount of terpene synthase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous terpene synthase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type terpene synthase such that the encoded modified or substituted terpene synthase protein has a reduced enzyme activity.

Modification of a dimethylallyltranstransferase activity

One component of terpene biosynthesis is the conversion of dimethylallyl diphosphate (dimethylallyl pyrophosphate) and isopentenyl diphosphate (isopentenyl pyrophosphate) into farnesyl diphosphate (farnesyl pyrophosphate). A dimethylallyltranstransferase enzyme (EC 2.5.1.1) can convert dimethylallyl diphosphate and isopentenyl diphosphate into farnesyl diphosphate. Dimethylallyltranstransferase may also be referred to as ERG20, farnesyl pyrophosphate synthase, farnesyl diphosphate synthase, or farnesyl diphosphate synthetase (EC 2.5.1.10). Geranyl diphosphate (geranyl pyrophosphate) often is created in an intermediate step. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a dimethylallyltranstransferase in a cell is modified. For

example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a dimethylallyltranstransferase, may be modified to decrease the amount and/or activity of a dimethylallyltranstransferase, or may be modified to alternately increase and decrease the amount and/or activity of a dimethylallyltranstransferase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a dimethylallyltranstransferase in a cell is increased. Increasing the amount and/or activity of a dimethylallyltranstransferase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., geranyl diphosphate production, farnesyl diphosphate production, geranylgeranyl diphosphate production, terpene production, carotenoid production, beta carotene production, lycopene production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, dimethylallyltranstransferase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host dimethylallyltranstransferase can be increased by increasing the number of copies of a nucleic acid encoding a dimethylallyltranstransferase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a dimethylallyltranstransferase; or by increasing the number of copies of a nucleic acid encoding a dimethylallyltranstransferase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a dimethylallyltranstransferase. In some embodiments, a dimethylallyltranstransferase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of dimethylallyltranstransferase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a dimethylallyltranstransferase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a dimethylallyltranstransferase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed. In some embodiments, a dimethylallyltranstransferase is not endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, a dimethylallyltranstransferase protein can be expressed in a cell by

introducing heterologous nucleic acid encoding a dimethylallyltranstransferase into a cell or microorganism.

In some embodiments, a dimethylallyltranstransferase enzyme can be a fungal or bacterial protein. In a particular embodiment, the dimethylallyltranstransferase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein. An example of a *Candida viswanathii* nucleotide sequence (SEQ ID NO: 378) encoding a polypeptide (*ERG20*; SEQ ID NO: 344) having a dimethylallyltranstransferase activity is provided herein. An additional nonlimiting example of a nucleotide sequence encoding a polypeptide having dimethylallyltranstransferase activity is *Candida tropicalis MYA-3404* farnesyl pyrophosphate synthetase (NCBI Reference Sequence: XM_002547254.1).

Presence, absence or amount of dimethylallyltranstransferase activity can be detected by any suitable method known in the art. For example, detection can be performed by using a dimethylallyltranstransferase ELISA kit (e.g., MYBIOSOURCE, MBS943684). Nucleic acid sequences encoding native and/or modified dimethylallyltranstransferase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding a dimethylallyltranstransferase can be modified. For example, the amount of a dimethylallyltranstransferase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some

embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of dimethylallyltranstransferase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing dimethylallyltranstransferase activity in a cell can be accomplished by modifying the amount of dimethylallyltranstransferase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous dimethylallyltranstransferase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type dimethylallyltranstransferase such that the encoded modified or substituted dimethylallyltranstransferase protein has a reduced enzyme activity.

Modification of a geranylgeranyl diphosphate synthase activity

Another component of terpene biosynthesis is the synthesis of geranylgeranyl diphosphate from farnesyl diphosphate and isopentenyl diphosphate. A geranylgeranyl diphosphate synthase enzyme (also referred to as BTS1, CrtE, GGPS1, GGPPS, GGPPS1, geranylgeranyl diphosphate synthase 1; EC 2.5.1.29) can catalyze the synthesis of geranylgeranyl diphosphate from farnesyl diphosphate and isopentenyl diphosphate. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a geranylgeranyl diphosphate synthase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a geranylgeranyl diphosphate synthase, may be modified to decrease the amount and/or activity of geranylgeranyl diphosphate synthase, or may be modified to alternately increase and decrease the amount and/or activity of an geranylgeranyl diphosphate synthase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a geranylgeranyl diphosphate synthase in a cell is increased. Increasing the amount and/or activity of a geranylgeranyl diphosphate synthase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., mevalonate production, isopentenyl

diphosphate production, geranyl diphosphate production, farnesyl diphosphate production, geranylgeranyl diphosphate production, terpene production, carotenoid production, beta carotene production, lycopene production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, geranylgeranyl diphosphate synthase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host geranylgeranyl diphosphate synthase can be increased by increasing the number of copies of a nucleic acid encoding a geranylgeranyl diphosphate synthase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a geranylgeranyl diphosphate synthase; or by increasing the number of copies of a nucleic acid encoding a geranylgeranyl diphosphate synthase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a geranylgeranyl diphosphate synthase. In some embodiments, a geranylgeranyl diphosphate synthase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of geranylgeranyl diphosphate synthase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a geranylgeranyl diphosphate synthase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a geranylgeranyl diphosphate synthase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed. In some embodiments, a geranylgeranyl diphosphate synthase is not endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, a geranylgeranyl diphosphate synthase protein can be expressed in a cell by introducing heterologous nucleic acid encoding a geranylgeranyl diphosphate synthase into a cell or microorganism.

In some embodiments, a geranylgeranyl diphosphate synthase enzyme can be a fungal or bacterial protein. In a particular embodiment, the geranylgeranyl diphosphate synthase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein. An example of a *Candida viswanathii* nucleotide sequence (SEQ ID NO: 379) encoding a polypeptide (*BTS1*; SEQ ID NO: 345) having a geranylgeranyl diphosphate synthase activity is provided herein. Additional nonlimiting examples of nucleotide sequences

encoding polypeptides having geranylgeranyl diphosphate synthase activity are provided herein which include: *Chronobacter sakazakii CrtE* gene (SEQ ID NO:357), which encodes *CsCrtE* (SEQ ID NO:326); *Xanthophyllomyces dendrorhous CrtE* gene (SEQ ID NO: 360, which encodes *XdCrtE* (SEQ ID NO:329); and *Pantoea ananatis CrtE* gene (SEQ ID NO: 363), which encodes *PaCrtE* (SEQ ID NO:332).

Presence, absence or amount of geranylgeranyl diphosphate synthase activity can be detected by any suitable method known in the art. For example, detection can be performed by using a geranylgeranyl diphosphate synthase ELISA kit (e.g., MYBIOSOURCE, MBS929545). Nucleic acid sequences encoding native and/or modified geranylgeranyl diphosphate synthase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding a geranylgeranyl diphosphate synthase can be modified. For example, the amount of a geranylgeranyl diphosphate synthase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of geranylgeranyl diphosphate synthase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing geranylgeranyl diphosphate synthase activity in a cell can be accomplished by modifying the amount of geranylgeranyl diphosphate synthase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous geranylgeranyl diphosphate synthase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type geranylgeranyl diphosphate synthase such that the encoded modified or substituted geranylgeranyl diphosphate synthase protein has a reduced enzyme activity.

Modification of a phytoene synthase activity

Another component of terpene biosynthesis is the conversion of geranylgeranyl diphosphate (geranylgeranyl pyrophosphate) to phytoene. A phytoene synthase enzyme (also referred to as CrtB, prephytoene-diphosphate synthase, 15-cis-phytoene synthase, PSase, geranylgeranyl-diphosphate geranylgeranyltransferase; EC 2.5.1.32) is a transferase enzyme that can catalyze the conversion of geranylgeranyl diphosphate to phytoene. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a phytoene synthase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a phytoene synthase, may be modified to decrease the amount and/or activity of a phytoene synthase, or may be modified to alternately increase and decrease the amount and/or activity of a phytoene synthase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a phytoene synthase in a cell is increased. Increasing the amount and/or activity of a phytoene synthase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., terpene production, carotenoid production, beta carotene production, lycopene production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, phytoene synthase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host phytoene synthase can be increased by increasing the number of copies of a nucleic acid encoding a phytoene synthase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by

increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a phytoene synthase; or by increasing the number of copies of a nucleic acid encoding a phytoene synthase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a phytoene synthase. In some embodiments, a phytoene synthase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of phytoene synthase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a phytoene synthase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a phytoene synthase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed. In some embodiments, a phytoene synthase is not endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, a phytoene synthase protein can be expressed in a cell by introducing heterologous nucleic acid encoding a phytoene synthase into a cell or microorganism.

In some embodiments, a phytoene synthase enzyme can be a fungal or bacterial protein. In a particular embodiment, the phytoene synthase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein. An example of a *Chronobacter sakazakii* nucleotide sequence (SEQ ID NO: 359) encoding a polypeptide (*CsCrtB*; SEQ ID NO: 328) having a phytoene synthase activity is provided herein. An example of a *Pantoea ananatis* nucleotide sequence (SEQ ID NO: 365) encoding a polypeptide (*PaCrtB*; SEQ ID NO: 334) having a phytoene synthase activity is provided herein.

Presence, absence or amount of phytoene synthase activity can be detected by any suitable method known in the art. For example, detection can be performed by using phytoene synthase activity assay (see e.g., Lopez-Emparan et al. (2014) PLoS One 9(12):e114878; Schledz et al. (1996) The Plant Journal 10(5):781-792). Nucleic acid sequences encoding native and/or modified phytoene synthase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or

organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding a phytoene synthase can be modified. For example, the amount of a phytoene synthase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of phytoene synthase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing phytoene synthase activity in a cell can be accomplished by modifying the amount of phytoene synthase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous phytoene synthase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type phytoene synthase such that the encoded modified or substituted phytoene synthase protein has a reduced enzyme activity.

Modification of a phytoene desaturase activity

Another component of terpene biosynthesis is the conversion of phytoene to lycopene. A phytoene desaturase enzyme can catalyze the conversion of phytoene to lycopene. In some embodiments, a phytoene desaturase is a 3,4-didehydrolycopene-forming phytoene desaturase (also referred to as 5-step phytoene desaturase, five-step phytoene desaturase, AI-1, 15-cis-phytoene:acceptor oxidoreductase (3,4-didehydrolycopene-forming); EC 1.3.99.30). In some embodiments, a phytoene desaturase is a neurosporene-forming phytoene desaturase (also

referred to as CrtI, 3-step phytoene desaturase, three-step phytoene desaturase, 15-cis-phytoene:acceptor oxidoreductase (neurosporene-forming); EC 1.3.99.28). In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a phytoene desaturase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a phytoene desaturase, may be modified to decrease the amount and/or activity of a phytoene desaturase, or may be modified to alternately increase and decrease the amount and/or activity of a phytoene desaturase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a phytoene desaturase in a cell is increased. Increasing the amount and/or activity of a phytoene desaturase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., terpene production, carotenoid production, beta carotene production, lycopene production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, phytoene desaturase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host phytoene desaturase can be increased by increasing the number of copies of a nucleic acid encoding a phytoene desaturase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a phytoene desaturase; or by increasing the number of copies of a nucleic acid encoding a phytoene desaturase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a phytoene desaturase. In some embodiments, a phytoene desaturase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of phytoene desaturase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a phytoene desaturase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a phytoene desaturase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed. In some embodiments, a phytoene desaturase is not endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, a phytoene desaturase protein can

be expressed in a cell by introducing heterologous nucleic acid encoding a phytoene desaturase into a cell or microorganism.

In some embodiments, a phytoene desaturase enzyme can be a fungal or bacterial protein. In a particular embodiment, the phytoene desaturase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein. An example of a *Chronobacter sakazakii* nucleotide sequence (SEQ ID NO: 358) encoding a polypeptide (*CsCrtI*; SEQ ID NO: 327) having a phytoene desaturase activity is provided herein. An example of a *Xanthophyllomyces dendrorhous* nucleotide sequence (SEQ ID NO: 361) encoding a polypeptide (*XdCrtI*; SEQ ID NO: 330) having a phytoene desaturase activity is provided herein. An example of a *Pantoea ananatis* nucleotide sequence (SEQ ID NO: 364) encoding a polypeptide (*PaCrtI*; SEQ ID NO: 333) having a phytoene desaturase activity is provided herein.

Presence, absence or amount of phytoene desaturase activity can be detected by any suitable method known in the art. For example, detection can be performed by using a phytoene desaturase activity assay (see e.g., Xu et al. (2007) Microbiology 153:1642-52). Nucleic acid sequences encoding native and/or modified phytoene desaturase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding a phytoene desaturase can be modified. For example, the amount of a phytoene desaturase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide

sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of phytoene desaturase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing phytoene desaturase activity in a cell can be accomplished by modifying the amount of phytoene desaturase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous phytoene desaturase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type phytoene desaturase such that the encoded modified or substituted phytoene desaturase protein has a reduced enzyme activity.

Modification of a lycopene cyclase activity and/or modification of a bifunctional lycopene cyclase/phytoene synthase activity

Another component of terpene biosynthesis is the conversion of lycopene to beta carotene. A lycopene cyclase enzyme (also referred to as lycopene beta-cyclase, CrtY, CrtL (beta-ionone end group producing), CrtL (eta-ionone end group producing) and CrtL (capsanthin/capsorubin synthase); EC 5.5.1.19) can catalyze the conversion of lycopene to beta carotene. The cyclization of lycopene is typically the final step in carotenoid biosynthesis and may proceed via one of two pathways: the formation of a beta ring by beta-cyclase, or an epsilon ring by epsilon-cyclase. Epsilon-cyclase adds one ring, forming a monocyclic delta-carotene, and beta-cyclase introduces a ring at both ends of lycopene to form a bicyclic beta-carotene. In some embodiments, a lycopene cyclase is a bifunctional lycopene cyclase/phytoene synthase (also referred to as CrtYB), which includes a lycopene beta-cyclase domain (EC 5.5.1.19) and a phytoene synthase domain (EC 2.5.1.32). A bifunctional lycopene cyclase can catalyze the reaction from geranylgeranyl diphosphate to phytoene (phytoene synthase) and the reaction from lycopene to beta-carotene via an intermediate gamma-carotene (lycopene cyclase). A bifunctional enzyme containing lycopene cyclase activity and phytoene synthase activity may be referred to herein as a bifunctional lycopene cyclase/phytoene synthase or a bifunctional lycopene cyclase.

In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a lycopene cyclase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a lycopene cyclase, may be modified to decrease the amount and/or activity of a lycopene cyclase, or may be modified to alternately increase and decrease the amount and/or activity of a lycopene cyclase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a bifunctional lycopene cyclase/phytoene synthase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a bifunctional lycopene cyclase/phytoene synthase, may be modified to decrease the amount and/or activity of a bifunctional lycopene cyclase/phytoene synthase, or may be modified to alternately increase and decrease the amount and/or activity of a bifunctional lycopene cyclase/phytoene synthase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a lycopene cyclase in a cell is increased. Increasing the amount and/or activity of a lycopene cyclase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., terpene production, carotenoid production, beta carotene production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain aspects, the amount and/or activity of a bifunctional lycopene cyclase/phytoene synthase in a cell is increased. Increasing the amount and/or activity of a bifunctional lycopene cyclase/phytoene synthase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., terpene production, carotenoid production, beta carotene production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, lycopene cyclase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host lycopene cyclase can be increased by increasing the number of copies of a nucleic acid encoding a lycopene cyclase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a lycopene cyclase; or by increasing the number of copies of a nucleic acid encoding a lycopene cyclase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a lycopene cyclase. In some embodiments, a lycopene cyclase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of lycopene cyclase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a lycopene cyclase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a lycopene cyclase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed. In some embodiments, a lycopene cyclase is not endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, a lycopene cyclase protein can be expressed in a cell by introducing heterologous nucleic acid encoding a lycopene cyclase into a cell or microorganism.

In certain embodiments, bifunctional lycopene cyclase/phytoene synthase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host bifunctional lycopene cyclase/phytoene synthase can be increased by increasing the number of copies of a nucleic acid encoding a bifunctional lycopene cyclase/phytoene synthase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a bifunctional lycopene cyclase/phytoene synthase; or by increasing the number of copies of a nucleic acid encoding a bifunctional lycopene cyclase/phytoene synthase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a bifunctional lycopene cyclase/phytoene synthase. In some embodiments, a bifunctional lycopene cyclase/phytoene synthase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of bifunctional lycopene cyclase/phytoene synthase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a bifunctional lycopene cyclase/phytoene synthase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a bifunctional lycopene cyclase/phytoene synthase can result in increased copy number of such nucleic acids and/or

provide for modification of the cellular location in which the protein is expressed. In some embodiments, a bifunctional lycopene cyclase/phytoene synthase is not endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, a bifunctional lycopene cyclase/phytoene synthase protein can be expressed in a cell by introducing heterologous nucleic acid encoding a bifunctional lycopene cyclase/phytoene synthase into a cell or microorganism.

In some embodiments, a lycopene cyclase enzyme and/or a bifunctional lycopene cyclase/phytoene synthase enzyme can be a fungal or bacterial protein. In a particular embodiment, the lycopene cyclase enzyme and/or the bifunctional lycopene cyclase/phytoene synthase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein. An example of a *Xanthophyllomyces dendrorhous* nucleotide sequence (SEQ ID NO: 362) encoding a polypeptide (*XdCrtYB*; SEQ ID NO: 331) having a bifunctional lycopene cyclase/phytoene synthase activity is provided herein.

Presence, absence or amount of lycopene cyclase activity and/or bifunctional lycopene cyclase/phytoene synthase activity can be detected by any suitable method known in the art. For example, detection can be performed by using a lycopene cyclase activity assay (see e.g., Yu et al. (2010) J. Biol. Chem. 285(16):12109-12120). Nucleic acid sequences encoding native and/or modified lycopene cyclase sequences and/or bifunctional lycopene cyclase/phytoene synthase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding a lycopene cyclase and/or a bifunctional lycopene cyclase/phytoene synthase can be modified. For example, the amount of a lycopene cyclase protein and/or a bifunctional lycopene cyclase/phytoene synthase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or

microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of lycopene cyclase expression and/or a bifunctional lycopene cyclase/phytoene synthase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing lycopene cyclase activity in a cell can be accomplished by modifying the amount of lycopene cyclase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous lycopene cyclase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type lycopene cyclase such that the encoded modified or substituted lycopene cyclase protein has a reduced enzyme activity.

Decreasing bifunctional lycopene cyclase/phytoene synthase activity in a cell can be accomplished by modifying the amount of bifunctional lycopene cyclase/phytoene synthase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous bifunctional lycopene cyclase/phytoene synthase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type bifunctional lycopene cyclase/phytoene synthase such that the encoded modified or substituted bifunctional lycopene cyclase/phytoene synthase protein has a reduced enzyme activity.

Modification of a beta carotene ketolase activity

Additional components of terpene biosynthesis include, for example, astaxanthin biosysnthesis (e.g., the conversion of beta carotene to astaxanthin), canthaxanthin biosynthesis, and zeaxanthin biosynthesis. Canthaxanthin, for example, may be produced using a beta carotene ketolase enzyme (also referred to as CrtW, beta carotene monoketolase, beta-carotene

oxygenase; EC 1.3.5.B4). Astaxanthin, for example, may be produced using a beta carotene ketolase enzyme in conjunction with a beta carotene hydroxylase (discussed below). In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a beta carotene ketolase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a beta carotene ketolase, may be modified to decrease the amount and/or activity of a beta carotene ketolase, or may be modified to alternately increase and decrease the amount and/or activity of a beta carotene ketolase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a beta carotene ketolase in a cell is increased. Increasing the amount and/or activity of a beta carotene ketolase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., terpene production, carotenoid production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments beta carotene ketolase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host beta carotene ketolase can be increased by increasing the number of copies of a nucleic acid encoding an beta carotene ketolase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a beta carotene ketolase; or by increasing the number of copies of a nucleic acid encoding a beta carotene ketolase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a beta carotene ketolase. In some embodiments, a beta carotene ketolase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of beta carotene ketolase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a beta carotene ketolase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a beta carotene ketolase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed. In some embodiments, a beta carotene ketolase is not endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, a beta carotene ketolase protein can be expressed in a cell by introducing heterologous nucleic acid encoding a beta carotene ketolase into a cell or microorganism.

In some embodiments, a beta carotene ketolase enzyme can be a fungal or bacterial protein. In a particular embodiment, the beta carotene ketolase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), *Pantoea* (e.g., *P. ananatis*), or *Agrobacterium* (e.g., *A. aurantiacum*) protein. An example of an *Agrobacterium aurantiacum* nucleotide sequence (SEQ ID NO: 369) encoding a polypeptide (*AaCrtW*; SEQ ID NO: 351) having a beta carotene ketolase activity is provided herein.

Presence, absence or amount of beta carotene ketolase activity can be detected by any suitable method known in the art. For example, detection can be performed by using an appropriate beta carotene ketolase activity assay. Nucleic acid sequences encoding native and/or modified beta carotene ketolase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding a beta carotene ketolase can be modified. For example, the amount of a beta carotene ketolase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of beta carotene ketolase expression by a carbon source or feedstock

include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing beta carotene ketolase activity in a cell can be accomplished by modifying the amount of beta carotene ketolase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous beta carotene ketolase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type beta carotene ketolase such that the encoded modified or substituted beta carotene ketolase protein has a reduced enzyme activity.

Modification of a beta carotene hydroxylase activity

In certain instances, astaxanthin and/or zeaxanthin may be produced using a beta carotene hydroxylase enzyme (also referred to as CrtZ, beta-carotene 3-hydroxylase, beta-carotene 3,3'-monooxygenase, beta-carotene, NADH:oxygen 3-oxidoreductase; EC 1.14.13.129). In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a beta carotene hydroxylase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a beta carotene hydroxylase, may be modified to decrease the amount and/or activity of a beta carotene hydroxylase, or may be modified to alternately increase and decrease the amount and/or activity of a beta carotene hydroxylase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a beta carotene hydroxylase in a cell is increased. Increasing the amount and/or activity of a beta carotene hydroxylase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., terpene production, carotenoid production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, beta carotene hydroxylase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host beta carotene hydroxylase can be increased by increasing the number of copies of a nucleic acid encoding a beta carotene hydroxylase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more

copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a beta carotene hydroxylase; or by increasing the number of copies of a nucleic acid encoding a beta carotene hydroxylase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a beta carotene hydroxylase. In some embodiments, a beta carotene hydroxylase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of beta carotene hydroxylase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a beta carotene hydroxylase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a beta carotene hydroxylase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed. In some embodiments, a beta carotene hydroxylase is not endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, a beta carotene hydroxylase protein can be expressed in a cell by introducing heterologous nucleic acid encoding a beta carotene hydroxylase into a cell or microorganism.

In some embodiments, a beta carotene hydroxylase enzyme can be a fungal or bacterial protein. In a particular embodiment, the beta carotene hydroxylase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), *Pantoea* (e.g., *P. ananatis*) or *Agrobacterium* (e.g., *A. aurantiacum*) protein. An example of an *Agrobacterium aurantiacum* nucleotide sequence (SEQ ID NO: 368) encoding a polypeptide (*AaCrtZ*; SEQ ID NO: 350) having a beta carotene hydroxylase activity is provided herein.

Presence, absence or amount of beta carotene hydroxylase activity can be detected by any suitable method known in the art. For example, detection can be performed by using an appropriate beta carotene hydroxylase activity assay. Nucleic acid sequences encoding native and/or modified beta carotene hydroxylase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding a beta carotene hydroxylase can be modified. For example, the amount of a beta carotene hydroxylase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of beta carotene hydroxylase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing beta carotene hydroxylase activity in a cell can be accomplished by modifying the amount of beta carotene hydroxylase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous beta carotene hydroxylase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type beta carotene hydroxylase such that the encoded modified or substituted beta carotene hydroxylase protein has a reduced enzyme activity.

Modification of an astaxanthin synthase activity

In certain instances, astaxanthin may be produced using an astaxanthin synthase enzyme (also referred to as CrtS, astaxanthin synthetase, cytochrome-450 hydroxylase, cytochrome-P450 hydroxylase/astaxanthin synthase). In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of an astaxanthin synthase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of an astaxanthin synthase, may be modified to decrease the amount and/or activity of an astaxanthin synthase, or may be modified to

alternately increase and decrease the amount and/or activity of an astaxanthin synthase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of an astaxanthin synthase in a cell is increased. Increasing the amount and/or activity of an astaxanthin synthase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., terpene production, carotenoid production, astaxanthin production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, astaxanthin synthase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host astaxanthin synthase can be increased by increasing the number of copies of a nucleic acid encoding an astaxanthin synthase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an astaxanthin synthase; or by increasing the number of copies of a nucleic acid encoding an astaxanthin synthase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an astaxanthin synthase. In some embodiments, an astaxanthin synthase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of astaxanthin synthase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding an astaxanthin synthase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding an astaxanthin synthase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed. In some embodiments, a astaxanthin synthase is not endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, a astaxanthin synthase protein can be expressed in a cell by introducing heterologous nucleic acid encoding a astaxanthin synthase into a cell or microorganism.

In some embodiments, an astaxanthin synthase enzyme can be a fungal or bacterial protein. In a particular embodiment, the astaxanthin synthase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein. An example

of a *Xanthophyllomyces dendrorhous* nucleotide sequence (SEQ ID NO: 367) encoding a polypeptide (*XdCrtS*; SEQ ID NO: 336) having an astaxanthin synthase activity is provided herein.

Presence, absence or amount astaxanthin synthase activity can be detected by any suitable method known in the art. For example, detection can be performed by using an appropriate astaxanthin synthase activity assay. Nucleic acid sequences encoding native and/or modified astaxanthin synthase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding an astaxanthin synthase can be modified. For example, the amount of an astaxanthin synthase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of astaxanthin synthase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing astaxanthin synthase activity in a cell can be accomplished by modifying the amount of astaxanthin synthase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous astaxanthin synthase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or

modifying a gene encoding a wild-type astaxanthin synthase such that the encoded modified or substituted astaxanthin synthase protein has a reduced enzyme activity.

Modification of a zeaxanthin glucosyltransferase activity

Another component of terpene biosynthesis is the conversion of zeaxanthin to zeaxanthin diglucoside. A zeaxanthin glucosyltransferase enzyme (also referred to as CrtX, UDP-glucose:zeaxanthin beta-D-glucosyltransferase; EC 2.4.1.276) can catalyze the conversion of zeaxanthin to zeaxanthin diglucoside. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a zeaxanthin glucosyltransferase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a zeaxanthin glucosyltransferase, may be modified to decrease the amount and/or activity of a zeaxanthin glucosyltransferase, or may be modified to alternately increase and decrease the amount and/or activity of a zeaxanthin glucosyltransferase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a zeaxanthin glucosyltransferase in a cell is increased. Increasing the amount and/or activity of a zeaxanthin glucosyltransferase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., terpene production, carotenoid production, xanthophyll production, zeaxanthin diglucoside production), and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, zeaxanthin glucosyltransferase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host zeaxanthin glucosyltransferase can be increased by increasing the number of copies of a nucleic acid encoding a zeaxanthin glucosyltransferase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a zeaxanthin glucosyltransferase; or by increasing the number of copies of a nucleic acid encoding a zeaxanthin glucosyltransferase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a zeaxanthin glucosyltransferase. In some embodiments, a zeaxanthin glucosyltransferase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods

provided herein, the amount of zeaxanthin glucosyltransferase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a zeaxanthin glucosyltransferase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a zeaxanthin glucosyltransferase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed. In some embodiments, a zeaxanthin glucosyltransferase is not endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, a zeaxanthin glucosyltransferase protein can be expressed in a cell by introducing heterologous nucleic acid encoding a zeaxanthin glucosyltransferase into a cell or microorganism.

In some embodiments, a zeaxanthin glucosyltransferase enzyme can be a fungal or bacterial protein. In a particular embodiment, the zeaxanthin glucosyltransferase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), or *Pantoea* (e.g., *P. ananatis*) protein.

Presence, absence or amount of zeaxanthin glucosyltransferase activity can be detected by any suitable method known in the art. For example, detection can be performed by using an appropriate zeaxanthin glucosyltransferase activity assay. Nucleic acid sequences encoding native and/or modified zeaxanthin glucosyltransferase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding a zeaxanthin glucosyltransferase can be modified. For example, the amount of a zeaxanthin glucosyltransferase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous

promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of zeaxanthin glucosyltransferase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing zeaxanthin glucosyltransferase activity in a cell can be accomplished by modifying the amount of zeaxanthin glucosyltransferase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous zeaxanthin glucosyltransferase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type zeaxanthin glucosyltransferase such that the encoded modified or substituted zeaxanthin glucosyltransferase protein has a reduced enzyme activity.

Modification of a valencene synthase activity

Another component of terpene biosynthesis is the synthesis of valencene. A valencene synthase enzyme (also referred to as TPS1, (2E,6E)-farnesyl-diphosphate diphosphate-lyase (valencene-forming); EC 4.2.3.73) can catalyze the conversion of farnesyl diphosphate (farnesyl pyrophosphate) to valencene. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a valencene synthase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a valencene synthase, may be modified to decrease the amount and/or activity of a valencene synthase, or may be modified to alternately increase and decrease the amount and/or activity of a valencene synthase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a valencene synthase in a cell is increased. Increasing the amount and/or activity of a valencene synthase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target

product pathway (e.g., terpene production, valencene production, nootkatone production) and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, valencene synthase activity is unchanged in a host or engineered cell or organism. In some embodiments, the amount and/or activity of a host valencene synthase can be increased by increasing the number of copies of a nucleic acid encoding a valencene synthase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid); by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a valencene synthase; or by increasing the number of copies of a nucleic acid encoding a valencene synthase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a valencene synthase. In some embodiments, a valencene synthase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount valencene synthase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a valencene synthase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a valencene synthase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed. In some embodiments, a valencene synthase is not endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, a valencene synthase protein can be expressed in a cell by introducing heterologous nucleic acid encoding a valencene synthase into a cell or microorganism.

In some embodiments, a valencene synthase enzyme can be a fungal or bacterial protein. In a particular embodiment, the valencene synthase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*), *Chronobacter* (e.g., *C. sakazakii*), *Xanthophyllomyces* (e.g., *X. dendrorhous*), *Pantoea* (e.g., *P. ananatis*), or *Callitropsis* (e.g., *C. nootkatensis*) protein. An example of a *Callitropsis nootkatensis* nucleotide sequence (SEQ ID NO: 370) encoding a polypeptide (*TPS1*; SEQ ID NO: 352) having a valencene synthase activity is provided herein.

Presence, absence or amount of valencene synthase activity can be detected by any suitable method known in the art. For example, detection can be performed by using an appropriate valencene synthase activity assay. Nucleic acid sequences encoding native and/or modified

valencene synthase sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a nucleic acid (e.g., a heterologous nucleic acid) encoding a valencene synthase can be modified. For example, the amount of a valencene synthase protein expressed in a particular cellular location may be increased by including in the nucleic acid (e.g., heterologous nucleic acid) a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* oleate-induced peroxisomal protein (*POX18*) gene promoter. Another example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequences of a *Candida viswanathii* HDE gene promoter and a *Candida* POX18 gene promoter are provided herein as are examples of additional fatty acid-inducible promoters. In some embodiments, the promoter is modified by replacing the endogenous promoter with a promoter comprising an alkane response element (ARE1). Non-limiting examples of assays suitable for assessing induction of valencene synthase expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing valencene synthase activity in a cell can be accomplished by modifying the amount of valencene synthase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous valencene synthase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type valencene synthase such that the encoded modified or substituted valencene synthase protein has a reduced enzyme activity.

Modification of RAS2 activity

RAS2 is a guanine nucleotide-binding protein that can be activated by binding GTP (e.g., in the presence of glucose). In some embodiments of the microorganisms, compositions and methods

provided herein, the amount and/or activity of RAS2 in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase RAS2 and/or RAS2 activity, may be modified to decrease RAS2 and/or RAS2 activity, or may be modified to alternately increase and decrease RAS2 and/or RAS2 activity depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a RAS2 in a cell is decreased. Reducing or eliminating the amount and/or activity of a RAS2 may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway and away from other cellular metabolic pathways involving activated fatty acids. In certain embodiments, an endogenous microbial gene encoding RAS2 (e.g., yeast *RAS2* gene) can be disrupted or deleted in a host microorganism to reduce or eliminate RAS2 activity in the host relative to a microorganism in which the gene has not been modified. Methods for decreasing the amount and/or activity of RAS2 in a cell include, but are not limited to, modifying the amount of RAS2 protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous RAS2 gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting one or both copies of an endogenous gene, and/or replacing or modifying a gene encoding a wild-type RAS2 such that the encoded modified or substituted RAS2 protein has a reduced enzyme activity. For example, expression of a host RAS2 activity can be decreased or eliminated by disruption (e.g., knockout, insertion mutagenesis, the like and combinations thereof) of a host gene encoding the protein, or by decreasing the activity of the promoter (e.g., through addition of repressor sequences to the promoter or 5'UTR or replacing the promoter) that controls transcription of a RAS2 gene using recombinant molecular biology techniques known in the art and/or described herein. In one embodiment, a diploid yeast, such as, for example, a *Candida* yeast, when used as a host microorganism can be subjected to genetic modification in which one of the two alleles of a RAS2 gene is disrupted or deleted. In so doing, a single allele of the gene remains for a reduced amount of RAS2 expression in the microorganism and a reduced amount of the protein in the cell.

One method for disrupting an endogenous RAS2 gene is by recombinantly inserting a heterologous nucleic acid (e.g., a nucleotide sequence encoding a selectable marker such as an enzyme that restores an auxotrophic host organism to prototrophy) into the endogenous gene, thereby generating an engineered organism deficient in RAS2 activity. This can be done, for example, through homologous recombination in which a heterologous nucleic acid

containing sequences of an endogenous RAS2 gene and a disrupting sequence (e.g., a knock-out gene cassette such as described herein) is introduced into a host cell or microorganism. Nucleic acids encoding a RAS2 can be obtained from a number of sources, including, for example, yeast cells. Genomic DNA from cell sources can be amplified using oligonucleotide primers based on the nucleotide sequence of a RAS2 encoding gene, including examples provided herein.

Presence, absence or amount of RAS2 activity can be detected by any suitable method known in the art. For example, detection can be performed by using an appropriate RAS2 activity assay (e.g., enzymatic assays, PCR based assays (e.g., qPCR, RT-PCR), immunological detection methods (e.g., antibodies specific for RAS2), the like and combinations thereof). Nucleic acid sequences encoding native and/or modified RAS2 sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

Carbon-containing products

Certain organic acids and polyketides are chemical intermediates in manufacturing processes used to make polyamides, polyurethanes and plasticizers, all of which have wide applications in producing items such as antiseptics, carpets, elastomers, food packaging, lubricants, top-grade coatings, hot-melt coating and adhesives, painting materials, corrosion inhibitor, surfactant, engineering plastics and can also be used as a starting material in the manufacture of fragrances. Some large-scale synthetic processes for making organic acids and polyketides include the use of noxious chemicals and/or solvents, some require high temperatures, and all require significant energy input. In addition, some of the processes emit toxic byproducts (such as nitrous oxide) and give rise to environmental concerns. Furthermore, chemical synthesis and extraction of desirable chemical compounds, such as terpenes, for example, from natural sources yields low product levels and is often not economically feasible. Provided herein are methods for producing organic acid and other organic chemical intermediate target molecules using biological systems provided herein. Such production systems may have significantly less

environmental impact and could be economically competitive with synthetic manufacturing systems.

Organic Acids

Examples of organic acid target molecules that can be produced using compositions and methods provided herein include, but are not limited to, fatty acids, diacids and β-hydroxy acids (e.g., hydroxyalkanoate monomers) and salts and esters thereof. Fatty acids generally tend to be aliphatic acids of varying carbon chain lengths. Naturally occurring fatty acids in biological systems generally contain an even number of carbon atoms, typically between about 12 to about 24, or about 14 to about 24, and most commonly, 16 or 18 carbon atoms. Based on the number of carbons in a fatty acid carbon chain, it can be categorized as a short-, medium- or long-chain fatty acid. Generally, short-chain fatty acids have a chain length of about 2 to about 6 carbon atoms, medium-chain fatty acids have a chain length of about 8 to about 10 carbon atoms, long-chain fatty acids have a chain length of about 12 to about 20 carbon atoms and very long-chain length fatty acids have a chain length of about 22 or about 24 or more carbon atoms. The carbon atom bonds in the alkyl chain may all be single bonds (i.e., a saturated fatty acid) or may contain one or more double bonds (i.e., an unsaturated fatty acid). Unsaturated fatty acids having one double bond are also referred to as monoenoic; unsaturated fatty acids having two or more double bonds in the carbon chain are also referred to as polyenoic and polyunsaturated (PUFA). The carbon chain in a fatty acid may also be substituted with hydroxyl, methyl, or other groups in place of a hydrogen. Carboxylic acids, such as fatty acids, can partially dissociate in aqueous media and exist as undissociated, uncharged molecules and as a dissociated, anionic form.

Fatty acids containing one carboxyl group can also be referred to as monocarboxylic fatty acids. A fatty acid containing two carboxyl groups (e.g., α,ω-dicarboxylic acids) is a fatty dicarboxylic acid, also referred to herein as a diacid. An example of a diacid is adipic acid (hexanedioic acid) which contains six carbon atoms. A diacid sometimes is a C4 to a C24 diacid (i.e., a diacid containing 4 carbons to 24 carbons) and sometimes is a C8, C10, C12, C14, C16, C18, or C20 diacid. Diacids can contain an even as well as an odd number (e.g., C5, C7, C9, C11, C13, C15, C17, C19, C21 or C23) of carbons. A hydrocarbon portion of a diacid sometimes is fully saturated and sometimes a diacid includes one or more unsaturations (e.g., double bonds).

Non-limiting examples of diacids include octadecanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid and other organic intermediates using biological systems. Non-limiting examples of fatty dicarboxylic acids include adipic acid (hexanedioic acid, 1,4-butanedicarboxylic acid), suberic acid (i.e., octanedioic acid, 1,8-octanedioic acid, octanedioic acid, octane-1,8-dioic acid, 1,6-hexanedicarboxylic acid, capryllic diacids), azelaic acid, sebacic acid (i.e., 1,10-decanedioic acid, decanedioic acid, decane-1,10-dioic acid, 1,8-octanedicarboxylic acid, capric diacid), undecanedioc acid, dodecanedioic acid (i.e., DDDA, 1,12-dodecanedioic acid, dodecanedioic acid, dodecane-1,12-dioic acid, 1,10-decanedicarboxylic acid, decamethylenedicaboxylic acid, 1,10-dicarboxydecane, lauric diacid), tetradecanedioic acid (i.e., TDDA, 1,14-tetradecanedioic acid, tetradecanedioic acid, tetradecane-1,14-dioic acid, 1,12-dodecanedicarboxylic acid, myristic diacid), thapsic acid (i.e., hexadecanedioic acid, 1,16-hexadecanedioic acid, hexadecanedioic acid, hexadecane-1,16-dioic acid, 1,14-tetradecanedicarboxylic acid, palmitic diacid), cis-9-hexadecenedioic acid (i.e., palmitoleic diacids), octanedioic acid (i.e., 1,18-octadecanedioic acid, octadecanedioic acid, octadecane-1,18-dioic acid, 1,16-hexadecanedicarboxylic acid, stearic diacid), cis-9-octadecenedioic acid (i.e., oleic diacids), cis-9,12-octadecenedioic acid (i.e., linoleic diacids), cis-9,12,15-octadecenedioic acid (i.e., linolenic diacids), arachidic diacid (i.e., eicosanoic diacid, icosanoic diacid), 11-eicosenoic diacid (i.e., cis-11-eicosenedioic acid), 13-eicosenoic diacids (i.e., cis-13-eicosenedioic acid), arachidonic diacid (i.e., cis-5,8,11,14-eicosatetraenedioic acid) and salts and esters of fatty acids, including, for example, any of the foregoing diacids.

Adipic acid and suberic acid are 6- and 8-carbon dicarboxylic acids, respectively, that are chemical intermediates in manufacturing processes used to make certain polyamides, polyurethanes and plasticizers. Azelaic acid, a 9-carbon dicarboxylic acid, is also used in therapeutic compositions due to its antibacterial and keratolytic activities. Sebacic acid, a 10-carbon dicarboxylic acid, is also used in cosmetics and candles and as an intermediate in producing aromatics and antiseptics. Dodecandioic acid (DDDA), a 12-carbon dicarboxylic acid, is widely used in forming polyamides, such as nylon. Some large-scale industrial processes for making adipic acid include (i) liquid phase oxidation of ketone alcohol oil (KA oil); (ii) air oxidation/hydration of cyclohexane with boric acid to make cyclohexanol, followed by oxidation with nitric acid; and (iii) hydrocyanation of butadiene to a pentenenitrile mixture, followed by hydroisomerization of adiponitrile, followed by hydrogenation. Suberic acid can be synthetically manufactured by oxidation of cyclooctene with ozone oxygen or ozone H_2O_2. Methods of

chemical synthesis of sebacic acid include alkaline cleavage of ricinoleic acid and electrolytic dimerization of monomethyl adipate. DDDA is synthetically produced from butadiene in a multistep chemical process. These energy-requiring processes involve the use and/or production of toxic chemicals and/or solvents.

3-hydroxypropionic acid (3-HP or 3HP, used interchangeably herein, which collectively refers to 3-hydroxypropionic acid, a 3-hydroxypropionate salt or ester thereof, or mixtures thereof in any proportion) is a platform chemical that can be converted into a variety of valuable products such as poly(hydroxypropionate), 1,3-propanediol, ethyl 3-ethoxypropionate (EEP), malonic acid and acrylic acid. For example, 3-HP can be dehydrated to produce acrylic acid, which in turn can be esterified to produce methyl acrylate or aminated to produce acrylamide. Acrylamide can further be converted by dehydration to acrylonitrile, acrylonitrile can be condensed to produce adiponitrile and adiponitrile can be hydrolysed to produce hexamethylenediamine (HMDA). In addition, polymerized acrylic acid (with itself or with other monomers such as acrylamide, acrylonitrile, vinyl, styrene, or butadiene) can produce a variety of homopolymers and copolymers that are used in the manufacture of various plastics, coatings, adhesives, elastomers, latex applications, emulsions, leather finishings, and paper coating, as well as floor polishes and paints. Acrylic acid also can be used as a chemical intermediate for the production of acrylic esters such as ethyl acrylate, butyl acrylate, methyl acrylate, and 2-ethyl hexyl acrylate and superabsorbent polymers (glacial acrylic acid).

Polyketides

Polyketides are secondary metabolites polymerized from short-chain carboxylic acid units (e.g., acetate, proprionate, malonate and butyrate). Many polyketide-derived molecules are valuable pharmaceuticals such as antibiotics, antitumor agents and cholesterol-lowering drugs. Non-limiting examples of polyketides include triacetic acid lactone (TAL or 4-hydroxy-6-methyl-2-pyrone, used interchangeably herein) and 6-methylsalicylic acid (6-MSA or 2-hydroxy-6-methylbenzoic acid). TAL can be converted into end products such as sorbic acid and 1,3-pentadiene, and can serve as a precursor in the synthesis of compounds (e.g, phloroglucinol and resorcinol) used in production of resin and adhesive formulations.

Terpenes

Terpenes are made up of units of isoprene (C_5H_8) (isoprene, methylbuta-1,3-diene, hemiterpene are used interchangeably herein) that can be joined together in a variety of different combinations to generate thousands of terpene compounds. Terpenes can be categorized according to the number of isoprene units contained in the molecule: monoterpenes (2 isoprene units), sesquiterpenes (3 isoprene units), diterpenes (4 isoprene units), sesterterpenes (5 isoprene units), triterpenes (6 isoprene units), sesquaterpenes (7 isoprene units), tetraterpenes (8 isoprene units), polyterpenes (many isoprene units). The isoprene units can be joined "head-to-tail" in a linear chain or arranged in rings. Terpenes can be hydrocarbons or can contain other atoms, such as oxygen (e.g., alcohols, aldehydes and ketones) which are typically referred to as terpenoids. Terpenes are commercially valuable compounds with a variety of uses in the healthcare, food, cosmetics and chemical industries, including, but not limited to, uses as pharmaceuticals (e.g., anticancer and antimalarial drugs), nutraceuticals, supplements, antioxidants, fragrances, flavoring agents, food colorants and agricultural pest control agents. Included in the tetraterpenes are organic pigments (e.g., β-carotene and astaxanthin) referred to as carotenoids. These terpenes have many uses such as, for example, additives in food and feed stocks, precursors to vitamin A, antioxidants and supplements (e.g., lutein and lycopene).

Cellular carbon flux

Cells can obtain carbon atoms from external carbon sources such as, for example, carbohydrates, hydrocarbons, acids and alcohols. Upon entering the cell, the source molecule is metabolized through various chemical reactions depending on the carbon source, as well as other factors (e.g., oxygen and nutrient availability). Carbon atoms flow or flux through these metabolic reactions and are utilized in generating energy and in the production of cellular materials. The multiple series of coordinated reactions involved in metabolizing different carbon sources for different purposes are referred to as metabolic pathways and can be catabolic or anabolic. In catabolic pathways, the carbon source is broken down through oxidative reactions in which electrons are removed from substrates or intermediates, and, in the process, energy is generated and stored as adenosine triphosphate (ATP). Glycolysis is an example of a catabolic pathway in which a carbohydrate carbon source (e.g., glucose) is converted to pyruvate which is oxidatively decarboxylated to acetyl-CoA by the pyruvate dehydrogenase multi-enzyme complex in the mitochondria of eukaryotic cells and in the cytosol of prokaryotic cells. In this series of

reactions, a multi-carbon source molecule is degraded into a 2-carbon acetyl group and carbon dioxide. The 2 carbons of the acetyl group are then incorporated into a citrate molecule in the tricarboxylic acid cycle (also referred to as the TCA, citric acid and Krebs cycle) in which additional ATP molecules are generated. Intermediates in the TCA cycle (e.g., citrate, α-ketoglutarate, succinyl-CoA and oxaloacetate) provide precursors in the synthesis of essential cellular components such as amino acids, fatty acids, nucleotides and porphyrins. The TCA cycle is considered an amphibolic pathway which combines both catabolic and anabolic functions. In another catabolic pathway, β-oxidation, a fatty acid carbon source is broken down into acetyl-CoA and chain-shortened acyl-CoA which in turn can enter another cycle of β-oxidation for further degradation. The acetyl-CoA molecules generated in β-oxidation, which occurs in peroxisomes in eukaryotic microorganisms and the cytosol of prokaryotic cells, are then utilized in ATP generation in the TCA cycle.

Acetyl-CoA generated through β-oxidation in microorganisms (e.g., yeast and bacteria) can also be used in the glyoxylate cycle, which is an anabolic pathway wherein 2-carbon acetyl units are converted to 4-carbon molecules that can be used for the biosynthesis of macromolecules. The glyoxylate cycle thus allows these microorganisms to utilize non-fermentable carbon sources, such as fatty acids, acetate and ethanol, as a sole carbon source. In the glyoxylate cycle, which is similar to the TCA cycle, isocitrate is cleaved directly into the 4-carbon succinate molecule, and the 2-carbon glyoxylate molecule, through the enzyme isocitrate lyase without the two decarboxylation steps that occur in the same conversion in the TCA cycle. Glyoxylate then condenses with acetyl-CoA generated through β-oxidation to produce malate which in turn is converted to oxaloacetate and then isocitrate. Succinate generated in the glyoxylate cycle can also reenter the TCA cycle to produce oxaloacetate. Malate and oxaloacetate produced in the glyoxylate cycle can be converted into phosphoenolpyruvate, which is the product of the first enzyme-catalyzed reaction in gluconeogenesis. Gluconeogenesis is another anabolic pathway and provides for synthesis of carbohydrates when non-carbohydrate carbon sources are available to cells. Microorganisms growing on non-fermentable carbon sources utilize gluconeogenesis to synthesize glucose-6-phosphate which is used in the synthesis of ribonucleotides and deoxyribonucleotides. The carbon skeletons for generation of glucose-6-phosphate are contained within oxaloacetate from the glyoxylate and TCA cycles. In gluconeogenesis, oxaloacetate is converted into pyruvate through phosphoenolpyruvate carboxykinase, followed by several reactions that ultimately yield glucose-6-phosphate.

Additional anabolic pathways in cells include reactions in the synthesis of lipids, including, for example, triacylglycerols (referred to interchangeably as triglycerides and TAG) and phospholipids. Lipids are a diverse group of compounds that are soluble in non-polar organic solvents but not in water. Fatty acids serve as building blocks in the synthesis of storage lipids (e.g., triacylglycerols and steryl esters) and membrane lipids (e.g., phospholipids and sphingolipids). For example, triacylglycerol is an ester of glycerol and three fatty acids. In the synthesis of triacylglycerols from free fatty acids internalized into microbial cells from the environment, the fatty acids are first activated with coenzyme A to form an acyl-CoA. The acyl-CoA is involved in two pathways of triacylglycerol synthesis: the glycerol-3-phosphate (G3P) pathway and the dihydroxyacetone phosphate (DHAP) pathway. Both pathways proceed through formation of phosphatidic acid and subsequently diacylglycerol which is then acetylated to form triacylglycerol. Phosphatidic acid can also be converted to cytidine diphosphate-diacylglycerol which is the precursor of all major phospholipids in cells.

Modification of Cellular Carbon Flux

The multiple routes of carbon metabolism in cells provide opportunities for loss of carbon from a production pathway for desired organic molecules in a cell. Such losses can result in decreased product yields, increased production times and costs, and overall decreased production process efficiency and economy. Cells, organisms and microorganisms and methods described herein provide systems for enhanced production of target molecules. In one aspect, production is enhanced through alteration of carbon flux in cell-based and microbial production systems. Through alteration of cellular carbon flux, carbon atoms that may have flowed or been transported into other metabolic processes (e.g., energy and/or cellular composition generation) in a cell are redirected and/or recycled and made available for use in organic target molecule production processes. In so doing, starting material loss is reduced and carbon sources are utilized to a fuller extent in the production of the desired molecules.

Acetyl coenzyme A

Acetyl coenzyme A (acetyl-CoA; used interchangeably herein) is a major precursor in cell-based or microbial production of many industrially important chemicals. The fatty acid biosynthesis pathway begins with the conversion of acetyl-CoA to malonyl-CoA. Similarly, organic acids, such as, for example 3-hydroxypropionic acid, and polyketides, such as triacetic acid lactone,

can be synthesized using acetyl-CoA as a starting material. Additional high-value products that can be synthesized in reactions beginning with acetyl-CoA include terpenes, which can be generated from isopentenyl diphosphate produced through microbial mevalonate pathways in cells. Acetyl-CoA is formed from an acetyl group and coenzyme A (a derivative of pantothenate and cysteine) which are linked through a thioester bond. Acetyl-CoA is a central metabolite in carbon metabolism. It is the final carbon form resulting from the catabolism of external carbon sources and is the initial precursor carbon form in many of the cellular anabolic pathways and energy generation processes. Acetyl-CoA is formed in multiple locations of a eukaryotic cell depending on the metabolic pathway and/or carbon source through which it is generated. For example, acetyl-CoA generated through glycolysis is localized in the mitochondria, whereas acetyl-CoA generated through peroxisomal β-oxidation is localized to peroxisomes. Acetyl-CoA generated through metabolism of acetate or ethanol is localized to the cytoplasm. Typically, acetyl-CoA formed in any of these cellular locations is transferred to the mitochondrial matrix for use in the TCA cycle for the generation of energy and precursors of cellular constituents, although some acetyl-CoA localized to the cytoplasm can be used in the synthesis of oxaloacetate via initial conversion to malate.

Acetyl-CoA, due to its chemical nature, cannot freely cross biological membranes. Therefore, acetyl-CoA formed in peroxisomes and the cytoplasm is modified for transport to the mitochondria. The primary mechanism for transfer of acetyl-CoA into the mitochondria in eukaryotic cells is the carnitine shuttle in which the acetyl group of acetyl-CoA is reversibly linked to a carrier molecule, carnitine, which is able to traverse biological membranes. Acetyl-carnitine is generated and degraded by the action of carnitine acetyltransferases (e.g., EC 2.3.1.7). Peroxisomal acetyl-CoA not destined for the glyoxylate cycle is converted to acetyl-carnitine by carnitine O-acetyltransferase. Due to its smaller size compared to acetyl-CoA, acetyl-carnitine is able to diffuse through pores in the peroxisomal membrane, across the cytoplasm to mitochondria where it is converted back to acetyl-CoA by mitochondrial carnitine O-acetyltransferase. Yeast also have carnitine acetyltransferases that localize to the cytosol and/or to the outer mitochondrial membrane which, in some species, are encoded by *YAT1* genes. The enzymes encoded by these genes may convert cytosolic acetyl-carnitine to acetyl-CoA and carnitine.

Acetyl-carnitine uptake into mitochondria involves an acetyl-carnitine translocase which, in some yeast species, is encoded by a *CRC1* gene. Mitochondria possess two membranes with

the outer membrane allowing free diffusion of metabolites and the inner membrane controlling metabolite transport with multiple membrane transport proteins. A mitochondrial inner-membrane transport protein (e.g., Crc1p) may function as an acetyl-carnitine transporter providing for transport of acetyl-carnitine into the mitochondrial matrix.

Certain aspects of the cells, microorganisms, compositions and methods provided herein involve cellular carbon flux modifications to capture the carbon atoms in the acetyl group of acetyl-CoA formed in cellular metabolism. In some embodiments, carbon flux is modified to capture acetyl group carbon atoms generated in organelles (e.g., peroxisomes) or membranes as they are transported through the cytosol in the form of acetyl-carnitine. In other embodiments, acetyl group carbons of organelle-generated acetyl-CoA are re-directed from the carnitine-assisted transport system and toward conversion to acetate. Acetate, unlike acetyl-CoA, is able to traverse membranes and enter the cytosol from organelles. In further embodiments, carbon atoms of acetyl groups in mitochondrial acetyl-CoA can also be captured from intermediates of the TCA cycle that move into the cytosol.

Acetyl-carnitine capture/conversion

Included in embodiments of the cells, microorganisms, compositions and methods provided herein are cell-based or microbial production platform systems and components thereof in which the amount of (a) acetyl-carnitine in the cell cytosol is modified and/or (b) carnitine acetyltransferase and/or carnitine acetyltransferase activity in the cell cytosol is/are modified. In some instances, the amount of (a) acetyl-carnitine in the cell cytosol is increased and/or decreased and/or (b) carnitine acetyltransferase and/or carnitine acetyltransferase activity in the cell cytosol is/are increased and/or decreased. For example, in some aspects, a cell or microorganism may be modified to increase cytosolic acetyl-carnitine, may be modified to decrease cytosolic acetyl-carnitine or may be modified to alternately increase and decrease cytosolic acetyl-carnitine depending on the conditions in which the modified cell or microorganism is cultured.

In one embodiment, the amount of acetyl group carbons in the cytosol in the form of acetyl-carnitine in transit from the peroxisome and other areas to the mitochondria is increased in a cell or microorganism through a reduction in, and/or slowing of, the entry of acetyl-carnitine into mitochondria from the cytosol. This provides an increased availability of substrate for cytosolic

carnitine acetyltransferase to convert to acetyl-CoA, and effectively results in an increase in the generation of cytosolic acetyl-CoA. Some of the acetyl group carbons are thereby diverted from the mitochondria, and from utilization in metabolic processes therein, and are instead retained in the cytosol. In some embodiments, the amount and/or activity of carnitine acetyltransferase in the cytosol of a cell or microorganism is/are increased. This provides an increased conversion of acetyl-carnitine, such as that which is in transit from the peroxisome to the mitochondria, into acetyl-CoA in the cytosol. In some embodiments, the entry of acetyl-carnitine into mitochondria from the cytosol is reduced in a cell or microorganism, and the amount and/or activity of carnitine acetyltransferase in the cytosol of the cell or microorganism is/are increased.

Acetate capture/conversion

Included in embodiments of the microorganisms, compositions and methods provided herein are microbial production platform systems and components thereof in which acetyl group carbons of organelle-generated acetyl-CoA are directed toward conversion to acetate. Modification of carbon flux in this manner provides for a tight and precise control of the movement of the acetyl carbons because acetate may pass through some membranes, e.g., peroxisomal membranes, more readily than other membranes, e.g., mitochondrial inner membranes. Therefore, carbon atoms captured in the form of cytosolic acetate will be less readily transported into mitochondria for further metabolism, unlike carbon atoms captured in the form of cytosolic acetyl-carnitine. Provided herein are cells, microorganisms, compositions and methods in which cellular carbon flux has been modified through the altered de novo generation of cellular acetate. In particular embodiments, cellular carbon flux has been modified to increase the production of acetate in a cell and/or a particular cellular location. For example, in certain aspects, cells are modified to increase the production of acetate in peroxisomes. In some embodiments, acetyl group carbons are directed toward conversion to acetate and away from the carnitine-carrier transport system.

In embodiments in which modification of cellular acetate generation yields increased amounts of cytosolic acetate, the amount and/or activity of cytosolic acetyl-CoA synthetase (also referred to as ACS or acetate-CoA ligase and used interchangeably herein) can also be increased to provide for increased conversion of acetate to acetyl-CoA. For example, the genomic copy number of nucleic acids encoding acetyl-CoA synthetase can be increased and/or the promoter

for the acetyl-CoA synthetase-encoding nucleic acid can be replaced with a stronger promoter or one that provides for a different pattern of expression in the cell or microorganism.

Citrate capture/conversion

Carbon atoms of acetyl groups in mitochondrial acetyl-CoA can also be captured from intermediates of the TCA cycle such as, for example, citrate molecules generated in the first step of the cycle through the citrate synthase-catalyzed condensation of acetyl-CoA and oxaloacetate. In another embodiment of the cell and microbial production systems and methods provided herein, carbon atoms incorporated into citrate that has been transferred to the cytosol are captured through the cleavage of citrate to oxaloacetate and acetyl-CoA by the enzyme ATP citrate lyase (i.e., ACL, used interchangeably herein). The capture of metabolite carbon in this manner diverts it from use in other metabolic processes and also results in an increase in the level cytoplasmic acetyl-CoA.

Acyl coenzyme A

Acyl-CoA synthetases (e.g., EC 6.2.1.3) are enzymes that catalyze the activation of free fatty acids in the cytoplasm into CoA esters (fatty acyl-CoA) which are involved in several metabolic pathways. For example, free fatty acids internalized into cells that become activated with coenzyme A to form an acyl-CoA are used in the synthesis of triacylglycerols via two pathways: the glycerol-3-phosphate (G3P) pathway and the dihydroxyacetone phosphate (DHAP) pathway. When free fatty acids are activated and used in cellular processes, such as lipid biosynthesis, the carbon atoms in the free fatty acids are not available for use in cell or microbial production of commercially important chemicals. Certain aspects of the cells, microorganisms, compositions and methods provided herein include one or more modifications to reduce or eliminate cytosolic activation of free fatty acids into acyl-CoA. An example of a modification to reduce or eliminate activation of cytosolic free fatty acids is reduction or elimination of the amount and/or activity of acyl-CoA synthetase in the cytoplasm.

Malonyl-CoA

Malonyl-CoA is a coenzyme A derivative of the dicarboxylic acid malonic acid that can serve as a precursor in the synthesis of numerous valuable organic molecules, including fatty acids and

polyketides. Cytoplasmic acetyl-CoA can be converted to malonyl-CoA by the enzyme acetyl-CoA carboxylase (e.g., EC 6.4.1.2). A modification of cellular carbon flux that increases cytosolic acetyl-CoA alone may not be optimal for enhancing fatty acid or other organic acid production in an engineered cell or microbial system if there is not a concurrent increase in conversion of acetyl-CoA to malonyl-CoA. To maximize production efficiencies, included in the cells, microorganisms, compositions and methods provided herein are cellular carbon flux modifications that increase the amount of cytosolic malonyl-CoA.

Cells and organisms

Provided herein are modified cells and organisms. In particular embodiments, the modified cells and organisms have been manipulated in ways designed to alter the cellular flux of carbon to direct carbon atoms toward one or more biochemical events or cellular locations and/or away from other metabolic pathways or locations. Also provided herein are methods of producing modified cells and organisms.

Host cells and organisms

Modified cells and organisms provided herein can be generated by manipulation of an existing cell or organism. The terms “host,” “starting” or “parental” as used herein in reference to a cell or organism refers to such an existing cell or organism. Host cells and organisms include, for example, wild-type or native cells or organisms as they occur in nature in their genetically unmodified, predominant form, and mutant cells or organisms that have one or more genetic differences compared to a wild-type cell or organism. A host cell or organism can also be a cell or organism that has been genetically modified. A host cell or organism thus serves as a reference cell or organism with respect to a modified or engineered cell or organism obtained by manipulation of a host. Organisms or cells that can be used as host organisms or cells, or as a source for a nucleic acid, are publicly available, from, for example, American Type Culture Collection (Manassas, Virginia), Thermo Fisher Scientific (Waltham, MA) and Agricultural Research Culture Collection (NRRL; Peoria, Illinois).

Host or modified organisms include multicellular and single cell, or unicellular, organisms. Microscopic organisms, referred to interchangeably herein as a “microorganism,” “microbial cell” or “microbe,” are an example of a host or modified organism and are included in the term

"organism." Many microorganisms are unicellular and often are capable of dividing and proliferating. Cells from non-microbial organisms can also be utilized as a host or modified organism or source for a heterologous polynucleotide.

Organisms can be prokaryotic (e.g., bacteria) and non-prokaryotic (e.g., eukaryotic). Examples of eukaryotic organisms include yeast, filamentous fungi, protists, plants, algae and amoeba. An organism or microorganism can include one or more of the following features: aerobe, anaerobe, filamentous, non-filamentous, monoploid, haploid, diploid, oleaginous, non-oleaginous, auxotrophic and/or non-auxotrophic.

Host cells or organisms or modified cells or organisms can be selected based on a variety of criteria depending, for example, on the methods of generating modified cells or organisms therefrom and the uses of the modified cells or organisms from which they are derived. Selection criteria can include inherent metabolic mechanisms, suitability for genetic manipulation, adaptability to a variety of or particular growth or culture conditions, and ease of large-scale maintenance for use in industrial production processes. For example, microorganisms often can be cultured at cell densities useful for industrial production of a target product, including in a fermentation device. Included among microorganisms that may be selected as a host or modified organism or source for a heterologous polynucleotide are fungi. Examples of fungi include, but are not limited to, yeast, *Aspergillus* fungi (e.g., *A. parasiticus*, *A. nidulans*), *Thraustochytrium* fungi, *Schizochytrium* fungi and *Rhizopus* fungi (e.g., *R. arrhizus*, *R. oryzae*, *R. nigricans*). In some embodiments, a host organism can be a fungus such as a yeast strain, an *A. parasiticus* strain that includes, but is not limited to, strain ATCC 24690, and in certain embodiments, a fungus is an *A. nidulans* strain that includes, but is not limited to, strain ATCC 38163.

In some embodiments, a modified cell or microorganism provided herein can be derived from any one of the following cell lines: ATCC 20362, ATCC 8862, ATCC 18944, ATCC 20228, ATCC 76982, LGAM S(7)1, ATCC 20336, ATCC 20913, SU-2 (ura3-/ura3-), ATCC 20962, ATCC 24690, ATCC 38164, ATCC 38163, H5343, ATCC 8661, ATCC 8662, ATCC 9773, ATCC 15586, ATCC 16617, ATCC 16618, ATCC 18942, ATCC 18943, ATCC 18944, ATCC 18945, ATCC 20114, ATCC 20177, ATCC 20182, ATCC 20225, ATCC 20226, ATCC 20228, ATCC 20237, ATCC 20255, ATCC 20287, ATCC 20297, ATCC 20306, ATCC 20315, ATCC 20320, ATCC 20324, ATCC 20341, ATCC 20346, ATCC 20348, ATCC 20362, ATCC 20363,

ATCC 20364, ATCC 20372, ATCC 20373, ATCC 20383, ATCC 20390, ATCC 20400, ATCC 20460, ATCC 20461, ATCC 20462, ATCC 20496, ATCC 20510, ATCC 20628, ATCC 20688, ATCC 20774, ATCC 20775, ATCC 20776, ATCC 20777, ATCC 20778, ATCC 20779, ATCC 20780, ATCC 20781, ATCC 20794, ATCC 20795, ATCC 20875, ATCC 22421, ATCC 22422, ATCC 22423, ATCC 22969, ATCC 32338, ATCC 32339, ATCC 32340, ATCC 32341, ATCC 32342, ATCC 32343, ATCC 32935, ATCC 34017, ATCC 34018, ATCC 34088, ATCC 34922, ATCC 38295, ATCC 42281, ATCC 44601, ATCC 46025, ATCC 46026, ATCC 46027, ATCC 46028, ATCC 46067, ATCC 46068, ATCC 46069, ATCC 46070, ATCC 46330, ATCC 46482, ATCC 46483, ATCC 46484, ATCC 48436, ATCC 60594, ATCC 62385, ATCC 64042, ATCC 74234, ATCC 76598, ATCC 76861, ATCC 76862, ATCC 90716, ATCC 90806, ATCC 90811, ATCC 90812, ATCC 90813, ATCC 90814, ATCC 90903, ATCC 90904, ATCC 90905, ATCC 96028, ATCC 201089, ATCC 201241, ATCC 201242, ATCC 201243, ATCC 201244, ATCC 201245, ATCC 201246, ATCC 201247, ATCC 201248, ATCC 201249, ATCC 201847, ATCC MYA-165, ATCC MYA-166, ATCC MYA-2613, and ATCC MYA-4467. That is, in certain embodiments, an engineered cell or microorganism described herein can be generated from one or more of the aforementioned ancestral cell lines.

Yeast that can serve as a host organism, and that can be modified organisms, include, but are not limited to, ascomycetes, non-Saccharomyces ascomycetes, and basidiomycetes. Non-limiting examples of yeast include *Yarrowia* yeast (e.g., *Y. lipolytica* (formerly classified as *Candida lipolytica*)), *Candida* yeast (e.g., *C. revkaufi*, *C. viswanathii*, *C. pulcherrima*, *C. tropicalis*, *C. utilis*), *Blastobotrys* (formerly classified as *Arxula*) (e.g., *Blastobotrys adeninivorans* (formerly classified as *Arxula adeninivorans*), *Blastobotrys mokoenaii*), *Rhodotorula* yeast (e.g., *R. glutinus*, *R. graminis*), *Rhodosporidium* yeast (e.g., *R. toruloides*), *Saccharomyces* yeast (e.g., *S. cerevisiae*, *S. bayanus*, *S. pastorianus*, *S. carlsbergensis*), *Cryptococcus* yeast, *Trichosporon* yeast (e.g., *T. pullans*, *T. cutaneum*), *Pichia* yeast (e.g., *P. pastoris*) and *Lipomyces* yeast (e.g., *L. starkeyii*, *L. lipoferus*). In some embodiments, a suitable yeast is of the genus *Arachniotus*, *Aspergillus*, *Aureobasidium*, *Auxarthron*, *Blastobotrys*, *Blastomyces*, *Candida*, *Chrysosporuim*, *Debaryomyces*, *Coccidiodes*, *Cryptococcus*, *Gymnoascus*, *Hansenula*, *Histoplasma*, *Issatchenkia, Kluyveromyces*, *Lipomyces*, *Lssatchenkia*, *Microsporum*, *Myxotrichum*, *Myxozyma*, *Oidiodendron*, *Pachysolen*, *Penicillium*, *Pichia*, *Rhodosporidium*, *Rhodotorula*, *Saccharomyces*, *Schizosaccharomyces*, *Scopulariopsis*, *Sepedonium*, *Trichosporon*, or *Yarrowia*. In some embodiments, a suitable yeast is of the species *Arachniotus flavoluteus*, *Aspergillus flavus*, *Aspergillus fumigatus*, *Aspergillus niger*,

Aureobasidium pullulans, *Auxarthron thaxteri*, *Blastobotrys adeninivorans*, *Blastomyces dermatitidis*, *Candida albicans*, *Candida dubliniensis*, *Candida famata*, *Candida glabrata*, *Candida guilliermondii*, *Candida kefyr*, *Candida krusei*, *Candida lambica*, *Candida lipolytica*, *Candida lustitaniae*, *Candida parapsilosis*, *Candida pulcherrima*, *Candida revkaufi*, *Candida rugosa*, *Candida tropicalis*, *Candida utilis*, *Candida viswanathii*, *Candida xestobii*, *Chrysosporuim keratinophilum*, *Coccidiodes immitis*, *Cryptococcus albidus* var. *diffluens*, *Cryptococcus laurentii*, *Cryptococcus neofomans*, *Debaryomyces hansenii*, *Gymnoascus dugwayensis*, *Hansenula anomala*, *Histoplasma capsulatum*, *Issatchenkia occidentalis*, *Isstachenkia orientalis*, *Kluyveromyces lactis*, *Kluyveromyces marxianus*, *Kluyveromyces thermotolerans*, *Kluyveromyces waltii*, *Lipomyces lipoferus*, *Lipomyces starkeyii*, *Microsporum gypseum*, *Myxotrichum deflexum*, *Oidiodendron echinulatum*, *Pachysolen tannophilis*, *Penicillium notatum*, *Pichia anomala*, *Pichia pastoris*, *Pichia stipitis*, *Rhodosporidium toruloides*, *Rhodotorula glutinus*, *Rhodotorula graminis*, *Saccharomyces cerevisiae*, *Saccharomyces kluyveri*, *Schizosaccharomyces pombe*, *Scopulariopsis acremonium*, *Sepedonium chrysospermum*, *Trichosporon cutaneum*, *Trichosporon pullans*, *Yarrowia lipolytica*, or *Yarrowia lipolytica* (formerly classified as *Candida lipolytica*). In some embodiments, a yeast is a *Y. lipolytica* strain that includes, but is not limited to, ATCC 20362, ATCC 8862, ATCC 18944, ATCC 20228, ATCC 76982 and LGAM S(7)1 strains (Papanikolaou S., and Aggelis G., *Bioresour. Technol*. 82(1):43-9 (2002)).

In certain embodiments, a yeast is a *C. tropicalis* strain, a *C. viswanathii* strain, a *Y. lipolytica* strain or a yeast strain that includes, but is not limited to, ATCC 20336, ATCC 20913, SU-2 (ura3-/ura3-), ATCC 20962, H5343 (beta oxidation blocked; US Patent No. 5648247) ATCC 20362, ATCC 8862, ATCC 18944, ATCC 20228, ATCC 76982, LGAM S(7)1, ATCC 8661, ATCC 8662, ATCC 9773, ATCC 15586, ATCC 16617, ATCC 16618, ATCC 18942, ATCC 18943, ATCC 18944, ATCC 18945, ATCC 20114, ATCC 20177, ATCC 20182, ATCC 20225, ATCC 20226, ATCC 20228, ATCC 20237, ATCC 20255, ATCC 20287, ATCC 20297, ATCC 20306, ATCC 20315, ATCC 20320, ATCC 20324, ATCC 20341, ATCC 20346, ATCC 20348, ATCC 20362, ATCC 20363, ATCC 20364, ATCC 20372, ATCC 20373, ATCC 20383, ATCC 20390, ATCC 20400, ATCC 20460, ATCC 20461, ATCC 20462, ATCC 20496, ATCC 20510, ATCC 20628, ATCC 20688, ATCC 20774, ATCC 20775, ATCC 20776, ATCC 20777, ATCC 20778, ATCC 20779, ATCC 20780, ATCC 20781, ATCC 20794, ATCC 20795, ATCC 20875, ATCC 22421, ATCC 22422, ATCC 22423, ATCC 22969, ATCC 32338, ATCC 32339, ATCC 32340, ATCC 32341, ATCC 32342, ATCC 32343, ATCC 32935, ATCC 34017, ATCC 34018,

ATCC 34088, ATCC 34922, ATCC 38295, ATCC 42281, ATCC 44601, ATCC 46025, ATCC 46026, ATCC 46027, ATCC 46028, ATCC 46067, ATCC 46068, ATCC 46069, ATCC 46070, ATCC 46330, ATCC 46482, ATCC 46483, ATCC 46484, ATCC 48436, ATCC 60594, ATCC 62385, ATCC 64042, ATCC 74234, ATCC 76598, ATCC 76861, ATCC 76862, ATCC 90716, ATCC 90806, ATCC 90811, ATCC 90812, ATCC 90813, ATCC 90814, ATCC 90903, ATCC 90904, ATCC 90905, ATCC 96028, ATCC 201089, ATCC 201241, ATCC 201242, ATCC 201243, ATCC 201244, ATCC 201245, ATCC 201246, ATCC 201247, ATCC 201248, ATCC 201249, ATCC 201847, ATCC MYA-165, ATCC MYA-166, ATCC MYA-2613, and ATCC MYA-4467.

In certain embodiments, a yeast is a *Candida* species (i.e., *Candida* spp.) yeast. In some embodiments, suitable *Candida* species include, but are not limited to *Candida albicans*, *Candida dubliniensis*, *Candida famata*, *Candida glabrata*, *Candida guilliermondii*, *Candida kefyr*, *Candida krusei*, *Candida lambica*, *Candida lipolytica*, *Candida lustitaniae*, *Candida parapsilosis*, *Candida pulcherrima*, *Candida revkaufi*, *Candida rugosa*, *Candida tropicalis*, *Candida utilis*, *Candida viswanathii*, *Candida xestobii* and any other *Candida* spp. yeast described herein. Non-limiting examples of *Candida* spp. strains include, but are not limited to, sAA001 (ATCC 20336), sAA002 (ATCC 20913), sAA003 (ATCC 20962), sAA496 (US2012/0077252), sAA106 (US2012/0077252), SU-2 (ura3-/ura3-), H5343 (beta oxidation blocked; US Patent No. 5648247) strains. Any suitable strains from *Candida* spp. yeast may be utilized as parental strains for modification.

Examples of ascomycetes fungi include, but are not limited to, *Candida* spp., *Yarrowia* spp., *Blastobotrys* spp., *Aspergillus* spp., *Penicillium* spp., *Saccharomyces* spp., *Debaryomyces* spp., *Lipomyces* spp., *Fusarium* spp., *Paecilomyces* spp., *Trichoderma* spp., *Cladosporium* spp., *Pichia* spp., and *Neurospora* spp. Examples of basidiomycetes fungi include, but are not limited to, *Trichosporon* spp., *Rhodotorula* spp., *Rhodosporidium* spp., *Cryptococcus* spp., *Phaffia* spp., and *Xanthophyllomyces* spp.

Prokaryote organisms that can serve as host organisms, and that can be modified organisms, include, for example, Gram negative or Gram positive bacteria. Examples of bacteria include, but are not limited to, *Bacillus* (e.g., *B. subtilis*, *B. megaterium*), *Acinetobacter*, *Norcardia*, *Xanthobacter*, *Escherichia* (e.g., *E. coli* (e.g., strains DH10B, Stbl2, DH5-alpha, DB3, DB3.1), DB4, DB5, JDP682 and ccdA-over (e.g., U.S. Application No. 09/518,188))), *Streptomyces*,

Erwinia, *Klebsiella*, *Serratia* (e.g., *S. marcessans*), *Pseudomonas* (e.g., P. aeruginosa), *Salmonella* (e.g., *S. typhimurium*, *S. typhi*), *Megasphaera* (e.g., *Megasphaera elsdenii*). Bacteria also include, but are not limited to, photosynthetic bacteria (e.g., green non-sulfur bacteria (e.g., *Choroflexus* (e.g., *C. aurantiacus*), *Chloronema* (e.g., *C. gigateum*)), green sulfur bacteria (e.g., *Chlorobium* bacteria (e.g., *C. limicola*), *Pelodictyon* (e.g., *P. luteolum*), purple sulfur bacteria (e.g., *Chromatium* (e.g., *C. okenii*)), and purple non-sulfur bacteria (e.g., *Rhodospirillum* (e.g., *R. rubrum*), *Rhodobacter* (e.g., *R. sphaeroides*, *R. capsulatus*), and *Rhodomicrobium* (e.g., *R. vanellii*)).

Examples of cells from non-microbial organisms that can be utilized as a host cell or organism, engineered cell or organism or source for a heterologous polynucleotide include, but are not limited to, insect cells (e.g., *Drosophila* (e.g., *D. melanogaster*), *Spodoptera* (e.g., *S. frugiperda* Sf9 or Sf21 cells) and *Trichoplusa* (e.g., High-Five cells); nematode cells (e.g., *C. elegans* cells); avian cells; amphibian cells (e.g., *Xenopus laevis* cells); reptilian cells; mammalian cells (e.g., NIH3T3, 293, CHO, COS, VERO, C127, BHK, Per-C6, Bowes melanoma and HeLa cells); and plant cells (e.g., *Arabidopsis thaliana*, *Nicotania tabacum*, *Cuphea acinifolia*, *Cuphea aequipetala*, *Cuphea angustifolia*, *Cuphea appendiculata*, *Cuphea avigera*, *Cuphea avigera* var. pulcherrima, *Cuphea axilliflora*, *Cuphea bahiensis*, *Cuphea baillonis*, *Cuphea brachypoda*, *Cuphea bustamanta*, *Cuphea calcarata*, *Cuphea calophylla*, *Cuphea calophylla* subsp. *mesostemon*, *Cuphea carthagenensis*, *Cuphea circaeoides*, *Cuphea confertiflora*, *Cuphea cordata*, *Cuphea crassiflora*, *Cuphea cyanea*, *Cuphea decandra*, *Cuphea denticulata*, *Cuphea disperma*, *Cuphea epilobiifolia*, *Cuphea ericoides*, *Cuphea flava*, *Cuphea flavisetula*, *Cuphea fuchsiifolia*, *Cuphea gaumeri*, *Cuphea glutinosa*, *Cuphea heterophylla*, *Cuphea hookeriana*, *Cuphea hyssopifolia* (Mexican-heather), *Cuphea hyssopoides*, *Cuphea ignea*, *Cuphea ingrata*, *Cuphea jorullensis*, *Cuphea lanceolata*, *Cuphea linarioides*, *Cuphea llavea*, *Cuphea lophostoma*, *Cuphea lutea*, *Cuphea lutescens*, *Cuphea melanium*, *Cuphea melvilla*, *Cuphea micrantha*, *Cuphea micropetala*, *Cuphea mimuloides*, *Cuphea nitidula*, *Cuphea palustris*, *Cuphea parsonsia*, *Cuphea pascuorum*, *Cuphea paucipetala*, *Cuphea procumbens*, *Cuphea pseudosilene*, *Cuphea pseudovaccinium*, *Cuphea pulchra*, *Cuphea racemosa*, *Cuphea repens*, *Cuphea salicifolia*, *Cuphea salvadorensis*, *Cuphea schumannii*, *Cuphea sessiliflora*, *Cuphea sessilifolia*, *Cuphea setosa*, *Cuphea spectabilis*, *Cuphea spermacoce*, *Cuphea splendida*, *Cuphea splendida* var. viridiflava, *Cuphea strigulosa*, *Cuphea subuligera*, *Cuphea teleandra*, *Cuphea thymoides*, *Cuphea tolucana*, *Cuphea urens*, *Cuphea utriculosa*, *Cuphea viscosissima*, *Cuphea watsoniana*, *Cuphea wrightii*, *Cuphea lanceolata*).

In some embodiments, host organisms, or modified organisms, can be hydrocarbon-utilizing (e.g. alkane-utilizing), fatty acid-utilizing and/or fatty alcohol-utilizing microorganisms. These organisms are able to assimilate hydrocarbons, fatty acids and/or fatty alcohols for energy and biomass generation. Many of these organisms are able to utilize hydrocarbons, fatty acids and/or fatty alcohols as a sole carbon source. Some examples of hydrocarbon-, fatty acid- and/or fatty alcohol-utilizing microorganisms include some species of fungi (including, e.g., yeast), bacteria and algae. Non-limiting examples of such organisms include *Yarrowia* (e.g., *Y. lipolytica* (formerly classified as *Candida lipolytica*)), *Candida* (e.g., *C. apicola*, *C. maltosa*, *C. tropicalis*, *C. utilis*, *C. viswanathii*, *C. catenulate*, *C. rugose*, *C. vini*, *C. entamophila*, *C. intermedia*), *Aspergillus* (e.g., *A. niger*, *A. versicolor*, *A. ustus*, *A. fumigatus*, *A. oryzae*, *A. flavus*, *A. ficuum*, *A. terricola*, *A. japonicas*, *A. wentii*, *A. clavatus*, *A. terreus*), *Penicillium* (*P. cyclopium*, *P. chrysogenum*, P. *italicum*), *Fusarium* (e.g., *F. oxysporum*, *F. moniliforme*, *F. solani*), *Paecilomyces* (e.g., *Paec. lilacinus*), *Trichoderma* (e.g., *T. koningii*, *T. viride*, *T. virens*), *Cladosporium* (e.g., *C. herbarum*), *Stachybotrys*, *Trichosporon* (e.g., *T. veenhuisii*, *T. asahii*, *T. jirovecii*, *T. monteviblankiideense*), *Rhodotorula* (e.g., *R. glutinous*, *R. mucilaginosa*), *Rhodosporidium* (e.g., *R. toruloides*), *Cryptococcus* (e.g., *C. neoformans*, *C. albidus*), *Pichia* (e.g., *P. farinosa*, *P. stipitis*), *Debaryomyces* (e.g., *D. hansenii*), *Blastobotrys* (e.g., *Blastobotrys adeninivorans*), *Saccharomyces* (e.g., *S. cerevisiae*, *S. bayanus*, *S. pastorianus*, *S. carlsbergensis*), *Lipomyces* (e.g., *L. starkeyii*, *L. lipoferus*) and *Chlorella* algae (e.g., *Chlorella protothecoides*).

In some embodiments, a host organism or modified organism can be an oleaginous organism (e.g., an oleaginous microorganism). As used herein, an “oleaginous” organism is an organism capable of accumulating at least about 20% or more of its cell mass (by dry weight) as intracellular lipids (e.g., oil). In oleaginous organisms, a significant carbon flux towards lipid synthesis occurs and is enhanced under certain conditions (e.g., limited supply of nitrogen). These lipid-accumulating organisms can be characterized by the endogenous expression of cytosolic ATP citrate lyase, which catalyzes the degradation of citrate generated in the TCA cycle into acetyl-CoA and oxaloacetate, and/or a dependence on AMP concentration for the activity of isocitrate dehydrogenase in the TCA cycle. Generally, under certain conditions (e.g., limited nitrogen), AMP deaminase is activated in oleaginous yeast which can lead to a decrease in mitochondrial AMP concentration and isocitrate dehydrogenase activity. This, in turn, can cause an accumulation of mitochondrial citrate from the TCA cycle which is then exported to the

cytosol and can serve as substrate for ATP citrate lyase. The acetyl-CoA that may be generated through the action of ATP citrate lyase can be used in synthesizing fatty acyl-CoA that can be converted into lipids which may be stored in lipid bodies in the cells. A "non-oleaginous" organism, as used herein, is an organism that is not capable of accumulating at least about 20% or more of its cell mass (by dry weight) as intracellular lipids. In some embodiments, a host organism or modified organism can be a non-oleaginous organism.

Oleaginous microorganisms include species of fungi, bacteria and algae. Examples of oleaginous fungi include, but are not limited to, *Blastobotrys* (e.g., *Blastobotrys adeninivorans*), *Yarrowia* (e.g., *Y. lipolytica*), *Trichosporon* (e.g., *T. fermentans*, *T. porosum*, *T. pullulan*), *Rhodotorula* (e.g., *R. graminis*, *R. glutinous*, *R. araucariae*, *R. minuta*, *R. bogoriensis*, *R. mucilaginosa*, *R. colostri*), *Rhodosporidium* (e.g., *R. toruloides*, *R. kractochvilovae*, *R. paludigenum*, *R. fluviale*, R. *babjevae*), *Lipomyces* (e.g., *L. starkeyii*, *L. lipofer*), *Debaryomyces* (e.g., *D. hansenii*), *Cryptococcus* (e.g., *C. podzolicus*, *C. phenolicus*, *C. curvatus*), *Pichia* (e.g., *P. segobiensis*), *Cystofilobasidium* (e.g., *C. informiminiatum*), *Leucosporidium* (e.g., *L. scottii*), *Sporobolomyces* (e.g., *S. singularis*, *S. poonsookiae*, *S. odoratus*, *S. metaroseus*, *S. bannaenis*), *Sporidiobolus* (e.g., *S. ruineniae*, *S. carnicolor*, S. pararoseus, *S. johnsonii*), *Schwanniomyces* (e.g., *S. occidentalis*), *Occultifur* (e.g., *O. externus*), *Blakeslea*, *Cunninghamella*, *Mortirella*, *Mucor*, *Phycomyces* and *Pythium*. Nonlimiting examples of oleaginous bacteria include *Morrococcus*, *Bacillus subtilis* and *Rhodococcus opacus*. Examples of oleaginous algae include, but are not limited to *Nannochloropsis* (e.g., *N. oceania*), *Chlorella* (e.g., *C. vulgaris*), *Thraustochyrtium* and *Schizochytrium*.

In particular embodiments, a host organism or modified organism can be a non-oleaginous yeast. In some embodiments, a host organism or modified organism can be a non-oleaginous, non-*Saccharomyces* yeast. Included among such yeast are non-oleaginous, non-*Saccharomyces* ascomycetes yeast as well as non-oleaginous, basidiomycetes yeast and non-oleaginous, ascomycetes yeast. In certain aspects, a host organism or modified organism can be an oleaginous, non-*Yarrowia* yeast, a non-*Yarrowia* ascomycetes yeast, or a non-*Yarrowia*, non-*Saccharomyces*, ascomycetes yeast. In another aspect, a host organism or modified organism can be an oleaginous yeast that accumulates 20% to 65% or 20% to 60% or 20% to 58%, or 20% to 55%, or 20% to 50% or 20% to 45%, or 20% to 40% or 20% to 35%, or 20% to 30% or 20% to 25% of its cell mass (by dry weight) as intracellular lipids (e.g., oil). In another embodiment, a host organism or modified organism can be an oleaginous yeast that

accumulates at least 20% or at least about 25% of its cell mass (by dry weight) as intracellular lipids (e.g., oil). In a further embodiment, a host organism or modified organism can be an oleaginous yeast that accumulates at least 20% but less than 70%, or at least 20% but less than 60%, or at least 20% but less than 50%, or at least 20% but less than 40%, or at least 20% but less than 30%, of its cell mass (by dry weight) as intracellular lipids (e.g., oil). In another embodiment, a host organism or modified organism can be an oleaginous yeast in which linoleic acid is less than 50% or less than 45% or less than 40% or less than 35% or less than 30% or less than 25% or less than 20% of the intracellular accumulated lipid composition.

In some embodiments, a host cell or organism or modified cell or organism is one that is capable of ω-oxidation of alkanes and/or fatty acids. Such cells or organisms can endogenously produce enzymes of the ω-oxidation pathway. This pathway includes steps of ω-hydroxylation, oxidation and dehydrogenation of ω-carbon. The ω-hydroxylation step can be catalyzed by a hydroxylase complex including a cytochrome P450 monooxygenase (such as, for example, an alkane-inducible cytochrome P450, e.g., CYP52) and a cytochrome P450:NADPH oxidoreductase which yields an alcohol. In a subsequent oxidation step, the alcohol is further oxidized to an aldehyde in a reaction catalyzed by a fatty alcohol oxidase. A dicarboxylic acid is generated through dehydrogenation of the aldehyde by a fatty aldehyde dehydrogenase. In some aspects, a host cell or organism endogenously expresses proteins having cytochrome P450 monooxygenase and cytochrome P450:NADPH oxidoreductase activity. In some embodiments, a host cell or organism or modified cell or organism is one that is capable of synthesizing dicarboxylic acids, such as, for example, α,ω-dicarboxylic acids.

In some embodiments, the host cell or organism or modified cell or organism is a diploid cell or organism. In certain aspects, the host cell or organism or modified cell or organism is an anamorphic ascomycetes yeast.

In some embodiments, a host organism or modified organism can be a thermotolerant and/or osmotolerant organism. As used herein, “thermotolerant,” in reference to an organism, e.g., a microorganism, refers to the ability of the organism to survive at elevated temperatures. For example, a thermotolerant organism, e.g., a microorganism, such as yeast, is one that is able to survive and/or grow and/or assimilate fatty acids and/or aliphatic carbon sources at temperatures greater than 30°C, greater than 31°C, greater than 32°C, greater than 33°C, greater than 34°C, greater than 35°C, greater than 36°C, greater than 37°C, greater than 38°C,

greater than 39°C, greater than 40°C, greater than 41°C, greater than 42°C, greater than 43°C, greater than 44°C, greater than 45°C, greater than 46°C, greater than 47°C, or greater than 48°C. A thermotolerant organism can be one that is able to survive and/or grow and/or assimilate fatty acids and/or aliphatic carbon sources at temperatures up to about 30°C, 32°C, 34°C, 35°C, 40°C, 41°C, 42°C, 43°C, 44°C, 45°C, 46°C, 47°C, 48°C, 49°C or 50°C or more. As used herein, "osmotolerant," in reference to an organism, e.g., a microorganism, refers to the ability of the organism to survive in elevated external osmotic pressure environments, e.g., high solute (such as salt or sugar) concentrations). For example, an osmotolerant organism, e.g., a microorganism, such as yeast, is one that is able to survive and/or grow and/or assimilate fatty acids and/or aliphatic carbon sources in media containing up to about 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25% or greater NaCl. Thermotolerant and/or osmotolerant microorganisms include, for example, species of *Blastobotrys* yeast (e.g., *Blastobotrys adeninivorans*), *Candida* yeast (e.g., *C. Mexicana, C. glycerinogenes*, *C. zemplinina*), *Pichia* yeast (e.g., *P. mississippiensis*, *P. mexicana*, *P. farinosa*, *P. sorbitophila*), *Clavispora* yeast (e.g., *C. opuntiae*, *C. lusitaniae*), *Kluyveromyces* yeast (e.g., *K. thermotolerans*), *Debaryomyces* (e.g., *D. hansenii*), *Rhodotorula* (e.g., *R. mucilaginosa*), *Zygosaccharomyces* (e.g., *Z. rouxii*) and *Issatchenkia* (e.g., *I. orientalis*). Thermotolerant and/or osmotolerant organisms can be well suited for use in industrial production systems operating at elevated temperatures and/or osmotic pressures that would impair growth and/or metabolism and/or completely inactivate organisms that are not thermotolerant and/or osmotolerant. Furthermore, in many instances, production efficiency can be improved and production costs reduced in using such organisms due to decreases in losses and avoidance of implementation of cooling processes.

Host cells and microorganisms and engineered cells and microorganisms may be provided in any suitable form. For example, such cells and microorganisms may be provided in liquid culture or solid culture (e.g., agar-based medium), which may be a primary culture or may have been passaged (e.g., diluted and cultured) one or more times. Microorganisms and cells also may be provided in frozen form or dry form (e.g., lyophilized). Microorganisms and cells may be provided at any suitable concentration.

Modified cells and organisms

Provided herein are cells and organisms (including microorgansims) that have been modified in one or more aspects relative to the unmodified cell or organism (i.e., the cell or organism prior to

the modification). For example, a cell or organism can be modified by altering one or more cellular activities and/or the sum total of a cell's or organism's activities. Thus, in this example, modifications can include alteration of cellular activities, addition of cellular activities and/or elimination of cellular activities. A "cellular activity," as used herein, refers to any process, functioning, or operation that can occur in a cell. In particular embodiments provided herein, a cell or organism has been modified to alter cellular carbon flux. Such modified cells and organisms have been manipulated in ways designed to direct carbon atoms toward one or more biochemical events, cellular activities or cellular locations and/or away from other metabolic pathways, cellular activities or locations. The alteration(s) can involve a single modification or multiple modifications of the original, or host, cell or organism in which carbon flux is altered. Also provided herein are methods of producing such modified cells and organisms. As described herein, there are multiple methods of altering cellular carbon flux by modifying one or more aspects of carbon processing in cells. Aspects of cellular carbon processing include, for example, but are not limited to, fatty acid metabolism, including fatty acid catabolism and synthesis, ω-oxidation, β-oxidation, fatty acid transport, acetyl group transfer/transport and processing, the TCA cycle, metabolite processing and triacylglyceride and lipid biosynthesis. For example, in altering carbon flux, certain cellular activities may be reduced, slowed or eliminated and/or other activities may be increased, accelerated, added or relocated. In particular embodiments, the amount and/or activity of one or more enzymes and/or transport proteins is/are modified in cells or microorganisms.

As such, the cells and organisms provided herein are "modified" or "engineered." The terms "engineered" or "modified," as used interchangeably herein, in reference to a cell, organism or microorganism refer to a cell or organism (including a microorganism) that has been manipulated or altered such that it is distinct (e.g., detectably changed or physically different) from a naturally occurring cell or organism. For example, the sum total of the cellular activities of a modified or engineered cell or microorganism can be distinct from those of a naturally occurring cell or microorganism, e.g., a modified cell or microorganism may include or lack one or more activities relative to the activities present in an unmodified cell or microorganism utilized as a starting point (e.g., host cell, host organism or host microorganism) for modification. In another example, one or more cellular activities of a modified or engineered cell or microorganism may be altered relative to the cellular activity or activities of the host cell or microorganism. A modified or engineered cell or organism can be genetically modified through any alteration in its genetic composition. For example, a genetically modified cell or organism

can include one or more heterologous polynucleotides, can have one or more endogenous nucleic acid deletions and/or can have one or more genetic mutations. Mutations include point mutations, insertions and deletions of a single or multiple residues in a nucleic acid. In some embodiments, an engineered cell, organism or microorganism includes a heterologous polynucleotide, and in certain embodiments, an engineered cell, organism or microorganism has been subjected to selective conditions that alter an activity, or introduce an activity, relative to the host cell or microorganism. Thus, a modified or engineered cell, organism or microorganism has been altered directly or indirectly by a human being. It is understood that the terms "modified cell," "modified organism," "modified microorganism," "engineered cell," " engineered organism," "engineered microorganism," refer not only to the particular cell or organism but to the progeny or potential progeny of such a cell or organism. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

For example, a "genetically modified" or "genetically engineered" cell, organism or microorganism is one in which the genetic make-up of the cell, organism or microorganism has been modified. Genetic modification encompasses a variety of alterations and can be accomplished in numerous ways. A genetic modification includes, but is not limited to, any of the following alterations: modification of the expression of an endogenous gene (e.g., the amount, pattern, timing and/or regulation (e.g., inducibility) of expression of a gene), disruption or deletion of an endogenous gene, increasing the copy number of an endogenous gene, mutation of an endogenous gene (including the regulatory components, exons, introns and/or peptide- or protein-encoding portions of a gene), and introduction of heterologous nucleic acid in to a cell or cells. These genetic modifications, and others, are described herein.

A genetic modification of a cell or organism can be one that modifies the expression of one or more nucleic acids or polypeptides in the cell or organism. A genetic modification of a cell or organism can be one that modifies the amount and/or activity of a polypeptide in the cell or organism. For example, modified expression of a nucleic acid or protein (e.g., modified rate, amount and/or level of expression) or modified amount or activity of a polypeptide may be a reduction, slowing, decrease or elimination, or increase, acceleration, addition or elevation in expression of a nucleic acid or protein or in the amount and/or activity of a polypeptide.

Modified expression of a nucleic acid or modification of the amount and/or activity of a polypeptide may be a relocation of expression or activity within a cell.

In one example, a genetic modification of a cell or organism can be one that alters the expression of, or the amount and/or activity of, a polypeptide involved in a reaction that generates a product (e.g., cytosolic acetyl-CoA, cytosolic malonyl-CoA, peroxisomal acetate) in a cell or organism. A "polypeptide involved in a reaction that generates," as used herein with respect to generation of a product, refers to a polypeptide that participates in the direct generation of the product from reactants. A reaction that directly generates a particular product can be a single-step reaction or a multi-step reaction involving transient reaction intermediates. For example, a polypeptide involved in a reaction that generates cytosolic acetyl-CoA is one that participates in a reaction that directly yields acetyl-CoA in the cytosol. Exemplary polypeptides (e.g., enzymes) involved in a reaction that generates cytosolic acetyl-CoA include, but are not limited to, cytosolic carnitine acetyltransferase, cytosolic acetyl-CoA synthetase and cytosolic ATP citrate lyase. A polypeptide involved in a reaction that generates cytosolic malonyl-CoA is one that participates in a reaction that directly yields malonyl-CoA in the cytosol. A non-limiting example of a polypeptide involved in a reaction that generates cytosolic malonyl-CoA is acetyl-CoA carboxylase. A polypeptide involved in a reaction that generates peroxisomal acetate is one that participates in a reaction that directly yields acetate in the peroxisome. A non-limiting example of a polypeptide involved in a reaction that generates peroxisomal acetate is acetyl-CoA hydrolase.

The term "endogenous," as used herein in reference to an aspect (e.g., a gene, nucleic acid, peptide, polypeptide, activity, genetic composition, gene expression, and the like) of a cell or organism or microorganism refers to the inherent aspect, or condition thereof, in the cell, organism or microorganism that has not been modified or engineered (i.e., the reference cell, organism or microorganism). The term "heterologous," "exogenous" or "foreign" as used herein with respect to a composition or quality (e.g., a gene, nucleic acid, peptide, polypeptide, cellular activity, genetic composition, gene expression, and the like) refers to the composition or quality not being a physically existing part or attribute of a reference cell, organism or microorganism. For example, a heterologous, exogenous or foreign nucleic acid can be any nucleic acid that is introduced into a cell or microorganism as part of a genetic modification of the cell or microorganism. A heterologous, exogenous or foreign composition (such as, for example, a nucleic acid) includes compositions that may be identical to an endogenous composition (e.g., a

nucleic acid gene sequence that is introduced into a cell or microorganism to increase the copy number and/or alter the positioning or expression of the same nucleic acid sequence existing therein) or may be different from an endogenous composition.

Coordination of carbon source, host organism and regulatory mechanisms to optimize carbon flux modification

In developing cell- and organism-based systems for enhanced production of target molecules, there are multiple factors, in addition to the design of cellular modifications for altering carbon flux, that can affect the overall efficiency and economics of the production process. These additional considerations include the sources of carbon available to the microorganism, the organism's ability to utilize various forms of carbon in the sources and the cellular regulatory systems that can be used in controlling carbon flux. Coordination of these factors can play a significant role in optimization of carbon flux alteration and, in turn, the efficiency of target molecule production.

Carbon sources used for culturing cells and microorganisms and/or fermentation processes sometimes are referred to as feedstocks. The term "feedstock" as used herein refers to a composition containing a carbon source that is provided to a cell or organism, which is used by the cell or organism to produce energy and metabolic products useful for growth. In order for cells and organisms to utilize carbon in vital processes, the carbon source is processed intracellularly in catabolic pathways to a form(s) that can be accommodated by energy generation and biosynthetic pathways. For example, glucose is processed in glycolytic pathways in cells whereas fatty acids are processed through β-oxidation. Thus, the carbon source used in microbial-based methods of target molecule production can influence which metabolic pathways will be involved in assimilating the carbon. A target molecule production system that incorporates elements of endogenous cellular metabolic pathways may not perform optimally if those pathways are not utilized in processing the carbon source.

Some cells and microorganisms are able to utilize a variety of carbon sources. However, many cells and microorganisms preferentially utilize particular carbon sources over others, and some cells and microorganisms are unable to utilize certain carbon sources. For example, *Saccharomyces cerevisiae* can utilize xylulose but not xylose. *Blastobotrys adeninivorans* and

Arxula terestre are able to utilize carbon- and nitrogen-containing compounds, e.g., adenine, uric acid, butylamine and pentylamine, as a sole source of carbon and nitrogen.

One consideration in the design of an economically feasible cell-based system for the production of target molecules is production-associated costs. The carbon source used in cultivating cells and organisms can be a significant factor contributing to production costs. Many microorganisms, including yeast, preferentially use glucose over other carbon sources. However, glucose is a relatively high-cost carbon source. Therefore, from an economic perspective, it can be beneficial to utilize lower-cost sources of carbon in bioproduction systems. Non-fermentable carbon sources, including, for example, glycerol and fatty acids, may be lower-cost alternatives to glucose and other carbohydrates in feedstocks. For example, waste materials, such as waste cooking oil, can be used as feedstocks containing non-fermentable carbon sources.

Therefore, in developing a cost-effective, efficient cell- or microbial-based target molecule production system, the modifications made to an organism to alter carbon flux should be coordinated and compatible with, and complementary to, the carbon source and cell or organism that will be employed in the production methods. Additionally, for optimal target molecule production, the regulatory mechanisms that are used in the cell or organism for controlling the individual elements (e.g., enzyme expression) being modified should provide for timing and extent of each element's activity that correlates with the desired carbon flux alterations at each stage of the production process.

For example, some embodiments of the cell- and microbial-based methods for producing target molecules provided herein include cells or organisms in which carbon processing activities have been engineered to enhance carbon flow through cellular oxidative metabolism pathways, e.g, ω-oxidation and/or β-oxidation. One advantage of such modified bioproduction systems is that they are well suited for use with lower cost, alternative carbon sources, including, for example, non-carbohydrate and non-fermentable carbon sources such as aliphatic compounds and hydrocarbons (e.g, alkanes, fatty acids and fatty alcohols). Use of such carbon sources is not only more cost-effective but can also have the added advantage of reducing the environmental impact of harmful wastes (e.g., agro-industrial by-products, waste cooking oil and waste motor oil) that can be used as feedstocks in target molecule production instead of being discarded. Cells or organisms particularly compatible with such methods are those that are able to utilize

non-fermentable, as well as fermentable, carbon sources. Generally, such cells and organisms contain endogenous metabolic pathways that form part of the basis for the desired carbon flux modifications. As also described herein, embodiments of the cell- and microbial-based systems in which carbon processing activities have been engineered to direct carbon flow through oxidative metabolism and away from mitochondrial metabolism can be controlled to provide for maximal, coordinated and highly efficient target molecule production based on, for example, use of carbon source-dependent transcription regulation of modified activities in the cells.

Transcription regulatory elements, including promoters, for some genes are responsive to the carbon source available to the cells. For example, transcription of some genes is subject to glucose repression in which the gene may not be expressed, or is less expressed, in the presence of glucose. Thus, in contrast to unregulated constitutive promoters, transcription regulatory elements for genes such as these are repressed, derepressible and/or inducible by varying carbon sources. When glucose is depleted, genes that were subject to glucose repression are then transcribed in a process referred to as glucose derepression. For some of these genes, this increase in transcription due to derepression represents the extent to which the genes will be expressed because they are not subject to induction and further increased transcription. For others of these genes, transcription may be increased several-fold over the derepressed level upon induction by, for example, certain carbon sources. Examples of such carbon sources include, but are not limited to, vegetable oils, triglycerides, fatty acids, e.g, oleic acid, esters of fatty acids and n-alkanes. Some genes encoding peroxisomal proteins (including enzymes involved in fatty acid catabolism) are subject to glucose repression/derepression. As described herein, the transcription regulatory elements for genes subject to glucose repression can advantageously be used in cell- and microbial-based methods for target molecule production involving alternative carbon sources.

Different carbon sources or feedstocks may be used in culturing cells or microorganisms at different phases of a target molecule production process. For example, one carbon source, e.g., glucose, may be used in preparing an initial starter culture of modified cells or microorganisms to establish a foundation of growing cells and a different carbon source (e.g., a lower-cost alternative such as fatty acids) may be used in a target molecule production phase subsequent to establishment of the starter culture. Accordingly, carbon source utilization can vary depending on the goal of a particular time or phase of a culture process.

In some embodiments of cell- and microbial-based target molecule production systems provided herein, modifications made to the cells or microorganisms include use of carbon source-dependent regulatory elements in altering carbon flux to enhance production efficiency. As described herein, in some embodiments, cytosolic activities for generating acetyl-CoA and/or malonyl-CoA (e.g., carnitine acetyltransferase, acetyl-CoA carboxylase, acetyl-CoA synthetase and/or ATP citrate lyase) can be increased for target molecule production during fatty acid or alkane assimilation, while mitochondrial and/or cytosolic activities for uptake and utilization of acetyl group carbons (e.g., cytosolic acyl-CoA synthetase, mitochondrial acetyl-carnitine transporters and/or mitochondrial carnitine acetyltransferase) are decreased. For optimal coordination of these activities with the overall production process, in some embodiments, the expression of proteins involved in some of the target molecule production activities can be engineered to be controlled by particular glucose-repressible and/or fatty acid-inducible transcription regulatory elements. For example, during initial cell-growth stages of a production method, a preferred carbon source may be, for example, glucose. Activities participating in target molecule production can be suppressed during this stage by using glucose-repressible elements, e.g., promoters, to regulate transcription of nucleic acids encoding proteins involved in those activities. At the same time, activities (e.g., mitochondrial metabolism) involved in cellular energy generation for growth can be permitted to function and/or elevated by using constitutive or glucose-inducible elements, e.g., promoters, to regulate transcription of nucleic acids encoding proteins involved in those activities. On the other hand, activities participating in target molecule production will be permited and/or increased following growth stages and during target molecule production stages when glucose is depleted and fatty acid carbon sources are provided by using glucose-repressible and/or fatty acid-inducible elements, e.g., promoters, to regulate transcription of nucleic acids encoding proteins involved in those activities. Also during those stages, the activities (e.g., mitochondrial metabolism) involved in cellular energy generation and growth can be unchanged or reduced or minimized by using weak, constitutive and/or glucose-inducible/fatty acid-inhibited elements, e.g., promoters, to regulate transcription of nucleic acids encoding proteins involved in those activities. This type of coordination of gene expression regulation with cellular modifications to alter carbon flux and use of alternative carbon sources can greatly enhance target molecule production efficiency and economy.

In some embodiments of cell- and microbial-based target molecule production systems provided herein, such as those in which target molecule production involves cellular oxidative metabolism pathways, e.g, ω-oxidation and/or β-oxidation, for processing of fatty acids and/or alkanes,

modifications made to the cells or microorganisms may include optimization of carbon source-dependent regulatory elements within the pathways. For example, although expression of unmodified genes encoding some of the polypeptides (e.g., enzymes) in these pathways is regulated by glucose-repressible and/or fatty acid (and/or alkane)-inducible transcriptional control elements, it may be beneficial to utilize heterologous stronger, more active, fatty acid (and/or alkane)-inducible transcriptional control elements to increase expression and/or activity of pathway polypeptides (e.g., enzymes) in modified cells and organisms. Thus, for example, the promoter of an endogenous glucose-repressible and/or fatty acid-inducible gene (e.g., an acyl-CoA oxidase (such as Pox5p), a peroxisomal protein (such as Pex11p), a cytochrome P450 monooxygenase or reductase (such as CYP52A17 or CPRB proteins)) can be replaced with a glucose-repressible and/or fatty acid-inducible promoter from another gene (e.g., an *HDE* gene) to enhance carbon processing through oxidative metabolism in modified cells or organisms.

In some embodiments of cell- and microbial-based target molecule production systems provided herein, carbon flux can be altered to reduce acetyl carbon processing in mitochondria during target molecule production occurring with fatty acid or alkane assimilation. In these embodiments, transcriptional control elements of some endogenous glucose-repressible and/or fatty acid-inducible genes encoding mitochondria-assoicated polypepties (e.g., mitochondrial transporter proteins (such as Crc1p) and carnitine acetyltransferases (such as Cat2p)) can be replaced with a promoter that is not fatty acid-inducible (e.g., a weaker and/or constitutive promoter) from another gene (e.g., glucose-6-phosphate isomerase gene) to reduce acetyl carbon uptake by and/or metabolism in mitochondria in modified cells or organisms. As also described herein, acetyl carbon uptake by and/or metabolism in mitochondria can be reduced in modified cells and organisms by replacing genes encoding one or more endogenous mitochondria-associated polypeptides (e.g., enzymes, such as carnitine acetyltransferase) with nucleic acid encoding a less active polypeptide. The nucleic acid encoding a less active polypeptide may also be linked to a transcriptional control element that provides for weak and/or not inducible expression of the polypeptide. If, however, the activity of the less-active polypeptide is insufficient for optimal cell functioning, a more active, inducible (e.g., fatty acid-inducible) promoter can be used to regulate expression of the less-active polypeptide.

Aliphatic and hydrophobic carbon sources

The ability of cells and microorganisms to utilize alternative carbon sources for energy generation and growth is based in the multiple enzyme-mediated metabolic pathways and gene regulation systems in microbial cells. In general, glucose often is a preferred carbon and energy source for many cells and microorganisms, e.g., yeast. A number of genes encoding products, such as certain enzymes, involved in cellular pathways and processes that are not used in carbohydrate metabolism may be repressed when glucose is present in culture media. If glucose is depleted in the media, some of these genes may then be derepressed. If alternative carbon sources, e.g., non-fermentable carbon sources, are available, some of the genes may be induced, and may be induced by several-fold. For example, when aliphatic or hydrophobic carbon sources, (e.g., alkanes, alkenes, fatty acids) are the external carbon source, the expression of genes encoding enzymes involved in β-oxidation and proteins involved in peroxisome proliferation can be induced. One example of a gene that is repressed in glucose media, derepressed in non-fermentable carbon source-containing media and induced in fatty acid-containing media is the gene encoding the peroxisomal trifunctional enzyme hydratase-dehydogenase-epimerase (HDE) in yeast, such as, for example, *Candida* (see, e.g., Sloots et al. (1991) *Gene* 105:129-134). The upstream regulatory regions of the HDE gene include a glucose-responsive region controlling glucose repression, a non-fermentable carbon-responsive region controlling derepression and an oleic acid-responsive region controlling fatty acid induction of transcription of the gene.

In some embodiments of the methods for producing a target molecule provided herein, a carbon source used during the production phase of a culturing step in the method includes an aliphatic or hydrophobic carbon source. In particular embodiments, an aliphatic or hydrophobic carbon source is the primary carbon source or may be the sole, or only, carbon source used during the production phase of a culturing step in the method. In some embodiments, the carbon source is a fatty acid and/or alkane. In certain aspects, the carbon source is a fatty acid. In some embodiments, the carbon source is an 18-carbon fatty acid such as, for example, oleic acid (C18:1), linoleic acid (C18:2) or linolenic acid (C18:3). Embodiments of the methods in which an aliphatic or hydrophobic carbon source is used are particularly advantageous when target molecule production involves oxidation pathways such as ω-oxidation and/or β-oxidation pathways and/or involves peroxisomes.

Additional advantages of using aliphatic or hydrophobic carbon sources in some embodiments of the methods of producing target molecules as provided herein are reduced costs and positive environmental effects realized through their use.

Methods of modifying cellular carbon flux

Included in the cells, organisms and microorganisms and methods described herein are those that provide for enhanced production of desired target molecules. In one aspect, production is enhanced through modification of carbon flux in cell or microbial production systems. Through modification of cellular carbon flux, carbon atoms that may have flowed or been transported into other metabolic processes (e.g., energy and/or cellular composition generation) in the cell are redirected for use in a target molecule production process. Described herein are multiple cellular modifications that can be employed to beneficially alter carbon flux. A modification can be used alone or in combination with one or more other modifications depending on the target molecule produced and the carbon flux alteration that is best suited to maximize its production.

Modification of acetyl-carnitine entry into mitochondria

Included in embodiments of the cells, microorganisms, compositions and methods provided herein are cell and microbial production platform systems and components thereof in which the amount of (a) acetyl-carnitine in the cell cytosol and/or (b) carnitine acetyltransferase and/or carnitine acetyltransferase activity in the cell cytosol is/are modified. Carnitine and carnitine acetyltransferase are the primary elements of the carnitine shuttle system in which acetyl carbons from acetyl-CoA are transferred across intracellular membranes and transported throughout eukaryotic cells. For example, the carnitine shuttle is a mechanism through which acetyl carbons from acetyl-CoA generated in peroxisomes flow to mitochondria. Acetyl groups from peroxisomal acetyl-CoA can be transferred to carnitine in a reaction catalyzed by carnitine acetyltransferase and then move across the peroxisomal membrane and into the cytosol in the form of acetyl-carnitine. Cytosolic acetyl-carnitine can then be transported into mitochondria where mitochondrial carnitine acetyltransferase catalyzes the reverse reaction to transfer the acetyl moiety from carnitine to free coenzyme A to generate acetyl-CoA. The carnitine shuttle provides a main “highway” for the flow of carbon atoms into mitochondria, particularly in the assimilation of fatty acids and other non-carbohydrate and non-fermentable carbon source

molecules by cells. It is thus one target for modification of carbon flux in cells and microorganisms provided herein.

In some embodiments of the cells, microorganisms, compositions and methods provided herein, the amount of (a) acetyl-carnitine in the cell cytosol is increased and/or decreased, and/or (b) carnitine acetyltransferase and/or carnitine acetyltransferase activity in the cell cytosol is/are increased and/or decreased. For example, in some aspects, a cell or microorganism may be modified to increase cytosolic acetyl-carnitine, may be modified to decrease cytosolic acetyl-carnitine or may be modified to alternately increase and decrease cytosolic acetyl-carnitine depending on the conditions in which the modified cell or microorganism is cultured.

In certain embodiments, a host cell or microorganism is modified to increase the amount of (a) acetyl-carnitine in the cell cytosol and/or (b) carnitine acetyltransferase and/or carnitine acetyltransferase activity in the cell cytosol. Increasing the amount and/or activity of cytosolic carnitine acetyltransferase provides for an increased conversion of acetyl-carnitine into acetyl-CoA in the cytosol. Increasing the amount of acetyl-carnitine in the cytosol provides an increased availability of substrate for cytosolic carnitine acetyltransferase to convert to acetyl-CoA. These modifications effectively result in an increase in the generation and amount of cytosolic acetyl-CoA which can then be used in the production of desired carbon-containing molecules.

Modification of a mitochondrial acetyl-carnitine transporter

In one embodiment, cells or organisms provided herein are modified to increase and/or decrease the amount of acetyl group carbons in the cytosol in the form of acetyl-carnitine in transit from the peroxisome and other areas to the mitochondria by increasing or reducing the entry of acetyl-carnitine into mitochondria from the cytosol. In one aspect, the amount and/or rate of acetyl-carnitine transfer into the mitochondria in a cell can be increased or reduced by increasing or decreasing the expression of an acetyl-carnitine translocase protein localized in mitochondrial inner membranes. This protein carrier transports acetyl-carnitine across the mitochondrial inner membrane and into the mitochondrial matrix in exchange for carnitine. For example, a mitochondrial carnitine translocase is encoded by a *CRC1* gene in *Saccharomyces* (see, e.g, Palmieri *et al*. (1999) *FEBS Lett* 462:472-276), which contains an oleate-responsive element in the promoter region, and by an *AcuH* gene in *Asperillgus* (see, e.g., De Lucas *et al*.

(2001) *FEMS Microbiol Lett* 201:193-198). Carnitine carrier proteins belong to a family of mitochondrial carrier proteins which generally contain three tandemly repeated ~100-amino acid domains. Each of the three domains typically contains two hydrophobic regions, spanning the membrane as α-helices, linked by a hydrophobic loop that extends into the mitochondrial matrix. Each domain also typically contains a version of a motif (PX[DE]XX[RK]XRK) involved in forming a salt bridge that closes off the matrix side of a channel generated by the α-helices (see, e.g., Indiveri et al. (2011) *Molecular Aspects of Medicine* 32:223-233).

There are a number of ways to increase or reduce expression of an acetyl-carnitine translocase in a cell. For example, a host acetyl-carnitine translocase activity can be decreased by disruption (e.g., knockout, insertion mutagenesis, the like and combinations thereof) of a host gene encoding the protein, or by decreasing the activity of the promoter (e.g., through addition of repressor sequences to the promoter or 5'UTR or replacing the promoter) that controls transcription of an acetyl-carnitine translocase gene using recombinant molecular biology techniques known in the art and/or described herein. One method for disrupting an endogenous acetyl-carnitine translocase gene is by recombinantly inserting a heterologous nucleic acid (e.g., a nucleotide sequence encoding a selectable marker such as an enzyme that restores an auxotrophic host cell or organism to prototrophy) into the endogenous gene, thereby generating an engineered cell or organism deficient in acetyl-carnitine translocase activity. This can be done, for example, through homologous recombination in which a heterologous nucleic acid containing sequences of the endogenous acetyl-carnitine translocase gene and a disrupting sequence (e.g., a knock-out gene cassette such as described herein) is introduced into a host cell or microorganism. Nucleic acids encoding an acetyl-carnitine translocase can be obtained from a number of sources, including, for example, yeast cells. For example, genomic DNA from cell sources can be amplified using oligonucleotide primers based on the nucleotide sequence of an acetyl-carnitine translocase-encoding gene. Provided herein, for example, are a nucleotide sequence (SEQ ID NO: 71) that encodes a *Candida viswanathii* acetyl-carnitine translocase and the corresponding amino acid sequence (SEQ ID NO: 14). Nucleotide sequences encoding additional acetyl-carnitine translocase proteins include, but are not limited to: *Saccharomyces cerevisiae CRC1* (Genbank accession number AJ250124) and *Aspergillus nidulans AcuH* (Genbank accession number AJ011563).

Presence, absence or amount of an acetyl-carnitine translocase activity can be detected by any suitable method known in the art and/or described herein. For example, detection can be

performed using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cells or organisms exhibit increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism. Methods of evaluating the activity of an acetyl-carnitine translocase include, for example, measuring carnitine uptake into and/or efflux from liposomes reconstituted with acetyl-carnitine translocase protein purified from microbial cells expressing the protein (see, e.g, , Palmieri *et al*. (1999) *FEBS Lett* 462:472-276).

In another example, a host acetyl-carnitine translocase activity can be increased, for example, by increasing the number of copies of an acetyl-carnitine translocase gene (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the gene), by increasing the activity of a promoter that regulates transcription of an acetyl-carnitine translocase gene, or by increasing the number of copies of an acetyl-carnitine translocase gene and increasing the activity of a promoter that regulates transcription of an acetyl-carnitine translocase gene. In some embodiments, an acetyl-carnitine translocase is endogenous to the host microorganism. Acetyl-carnitine translocase activities can also be increased, for example, by using an inducible promoter, e.g, a glucose- or fatty acid-inducible promoter for regulating transcription of an acetyl-carnitine translocase-encoding nucleic acid, and culturing the recombinant cell or microorganism in media containing a transcription-inducing carbon source.

Thus, in another example of modifying the expression of an acetyl-carnitine translocase in a cell, the promoter used for expression of nucleic acid encoding a mitochondrial acetyl-carnitine transport protein can be modified relative to an endogenous promoter encoding a transport protein. A promoter that is weaker, stronger and/or differently regulated than any endogenous mitochondrial acetyl-carnitine translocase gene promoter will provide for modified expression levels of the translocase protein. To achieve such modified expression, an endogenous promoter of a gene encoding a mitochondrial inner membrane acetyl-carnitine translocase can, in effect, be replaced with another promoter. This can be accomplished, for example, by introducing into a cell or microorganism a heterologous nucleic acid construct that includes a translocase-encoding sequence of nucleotides operably linked to a promoter that provides modified transcription or expression in the cell or microorganism relative to the endogenous promoter. The cell or microorganism can be one in which the endogenous gene(s) encoding a mitochondrial acetyl-carnitine translocase has been disrupted or deleted. For example, a host

organism could be a yeast, e.g., a *Candida* yeast, in which the endogenous promoter of the mitochondrial acetyl-carnitine translocase includes an oleate-responsive element allowing for fatty acid induction. An example of a weaker promoter that would not be fatty acid inducible and provide for decreased acetyl-carnitine translocase expression, particularly when exposed to fatty acids as a carbon source, could be a yeast glucose-6-phosphate isomerase gene promoter. Modifying a promoter in this way provides another method for decreasing the amount and/or activity of acetyl-carnitine translocase protein in a host. This method is particularly advantageous when decreasing or eliminating acetyl-carnitine translocase activity through gene disruption is detrimental to cell growth and/or viability.

Modification of mitochondrial carnitine acetyltransferase activity

Acetyl-carnitine can be generated and degraded by the action of carnitine acetyltransferases (e.g., EC 2.3.1.7). In another embodiment provided herein, the amount of acetyl group carbons in the cytosol in the form of acetyl-carnitine in transit from the peroxisome to the mitochondria can be modified through altering the amount and/or activity of mitochondrial carnitine acetyltransferase. For example, by decreasing the activity level of mitochondrial carnitine acetyltransferase, there can be a corresponding decrease in conversion of acetyl-carnitine to acetyl-CoA in the mitochondria. This can introduce a bottleneck in acetyl-carnitine processing in the mitochondria which can have the effect of diverting acetyl-carnitine from entering the mitochondria from the cytoplasm. Alternatively, by increasing the activity level of mitochondrial carnitine acetyltransferase, there can be a corresponding increase in conversion of acetyl-carnitine to acetyl-CoA in the mitochondria which can augment acetyl-carnitine processing and avoid slowing of mitochondrial entry of acetyl-carnitine due to bottlenecks that might occur in the presence of increased amounts of acetyl-carnitine in the cytoplasm.

In some organisms, such as, for example, certain yeast species, carnitine acetyltransferase is dually targeted to mitochondria and peroxisomes by N-terminal and C-terminal targeting signals, respectively (see, e.g., Elgersma *et al.* (1995) *EMBO J.* 14: 3472-3479 and Kawachi *et al.* (1996) *Eur. J. Biochem.* 238: 845-852). An N-terminal sequence is referred to as the mitochondrial targeting signal (mts) and a C-terminal sequence is referred to as the peroxisomal targeting sequence (pts). An example of such an enzyme is the carnitine O-acetyltransferase enzyme encoded by some yeast *CAT2* genes.

Modifying carnitine acetyltransferase activity in mitochondria can be accomplished by modifying the amount of mitochondrial carnitine acetyltransferase protein expression in a cell, for example, by replacing the wild-type promoter of an endogenous gene in a cell or organism with a weaker or stronger heterologous promoter, and/or replacing or modifying a gene encoding a wild-type carnitine acetyltransferase such that the encoded modified or substituted carnitine acetyltransferase protein has a reduced or increased enzyme activity. For example, a host carnitine acetyltransferase activity can be decreased by disruption (e.g., knockout, insertion mutagenesis, the like and combinations thereof) of a host gene encoding the protein, or by decreasing the activity of the promoter (e.g., through addition of repressor sequences to the promoter or 5'UTR or replacing the promoter) that controls transcription of a carnitine acetyltransferase gene using recombinant molecular biology techniques known in the art and/or described herein. In one embodiment, a diploid yeast, such as, for example, a *Candida* yeast, when used as a host microorganism can be subjected to genetic modification in which one of the two alleles of a mitochondrial carnitine acetyltransferase gene is disrupted or deleted. In so doing, a single allele of the gene remains for a reduced amount of carnitine acetyltransferase expression in the microorganism and a reduced amount of the protein in the mitochondria. This can effectively reduce and/or slow the amount of acetyl-carnitine that is processed into acetyl-CoA in the mitochondria without completely eliminating a minimal supply of acetyl carbons that may be required for cellular respiration that occurs in the mitochondria, yet provides for increased retention of acetyl-carnitine in the cytosol. In some instances, the amount of carnitine acetyltransferase activity remaining after disruption of a single allele of a mitochondrial carnitine acetyltransferase gene of a diploid cell or organism may be at a higher level than desired. In such cases, both alleles may be disrupted. A heterologous nucleic acid encoding a carnitine acetyltransferase that is less active than an endogenous carnitine acetyltransferase (or nucleic acid encoding a carnitine acetyltransferase that is linked to a weak promoter) can be introduced into host cells or organisms in which all alleles of the endogenous gene have been disrupted.

One method for disrupting an endogenous carnitine acetyltransferase gene is by recombinantly inserting a heterologous nucleic acid (e.g., a nucleotide sequence encoding a selectable marker such as an enzyme that restores an auxotrophic host cell or organism to prototrophy) into the endogenous gene, thereby generating an engineered cell or organism deficient in carnitine acetyltransferase activity. This can be done, for example, through homologous recombination in which a heterologous nucleic acid containing sequences of the endogenous carnitine acetyltransferase gene and a disrupting sequence (e.g., a knock-out gene cassette such as

described herein) is introduced into a host cell or microorganism. Nucleic acids encoding a carnitine acetyltransferase can be obtained from a number of sources, including, for example, yeast cells. For example, genomic DNA from cell sources can be amplified using oligonucleotide primers based on the nucleotide sequence of a carnitine acetyltransferase-encoding gene. Provided herein, for example, are a nucleotide sequence (SEQ ID NO: 59) that encodes a *Candida viswanathii* carnitine acetyltransferase (*CAT2* gene) and the corresponding amino acid sequence (SEQ ID NO: 2). Nucleotide sequences encoding additional carnitine acetyltransferase proteins include, but are not limited to: *Saccharomyces cerevisiae CAT2* (Genbank accession numbers Z14021, NM_001182400), *Candida tropicalis CAT2* (Genbank accession number D84549), *Candida glabrata CAT2* (Genbank accession number AF2811), *Candida albicans CAT2* (Genbank accession numbers AF525684), *Aspergillus nidulans AcuJ* (Genbank accession number XM_658791) and *Cyberlindnera jadinii* (Genbank accession number AB641826).

In cells or organisms in which a mitochondrial carnitine acetyltransferase is encoded by a gene that generates a protein containing mitochondrial and peroxisomal targeting sequences, it may be desired to modify only the mitochondrial enzyme, and continue expression of the peroxisomal enzyme. In this instance, an endogenous mitochondrial/peroxisomal carnitine acetyltransferase gene can be disrupted or deleted and heterologous nucleic acids separately encoding a mitochondrial-targeted enzyme and a peroxisomal-targeted enzyme can be introduced into the cell or microorganism. For example, a peroxisomal-targeted enzyme that would not be expressed in the mitochondria can be produced in a cell or microorganism by introducing a heterologous nucleic acid that encodes a carnitine acetyltransferase that includes a peroxisomal targeting sequence of amino acids but lacks a mitochondrial targeting sequence of amino acids. An example of such a modified *Candida viswanathii* nucleic acid sequence (*CAT2*$^{\Delta mts}$; SEQ ID NO: 60), and the amino acid sequence encoded thereby (Cat2p$^{\Delta mts}$; SEQ ID NO: 3), are provided herein. A mitochondrial-targeted carnitine acetyltransferase that would not be expressed in peroxisomes can be produced in a cell or microorganism by introducing a heterologous nucleic acid that encodes a carnitine acetyltransferase that includes a mitochondrial targeting sequence of amino acids but lacks a peroxisomal targeting sequence of amino acids. An example of such a modified *Candida viswanathii* nucleic acid sequence (*CAT2*$^{\Delta pts}$; SEQ ID NO: 62), and the amino acid sequence encoded thereby (Cat2p$^{\Delta pts}$; SEQ ID NO: 5), are provided herein. A heterologous nucleic acid encoding a mitochondrial-targeted carnitine acetyltransferase that would not be expressed in peroxisomes can also include

modifications that alter its expression and or activity in the mitochondria as described herein. For example, regulatory sequences of nucleic acids (e.g, promoter sequences, repressor sequences) can be included that provide for decreased or increased expression of the enzyme and/or an altered pattern of expression of the enzyme. A heterologous nucleic acid encoding a mitochondrial-targeted carnitine acetyltransferase can include modifications that alter its activity, e.g., providing for more active or less active enzymatic activity relative to an endogenous mitochondrial carnitine acetyltransferase. The carnitine acetyltransferase activities of host and modified cells and microorganisms can be evaluated and monitored using methods known in the art. For example, methods of isolating peroxisomal and mitochondrial components of yeast cells and of extracting carnitine acetyltransferase from subcellular fractions have been described by Ueda et al. [(1982) *Eur. J. Biochem.*124:205-210] and Kozulic et al. [(1987) *Eur. J. Biochem.*168:245-250]. Methods of measuring the enzymatic activity of carnitine acetyltransferase are also known in the art, see, e.g., Fritz and Schultz (1965) *J. Biol. Chem.* 240:2188-2192; Chase (1969) *Meth. Enzymol.*13:387-393.

In one embodiment provided herein, a heterologous nucleic acid encoding a yeast cytoplasmic carnitine acetyltransferase that has a reduced carnitine acetyltransferase activity relative to the activity of the enzyme encoded by a host microorganism's endogenous mitochondrial carnitine acetyltransferase gene can be introduced into a microbial host in which the endogenous mitochondrial carnitine acetyltransferase gene(s) has been disrupted or deleted. The heterologous nucleic acid encoding the less active carnitine acetyltransferase can be modified to include nucleotides encoding a mitochondrial targeting sequence for expression of the enzyme in the mitochondria. For example, in one aspect, a heterologous nucleic acid encoding a *Candida viswanathii* cytosolic carnitine acetyltransferase (*YAT1*) with added nucleotides encoding a mitochondrial targeting sequence (*YAT1*$^{+mts}$) can be introduced into a host cell or microorganism (e.g., a *Candida viswanathii* cell). Any sequence encoding a mitochondrial targeting from a protein that is localized to mitochondria can be used in generating the heterologous nucleic acid. Examples include, but are not limited to, nucleotides encoding mitochondrial targeting sequences from mitochondrial cytochrome oxidase subunit IV (Cox4p), mitochondrial citrate synthase (Cit1p) and mitochondrial carnitine acetyltransferase (Cat2p) proteins. Nucleotide sequences encoding (and the amino acid sequences of) *Candida viswanathii* Yat1p (amino acid SEQ ID NO: 6 and nucleotide SEQ ID NO: 63), YAT1^{+mts}p (amino acid SEQ ID NOS: 10, 11 and 12 and nucleotide SEQ ID NOS: 67, 68 and 69), and the mitochondrial targeting sequences of Cox4p (amino acid SEQ ID NO: 7 and nucleotide SEQ ID

NO: 64), Cit1p (amino acid SEQ ID NO: 8 and nucleotide SEQ ID NO: 65) and Cat2p (amino acid SEQ ID NO: 9 and nucleotide SEQ ID NO: 66) are provided herein. Additional non-limiting examples of nucleic acids encoding cytoplasmic carnitine acetyltransferase include: *Saccharomyces cerevisiae YAT1* (Genbank accession number X74553), *Aspergillus nidulans FacC* (Genbank accession number AF023156), *Cyberlindnera jadinii YAT1* (Genbank accession number AB641829), *Candida dubliniensis YAT1* (Genbank accession number XM_002416790) and *Candida albicans* (Genbank accession number AF525683). Additional non-limiting examples of nucleic acids encoding mitochondrial targeting sequences include *Saccharomyces cerevisiae Cit1* (nucleotides in Genbank accession number NM_001183178 encoding N-terminal 37 amino acids) and *Saccharomyces cerevisiae Cox4* (nucleotides in Genbank accession number NM_001181052 encoding N-terminal 25 amino acids).

In another example, the amount and/or activity of carnitine acetyltransferase in a cell or microorganism can be increased, for example, by increasing the number of copies of a carnitine acetyltransferase gene (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the gene), by increasing the activity of a promoter that regulates transcription of a carnitine acetyltransferase gene, or by increasing the number of copies of a carnitine acetyltransferase gene and increasing the activity of a promoter that regulates transcription of a carnitine acetyltransferase gene. In some embodiments, a carnitine acetyltransferase is endogenous to the host cell or microorganism. Carnitine acetyltransferase activities can also be increased, for example, by using an inducible promoter, e.g, a glucose- or fatty acid-inducible promoter for regulating transcription of a carnitine acetyltransferase-encoding nucleic acid, and culturing the recombinant cell or microorganism in media containing a transcription-inducing carbon source.

Thus, in another example of modifying the expression of a carnitine acetyltransferase in a cell, the promoter used for expression of nucleic acid encoding a carnitine acetyltransferase protein can be modified relative to an endogenous promoter encoding a carnitine acetyltransferase protein. A promoter that is weaker, stronger and/or differently regulated than any endogenous carnitine acetyltransferase gene promoter will provide for modified expression levels of the protein. To achieve such modified expression, an endogenous promoter of a gene encoding a carnitine acetyltransferase can, in effect, be replaced with another promoter. This can be accomplished, for example, by introducing into a cell or microorganism a heterologous nucleic acid construct that includes a carnitine acetyltransferase-encoding sequence of nucleotides operably linked to a promoter that provides modified expression in the cell or microorganism

relative to the endogenous promoter. The cell or microorganism can be one in which the endogenous gene(s) encoding a carnitine acetyltransferase has been disrupted or deleted. For example, a host organism could be a yeast, e.g., a *Candida* yeast, in which the endogenous promoter includes an oleic acid-responsive element allowing for fatty acid induction. An example of a weaker promoter that would not be fatty acid inducible and provide for decreased carnitine acetyltransferase expression, particularly when exposed to fatty acids as a carbon source, could be a yeast glucose-6-phosphate isomerase gene promoter.

Different combinations of transcription regulatory elements (e.g., promoters) and enzymes can be utilized to achieve an optimal level of activity of carnitine acetyltransferase (or other enzyme being modified) in a cell or microorganism modified to alter carbon flux therein. For example, in embodiments in which a decreased level, but not an elimination, of an activity, such as mitochondrial carnitine acetyltransferase, in a cell or organism is desired, an optimal activity level may be achieved by using a strong and/or inducible promoter to express nucleic acid encoding a protein having a decreased activity. In one embodiment described herein, the mitochondrial carnitine acetyltransferase activity of a host organism (e.g., *Candida* yeast) is decreased by disrupting both alleles of the endogenous gene encoding mitochondrial carnitine acetyltransferase and introducing heterologous nucleic acid encoding a mitochondrial-targeted carnitine acetyltransferase having a lower enzyme activity than the endogenous mitochondrial carnitine acetyltransferase. To ensure that the level of enzyme activity provided by the less active mitochondrial-targeted carnitine acetyltransferase is sufficient and optimal in the modified cell, a strong, fatty acid-inducible promoter (e.g., an *HDE* gene promoter) can be linked to the nucleic acid encoding the less active enzyme to regulate transcription and production of a desired amount of the enzyme.

Modification of cytosolic carnitine acetyltransferase activity

Included in embodiments of the cells, microorganisms, compositions and methods provided herein are microbial production platform systems and components thereof in which the amount of carnitine acetyltransferase in the cell cytosol and/or carnitine acetyltransferase activity in the cell cytosol is modified. In some instances, the amount of carnitine acetyltransferase in the cell cytosol is increased and/or decreased, and/or carnitine acetyltransferase activity in the cell cytosol is increased and/or decreased. For example, in some aspects, a cell or microorganism may be modified to increase cytosolic carnitine acetyltransferase and/or carnitine

acetyltransferase activity, may be modified to decrease cytosolic carnitine acetyltransferase and/or carnitine acetyltransferase activity, or may be modified to alternately increase and decrease cytosolic carnitine acetyltransferase and/or carnitine acetyltransferase activity depending on the conditions in which the modified cell or microorganism is cultured.

In some embodiments, the capture of carbon atoms in the acetyl group of acetyl-CoA generated from metabolic processes such as peroxisomal β-oxidation can be accomplished by increasing the amount of carnitine acetyltransferase protein and/or activity in the cell cytosol of a cell or microorganism. In so doing, there is an increased conversion of acetyl-carnitine, such as that which is in transit from the peroxisome to the mitochondria, into acetyl-CoA in the cytoplasm. In one aspect, the amount and/or activity of a host cytosolic carnitine acetyltransferase can be increased, for example, by increasing the number of copies of a gene encoding a cytoplasmic carnitine acetyltransferase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the gene), by increasing the activity of a promoter that regulates transcription of a gene encoding a cytoplasmic carnitine acetyltransferase, or by increasing the number of copies of a gene encoding a cytoplasmic carnitine acetyltransferase and increasing the activity of a promoter that regulates transcription of a gene encoding a cytoplasmic carnitine acetyltransferase. In some embodiments, a cytoplasmic carnitine acetyltransferase is endogenous to the host cell or microorganism.

In one embodiment of the cell and microbial systems and methods provided herein, the amount of carnitine acetyltransferase protein expressed in the cytosol can be increased by introducing heterologous nucleic acid encoding a cytoplasmic carnitine acetyltransferase into a cell or microorganism. In some cells and microorganisms, e.g., some yeast strains, cytoplasmic carnitine acetyltransferase is encoded by a gene that is distinct from the gene(s) encoding mitochondrial and/or peroxisomal carnitine acetyltransferase. For example, in some yeast strains, a cytoplasmic carnitine acetyltransferase is encoded by a *YAT* gene, whereas a mitochondrial and/or peroxisomal carnitine acetyltransferase is encoded by a *CAT* gene. Nucleotide sequences encoding (and the amino acid sequences of) *Candida viswanathii* Yat1p (amino acid SEQ ID NO: 6 and nucleotide SEQ ID NO: 63) are provided herein. Additional non-limiting examples of nucleic acids encoding cytoplasmic carnitine acetyltransferase include *Saccharomyces cerevisiae YAT1* (Genbank accession number X74553), *Aspergillus nidulans FacC* (Genbank accession number AF023156), *Cyberlindnera jadinii YAT1* (Genbank accession

number AB641829), *Candida dubliniensis YAT1* (Genbank accession number XM_002416790) and *Candida albicans* (Genbank accession number AF525683).

In another embodiment, the amount of cytoplasmic carnitine acetyltransferase can be modified by modifying the activity of a promoter that regulates transcription of a gene encoding a cytoplasmic carnitine acetyltransferase. Thus, in another example of modifying the expression of a cytosolic carnitine acetyltransferase in a cell, the promoter used for expression of nucleic acid encoding a cytosolic carnitine acetyltransferase protein can be modified relative to an endogenous promoter encoding a cytosolic carnitine acetyltransferase protein. A promoter that is weaker, stronger and/or differently regulated than any endogenous cytosolic carnitine acetyltransferase gene promoter will provide for modified expression levels of the protein. To achieve such modified expression, an endogenous promoter of a gene encoding a cytosolic carnitine acetyltransferase can, in effect, be replaced with another promoter. This can be accomplished, for example, by introducing into a cell or microorganism a heterologous nucleic acid construct that includes a cytosolic carnitine acetyltransferase-encoding sequence of nucleotides operably linked to a promoter that provides modified expression in the cell or microorganism relative to the endogenous promoter. The cell or microorganism can be one in which the endogenous gene(s) encoding a cytosolic carnitine acetyltransferase has been disrupted or deleted. For example, a host organism could be a yeast, e.g., a *Candida* yeast, in which the endogenous promoter does not include an oleic acid-responsive element allowing for fatty acid induction. An example of a stronger promoter that would be fatty acid inducible and provide for increased cytosolic carnitine acetyltransferase expression, particularly when exposed to fatty acids as a carbon source, is a peroxisomal protein gene and/or β-oxidation enzyme gene promoter, e.g., a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequence of a *Candida viswanathii* HDE gene promoter (SEQ ID NO: 113) is provided herein as are examples of additional fatty acid-inducible promoters.

In a further embodiment, cytosolic carnitine acetyltransferase activity can be modified by introducing into a cell or microorganism a heterologous nucleic acid encoding a carnitine acetyltransferase that is more active or less active than an endogenous cytosolic carnitine acetyltransferase. For example, a heterologous nucleic acid encoding an enzyme that has an increased carnitine acetyltransferase activity relative to the activity of a cytosolic carnitine acetyltransferase expressed in the host cell or microorganism can be introduced into a host to provide for increased generation of cytosolic acetyl-CoA from acetyl-carnitine. The host can be

one in which the endogenous cytosolic carnitine acetyltransferase gene(s) has been disrupted or deleted. The heterologous nucleic acid encoding the more active carnitine acetyltransferase can, if required, be altered to exclude any nucleotides encoding a cell localization (e.g., mitochondria, peroxisomes) targeting sequence in order to provide for expression of the enzyme in the cytosol.

Thus, for example, engineered carnitine O-acetyltransferase proteins lacking amino acid sequence targeting signals that direct the enzyme to one or more cellular locations other than the cytoplasm can be expressed in host cells thereby increasing the amount of carnitine O-acetyltransferase in the cytoplasm. Such engineered proteins will remain in the cytoplasm after being produced by the cell or organism. One such modified carnitine O-acetyltransferase protein is a yeast Cat2p$^{\Delta mts\Delta pts}$ lacking a mitochondrial targeting signal (mts) and a perioxisomal targeting signal (pts). In some instances, a mitochondrial and/or peroxisomal carnitine acetyltransferase (e.g., a yeast Cat2p) may be more active than an endogenous cytosolic carnitine acetyltransferase (e.g., a yeast Yat1p). A more active Cat2p enzyme can be expressed in the cytosol of a host upon introduction of heterologous nucleic acid (e.g., *CAT2*$^{\Delta mts\Delta pts}$) encoding the more active enzyme lacking mitochondrial- and peroxisomal-targeting sequences. In a particular embodiment, the Cat2p enzyme can be a *Candida* yeast protein. An example of a *Candida viswanathii* nucleotide sequence (*CAT2*$^{\Delta mts\Delta pts}$; SEQ ID NO: 61) encoding a carnitine acetyltransferase lacking mitochondrial- and peroxisomal-targeting sequences (Cat2p$^{\Delta mts\Delta pts}$; SEQ ID NO: 4) is provided herein.

Additional examples of nucleotide sequences encoding carnitine acetyltransferase proteins include: *Saccharomyces cerevisiae CAT2* (Genbank accession numbers Z14021, NM_001182400), *Candida tropicalis CAT2* (Genbank accession number D84549), *Candida glabrata CAT2* (Genbank accession number AF2811), *Candida albicans CAT2* (Genbank accession numbers AF525684), *Aspergillus nidulans AcuJ* (Genbank accession number XM_658791), *Neurospora crassa* (Genbank accession number XM_957579) and *Cyberlindnera jadinii CAT2* (Genbank accession number AB641826). Any of these, and other such carnitine acetyltransferase-encoding nucleic acids, can be analyzed for the presence of 5' and 3' ORF nucleotides encoding possible mitochondrial- or peroxisomal-targeting sequences of amino acids and modified to eliminate such sequences. For example, the initial approximately 66 base pairs of the *Saccharomyces cerevisiae CAT2* or the *Candida tropicalis CAT2* coding sequence can be excluded to eliminate mitochondrial targeting of the enzyme, while deletion of the

terminal 9 base pairs of the coding sequence that encode a 3-amino acid persoxisomal targeting sequence, i.e., PTS1, (AKL or PKL motif) eliminates peroxisomal targeting of the enzyme (see, e.g., Elgersma *et al.* (1995) *EMBO J.* 14: 3472-3479 and Kawachi et al. (1996) *J. Biochem.* 120:731-735). Similarly, the initial approximately 120 base pairs of the *Aspergillus nidulans AcuJ* coding sequence can be excluded to possibly eliminate mitochondrial targeting of the enzyme, while deletion of the terminal 9 base pairs of the coding sequence that encode a 3-amino acid PTS1 (AKL motif) may eliminate peroxisomal targeting of the enzyme (see, e.g., Hynes et al. (2011) *Eukarot. Cell* 10:547-555). In general, yeast mitochondrial targeting sequences occur within the initial 10-90 N-terminal amino acid residues of a mitochondrial protein, have a significant arginine composition and very few to no negatively charged residues. Prediction tools, e.g., MitoProt, TargetP, Predotar and TPpred2, can be used in evaluating an amino acid sequence for identification of possible mitochondrial targeting sequences (see, e.g., Claros (1995) *Comput. Appl. Sci.* 11:441-447; Emanuelsson et al. (2000) *J. Mol. Biol.* 300:1005-1016; Small et al. (2004) *Proteomics* 4:1581-1590; Savojardo et al. (2014) *Bioinformatics* 30:2973-2974). Yeast peroxisomal targeting sequences generally occur at the C-terminus of a peroxisomal protein. Generally, the 3-amino acid consensus sequence of a yeast PTS1 has an initial amino acid containing a small, uncharged side chain (e.g., serine, alanine and cysteine), followed by a positively charged residue (e.g., lysine, arginine and histidine) and ending with a leucine residue; however, variants (e.g., PKL and others) of the consensus sequence do occur. Another example of a peroxisomal targeting signal sequence is the tripeptide AKI of the *Candida tropicalis* trifunctional enzyme hydratase-dehydrogenase-epimerase (HDE) (see, e.g., Aitchison et al. (1991) *J. Biol. Chem.* 266(34):23197-23203. Prediction tools, e.g., PTSI Predictor (mendel.imp.ac.at/mendeljsp/sat/pts1/PTS1predictor.jsp), can be used in evaluating an amino acid sequence for identification of possible peroxisomal targeting signal sequences (see, e.g., Brocard and Hartig (2006) *Biochim. Biophys. ACTA* 1763:1565-1573).

The promoter used for regulating transcription of a heterologous nucleic acid encoding a carnitine acetyltransferase that is more active or less active than an endogenous cytosolic carnitine acetyltransferase can also be modified. For example, the amount of a more active carnitine acetyltransferase protein expressed in the cytosol may be increased by including in the heterologous nucleic acid a stronger heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism.

Alternatively, decreasing carnitine acetyltransferase activity in the cytosol can be accomplished by modifying the amount of cytosolic carnitine acetyltransferase protein expression in a cell, for example, by replacing the wild-type promoter of an endogenous cytosolic carnitine acetyltransferase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type cytosolic carnitine acetyltransferase such that the encoded modified or substituted carnitine acetyltransferase protein has a reduced enzyme activity. For example, expression of a host cytosolic carnitine acetyltransferase activity can be decreased by disruption (e.g., knockout, insertion mutagenesis, the like and combinations thereof) of a host gene encoding the protein, or by decreasing the activity of the promoter (e.g., through addition of repressor sequences to the promoter or 5'UTR or replacing the promoter) that controls transcription of a cytosolic carnitine acetyltransferase gene using recombinant molecular biology techniques known in the art and/or described herein. In one embodiment, a diploid yeast, such as, for example, a *Candida* yeast, when used as a host microorganism can be subjected to genetic modification in which one of the two alleles of a cytosolic carnitine acetyltransferase gene is disrupted or deleted. In so doing, a single allele of the gene remains for a reduced amount of carnitine acetyltransferase expression in the microorganism and a reduced amount of the protein in the cytosol. This effectively reduces and/or slows the amount of acetyl-carnitine that is processed into acetyl-CoA in the cytosol.

One method for disrupting an endogenous carnitine acetyltransferase gene is by recombinantly inserting a heterologous nucleic acid (e.g., a nucleotide sequence encoding a selectable marker such as an enzyme that restores an auxotrophic host organism to prototrophy) into the endogenous gene, thereby generating an engineered organism deficient in cytosolic carnitine acetyltransferase activity. This can be done, for example, through homologous recombination in which a heterologous nucleic acid containing sequences of the endogenous cytosolic carnitine acetyltransferase gene and a disrupting sequence (e.g., a knock out gene cassette as described herein) is introduced into a host cell or microorganism. Nucleic acids encoding a cytosolic carnitine acetyltransferase can be obtained from a number of sources, including, for example, yeast cells. Genomic DNA from cell sources can be amplified using oligonucleotide primers based on the nucleotide sequence of a cytosolic carnitine acetyltransferase-encoding gene. For example, in some yeast strains, a cytosolic carnitine acetyltransferase is encoded by a *YAT* gene. Nucleotide sequences encoding (and the amino acid sequences of) *Candida viswanathii* Yat1p (amino acid SEQ ID NO: 6 and nucleotide SEQ ID NO: 63) are provided herein.

Additional non-limiting examples of nucleic acids encoding cytoplasmic carnitine acetyltransferase include *Saccharomyces cerevisiae YAT1* (Genbank accession number X74553), *Aspergillus nidulans FacC* (Genbank accession number AF023156), *Cyberlindnera jadinii YAT1* (Genbank accession number AB641829), *Candida dubliniensis YAT1* (Genbank accession number XM_002416790) and *Candida albicans* (Genbank accession number AF525683).

Modification of acetyl-CoA generation through oxidative metabolism

Included in the biological production platform systems and components thereof provided herein are embodiments in which the generation of acetyl-CoA in a cell or organism is modified. In some embodiments, the cellular processing of fatty acids, such as those obtained from external carbon sources (e.g., non-fermentable carbon sources) and internal, cell-generated sources (including, for example, but not limited to, fatty acids generated by catabolism of alkanes, fatty alcohols and fatty aldehydes), can be directed toward acetyl-CoA-generating oxidative metabolism pathways in cells. In some embodiments, the processing of fatty acids can be directed toward oxidative metabolism (e.g., ω- and/or β-oxidation) and away from cellular pathways, such as lipid synthesis pathways, that may not be involved in target molecule production. Accordingly, provided herein are cells, organisms, compositions and methods in which cellular carbon flux has been modified through alterations in cellular oxidative metabolism and/or fatty acid activation. In particular embodiments, cellular carbon flux has been modified to increase the production of acetyl-CoA in a cell through altering oxidative metabolism and/or fatty acid activation. Carbon flux modifications involving oxidative metabolism are particularly useful in embodiments in which alternative, non-carbohydrate carbon sources (e.g., some non-fermentable carbon sources) are used as a feedstock for modified cells and organisms in target molecule production.

For example, some organisms (e.g., some species of *Candida*, *Yarrowia*, *Pichia*, *Debaryomyces*, *Acinetobacter, Bacillus, Mycobaterium, Pseudomonas, Sphingomonas, Alcanivorax and Rhodococcus*) are able to endogenously assimilate alkanes as a carbon source. A primary pathway for alkane assimilation (also referred to as monoterminal alkane oxidation), which occurs in association with the endoplasmic reticulum and peroxisomes in eukaryotes, is through the initial conversion of alkanes to fatty acids which can then be metabolized in the cells. The conversion can occur through a three-step process as follows: (1)

terminal hydroxylation of alkane by a cytochrome P450-dependent monooxygenase system (e.g, *ALK* gene products of the CYP52 family as a terminal oxidase and an NADPH-dependent cytochrome P450 reductase (e.g., *CPR1*-encoded) for electron transfer) which yields a fatty alcohol; (2) conversion of the terminal hydroxy group of the alcohol to a fatty aldehyde in reactions involving fatty alcohol dehydrogenase (e.g., *ADH*) or fatty alcohol oxidase (e.g., *FAO*); and (3) conversion of the fatty aldehyde to a fatty acid by a fatty aldehyde dehydrogenase. The resulting fatty acid can then be subject to the same metabolic processing as is a fatty acid taken up directly by the organism.

Fatty acids can be metabolized in several ways depending on the type of cell or organism. Many fatty acid metabolic pathways, including β-oxidation, lipid biosynthesis, and protein acylation, require that a fatty acid be activated by thioesterification to coenzyme A (i.e., acyl-CoA), or to an acyl carrier protein (i.e., acyl-ACP), prior to being metabolized. As used herein, "activation" with reference to fatty acids refers to the thioesterification of a fatty acid with a carrier molecule such as coenzyme A (Co-A) or acyl carrier protein (ACP). A fatty acid that has undergone activation into an acyl-CoA or fatty acid-ACP molecule is referred to as an activated fatty acid. The thioesterification reaction can be catalyzed by acyl-CoA synthetase enzymes. There are multiple enzymes having acyl-CoA synthetase activity in cells which differ based on cellular localization (e.g., plasma membrane, cytosol, endoplasmic reticulum membrane, peroxisomes) and substrate (e.g., fatty acid carbon chain length) specificity. In general, plasma membrane-associated acyl-CoA synthetases often are more specific for very long chain fatty acids and are involved in transport of these hydrophobic molecules across the membrane concurrent with activation of the fatty acids to acyl-CoA. Once activated, the acyl-CoA can then be used in a number of metabolic pathways, only one of which is β-oxidation. Thus, fatty acids activated at the plasma membrane and/or in the cytosol can represent possible "losses" of carbon atoms to cellular synthesis pathways (e.g., lipid synthesis) at the expense of other target molecule production pathways (e.g., oxidative metabolism, malonyl-CoA production and organic acid, polyketide and/or terpene synthesis). Therefore, in some embodiments of cells and organisms provided herein for use in target molecule production systems, it can be beneficial to capture the carbon atoms of free fatty acids for use in target molecule production and decrease activated fatty acid (e.g., acyl-CoA) flow into cellular pathways not associated with target molecule production. As described herein, methods of enhancing fatty acid carbon flow through oxidative metabolism include, but are not limited to, modification of activities of cellular ω- and/or β-oxidation systems and acyl-CoA synthetase activities.

Modification of ω-oxidation

In the oxidative metabolism pathway referred to as ω-oxidation (or diterminal oxidation), fatty acids can be converted to dicarboxylic acids (diacids). Several enzyme activities (e.g., cytochrome P450 hydroxylase complex, fatty alcohol oxidase and fatty aldehyde dehydrogenase) can be involved in the process of ω-oxidation. The term "ω-oxidation pathway" as used herein, refers to a cellular metabolic pathway constituted by a series of enzymatic activities through which fatty acids and alkanes are converted to dicarboxylic acids. Some cells and microorganisms (e.g., species of yeast and bacteria) endogenously express the enzyme activities, and those that do not can be genetically modified to provide a heterologous ω-oxidation pathway by introducing nucleic acids encoding the required enzymes into cells and expressing the proteins therein. Free fatty acids internalized into cells, or generated within cells (e.g., by oxidation of internalized alkanes), can directly enter into and be processed in the ω-oxidation pathway without prior activation to acyl-CoA. In some embodiments of the cell-based production systems and methods provided herein, dicarboxylic acids can be a target molecule. In some embodiments, e.g., in which some shorter-chain dicarboxylic acids are a target molecule, or dicarboxylic acids are not a target molecule (or a co-target molecule along with one or more other desired products), dicarboxylic acids generated by ω-oxidation can be further oxidized through β-oxidation. Dicarboxylic acids can traverse peroxisomal membranes in eukaryotic cells and be metabolized to yield acetyl-CoA that can be used in target molecule generation. Because free fatty acids can be processed through ω-oxidation without being activated by thioesterification with Co-A, and diacids resulting from ω-oxidation of fatty acids readily move into peroxisomes, the ω-oxidation pathway can serve as a cellular gateway for funneling internalized fatty acids into oxidative metabolism and away from cytosolic activation that is required for use of fatty acids in other cellular pathways that may not be involved in target molecule production.

The term "ω-oxidation activity" refers to any of the activities in the ω-oxidation pathway utilized to metabolize alkanes and fatty acids. The activities that may be utilized in this metabolic pathway include, but are not limited to, monooxygenase activity (e.g., cytochrome P450 activity), monooxygenase reductase activity (e.g., cytochrome P450 reductase activity), alcohol dehydrogenase activity (e.g., fatty alcohol dehydrogenase activity or long-chain alcohol dehydrogenase activity), fatty alcohol oxidase activity and fatty aldehyde dehydrogenase

activity. In some embodiments of the cells, organisms, compositions and methods provided herein, the ω-oxidation activity of a cell or organism is modified. In one embodiment, one or more of the activities in the ω-oxidation pathway can be modified. In particular embodiments, one or more of the activities in the ω-oxidation pathway can be increased.

Modification of a monooxygenase activity

The initial step in the ω-oxidation pathway is the conversion of a fatty acid to a corresponding fatty alcohol and involves NADPH and molecular oxygen. A cytochrome P450 enzyme (e.g., monooxygenase activity, EC 1.14.14.1) often catalyzes the insertion of one atom of oxygen bound to the heme group in cytochrome P450 into an organic substrate (RH) while the other oxygen atom is reduced to water. A cytochrome P450 reductase catalyzes the reductive splitting of the oxygen and transfer of electrons from NADPH to cytochrome P450. Insertion of the oxygen atom near the omega carbon of a substrate yields an alcohol derivative of the original starting substrate (e.g., yields a fatty alcohol). In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a monooxygenase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a monooxygenase, may be modified to decrease the amount and/or activity of a monooxygenase, or may be modified to alternately increase and decrease the amount and/or activity of a monooxygenase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a monooxygenase in a cell is increased. Increasing the amount and/or activity of a monooxygenase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway involving oxidative metabolism and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, the monooxygenase activity is unchanged in a host or engineered cell or organism. In one embodiment, the amount and/or activity of a host monooxygenase can be increased, for example, by increasing the number of copies of a nucleic acid encoding a monooxygenase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid), by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a

monooxygenase, or by increasing the number of copies of a nucleic acid encoding a monooxygenase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a monooxygenase. In some embodiments, a monooxygenase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of monooxygenase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a monooxygenase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a monooxygenase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

In some embodiments, a cytochrome P450 monooxygenase enzyme can be a fungal or bacterial protein. In a particular embodiment, the monooxygenase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*), *Fusarium* (e.g., *F. oxysporum*), *Bacillus* (e.g., *B. megaterium*, *B. subtilis*) protein. *Candida tropicalis* contains a family of cytochrome P450 genes referred to as *CYP* genes. Examples of *Candida viswanathii* nucleotide sequences encoding polypeptides having monooxygenase activities are provided herein (nucleotide SEQ ID NO: 99 and amino acid SEQ ID NO: 45) and in International patent application no. PCT/US2012/045615 (publication no. WO 2013/106730). Additional nonlimiting examples of nucleotide sequences encoding polypeptides having monooxygenase activity include: *Candida tropicalis CYP52A12* (Genbank accession no. AY230498), *Candida tropicalis CYP52A13* (Genbank accession no. AY230499), *Candida tropicalis CYP52A14* (Genbank accession no. AY230500), *Candida tropicalis CYP52A15* (Genbank accession no. AY230501), *Candida tropicalis CYP52A16* (Genbank accession no. AY230502), *Candida tropicalis CYP52A17* (Genbank accession no. AY230504), *Candida tropicalis CYP52A18* (Genbank accession no. AY230505), *Candida tropicalis CYP52A19* (Genbank accession no. AY230506), *Candida tropicalis CYP52A20* (Genbank accession no. AY230507), *Candida tropicalis CYP52D2* (Genbank accession no. AY230503), *Bacillus megaterium CYPBM3* (Genbank accession no. KC839476) and *Fusarium oxysporum CYP505* (Genbank accession no. AB030037).

Monooxygenase activity can be provided by any suitable polypeptide, such as a cytochrome P450 polypeptide (CYP450) in certain embodiments. Examples of a polypeptide having CYP450 activity include CYP52A12, a CYP52A13, a CYP52A14, a CYP52A15, a CYP52A16, a CYP52A17, a CYP52A18, a CYP52A19, a CYP52A20, a CYP52D2, and/or a BM3. In some

embodiments, the activity can be a single polypeptide with both monooxygenase and monooxygenase reductase activities (e.g., *B. megaterium* cytochrome P450:NADPH P450 reductase, *Fusarium oxysporum* CYP505). Presence, absence or amount of cytochrome P450 activity can be detected by any suitable method known in the art. For example, detection can be performed by assaying a reaction containing cytochrome P450 (CYP52A family) and NADPH - cytochrome P450 reductase (see, e.g., Craft et al. (2003) *Appl. Environ. Microbiol.* 69: 5983 and 5992). Nucleic acid sequences encoding native and/or modified CYP450 sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a heterologous nucleic acid encoding a monooxygenase can also be modified. For example, the amount of a monooxygenase protein expressed in a particular cellular location may be increased by including in the heterologous nucleic acid a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequence of a *Candida viswanathii* HDE gene promoter is provided herein as are examples of additional fatty acid-inducible promoters. Promoter elements from different monooxygenase-encoding genes can have differing responsiveness to induction by various carbon sources. Thus, the amount of a monooxygenase protein expressed in a cell or organism can be modified by using heterologous promoters from different cytochrome P450 monoxygenase genes to regulate transcription of a monoxgenase-encoding nucleic acid that is introduced into a host cell and by the carbon source provided to the modified cell or organism. Non-limiting examples of assays suitable for assessing induction of cytochrome P450 (or other protein) expression by a carbon source or feedstock include RT-PCR or qRT-PCR after the host cell or microorganism has been exposed to the chosen carbon source or feedstock for varying amounts of time.

Decreasing monooxygenase activity in a cell can be accomplished by modifying the amount of monooxygenase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous monooxygenase gene in a cell or organism with a weaker heterologous

promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type monooxygenase such that the encoded modified or substituted monooxygenase protein has a reduced enzyme activity.

Modification of a cytochrome P450 reductase activity

A cytochrome P450 reductase (e.g., monooxygenase reductase activity or NADPH:cytochrome oxidoreductase (NCP); EC 1.6.2.4) can catalyze the reduction of the heme-thiolate moiety in cytochrome P450 by transferring electrons to the cytochrome P450. This activity recycles cytochrome P450 and makes it available for further use in catalyzing reactions that occur in ω-oxidation of fatty acids. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a cytochrome P450 reductase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a cytochrome P450 reductase, may be modified to decrease the amount and/or activity of a cytochrome P450 reductase, or may be modified to alternately increase and decrease the amount and/or activity of a cytochrome P450 reductase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a cytochrome P450 reductase in a cell is increased. Increasing the amount and/or activity of a cytochrome P450 reductase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway involving oxidative metabolism and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, the cytochrome P450 reductase activity is unchanged in a host or engineered cell or organism. In one embodiment, the amount and/or activity of a host cytochrome P450 reductase can be increased, for example, by increasing the number of copies of a nucleic acid encoding a cytochrome P450 reductase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid), by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a cytochrome P450 reductase, or by increasing the number of copies of a nucleic acid encoding a cytochrome P450 reductase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a cytochrome P450 reductase. In some embodiments, a cytochrome P450 reductase is endogenous to the host cell

or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of cytochrome P450 reductase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a cytochrome P450 reductase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a cytochrome P450 reductase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

In some embodiments, a cytochrome P450 reductase enzyme can be a yeast or bacterial protein. In a particular embodiment, the reductase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*) or *Bacillus* (e.g., *B. megaterium*) protein. In a particular embodiment, the cytochrome P450 reductase enzyme can be a *Candida* yeast protein. *Candida tropicalis* contains two alleles of a cytochrome P450 reductase gene referred to as *CPRa* and *CPRb*. Examples of *Candida viswanathii* nucleotide sequences encoding cytochrome P450 reductase activities are provided herein (nucleotide SEQ ID NO: 90 and amino acid SEQ ID NO: 45) and in International patent application no. PCT/US2012/045615 (publication no. WO 2013/106730). Additional non-limiting examples of nucleotide sequences encoding polypeptides having cytochrome P450 reductase activity include: *Candida tropicalis* (Genbank accession nos. AY705446, AY823228), *Candida bombicola* (Genbank accession no. EF050789), *Bacillus megaterium CYPBM3* (Genbank accession no. KC839476), *Bacillus megaterium* (Genbank accession no. FJ859036).

Presence, absence or amount of cytochrome P450 reductase activity can be detected by any suitable method known in the art. For example, detection can be performed by assaying a reaction containing cytochrome c and NADPH and measuring the rate of cytochrome reduction by monitoring absorbance (see, e.g., He and Chen (2005) *Yeast* 22:481-491; Van Bogaert et al. (2007) *Yeast* 7:922-928). Nucleic acid sequences encoding native and/or modified cytochrome P450 reductase also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a heterologous nucleic acid encoding a cytochrome P450 reductase can also be modified. For example, the amount of a cytochrome

P450 reductase protein expressed in a particular cellular location may be increased by including in the heterologous nucleic acid a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequence of a *Candida viswanathii* HDE gene promoter is provided herein as are examples of additional fatty acid-inducible promoters.

Alternatively, decreasing cytochrome P450 reductase activity in a cell can be accomplished by modifying the amount of cytochrome P450 reductase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous cytochrome P450 reductase gene in an organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type cytochrome P450 reductase such that the encoded modified or substituted cytochrome P450 reductase protein has a reduced enzyme activity.

Modification of an alcohol dehydrogenase activity

A second step in the ω-oxidation pathway generally is the conversion of a fatty alcohol to a corresponding fatty aldehyde and involves NAD^+- or $NADP^+$-dependent fatty alcohol dehydrogenases and/or hydrogen peroxide-producing fatty alcohol oxidases. Oxidation of the alcohol to an aldehyde may be performed by an enzyme in the fatty alcohol oxidase family (e.g., long-chain fatty alcohol oxidase EC 1.1.3.20), or an enzyme in the alcohol dehydrogenase family (e.g., fatty alcohol dehydrogenase; EC 1.1.1.1). An alcohol dehydrogenase (e.g., fatty alcohol dehydrogenase, long-chain alcohol dehydrogenase) can catalyze the removal of a hydrogen from an alcohol to yield an aldehyde or ketone and a hydrogen atom and NADH.

In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of an alcohol dehydrogenase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of an alcohol dehydrogenase, may be modified to decrease the amount and/or activity of an alcohol dehydrogenase, or may be modified to alternately increase and decrease the amount and/or activity of an alcohol dehydrogenase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of an alcohol dehydrogenase in a cell is increased. Increasing the amount and/or activity of an alcohol dehydrogenase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway involving oxidative metabolism and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, the alcohol dehydrogenase activity is unchanged in a host or engineered cell or organism. In one embodiment, the amount and/or activity of a host alcohol dehydrogenase can be increased, for example, by increasing the number of copies of a nucleic acid encoding an alcohol dehydrogenase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid), by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an alcohol dehydrogenase, or by increasing the number of copies of a nucleic acid encoding an alcohol dehydrogenase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an alcohol dehydrogenase. In some embodiments, an alcohol dehydrogenase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of alcohol dehydrogenase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding an alcohol dehydrogenase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding an alcohol dehydrogenase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

In some embodiments, an alcohol dehydrogenase enzyme can be a yeast or bacterial protein. In a particular embodiment, the alcohol dehydrogenase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*) or *Bacillus* (*B. stearothermophilus*) protein. In a particular embodiment, the alcohol dehydrogenase enzyme can be a *Candida* yeast protein. *Candida tropicalis* contains at least 6 genes encoding alcohol dehydrogenases. Examples of *Candida viswanathii* nucleotide sequences encoding polypeptides having alcohol dehydrogenase activities are provided herein (nucleotide SEQ ID NO: 100 and amino acid SEQ ID NO: 46) and in International patent application no. PCT/US2012/045615 (publication no. WO 2013/106730). Additional examples of nucleotide sequences encoding polypeptides having alcohol dehydrogenase activity include, but are not limited to: *Candida tropicalis ADH1* (Genbank accession no. XM_002546589), *Candida utilis ADH1* (Genbank accession no. DQ397054), *Candida albicans ADH1* (Genbank accession no.

X81694), *Aspergillus flavus ADH1* (Genbank accession no. L27434), *Yarrowia lipolytica ADH1* (Genbank accession no. AF175271), *Yarrowia lipolytica ADH2* (Genbank accession no. AF175272), *Yarrowia lipolytica ADH3* (Genbank accession no. AF175273), *Bacillus stearothermophilus ADH-HT* (Genbank accession no. Z27089), *Pseudomonas putida ADHA* (Genbank accession no. AF052750).

Presence, absence or amount of alcohol dehydrogenase activity can be detected by any suitable method known in the art. For example, detection can be performed using spectrophotometric assays (see, e.g., Gatter et al. (2014) *FEMS Yeast Res.* 14:858-872). Nucleic acid sequences encoding native and/or modified alcohol dehydrogenase also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a heterologous nucleic acid encoding an alcohol dehydrogenase can also be modified. For example, the amount of an alcohol dehydrogenase protein expressed in a particular cellular location may be increased by including in the heterologous nucleic acid a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequence of a *Candida viswanathii* HDE gene promoter is provided herein as are examples of additional fatty acid-inducible promoters.

Alternatively, decreasing alcohol dehydrogenase activity in a cell can be accomplished by modifying the amount of alcohol dehydrogenase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous alcohol dehydrogenase gene in an organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type alcohol dehydrogenase such that the encoded modified or substituted alcohol dehydrogenase protein has a reduced enzyme activity.

Modification of fatty alcohol oxidase activity

A fatty alcohol oxidase (e.g., long-chain alcohol oxidase, EC 1.1.3.20) enzyme can catalyze the oxidation of a fatty alcohol to yield a fatty aldehyde in the peroxisome of a cell. In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a fatty alcohol oxidase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a fatty alcohol oxidase, may be modified to decrease the amount and/or activity of a fatty alcohol oxidase, or may be modified to alternately increase and decrease the amount and/or activity of a fatty alcohol oxidase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a fatty alcohol oxidase in a cell is increased. Increasing the amount and/or activity of a fatty alcohol oxidase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway involving oxidative metabolism and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, the fatty alcohol oxidase activity is unchanged in a host or engineered cell or organism. In one embodiment, the amount and/or activity of a host fatty alcohol oxidase can be increased, for example, by increasing the number of copies of a nucleic acid encoding a fatty alcohol oxidase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid), by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a fatty alcohol oxidase, or by increasing the number of copies of a nucleic acid encoding a fatty alcohol oxidase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a fatty alcohol oxidase. In some embodiments, a fatty alcohol oxidase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of fatty alcohol oxidase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a fatty alcohol oxidase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a fatty alcohol oxidase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

In some embodiments, a fatty alcohol oxidase enzyme can be a yeast or bacterial protein. In a particular embodiment, the fatty alcohol oxidase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*), *Yarrowia* (e.g., *Y. lipolytica*) or *Bacillus* (e.g, *B.*

stearothermophilus) protein. In a particular embodiment, the fatty alcohol oxidase enzyme can be a *Candida* yeast protein. *Candida tropicalis* contains two genes encoding fatty alcohol oxidase. Examples of *Candida viswanathii* nucleotide sequences encoding polypeptides having fatty alcohol oxidase activities are provided herein (nucleotide SEQ ID NO: 101 and amino acid SEQ ID NO: 47) and in International patent application no. PCT/US2012/045615 (publication no. WO 2013/106730). Additional examples of nucleotide sequences encoding polypeptides having fatty alcohol oxidase activity include, but are not limited to: *Candida tropicalis FAO1* (Genbank accession no. AY538780), *Candida tropicalis FAO2a* (Genbank accession no. AY538781), *Candida tropicalis FAO2b* (Genbank accession no. AY538782).

Presence, absence or amount of fatty alcohol oxidase activity can be detected by any suitable method known in the art. For example, detection can be performed using a two-enzyme coupled reaction assay (see, e.g., Eirich et al. (2004) *Appl. Environ. Microbiol.* 70(8):4872-4879). Nucleic acid sequences encoding native and/or modified fatty alcohol oxidase also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a heterologous nucleic acid encoding a fatty alcohol oxidase protein expressed in a particular cellular location may be increased by including in the heterologous nucleic acid a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequence of a *Candida viswanathii* HDE gene promoter is provided herein as are examples of additional fatty acid-inducible promoters.

Alternatively, decreasing fatty alcohol oxidase activity in a cell can be accomplished by modifying the amount of fatty alcohol oxidase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous fatty alcohol oxidase gene in an organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type fatty alcohol oxidase such that the encoded modified or substituted fatty alcohol oxidase protein has a reduced enzyme activity.

Modification of aldehyde dehydrogenase activity

A third step in the ω-oxidation pathway is generally the conversion of a fatty aldehyde to a corresponding fatty acid and involves NAD^+- or $NADP^+$-dependent fatty aldehyde dehydrogenases (e.g., long-chain-aldehyde dehydrogenase or fatty aldehyde dehydrogenase; EC 1.2.1.48). In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a fatty aldehyde dehydrogenase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a fatty aldehyde dehydrogenase, may be modified to decrease the amount and/or activity of a fatty aldehyde dehydrogenase, or may be modified to alternately increase and decrease the amount and/or activity of a fatty aldehyde dehydrogenase depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a fatty aldehyde dehydrogenase in a cell is increased. Increasing the amount and/or activity of a fatty aldehyde dehydrogenase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway involving oxidative metabolism and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, the fatty aldehyde dehydrogenase activity is unchanged in a host or engineered cell or organism. In one embodiment, the amount and/or activity of a host fatty aldehyde dehydrogenase can be increased, for example, by increasing the number of copies of a nucleic acid encoding a fatty aldehyde dehydrogenase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid), by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a fatty aldehyde dehydrogenase, or by increasing the number of copies of a nucleic acid encoding a fatty aldehyde dehydrogenase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a fatty aldehyde dehydrogenase. In some embodiments, a fatty aldehyde dehydrogenase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of fatty aldehyde dehydrogenase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a fatty aldehyde dehydrogenase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a

fatty aldehyde dehydrogenase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

In some embodiments, a fatty aldehyde dehydrogenase enzyme can be a yeast protein. In a particular embodiment, the fatty aldehyde dehydrogenase enzyme can be a *Candida* (e.g., *C. tropicalis*, *C. viswanathii*, *C. maltosa*, *C. cloacae*) or a *Yarrowia* (e.g., *Y. lipolytica*) yeast protein. In a particular embodiment, the fatty aldehyde dehydrogenase enzyme can be a *Candida* yeast protein. Examples of *Candida viswanathii* nucleotide sequences encoding polypeptides having fatty aldehyde dehydrogenase activities are provided, for example, in International patent application no. PCT/US2012/045615 (publication no. WO 2013/106730). Additional examples of nucleotide sequences encoding polypeptides having fatty aldehyde dehydrogenase activity include, but are not limited to: *Yarrowia lipolytica HFD1* (Genbank accession no. AB935099), *Yarrowia lipolytica HFD2A* (Genbank accession no. AB935101), *Yarrowia lipolytica HFD2B* (Genbank accession no. AB935103), *Yarrowia lipolytica HFD3* (Genbank accession no. AB935104), *Yarrowia lipolytica HFD4* (Genbank accession no. AB935106).

Presence, absence or amount of fatty aldehyde dehydrogenase activity can be detected by any suitable method known in the art. For example, detection can be performed using enzyme activity assays (see, e.g., Iwama et al. (2014) *J. Biol. Chem.* 289(48):33275-33286). Nucleic acid sequences encoding native and/or modified fatty aldehyde dehydrogenase also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cell or organism exhibits decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a heterologous nucleic acid encoding an fatty aldehyde dehydrogenase can also be modified. For example, the amount of a fatty aldehyde dehydrogenase protein expressed in a particular cellular location may be increased by including in the heterologous nucleic acid a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequence of a *Candida viswanathii* HDE gene promoter is provided herein as are examples of additional fatty acid-inducible promoters.

Alternatively, decreasing fatty aldehyde dehydrogenase activity in a cell can be accomplished by modifying the amount of fatty aldehyde dehydrogenase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous fatty aldehyde dehydrogenase gene in an organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type fatty aldehyde dehydrogenase such that the encoded modified or substituted fatty aldehyde dehydrogenase protein has a reduced enzyme activity.

Modification of β-oxidation

Another oxidative metabolism pathway, referred to as β-oxidation, is generally a degradative pathway through which fatty acids, typically in the form of fatty acid-CoA esters, can be broken down to shorter chain acyl-CoA and acetyl-CoA. In fungi and plant cells, β-oxidation can occur in peroxisomes, whereas in animal cells it additionally can take place in mitochondria. The β-oxidation pathway generally includes four main reaction steps resulting in an acyl-CoA that is shortened by two carbon atoms which are released as acetyl-CoA. The shortened acyl-CoA molecule can re-enter the pathway after each cycle and be subjected to another removal of two carbons from the acyl carbon chain. As such, the β-oxidation pathway can generate significant amounts of acetyl-CoA and is a major source of acetyl-CoA in cells. Alteration of enzyme activities in the the β-oxidation pathway can also provide for the generation of fatty acid or diacid target molecules including, but not limited to, adipic acid, suberic acid, sebacic acid and dodecanedioic acid (DDDA). Provided herein are cells, organisms, compositions and methods in which cellular carbon flux has been modified through one or more alterations in the β-oxidation pathway. In some embodiments, the β-oxidation pathway is modified by modifying one or more activities in the pathway. In particular embodiments, the β-oxidation pathway is modified to increase the generation of acetyl-CoA in a cell for use in target molecule production. For example, the β-oxidation pathway can be modified to increase one or more activities in the pathway. In some embodiments, β-oxidation can be manipulated (e.g., decreasing one or more pathway activities and/or altering the specificity of one or more activities) to be used as a pathway for production of target fatty acids and diacids (e.g., adipic acid) of a particular carbon chain length.

The term "β-oxidation pathway" as used herein, refers to a series of cellular enzymatic activities utilized to metabolize fatty alcohols, fatty acids, or dicarboxylic acids. The activities utilized to metabolize fatty alcohols, fatty acids, or dicarboxylic acids can include, but are not limited to, acyl-CoA oxidase activity, acyl-CoA hydrolase activity, enoyl-CoA hydratase activity, 3-hydroxyacyl-CoA dehydrogenase activity, and acetyl-CoA C-acyltransferase activity. The term "β-oxidation activity" refers to any and/or all of the activities in the β-oxidation pathway utilized to metabolize fatty alcohols, fatty acids or dicarboxylic acids. Additional activities, referred to as β-oxidation peripheral or auxillary activities, can be involved in degradation of unsaturated fatty acids (i.e., fatty acid chains containing double bonds) and fatty acids containing modifications (e.g., hydroxyl, methyl, phenoxy groups) including, but not limited to enoyl CoA isomerase ((ECI) or enoyl-CoA delta isomerase 1, dodecenoyl-CoA isomerase, 3,2 trans-enoyl-CoA isomerase, acetylene-allene isomerase, delta3, delta2-enoyl-CoA isomerase, dodecenoyl-CoA delta isomerase, and EC 5.3.3.8), dienoyl CoA Isomerase (DCI, e.g., EC 5.3.3, Δ3,5,Δ2,4-dienoyl-CoA isomerase, Δ3,5,Δ2,4-dienoyl-coenzyme A isomerase) and 2,4-dienoyl-CoA reductase (DCR, e.g., EC 1.3.1.34).

There are also cellular compositions and activities that can be closely associated with β-oxidation. These include peroxisomal- and mitochondrial-related compositions and activities. For example, as described herein, such compositions and activities include, but are not limited to, acyl-CoA synthetases, thioesterases, peroxisomal transport proteins and peroxisomal biogenesis factors. Included in the cells, organisms, systems and methods provided herein are embodiments in which one or more of these β-oxidation-associated compositions and/or activities are modified. In some embodiments, a β-oxidation-associated composition or activity is modified to enhance β-oxidation activity.

Modification of acyl-CoA oxidase activity

Typically, the first step in the β-oxidation pathway is oxidation of acyl-CoA, which is carried out by the enzyme acyl-CoA oxidase (e.g., EC 1.3.3.6). This step can be a rate-limiting step in β-oxidation. The term "acyl-CoA oxidase activity" as used herein refers to the enzymatic activity (e.g., catalytic activity) of an acyl-CoA oxidase. An acyl-CoA oxidase can catalyze the following chemical reaction:

$$\text{acyl-CoA} + O_2 \leftrightarrow \text{trans-2-enoyl-CoA} + H_2O_2$$

Acyl-CoA oxidase enzymes generally contain FAD from which two electrons are transferred to oxygen to generate H_2O_2.

Different cells contain different types, and numbers of types, of acyl-CoA oxidase activities. For example, *Saccharomyces cerevisiae* expresses only one acyl-CoA oxidase, Pox1p/Fox1p, which has activity on acyl-CoA substrates with a broad range of carbon chain lengths. In contrast, other organisms, e.g., species of *Candida*, *Yarrowia*, *Arabidopsis*, can have multiple genes encoding different proteins having acyl-CoA oxidase activities with varying substrate specificities. In some embodiments, acyl-CoA oxidase activity refers to its enzyme activity (or lack thereof) on a selective set of substrates. The activity of an acyl-CoA oxidase can be affected by its ability to bind a substrate, oxidize a substrate and/or release a product. In some embodiments, an acyl-CoA oxidase is active in one compartment of a cell and not in another compartment of the cell. In some embodiments, an acyl-CoA oxidase activity is from a peroxisome.

Different Acyl-CoA oxidases can display different carbon chain-length substrate specificities. Some acyl-CoA oxidases display broad chain-length specificity and can accept any fatty acyl-CoA (or diacyl-CoA) as a substrate. However, some acyl-CoA oxidases can display narrow chain-length specificity. For example, the acyl-CoA oxidase activity encoded by the *POX4* gene of *Candida* strain ATCC 20336 has a relatively broad carbon chain-length specificity and exhibits a higher specific activity for acyl-CoA molecules with shorter carbon chain lengths (e.g., less than 10 carbons). The Pox5 enzyme from *Candida* strain ATCC 20336 displays optimal activity on fatty acid substrates having 12 to 18 carbons (C12 - C18) in the carbon chain, a decreased activity on substrates having less than 10 carbons (C10) in the carbon chain and has low activity on C6 and C8 substrates. In a cell with such a *Candida* Pox5 as the only functional acyl-CoA oxidase, long chain fatty acyl-CoA or diacyl-CoA substrates can be shortened to about 8 carbons and do not typically enter another cycle of β-oxidation. The shorter substrates (e.g., a C8 fatty acid or dicarboxylic acid) are not typically recognized as a substrate by Pox5. In this instance, an acyl-CoA would not be completely broken down to acetyl-CoA units. Instead, the chain-length substrate specificity of the acyl-CoA oxidase in such a cell (which would limit further degradation of an acyl-CoA once it has been broken down into about an 8-carbon chain) effectively controls the chain length of an acid or diacid produced through break down of fatty acids through β-oxidation. The shorter substrates (e.g., a C8 acyl-CoA) would remain intact, the

CoA would be removed by peroxisomal thioesterases and the fatty acid or dicarboxylic acid (e.g., an α,ω - dicarboxylic acid) product is secreted from the cell. In this way, β-oxidation can be manipulated to be used as a pathway for production of target fatty acids and diacids (e.g., adipic acid) of a particular carbon chain length.

In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of one or more acyl-CoA oxidases in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of an acyl-CoA oxidase, may be modified to decrease the amount and/or activity of an acyl-CoA oxidase, or may be modified to alternately increase and decrease the amount and/or activity of one or more acyl-CoA oxidases depending, for example, on the substrate specificity, target molecule(s) being produced, cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of an acyl-CoA oxidase in a cell is increased. Increasing the amount and/or activity of an acyl-CoA oxidase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway involving oxidative metabolism and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, the acyl-CoA oxidase activity is unchanged in a host or engineered cell or organism. In one embodiment, the amount and/or activity of a host acyl-CoA oxidase can be increased, for example, by increasing the number of copies of a nucleic acid encoding an acyl-CoA oxidase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid), by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an acyl-CoA oxidase, or by increasing the number of copies of a nucleic acid encoding an acyl-CoA oxidase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an acyl-CoA oxidase. In some embodiments, an acyl-CoA oxidase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of an acyl-CoA oxidase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding an acyl-CoA oxidase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding an acyl-CoA oxidase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

Non-limiting examples of organisms that include, or can be used as donors for, an acyl-CoA oxidase enzyme include yeast (e.g., *Candida*, *Saccharomyces*, *Debaryomyces*, *Meyerozyma*, *Lodderomyces*, *Scheffersomyces*, *Clavispora*, *Yarrowia*, *Pichia*, *Kluyveromyces*, *Eremothecium*, *Zygosaccharomyces*, *Lachancea*, *Nakaseomyces*), animals (e.g., *Homo*, *Rattus*), bacteria (e.g., *Escherichia*, *Pseudomonas*, *Bacillus*), or plants (e.g., *Arabidopsis*, *Nictotania*, *Cuphea*). In a particular embodiment, an acyl-CoA oxidase enzyme can be a *Candida* yeast protein. Examples of *Candida viswanathii* nucleotide sequences encoding acyl-CoA oxidases are provided herein (nucleotide SEQ ID NOS: 92 and 93 and amino acid SEQ ID NOS: 36, 37, 38 and 39) and in International patent application no. PCT/US2012/045622 (publication no. WO 2013/006733) and International patent application no. PCT/US2013/076739 (publication no. WO 2014/100504). Additional examples of nucleotide sequences encoding polypeptides having acyl-CoA oxidase activity include: *Saccharomyces cerevisiae POX1* (Genbank accession no. M27515), *Candida albicans POX1-3* (Genbank accession no. XM_716636), *Candida tropicalis POX2* (Genbank accession no. XM_002548031), *Candida tropicalis POX5* (Genbank accession no. XM_002548378),*Candida tropicalis POX4* (Genbank accession nos. AB031271, AB031272), *Candida maltosa POX2* (Genbank accession no. D21228), *Yarrowia lipolytica ACO1* (Genbank accession no. AJ001299), *Yarrowia lipolytica ACO2* (Genbank accession no. A001300), *Yarrowia lipolytica ACO3* (Genbank accession no. AJ001301), *Yarrowia lipolytica ACO4* (Genbank accession no. AJ001302), *Yarrowia lipolytica ACO5* (Genbank accession no. AJ001303), *Debaryomyces fabyri* (Genbank accession no. XM_015613952).

Presence, absence or amount of acyl-CoA oxidase activity can be detected by any suitable method known in the art and/or described herein. For example, detection can be performed using enzyme activity assays (see, e.g., Shimizu et al. (1979) *Biochem. Biophys. Res. Commun.* 91:108-113; Yao et al. (2014) *J. Braz. Chem. Soc.* 25(4):777-782); Kvannes and Flatmark (1991) *J. Biochem. Biophys. Methods* 23(2):135-149). Native and/or disrupted nucleic acid sequences encoding acyl-CoA oxidase (or other polypeptide) also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), and the amount of the nucleic acids or encoded proteins can be asessed using quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cells or organisms exhibit increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a heterologous nucleic acid encoding an acyl-CoA oxidase can also be modified. For example, the amount of an acyl-CoA oxidase protein expressed in a particular cellular location may be increased by including in the heterologous nucleic acid a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequence of a *Candida viswanathii* HDE gene promoter is provided herein as are examples of additional fatty acid-inducible promoters.

Alternatively, decreasing acyl-CoA oxidase activity in a cell can be accomplished by modifying the amount of acyl-CoA oxidase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous acyl-CoA oxidase gene in an organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type acyl-CoA oxidase such that the encoded modified or substituted acyl-CoA oxidase protein has a reduced enzyme activity. Reducing or eliminating the amount and/or activity of an acyl-CoA oxidase may be particularly beneficial in embodiments in which a target molecule and/or precursor or intermediate in the production of a target molecule contains a carbon chain of a particular length. In this case, the processing of fatty acids of particular chain lengths may be decreased or eliminated by decreasing the amount and/or activity of a particular acyl-CoA oxidase having activity on fatty acids of the particular chain length in a cell. Certain aspects of the cells, microorganisms, compositions and methods provided herein include one or more modifications to reduce or eliminate an acyl-CoA oxidase. One approach to reducing or eliminating the amount and/or activity of an acyl-CoA oxidase is by disrupting or deleting nucleic acid encoding the acyl-CoA oxidase in a host cell or microorganism to reduce or eliminate the acyl-CoA oxidase activity in the host relative to a cell or microorganism in which the gene(s) have not been modified. For example, expression of a host acyl-CoA oxidase activity can be decreased or eliminated by disruption (e.g., knockout, insertion mutagenesis, the like and combinations thereof) of a host gene encoding the protein, or by decreasing the activity of the promoter (e.g., through addition of repressor sequences to the promoter or 5'UTR or replacing the promoter) that controls transcription of an acyl-CoA oxidase gene using recombinant molecular biology techniques known in the art and/or described herein.

One method for disrupting an endogenous acyl-CoA oxidase gene is by recombinantly inserting a heterologous nucleic acid (e.g., a nucleotide sequence encoding a selectable marker such as an enzyme that restores an auxotrophic host organism to prototrophy) into the endogenous gene, thereby generating an engineered organism deficient in acyl-CoA oxidase activity. This can be done, for example, through homologous recombination in which a heterologous nucleic acid containing sequences of an endogenous acyl-CoA oxidase gene and a disrupting sequence (e.g., a knock-out gene cassette such as described herein) is introduced into a host cell or microorganism. In some embodiments, the nucleotide sequence of one or more acyl CoA oxidases (e.g., a yeast POX4, POX 5, or POX4 and POX5) can be disrupted with a URA3 nucleotide sequence encoding a selectable marker, and introduced to a host cell or microorganism, thereby generating an engineered cell or organism deficient in an acyl-CoA oxidase activity. Nucleic acids encoding an acyl-CoA oxidase can be obtained from a number of sources, including, for example, yeast cells. Genomic DNA from cell sources can be amplified using oligonucleotide primers based on the nucleotide sequence of an acyl-CoA oxidase encoding gene, including examples provided herein.

In some embodiments, of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a peroxisomal acyl-CoA oxidase in a cell is modified, e.g., a POX activity of a POX polypeptide. In particular embodiments, the acyl-CoA oxidase activity to be modified is encoded by the *POX4* and/or *POX5* genes of a species of *Candida* (e.g., ATCC 20336). In certain embodiments, the amount and/or activity of an endogenous acyl-CoA oxidase can be increased. In some embodiments, the amount and/or activity of acyl-CoA oxidases in a cell or organism containing one or more acyl-CoA oxidases can be independently modified (e.g., one or more acyl-CoA oxidases can be modified). In some embodiments, the amount and/or activity of POX4 acyl-CoA oxidase and a POX5 acyl-CoA oxidase can be altered independently of each other (e.g., increase amount and/or activity of POX4 alone, POX5 alone, increase amount and/or activity of one and decrease or eliminate the amount and/or activity of the other, and the like). Increasing the amount and/or activity of one acyl-CoA oxidase, while decreasing or eliminating the amount and/or activity of another acyl-CoA oxidase, may alter the specific activity of acyl-CoA oxidase in a cell or organism with respect to carbon chain length, while maintaining or increasing overall carbon flux through the β-oxidation pathway, in certain embodiments. Disruption of nucleotide sequences encoding one or more acyl-CoA oxidases (e.g., POX4, POX 5, or POX4 and POX5) sometimes can alter pathway efficiency, specificity and/or specific activity with respect to metabolism of carbon chains of different lengths (e.g.,

carbon chains including fatty alcohols, fatty acids, paraffins, dicarboxylic acids, aliphatic molecules of between about 1 and about 26 carbons in length).

In some embodiments of the modified cells or organisms provided herein, a β-oxidation pathway in a yeast is active and includes a genetically modified acyl-CoA oxidase. In some embodiments, an acyl-CoA oxidase is genetically modified to prevent complete oxidation of fatty acyl-CoA or diacyl-CoA substrates. Genetic modification of an acyl-CoA oxidase can increase the production yield of a desired fatty acid or fatty dicarboxylic acid product. Therefore, in some embodiments, metabolic degradation of a fatty acid of a specified chain length (e.g., the chain length of a desired or target fatty acid or fatty dicarboxylic acid product) is reduced significantly, when an acyl-CoA oxidase is genetically modified. For example, metabolic degradation of a fatty dicarboxylic acid product (e.g., dodecanedioic acid or DDDA) by beta-oxidation can be reduced significantly when an acyl-CoA oxidase is genetically modified. This can be accomplished by modifying the substrate specificity of an acyl-CoA oxidase such that the enzyme has low activity (e.g., enzymatic activity) on chain lengths equal to or less than that of a desired product.

Nucleic acids encoding a genetically modified acyl-CoA oxidase can be engineered and expressed in a suitable organism (e.g., bacteria (e.g., *E. coli*) or a yeast) to test the substrate specificity of the modified enzyme in vitro. In some embodiments, nucleic acids encoding a genetically modified acyl-CoA oxidase are engineered and expressed in a suitable yeast and the substrate specificity is tested. Yeast that express a modified acyl-CoA oxidase can be tested for production of a desired molecule, e.g., a fatty acid or fatty dicarboxylic acid product. A modified acyl-CoA oxidase can be generated in any suitable manner (e.g., random or rational mutagenesis), non-limiting examples of which are provided herein and, for example, in International patent application no. PCT/US2012/045622 (publication no. WO 2013/006733) and International patent application no. PCT/US2013/076739 (publication no. WO 2014/100504).

In some instances, a modified endogenous acyl-coA oxidase polypeptide is a modified POX4 or POX5 polypeptide from a *Candida* spp. yeast (e.g., strain ATCC 20336 or ATCC 20962). In some cases a modified POX4 polypeptide contains a modified amino acid sequence of the wild-type *Candida* strain ATCC 20336 Pox4p sequence provided herein. Sometimes the POX4 polypeptide contains an amino acid modification at one or more amino acid positions chosen from 88, 90, 96, 98, 99, 100, 102, 103, 302, 309, 310, 473, 474, 475, 476, 477, 478, 479, 480,

481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504 and 505. A modified endogenous acyl-coA oxidase polypeptide that is not a modified POX4 polypeptide can include an amino acid modification at one or more positions corresponding to one or more of the foregoing positions in the POX4 polypeptide. In some instances a modified POX5 polypeptide contains a modified amino acid sequence of the wild-type *Candida* strain ATCC 20336 Pox5p sequence provided herein. Sometimes the POX5 polypeptide contains an amino acid modification at one or more amino acid positions chosen from 81, 82, 83, 84, 85, 86, 88, 93, 94, 95, 96, 98, 102, 284, 287, 290, 291, 292, 294, 295, 436, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462 and 463. A modified endogenous acyl-coA oxidase polypeptide that is not a modified POX5 polypeptide can include an amino acid modification at one or more positions corresponding to one or more of the foregoing positions in the POX5 polypeptide.

In some embodiments, the substrate specificity of an acyl-CoA oxidase is modified such that the enzyme has low activity for aliphatic molecules with chain lengths less than C24 (i.e., 24 carbons). In some embodiments, the substrate specificity of an acyl CoA oxidase is modified such that the enzyme has very low activity with chain lengths less than 24, 22, 20, 18, 16, 14, 12, 10, 8, 6 or 4 carbons. In some embodiments, the substrate specificity of an acyl-CoA oxidase is modified such that the enzyme has very low activity with chain lengths less than 18, 16, 14, 12, 10 or 8 carbons. In some embodiments, the substrate specificity of an acyl-CoA oxidase is modified such that the enzyme has very low activity with chain lengths less than C12. In some embodiments, the substrate specificity of an acyl-CoA oxidase is modified such that the enzyme has very low activity with chain lengths less than C10 or C8. For example, in one embodiment for producing a 6-carbon dicarboxylic acid (e.g., adipic acid), a host cell or organism can be modified to decrease or eliminate acyl-CoA oxidase activities that are active on a broad range of substrate chain lengths (e.g., Pox4p acyl-CoA oxidase of *Candida viswanathii* ATCC 20336), and, to further increase productivity, can additionally be modified to express a mutant acyl-CoA oxidase activity that is more active on substrates with chain lengths of C8 and greater with little or no activity on substrates with chain lengths less than C8 (e.g., *Candida viswanathii* ATCC 20336 Pox5p(F98G); SEQ ID NO: 37).

As described herein (and in International patent application no. PCT/US2012/045622 (publication no. WO 2013/006733) and International patent application no. PCT/US2013/076739 (publication no. WO 2014/100504)), catalytic specificity of acyl-CoA oxidases (e.g., POX4,

POX5) can be altered by a variety of methods. Altering the binding and/or catalytic specificity of acyl-CoA oxidases may prove advantageous for generating novel acyl-CoA oxidases with altered chain length recognition, altered chain length catalytic activity, and/or generation of an acyl-CoA oxidase activity with a narrow or specific chain length specificity, thereby allowing further increases in pathway efficiency, specificity and/or specific activity with respect to metabolism of carbon chains of different lengths or metabolism of carbon chain distributions found in a particular chosen feedstock. In some embodiments the altered acyl-CoA oxidase sequences are identified and/or generated by; (i) screening naturally occurring variant populations; (ii) mutagenesis of endogenous sequences; (iii) introduction of heterologous sequences having a desired specificity; (iv) generation of chimeric sequences having a portion of the coding sequence from one polynucleotide source (e.g., gene, organism) and a portion of the coding sequence from another source and/or (v) intelligent design using nucleotide sequences and three dimensional structure analysis from an acyl-CoA oxidase having a desired specificity to remodel an endogenous acyl-CoA oxidase, thereby generating a novel specificity enzyme. In some embodiments, a chimeric acyl-CoA oxidase nucleic acid sequence can have polynucleotide sequence contributions from two or more sources. In some embodiments, a chimeric acyl-CoA oxidase nucleic acid sequence comprises a portion of the coding sequences from an endogenous polynucleotide and a portion of the coding sequence from a heterologous polynucleotide.

One method for generating modified acyl-CoA oxidase proteins having altered substrate specificity is through random mutagenesis. A library of genetically modified acyl-CoA oxidases can be generated using several methods known in the art (e.g., site-directed mutagenesis). Genetically modified acyl-CoA oxidase genes can then be transformed into a β-oxidation blocked strain of a suitable yeast strain (e.g., *Candida* spp. (e.g., *Candida viswanathii* or *Candida tropicalis*)). In some embodiments, a genetically modified acyl-CoA oxidase is expressed under the control of the *POX4* promoter or another strong constitutive or inducible promoter in a pox4Δ/ pox4Δ pox5Δ/pox5Δ (e.g., an organism that lacks some or all endogenous acyl-CoA oxidase activity) background. In some embodiments, the genetically modified acyl-CoA oxidase is expressed under the control of an endogenous promoter. In some embodiments, the genetically modified acyl-CoA oxidase is expressed under the control of a heterologous promoter. The transformants can be selected by growth in media containing a fatty acid or methyl-derivate fatty acid containing fatty acids with two more carbons than a fatty acid product of interest. For example, for an adipic acid product, the transformants can be

grown in caprylic acid or methyl-caprylate. For example, for a dodecanedioic acid product, the transformants can be grown in tetradecanedioic acid. The group of transformants can then be moved to a medium with a carbon source of a fatty acid of interest (for example dodecanedioic acid) in the presence of an agent that kills growing cells (e.g., Nystatin) and cells that cannot metabolize the carbon source (e.g., dodecanedioic acid in this example) can be selected. The resulting modified strains can then be further characterized for acyl-CoA oxidase activity. This method can be used to select for any modified acyl CoA oxidase (e.g., those listed and/or described in International patent application no. PCT/US2012/045622 (publication no. WO 2013/006733) and International patent application no. PCT/US2013/076739 (publication no. WO 2014/100504)). In addition, this method can be used to select for any heterologous acyl-CoA oxidase expressed in a suitable organism.

Another method for generating modified acyl-CoA oxidase proteins having altered substrate specificity is through rational mutagenesis. Structural and sequence information and experimental data can be combined to determine specific mutations for testing in an acyl-CoA oxidase for altered specificity. For example, primary sequences of acyl-CoA oxidases tested can be compared and correlated with substrate specificity. Based on such an analysis, single amino-acids, small numbers of contiguous amino acids and/or domains can be proposed for providing a desired substrate specificity. Those amino acids positions can be targeted for specific or random mutations for improve specificity.

Acyl-CoA oxidase structure also can be modeled against a known tertiary structure using modeling methods known in the art. The models can be used to propose amino acids and regions pertaining to substrate selectivity. For example, biochemical, structure and sequence data suggest that the N-terminus of acyl-CoA oxidases often, in part, determines substrate specificity. Mutations or region replacements can be introduced based on such analyses and the specificity of the new acyl-CoA oxidase tested as described before. The resulting information can be used to go back to the models to postulate new potential mutations. As for random mutagenesis, any suitable acyl-CoA oxidase can be modified to alter substrate specificity (e.g., those listed in International patent application no. PCT/US2012/045622 (publication no. WO 2013/006733) and International patent application no. PCT/US2013/076739 (publication no. WO 2014/100504)).

Examples of modified Pox5 enzymes encoded by mutated *POX5* genes from *Candida viswanathii* include Pox5p (F98G) and Pox5p(W249F) are provided herein. The design, generation and analysis of modified Pox5 enzymes encoded by mutated *POX5* genes from *Candida viswanathii* are described in the examples.

Modification of multifunctional enzyme activities

Next (e.g., second and third) steps of the β-oxidation pathway can be catalyzed by a multifunctional enzyme (referred to, for example, as Mfe2, Fox2 and HDE in fungi) having hydratase and dehydrogenase activities, or by separate hydratase and dehydrogenase enzymes. In these steps, a trans-2-enoyl-CoA can be converted to 3-ketoacyl-CoA via a (3*R*)-hydroxy intermediate. An enoyl-CoA hydratase enzyme (e.g., EC 4.2.1.17) can catalyze the addition of a hydroxyl group and a proton to the unsaturated β-carbon on a fatty-acyl CoA in a second step of the pathway to generate 3-hydroxyacyl-CoA. In a next (e.g., third) step, a 3-hydroxyacyl-CoA dehydrogenase enzyme (e.g., EC 1.1.1.35) can catalyze the formation of a 3-ketoacyl-CoA by removal of a hydrogen from the newly formed hydroxyl group created by the activity of an enoyl-CoA hydratase. Typically, fungi have one peroxisomal multifunctional enzyme (HDE, Mfe2 or Fox2), mammalian cells have two peroxisomal multifunctional enzymes (Mfe1 and Mfe2) and bacteria have a single multifunctional enzyme, Mfe1. In the yeast *Candida tropicalis*, the N-terminal portion of the MFE polypeptide typically contains two duplicate 3-hydroxyacyl-CoA dehydrogenase domains, referred to as the A and B domains, which have differing substrate specificities. The A domain can catalyze the reaction for substrates with medium-to-long carbon chains (e.g., C10 – C16). The catalytic activity of the B domain often is more active on substrates having shorter carbon chains (e.g., C4). The hydratase domain is generally located at the C-terminal region of the polypeptide. Thus, each Mfe2 monomer can contain a dehydrogenase heterodimer and a hydratase monomer.

Some multifunctional enzymes involved in the β-oxidation pathway have additional enzymatic activities, including, but not limited to, an isomerase (e.g., a Δ3,Δ2-enoyl-CoA isomerase) activity and/or an epimerase (e.g., 3-hydroxyacyl-CoA epimerase; EC 5.1.2.3) activity. These enzymes function as auxillary enzymes in the oxidation of polyunsaturated fatty acids. For example, 3-hydroxyacyl-CoA epimerase catalyzes the reversible conversion of *S*-3-hydroxyacyl-CoA to *R*-3-hydroxyacyl-CoA, which (unlike *S*-3-hydroxyacyl-CoA) is a substrate for 3-

hydroxyacyl-CoA dehydrogenase contained within Mfe2-type enzymes. Therefore, β-oxidation can proceed through the third step once the *R* isomer has been generated.

In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a multifunctional enzyme (or an enoyl-CoA hydratase or 3-hydroxyacyl-CoA dehydrogenase) in a cell is modified. In particular embodiments, the multifunctional enzyme is a peroxisomal protein. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a multifunctional enzyme (or an enoyl-CoA hydratase or 3-hydroxyacyl-CoA dehydrogenase), may be modified to decrease the amount and/or activity of a multifunctional enzyme (or an enoyl-CoA hydratase or 3-hydroxyacyl-CoA dehydrogenase), or may be modified to alternately increase and decrease the amount and/or activity of a multifunctional enzyme (or an enoyl-CoA hydratase or 3-hydroxyacyl-CoA dehydrogenase) depending, for example, on the substrate specificity, target molecule(s) being produced, cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured. In some embodiments, the amount and/or activity of one or more of the hydratase and dehydrogenase enzymes of a multifunctional enzyme may be independently modified.

In certain aspects, the amount and/or activity of one or more of the hydratase and dehydrogenase enzymes of a multifunctional enzyme in a cell is increased. Increasing the amount and/or activity of one or more of the hydratase and dehydrogenase enzymes of a multifunctional enzyme may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway involving oxidative metabolism and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, the multifunctional enzyme (or an enoyl-CoA hydratase or 3-hydroxyacyl-CoA dehydrogenase) activity is unchanged in a host or engineered cell or organism. In one embodiment, the amount and/or activity of one or more of a hydratase and/or dehydrogenase enzyme, for example, of a multifunctional enzyme, can be increased, for example, by increasing the number of copies of a nucleic acid encoding one or more of a hydratase and/or dehydrogenase enzyme (for example, of a multifunctional enzyme) (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid), by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding one or more of a hydratase

and dehydrogenase enzyme (for example, of a multifunctional enzyme), or by increasing the number of copies of a nucleic acid encoding one or more of a hydratase and/or dehydrogenase enzyme (for example, of a multifunctional enzyme) and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding one or more of a hydratase and/or dehydrogenase enzyme (for example, of a multifunctional enzyme). In some embodiments, a multifunctional enzyme (or an enoyl-CoA hydratase and/or 3-hydroxyacyl-CoA dehydrogenase) is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of one or more of the hydratase and dehydrogenase enzymes of a multifunctional enzyme protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding one or more of the hydratase and dehydrogenase enzymes of a multifunctional enzyme into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding one or more of the hydratase and dehydrogenase enzymes of a multifunctional enzyme can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

Non-limiting examples of organisms that include, or can be used as donors for, one or more of the hydratase and dehydrogenase enzymes of a multifunctional enzyme include yeast (e.g., *Candida*, *Saccharomyces*, *Yarrowia*), animals (e.g., *Homo*, *Rattus*), bacteria. In a particular embodiment, one or more of the hydratase and dehydrogenase enzymes of a multifunctional enzyme can be a *Candida* yeast protein. Additional examples of nucleotide sequences encoding multifunctional enzyme polypeptides include: *Saccharomyces cerevisiae FOX2* (Genbank accession nos. NM_001179799, M86456), *Candida tropicalis* (strain PK 233) *HDE* (Genbank accession nos. X57854, M22765), *Yarrowia lipolytica MFE2* (Genbank accession no. AF198225).

Presence, absence or amount of one or more of the hydratase and dehydrogenase enzymes of a multifunctional enzyme can be detected by any suitable method known in the art and/or described herein. For example, detection can be performed using enzyme activity assays (see, e.g., Hiltunen et al. (1992) J. *Biol. Chem.* 267(10):6646-6653). Nucleic acid sequences representing native and/or modified multifunctional enzyme sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or the amounts of the nucleic acids or encoded proteins can be assessed using quantitative expression based analysis (e.g., RT-PCR, western blot analysis,

northern blot analysis, the like and combinations thereof), where the engineered cells or organisms exhibit increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a heterologous nucleic acid encoding one or more of the hydratase and dehydrogenase enzymes (for example, of a multifunctional enzyme) can also be modified. For example, the amount of one or more of the hydratase and dehydrogenase enzymes (for example, of a multifunctional enzyme protein) expressed in a particular cellular location may be increased or decreased by including in the heterologous nucleic acid a stronger or weaker heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism.

Alternatively, decreasing the activity of one or more of the hydratase and dehydrogenase enzymes (for example, of a multifunctional enzyme) in a cell can be accomplished by modifying the amount of expression of one or more of the hydratase and dehydrogenase enzymes (for example, of a multifunctional enzyme) in the cell, for example, by replacing the wild-type promoter of an endogenous multifunctional enzyme gene (or an enoyl-CoA hydratase or 3-hydroxyacyl-CoA dehydrogenase gene) in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type multifunctional enzyme (or an enoyl-CoA hydratase or 3-hydroxyacyl-CoA dehydrogenase) such that the encoded modified or substituted protein has a reduced enzyme(s) activity.

Modification of 3-ketoacyl-CoA thiolase activity

In a final step of the β-oxidation pathway, 3-ketoacyl-CoA can undergo thiolytic cleavage to yield a fatty acyl-CoA shortened by 2 carbons and acetyl-CoA. The reaction can be catalyzed by 3-ketoacyl-CoA thiolase (e.g., EC 2.3.1.16; also referred to as β-ketothiolase, acetyl-CoA acyltransferase) and involves cleavage of the 3-ketoacyl-CoA by the thiol group of another molecule of CoA. The thiol is inserted between C-2 and C-3, which yields an acetyl CoA molecule and an acyl CoA molecule that is two carbons shorter.

In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a 3-ketoacyl-CoA thiolase in a cell is modified. For example, in some

aspects, a cell or microorganism may be modified to increase the amount and/or activity of a 3-ketoacyl-CoA thiolase, may be modified to decrease the amount and/or activity of a 3-ketoacyl-CoA thiolase, or may be modified to alternately increase and decrease the amount and/or activity of a 3-ketoacyl-CoA thiolase depending, for example, on the substrate specificity, cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a 3-ketoacyl-CoA thiolase in a cell is increased. Increasing the amount and/or activity of a 3-ketoacyl-CoA thiolase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway involving oxidative metabolism and away from other cellular metabolic pathways not involved in target molecule production.

In certain embodiments, the 3-ketoacyl-CoA thiolase activity is unchanged in a host or engineered cell or organism. In one embodiment, the amount and/or activity of a host 3-ketoacyl-CoA thiolase can be increased, for example, by increasing the number of copies of a nucleic acid encoding a 3-ketoacyl-CoA thiolase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid), by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a 3-ketoacyl-CoA thiolase, or by increasing the number of copies of a nucleic acid encoding a 3-ketoacyl-CoA thiolase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a 3-ketoacyl-CoA thiolase. In some embodiments, a 3-ketoacyl-CoA thiolase is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of a 3-ketoacyl-CoA thiolase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a 3-ketoacyl-CoA thiolase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding a 3-ketoacyl-CoA thiolase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

Non-limiting examples of organisms that include, or can be used as donors for, a 3-ketoacyl-CoA thiolase enzyme include yeast (e.g., *Candida*, *Saccharomyces*, *Debaryomyces*, *Meyerozyma*, *Lodderomyces*, *Scheffersomyces*, *Clavispora*, *Yarrowia*, *Pichia*, *Kluyveromyces*, *Eremothecium*, *Zygosaccharomyces*, *Lachancea*, *Nakaseomyces*), animals (e.g., *Homo*, *Rattus*), bacteria (e.g., *Escherichia*, *Pseudomonas*, *Bacillus*), or plants (e.g., *Arabidopsis*,

Nictotania, *Cuphea*). In a particular embodiment, a 3-ketoacyl-CoA thiolase enzyme can be a *Candida* yeast protein. Examples of nucleotide sequences encoding polypeptides having 3-ketoacyl-CoA thiolase activity include, but are not limited to: *Saccharomyces cerevisiae FOX1* (Genbank accession no. NM_001179508), *Candida tenuis* (Genbank accession no. XM_006688917), *Candida tropicalis CT-T3A* (Genbank accession no. AB025647), *Candida tropicalis CT-T3B* (Genbank accession no. AB025648), *Yarrowia lipolytica POT1* (Genbank accession no. XM_504109, X69988), *Scheffersomyces stipitis POT11* (Genbank accession no. XM_001386372), *Debaryomyces fabyri* (Genbank accession no. XM_015611011), *Arabidopsis thaliana KAT2* (Genbank accession no. NM_128874), *Lillium* cultivar Belladonna (Genbank accession no. KR998331) and *Populus davidianna KAT* (Genbank accession no. KU297273).

Presence, absence or amount of 3-ketoacyl-CoA thiolase activity can be detected by any suitable method known in the art and/or described herein. For example, detection can be performed using enzyme activity assays (see, e.g., Staack et al. (1978) *J. Biol. Chem* 253: 1827-1831; Kurihara et al. (1988) *FEBS Lett.* 229(1):215-218). Nucleic acid sequences representing native and/or modified 3-ketoacyl-CoA thiolase-encoding sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or the amounts of the nucleic acids or encoded proteins can be assessed using quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cells or organisms exhibit increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a heterologous nucleic acid encoding a 3-ketoacyl-CoA thiolase can also be modified. For example, the amount of a 3-ketoacyl-CoA thiolase protein expressed in a particular cellular location may be increased by including in the heterologous nucleic acid a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequence of a *Candida viswanathii* HDE gene promoter is provided herein as are examples of additional fatty acid-inducible promoters.

Alternatively, decreasing 3-ketoacyl-CoA thiolase activity in a cell can be accomplished by modifying the amount of 3-ketoacyl-CoA thiolase protein expression in the cell, for example, by

replacing the wild-type promoter of an endogenous 3-ketoacyl-CoA thiolase gene in an organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type 3-ketoacyl-CoA thiolase such that the encoded modified or substituted 3-ketoacyl-CoA thiolase protein has a reduced enzyme activity.

Modification of enoyl-CoA isomerase activity

Feedstocks, such as, for example, fatty acid distillates and soapstocks, can comprise unsaturated fatty acids, for example, such as oleic acid (C18:1), linoleic acid (C18:2), and linolenic acid (C18:3). In some embodiments, unsaturated fatty acids are converted to dicarboxylic acids that maintain the position and orientation of the double bonds. Unsaturated fatty acids generally are degraded through the same reactions that degrade saturated fatty acids until a Δ3-cis-acyl-CoA or Δ2-cis-acyl-CoA is formed in the process of β-oxidation. Cells can employ additional enzymes to allow the oxidation of these types of unsaturated fatty acids or diacids. In some instances, an enzyme enoyl-CoA isomerase (ECI) is required for the beta-oxidation of substrates with double bonds at odd numbered positions. In some instances, the enzyme dienoyl-CoA reductase (DCR) is required for the beta-oxidation of substrates with double bonds at even numbered positions.

Enoyl-CoA isomerase (ECI) can also be known as enoyl-CoA delta isomerase 1, dodecenoyl-CoA isomerase, 3,2 trans-enoyl-CoA isomerase, acetylene-allene isomerase, Δ3 Δ2-enoyl-CoA isomerase, dodecenoyl-CoA delta isomerase, and EC 5.3.3.8 (in human for example). Several alternatively spliced transcript variants are also known. ECI is a member of the hydratase/isomerase superfamily. ECI can be a key mitochondrial enzyme involved in beta-oxidation of unsaturated fatty acids. This enzyme can isomerize both 3-cis and 3-trans double bonds into the 2-trans form in a range of ECI enzymes from different species. ECI can catalyze the transformation of 3-cis and 3-trans-enoyl-CoA esters arising during stepwise degradation of cis-, mono-, and polyunsaturated fatty acids to the 2-trans-enoyl-CoA intermediates. ECI is present in many microorganisms and several species of yeast have at least two ECI enzymes. Nucleotide sequences (and corresponding amino acid sequences) encoding enoyl-CoA isomerase enzymes from *Candida* strain ATCC 20336 are provided herein (nucleotide SEQ ID NOS: 106 and 107 and animo acid SEQ ID NOS: 50 and 51). Examples of nucleotide sequences encoding polypeptides having enoyl-CoA isomerase activity include, but are not

limited to: *Saccharomyces cerevisiae ECI1* (Genbank accession no. AF090442) and *Candida albicans* (Genbank accession no. XM_711189).

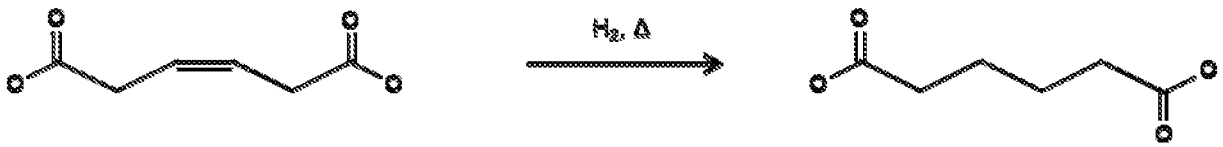
In some embodiments, ECI is utilized in generating a target fatty acid product through β-oxidation of an unsaturated fatty because of its activity and the normal position of double bonds in some feedstocks (e.g., soapstocks and fatty acid distillates). Many unsaturated fatty acids have a cis-Δ9 double bond. During the β -oxidation of an 18-carbon diacid with a cis-Δ9 double bond, the double bond is encountered when it has been chain shortened to 12 carbons. At this stage the 12-carbon molecule can have a cis-Δ3 double bond that is not a substrate for an acyl-CoA oxidase. ECI can convert the cis-Δ3 double bond to a trans-Δ2 double bond. In some instances, the product of the ECI reaction is a substrate for the second step in beta-oxidation (e.g., a substrate for enoyl-CoA hydratase), and ECI can effectively bypass acyl-CoA oxidase in a particular round of beta-oxidation. In some instances, even if a yeast strain lacks any acyl-CoA oxidase that is active on fatty acids of less than or equal to C12 (i.e., 12 carbons), an active ECI can effect the shortening of one more rounds of β-oxidation, which can produce a 10-carbon product for substrates with a cis-Δ9 double bond. Therefore, in some embodiments, the ECI gene is disrupted (e.g., knocked out or deleted) in a yeast (e.g., in a *Candida* strain) to prevent chain shortening past a desired chain-length (e.g., in this instance, 12 carbons). In some embodiments, disrupting the expression (e.g. knocking out the expression) of an ECI gene can result in an increase in the production of a fatty dicarboxylic acid containing 10 to 18 carbons. In some embodiments, disrupting the expression (e.g. knocking out the expression) of an ECI gene can result in an increase in the production of a fatty dicarboxylic acid containing 10, 12, 14, 16 or 18 carbons. In some embodiments, disrupting the expression of an enoyl-CoA isomerase can increase the production of fatty dicarboxylic acid containing 10, 12, 14, 16 or 18 carbons when using certain feedstocks (e.g., certain soapstocks or fatty acid distillates).

In some embodiments, an ECI knock out (i.e., eciΔ or Eci^{-}) strain is able to produce DDDA from from fatty acid feedstocks containing unsaturated fatty acids (e.g., oleic acid, linoleic acid, linolenic acid) even in the presence of acyl-CoA oxidase with activity on substrates of chain-length less than 12 carbons (but with little or nor activity on substrates having 12 carbons in the chain). This can be accomplished, for example, by discontinuation of β-oxidation after obtaining 3-dodecendioic acid (e.g., from oleic acid feedstock), 3,6-dodecenedioc acid (e.g., from linoleic acid feedstock) or 3,6,9-dodecenedioc acid (e.g., from linolenic acid feedstock) through an initial

three rounds of β-oxidation (due to the lack of enoyl-CoA isomerase activity), and then hydrogenation of the dodecendioic acids to yield a fully saturated DDDA.

Thus, in some embodiments, a 12-carbon dicarboxylic acid produced from fatty acid feedstocks containing unsaturated fatty acids can be hydrogenated to generate a the fully saturated DDDA product. An unsaturated diacid sometimes is produced from a feedstock containing an unsaturated fatty acid, and production of a fully saturated diacid in such situations can involve hydrogenation of the unsaturated diacid. For example, an unsaturated C6:1 diacid generated from one or more long chain unsaturated fatty acids in an Eci$^-$ yeast strain which also lacks acyl-CoA oxidase enzymes having activity on substrates of 6-carbon chain lengths (e.g., a pox4Δ yeast strain) can be converted to a fully saturated C6:0 diacid by reducing the double bond by a suitable method. Non-limiting examples of hydrogenation methods include the use of a metallic chemical catalyst, non-metallic chemical catalyst, enzymatic catalyst, the like or combination thereof.

A non-limiting example of a hydrogenation reaction is shown below. Sometimes source hydrogen is provided from molecular hydrogen (e.g., in the case of chemical catalysis) and sometimes source hydrogen is provided from enzymatic cofactors, non-limiting examples of which include NADH, NADPH, FADH2, the like or combination thereof (e.g., in the case of enzymatic catalysis).



In some embodiments, catalytic hydrogenation is carried out with a suitable metallic catalyst, non-limiting examples of which include platinum, palladium, rhodium, ruthenium, nickel, the like or combination thereof. Sometimes a catalyst is a homogenous catalyst and sometimes a catalyst is a heterogeneous catalyst. An elevated temperature and/or pressure can be employed to increase reaction rate. For example, an unsaturated diacid (e.g., cis, cis-muconic acid) can be hydrogenated and converted to adipic acid using a 10% Pt on carbon catalyst at 3400 kPa for 2.5 hours at ambient temperature (Niu et al., (2002) *Biotechnol.Prog.* 18:201-211). In some embodiments, catalytic hydrogenation can occur with nonmetallic catalysts such as frustrated Lewis pair compounds (Welch et al., (2006) Science 314:1124-1126).

In certain embodiments, enzymatic hydrogenation is conducted in vivo or in vitro with a suitable native or engineered enzyme that can catalyze a redox reaction with an unsaturated diacid or fatty acid as a substrate or a product. An enzyme can be utilized in vivo in some embodiments by increasing expression of a native enzyme or expressing a non-native enzyme capable of catalyzing a desired hydrogenation reaction in an organism that produces an unsaturated diacid precursor of a saturated diacid product. A lysate of an organism containing an enzyme capable of catalyzing a desired hydrogenation reaction, or a purified or isolated enzyme preparation, sometimes is utilized in an in vitro reaction. Non-limiting examples of a suitable native or engineered enzyme include acyl-CoA dehydrogenase (EC 1.3.1.8), trans-2-enoyl-CoA reductase (EC 1.3.1.44), stearoyl-CoA 9-desaturase (EC 1.14.19.1), the like or combination thereof. In some embodiments, a desired reaction product (e.g., a saturated diacid) is produced by an enzyme operating in a forward or a reverse direction (e.g., a forward or reverse reaction).

Modification of dienoyl-CoA reductase activity

Dienoyl CoA reductase (DCR, e.g., EC 1.3.1.34) is a peripheral enzyme that can convert trans-2, cis-4 dienoyl-CoA substrates to trans-3-enoyl-CoA products (Gurvitz A, et al., (1997) *J.Biol. Chem.* 272:22140-22147).

The trans-3-enoyl-CoA is then converted by the enzyme enoyl-CoA isomerase (ECI) to trans-2-enoyl-CoA which is then the substrate for the second enzyme (enoyl-CoA hydratase) in β-oxidation. Complete β-oxidation of fatty acids, including diacids, with double bonds at even numbered positions (e.g., linoleic acid (C18:2) and linolenic acid (C18:3)) can be achieved by including the DCR reaction in the β-oxidation pathway. Diacids are capable of being oxidized starting from either end (diterminal β-oxidation), and sometimes the enzymes used to rearrange and degrade the double bonds are the same from either direction. This is because even-numbered diacids with double bonds at even-numbered positions maintain the even-numbered position from either end (similarly with double bonds at odd-numbered positions).

The DCR reaction can be used for complete β-oxidation of fatty acids with double bonds at even numbered positions, such as linoleic acid and linolenic acid. Depending on the carbon chain length of a desired final diacid product, it may be useful to either amplify or reduce the activity of one or more DCR enzymes in the host cell or engineered organism. For diacid products that have a carbon chain length of eight or greater, it may be useful or desirable to reduce or eliminate one or all DCR enzymes in the host cell. For diacid products that have a carbon chain length of less than eight carbons, it may be useful or desirable to amplify the activity of one or more DCR enzymes in the host cell or engineered organism.

Table 1 is a table of diacid products that may be produced from unsaturated fatty acids using a yeast strain in which the *ECI* and/or *DCR* genes have been disrupted or deleted. A Dcr^- strain that does not include mutations of other genes encoding enzyme activities of the β-oxidation pathway typically can produce exclusively a C8:3 diacid. In some embodiments, a DCR polypeptide is not decreased, such as by disrupting a Dcr-encoding polynucleotide, in a strain utilized to produce a diacid product (e.g., adipic acid sebacic acid, DDDA). In certain embodiments, Dcr polypeptide production is increased (e.g., introducing additional copy numbers of an endogenous Dcr-encoding polynucleotide; introducing one or more copies of a heterologous Dcr-encoding polynucleotide) to produce adipic acid from polyunsaturated fatty acids, such as those prevalent in soybean or corn oil.

TABLE 1 **Diacids Produced in a $Pox4^+$, $Pox5^+$ Background**

Fatty Acid Carbon Source	**Eci^- Product**	**Dcr^- Product**	**Eci^-,Dcr^- Product**
Oleic acid (C18:1)	C6:1 (3-hexenedioic acid)	None	C6:1 (3-hexenedioic acid)
Linoleic acid (C18:2)	C10:2 (3,7-decenedioc acid)	C8:3 (2,4,6-octenedioic acid)	C10:3 (2,4,7-decenedioic acid)
Linolenic acid (C18:3)	C12:3 (3,6,9-dodecenedioc acid)	C8:3 (2,4,6-octenedioic acid)	C12:3 (3,6,9-dodecenedioic acid)
Eicosenoic acid (C20:1)	C6:1 (3-hexenedioic acid)	None	C6:1 (3-hexenedioic acid)
Erucic acid (C22:1)	C6:1 (3-hexenedioic acid)	None	C6:1 (3-hexenedioic acid)

In yeast such as *Candida tropicalis* and *Candida viswanathii*, there are two *DCR* homologs, often referred to as *DCR1* and *DCR2*. The yeast *Saccharomyces cerevisiae* includes one Dcr enzyme, while the yeast *Yarrowia lipolytica* includes at least three *DCR* homologs, referred to herein as "*DCR1*", "*DCR2*", and "*DCR3*". Nucleotide sequences (and corresponding amino acid sequences) encoding dienoyl-CoA reductase enzymes from *Candida* strain ATCC 20336 are provided herein (nucleotide SEQ ID NOS: 108 and 109 and amino acid SEQ ID NOS: 52 and 53). Examples of nucleotide sequences encoding polypeptides having dienoyl-CoA reductase activity include, but are not limited to: *Saccharomyces cerevisiae SPS19* (Genbank accession no. NM_001183040), *Candida tropicalis SPS19* (Genbank accession no. XM_002545237) and *Yarrowia lipolytica* (Genbank accession nos. XM_501382, XM_503937, XM_502296).

Accordingly, there are multiple possible genotypes of yeast strains (having varying combinations of wild-type and mutant acyl-CoA oxidase activity specificities and functional or non-functional ECI and/or DCR genes) for the production of fatty acids and diacids of differing carbon chain lengths and degrees of saturation/unsaturation. The fatty acid or diacid produced can depend on the carbon source in the feedstock. Table 2 provides non-limiting examples of some of the yeast strain (e.g., *Candida* spp.) genotype combinations and carbon sources for the production of adipic acid, suberic acid, sebacic acid and DDDA.

TABLE 2	**STRAIN GENOTYPE (with respect to *POX4/ECI/DCR*) AND CARBON SOURCE COMBINATIONS**		
DIACID PRODUCT	**OLEIC ACID**	**LINOLEIC ACID**	**LINOLENIC ACID**
Adipic acid (C6)	*pox4Δ/ECI/DCR* (yields saturated diacid)	*pox4Δ/ECI/DCR* (yields saturated diacid)	*pox4Δ/ECI/DCR* (yields saturated diacid*)
	POX4/Eci⁻/DCR (yields 3-hexenedioic acid*)		
Suberic acid (C8)	*pox4Δ/ECI/DCR* (yields saturated diacid)	*pox4Δ/ECI/DCR* (yields saturated diacid)	
	pox4Δ/ Eci⁻/DCR (yields 3-octenedioic acid*)	*POX4/ECI/Dcr⁻* (yields 2,4,6-octenedioic acid*)	*POX4/ECI/Dcr⁻* (yields 2,4,6-octenedioic acid*)

TABLE 2	STRAIN GENOTYPE (with respect to *POX4/ECI/DCR*) AND CARBON SOURCE COMBINATIONS		
DIACID PRODUCT	**OLEIC ACID**	**LINOLEIC ACID**	**LINOLENIC ACID**
		POX4/Eci⁻/DCR (yields 3,6-octenedioic acid*)	
Sebacic acid (C10)		*POX4/Eci⁻/DCR* or *pox4Δ/ Eci⁻/DCR* (yields 3,7-decenedioic acid*)	
		POX4/Eci⁻/ Dcr⁻ (yields 2,4,7-decenedioic acid*)	
Dodecanedioic acid (C12)			*POX4/Eci⁻/DCR* or *pox4Δ/ Eci⁻/DCR* (yields 3,6,9-dodecenedioic acid*)

* can be hydrogenated to saturate

Modification of dienoyl-CoA isomerase activity

Dienoyl-CoA isomerase (DCI, e.g., EC 5.3.3, Δ3,5,Δ2,4-dienoyl-CoA isomerase, Δ3,5,Δ2,4-dienoyl-coenzyme A isomerase) is a peripheral β-oxidation enzyme that catalyzes the isomerization of a Δ3,5-dienoyl-CoA to a Δ2,4-dienoyl-CoA. This reaction is part of a minor β-oxidation pathway that occurs when the 3,2-enoyl-CoA isomerase (ECI) converts a Δ2,5-dienoyl-CoA to a Δ3,5-dienoyl-CoA. In order to fully oxidize this product DCI converts the Δ3,5-dienoyl-CoA to a Δ2,4-dienoyl-CoA, the latter of which is a substrate for the 2,4-dienoyl-CoA reductase (DCR). The product of the DCR reaction is a 3-enoyl-CoA, which is a substrate for ECI that converts it to a 2-enoyl-CoA that can be fully oxidized through β-oxidation.

In some embodiments, the amount and/or activity of a Dci enzyne in a cell or organism is decreased or increased, depending upon the chain-length of a desired target diacid product to be generated through β-oxidation of a fatty acid. For example, for adipic acid production, a DCI

activity can be increased to improve productivity of unsaturated fatty acids in a host cell or organism (e.g., DCI activity can be increased by introducing one or more copies of a polynucleotide encoding a polypeptide having DCI activity into the cell or organism (e.g., introducing one or more copies of an endogenous or exogenous polynucleotide)). In some embodiments, for production of C8 and longer diacids, the amount and/or activity of a DCI enzyme in a cell or organism can be decreased (e.g., by introducing a disruption, deletion or knockout of a polynucleotide that encodes a polypeptide having DCI activity, or replacing a promoter of a DCI gene with a weaker promoter (for example, introducing a nucleic acid containing a weak promoter operably linked to a polynucleotide that encodes a polypeptide having DCI activity into a cell in which an endogenous DCI gene has been disrupted or deleted). An example of a nucleotide sequence encoding a *Saccharomyces cerevisiae* DCI enzyme is Genbank accession no. NM_001183599.

Modification of β-oxidation-associated activities

There are also cellular compositions and activities that are closely associated with β-oxidation and support the core degradative functioning of the pathway. These include peroxisomal- and mitochondrial-related compositions and activities. For example, as described herein, such compositions and activities include, but are not limited to, compositions and activities involved in: generating acyl-CoA through thioesterification of fatty acids, movement of fatty acids and/or acyl-CoA into cellular sites of β-oxidation (e.g., peroxisomes), regulation of β-oxidation activities, synthesis of compositions involved in β-oxidation, and maintenance/amount of sites of β-oxidation (e.g., peroxisomes). Included in the cells, organisms, systems and methods provided herein are embodiments in which one or more of these β-oxidation-associated compositions and/or activities are modified. In some embodiments, a β-oxidation-associated composition or activity is modified to enhance β-oxidation activity.

Modification of peroxisomal transport activity

In order for fatty acids to undergo peroxisomal β-oxidation, they must first move into the peroxisomes. Generally, medium-chain free fatty acids present in the cytosol can traverse the peroxisomal membrane and become activated once in the peroxisome by a persoxisomal acyl-CoA synthetase to then be processed as an acyl-CoA in β-oxidation. Long-chain fatty acids that have entered a cell from the extracellular medium tend to quickly be activated by acyl-CoA

synthetases located at or near the cell membrane or in the cytosol. These acyl-CoA esters typically are not able to traverse the peroxisomal membrane and thus require a peroxisomal transporter in order to move into peroxisomes. Peroxisomal transporter proteins can be a target for modifying entry of fatty acids into peroxisomes. Free fatty acids internalized into cells, or generated within cells (e.g., by oxidation of internalized alkanes), can directly enter into and be processed in the ω-oxidation pathway without prior activation to acyl-CoA.

In some embodiments of the microorganisms, compositions and methods provided herein, the amount and/or activity of a peroxisomal transporter protein in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount of a peroxisomal transporter protein and/or a peroxisomal transporter protein activity, may be modified to decrease the amount of a peroxisomal transporter protein and/or a peroxisomal transporter protein activity, or may be modified to alternately increase and decrease the amount of a peroxisomal transporter protein and/or a peroxisomal transporter protein activity depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of a peroxisomal transporter protein in a cell is decreased. Reducing or eliminating the amount and/or activity of a peroxisomal transporter protein may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway (e.g., in peroxisomes) and away from other cellular metabolic pathways involving activated fatty acids (acyl-CoA). For example, in embodiments of the production systems in which a target molecule, or intermediate/precursor of a target molecule, is a dicarboxylic acid, it may be optimal to decrease or eliminate fatty acid entry into peroxisomes through modes other than as a dicarboxylic acid which moves freely into peroxisomes after formation though an initial ω-oxidation of a free fatty acid. Certain aspects of the cells, microorganisms, compositions and methods provided herein include one or more modifications to reduce or eliminate transport of acyl-CoA into peroxisomes. One approach to reducing or eliminating such transport is to decrease the amount and/or activity of a peroxisomal transporter protein. For example, one or more endogenous genes encoding a peroxisomal transporter protein (e.g., yeast *PXA1* and/or *PXA2*) can be disrupted or deleted in a host cell or microorganism to reduce or eliminate the amount of and/or activity of a peroxisomal transporter protein in the host relative to a cell or microorganism in which the gene(s) have not been modified.

Methods for decreasing the amount and/or activity of a peroxisomal transporter protein in a cell include, but are not limited to, modifying the amount of peroxisomal transporter protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous a peroxisomal transporter protein gene in an organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type peroxisomal transporter protein such that the encoded modified or substituted peroxisomal transporter protein has a reduced activity. For example, expression of a host peroxisomal transporter protein activity can be decreased or eliminated by disruption (e.g., knockout, insertion mutagenesis, the like and combinations thereof) of a host gene encoding the protein, or by decreasing the activity of the promoter (e.g., through addition of repressor sequences to the promoter or 5'UTR or replacing the promoter) that controls transcription of a peroxisomal transporter protein gene using recombinant molecular biology techniques known in the art and/or described herein. In one embodiment, a diploid yeast, such as, for example, a *Candida* yeast, when used as a host microorganism can be subjected to genetic modification in which one of the two alleles of a peroxisomal transporter protein gene is disrupted or deleted. In so doing, a single allele of the gene remains for a reduced amount of peroxisomal transporter protein expression in the microorganism and a reduced amount of the protein in the cell.

One method for disrupting an endogenous peroxisomal transporter protein gene is by recombinantly inserting a heterologous nucleic acid (e.g., a nucleotide sequence encoding a selectable marker such as an enzyme that restores an auxotrophic host organism to prototrophy) into the endogenous gene, thereby generating an engineered organism deficient in a peroxisomal transporter protein activity. This can be done, for example, through homologous recombination in which a heterologous nucleic acid containing sequences of an endogenous peroxisomal transporter protein gene and a disrupting sequence (e.g., a knock-out gene cassette such as described herein) is introduced into a host cell or microorganism. Nucleic acids encoding a peroxisomal transporter protein can be obtained from a number of sources, including, for example, yeast cells. Genomic DNA from cell sources can be amplified using oligonucleotide primers based on the nucleotide sequence of a peroxisomal transporter protein encoding gene, including examples provided herein. Nucleotide sequences encoding the subunits of (and the amino acid sequences of) a *Candida viswanathii* peroxisomal transporter protein, Pxa1 and Pxa2, are provided herein (nucleotide SEQ ID NOS: 94 and 95 and amino acid SEQ ID NOS: 40 and 41). Additional non-limiting examples of nucleic acids encoding a

peroxisomal transporter protein include *Saccharomyces cerevisiae PXA1* (Genbank accession numbers NM_001183961 and U17065), *Saccharomyces cerevisiae PXA2* (Genbank accession numbers NM_001179754 and U93584), *Schizosaccharomyces pombe PXA1* (Genbank accession number NM_001018794), *Candida albicans PXA1* (Genbank accession number XM_713564), *Yarrowia lypolytica PXA1* (Genbank accession number XM_499814), *Yarrowia lypolytica PXA2* (Genbank accession number XM_502396), *Candida orthopsilosis PXA1* (Genbank accession number XM_003865834), *Aspergillus nomius PXA1* (Genbank accession number XM_015554863), *Clavispora lusitaniae PXA1* (Genbank accession number JQ710938), *Aspergillus niger PXA1* (Genbank accession number XM_001388761) and *Arabidopsis thalina ABCD1* (Genbank accession number NM_001204043).

In other embodiments, the amount and/or activity of a host peroxisomal transporter protein can be increased, for example, by increasing the number of copies of a gene encoding a peroxisomal transporter protein (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the gene), by increasing the activity of a promoter that regulates transcription of a gene encoding a peroxisomal transporter protein, or by increasing the number of copies of a gene encoding a peroxisomal transporter protein and increasing the activity of a promoter that regulates transcription of a gene encoding a peroxisomal transporter protein. In some embodiments, a peroxisomal transporter protein is endogenous to the host cell or microorganism. In particular embodiments, the amount and/or activity of a host peroxisomal transporter protein is increased.

The presence, absence or amount of peroxisomal transporter protein can be detected by any suitable method known in the art. Non-limiting examples of suitable detection methods include nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cells or organisms exhibit increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

Modification of peroxisome biogenesis activity

Peroxisomes can be found in eukaryotic cells and are a cellular location for β-oxidation (i.e., the site for β-oxidation in fungi and plant cells and one of two sites (the other being mitochondria) for β-oxidation of fatty acids in animal cells). Consistent with this function, peroxisome

proliferation may occur in cells exposed to fatty acids as a sole source of carbon, and peroxisome degradation may occur in cells in the presence of glucose. Thus, the number of and volume of peroxisomes in cells can be regulated. Although most of the more than 30 peroxisomal membrane proteins, referred to as peroxins or Pex proteins, play a role in importing proteins into the peroxisomal matrix from the cytosol (e.g., Pex5, Pex7, Pex13, Pex14 Pex16, Pex17), some (e.g. Pex 11, Pex 25, Pex 27, Pex 34) are involved in peroxisome proliferation.

In some embodiments of the cells, organisms, compositions and methods provided herein, the amount and/or activity of a Pex protein in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount and/or activity of a Pex protein, to decrease the amount and/or activity of a Pex protein, or to alter the pattern of expression of a Pex protein. In particular embodiments, the Pex protein is one that is involved in peroxisome proliferation, e.g., Pex11.

In certain aspects, the amount and/or activity of a Pex protein in a cell or organism is increased. Increasing the amount and/or activity of a Pex protein may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed toward a particular target product pathway involving oxidative metabolism and away from other cellular metabolic pathways not involved in target molecule production. In some embodiments, the amount and/or activity of a Pex protein involved in peroxisome proliferation is increased in a cell or organism to provide for increased numbers of peroxisomes as sites for β-oxidation. In particular embodiments, the Pex protein is Pex11.

In certain embodiments, the Pex protein activity is unchanged in a host or engineered cell or organism. In one embodiment, the amount and/or activity of a host Pex protein can be increased, for example, by increasing the number of copies of a nucleic acid encoding a Pex protein (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid), by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a Pex protein, or by increasing the number of copies of a nucleic acid encoding a Pex protein and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding a Pex protein. In some embodiments, a Pex protein is endogenous to the host cell or microorganism. In one aspect of the cell or microbial systems and methods provided herein, the amount of a Pex protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding a Pex protein into a cell or microorganism. For example, introduction of heterologous

nucleic acid encoding a Pex protein can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

Non-limiting examples of organisms that include, or can be used as donors for, a Pex protein include yeast (e.g., *Candida*, *Saccharomyces*, *Debaryomyces*, *Meyerozyma*, *Lodderomyces*, *Scheffersomyces*, *Clavispora*, *Yarrowia*, *Pichia*, *Kluyveromyces*, *Eremothecium*, *Zygosaccharomyces*, *Lachancea*, *Nakaseomyces*), animals (e.g., *Homo*, *Rattus*), or plants (e.g., *Arabidopsis*, *Nictotania*, *Cuphea*). In a particular embodiment, a Pex protein can be a *Candida* yeast protein. An example of a *Candida viswanathii* nucleotide sequence (and corresponding amino acid sequence) encoding a Pex11 protein is provided herein (nucleotide SEQ ID NO: 89 and amino acid SEQ ID NO: 33). Additional examples of nucleotide sequences encoding polypeptides having Pex protein activity include, but are not limited to: *Saccharomyces cerevisiae PEX11* (Genbank accession no. NM_001183401), *Candida albicans* (Genbank accession no. XM_707009), *Candida orthopsilosiis PEX11* (Genbank accession no. XM_003870517), *Yarrowia lipolytica PEX11* (Genbank accession nos. XM_503276, XM_501447, XM_501425), *Arabidopsis thaliana PEX11A* (Genbank accession no. NM_103668), *Neurospora crassa PEX11* (Genbank accession no. XM_011396615), *Pichia angusta PEX11* (Genbank accession no. DQ645582).

Presence, absence or amount of Pex protein or nucleic acids encoding the protein can be detected by any suitable method known in the art and/or described herein. For example, detection can be performed using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cells or organisms exhibit increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

The promoter used for regulating transcription of a heterologous nucleic acid encoding a Pex protein can also be modified. For example, the amount of a Pex protein expressed in a particular cellular location may be increased by including in the heterologous nucleic acid a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequence

of a *Candida viswanathii* HDE gene promoter is provided herein as are examples of additional fatty acid-inducible promoters.

Modification of acetyl-CoA processing in peroxisomes

Included in the cell-based and microbial production platform systems and components thereof provided herein are embodiments in which the processing of acetyl-CoA in organelles of a cell or microorganism is modified. In some embodiments, acetyl group carbons of organelle-generated acetyl-CoA are directed toward conversion to acetate. In particular embodiments, acetyl group carbons are directed toward conversion to acetate and away from the carnitine-carrier transport system. Accordingly, provided herein are cells, microorganisms, compositions and methods in which cellular carbon flux has been modified through the altered (e.g., increased or decreased) de novo generation of cellular acetate. In particular embodiments, cellular carbon flux has been modified to increase the production of acetate in a cell and/or a particular cellular location. In certain aspects, cells or microorganisms are modified to increase the production of acetate in peroxisomes.

Modification of acetyl-CoA hydrolase activity

In some embodiments of the cells, microorganisms, compositions and methods provided herein, the amount and/or activity of acetyl-CoA hydrolase in a cell is modified. Acetyl-CoA hydrolase (e.g. EC 3.1.2.1) is an enzyme that catalyzes the hydrolysis of acetyl-CoA to form acetate and CoA. For example, in some aspects, a cell or microorganism may be modified to increase acetyl-CoA hydrolase and/or acetyl-CoA hydrolase activity, may be modified to decrease acetyl-CoA hydrolase and/or acetyl-CoA hydrolase activity, or may be modified to alternately increase and decrease acetyl-CoA hydrolase and/or acetyl-CoA hydrolase activity depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In some aspects, the amount and/or activity of acetyl-CoA hydrolase in a cell and/or a particular cellular location is increased. For example, the amount and/or activity of acetyl-CoA hydrolase in peroxisomes of a cell can be increased. In some embodiments, the pattern of expression of acetyl-CoA hydrolase can be modified such that the enzyme is produced in a cellular location

where it is not produced in an unmodified cell and/or is no longer produced in a cellular location where it is produced in an unmodified cell.

In one aspect, the amount and/or activity of a host acetyl-CoA hydrolase can be increased, for example, by increasing the number of copies of a nucleic acid encoding an acetyl-CoA hydrolase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the nucleic acid), by increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an acetyl-CoA hydrolase, or by increasing the number of copies of a nucleic acid encoding an acetyl-CoA hydrolase and increasing the activity of a promoter that regulates transcription of a nucleic acid encoding an acetyl-CoA hydrolase. In some embodiments, an acetyl-CoA hydrolase is endogenous to the host cell or microorganism. In one aspect of the cell-based and microbial systems and methods provided herein, the amount of acetyl-CoA hydrolase protein expressed in a cell can be increased by introducing heterologous nucleic acid encoding acetyl-CoA hydrolase into a cell or microorganism. For example, introduction of heterologous nucleic acid encoding acetyl-CoA hydrolase can result in increased copy number of such nucleic acids and/or provide for modification of the cellular location in which the protein is expressed.

Acetyl-CoA hydrolase is typically localized to the mitochondrial compartment in eukaryotes. In one embodiment of the cells, microorganisms and methods provided herein, host cells are modified to express engineered acetyl-CoA hydrolase proteins that include targeting signals that direct the enzyme to peroxisomes, thereby introducing, or increasing the amount of, peroxisomal acetyl-CoA hydrolase in the cells. In particular embodiments, the engineered acetyl-CoA hydrolase protein has also been modified to exclude amino acids of a mitochondrial targeting sequence. One such modified acetyl-CoA hydrolase protein is a yeast Ach1p $^{\Delta mts+pts}$ which includes a heterologous peroxisomal targeting signal (pts) and excludes a mitochondrial targeting sequence (mts). In order to express engineered acetyl-CoA hydrolase in a targeted location, such as the peroxisomes, heterologous nucleic acid encoding the modified enzyme can be introduced into host cells. Acetate generated through the action of peroxisomal acetyl-CoA hydrolase can freely diffuse out of the peroxisome into the cytosol where it can be converted back to acetyl-CoA by the enzyme acetyl-CoA synthetase (e.g., EC 6.2.1.1), thereby increasing the generation and amount of cytosolic acetyl-CoA. In a particular embodiment, the acetyl-CoA hydrolase enzyme can be a *Candida* yeast protein. An example of a *Candida viswanathii* nucleotide sequence (*ACH1*$^{\Delta mts+pts}$; SEQ ID NO: 73) encoding a modified acetyl-CoA hydrolase lacking a mitochondrial-targeting sequence and including a peroxisomal-targeting

sequence (Ach1p$^{\Delta mts+pts}$; SEQ ID NO: 16) is provided herein. Additional examples of nucleotide sequences encoding acetyl-CoA hydrolase proteins include but are not limited to: *Saccharomyces cerevisiae ACH1* (Genbank accession numbers M31036, NM_001178255), *Candida tropicalis ACH1* (Genbank accession number XM_002550976), *Candida orthopsilosis ACH1* (Genbank accession number XM_003870486), *Candida albicans ACH1* (Genbank accession number XM_709496), *Aspergillus flavus ACH1* (Genbank accession number XM_002372714), *Neurospora crassa ACU8* (Genbank accession number XM_953261), *Cyberlindnera jadinii ACH1* (Genbank accession number AB641818), *Debaryomyces fabryi ACH1* (Genbank accession number XM_015614474), *Schizosaccharomyces octosporus ACH1* (Genbank accession numberXM_013163018), *Schizosaccharomyces japonicus ACH1* (Genbank accession numberXM_002173925), *Penicillium digitatum ACH1* (Genbank accession number XM_014683672), *Penicillium marneffei ACH1* (Genbank accession number XM_002152968) and *Talaromyces stipitatis ACH1* (Genbank accession number XM_002487448). Any of these, and other such acetyl-CoA hydrolase-encoding nucleic acids, can be analyzed for the presence of 5' ORF nucleotides encoding possible mitochondrial-targeting sequences of amino acids and modified to eliminate such sequences. Nucleotides encoding a peroxisomal-targeting sequence (e.g, a PTS1 sequence such as AKL or SKL) can also be added to the 3' terminus of the coding sequences of the nucleic acids.

The promoter used for regulating transcription of a heterologous nucleic acid encoding an acetyl-CoA hydrolase can also be modified. For example, the amount of an acetyl-CoA hydrolase protein expressed in a particular cellular location may be increased by including in the heterologous nucleic acid a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a fatty acid-inducible promoter that can provide for increased acetyl-CoA hydrolase expression, particularly when exposed to fatty acids as a carbon source. Such promoter elements include those that regulate expression of peroxisomal proteins and/or β-oxidation enzymes in microbes, e.g., a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequence of a *Candida viswanathii* HDE gene promoter is provided herein as are examples of additional fatty acid-inducible promoters.

The acetyl-CoA hydrolase activities of host and modified cells and microorganisms can be evaluated and monitored using methods known in the art. Examples of acetyl-CoA hydrolase activity assays include colorimetric assays (see, e.g., Connerton et al. (1992) *J. Gen. Microbiol.*

138:1797-1800; Robinson et al (1976) *Biochem. Biophys. Res. Commun.* 21:959-965) and radioactivity-based and acetylation inhibition assays (see, e.g., U.S. Patent no. 5,487,990 to Smith et al.). Nucleic acid sequences representing native and/or modified acetyl-CoA hydrolase-encoding sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or the amounts of the nucleic acids or encoded proteins can be assessed using quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cells or organisms exhibit increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

Modification of peroxisomal carnitine acetyltransferase activity

Also provided herein are cells, microorganisms, compositions and methods in which the amount and/or activity of peroxisomal carnitine acetyltransferase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount of peroxisomal carnitine acetyltransferase and/or peroxisomal carnitine acetyltransferase activity, may be modified to decrease the amount of peroxisomal carnitine acetyltransferase and/or peroxisomal carnitine acetyltransferase activity, or may be modified to alternately increase and decrease the amount of peroxisomal carnitine acetyltransferase and/or peroxisomal carnitine acetyltransferase activity depending, for example, on the conditions in which the modified cell or microorganism is cultured.

In some aspects, the amount and/or activity of peroxisomal carnitine acetyltransferase in a cell is decreased. Reducing or eliminating the amount and/or activity of peroxisomal carnitine acetyltransferase may be particularly beneficial in embodiments in which the flux of peroxisomal acetyl moiety carbons is directed toward generation of acetate within peroxisomes. In these embodiments, reducing or eliminating the amount and/or activity of peroxisomal carnitine acetyltransferase decreases the amount of peroxisomal acetyl group carbon atoms that are converted to acetyl-carnitine and provides increased peroxisomal acetyl-CoA availability for generation of peroxisomal acetate. Methods for decreasing peroxisomal carnitine acetyltransferase activity in a cell include, but are not limited to, modifying the amount of peroxisomal carnitine acetyltransferase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous peroxisomal carnitine acetyltransferase gene in an organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene,

and/or replacing or modifying a gene encoding a wild-type peroxisomal carnitine acetyltransferase such that the encoded modified or substituted peroxisomal carnitine acetyltransferase protein has a reduced enzyme activity. For example, expression of a host peroxisomal carnitine acetyltransferase activity can be decreased by disruption (e.g., knockout, insertion mutagenesis, the like and combinations thereof) of a host gene encoding the protein, or by decreasing the activity of the promoter (e.g., through addition of repressor sequences to the promoter or 5'UTR or replacing the promoter) that controls transcription of a peroxisomal carnitine acetyltransferase gene using recombinant molecular biology techniques known in the art and described herein. In one embodiment, a diploid yeast, such as, for example, a *Candida* yeast, when used as a host microorganism can be subjected to genetic modification in which one of the two alleles of a peroxisomal carnitine acetyltransferase gene is disrupted or deleted. In so doing, a single allele of the gene remains for a reduced amount of peroxisomal carnitine acetyltransferase expression in the microorganism and a reduced amount of the protein in the cell.

One method for disrupting an endogenous peroxisomal carnitine acetyltransferase gene is by recombinantly inserting a heterologous nucleic acid (e.g., a nucleotide sequence encoding a selectable marker such as an enzyme that restores an auxotrophic host organism to prototrophy) into the endogenous gene, thereby generating an engineered organism deficient in peroxisomal carnitine acetyltransferase activity. This can be done, for example, through homologous recombination in which a heterologous nucleic acid containing sequences of the endogenous peroxisomal carnitine acetyltransferase gene and a disrupting sequence (e.g., a knock-out gene cassette such as described herein) is introduced into a host cell or microorganism. Nucleic acids encoding a peroxisomal carnitine acetyltransferase can be obtained from a number of sources, including, for example, yeast cells. Genomic DNA from cell sources can be amplified using oligonucleotide primers based on the nucleotide sequence of a peroxisomal carnitine acetyltransferase encoding gene, including examples provided herein.

In some instances, a host gene, e.g., certain yeast genes, encoding a peroxisomal carnitine acetyltransferase also encodes a mitochondrial carnitine acetyltransferase. In these organisms, a peroxisomal carnitine acetyltransferase is encoded by a gene that generates a protein containing mitochondrial and peroxisomal targeting sequences. Therefore, in such an instance, disruption or deletion of a gene encoding a carnitine acetyltransferase that is localized to peroxisomes will result in reducing or eliminating mitochondrial, as well as peroxisomal,

carnitine acetyltransferase protein expression. In order to reduce or eliminate peroxisomal carnitine acetyltransferase expression in such cells without eliminating mitochondrial carnitine acetyltransferase expression, a heterologous nucleic acid encoding a mitochondria-targeted carnitine acetyltransferase can be introduced into the cell after disruption of the endogenous gene. For example, a mitochondrial-targeted enzyme that would not be expressed in peroxisomes can be produced in a cell or microorganism by introducing a heterologous nucleic acid that encodes a carnitine acetyltransferase that includes a mitochondrial targeting sequence of amino acids but lacks a peroxisomal targeting sequence of amino acids. An example of such a modified *Candida viswanathii* nucleic acid sequence (*CAT2*$^{\Delta pts}$; SEQ ID NO: 62), and the amino acid sequence encoded thereby (Cat2p$^{\Delta pts}$; SEQ ID NO: 5), are provided herein.

In another embodiment provided herein, a heterologous nucleic acid encoding a peroxisomal carnitine acetyltransferase that has a reduced carnitine acetyltransferase activity relative to the activity of the enzyme encoded by a host cell's or microorganism's endogenous peroxisomal carnitine acetyltransferase gene can be introduced into a host cell in which the endogenous peroxisomal carnitine acetyltransferase gene(s) has been disrupted or deleted. The heterologous nucleic acid encoding the less active carnitine acetyltransferase can be modified to include nucleotides encoding a peroxisomal targeting sequence for expression of the enzyme specifically in peroxisomes and not in other areas, such as mitochondria. For example, in one aspect, a heterologous nucleic acid encoding a *Candida viswanathii* cytoplasmic carnitine acetyltransferase (*YAT1*) with added nucleotides encoding a peroxisomal targeting sequence (e.g, a PTS1 sequence such as AKL or SKL, or slight variant thereof (PKL, PKF)) can be introduced into a host cell or microorganism (e.g., a *Candida viswanathii* cell). A nucleotide sequence encoding (and the amino acid sequence of) a *Candida viswanathii* YAT1^{+pts}p are provided herein (nucleotide SEQ ID NO: 70 and amino acid SEQ ID NO: 13). Additional non-limiting examples of nucleic acids encoding cytoplasmic carnitine acetyltransferase include *Saccharomyces cerevisiae YAT1* (Genbank accession number X74553), *Aspergillus nidulans FacC* (Genbank accession number AF023156), *Cyberlindnera jadinii YAT1* (Genbank accession number AB641829), *Candida dubliniensis YAT1* (Genbank accession number XM_002416790) and *Candida albicans* (Genbank accession number AF525683). A sequence of nucleotides encoding a peroxisomal targeting sequence can be added to the 3' end of the coding sequence of any such nucleic acid using methods as described herein.

The peroxisomal carnitine acetyltransferase activities of host and modified cells and microorganisms can be evaluated and monitored using methods known in the art. For example, methods of isolating peroxisomal and mitochondrial components of yeast cells and of extracting carnitine acetyltransferase from subcellular fractions have been described by Ueda et al. [(1982) *Eur. J. Biochem.*124:205-210] and Kozulic et al. [(1987) *Eur. J. Biochem.*168:245-250]. Methods of measuring the enzymatic activity of carnitine acetyltransferase are also known in the art, see, e.g., Fritz and Schultz (1965) *J. Biol. Chem.* 240:2188-2192; Chase (1969) *Meth. Enzymol.*13:387-393. Nucleic acid sequences representing native and/or modified peroxisomal carnitine acetyltransferase-encoding sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or the amounts of the nucleic acids or encoded proteins can be assessed using quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cells or organisms exhibit increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

In other embodiments, the amount and/or activity of a host peroxisomal carnitine acetyltransferase can be increased, for example, by increasing the number of copies of a gene encoding a peroxisomal carnitine acetyltransferase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the gene), by increasing the activity of a promoter that regulates transcription of a gene encoding a peroxisomal carnitine acetyltransferase, or by increasing the number of copies of a gene encoding a peroxisomal carnitine acetyltransferase and increasing the activity of a promoter that regulates transcription of a gene encoding a peroxisomal carnitine acetyltransferase. In some embodiments, a peroxisomal carnitine acetyltransferase is endogenous to the host cell or microorganism.

Modification of acetyl-CoA synthetase

Acetyl-CoA synthetase (EC 6.2.1.1) is an enzyme that can catalyze the ligation of acetate and coenzyme A to produce acetyl-CoA. In many cells and organisms, the enzyme is encoded by one or more *ACS* genes. For example, in some yeast, acetyl-CoA synthetase is encoded by two genes, *ACS1* and *ACS2*, which may be differentially expressed in response to growth on differing carbon sources. In some cells, the proteins encoded by the two genes may also be differentially distributed within the nucleus, mitochondria, peroxisomes and cytoplasm of cells.

Acs1p and Acs2p are expressed in the cytoplasm, but, in some cells, only Acs2p is present when cells are grown in glucose. *ACS1* expression may be repressed in some cells grown in glucose and derepressed when glucose is limited and/or in the presence of non-fermentable carbon sources, e.g., acetate and ethanol. Generally, *ACS2* is constitutively expressed in yeast cells. In some instances, the affinity of Acs1p for acetate may be higher, e.g, about 30-fold higher, than that of Acs2p.

Provided herein are cells, microorganisms, compositions and methods in which the amount and/or activity of acetyl-CoA synthetase (also referred to as ACS or acetate-CoA ligase and used interchangeably herein) in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount of acetyl-CoA synthetase and/or acetyl-CoA synthetase activity, may be modified to decrease the amount of acetyl-CoA synthetase and/or acetyl-CoA synthetase activity, or may be modified to alternately increase and decrease the amount of acetyl-CoA synthetase and/or acetyl-CoA synthetase activity depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

For example, in embodiments in which modification of cellular acetate generation yields increased amounts of cytosolic acetate, the amount and/or activity of cytosolic acetyl-CoA synthetase can also be increased to provide for increased conversion of acetate to acetyl-CoA. Heterologous nucleic acid encoding Acs1p and/or Acs2p can be introduced into a host cell to increase the amount of cytosolic acetyl-CoA synthetase. For example, the amount and/or activity of a host cytosolic acetyl-CoA synthetase can be increased by increasing the number of copies of a gene encoding a cytosolic acetyl-CoA synthetase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the gene), by increasing the activity of a promoter that regulates transcription of a gene encoding a cytosolic acetyl-CoA synthetase, or by increasing the number of copies of a gene encoding a cytosolic acetyl-CoA synthetase and increasing the activity of a promoter that regulates transcription of a gene encoding a cytosolic acetyl-CoA synthetase. In some embodiments, a cytosolic acetyl-CoA synthetase is endogenous to the host cell or microorganism. Additionally, a heterologous promoter can be used to regulate expression of a recombinant acetyl-CoA synthetase-encoding nucleic acid. An example of one such heterologous promoter is a fatty acid-inducible promoter that can provide for increased acetyl-CoA synthetase expression, particularly when exposed to fatty acids as a carbon source. Such promoter elements include those that regulate expression of peroxisomal proteins and/or β-

oxidation enzymes in microbes, e.g., a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter.

In other aspects, the amount and/or activity of acetyl-CoA synthetase in a cell is decreased. Methods for decreasing acetyl-CoA synthetase activity in a cell include, but are not limited to, modifying the amount of acetyl-CoA synthetase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous acetyl-CoA synthetase gene in an organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type acetyl-CoA synthetase such that the encoded modified or substituted acetyl-CoA synthetase protein has a reduced enzyme activity. For example, in some instances, it may be desirable to decrease the amount and/or activity of a peroxisomal protein having acetyl-CoA synthetase activity.

In a particular embodiment, the acetyl-CoA synthetase enzyme can be a *Candida* yeast protein. Examples of *Candida viswanathii* nucleotide sequences (and corresponding amino acid sequences) encoding acetyl-CoA synthetase are provided herein (nucleotide SEQ ID NOS: 76 and 77 and amino acid SEQ ID NOS: 20 and 21). Additional examples of nucleotide sequences encoding acetyl-CoA synthetase proteins include, but are not limited to: *Saccharomyces cerevisiae ACS1* (Genbank accession number NM_001178197), *Saccharomyces cerevisiae ACS2* (Genbank accession number NM_001182040), *Candida tropicalis ACS1* (Genbank accession number XM_002547679), *Candida albicans ACS2* (Genbank accession number AF535132), *Cyberlindnera jadinii ACS1* (Genbank accession number AB641819), *Cyberlindnera jadinii ACS2* (Genbank accession number AB641820), *Kluyveromyces lactis ACS2* (Genbank accession number AF134491).

The acetyl-CoA synthetase activities of host and modified cells and microorganisms can be evaluated and monitored using methods known in the art. Examples of acetyl-CoA synthetase activity assays include a continuous coupled enzymatic assay (see, e.g., Castano-Cerezo et al. (2012) *Bio-protocol* 2(17) and Frenkel and Kitchens (1977) *J. Biol. Chem.* 252(2): 504-507). Nucleic acid sequences representing native and/or modified acetyl-CoA synthetase-encoding sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or the amounts of the nucleic acids or encoded proteins can be assessed using quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations

thereof), where the engineered cells or organisms exhibit increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

Modification of citrate processing

Carbon atoms of acetyl groups in mitochondrial acetyl-CoA can also be captured from intermediates of the TCA cycle such as, for example, citrate molecules generated in the first step of the cycle through the citrate synthase-catalyzed condensation of acetyl-CoA and oxaloacetate. Under certain conditions, citrate can be transported from mitochondria into the cytoplasm via a mitochondrial inner membrane citrate transport protein (CTP). This transport protein provides for the efflux of citrate from mitochondria generally in exchange for the influx of a carboxylate molecule (e.g., malate) from the cytosol. Cytosolic citrate can be converted to isocitrate which can serve as a substrate in the NADPH-generating oxidation reaction through which it is converted to α-ketoglutarate. Some yeast, typically oleaginous yeast, express an endogenous ATP citrate lyase which can catalyze the cleavage of citrate into oxaloacetate and acetyl-CoA. Thus, in such instances, cytosolic citrate can serve as a source of acetyl carbons that can be converted to acetyl-CoA. In general, yeast ATP citrate lysate is a dimer and can be heterodimeric (e.g., Acl1p/Acl2p) or homomeric.

In another embodiment of the cells and microorganisms, target molecule production systems and methods provided herein, carbon atoms incorporated into citrate that has been transferred to the cytosol can be captured through the cleavage of citrate to oxaloacetate and acetyl-CoA by the enzyme ATP citrate lyase (i.e., ACL, used interchangeably herein; e.g., EC 2.3.3.8). The capture of metabolite carbon in this manner can divert it from use in other metabolic processes and also can result in an increase in the level cytoplasmic acetyl-CoA. In one aspect of this embodiment, the amount and/or activity of ATP citrate lyase in the cytosol of modified cells is increased relative to the unmodified host cell. The amount and/or activity of a host ATP citrate lyase can be increased, for example, by introducing and/or increasing the number of copies of a gene encoding an ATP citrate lyase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the gene), by increasing the activity of a promoter that regulates transcription of a gene encoding an ATP citrate lyase, or by increasing the number of copies of a gene encoding an ATP citrate lyase and increasing the activity of a promoter that regulates transcription of a gene encoding an ATP citrate lyase. In some embodiments, an ATP citrate lyase is endogenous to

the host cell or microorganism. In other embodiments, a host cell or microorganism does not express an endogenous cytosolic ATP citrate lyase.

Thus, for example, heterologous nucleic acids encoding an ATP citrate lyase can be introduced into a host cell or microorganism to provide for an increased amount and/or activity of cytosolic ATP citrate lyase. In a particular embodiment, the ATP citrate lyase enzyme encoded by the heterologous nucleic acid can be an oleaginous yeast protein. An example of an oleaginous yeast ATP citrate lyase is formed by the *Yarrowia lipolytica* Acl1 and Acl2 proteins. Examples of *Y. lipolytica* Acl1p and Acl2p amino acid sequences are provided herein (SEQ ID NOS: 42 and 43). If a host cell or microorganism is a different species than the heterologous ATP citrate lyase that will be expressed in the host, it can be desirable to introduce nucleic acids encoding the ATP citrate lyase proteins that have been optimized for codons used in the host species. As a non-limiting example, the nucleotide sequences encoding *Yarrowia lipolytica* Acl1p and Acl2p that have been optimized for expression in a different yeast species (*Candida viswanathii*) are provided herein (SEQ ID NOS: 96 and 97). Additional examples of nucleotide sequences encoding ATP citrate lyase proteins include, but are not limited to: *Phaffia rhodozyma ACL1 and ACL2* (Genbank accession numbers KM503045, KM510496) and *Sordaria macrospora ACL1* and *ACL2* (Genbank accession numbers AJ224922, XM_003344949).

The promoter used for regulating transcription of a heterologous nucleic acid encoding an ATP citrate lyase can also be modified. For example, the amount of an ATP citrate lyase protein expressed in a cell may be increased by including in the heterologous nucleic acid a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a fatty acid-inducible promoter that can provide for increased ATP citrate lyase expression, particularly when exposed to fatty acids as a carbon source. Such promoter elements include those that regulate expression of peroxisomal proteins and/or β-oxidation enzymes in microbes, e.g., a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter.

ATP citrate lyase activity can be determined using assays known in the art and/or described herein. Such assays include, for example, methods described by Linn and Srere [(1979) *J. Biol. Chem.* 254:1691-1698] and Pentyala and Benjamin [(1995) *Biochemistry* 34:10961-10969]. Nucleic acid sequences representing native and/or modified ATP citrate lyase-encoding sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer

extension, nucleic acid hybridization, the like and combinations thereof), or the amounts of the nucleic acids or encoded proteins can be assessed using quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cells or organisms exhibit increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

In some embodiments of the cells, organisms and methods provided herein involving capture of acetyl carbons from cytosolic citrate, it may be beneficial to increase the amount of citrate in the cytosol. One approach to increasing cytosolic citrate levels is by increasing efflux of citrate from mitochondria into the cytosol. One method of increasing mitochondrial citrate efflux involves increasing the amount and/or activity of citrate transporter protein (CTP) in mitochondria of the modified cells. The amount and/or activity of a host citrate transporter protein can be increased, for example, by introducing and/or increasing the number of copies of a gene encoding a mitochondrial citrate transporter (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the gene), by increasing the activity of a promoter that regulates transcription of a gene encoding a mitochondrial citrate transporter, or by increasing the number of copies of a gene encoding a mitochondrial citrate transporter and increasing the activity of a promoter that regulates transcription of a gene encoding a mitochondrial citrate transporter. In some embodiments, a mitochondrial citrate transporter is endogenous to the host cell or microorganism.

Thus, for example, heterologous nucleic acids encoding a mitochondrial citrate transporter can be introduced into a host cell or microorganism to provide for an increased amount and/or activity of a mitochondrial citrate transporter. In a particular embodiment, the mitochondrial citrate transporter encoded by the heterologous nucleic acid can be a yeast protein. If a host cell or microorganism is a different species than the heterologous mitochondrial citrate transporter that will be expressed in the host, it can be desirable to introduce nucleic acids encoding the mitochondrial citrate transporter that have been optimized for codons used in the host species. Examples of nucleotide sequences encoding a protein that may have mitochondrial citrate transporter activity include, but are not limited to: *Candida albicans CTP1* (Genbank accession number XM_019475315), *Candida orthopsilosis CTP1* (Genbank accession number XM_003868562), *Saccharomyces cerevisiae CTP1* (Genbank accession number NM_001178639) and *Candida tropicalis* (Genbank accession number XM_002548023).

Modification of acyl-CoA formation, hydrolysis and use

Acyl-CoA is a molecule containing a carboxylic acid and coenzyme A joined through a thioester bond. In cells, acyl-CoA can be generated from carboxylic acids entering the cytosol from the extracellular environment, fatty acids synthesized within cells, lipid sidechains resulting from membrane turnover, products of the hydrolysis of triglyceride and sterol esters and carboxylation of acetyl-CoA. Acyl-CoA participates in multiple cellular pathways including lipid synthesis, β-oxidation, fatty acid synthesis and protein acylation. One reaction through which acyl-CoA is formed is the condensation between a thiol group of coenzyme A and a carboxy group of a carboxylic acid. This reaction between a fatty acid and coenzyme A is referred to as activation of the free fatty acid and can be catalyzed by an acyl-CoA synthetase enzyme (e.g. EC 6.2.1.3). A short-chain acyl-CoA, for example, malonyl-CoA, can also be generated through carboxylation of acetyl-CoA in a reaction catalyzed by an acetyl-CoA carboxylase (e.g., EC 6.4.1.2). Conversely, free fatty acids can be liberated from acyl-CoA through the action of a thioesterase (e.g., EC 3.1.2.20). Because acyl-CoA is a major carrier molecule of cellular carbons, its formation and hydrolysis represent certain aspects of methods of modifying carbon flux in cells.

Modification of acyl-CoA synthetase activity

Acyl-CoA synthetases (also referred to as fatty acid or acyl Co-A ligases and used interchangeably herein), are a family of enzymes in the enzyme classification subgroup 6.2.1 with varying substrate affinities, expression patterns and cellular localizations. In many microorganisms, there are multiple, distinct genes encoding separate acyl-CoA synthetases. Many yeast species (e.g., *Candida* spp. and *Saccharomyces* spp.) have five or six or more acyl-CoA synthetase genes encoding distinct enzymes. For example, *Saccharomyces cerevisiae* has 4 *FAA* genes (*FAA1*, *FAA2*, *FAA3* and *FAA4*) and 2 *FAT* genes (*FAT1* and *FAT2*) encoding acyl-CoA synthetase enzymes. Generally, *FAA* gene-encoded enzymes catalyze activation of acyl chains containing about 8-20 carbon atoms whereas the enzyme encoded by *FAT1* typically catalyzes activation of acyl chains containing 20 or more carbon atoms. Faa1p and Faa4p, which tend to be located in the cytosol and associated with membranes, are involved in activation of fatty acids internalized into cells from the extracellular medium and intracellular fatty acids arising from degradation of lipids, triacylglycerides and steryl esters. The Faa1p isozyme can exhibit broad substrate chain-length specificity, represents 90% of the cellular acyl-

CoA synthetase activity, and is localized in the cytosolic and microsomal fractions. Faa4p has broad chain-length specificity and has been shown to be important in protein myristoylation. Faa2p is localized to peroxisomes, has broad chain-length specificity, and participates in activation of fatty acids occurring during β-oxidation. Fat1p is typically a dual function protein localized to the cellular membrane that has activity for both fatty acid transport and fatty acid activation. Fat2p tends to be targeted to the peroxisomal membrane for medium chain fatty acid transport and activation.

Homologs for *FAA1* and *FAT1* have been identified in *Candida* strains. Acyl-CoA synthetase has six isoforms encoded by FAA1, FAT1, ACS2A, ACS2B, ACS2C and ACS2D, respectively, in *Candida* spp. (e.g., homologous to FAA1, FAT1, and FAA2 in *S. cerevisiae*). Two of the homologs display 95% identity to one another and are most likely alleles of the same gene. Four *FAA2* homologs have been identified in *Candida* strain ATCC 20336 (also referred to in the art as acyl-CoA synthetase-encoding genes *ACS2A* through *ACS2D* in *Candida*). Examples of *Candida viswanathii* nucleotide sequences (and corresponding amino acid sequences) encoding acyl-CoA synthetases are provided herein (nucleotide SEQ ID NOS: 91 and 98 and amino acid SEQ ID NOS: 35 and 44) and in International patent application no. PCT/US2012/045615 (publication no. WO 2013/106730). Acetyl-CoA synthetase-encoding genes are also referred to as *ACS* genes, as described herein. For clarity, acyl-CoA synthetase-encoding genes are referred to as *FAA* or *FAT* herein and not as *ACS* genes (which herein refer to acetyl-CoA synthetase-encoding genes).

In some embodiments of the microorganisms, compositions and methods provided herein, the amount and/or activity of acyl-CoA synthetase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase acyl-CoA synthetase and/or acyl-CoA synthetase activity, may be modified to decrease acyl-CoA synthetase and/or acyl-CoA synthetase activity, or may be modified to alternately increase and decrease acyl-CoA synthetase and/or acyl-CoA synthetase activity depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

In certain aspects, the amount and/or activity of an acyl-CoA synthetase in a cell is decreased. Reducing or eliminating the amount and/or activity of an acyl-CoA synthetase may be particularly beneficial in embodiments in which the flux of carbons from fatty acids is directed

toward a particular target product pathway and away from other cellular metabolic pathways involving activated fatty acids. When free internalized or cytosolic fatty acids are activated by acyl-CoA synthetase and used in cellular processes, such as lipid biosynthesis, the carbon atoms in the free fatty acids are not available for use in the cell or microorganism production of commercially important chemicals. Without being limited by theory, it is believed that reduction in the amount of fatty-acyl-CoA available for various cellular processes can increase the amount of fatty acids available for conversion into target molecules, for example, a fatty dicarboxylic acid (e.g., adipic acid, suberic acid, sebacic acid and dodecanedioic acid) by other engineered pathways in the same host cell or organism (e.g., omega oxidation pathway, beta oxidation pathway, omega oxidation pathway and beta oxidation pathway).

In some embodiments, one strategy is to control the subcellular location of acyl-CoA synthetase enzyme activity so that it is present only in the peroxisome. FAA1 and FAT1 mutants, faa1Δ and fat1Δ, of *Candida* were constructed and should have very little acyl-CoA synthetase activity targeted to the cytoplasm. In these mutant strains, exogenously supplied long-chain free fatty acids tend to accumulate in the cytoplasm since they cannot be transported into the peroxisome unless they are activated to the acyl-CoA thioester. High concentrations of free fatty acid can be toxic, so the cell acts to detoxify itself by oxidizing the free fatty acids to dicarboxylic acids that are much less toxic. Unlike long-chain fatty acids, long-chain dicarboxylic acids are able to diffuse into the peroxisomal compartment where they can then be activated to diacyl-CoA thioesters and enter into the beta-oxidation pathway. With multiple peroxisomal acyl-CoA synthetase isozymes it may be that each isozyme has different substrate specificity. In some embodiments, it is desired to retain those peroxisomal acyl-CoA synthetase enzymes with substrate specificity matching the chain-length of the fatty acid feedstock but without activity (or low activity) on diacids of chain-length ≤ 6, 8, 10, 12, 14, 16, 18 or 20 carbons. With this strategy, any long-chain dicarboxyl-CoA that is chain-shortened by beta-oxidation to 12 carbons, for example, that is subsequently hydrolyzed to a dicarboxylic acid and free CoA cannot be reactivated to a dicarboxyl-CoA for re-entry into beta-oxidation for further chain shortening. In some embodiments, in combination with controlling the substrate chain-length specificity of the peroxisomal acyl-CoA synthetase, a peroxisomal thioesterase activity is amplified with maximum activity at the desired chain-length of a target product. This strategy can control the chain-length of the dicarboxylic acid produced by beta-oxidation.

Certain aspects of the microorganisms, compositions and methods provided herein include one or more modifications to reduce or eliminate cytosolic activation of free fatty acids into acyl-CoA. One approach to reducing or eliminating cytosolic free fatty acid activation is to decrease the amount and/or activity of an acyl-CoA synthetase. For example, endogenous microbial genes encoding one or more acyl-CoA synthetases (e.g., yeast *FAA1, FAA4* and/or *FAT1* gene) can be disrupted or deleted in a host cell or microorganism to reduce or eliminate acyl-CoA synthetase activity in the host relative to a cell or microorganism in which the gene(s) have not been modified. Methods for decreasing the amount and/or activity of one or more acyl-CoA synthetases, such as acyl-CoA synthetases involved in activation of cytosolic free fatty acids, in a cell include, but are not limited to, modifying the amount of acyl-CoA synthetase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous acyl-CoA synthetase gene in a cell or organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type acyl-CoA synthetase such that the encoded modified or substituted acyl-CoA synthetase protein has a reduced enzyme activity. For example, expression of a host acyl-CoA synthetase activity can be decreased or eliminated by disruption (e.g., knockout, insertion mutagenesis, the like and combinations thereof) of a host gene encoding the protein, or by decreasing the activity of the promoter (e.g., through addition of repressor sequences to the promoter or 5'UTR or replacing the promoter) that controls transcription of an acyl-CoA synthetase gene using recombinant molecular biology techniques known in the art and/or described herein. In one embodiment, a diploid yeast, such as, for example, a *Candida* yeast, when used as a host microorganism can be subjected to genetic modification in which one of the two alleles of an acyl-CoA synthetase gene is disrupted or deleted. In so doing, a single allele of the gene remains for a reduced amount of acyl-CoA synthetase expression in the microorganism and a reduced amount of the protein in the cell.

One method for disrupting an endogenous acyl-CoA synthetase gene is by recombinantly inserting a heterologous nucleic acid (e.g., a nucleotide sequence encoding a selectable marker such as an enzyme that restores an auxotrophic host organism to prototrophy) into the endogenous gene, thereby generating an engineered organism deficient in acyl-CoA synthetase activity. This can be done, for example, through homologous recombination in which a heterologous nucleic acid containing sequences of an endogenous acyl-CoA synthetase gene and a disrupting sequence (e.g., a knock-out gene cassette such as described herein) is introduced into a host cell or microorganism. Nucleic acids encoding an acyl-CoA synthetase

can be obtained from a number of sources, including, for example, yeast cells. Genomic DNA from cell sources can be amplified using oligonucleotide primers based on the nucleotide sequence of an acyl-CoA synthetase encoding gene, including examples provided herein. Nucleotide sequences encoding (and the amino acid sequences of) *Candida viswanathii* acyl-CoA synthetase Faa1p and Fat1p are provided herein (nucleotide SEQ ID NOS: 91 and 98 and amino acid SEQ ID NOS: 35 and 44). Additional non-limiting examples of nucleic acids encoding acyl-CoA synthetases include *Saccharomyces cerevisiae FAA1* (Genbank accession numbers NM_001183737 and M96371), *Saccharomyces cerevisiae FAA4* (Genbank accession number NM_001182754), *Saccharomyces cerevisiae FAA2* (Genbank accession number NM_001178906), *Saccharomyces cerevisiae FAA3* (Genbank accession number NM_001179359), *Yarrowia lypolytica YAL1* (Genbank accession number XM_502959), *Yarrowia lypolytica FAT1* (Genbank accession number NC_006071), *Candida albicans FAA4* (Genbank accession number XM_714261), *Aspergillus nomius FAA4* (Genbank accession number XM_015551345), *Coccidioides immitis FAA4* (Genbank accession number XM_001240655) and *Aspergillus niger FAA4* (Genbank accession number XM_001397786).

In other embodiments, the amount and/or activity of a host acyl-CoA synthetase can be increased, for example, by increasing the number of copies of a gene encoding an acyl-CoA synthetase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the gene), by increasing the activity of a promoter that regulates transcription of a gene encoding an acyl-CoA synthetase, or by increasing the number of copies of a gene encoding an acyl-CoA synthetase and increasing the activity of a promoter that regulates transcription of a gene encoding an acyl-CoA synthetase. In some embodiments, an acyl-CoA synthetase is endogenous to the host cell or microorganism. In particular embodiments, the amount and/or activity of a host peroxisomal acyl-CoA synthetase is increased.

The presence, absence or amount of acyl-CoA synthetase activity can be detected by any suitable method known in the art. Non-limiting examples of suitable detection methods include enzymatic assays (e.g., Lageweg et al. (1991) *Anal. Biochem.* 197(2):384-388, Erland et al. (2001) *Anal. Biochem.* 295(1):38-44), PCR based assays (e.g., qPCR, RT-PCR), immunological detection methods (e.g., antibodies specific for acyl-CoA synthetase), the like and combinations thereof. Methods for determining acyl-CoA synthetase activities also include assays described by Trigatti et al. [(1992) *Biochem. Cell. Biol.* 70:76-80] and Kamiryo et al. [(1977) *Proc. Natl. Acad. Sci. USA* 74:4947-4950].

Modification of acetyl-CoA carboxylase activity

Malonyl-CoA is a coenzyme A derivative of the dicarboxylic acid malonic acid that can serve as a precursor in the synthesis of numerous valuable organic molecules, including fatty acids and polyketides. In the cytosol, malonyl-CoA can be generated by carboxylation of acetyl-CoA through the addition of CO_2 (e.g., derived from bicarbonate) in a reaction catalyzed by the enzyme acetyl-CoA carboxylase (e.g., EC 6.4.1.2). Acetyl-CoA carboxylase sometimes is also referred to as "acetyl-CoA:carbon-dioxide ligase (ADP-forming)" and "acetyl coenzyme A carboxylase". In eukaryotes, acetyl-CoA carboxylase is a multifunctional polypeptide containing a biotin carrier protein domain, a biotin carboxylase domain and a carboxyl-transferase domain. Biotin joined to the biotin carrier protein is a co-factor in malonyl-CoA formation. It receives CO_2 which becomes attached to it at a biotin ring nitrogen in an ATP-dependent reaction catalyzed by the biotin carboxylase of the acetyl-CoA carboxylase. The activated CO_2 is then transferred from biotin to acetyl-CoA by the carboxyl-transferase domain to form malonyl-CoA. Malonyl-CoA can serve as a carbon donor in the synthesis of a fatty acid chain in repeated cycles of the addition of 2 carbon atoms per cycle to extend the chain and generate a fatty acid. The reactions of each cycle are catalyzed by fatty acid synthase (FAS) and continue until typically a 16-carbon fatty acid (palmitic acid) or 18-carbon fatty acid (stearic acid) is completed in the form of palmitoyl-CoA or stearoyl-CoA, respectively. Accordingly, a supply of malonyl-CoA, and/or precursors and enzymes (e.g., acetyl-CoA carboxylase) that generate malonyl-CoA, can be required for fatty acid synthesis.

In some embodiments of the cells, microorganisms, compositions and methods provided herein, the amount and/or activity of acetyl-CoA carboxylase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase acetyl-CoA carboxylase and/or acetyl-CoA carboxylase activity, may be modified to decrease acetyl-CoA carboxylase and/or acetyl-CoA carboxylase activity, or may be modified to alternately increase and decrease acetyl-CoA carboxylase and/or acetyl-CoA carboxylase activity depending, for example, on the target molecule(s) produced and/or on the conditions in which the modified cell or microorganism is cultured.

A modification of cellular carbon flux that increases cytosolic acetyl-CoA alone may not be optimal for enhancing fatty acid or other target molecule production in an engineered, cell-based

or microbial system if there is not a concurrent increase in conversion of acetyl-CoA to malonyl-CoA. To maximize production efficiencies, included in the cells, microorganisms, compositions and methods provided herein are cellular carbon flux modifications that increase the amount of cytosolic malonyl-CoA. In one embodiment, the amount and/or activity of acetyl-CoA carboxylase is increased in the cytosol to direct carbon flux towards generation of malonyl-CoA. The amount and/or activity of a host acetyl-CoA carboxylase can be increased, for example, by increasing the number of copies of a gene encoding an acetyl-CoA carboxylase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the gene), by increasing the activity of a promoter that regulates transcription of a gene encoding an acetyl-CoA carboxylase, or by increasing the number of copies of a gene encoding an acetyl-CoA carboxylase and increasing the activity of a promoter that regulates transcription of a gene encoding an acetyl-CoA carboxylase. In some embodiments, an acetyl-CoA carboxylase is endogenous to the host cell or microorganism. An acetyl-CoA carboxylase activity may be amplified by over-expression of an acetyl-CoA carboxylase gene by any suitable method. Non-limiting examples of methods suitable to amplify or over express a gene include amplifying the number of acetyl-CoA carboxylase genes in yeast, for example, following transformation with a high-copy number plasmid (e.g., such as one containing a 2u origin of replication), integration of multiple copies of the gene into the host genome, over-expression of the gene directed by a strong promoter, the like or combinations thereof. An acetyl-CoA carboxylase gene may be native to *Candida tropicalis* or *Candida viswanithii*, for example, or it may be obtained from a heterologous source. Examples of a *Candida viswanathii* acetyl-CoA carboxylase polypeptide amino acid sequence (Acc1p), and nucleotide sequence encoding it (*ACC1*), are provided herein (nucleotide SEQ ID NO: 74 and amino acid SEQ ID NOS: 18 and 19). Additional non-limiting examples of nucleic acids encoding an acetyl-CoA carboxylase include *Yarrowia lypolytica ACC1* (Genbank accession NC_006069), *Saccharomyces cerevisiae ACC1* (Genbank accession NM_001183193), *Candida tropicalis ACC* (Genbank accession number XM_002546179), *Candida albicans ACC1* (Genbank accession number XM_713531), *Aspergillus nidulans ACCA* (Genbank accession number Y15996), *Aspergillus niger ACCA* (Genbank accession number XM_001395439), *Aspergillus oryzae ACC* (Genbank accession number XM_001826359), *Schizosaccharomyces pombe ACC* (Genbank accession D78169), *Neurospora crassa ACC* (Genbank accession XM_957924), *Lipomyces starkeyi ACC1* (Genbank accession KJ948118), *Debaryomyces hansenii ACC1* (Genbank accession XM_457211), *Amylomyces rouxii ACC* (Genbank accession EF397565) and *Coccidioides immitis ACC* (Genbank accession number XM_001247055).

The promoter used for regulating transcription of a heterologous nucleic acid encoding an acetyl-CoA carboxylase can also be modified. For example, the amount of an acetyl-CoA carboxylase protein expressed in a particular cellular location may be increased by including in the heterologous nucleic acid a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a fatty acid-inducible promoter that can provide for increased acetyl-CoA carboxylase expression, particularly when exposed to fatty acids as a carbon source. Such promoter elements include those that regulate expression of peroxisomal proteins and/or β-oxidation enzymes in microbes, e.g., a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene promoter. The nucleotide sequence of a *Candida viswanathii* HDE gene promoter is provided herein as are examples of additional fatty acid-inducible promoters.

The reverse activity (e.g., decarboxylation of malonyl-CoA) is carried out by a separate enzyme, malonyl-CoA decarboxylase. In some embodiments, to further increase carbon flux through a particular reaction or through a metabolic pathway, one or more reverse activities in the pathway can be altered to inhibit the back conversion of a desired product into its starting reactants. In certain embodiments, a malonyl-CoA decarboxylase activity is reduced or eliminated to further increase the carbon flux through an acetyl-CoA carboxylase activity in the direction of malonyl-CoA production.

Acetyl-CoA carboxylase is regulated by feedback inhibition of acyl-CoA (e.g., palmitoyl-CoA) and by phosphorylation. As such, increasing the copy number of acetyl-CoA carboxylase-encoding nucleic acids in a cell may not alone be sufficient in increasing the acetyl-CoA carboxylase activity in the cell. Because the dephosphorylated state is the active state of the enzyme, one approach for increasing the activity of acetyl-CoA carboxylase is to reduce or eliminate phosphorylation of the protein. Provided herein are modified acetyl-CoA carboxylase proteins (and mutant nucleic acids encoding the proteins) in which one or more phosphorylatable serine residues have been substituted with alanine residues thereby relieving the regulation by phosphorylation. In a particular embodiment, the modified acetyl-CoA carboxylase is a modified yeast enzyme. For example, as described herein, an endogenous *Candida viswanathii* acetyl-CoA carboxylase wild-type enzyme was modified to substitute alanine residues for one or more of the following serine amino acid residues: S652, S1131, S1138, S1153, S1158. The modified amino acid sequences are provided herein. Also provided

herein are cells and microorganisms containing heterologous nucleic acid encoding a modified acetyl-CoA carboxylase protein and methods of increasing cytosolic malonyl-CoA generation and producing organic acids and other target products using the compositions provided herein. In another aspect of this embodiment, the amount and/or activity of acetyl-CoA carboxylase is increased in the cytosol of modified cells relative to an unmodified host cell by introducing multiple copies of the modified heterologous nucleic acid encoding acetyl-CoA carboxylase into a host cell to provide for increased acetyl-CoA carboxylase protein and/or by increasing the expression of the modified acetyl-CoA carboxylase in the cell through the use of a strong heterologous promoter.

In other aspects, the amount and/or activity of acetyl-CoA carboxylase in a cell is decreased. Methods for decreasing acetyl-CoA carboxylase activity in a cell include, but are not limited to, modifying the amount of acetyl-CoA carboxylase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous acetyl-CoA carboxylase gene in an organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type acetyl-CoA carboxylase such that the encoded modified or substituted acetyl-CoA carboxylase protein has a reduced enzyme activity.

The presence, absence or amount of acyl-CoA carboxylase activity can be detected by any suitable method known in the art. Non-limiting examples of suitable detection methods include radioactive HCO_3^- incorporation assays and coupled enzyme assays (e.g., Diacovich et al. (2002) *J. Biol. Chem.* 277(34):31228-31236), PCR based assays (e.g., qPCR, RT-PCR), immunological detection methods (e.g., antibodies specific for acyl-CoA carboxylase), the like and combinations thereof.

Modification of thioesterase activity

A thioesterase is an enzyme that catalyzes the hydrolysis of a thioester bond between a carbonyl group and a sulfur atom. In cells, certain thioesterases (e.g., acyl-CoA thioesterase activity, acyl-ACP thioesterase activity) catalyze the removal of Coenzyme A or acyl carrier protein (e.g., ACP) from a fatty acid yielding a free fatty acid and unesterified carrier, e.g., Coenzyme A (CoASH). The reaction occurs in the presence of water, and Coenzyme A or acyl carrier protein is specifically removed at a thiol group. The released CoA can then be reused for other cellular activities. A non-limiting example of an enzyme with thioesterase activity is acyl-

CoA hydrolase (e.g., EC 3.1.2.20; also referred to as acyl coenzyme A thioesterase, acyl-CoA thioesterase, acyl coenzyme A hydrolase, thioesterase B, thioesterase II, lecithinase B, lysophopholipase L1, acyl-CoA thioesterase 1, and acyl-CoA thioesterase). In eukaryotic microorganisms, acyl-CoA thioesterases are generally localized in peroxisomes but may also occur in mitochondria.

In some embodiments of the cells, microorganisms, compositions and methods provided herein, the amount and/or activity of a thioesterase in a cell is modified. For example, in some aspects, a cell or microorganism may be modified to increase the amount of thioesterase and/or thioesterase activity, may be modified to decrease thioesterase and/or thioesterase activity, or may be modified to alternately increase and decrease thioesterase and/or thioesterase activity depending, for example, on the cellular location(s) of the enzyme and/or on the conditions in which the modified cell or microorganism is cultured.

Embodiments of cells and microorganisms provided herein in which carbon flux is modified to increase acetyl-CoA carboxylase and/or cytosolic malonyl-CoA, may further benefit from also modifying the amount of medium-to-long chain fatty acids present in the cytosol in the esterified form as acyl-CoA (e.g., palmitoyl-CoA). In some of these embodiments, the increased generation of malonyl-CoA can lead to increased fatty acid synthesis in the presence of cytosolic fatty acid synthase (FAS). The end-product of cytosolic fatty acid synthesis in yeast cells is typically an acyl-CoA, e.g., palmitoyl-CoA, which can then be used in cellular metabolic pathways other than desired engineered target molecule production processes. This represents a loss of the carbon atoms in the acyl-CoA which could have been incorporated into target products. Additionally, high levels of cytosolic acyl-CoA end products of fatty acid synthesis (e.g., palmitoyl-CoA) can inhibit acetyl-CoA carboxylase. Therefore, production efficiency may be enhanced in some embodiments by decreasing the amount of fatty acids present in the cytosol in the esterified form as acyl-CoA.

Included in the cells, microorganisms, compositions and methods provided herein are cellular carbon flux modifications that decrease the amount of fatty acids present in the cytosol in the esterified form as acyl-CoA. In one embodiment, the amount and/or activity of thioesterase is increased in the cytosol of cells or microorganisms. The amount and/or activity of a thioesterase can be increased, for example, by increasing the number of copies of a gene encoding a thioesterase (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more copies of the gene),

by increasing the activity of a promoter that regulates transcription of a gene encoding a thioesterase, or by increasing the number of copies of a gene encoding a thioesterase and increasing the activity of a promoter that regulates transcription of a gene encoding a thioesterase. In certain aspects, the amount and/or activity of a thioesterase in a cell and/or a particular cellular location is increased. For example, the amount and/or activity of a thioesterase in the cytosol of a cell can be increased. In some embodiments, the pattern of expression of a thioesterase can be modified such that the enzyme is produced in a cellular location where it is not produced in an unmodified cell and/or is no longer produced in a cellular location where it is produced in an unmodified cell.

In yeast, thioesterases are generally present in the peroxisomal compartment of the cells to ensure that free coenzyme A is available for beta-oxidation. Without being bound or limited by theory, this enzyme should not normally be present in the cytoplasm because producing fatty acyl-CoA via cytosolic fatty acid synthesis is an energy intensive process, and removing the CoA from the synthesized acyl-CoA would waste the energy put into the process. In order to provide for generation of free fatty acids in the cytoplasm of modified cells or organisms, peroxisomal thioesterase with activity on long chain fatty acids can be re-targeted to the cytoplasm. In one embodiment of the cells, microorganisms and methods provided herein, host cells are modified to express engineered thioesterase polypeptides that lack targeting signals that direct the enzyme to peroxisomes (i.e, PTS), thereby introducing, or increasing the amount of, cytosolic thioesterase in the cells. In a particular embodiment, a thioesterase lacking a PTS that has activity on medium and long chain fatty acids is heterologously expressed in the cytoplasm of modified cells. This can be accomplished by modifying nucleic acids encoding a peroxisomal thioesterase to delete the portion of the nucleic acid encoding the peroxisomal targeting signal at the C-terminus of the protein, and introducing the modified heterologous nucleic acid into host cells. An example of one such modified thioesterase protein is a yeast Tes3p$^{\Delta pts}$ which excludes a peroxisomal targeting signal (PTS). As described herein, generally, a yeast peroxisomal targeting sequence is a 3-amino acid consensus sequence (PTS1). In a particular embodiment, the thioesterase enzyme can be a *Candida* yeast protein. For example, *Candida* strain ATCC 20336 contains eight genes encoding peroxisomal thioesterases (*TES1-TES8*), each of which contains a C-terminal 3-amino acid PTS1 consensus sequence (SRL, ARL) or slight variant thereof (PKL, PKF). Nucleotide sequences encoding the 8 thioesterases, and the amino acid acid sequences of the thioesterases, are provided herein (nucleotide SEQ ID NOS: 78, 79, 80, 81, 82, 83, 84, 85, 86 and 87 and amino acid SEQ ID NOS: 22, 23, 24, 25,

26, 27, 28, 29, 30 and 31). An example of a *Candida viswanathii* nucleotide sequence encoding a modified Tes3p lacking a peroxisomal-targeting sequence (*TES3*$^{\Delta pts}$), and the amino acid sequence of the modified Tes3p, are also provided herein (nucleotide SEQ ID NO: 88 and amino acid SEQ ID NO: 32). In some embodiments, the thioesterase polypeptide is from a different species than a host microorganism in which it is expressed.

Nucleic acid sequences encoding polypeptides conferring thioesterase activity can be obtained from a number of sources, including, for example, yeast (e.g., *Candida*, *Saccharomyces*, *Debaryomyces*, *Meyerozyma*, *Lodderomyces*, *Scheffersomyces*, *Clavispora*, *Yarrowia*, *Pichia,* *Kluyveromyces*, *Eremothecium*, *Zygosaccharomyces*, *Lachancea*, *Nakaseomyces*), animals (e.g., *Homo*, *Rattus*), bacteria (e.g., *Escherichia*, *Pseudomonas*, *Bacillus*), and plants (e.g., *Arabidopsis*, *Nictotania*, *Cuphea*). Examples of nucleotide sequences encoding polypeptides having thioesterase activity include, without limitation, *Saccharomyces cerevisiae PTE1* (Genbank accession no. AF124265), *Debaryomyces hansenii* (Genbank accession nos. XM_456353, XM_459767), *Aspergillus niger* (Genbank accession nos. XM_001392518, XM_011389712, XM_011395790), *Aspergillus fumigatus* (Genbank accession no. XM_742375), *Candida albicans* (Genbank accession nos. XM_705831, XM_705833), *Candida dubliniensis* (Genbank accession no. XM_002418475), *Candida orthopsilosis* (Genbank accession nos. XM_003866686, XM_003866684), *Neurospora crassa* (Genbank accession nos. XM_956915, XM_960627), *Rhodotorula toruloides* (Genbank accession no. XM_016414800), *Cryptococcus neoformans* (Genbank accession no. XM_012196078, XM_012195836), *Escherichia coli TesA* (Genbank accession no. L06182) and acyl-(ACP) thioesterase type B from *Cuphea lanceolata* (Genbank accession no. CAB60830).

The promoter used for regulating transcription of a heterologous nucleic acid encoding a thioesterase can also be modified. For example, the amount of a thioesterase protein expressed in a particular cellular location may be increased by including in the heterologous nucleic acid a strong heterologous promoter and/or a promoter that provides for a different pattern of expression in the cell or microorganism. An example of one such heterologous promoter is a fatty acid inducible promoter that can provide for increased thioesterase expression, particularly when exposed to fatty acids as a carbon source. Such promoter elements include those that regulate expression of peroxisomal proteins and/or β-oxidation enzymes in microbes, e.g., a *Candida* hydratase-dehydrogenase-epimerase (*HDE*) gene

promoter. The nucleotide sequence of a *Candida viswanathii* HDE gene promoter is provided herein as are examples of additional fatty acid-inducible promoters.

In other aspects, the amount and/or activity of a thioesterase in a cell is decreased. Methods for decreasing thioesterase activity in a cell include, but are not limited to, modifying the amount of thioesterase protein expression in the cell, for example, by replacing the wild-type promoter of an endogenous thioesterase gene in an organism with a weaker heterologous promoter, deleting or disrupting an endogenous gene, and/or replacing or modifying a gene encoding a wild-type such that the encoded modified or substituted thioesterase protein has a reduced enzyme activity.

Presence, absence or amount of thioesterase activity can be detected by any suitable method known in the art or described herein (see, e.g., Jones et al. (1999) *J. Biol. Chem.* 274(14):9216-9223 and Chemistry and Biology 9: 981-988). Nucleic acid sequences representing native and/or modified thioesterase-encoding sequences also can be detected using nucleic acid detection methods (e.g., PCR, primer extension, nucleic acid hybridization, the like and combinations thereof), or the amounts of the nucleic acids or encoded proteins can be assessed using quantitative expression based analysis (e.g., RT-PCR, western blot analysis, northern blot analysis, the like and combinations thereof), where the engineered cells or organisms exhibit increased or decreased RNA and/or polypeptide levels as compared to the host cell or organism.

Methods of modifying cells and organisms

Provided herein are cells and organisms (including microorganisms) that have been modified in one or more aspects relative to the unmodified cell or organism (i.e., the cell or organism prior to the modification). For example, a cell or organism can be modified by altering one or more cellular activities and/or the sum total of a cell's or organism's activities. Thus, modifications can include, for example, alteration of cellular activities, addition of cellular activities and/or elimination of cellular activities. A cell or organism may be modified, for example, by altering the amount of one or more cellular compositions, e.g, polynucleotides and/or polypeptides. In some embodiments, an activity and/or amount of a composition can be altered by genetically modifying a host cell or microorganism which yields an engineered cell or microorganism having added, increased, reduced, decreased or removed activity or composition. Genetic

modifications can be achieved in several ways, including, for example, introducing heterologous nucleic acids into host cells or organisms using molecular biological techniques known in the art and/or described herein.

Polynucleotides

The term "polynucleotide" is used herein interchangeably with the term "nucleic acid" and refers to an organic polymer composed of two or more monomers including nucleotides, nucleosides or analogs thereof, including but not limited to single stranded or double stranded, sense or antisense deoxyribonucleic acid (DNA) of any length and, where appropriate, single stranded or double stranded, sense or antisense ribonucleic acid (RNA) of any length, including siRNA. The term "nucleotide" refers to any of several compounds that consist of a ribose or deoxyribose sugar joined to a purine or a pyrimidine base (nitrogenous base) and to a phosphate group, and that are the basic structural units of nucleic acids. The term "nucleoside" refers to a compound (as guanosine or adenosine) that consists of a purine or pyrimidine base (nitrogenous base) combined with deoxyribose or ribose and is found especially in nucleic acids. The term "nucleotide analog" or "nucleoside analog" refers, respectively, to a nucleotide or nucleoside in which one or more individual atoms have been replaced with a different atom or with a different functional group. Accordingly, the term polynucleotide includes nucleic acids of any length, DNA, RNA, analogs and fragments thereof. A polynucleotide of three or more nucleotides is also called nucleotidic oligomer or oligonucleotide.

A nucleic acid (e.g., a nucleic acid reagent, target nucleic acid, target nucleotide sequence, nucleic acid sequence of interest or nucleic acid region of interest) can be from any source or composition, such as DNA, cDNA, gDNA (genomic DNA), RNA, siRNA (short inhibitory RNA), RNAi, tRNA or mRNA, for example, and can be in any form (e.g., linear, circular, supercoiled, single-stranded, double-stranded, and the like). A nucleic acid can also comprise DNA or RNA analogs (e.g., containing base analogs, sugar analogs and/or a non-native backbone and the like). It is understood that the term “nucleic acid” does not refer to or infer a specific length of the polynucleotide chain, thus polynucleotides and oligonucleotides are also included in the definition. Deoxyribonucleotides include deoxyadenosine, deoxycytidine, deoxyguanosine and deoxythymidine. For RNA, the uracil base is uridine.

The terms “endogenous polynucleotide,” “endogenous nucleic acid,” “native polynucleotide” and “native nucleic acid,” used interchangeably herein, refer to a polynucleotide of a cell or organism that exists, or is inherent, in the genetic composition of the cell or organism prior to modification.

The terms "heterologous polynucleotide," "heterologous nucleic acid," “exogenous polynucleotide,” “exogenous nucleic acid,” “foreign polynucleotide” and “foreign nucleic acid,” used interchangeably herein, refer to a polynucleotide as it relates to a particular reference cell or organism (e.g., a host cell or organism) and is one that is not present in the genetic composition of the reference cell or organism. A heterologous polynucleotide includes a polynucleotide that may be identical in nucleotide sequence to an endogenous polynucleotide present in a cell, but if introduced into the cell would alter the genetic composition of the cell by, for example, increasing the copy number of the polynucleotide in the cell, altering the position(s) of the polynucleotide in the cell genome, altering the expression of the polynucleotide in the cell, and the like. Thus, such a heterologous nucleic acid thereby genetically modifies the cell into which it is introduced. A heterologous polynucleotide in a host cell may exist in a nucleic acid autonomous of the host chromosome or may be inserted into a host chromosome. A heterologous polynucleotide can also be a polynucleotide with a different nucleotide sequence relative to any nucleic acid in a particular reference cell and can also be obtained from a different cell type or species of organism. A heterologous nucleic acid can also be generated by synthetic methods known in the art and/or described herein.

The term "expression" with respect to a nucleic acid sequence or protein refers to transcription of the nucleic acid and/or, as appropriate, translation of an mRNA transcript to a protein (protein synthesis). Thus, as will be clear from the context, expression of a protein results from transcription and translation of an open reading frame (ORF) sequence. The level of expression of a nucleic acid and/or protein in a cell may be determined, for example, on the basis of either the amount of RNA transcript of a nucleic acid that is present in the cell and/or the amount of the product encoded by the nucleic acid. For example, mRNA transcribed from a selected sequence can be quantitated by qRT-PCR or by Northern hybridization (see Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1989)). Protein encoded by a nucleic acid can be quantitated by various methods, e.g., by ELISA, by assaying for the biological activity of the protein, or by employing assays that are independent of such activity, such as western blotting or radioimmunoassay, using antibodies that recognize and bind the protein (see Sambrook et al., 1989, supra).

A nucleic acid sometimes is a plasmid, phage, autonomously replicating sequence (ARS), centromere, artificial chromosome, yeast artificial chromosome (e.g., YAC) or other nucleic acid able to replicate or be replicated in a host cell. In certain embodiments a nucleic acid can be from a library or can be obtained from enzymatically digested, sheared or sonicated genomic DNA (e.g., fragmented) from an organism of interest. Fragments can be generated by any suitable method in the art, and the average, mean or nominal length of nucleic acid fragments can be controlled by selecting an appropriate fragment-generating procedure. In some embodiments, the fragmented DNA can be size selected to obtain nucleic acid fragments of a particular size range. Nucleic acid can be fragmented by various methods known in the art, which include without limitation, physical, chemical and enzymic processes. Examples of such processes are described in U.S. Patent Application Publication No. 20050112590 (published on May 26, 2005, entitled "Fragmentation-based methods and systems for sequence variation detection and discovery," naming Van Den Boom et al.). Certain processes can be selected to generate non-specifically cleaved fragments or specifically cleaved fragments. Examples of processes that can generate non-specifically cleaved fragment sample nucleic acid include, without limitation, contacting sample nucleic acid with apparatus that expose nucleic acid to shearing force (e.g., passing nucleic acid through a syringe needle; use of a French press); exposing sample nucleic acid to irradiation (e.g., gamma, x-ray, UV irradiation; fragment sizes can be controlled by irradiation intensity); boiling nucleic acid in water (e.g., yields about 500 base pair fragments) and exposing nucleic acid to an acid and base hydrolysis process.

Nucleic acid may be specifically cleaved by contacting the nucleic acid with one or more specific cleavage agents. The term "specific cleavage agent" as used herein refers to an agent, sometimes a chemical or an enzyme that can cleave a nucleic acid at one or more specific sites. Specific cleavage agents often will cleave specifically according to a particular nucleotide sequence at a particular site. Examples of enzyme specific cleavage agents include without limitation endonucleases (e.g., DNase (e.g., DNase I, II); RNase (e.g., RNase E, F, H, P); Cleavase™ enzyme; Taq DNA polymerase; E. coli DNA polymerase I and eukaryotic structure-specific endonucleases; murine FEN-1 endonucleases; type I, II or III restriction endonucleases such as Acc I, Afl III, Alu I, Alw44 I, Apa I, Asn I, Ava I, Ava II, BamH I, Ban II, Bcl I, Bgl I. Bgl II, Bln I, Bsm I, BssH II, BstE II, Cfo I, Cla I, Dde I, Dpn I, Dra I, EclX I, EcoR I, EcoR I, EcoR II, EcoR V, Hae II, Hae II, Hind II, Hind III, Hpa I, Hpa II, Kpn I, Ksp I, Mlu I, MluN I, Msp I, Nci I, Nco I, Nde I, Nde II, Nhe I, Not I, Nru I, Nsi I, Pst I, Pvu I, Pvu II, Rsa I, Sac I, Sal I, Sau3A I,

Sca I, ScrF I, Sfi I, Sma I, Spe I, Sph I, Ssp I, Stu I, Sty I, Swa I, Taq I, Xba I, Xho I); glycosylases (e.g., uracil-DNA glycolsylase (UDG), 3-methyladenine DNA glycosylase, 3-methyladenine DNA glycosylase II, pyrimidine hydrate-DNA glycosylase, FaPy-DNA glycosylase, thymine mismatch-DNA glycosylase, hypoxanthine-DNA glycosylase, 5-Hydroxymethyluracil DNA glycosylase (HmUDG), 5-Hydroxymethylcytosine DNA glycosylase, or 1,N6-etheno-adenine DNA glycosylase); exonucleases (e.g., exonuclease III); ribozymes, and DNAzymes. Nucleic acid may be treated with a chemical agent, or synthesized using modified nucleotides, and the modified nucleic acid may be cleaved. In non-limiting examples, nucleic acid may be treated with (i) alkylating agents such as methylnitrosourea that generate several alkylated bases, including N3-methyladenine and N3-methylguanine, which are recognized and cleaved by alkyl purine DNA-glycosylase; (ii) sodium bisulfite, which causes deamination of cytosine residues in DNA to form uracil residues that can be cleaved by uracil N-glycosylase; and (iii) a chemical agent that converts guanine to its oxidized form, 8-hydroxyguanine, which can be cleaved by formamidopyrimidine DNA N-glycosylase. Examples of chemical cleavage processes include without limitation alkylation, (e.g., alkylation of phosphorothioate-modified nucleic acid); cleavage of acid lability of P3'-N5'-phosphoroamidate-containing nucleic acid; and osmium tetroxide and piperidine treatment of nucleic acid.

As used herein, the term “complementary cleavage reactions” refers to cleavage reactions that are carried out on the same nucleic acid using different cleavage reagents or by altering the cleavage specificity of the same cleavage reagent such that alternate cleavage patterns of the same target or reference nucleic acid or protein are generated. In certain embodiments, nucleic acids of interest may be treated with one or more specific cleavage agents (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more specific cleavage agents) in one or more reaction vessels (e.g., nucleic acid of interest is treated with each specific cleavage agent in a separate vessel).

A nucleic acid suitable for use in the embodiments described herein sometimes is amplified by any amplification process known in the art (e.g., PCR, RT-PCR and the like). Nucleic acid amplification may be particularly beneficial when using organisms that are typically difficult to culture (e.g., slow growing, require specialize culture conditions and the like). The terms “amplify”, “amplification”, “amplification reaction”, or “amplifying” as used herein refer to any in vitro processes for multiplying the copies of a target sequence of nucleic acid. Amplification sometimes refers to an “exponential” increase in target nucleic acid. However, “amplifying” as used herein can also refer to linear increases in the numbers of a select target sequence of

nucleic acid, but is different than a one-time, single primer extension step. In some embodiments, a limited amplification reaction, also known as pre-amplification, can be performed. Pre-amplification is a method in which a limited amount of amplification occurs due to a small number of cycles, for example 10 cycles, being performed. Pre-amplification can allow some amplification, but stops amplification prior to the exponential phase, and typically produces about 500 copies of the desired nucleotide sequence(s). Use of pre-amplification may also limit inaccuracies associated with depleted reactants in standard PCR reactions.

In some embodiments, a nucleic acid reagent sometimes is stably integrated into the chromosome of a host cell or organism, or a nucleic acid reagent can be a deletion of a portion of a host chromosome, in certain embodiments (e.g., genetically modified cells or organisms, where alteration of the host genome confers the ability to selectively or preferentially maintain the desired cell or organism carrying the genetic modification). Such nucleic acid reagents (e.g., nucleic acids or genetically modified cells or organisms whose altered genome confers a selectable trait to the cell or organism) can be selected for their ability to guide production of a desired protein or nucleic acid molecule. When desired, the nucleic acid reagent can be altered such that codons encode for (i) the same amino acid, using a different tRNA than that specified in the native sequence, or (ii) a different amino acid than is normal, including unconventional or unnatural amino acids (including detectably labeled amino acids).

A nucleic acid or nucleic acid reagent can comprise certain elements often selected according to the intended use of the nucleic acid. Any of the following elements can be included in or excluded from a nucleic acid reagent. A nucleic acid reagent, for example, may include one or more or all of the following nucleotide elements: one or more promoters, one or more 5' untranslated regions (5'UTRs), one or more regions into which a target nucleotide sequence may be inserted (an "insertion element"), one or more target nucleotide sequences, one or more terminator elements, one or more 3' untranslated regions (3'UTRs), and one or more selection elements. A nucleic acid reagent can be provided with one or more of such elements and other elements may be inserted into the nucleic acid before the nucleic acid is introduced into the desired cell or organism. In some embodiments, a provided nucleic acid reagent comprises a promoter, 5'UTR, optional 3'UTR and insertion element(s) by which a target nucleotide sequence is inserted (i.e., cloned) into the nucleotide acid reagent. In certain embodiments, a provided nucleic acid reagent comprises a promoter, insertion element(s) and optional 3'UTR, and a 5' UTR/target nucleotide sequence is inserted with an optional 3'UTR. The elements can

be arranged in any order suitable for expression in the chosen expression system (e.g., expression in a chosen cell or organism, or expression in a cell free system, for example), and in some embodiments a nucleic acid reagent comprises the following elements in the 5' to 3' direction: (1) promoter, 5'UTR, and insertion element(s); (2) promoter, 5'UTR, and target nucleotide sequence; (3) promoter, 5'UTR, insertion element(s) and 3'UTR; and (4) promoter, 5'UTR, target nucleotide sequence and 3'UTR.

Promoters

A promoter typically is required for cellular DNA synthesis and/or RNA synthesis. A promoter often contains a region of DNA that can facilitate the transcription of a particular gene, by providing a start site for the synthesis of RNA corresponding to a gene. Promoters generally are located near the genes they regulate, are located upstream (i.e., 5') of the START codon of the structural gene, and are on the same strand of DNA as the sense strand of the gene, in some instances. Eukaryotic promoters generally include a core promoter element that may contain a TATA box, a proximal sequence and transcription enhancer sequences positioned farther upstream (referred to, e.g., with respect to yeast, as upstream activating sequences or UAS located several hundred to thousands of kilobases upstream from a transcriptional start site (TSS)). The types and combination of these elements can influence promoter strength (see, e.g., Hussain et al. (2016) *ACS Synth. Biol.* 5:213-223). As used herein, "promoter," "promoter sequence" and "promoter region" are used interchangeably to refer to nucleotide sequences that can regulate gene transcription. Such sequences can include, but are not limited to, core promoter (e.g., extending upstream from the transcription START site (TSS)) elements (e.g., TATA box, RNA polymerase binding site, CCAAT box), proximal cis-acting sequences that bind proteins and can facilitate binding of RNA polymerase to DNA, and distant cis-regulatory sequences (e.g, enhancers and silencers) that can bind transcription factors and influence (e.g., activate, increase, elevate, decrease, reduce) transcription.

In some embodiments, a promoter sequence can be isolated from a nucleic acid or cell or organism and combined in functional connection or operable linkage with a polynucleotide sequence to allow altered and/or regulated expression. A non-native promoter (e.g., promoter not normally associated with a given nucleic acid sequence) used for expression of a nucleic acid often is referred to as a heterologous promoter. In certain embodiments, a heterologous promoter and/or a 5'UTR can be combined in functional connection with a polynucleotide that

encodes a polypeptide having a desired activity as described herein. The terms "operably linked" and "in functional connection with" as used herein with respect to promoters, refer to a relationship between a nucleic acid coding sequence and a promoter element. The promoter is operably linked or in functional connection with the coding sequence when expression from the coding sequence via transcription is regulated, or controlled by, the promoter element. The terms "operably linked" and "in functional connection with" are utilized interchangeably herein with respect to promoter elements.

A promoter often interacts with a RNA polymerase. A polymerase is an enzyme that catalyzes synthesis of nucleic acids using a preexisting nucleic acid reagent. When the template is a DNA template, an RNA molecule is transcribed before protein is synthesized. Enzymes having polymerase activity suitable for use in the present methods include any polymerase that is active in the chosen system with the chosen template to synthesize protein. In some embodiments, a promoter (e.g., a heterologous promoter), can be operably linked to a nucleotide sequence or an open reading frame (ORF). Transcription from the promoter element can catalyze the synthesis of an RNA corresponding to the nucleotide sequence or ORF sequence operably linked to the promoter, which in turn leads to synthesis of a desired peptide, polypeptide or protein.

There are generally several types of promoters, e.g., constitutive, repressible and inducible. Constitutive promoters can be considered as unregulated (i.e., regulated essentially only by RNA polymerase levels) and provide for consistent expression of a gene that is under the transcriptional control of the promoter. Repressible and inducible promoters are regulatable by various cellular conditions. A repressible promoter is one that can be silenced, or "turned off," by the binding of a repressor molecule to a particular nucleotide sequence which serves to inhibit the functional interaction of RNA polymerase with the promoter and inhibits transcription. This is referred to as negative control or regulation and is in contrast to positive control of transcription which can occur via activator molecules binding to DNA and increasing the rate of transcription. An inducible promoter is one in which transcription can be induced in the presence of an effector molecule that, for example, binds to a regulatory transcription factor and results in increased rates of transcription. As used herein, a "non-inducible" promoter is a promoter that does not exhibit increased activity, in terms of transcription activation of an operably linked nucleic acid, in response to the presence of an effector or inducing agent. A non-inducible promoter can be one that is not induced by one agent but is induced by another.

For example, the transcription-regulating activity of a non-fatty acid-inducible promoter is not detectably increased in the presence of a fatty acid, although there may be other agents that do induce the promoter and increase transcription of a nucleic acid operably linked to the promoter.

Promoters sometimes exhibit responsiveness to regulatory control. Promoters also sometimes can be regulated by a selective agent. That is, transcription from promoters sometimes can be turned on, turned off, up-regulated or down-regulated, in response to a change in environmental, nutritional or internal conditions or signals (e.g., heat inducible promoters, light regulated promoters, feedback regulated promoters, hormone influenced promoters, tissue specific promoters, oxygen and pH influenced promoters, promoters that are responsive to selective agents (e.g., kanamycin) and promoters responsive to certain carbon sources (e.g., fatty acids) and the like, for example). Promoters influenced by environmental, nutritional or internal signals frequently are influenced by a signal (direct or indirect) that binds at or near the promoter and increases or decreases expression of the target sequence under certain conditions.

The strength of a promoter sequence can be measured as the amount of transcription of a gene product initiated at the promoter relative to a reference or control. For example, a reference or control can be the amount of transcription of the same gene product (e.g., a reporter gene product) initiated from a standard or reference promoter under the same conditions. In assessing the strength of an inducible promoter, the amount of transcription of a gene product that occurs from the promoter in the absence (non-inducing conditions) and presence (inducing conditions) of an inducing factor, or environment or condition, can be compared to determine the degree of inducibility. The difference in those transcription amounts can also be compared to the difference in transcription amounts under the same non-inducing and inducing conditions of a reference or control promoter to determine relative strength and inducibility. Methods for evaluating promoter strength using quantitative techniques for measuring gene product expression include, for example, RT-qPCR, northern blot techniques, and reporter gene product expression assays (see, e.g., Teste et al. (2009) *BMC Molecular Biology* 10:99; Wang et al. (2016) *Yeast* 33:99-106; Peng et al. (2015) Microb. Cell Fact. 14:91). For example, transcription (e.g., measured in ways known in the art) can sometimes be increased by at least about the following percentages when an inducible promoter controlling transcription of a nucleic acid is subjected to inducing conditions as compared to transcription under non-inducing conditions: by at least about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%,

14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100% or more. In some instances, transcription can sometimes be increased by at least about the following fold when an inducible promoter controlling transcription of a nucleic acid is subjected to inducing conditions as compared to transcription under non-inducing conditions: at least about 1.5-fold, 2-fold, 2.5-fold, 3-fold, 3.5-fold, 4-fold, 4.5-fold, 5-fold, 5.5-fold, 6-fold, 6.5-fold, 7-fold, 7.5-fold, 8-fold, 8.5-fold, 9-fold, 9.5-fold,10-fold, 11-fold, 12-fold, 13-fold, 14-fold, 15-fold, 20-fold or more.

In some instances, the use of a stronger heterologous inducible promoter to control transcription can increase the amount of induced transcription of a product-encoding nucleic acid by at least about the following percentages over the amount of induced transcription of the same nucleic acid controlled by a weaker inducible promoter: by at least about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100% or more. In some instances, the use of a stronger heterologous inducible promoter to control transcription can increase the amount of induced transcription of a product-encoding nucleic acid by at least about the following fold over the amount of induced transcription of the same nucleic acid controlled by a weaker inducible promoter: by at least about 1.5-fold, 2-fold, 2.5-fold, 3-fold, 3.5-fold, 4-fold, 4.5-fold, 5-fold, 5.5-fold, 6-fold, 6.5-fold, 7-fold, 7.5-fold, 8-fold, 8.5-fold, 9-fold, 9.5-fold,10-fold, 11-fold, 12-fold, 13-fold, 14-fold, 15-fold, 20-fold or more.

In some embodiments, regulation of a promoter can be used to alter (e.g., increase, add, decrease or substantially eliminate) the activity of a peptide, polypeptide or protein (e.g., enzyme activity). For example, a cell or microorganism can be engineered by genetic modification to express a nucleic acid reagent that can add a novel activity (e.g., an activity not normally found in the host cell or organism) or increase the expression of an existing activity by

increasing transcription from a homologous or heterologous promoter operably linked to a nucleotide sequence of interest (e.g., heterologous nucleotide sequence of interest), in certain embodiments. In some embodiments, a cell or microorganism can be engineered by genetic modification to express a nucleic acid reagent that can decrease expression of an activity by decreasing or substantially eliminating transcription from a homologous or heterologous promoter operably linked to a nucleotide sequence of interest, in certain embodiments. In some embodiments, an inducible heterologous promoter can be used to regulate transcription of a protein-encoding nucleic acid that is a stronger, or more strongly inducible, promoter than an endogenous inducible promoter that regulates expression of the protein-encoding nucleic acid in its endogenous state.

In some embodiments the activity can be altered using recombinant DNA and genetic techniques known to the artisan. Methods for engineering cells and microorganisms are further described herein. Also provided herein are non-limiting examples of regulated promoters, e.g., promoters that are up-regulated by oxygen, promoters that are down-regulated by oxygen, promoters that are repressed in the presence of certain carbon sources (e.g, glucose), promoters that are de-repressed under certain carbon source conditions (e.g., limited or depleted glucose and/or non-fermentable carbon sources), promoters that are induced in the presence of certain carbon sources (e.g., fatty acids), transcriptional repressors and their associated genes, DNA binding motifs as determined using the MEME sequence analysis software. Potential regulator binding motifs can be identified using the program MEME to search intergenic regions bound by regulators for overrepresented sequences. For each regulator, the sequences of intergenic regions bound with p-values less than 0.001 are extracted to use as input for motif discovery. The MEME software can be run, for example, using the following settings: a motif width ranging from 6 to 18 bases, the “zoops” distribution model, a 6th order Markov background model and a discovery limit of 20 motifs. The discovered sequence motifs can be scored for significance by two criteria: an E-value calculated by MEME and a specificity score. The motif with the best score using each metric is shown for each regulator.

Carbon source-dependent gene regulatory elements

Many cells and organisms, including, for example, many yeast species, preferentially use glucose over other carbon sources. Often, cell and organism growth is maximal in the presence

of glucose. However, some cells and organisms are able to use alternative carbon sources for the production of metabolic energy and cellular biomass. In doing so, cellular metabolism can undergo substantial changes as certain pathways (such as, for example, oxidative metabolism, the TCA cycle, glyoxylate cycle and gluconeogenesis) required for utilizing non-glucose carbon sources are activated. Genes encoding such pathway-specific components can be subject to carbon source regulation of transcription. When glucose is present as a carbon source, some components (e.g. enzymes) of these other pathways may not be expressed, or are less expressed, because the pathways are not essential, or are used to a lesser extent, in the presence of glucose. This is referred to as glucose repression. Thus, in contrast to unregulated constitutive promoters, transcription regulatory elements for genes such as these are repressed, derepressible and/or inducible by varying carbon sources. When glucose is depleted, genes that were subject to glucose repression are then transcribed in a process referred to as glucose derepression. For some of these genes, this increase in transcription due to derepression represents the extent to which the genes will be expressed because they are not subject to induction and further increased transcription. For others of these genes, transcription may be increased (e.g., several-fold) over the derepressed level upon induction by, for example, certain carbon sources. Examples of such carbon sources include fatty acids (e.g, oleic acid) and n-alkanes. Some genes encoding peroxisomal proteins (including enzymes involved in fatty acid catabolism) are subject to glucose repression/derepression. *Cis*-acting regulatory elements have been identified for some of these genes. For example, sequences located upstream of the TATA boxes in the *Saccharomyces cerevisiae FOX1* gene encoding an acyl-CoA oxidase and *FOX3* gene encoding 3-oxoacyl-CoA thiolase have been reported as glucose response elements (see, e.g., Wang et al. (1992) *Nucleic Acids Res*. 20:3495; Wang et al. (1994) *J. Biol. Chem*. 269:24480; and Einerhand et al. (1991) *Eur. J. Biochem*. 200:113).

As described herein, engineered alteration of carbon flux in cells and organisms can involve directing internalized carbon sources toward particular cellular processing pathways and/or away from particular pathways. Some host cell modifications made in engineered bioproduction systems described herein can depend in part on the carbon source or sources used and the target molecule being produced. For example, in some embodiments provided herein, cells or organisms are modified to enhance carbon flux through oxidative metabolism pathways (e.g., β-oxidation and/or ω-oxidation) and/or fatty acid synthesis for production of organic acid, polyketide, terpene and/or other target molecules. In particular embodiments, the modified cells or organisms are provided with non-fermentable carbon sources (e.g., fatty acids, alkanes)

and/or limited amounts of, or no, glucose for production of target molecule production. In these and other embodiments described herein, genetic modifications may be made to the cells or organisms to, for example, modify the amount and/or activity of one or more enzymes (e.g., acetyl-CoA carboxylase, ATP citrate lyase, carnitine acetyltransferase, acyl-CoA thioesterase, acetyl-CoA hydrolase, acetyl-CoA synthetase) in carbon-processing pathways. As also described herein, in some of these genetic modifications, it may be beneficial to use heterologous transcription-regulating nucleic acid elements that are differentially responsive to certain carbon sources for controlling expression of the enzyme(s). For example, promoters and other regulatory nucleic acid elements that are repressed when glucose is present, derepressed in glucose-limiting, or depleted, conditions and/or induced in the presence of alternative carbon sources can provide for optimized and regulatable production of target molecules, such as in embodiments involving use of non-glucose carbon sources. This is particularly useful in instances where target molecules may be toxic to cells or organisms in high levels. In this example, modified cells could initially be cultured in the presence of glucose, if desired to build up cell mass during a growth phase, and then switched to an alternative carbon source for target molecule production through engineered pathways during which time expression of modified enzymes would be derepressed and/or induced.

In some embodiments of the engineered cells and organisms provided herein, carbon flux alterations may include diversion of carbon atoms (e.g., acetyl groups) away from particular cellular pathways (e.g. the TCA cycle) to minimize carbon atom loss to those pathways at the expense of target molecule-producing pathways. In these instances, it may be beneficial to modify and/or replace promoters and other transcription regulatory elements that control expression of components (e.g., mitochondrial proteins such as carnitine acetyltransferase and carnitine transporters) of pathways not involved in target molecule production such that the components are not expressed, or are expressed at reduced levels, under glucose-limiting conditions and/or in the presence of alternative carbon sources. Heterologous transcription-regulatory nucleic acid elements suitable in achieving such control include, but are not limited to, weak, constitutive promoters and promoters that are repressed when non-glucose (or non-fermentable) carbon sources are available, derepressed when glucose is present and/or induced in the presence of glucose. Examples of such transcriptional control elements include, but are not limited to, promoter sequences regulating transcription of genes encoding phosphoglycerate kinase (PGK), glyceraldehyde-3-phosphate dehydrogenase (GPD), translation elongation factor (TEF) and glucose-6-phosphate isomerase (G6PI; also referred to

as phosphoglucose isomerase or PGI). Nucleotide sequences for promoters of the *PGK* (SEQ ID NO: 114), *GPD* (SEQ ID NO: 119), *TEF* (SEQ ID NO: 120) and *G6PI* (SEQ ID NO: 118) genes of *Candida* strain ATCC 20336 are provided herein. Additional examples of sources of PGK, GPD, TEF and G6PI gene sequences include, but are not limited to: *Saccharomyces cerevisiae PGI1* (Genbank accession no. X13977), *Aspergillus oryzae PGK* (Genbank accession no. E04898), *Yarrowia lipolytica PGK* (Genbank accession no. M91598), *Candida albicans PGK* (Genbank accession no. U25180), *Candida maltosa C-PGK1* (Genbank accession no. D12474), *Saccharomyces cerevisiae GPD* (Genbank accession no. M13807), *Cyberlindnera jadinii GAP* (Genbank accession no. FJ664342), *Yarrowia lipolytica TEF1* (Genbank accession no. AF054508), *Debaryomyces hansenii TEF1* (Genbank accession no. AM600962).

Genes that are transcriptionally regulated by carbon source availability to cells (see, e.g., Turcotte et al. (2010) *FEMS Yeast Res.* 10:2-13; Weinhandl et al. (2014) *Microbial Cell Factories* 13(5):1-17) are possible sources of carbon source-dependent heterologous promoters for use in modification of cells and organisms as described herein. For example, genes encoding glycerol kinase and glycerol-3-phosphate dehydrogenase, such as the *GUT1* and *GUT2* genes of yeast, respectively, can be repressed in the presence of fermentable carbon sources such as glucose and expressed in the presence of non-fermentable carbon sources, e.g., glycerol or ethanol. When glucose is present, repression of the *Saccharomyces cerevisiae GUT1* and *GUT2* genes occurs in cells and is mediated by the negative regulator Opi1. The promoter region of the *Saccharomyces cerevisiae GUT1* gene contains two upstream transcription activation sequences, UAS_{ADR1} and UAS_{INO}, that can be binding sites for Adr1p (a zinc finger transcription factor) and Ino2p/Ino4p (basic helix-loop-helix factors), respectively, which are responsible for about 90% of the *GUT1* gene expression in the presence of glycerol (see, e.g., Grauslund et al. (1999) *Nucleic Acids Res.* 27(22):4391-4398). Similarly, the promoter region of the *S. cerevisiae GUT2* gene contains a UAS_{HAP} upstream sequence that can be a binding site for the Hap2/3/4/5 protein complex which activates transcription of several genes with mitochondrial functions (see, e.g., Grauslund and Ronnow (2000) *Can. J. Microbiol.* 46:1096-1100). The UAS_{HAP} element is required for full expression of the *GUT2* gene in the presence of glycerol or ethanol.

Additional examples of carbon source-dependent promoters include regulatory nucleic acid sequences controlling the transcription of genes encoding some polypeptides involved in fatty

acid metabolism, peroxisomal transport/biogenesis and/or the glyoxylate cycle. Some of these genes undergo significant induction of transcription in the presence of fatty acids (e.g., oleic acid) and/or n-alkanes. This phenomenon is referred to as fatty acid or oleic acid (or oleate) induction or alkane induction. In yeast, some of the genes subject to fatty acid induction are transcriptionally controlled by promoter regions containing an oleate response element (ORE). In *Saccharomyces cerevisiae*, for example, the ORE of fatty acid-inducible gene promoters binds the positive transcription factor of zinc cluster proteins, Pip2p-Oaf1p. The promoter region of such genes typically contains a palindrome sequence of two CGG triplets with a sequence of 15-18 nucleotides between them that includes at least one half site containing a TNA triplet (where "N" represents any nucleotide) and thus has the sequence 5'-CGGNNNTNA(N_{9-12})CCG-3' (see, e.g., Gurvitz and Rottensteiner (2006) *Biochim. Biophys. Acta* 1763:1392-1402). Variants of this sequence in fatty acid-inducible *S. cerevisiae* gene promoters have been identified in connection with the *ANT1* and *PEX25* genes leading to the following sequence as being considered the minimal ORE: CGGNNNTN$^{A}/_{R}$(N_{8-12})CCG (see, e.g., Rottensteiner et al. (2003) *Eur. J. Biochem.* 270:2013-2022). Some of the promoter regions of fatty acid-inducible *S. cerevisiae* genes also include a type 1 upstream activation sequence (UAS1) having a consensus sequence of CYCCR(A/T/G)N_{4-36}(T/A/C)YGGRG that binds the Adr1 transcription factor and directly regulates some *S. cerevisiae* genes including *SPS19*, *POX1*, *CTA1*, *PEX11*, *PIP2* which encode peroxisomal proteins (e.g., peroxisomal 2,4-dienoyl-CoA reductase, or SPS19p, and Pex11p) and/or proteins involved in β-oxidation (e.g., acyl-CoA oxidase or Pox1p) or involved in the regulation of genes associated with fatty acid metabolism (see, e.g., Gurvitz et al. (2000) *Mol. Cell. Biol. Res. Commun.* 4:81-89; Gurvitz et al. (2001) *J. Biol. Chem.* 276:31825-31830; Rottensteiner et al. (2003) *J. Biol. Chem.* 278:27605-276110). Additional *S. cerevisiae* gene promoter regions containing an ORE include those controlling transcription of *MDH3* (peroxisomal malate dehydrogenase), *YCAT* (peroxisomal and mitochondrial carnitine acetyltransferase), *CRC1* (mitochondrial carnitine transporter) and *TES1* (peroxisomal thioesterase) (see, e.g., Karpichev and Small (1998) *Mol. Cell. Bio.* 18:6560-6570).

There are numerous DNA-binding factors and regulatory proteins involved in transcriptional regulation associated with carbon source utilization. For example, in yeast such as *S. cerevisiae*, glucose repression is mediated by repressors such as, for example, members of the Mig family of C_2H_2-zinc-finger DNA-binding proteins, and some zinc cluster proteins, e.g.,Oaf3. The promoter regions of genes subject to glucose repression typically contain a GC-rich recognition site (e.g., SYGGGG) to which a repressor, e.g., Mig1, binds in the presence of high

levels of glucose (see, e.g., Gancedo (1998) *Microbiol. Mol. Biol. Rev.* 62(2):334-361). The repressor recruits a repressor complex, e.g., Ssn6-Tup1, resulting in conformational changes in the chromatin structure that prevent transcription initiation factors (e.g., the Sip4 and Adr1 activators of genes encoding, for example, gluconeogenic and glycerol or ethanol utilization proteins) from binding to sites in the DNA. Derepression occurring when glucose is depleted can result in activation of a protein kinase, Snf1, which participates in phosphorylation and release of the repressor complex thereby allowing for the activator to bind DNA in the promoter region. A shift from glucose-repressing to derepressing conditions typically results in an increased binding of Oaf1-Pip2 to oleate-response elements in promoters of fatty acid-inducible genes; however, in the presence of inducer (e.g., oleic acid), there may be only a marginal increase in this binding. This is because under derepressed conditions, Oaf1-Pip2 may be constitutively bound to target gene promoters. Activation of Oaf1-Pip2 involves binding of oleate to Oaf1 which is hyper-phosphorylated in the presence of oleate. In the activation of some fatty acid-inducible genes, Adr1 may also be involved. For example, promoter regions of genes encoding peroxisomal proteins often include a UAS1 that binds Adr1. DNA motifs for regulator protein (e.g., Adr1p, Hap2, Mig1) binding in carbon source-dependent promoters have been identified (see e.g., Weinhandl et al. (2014) *Microbial Cell Factories* 13(5):1-17) as have entire carbon source-dependent promoter nucleic acid sequences.

In other fungi, there can be different DNA-binding factors and regulatory proteins involved in transcriptional regulation associated with carbon source utilization. For example, in the filamentous fungus *Aspergillus nidulans*, glucose repression is mediated by the CreA repressor. In order to grow on fatty acids as a sole carbon source, this fungus typically requires two Zn_2Cys_6 proteins, FarAp and FarBp. These proteins are transcription factors that bind to a CCTCGG motif contained in the promoter region of genes encoding proteins involved in β-oxidation, peroxisomal functions and the glyoxylate cycle in this fungus. Specifically, FarAp is required for oxidation of short- and long-chain fatty acids and FarBp is required for oxidation of short-chain fatty acids. A homolog of FarA/FarB in *Candida albicans*, referred to as Ctf1p, is typically required for growth of *C. albicans* on fatty acids and regulates expression of some of the genes encoding proteins involved in β-oxidation.

Fatty acid- and/or alkane-inducible promoters from other organisms include, but are not limited to, those regulating transcription of the following genes: *Yarrowia lipolytica POX2* (acyl-CoA oxidase; Genbank accession no. AJ001300), *Yarrowia lipolytica POT1* (3-oxo-acyl-CoA

thiolase; Genbank accession no. X69988), *Yarrowia lipolytica ICL1* (isocitrate lyase; Genbank accession no. CQ771439) and *Candida tropicalis HDE* (hydratase-dehydrogenase-epimerase; Genbank accession no. X57854), *Candida tropicalis POX4* (acyl-CoA oxidase; Genbank accession no. AB031271), *Candida tropicalis POX18* (peroxisomal 18-kDa protein; Genbank accession no. X53633), *Candida tropicalis SPS19* (2,4-dienoyl-CoA reductase; Genbank accession no. XM_002545237), *Candida albicans PEX11* (peroxisomal protein; Genbank accession no. XM_707009), *Candida tropicalis P450alk* (alkane-inducible cytochrome P450; Genbank accession no. M24894), and *Candida tropicalis CATL* (catalase; Genbank accession no. AB181391) (see, e.g., Hussain et al. (2016) *ACS Synth. Biol*. 5:213-223 and Sloots et al. (1991) *Gene* 105:129-134). Sequences of promoter elements of fatty acid-inducible genes (e.g., *HDE*, *POX4*, *PEX11*) from *Candida* strain ATCC 20336 are also provided herein (SEQ ID NOS: 113, 117 and 121).

The promoter region controlling transcription of the *Candida tropicalis* peroxisomal *HDE* gene includes a sequence similar to, but with deviations from, the *S. cerevisiae* ORE consensus sequence, and is as follows: $CGGNNNTTAN_{12}CAG$. This sequence, located in a region between nucleotides -393 and -341 (relative to the A nucleotide of the translation START codon), contains a 3' triplet of CAG in contrast to the CCG 3' triplet of the *S. cerevisiae* ORE consensus sequence. Specific nucleotides of the *C. tropicalis HDE* gene promoter ORE are CGGTTATTACGCCTGGGGGGGCAG. Similar sequences occur in the upstream promoter regions of *C. tropicalis* genes *POX4*, *POX18*, *P450alk* and *CATL* (see Sloots et al. (1991) *Gene* 105:129-134). The promoter regions for these genes (and the *HDE* gene) can also contain sequences similar to a 7-nucleotide consensus sequence ($ATTTCC_TC_T$) for regulation of the *S. cerevisiae SUC2* gene by glucose. This glucose-responsive region of the *C. tropicalis* HDE gene is located between nucleotides -526 and -393.

Alkane-assimilating organisms, such as, for example, *Candida tropicalis*, *Candida maltosa*, *Candida albicans*, *Candida bombicola*, *Candida parapsilosis*, *Yarrowia lipolytica*, *Pichia stipitis* and *Debaryomyces hansenii*, can utilize alkanes by first converting them to fatty alcohols through oxidation catalyzed by cytochrome P450. The fatty alcohols are then oxidized to fatty aldehydes which are in turn oxidized to fatty acids. Promoters for some of the genes in alkane-assimilating yeast have been shown to contain alkane-responsive elements. For example, an upstream activating sequence referred to as ARE1 and having a sequence $CTTGTGN_xCATGTG$ (where N represents any nucleotide and x refers to the number of

nucleotides) has been identified as an alkane-responsive element present in the promoter of the *Yarrowia lipolytica ALK1* gene (cytochrome P450; Genbank accession no. AB010388) (see, e.g., Sumita et al (2002) *Biochem. Biophys. Res. Commun.* 294:1071-1078). Similar ARE1-like sequences (and/or conserved repeating motif: TGTG, or the CACA complement) occur in promoters of other genes encoding enzymes involved in alkane degradation, including, for example, cytochrome P450 genes of *Candida tropicalis*, (see, e.g., Seghezzi et al. (1992) *DNA Cell Biol.* 11:767-780), *Candida maltosa* (Genbank accession no. X55881), *Debaryomces hansenii* (Genbank accession no. AF103948) and also thiolase genes such as the acetoacetyl-CoA thiolase encoded by the *Y. lipolytica PAT1* gene (Genbank accession no. AB1020846) and the peroxisomal 3-ketoacyl-CoA thiolase encoded by the *C. tropicalis CT-T3A* gene (Genbank accession no. AB025647).

Possible additional candidate fatty acid- and/or alkane-inducible promoter sequences may be identified by searching genome databases for ORE consensus sequences located within about 500-1000 bp upstream of the START codon of an ORF and operably linking identified sequences with a reporter protein-encoding nucleic acid sequence for introduction into a host cell and analysis of reporter protein expression in the presence of varying carbon sources (such as fermentable and non-fermentable carbon sources and, in particular, fatty acids). Induced reporter protein expression in the presence of fatty acids and/or alkanes is indicative of a regulable, fatty acid- and/or alkane-inducible promoter sequence being linked to the reporter protein-encoding nucleic acid. Computer-assisted bioinformatics search programs are also available for use in identifying candidate transcription regulatory elements for genes (see, e.g., Worldwide Web uniform Resource Locator (URL) yeastract.com; Worldwide Web uniform Resource Locator (URL) pepper.molgenrug.nl/; Worldwide Web uniform Resource Locator (URL) rulai.cshl.edu/SCPD/; Worldwide Web uniform Resource Locator (URL) bimas.cit.nih.gov/molbio/proscan/; Worldwide Web uniform Resource Locator (URL) bioit.dmbr.ugent.be/contrav2/index.php).

Untranslated regions (UTR)

Nucleic acid reagents may also contain one or more 5' UTRs, and one or more 3'UTRs. Untranslated regions of a gene are sequences that are transcribed but are not translated into protein. A 5' UTR generally extends from the transcription start initiation site up to the first nucleotide of the translation START codon. A 3' UTR generally extends from the translation

STOP codon to the polyA tail. Untranslated sequences can play important roles in post-transcriptional gene expression, including, for example, transport of a transcript out of the nucleus, translation efficiency, subcellular localization and mRNA stability.

A 5' UTR used in a nucleic acid reagent in genetically modifying cells may include one or more elements that are associated with it in an endogenous state, e.g., in a cell from which it originates, and sometimes includes one or more exogenous elements. A 5' UTR can originate from any suitable nucleic acid, such as genomic DNA, plasmid DNA, RNA or mRNA, for example, from any suitable organism (e.g., virus, bacterium, yeast, fungi, plant, insect or mammal). Appropriate elements for the 5' UTR can be selected based upon the chosen expression system (e.g., expression in a chosen organism, or expression in a cell free system, for example). A 5' UTR sometimes contains one or more of the following elements: enhancer sequences (e.g., translational), translation regulation site, translation initiation site, translation factor binding site, accessory protein binding site, feedback regulation agent binding sites, ribosome binding site, replicon, internal ribosome entry site (IRES), silencer element and the like. In some embodiments, a promoter element may be isolated such that all 5' UTR elements necessary for proper conditional regulation are contained in the promoter element fragment, or within a functional subsequence of a promoter element fragment.

A 5 'UTR in the nucleic acid reagent can include a translational enhancer nucleotide sequence. A translational enhancer nucleotide sequence often is located between the promoter and the target nucleotide sequence in a nucleic acid reagent. A translational enhancer sequence often binds to a ribosome, sometimes is an 18S rRNA-binding ribonucleotide sequence (i.e., a 40S ribosome binding sequence) and sometimes is an internal ribosome entry sequence (IRES). An IRES generally forms an RNA scaffold with precisely placed RNA tertiary structures that contact a 40S ribosomal subunit via a number of specific intermolecular interactions. Examples of ribosomal enhancer sequences are known and can be identified by the artisan (e.g., Mignone et al. (2005) *Nucleic Acids Research 33*: D141-D146; Paulous et al. (2003) *Nucleic Acids Research 31*:722-733; Akbergenov et al. (2004) *Nucleic Acids Research 32*:239-247; Mignone et al. (2002) *Genome Biology 3*(3): reviews0004.1-0001.10; Gallie (2002) *Nucleic Acids Research 30*:3401-3411; Shaloiko et al., Worldwide Web uniform Resource Locator (URL) interscience.wiley.com, DOI: 10.1002/bit.20267; and Gallie et al. (1987) *Nucleic Acids Research 15*:3257-3273).

A translational enhancer sequence sometimes is a eukaryotic sequence, such as a Kozak consensus sequence or other sequence (e.g., hydroid polyp sequence, GenBank accession no. U07128). A translational enhancer sequence sometimes is a prokaryotic sequence, such as a Shine-Dalgarno consensus sequence. In certain embodiments, the translational enhancer sequence is a viral nucleotide sequence. A translational enhancer sequence sometimes is from a 5' UTR of a plant virus, such as Tobacco Mosaic Virus (TMV), Alfalfa Mosaic Virus (AMV); Tobacco Etch Virus (ETV); Potato Virus Y (PVY); Turnip Mosaic (poty) Virus and Pea Seed Borne Mosaic Virus, for example. In certain embodiments, an omega sequence about 67 bases in length from TMV is included in the nucleic acid reagent as a translational enhancer sequence (e.g., devoid of guanosine nucleotides and includes a 25 nucleotide long poly (CAA) central region).

A 3' UTR used in a nucleic acid reagent in genetically modifying cells may include one or more elements that are associated with it in an endogenous state, e.g., in a cell from which it originates, and sometimes includes one or more exogenous elements. A 3' UTR may originate from any suitable nucleic acid, such as genomic DNA, plasmid DNA, RNA or mRNA, for example, from any suitable organism (e.g., a virus, bacterium, yeast, fungi, plant, insect or mammal). Appropriate elements for the 3' UTR can be selected based upon the chosen expression system (e.g., expression in a chosen organism, for example). A 3' UTR sometimes comprises one or more of the following elements: translation regulation site, translation termination site, translation initiation site, translation factor binding site, ribosome binding site, replicon, enhancer element, silencer element and polyadenosine tail. A 3' UTR often includes a polyadenosine tail and sometimes does not, and if a polyadenosine tail is present, one or more adenosine moieties may be added or deleted from it (e.g., about 5, about 10, about 15, about 20, about 25, about 30, about 35, about 40, about 45 or about 50 adenosine moieties may be added or subtracted).

In some embodiments, modification of a 5' UTR and/or a 3' UTR can be used to alter (e.g., increase, add, decrease or substantially eliminate) gene expression activity. This can in turn alter the activity of a peptide, polypeptide or protein (e.g., enzyme activity for example), by a change in transcription of the nucleotide sequence(s) of interest from an operably linked promoter element comprising the modified 5' or 3' UTR. For example, a microorganism can be engineered by genetic modification to express a nucleic acid reagent comprising a modified 5' or 3' UTR that can add a novel activity (e.g., an activity not normally found in the host organism)

or increase the expression of an existing activity by increasing transcription from a homologous or heterologous promoter operably linked to a nucleotide sequence of interest (e.g., homologous or heterologous nucleotide sequence of interest), in certain embodiments. In some embodiments, a microorganism can be engineered by genetic modification to express a nucleic acid reagent comprising a modified 5' or 3' UTR that can decrease the expression of an activity by decreasing or substantially eliminating transcription from a homologous or heterologous promoter operably linked to a nucleotide sequence of interest, in certain embodiments.

Nucleic acid or protein similarity

In addition to the nucleotide and amino acid sequences provided herein, a polynucleotide or polypeptide sequence may also be one that is substantially similar to those provided herein, including, but not limited to, promoter sequences, regulatory sequences, coding polynucleotides, amino acid signal sequences and amino acid protein sequences provided herein. Similarity between two nucleic acids or polypeptides refers to the relatedness between nucleotide sequences or amino acid sequences. Similarity can be based on the degree of identity and/or homology of sequences and the residues contained therein. Methods of assessing the degree of similarity between nucleic acids or proteins are known to those of skill in the art. For example, in one method of assessing sequence similarity, two nucleotide or amino acid sequences are aligned in a manner that yields a maximal level of identity between the sequences. Identity refers to the extent to which the sequences are invariant. Alignment of amino acid sequences, and, to some extent, nucleotide sequences, also can take into account conservative differences and/or frequent substitutions in amino acids (or nucleotides). Conservative differences are those that conserve the physico-chemical properties of the residues involved. Alignments can be global (alignment of the compared sequences over the entire length of the sequences and including all residues) or local (alignment of a portion of the compared sequences e.g., a portion or portions that includes only the most similar region or regions). Homology, with reference to polynucleotide or polypeptide sequences, refers to nucleotide or amino acid sequence similarity that takes into account identical residues and residues that can substitute for one another.

Percent identity and/or homology may be determined, for example, by comparing sequence information using any of a number of computer algorithms known in the art. In one example, calculations of sequence identity can be performed as follows. Sequences are aligned for

optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment, and non-homologous sequences can be disregarded for comparison purposes). The length of a reference sequence aligned for comparison purposes is sometimes 30% or more, 40% or more, 50% or more, often 60% or more, and more often 70% or more, 80% or more, 90% or more, or 100% of the length of the reference sequence. The nucleotides or amino acids at corresponding nucleotide or polypeptide positions, respectively, are then compared among the two sequences. When a position in the first sequence is occupied by the same nucleotide or amino acid as the corresponding position in the second sequence, the nucleotides or amino acids are deemed to be identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, introduced for optimal alignment of the two sequences. Examples of sequence alignment and analysis software that can be used to calculate sequence identity include BLAST (Worldwide Web uniform Resource Locator (URL) blast.ncbi.nlm.nih.gov/Blast.cgi), MUSCLE (Worldwide Web uniform Resource Locator (URL) ebi.ac.uk/Tools/msa/muscle/ and Worldwide Web uniform Resource Locator (URL) drive5.com/muscle/) and MAFFT (Worldwide Web uniform Resource Locator (URL) mafft.cbrc.jp/alignment/server/ and Worldwide Web uniform Resource Locator (URL) ebi.ac.uk/Tools/msa/mafft/) for comparing nucleotide sequences and SIM (Worldwide Web uniform Resource Locator (URL) web.expasy.org/sim/) and BLAST for comparison of amino acid sequences. Nucleic acid sequence identity can also be determined by hybridization assays conducted under stringent conditions. As used herein, the term “stringent conditions” refers to conditions for hybridization and washing. Stringent conditions are known to those skilled in the art and can be found, for example, in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y., 6.3.1-6.3.6 (1989). Aqueous and non-aqueous methods are described in that reference and either can be used. An example of stringent hybridization conditions is hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45ºC, followed by one or more washes in 0.2X SSC, 0.1% SDS at 50ºC. Another example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45ºC, followed by one or more washes in 0.2X SSC, 0.1% SDS at 55ºC. A further example of stringent hybridization conditions is hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45ºC, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60ºC. Often, stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45ºC, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65ºC. More often, stringency

conditions are 0.5M sodium phosphate, 7% SDS at 65ºC, followed by one or more washes at 0.2X SSC, 1% SDS at 65ºC.

A nucleic acid or polypeptide for use in developing cells and organisms and/or methods described herein may be, for example, a polynucleotide or amino acid sequence that is homologous or identical to a nucleotide sequence (or complement thereof) or amino acid sequence provided herein over at least about 75%, or at least about 77%, or at least about 78%, or at least about 79%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95% or more of the provided reference sequence. For example, a polynucleotide or polypeptide can be one that is at least about 50%, or at least about 51%, or at least about 52%, or at least about 54%, or at least about 55%, or at least about 58%, or at least about 60%, or at least about 62%, or at least about 65%, or at least about 70%, or at least about 75% or at least about 80% or more homologous or identical to a nucleic acid (or complement thereof) or polypeptide provided herein over the specified extent of a nucleic acid or polypeptide provided herein. In another embodiment, a nucleic acid or polypeptide can be one that is homologous or identical to a nucleic acid (or complement thereof) or polypeptide provided herein over at least about 86%, or at least about 87%, or at least about 88%, or at least about 89%, or at least about 92%, or at least about 97% or more of the provided reference nucleic acid (or complement thereof) or polypeptide. For example, the protein can be one that is at least about 67%, or at least about 68%, or at least about 69%, or at least about 72%, or at least about 77%, or at least about 82%, or at least about 87%, or at least about 90%, or at least about 95% or at least about 96%, or at least about 97%, or at least about 98%, or at least about 99% or more homologous or identical to a reference nucleic acid (or complement thereof) or polypeptide provided herein over the specified extent of the nucleic acid (or complement thereof) or polypeptide.

In some embodiments, a nucleotide or amino acid sequence that is at least 80% or more, 81% or more, 82% or more, 83% or more, 84% or more, 85% or more, 86% or more, 87% or more, 88% or more, 89% or more, 90% or more, 91% or more, 92% or more, 93% or more, 94% or more, 95% or more, 96% or more, 97% or more, 98% or more, or 99% or more identical to a nucleotide sequence (or complement thereof) or amino acid sequence described herein can be utilized. The term “identical” as used herein refers to two or more nucleotide or amino acid sequences having substantially the same nucleotide or amino acid sequence when compared to each other. One test for determining whether two nucleotide sequences or amino acids

sequences are substantially identical is to determine the percent of identical nucleotide sequences or amino acid sequences shared.

Target nucleotide sequence

A nucleic acid reagent sometimes can comprise a target nucleotide sequence. A “target nucleotide sequence” as used herein encodes a nucleic acid, peptide, polypeptide or protein of interest, and may be a ribonucleotide sequence or a deoxyribonucleotide sequence. A target nucleic acid sometimes is an untranslated ribonucleic acid and sometimes is a translatable ribonucleic acid. An untranslated ribonucleic acid may include, but is not limited to, a small interfering ribonucleic acid (siRNA), a short hairpin ribonucleic acid (shRNA), other ribonucleic acid capable of RNA interference (RNAi), an antisense ribonucleic acid, or a ribozyme. A translatable target nucleotide sequence (e.g., a target ribonucleotide sequence) sometimes encodes a peptide, polypeptide or protein, which are sometimes referred to herein as “target peptides,” “target polypeptides” or “target proteins.”

Any peptides, polypeptides or proteins, or an activity catalyzed by one or more peptides, polypeptides or proteins, may be encoded by a target nucleotide sequence and may be selected by a user. Representative proteins include enzymes (e.g., acetyl-CoA carboxylase, acyl-CoA oxidase, thioesterase, monooxygenase, monooxygenase reductase, fatty alcohol oxidase, acyltransferase and the like, for example), antibodies, serum proteins (e.g., albumin), membrane bound proteins, hormones (e.g., growth hormone, erythropoietin, insulin, and the like), cytokines, and the like, and include both naturally occurring and exogenously expressed polypeptides. Representative activities (e.g., enzymes or combinations of enzymes which are functionally associated to provide an activity) include thioesterase activity, monooxygenase activity, monooxygenase reductase activity, acetyltransferase activity, omega hydroxyl fatty acid dehydrogenase activity, beta-oxidation activity, omega-oxidation activity and the like, for example. The term “enzyme” as used herein refers to a protein which can act as a catalyst to induce a chemical change in other compounds, thereby producing one or more products from one or more substrates.

Specific polypeptides (e.g., enzymes) useful for embodiments described herein are listed herein. The term "protein" as used herein refers to a molecule having a sequence of amino acids linked by peptide bonds. This term includes fusion proteins, oligopeptides, peptides,

cyclic peptides, polypeptides and polypeptide derivatives, whether native or recombinant, and also includes fragments, derivatives, homologs, and variants thereof. A protein or polypeptide sometimes is of intracellular origin (e.g., located in the nucleus, cytosol, organelle (e.g., mitochondria or peroxisome) or interstitial space of host cells in vivo) and sometimes is a cell membrane protein in vivo. In some embodiments (described in further detail herein), a genetic modification can result in a modification (e.g., increase, substantially increase, decrease or substantially decrease) of a target activity and/or in a modification of a cellular location for a protein.

A translatable nucleotide sequence generally is located between a start codon (AUG in ribonucleic acids and ATG in deoxyribonucleic acids) and a stop codon (e.g., UAA (ochre), UAG (amber) or UGA (opal) in ribonucleic acids and TAA, TAG or TGA in deoxyribonucleic acids), and sometimes is referred to herein as an "open reading frame" (ORF). A translatable nucleotide sequence (e.g., ORF) sometimes is encoded differently in one organism (e.g., most organisms encode CTG as leucine) than in another organism (e.g., *Candida tropicalis* and *Candida viswanathii* encode CTG as serine). In some embodiments, a translatable nucleotide sequence is altered to correct alternate genetic code (e.g., codon usage) differences between a nucleotide donor organism and a nucleotide recipient organism (e.g., engineered organism). In certain embodiments, a translatable nucleotide sequence is altered to improve; (i) codon usage, (ii) transcriptional efficiency, (iii) translational efficiency, (iv) the like, and combinations thereof.

Nucleic acid reagents and tools

A nucleic acid reagent sometimes includes one or more ORFs. An ORF may be from any suitable source, sometimes from genomic DNA, mRNA, reverse transcribed RNA or complementary DNA (cDNA) or a nucleic acid library comprising one or more of the foregoing, and is from any organism species that contains a nucleic acid sequence of interest, protein of interest, or activity of interest. Non-limiting examples of organisms from which an ORF can be obtained include bacteria, yeast, fungi, plant, human, insect, nematode, bovine, equine, canine, feline, rat or mouse, for example.

A nucleic acid reagent sometimes contains a nucleotide sequence adjacent to an ORF that is translated in conjunction with the ORF and encodes an amino acid tag. The tag-encoding nucleotide sequence can be located 3' and/or 5' of an ORF in the nucleic acid reagent, thereby

encoding a tag at the C-terminus or N-terminus of the protein or peptide encoded by the ORF. Any tag that does not abrogate in vitro transcription and/or translation may be utilized and may be appropriately selected by the artisan. Tags may facilitate isolation and/or purification of the desired ORF product from culture or fermentation media.

A tag sometimes specifically binds a molecule or moiety of a solid phase or a detectable label, for example, thereby having utility for isolating, purifying and/or detecting a protein or peptide encoded by the ORF. In some embodiments, a tag includes one or more of the following elements: FLAG (e.g., DYKDDDDKG), V5 (e.g., GKPIPNPLLGLDST), c-MYC (e.g., EQKLISEEDL), HSV (e.g., QPELAPEDPED), influenza hemaglutinin, HA (e.g., YPYDVPDYA), VSV-G (e.g., YTDIEMNRLGK), bacterial glutathione-S-transferase, maltose binding protein, a streptavidin- or avidin-binding tag (e.g., pcDNA™6 BioEase™ Gateway® Biotinylation System (ThermoFisher Scientific)), thioredoxin, β-galactosidase, VSV-glycoprotein, a fluorescent protein (e.g., green fluorescent protein or one of its many color variants (e.g., yellow, red, blue)), a polylysine or polyarginine sequence, a polyhistidine sequence (e.g., His6) or other sequence that chelates a metal (e.g., cobalt, zinc, copper), and/or a cysteine-rich sequence that binds to an arsenic-containing molecule. In certain embodiments, a cysteine-rich tag comprises the amino acid sequence CC-Xn-CC, where X is any amino acid and n is 1 to 3, and the cysteine-rich sequence sometimes is CCPGCC. In certain embodiments, the tag contains a cysteine-rich element and a polyhistidine element (e.g., CCPGCC and His6).

A tag often conveniently binds to a binding partner. For example, some tags bind to an antibody (e.g., FLAG) and sometimes specifically bind to a small molecule. For example, a polyhistidine tag specifically chelates a bivalent metal, such as copper, zinc and cobalt; a polylysine or polyarginine tag specifically binds to a zinc finger; a glutathione S-transferase tag binds to glutathione; and a cysteine-rich tag specifically binds to an arsenic-containing molecule. Arsenic-containing molecules include LUMIO™ agents (ThermoFisher Scientific), such as FlAsH™ (EDT2[4',5'-bis(1,3,2-dithioarsolan-2-yl)fluorescein-(1,2-ethanedithiol)2]) and ReAsH reagents (e.g., U.S. Patent 5,932,474 to Tsien et al., entitled “Target Sequences for Synthetic Molecules;” U.S. Patent 6,054,271 to Tsien et al., entitled “Methods of Using Synthetic Molecules and Target Sequences;” U.S. Patents 6,451,569 and 6,008,378; published U.S. Patent Application 2003/0083373, and published PCT Patent Application WO 99/21013, all to Tsien et al. and all entitled “Synthetic Molecules that Specifically React with Target

Sequences"). Such antibodies and small molecules sometimes are linked to a solid phase for convenient isolation of the target protein or target peptide.

A tag sometimes includes a sequence that localizes a translated protein or peptide to a component in a system, which may be referred to as a "signal sequence," "targeting sequence" or "localization signal sequence" herein. A signal sequence often is incorporated at the N-terminus of a target protein or target peptide, and sometimes is incorporated at the C-terminus. Examples of signal sequences are known to the artisan, are readily incorporated into a nucleic acid reagent, and often are selected according to the organism in which expression of the nucleic acid reagent is performed. A signal sequence in some embodiments localizes a translated protein or peptide to a cell membrane. Examples of signal sequences include, but are not limited to, a nucleus targeting signal (e.g., steroid receptor sequence and N-terminal sequence of SV40 virus large T antigen); mitochondrial targeting signal (e.g., amino acid sequence that forms an amphipathic helix); peroxisome targeting signal (e.g., C-terminal sequence in YFG from *S.cerevisiae*); and a secretion signal (e.g., N-terminal sequences from invertase, mating factor alpha, PHO5 and SUC2 in *S.cerevisiae*; multiple N-terminal sequences of *B. subtilis* proteins (e.g., Tjalsma et al., *Microbiol.Molec. Biol. Rev*. 64: 515-547 (2000)); alpha amylase signal sequence (e.g., U.S. Patent No. 6,288,302); pectate lyase signal sequence (e.g., U.S. Patent No. 5,846,818); precollagen signal sequence (e.g., U.S. Patent No. 5,712,114); OmpA signal sequence (e.g., U.S. Patent No. 5,470,719); Iam beta signal sequence (e.g., U.S. Patent No. 5,389,529); *B. brevis* signal sequence (e.g., U.S. Patent No. 5,232,841); and *P. pastoris* signal sequence (e.g., U.S. Patent No. 5,268,273)).

A tag sometimes is directly adjacent to the amino acid sequence encoded by an ORF (i.e., there is no intervening sequence) and sometimes a tag is substantially adjacent to an ORF encoded amino acid sequence (e.g., an intervening sequence is present). An intervening sequence sometimes includes a recognition site for a protease, which is useful for cleaving a tag from a target protein or peptide. In some embodiments, the intervening sequence is cleaved by Factor Xa (e.g., recognition site I (E/D)GR), thrombin (e.g., recognition site LVPRGS), enterokinase (e.g., recognition site DDDDK), TEV protease (e.g., recognition site ENLYFQG) or PreScission™ protease (e.g., recognition site LEVLFQGP), for example.

An intervening sequence sometimes is referred to herein as a "linker sequence," and may be of any suitable length selected by the artisan. A linker sequence sometimes is about 1 to about 20

amino acids in length, and sometimes about 5 to about 10 amino acids in length. The linker length can be selected to substantially preserve target protein or peptide function (e.g., a tag may reduce target protein or peptide function unless separated by a linker), to enhance disassociation of a tag from a target protein or peptide when a protease cleavage site is present (e.g., cleavage may be enhanced when a linker is present), and to enhance interaction of a tag/target protein product with a solid phase. A linker can be of any suitable amino acid content, and often comprises a higher proportion of amino acids having relatively short side chains (e.g., glycine, alanine, serine and threonine).

A nucleic acid reagent sometimes includes a stop codon between a tag element and an insertion element or ORF, which can be useful for translating an ORF with or without the tag. Mutant tRNA molecules that recognize stop codons suppress translation termination and thereby are designated “suppressor tRNAs.” Suppressor tRNAs can result in the insertion of amino acids and continuation of translation past stop codons (e.g., U.S. Patent Application No. 60/587,583, filed July 14, 2004, entitled "Production of Fusion Proteins by Cell-Free Protein Synthesis,"; Eggertsson, et al., (1988) *Microbiological Review* 52(3):354-374, and Engleerg-Kukla, et al. (1996) in *Escherichia coli* and *Salmonella* Cellular and Molecular Biology, Chapter 60, pps 909-921, Neidhardt, et al. eds., ASM Press, Washington, DC). A number of suppressor tRNAs are known, including but not limited to, supE, supP, supD, supF and supZ suppressors, which suppress the termination of translation of the amber stop codon; supB, glT, supL, supN, supC and supM suppressors, which suppress the function of the ochre stop codon and glyT, trpT and Su-9 suppressors, which suppress the function of the opal stop codon. In general, suppressor tRNAs contain one or more mutations in the anti-codon loop of the tRNA that allows the tRNA to base pair with a codon that ordinarily functions as a stop codon. The mutant tRNA is charged with its cognate amino acid residue and the cognate amino acid residue is inserted into the translating polypeptide when the stop codon is encountered. Mutations that enhance the efficiency of termination suppressors (i.e., increase stop codon read-through) have been identified. These include, but are not limited to, mutations in the uar gene (also known as the prfA gene), mutations in the ups gene, mutations in the sueA, sueB and sueC genes, mutations in the rpsD (ramA) and rpsE (spcA) genes and mutations in the rplL gene.

Thus, a nucleic acid reagent containing a stop codon located between an ORF and a tag can yield a translated ORF alone when no suppressor tRNA is present in the translation system, and can yield a translated ORF-tag fusion when a suppressor tRNA is present in the system.

Suppressor tRNA can be generated in cells transfected with a nucleic acid encoding the tRNA (e.g., a replication incompetent adenovirus containing the human tRNA-Ser suppressor gene can be transfected into cells, or a YAC containing a yeast or bacterial tRNA suppressor gene can be transfected into yeast cells, for example). Vectors for synthesizing suppressor tRNA and for translating ORFs with or without a tag are available to the artisan (e.g., Tag-On-Demand™ kit (ThermoFisher Scientific); Tag-On-Demand™ Suppressor Supernatant Instruction Manual, Version C, 31 October 2010, World Wide Web Uniform Resource Locator (URL) tools.thermofisher.com/content/sfs/manuals/tagondemand_supernatant_man.pdf; Tag-On-Demand™ Gateway® Vector Instruction Manual, Version D, 31 October 2010 World Wide Web Uniform Resource Locator (URL) tools.thermofisher.com/content/sfs/manuals/tagondemand_vectors_man.pdf; and Capone et al. (1985) Amber, ochre and opal suppressor tRNA genes derived from a human serine tRNA gene. *EMBO J.* 4:213).

Any convenient cloning strategy known in the art may be utilized to incorporate an element, such as an ORF, into a nucleic acid reagent. Known methods can be utilized to insert an element into the template independent of an insertion element, such as (1) cleaving the template at one or more existing restriction enzyme sites and ligating an element of interest and (2) adding restriction enzyme sites to the template by hybridizing oligonucleotide primers that include one or more suitable restriction enzyme sites and amplifying by polymerase chain reaction. Other cloning strategies take advantage of one or more insertion sites present or inserted into the nucleic acid reagent, such as an oligonucleotide primer hybridization site for PCR, for example, and others described herein. In some embodiments, a cloning strategy can be combined with genetic manipulation such as recombination (e.g., recombination of a nucleic acid reagent with a nucleic acid sequence of interest into the genome of the organism that is modified, as described further herein). In some embodiments, the cloned ORF(s) can produce (directly or indirectly), for example, a fatty acid or dicarboxylic acid (e.g., adipic acid, octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid), 3-hydroxyproprionic acid, triacetic acid lactone, by engineering a cell or microorganism with one or more ORFs of interest, which cell or microorganism may include one or more altered activities such as, for example, carnitine acetyltransferase activity, acetyl-CoA carboxylase activity, ATP citrate lyase activity, acetyl-CoA synthetase activity, cytochrome P450 reductase activity, acetyl-CoA hydrolase activity, 6-oxohexanoic acid dehydrogenase activity, 6-hydroxyhexanoic acid dehydrogenase activity,

glucose-6-phosphate dehydrogenase activity, hexanoate synthase activity, lipase activity, fatty acid synthase activity, omega hydroxyl fatty acid dehydrogenase activity, acyl-CoA oxidase activity, acyltransferase activity, thioesterase activity, monooxygenase activity and monooxygenase reductase activity.

In some embodiments, a nucleic acid reagent includes one or more recombinase insertion sites. A recombinase insertion site is a recognition sequence on a nucleic acid molecule that participates in an integration/recombination reaction by recombination proteins. For example, the recombination site for Cre recombinase is loxP, which is a 34 base pair sequence comprised of two 13 base pair inverted repeats (serving as the recombinase binding sites) flanking an 8 base pair core sequence (e.g., Figure 1 of Sauer, B., *Curr. Opin. Biotech. 5*:521-527 (1994)). Other examples of recombination sites include attB, attP, attL, and attR sequences, and mutants, fragments, variants and derivatives thereof, which are recognized by the recombination protein λ Int and by the auxiliary proteins integration host factor (IHF), FIS and excisionase (Xis) (e.g., U.S. Patent Nos. 5,888,732; 6,143,557; 6,171,861; 6,270,969; 6,277,608; and 6,720,140; U.S. Patent Appln. Nos. 09/517,466, filed March 2, 2000, and 09/732,914, filed August 14, 2003, and in U.S. patent publication no. 2002-0007051-A1; Landy, Curr. Opin. Biotech. 3:699-707 (1993)).

Examples of recombinase cloning nucleic acids are in Gateway® systems (ThermoFisher Scientific), which include at least one recombination site for cloning a desired nucleic acid molecules in vivo or in vitro. In some embodiments, the system utilizes vectors that contain at least two different site-specific recombination sites, often based on the bacteriophage lambda system (e.g., att1 and att2), and are mutated from the wild-type (att0) sites. Each mutated site has a unique specificity for its cognate partner att site (i.e., its binding partner recombination site) of the same type (for example attB1 with attP1, or attL1 with attR1) and will not cross-react with recombination sites of the other mutant type or with the wild-type att0 site. Different site specificities allow directional cloning or linkage of desired molecules thus providing desired orientation of the cloned molecules. Nucleic acid fragments flanked by recombination sites are cloned and subcloned using the Gateway® system by replacing a selectable marker (for example, ccdB) flanked by att sites on the recipient plasmid molecule, sometimes termed the Destination Vector. Desired clones are then selected by transformation of a ccdB sensitive host strain and positive selection for a marker on the recipient molecule. Similar strategies for

negative selection (e.g., use of toxic genes) can be used in other organisms such as thymidine kinase (TK) in mammals and insects.

A recombination system useful for engineering yeast is outlined briefly. The system makes use of the *URA3* gene (e.g., for *Candida tropicalis*, *Candida viswanathii*, *Saccharomyces cerevisiae* and *Candida albicans*, for example) or *URA4* and *URA5* genes (e.g., for *S. pombe*, for example) and toxicity of the nucleotide analogue 5-Fluoroorotic acid (5-FOA). The *URA3* or *URA4* and *URA5* genes encode orotidine-5'-monophosphate (OMP) dicarboxylase. Yeast with an active *URA3* or *URA4* and *URA5* gene (phenotypically Ura+) convert 5-FOA to fluorodeoxyuridine, which is toxic to yeast cells. Yeast carrying a mutation in the appropriate gene(s) or having a knock out of the appropriate gene(s) can grow in the presence of 5-FOA, if the media is also supplemented with uracil.

A nucleic acid engineering construct can be made which may contain the *URA3* gene or cassette (for *C. tropicalis*, *C. viswanathii* or *S. cerevisiae*, for example), flanked on either side by the same nucleotide sequence in the same orientation. The *URA3* cassette can include a promoter, the *URA3* gene and a functional transcription terminator. Target sequences which direct the construct to a particular nucleic acid region of interest in the cell or organism to be engineered are added such that the target sequences are adjacent to and abut the flanking sequences on either side of the *URA3* cassette. Yeast can be transformed with the engineered construct and plated on minimal media without uracil. Colonies can be screened by PCR to determine those transformants that have the engineering construct inserted in the proper location in the genome. Checking insertion location prior to selecting for recombination of the ura3 cassette may reduce the number of incorrect clones carried through to later stages of the procedure. Correctly inserted transformants can then be grown and plated on minimal media containing 5-FOA to select for recombination of the URA3 cassette out of the construct, leaving a disrupted gene and an identifiable footprint (e.g., nucleic acid sequence) that can be used to verify the presence of the disrupted gene. The technique described is useful for disrupting or "knocking out" gene function, but also can be used to insert genes or constructs into a host cell genome in a targeted, sequence specific manner.

In certain embodiments, a nucleic acid reagent includes one or more topoisomerase insertion sites. A topoisomerase insertion site is a defined nucleotide sequence recognized and bound by a site-specific topoisomerase. For example, the nucleotide sequence 5'-(C/T)CCTT-3' is a

topoisomerase recognition site bound specifically by most poxvirus topoisomerases, including vaccinia virus DNA topoisomerase I. After binding to the recognition sequence, the topoisomerase cleaves the strand at the 3'-most thymidine of the recognition site to produce a nucleotide sequence comprising 5'-(C/T)CCTT-PO4-TOPO, a complex of the topoisomerase covalently bound to the 3' phosphate via a tyrosine in the topoisomerase (e.g., Shuman (1991) *J. Biol. Chem. 266*:11372-11379; Sekiguchi and Shuman (1994) *Nucl. Acids Res*. 22:5360-5365; U.S. Pat. No. 5,766,891; PCT/US95/16099; and PCT/US98/12372). In comparison, the nucleotide sequence 5'-GCAACTT-3' is a topoisomerase recognition site for type IA *E. coli* topoisomerase III. An element that is inserted often is combined with topoisomerase-reacted template and thereby incorporated into the nucleic acid reagent (e.g., World Wide Web Uniform Resource Locator (URL) tools.thermofisher.com/downloads/F-13512_Topo_Flyer.pdf; World Wide Web Uniform Resource Locator (URL) tools.thermofisher.com/content/sfs/brochures/topo-pcr-cloning-brochure.pdf; TOPO TA Cloning® Kit and Zero Blunt® TOPO® Cloning Kit product information).

A nucleic acid reagent sometimes contains one or more origin of replication (ORI) elements. In some embodiments, a template comprises two or more ORIs, where one functions efficiently in one organism (e.g., a bacterium) and another functions efficiently in another organism (e.g., a eukaryote, like yeast for example). In some embodiments, an ORI may function efficiently in one species (e.g., *S. cerevisiae*, for example) and another ORI may function efficiently in a different species (e.g., *S. pombe*, for example). A nucleic acid reagent also sometimes includes one or more transcription regulation sites.

A nucleic acid reagent can include one or more selection elements (e.g., elements for selection of the presence of the nucleic acid reagent, and not for activation of a promoter element which can be selectively regulated). Selection elements often are utilized using known processes to determine whether a nucleic acid reagent is included in a cell. In some embodiments, a nucleic acid reagent includes two or more selection elements, where one functions efficiently in one organism and another functions efficiently in another organism. Examples of selection elements include, but are not limited to, (1) nucleic acid segments that encode products that provide resistance against otherwise toxic compounds (e.g., antibiotics); (2) nucleic acid segments that encode products that are otherwise lacking in the recipient cell (e.g., essential products, tRNA genes, auxotrophic markers); (3) nucleic acid segments that encode products that suppress the activity of a gene product; (4) nucleic acid segments that encode products that can be readily

identified (e.g., phenotypic markers such as antibiotics (e.g., β-lactamase), β-galactosidase, green fluorescent protein (GFP), yellow fluorescent protein (YFP), red fluorescent protein (RFP), cyan fluorescent protein (CFP), and cell surface proteins); (5) nucleic acid segments that bind products that are otherwise detrimental to cell survival and/or function; (6) nucleic acid segments that otherwise inhibit the activity of any of the nucleic acid segments described in Nos. 1-5 above (e.g., antisense oligonucleotides); (7) nucleic acid segments that bind products that modify a substrate (e.g., restriction endonucleases); (8) nucleic acid segments that can be used to isolate or identify a desired molecule (e.g., specific protein binding sites); (9) nucleic acid segments that encode a specific nucleotide sequence that can be otherwise non-functional (e.g., for PCR amplification of subpopulations of molecules); (10) nucleic acid segments that, when absent, directly or indirectly confer resistance or sensitivity to particular compounds; (11) nucleic acid segments that encode products that either are toxic or convert a relatively non-toxic compound to a toxic compound (e.g., Herpes simplex thymidine kinase, cytosine deaminase) in recipient cells; (12) nucleic acid segments that inhibit replication, partition or heritability of nucleic acid molecules that contain them; and/or (13) nucleic acid segments that encode conditional replication functions, e.g., replication in certain hosts or host cell strains or under certain environmental conditions (e.g., temperature, nutritional conditions, and the like). In some embodiments, the regulatory or selective agent can be added to change the existing growth conditions to which a cell or organism is subjected (e.g., growth in liquid culture, growth in a fermenter, growth on solid nutrient plates and the like for example).

A nucleic acid reagent can sometimes be of any form useful for in vivo transcription and/or translation. A nucleic acid sometimes is a plasmid, such as a supercoiled plasmid, sometimes is a yeast artificial chromosome (e.g., YAC), sometimes is a linear nucleic acid (e.g., a linear nucleic acid produced by PCR or by restriction digest), sometimes is single-stranded and sometimes is double-stranded. A nucleic acid reagent sometimes is prepared by an amplification process, such as a polymerase chain reaction (PCR) process or transcription-mediated amplification process (TMA). In TMA, two enzymes are used in an isothermal reaction to produce amplification products detected by light emission (see, e.g., *Biochemistry* 1996 Jun 25;35(25):8429-38). Standard PCR processes are known (e.g., U. S. Patent Nos. 4,683,202; 4,683,195; 4,965,188; and 5,656,493), and generally are performed in cycles. Each cycle includes heat denaturation, in which hybrid nucleic acids dissociate; cooling, in which primer oligonucleotides hybridize; and extension of the oligonucleotides by a polymerase (i.e., Taq polymerase). An example of a PCR cyclical process is treating the sample at 95°C for 5

minutes; repeating forty-five cycles of 95°C for 1 minute, 59°C for 1 minute, 10 seconds, and 72°C for 1 minute 30 seconds; and then treating the sample at 72°C for 5 minutes. Multiple cycles frequently are performed using a commercially available thermal cycler. PCR amplification products sometimes are stored for a time at a lower temperature (e.g., at 4°C) and sometimes are frozen (e.g., at –20°C) before analysis.

In some embodiments, a nucleic acid reagent, protein reagent, protein fragment reagent or other reagent described herein is isolated or purified. The term “isolated” as used herein refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring, or a host cell if expressed exogenously), and thus is altered "by the hand of man" from its original environment. The term “purified” as used herein with reference to molecules does not refer to absolute purity. Rather, “purified” refers to a substance in a composition that contains fewer substance species in the same class (e.g., nucleic acid or protein species) other than the substance of interest in comparison to the sample from which it originated. “Purified,” if a nucleic acid or protein for example, refers to a substance in a composition that contains fewer nucleic acid species or protein species other than the nucleic acid or protein of interest in comparison to the sample from which it originated. Sometimes, a protein or nucleic acid is “substantially pure,” indicating that the protein or nucleic acid represents at least 50% of protein or nucleic acid on a mass basis of the composition. Often, a substantially pure protein or nucleic acid is at least 75% on a mass basis of the composition, and sometimes at least 95% on a mass basis of the composition.

Genetic engineering methods

Methods and compositions (e.g., nucleic acid reagents) described herein can be used to generate modified or engineered cells or organisms. For example, a cell or organism can be modified by altering one or more cellular activities and/or the sum total of a cell’s or organism’s activities. Modifications can be, for example, any alteration of cellular activities, including addition of cellular activities and/or elimination of cellular activities. The term "altered activity" as used herein refers to an activity in an engineered cell or microorganism that is added, removed or modified in any way relative to the host cell or microorganism (e.g., added, increased, reduced, decreased, inhibited, removed or redirected activity). In some embodiments, the methods and nucleic acid reagents described herein can be used to generate genetically modified cells and organisms with altered activities in cellular carbon processing.

For example, the methods of genetic modification can be used to alter fatty acid (e.g., oleic acid, adipic acid, sebacic acid, suberic acid, octanedioic acid, decanedioic acid, dodecanedioic acid, tetradecanedioic acid, hexadecanedioic acid, octadecanedioic acid, eicosanedioic acid) synthesis and/or catabolism. In some embodiments, an engineered cell or organism described herein may include an increased number of copies of one or more polynucleotides encoding one or more polypeptides having carnitine acetyltransferase, acetyl-CoA carboxylase, ATP citrate lyase, thioesterase, acetyl-CoA hydrolase, acetyl-CoA synthetase, acyl-CoA oxidase, cytochrome P450 reductase, monooxygenase, peroxisomal biogenesis factor, alcohol dehydrogenase, alcohol oxidase, aldehyde dehydrogenase, 3-ketoacyl-CoA thiolase, and/or multifunctional enzyme (e.g., enoyl-CoA hydratase and/or 3-hydroxyacyl-CoA dehydrogenase) activity. In certain embodiments, an engineered cell or microorganism described herein may include one or more genetic modifications that reduce one or more of the following activities: carnitine acetyltransferase (e.g., mitochondrial), acetyl-carnitine translocase (e.g., mitochondrial), acyl-CoA synthetase, acyl-CoA oxidase and peroxisomal transporter activity.

In some embodiments, the engineered cell or organism can be a prokaryote. In certain embodiments, the prokaryote can be a bacterium, e.g., *Escherichia coli*. In some embodiments, the engineered cell or organism can be a eukaryote. In some embodiments, the eukaryote may be a fungus. In certain embodiments, the eukaryote can be a yeast. In certain embodiments, the yeast can be a *Candida* yeast. In some embodiments, the *Candida* yeast may be *C. viswanathii* or *C. troplicalis*. In certain embodiments, the fungus can be a *Yarrowia* fungus. In some embodiments the *Yarrowia* fungus may be *Y. lipolytica*. In some embodiments, the fungus can be a *Blastobotrys* yeast, e.g., *B. adeninivorans*. In certain embodiments, the fungus can be an *Aspergillus* fungus. In some embodiments, the *Aspergillus* fungus may be *A. parasiticus* or A. *nidulans*.

In some embodiments, an activity and/or amount of a composition can be altered by genetically modifying a host cell or organism which yields an engineered cell or organism having added, increased, reduced, decreased, inhibited, redirected, removed and/or otherwise modified activity or composition. A cell or organism may be modified, for example, by altering the amount of one or more cellular compositions, e.g, polynucleotides and/or polypeptides. Engineered cells or organisms typically arise as a result of a genetic modification, usually introduced by one of skill in the art using readily available techniques. Such cells or organisms are referred to herein as genetically modified or genetically engineered cells, microorganisms or organisms. The term

"genetic modification" as used herein refers to any alteration in the genetic make-up of a cell or organism, including, for example, any nucleic acid addition, removal or alteration. Genetic modifications include, without limitation, insertion of one or more nucleotides in an endogenous nucleic acid of a host cell or organism in one or more locations, deletion of one or more nucleotides in an endogenous nucleic acid of a host cell or organism in one or more locations, modification or substitution of one or more nucleotides in an endogenous nucleic acid of a host cell or organism in one or more locations. In some embodiments, a portion of a host genome can be replaced with a heterologous nucleic acid. A genetic modification can also be insertion of a nucleic acid into a host cell organism that is distinct from the host endogenous genome (e.g., insertion of an autonomously replicating vector), and removal of a nucleic acid that is distinct from the endogenous host genome (e.g., removal of a vector).

Non-limiting examples of methods useful for genetically modifying a cell or organism include, introducing a heterologous polynucleotide (e.g., nucleic acid or gene integration, also referred to as “knock in”), removing an endogenous polynucleotide, altering the sequence of an existing endogenous nucleic acid sequence (e.g., site-directed mutagenesis), disruption of an existing endogenous nucleic acid sequence (e.g., knock outs and transposon or insertion element mediated mutagenesis), selection for an altered activity where the selection causes a change in a naturally occurring activity that can be stably inherited (e.g., causes a change in a nucleic acid sequence in the genome of the cell or organism or in an epigenetic nucleic acid that is replicated and passed on to daughter cells), PCR-based mutagenesis, and the like. The terms “mutant” and “mutagenesis” as used herein refer to any modification to a nucleic acid (e.g., nucleic acid reagent or host chromosome) and/or polypeptide which results in an altered nucleic acid and/or polypeptide. Non-limiting examples of mutagenesis include, deletion, insertion, substitution, rearrangement, point mutations, suppressor mutations and the like of a single or multiple residues in a polynucleotide. Mutagenesis methods are known in the art and are readily available to the artisan. Non-limiting examples of mutagenesis methods are described herein and can also be found in Maniatis, T., E. F. Fritsch and J. Sambrook (1982) *Molecular Cloning: a Laboratory Manual*; Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. Another non-limiting example of mutagenesis can be conducted using an Agilent (Santa Clara, CA) “QuickChange” kit according to the manufacturer’s instructions.

Decreasing an amount of a composition and/or activity in a cell

An altered activity or composition sometimes is an activity or composition detectable in a host cell or organism and that is reduced, decreased, inhibited or removed (i.e., not detectable) in an engineered cell or organism. For example, a genetic modification that disrupts cellular synthesis of a composition (e.g., acyl-CoA synthetase protein) and/or or disrupts an activity, such as activation of fatty acids, or disrupts a polynucleotide that encodes a polypeptide that carries out a forward reaction in the activity (e.g., acyl-CoA synthetase activity), may render the composition (e.g., acyl-CoA synthetase protein) or activity, such as fatty acid activation, undetectable. The term "undetectable" as used herein refers to an amount of an analyte (including an activity) that is below the limits of detection, using know detection methods or assays (e.g., described herein). In certain embodiments, the genetic modification may partially reduce or decrease a composition or an activity. The term "reduces" or "decreases" with reference to a composition or an activity as used herein refers to a level of the composition or activity in an engineered cell or organism that is lower than the level of the composition or activity found in the host or starting cell or organism. A "lower" level can be a level that is detectable or undetectable. The term "partially reduces" or "partially decreases" with reference to a composition or an activity as used herein refers to a level of the composition or activity in an engineered cell or organism that is lower than the level of the composition or activity found in the host or starting cell or organism but that is still detectable. Thus, an activity or composition can be reduced to undetectable levels in some embodiments, or detectable levels in certain embodiments. An activity or composition can be decreased to any suitable level for production of a target molecule product (e.g., an organic acid), including but not limited to less than 2-fold (e.g., about 10% decrease to about 99% decrease; about 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% decrease), 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, of 10-fold decrease, or greater than about 10-fold decrease.

The term "level", as used herein, often refers to an amount (e.g., a quantitative or relative amount) of a nucleic acid (e.g. an RNA (e.g. an mRNA) or DNA), polypeptide or activty.

In some embodiments, an activity or composition may be reduced or removed by decreasing the number of copies of a polynucleotide that encodes a composition polypeptide or polypeptide having a target activity. In some embodiments, an activity or composition can be reduced or removed by (i) inserting a polynucleotide within a polynucleotide that encodes a protein having

the target activity or the target composition (disruptive insertion), and/or (ii) removing a portion of or all of a polynucleotide that encodes a polypeptide having the target activity or the target composition (deletion or knock out, respectively). In certain embodiments, an activity or composition can be reduced or removed by inserting into a host cell or microorganism a heterologous polynucleotide that is (i) operably linked to another polynucleotide that encodes a polypeptide having the target activity or target composition, and (ii) down regulates production of the polypeptide. Thus, an activity or composition can be reduced or removed by inserting or modifying a regulatory polynucleotide operably linked to another polynucleotide that encodes a polypeptide having the target activity or target composition.

An activity or composition also can be reduced or removed by (i) inhibiting a polynucleotide that encodes a polypeptide having the activity or the targeted composition or (ii) inhibiting a polynucleotide operably linked to another polynucleotide that encodes a polypeptide having the activity or targeted composition. A polynucleotide can be inhibited by a suitable technique known in the art, such as by contacting an RNA encoded by the polynucleotide with a specific inhibitory RNA (e.g., RNAi, siRNA, ribozyme). An activity also can be reduced or removed by contacting a polypeptide having the activity with a molecule that specifically inhibits the activity (e.g., enzyme inhibitor, antibody). In certain embodiments, an activity or composition can be reduced or removed by subjecting a host cell or organism to a selective environment and screening for cells or organisms that have a reduced level or removal of the activity or composition.

In some embodiments, an untranslated ribonucleic acid or a cDNA can be used to reduce the expression of a particular activity or enzyme. For example, a host cell or organism can be engineered by genetic modification to express a nucleic acid reagent that reduces the expression of an activity by producing an RNA molecule that is partially or substantially homologous to a nucleic acid sequence of interest which encodes the activity of interest. The RNA molecule can bind to the nucleic acid sequence of interest and inhibit the nucleic acid sequence from performing its natural function, in certain embodiments. In some embodiments, the RNA may alter the nucleic acid sequence of interest which encodes the activity of interest in a manner that the nucleic acid sequence of interest is no longer capable of performing its natural function (e.g., the action of a ribozyme for example).

In some embodiments, an activity and/or composition may be reduced in, or removed from, a host cell or organism by increasing or adding a separate activity or composition in the host cell or organism. For example, an activity and/or composition that inhibits a targeted activity or composition in a host cell or organism can be increased or added thereby reducing or eliminating the targeted activity or composition by adding or increasing an inhibiting activity or composition. Methods of increasing or adding an activity or composition in a cell or organism are described herein.

Increasing an amount of a composition and/or activity in a cell

An altered activity and/or composition in an engineered cell or organism is sometimes an added composition or activity that is not detectable in a host cell or organism. An altered activity or composition can also be an increased or elevated activity or amount of a composition in an engineered cell or organism. An increased or elevated activity or composition generally is an activity or an amount of the composition that is greater than the activity or composition amount detectable in a host cell or organism. However, an increased or elevated activity or amount of a composition in an engineered cell or organism can also be a detectable activity or detectable composition that is not detectable in a host cell or organism. An activity or amount of a composition can be increased to any suitable level for example, for production of a target molecule product (e.g., an organic acid), including but not limited to less than 2-fold (e.g., about 10% increase to about 99% increase; about 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% increase), 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, of 10-fold increase, or greater than about 10-fold increase.

In some embodiments, an activity and/or composition may be added to or increased in a host cell or organism by increasing the number of copies of a polynucleotide that encodes a polypeptide composition or polypeptide having the activity. In some embodiments, the activity and/or amount of a native or endogenous polypeptide can be increased by introducing heterologous nucleic acid into a host cell or organism that includes copies of a polynucleotide that encodes the polypeptide, for example, introducing 1 to about 100 additional heterologous copies of the polynucleotide (e.g., introducing 1 or more, 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, 11 or more, 12 or more, 13 or more, 14 or more, 15 or more, 16 or more, 17 or more, 18 or more, 19 or more, 20 or more, 22 or more, 24 or more, 25 or more, 26 or more, 28 or more, 30 or more additional copies of the

polynucleotide). In certain embodiments, an activity and/or composition can be added or increased by inserting into a host cell or organism a polynucleotide that encodes a heterologous polypeptide from a different species having the added activity or composition, or encodes a heterologous polypeptide that is a modified version of an endogenous polypeptide. In such embodiments, 1 to about 100 copies of the polynucleotide can be introduced (e.g., introducing 1 or more, 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, 11 or more, 12 or more, 13 or more, 14 or more, 15 or more, 16 or more, 17 or more, 18 or more, 19 or more, 20 or more, 22 or more, 24 or more, 25 or more, 26 or more, 28 or more, 30 or more copies). A heterologous polypeptide that is a "modified endogenous polypeptide" often has an activity different than an activity of a native polypeptide counterpart (e.g., different catalytic activity and/or different substrate specificity), and often is active (e.g., an activity (e.g., substrate turnover) is detectable). A heterologous polypeptide that is a "modified endogenous polypeptide" also often includes or lacks a cell location-targeting amino acid sequence that a native polypeptide counterpart has or doesn't have (e.g., in order to modify the cellular location of the expressed polypeptide). In certain embodiments, an activity or composition can be added or increased by inserting into a host cell or organism a heterologous polynucleotide that is (i) operably linked to another polynucleotide that encodes a polypeptide composition or a polypeptide having the added activity, and (ii) up regulates production of the polypeptide. Thus, a composition or an activity can be added or increased by inserting or modifying a regulatory polynucleotide operably linked to another polynucleotide that encodes a composition polypeptide or polypeptide having the targeted activity. In certain embodiments, an activity or composition can be added or increased by subjecting a host cell or organism to a selective environment and screening for cells or organisms that have a detectable level of the activity or composition. Examples of a selective environment include, without limitation, a medium containing a substrate that a host cell or organism can process and a medium lacking a substrate that a host cell or organism can process.

In some embodiments, an activity and/or composition may be added to or increased in a host cell or organism by decreasing or removing a separate activity or composition in a host cell or organism. For example, an activity and/or composition in a host cell or organism that inhibits a desired target activity or composition can be decreased or removed thereby reducing or eliminating the inhibition of the desired activity or composition and adding or increasing the desired activity. Methods of decreasing or removing an activity or composition in a cell or organism are described herein.

Nucleic acid manipulation

In certain embodiments, nucleotide sequences sometimes are added to, modified or removed from one or more of the nucleic acid reagent elements, such as the promoter, 5' UTR, target sequence, or 3' UTR elements, to enhance, potentially enhance, reduce, or potentially reduce transcription and/or translation before or after such elements are incorporated in a nucleic acid reagent. In some embodiments, one or more of the following sequences may be modified or removed if they are present in a 5' UTR: a sequence that forms a stable secondary structure (e.g., quadruplex structure or stem loop stem structure (e.g., EMBL sequences X12949, AF274954, AF139980, AF152961, S95936, U194144, AF116649 or substantially identical sequences that form such stem loop stem structures)); a translation initiation codon upstream of the nucleotide sequence start codon; a stop codon upstream of the nucleotide sequence translation initiation codon; an ORF upstream of the nucleotide sequence translation initiation codon; an iron responsive element (IRE) or like sequence; and a 5' terminal oligopyrimidine tract (TOP, e.g., consisting of 5-15 pyrimidines adjacent to the cap). Computer-assisted software programs are available for nucleic acid sequence evaluation to optimize untranslated region sequences (see, e.g., World Wide Web Uniform Resource Locator (URL) bioinformatics.ua.pt/software/mrna-optimiser/). A translational enhancer sequence and/or an internal ribosome entry site (IRES) sometimes is inserted into a 5'UTR (e.g., EMBL nucleotide sequences J04513, X87949, M95825, M12783, AF025841, AF013263, AF006822, M17169, M13440, M22427, D14838 and M17446 and substantially identical nucleotide sequences).

An AU-rich element (ARE, e.g., AUUUA repeats) and/or splicing junction that follows a non-sense codon sometimes is removed from or modified in a 3' UTR. A polyadenosine tail sometimes is inserted into a 3' UTR if none is present, sometimes is removed if it is present, and adenosine moieties sometimes are added to or removed from a polyadenosine tail present in a 3' UTR. Thus, some embodiments are directed to a process that includes: determining whether any nucleotide sequences that increase, potentially increase, reduce or potentially reduce translation efficiency are present in the elements, and adding, removing or modifying one or more of such sequences if they are identified. Certain embodiments are directed to a process that includes: determining whether any nucleotide sequences that increase or potentially increase translation efficiency are not present in the elements, and incorporating such sequences into the nucleic acid reagent.

In some embodiments, an activity and/or composition can be altered by modifying the nucleotide sequence of an ORF. An ORF sometimes is mutated or modified (for example, by point mutation, deletion mutation, insertion mutation, PCR based mutagenesis and the like) to alter, enhance or increase, reduce, substantially reduce or eliminate the activity of the encoded protein or peptide. The protein or peptide encoded by a modified ORF sometimes is produced in a lower amount or may not be produced at detectable levels, and in other embodiments, the product or protein encoded by the modified ORF is produced at a higher level (e.g., codons sometimes are modified so they are compatible with tRNA's preferentially used in the host or engineered cell or organism). To determine the relative activity, the activity from the product of the mutated ORF (or cell containing it) can be compared to the activity of the product or protein encoded by the unmodified ORF (or cell containing it).

In some embodiments, an ORF nucleotide sequence sometimes is mutated or modified to alter the triplet nucleotide sequences used to encode amino acids (e.g., amino acid codon triplets, for example). Modification of the nucleotide sequence of an ORF to alter codon triplets sometimes is used to change the codon found in the original sequence to better match the preferred codon usage of the organism in which the ORF or nucleic acid reagent will be expressed. The codon usage, and therefore the codon triplets encoded by a nucleic acid sequence, in bacteria may be different from the preferred codon usage in eukaryotes, like yeast or plants for example. Preferred codon usage also may be different between bacterial species. In certain embodiments, an ORF nucleotide sequence sometimes is modified to eliminate codon pairs and/or eliminate mRNA secondary structures that can cause pauses during translation of the mRNA encoded by the ORF nucleotide sequence. Translational pausing sometimes occurs when nucleic acid secondary structures exist in an mRNA, and sometimes occurs due to the presence of codon pairs that slow the rate of translation by causing ribosomes to pause. In some embodiments, the use of lower abundance codon triplets can reduce translational pausing due to a decrease in the pause time needed to load a charged tRNA into the ribosome translation machinery. Therefore, to increase transcriptional and translational efficiency in bacteria (e.g., where transcription and translation are concurrent, for example) or to increase translational efficiency in eukaryotes (e.g., where transcription and translation are functionally separated), the nucleotide sequence of a nucleotide sequence of interest can be altered to better suit the transcription and/or translational machinery of the host and/or genetically modified cell or organism. In certain embodiments, slowing the rate of translation by the use of

lower abundance codons, which slow or pause the ribosome, can lead to higher yields of the desired product due to an increase in correctly folded proteins and a reduction in the formation of inclusion bodies.

Codons can be altered and optimized according to the preferred usage by a given organism by determining the codon distribution of the nucleotide sequence donor organism and comparing the distribution of codons to the distribution of codons in the recipient or host organism. Techniques described herein (e.g., site directed mutagenesis and the like) can then be used to alter the codons accordingly. Comparisons of codon usage can be done by hand, or using nucleic acid analysis software commercially available to the artisan (see, e.g., World Wide Web Uniform Resource Locator (URL) kazusa.or.jp/codon/, World Wide Web Uniform Resource Locator (URL) jcat.de, World Wide Web Uniform Resource Locator (URL) idtdna.com/CodonOpt).

Modification of the nucleotide sequence of an ORF also can be used to correct codon triplet sequences that have diverged in different organisms. For example, certain yeast (e.g., *Candida tropicalis*, *Candida viswanathii* and *Candida maltosa*) use the amino acid triplet CUG (e.g., CTG in the DNA sequence) to encode serine. CUG typically encodes leucine in most organisms. In order to maintain the correct amino acid in the resultant polypeptide or protein, the CUG codon must be altered to reflect the organism in which the nucleic acid reagent will be expressed. Thus, if an ORF from a bacterial donor is expressed in such a *Candida* yeast strain mentioned above, the heterologous nucleotide sequence must first be altered or modified to the appropriate leucine codon. Therefore, in some embodiments, the nucleotide sequence of an ORF sometimes is altered or modified to correct for differences that have occurred in the evolution of the amino acid codon triplets between different organisms. In some embodiments, the nucleotide sequence can be left unchanged at a particular amino acid codon, if the amino acid encoded is a conservative or neutral change in amino acid when compared to the originally encoded amino acid.

In some embodiments, an activity can be altered by modifying translational regulation signals, like a stop codon for example. A stop codon at the end of an ORF sometimes is modified to another stop codon, such as an amber stop codon. In some embodiments, a stop codon is introduced within an ORF, sometimes by insertion or mutation of an existing codon. An ORF comprising a modified terminal stop codon and/or internal stop codon often is translated in a

system comprising a suppressor tRNA that recognizes the stop codon. An ORF comprising a stop codon sometimes is translated in a system comprising a suppressor tRNA that incorporates an unnatural amino acid during translation of the target protein or target peptide. Methods for incorporating unnatural amino acids into a target protein or peptide are known, which include, for example, processes utilizing a heterologous tRNA/synthetase pair, where the tRNA recognizes an amber stop codon and is loaded with an unnatural amino acid (e.g., World Wide Web Uniform Resource Locator (URL) iupac.org/news/prize/2003/wang.pdf).

Depending on the portion of a nucleic acid reagent (e.g., promoter, 5' or 3' UTR, ORI, ORF, and the like) chosen for alteration (e.g., by mutagenesis, introduction or deletion, for example), the modifications described above can alter a given activity by (i) increasing or decreasing feedback inhibition mechanisms, (ii) increasing or decreasing promoter initiation, (iii) increasing or decreasing translation initiation, (iv) increasing or decreasing translational efficiency, (v) modifying localization of peptides or products expressed from nucleic acid reagents described herein, (vi) increasing or decreasing the copy number of a nucleotide sequence of interest, or (vii) expression of an anti-sense RNA, RNAi, siRNA, ribozyme and the like. In some embodiments, alteration of a nucleic acid reagent or nucleotide sequence can alter a region involved in feedback inhibition (e.g., 5' UTR, promoter and the like). A modification sometimes is made that can add or enhance binding of a feedback regulator and sometimes a modification is made that can reduce, inhibit or eliminate binding of a feedback regulator.

In certain embodiments, alteration of a nucleic acid reagent or nucleotide sequence can alter sequences involved in transcription initiation (e.g., promoters, 5' UTR, and the like). A modification sometimes can be made that can enhance or increase initiation from an endogenous or heterologous promoter element. A modification sometimes can be made that removes or disrupts sequences that increase or enhance transcription initiation, resulting in a decrease or elimination of transcription from an endogenous or heterologous promoter element.

In some embodiments, alteration of a nucleic acid reagent or nucleotide sequence can alter sequences involved in translational initiation or translational efficiency (e.g., 5' UTR, 3' UTR, codon triplets of higher or lower abundance, translational terminator sequences and the like, for example). A modification sometimes can be made that can increase or decrease translational initiation, modifying a ribosome binding site for example. A modification sometimes can be made that can increase or decrease translational efficiency. Removing or adding sequences

that form hairpins and changing codon triplets to a more or less preferred codon are non-limiting examples of genetic modifications that can be made to alter translation initiation and translation efficiency.

In certain embodiments, alteration of a nucleic acid reagent or nucleotide sequence can alter sequences involved in localization of peptides, proteins or other desired products (e.g., an organic acid, for example). A modification sometimes can be made that can alter, add or remove sequences responsible for targeting a polypeptide, protein or product to an intracellular organelle, the periplasm, cellular membranes, or extracellularly. Transport of a heterologous product to a different intracellular space or extracellularly sometimes can reduce or eliminate the formation of inclusion bodies (e.g., insoluble aggregates of the desired product).

In some embodiments, alteration of a nucleic acid reagent or nucleotide sequence can alter sequences involved in increasing or decreasing the copy number of a nucleotide sequence of interest. A modification sometimes can be made that increases or decreases the number of copies of an ORF stably integrated into the genome of an organism or on an epigenetic nucleic acid reagent. Non-limiting examples of alterations that can increase the number of copies of a sequence of interest include, adding copies of the sequence of interest by duplication of regions in the genome (e.g., adding additional copies by recombination or by causing gene amplification of the host genome, for example), cloning additional copies of a sequence onto a nucleic acid reagent, or altering an ORI to increase the number of copies of an epigenetic nucleic acid reagent. Non-limiting examples of alterations that can decrease the number of copies of a sequence of interest include, removing copies of the sequence of interest by deletion or disruption of regions in the genome, removing additional copies of the sequence from epigenetic nucleic acid reagents, or altering an ORI to decrease the number of copies of an epigenetic nucleic acid reagent.

In certain embodiments, increasing or decreasing the expression of a nucleotide sequence of interest can also be accomplished by altering, adding or removing sequences involved in the expression of an anti-sense RNA, RNAi, siRNA, ribozyme and the like. The methods described herein can be used to modify expression of anti-sense RNA, RNAi, siRNA, ribozyme and the like.

Nucleic acid sequences of interest can be genetically modified using methods known in the art. Mutagenesis techniques are particularly useful for small scale (e.g., 1, 2, 5, 10 or more nucleotides) or large scale (e.g., 50, 100, 150, 200, 500, or more nucleotides) genetic modification. Mutagenesis allows the artisan to alter the genetic information of a cell or organism in a stable manner, either naturally (e.g., isolation using selection and screening) or experimentally by the use of chemicals, radiation or inaccurate DNA replication (e.g., PCR mutagenesis). In some embodiments, genetic modification can be performed by whole scale synthetic synthesis of nucleic acids, using a native nucleotide sequence as the reference sequence, and modifying nucleotides that can result in the desired alteration of activity. Mutagenesis methods sometimes are specific or targeted to specific regions or nucleotides (e.g., site-directed mutagenesis, PCR-based site-directed mutagenesis, and in vitro mutagenesis techniques such as transplacement and in vivo oligonucleotide site-directed mutagenesis, for example). Mutagenesis methods sometimes are non-specific or random with respect to the placement of genetic modifications (e.g., chemical mutagenesis, insertion element (e.g., insertion or transposon elements) and inaccurate PCR based methods, for example).

Site directed mutagenesis is a procedure in which a specific nucleotide or specific nucleotides in a nucleic acid molecule are mutated or altered. Site directed mutagenesis typically is performed using a nucleic acid sequence of interest cloned into a circular plasmid vector. Site-directed mutagenesis requires that the wild type sequence be known and used a platform for the genetic alteration. Site-directed mutagenesis sometimes is referred to as oligonucleotide-directed mutagenesis because the technique can be performed using oligonucleotides which have the desired genetic modification incorporated into the complement of a nucleotide sequence of interest. The wild type sequence and the altered nucleotide are allowed to hybridize and the hybridized nucleic acids are extended and replicated using a DNA polymerase. The double stranded nucleic acids are introduced into a host (e.g., *E. coli*, for example) and further rounds of replication are carried out in vivo. The transformed cells carrying the mutated nucleic acid sequence are then selected and/or screened for those cells carrying the correctly mutagenized sequence. Cassette mutagenesis and PCR-based site-directed mutagenesis are further modifications of the site-directed mutagenesis technique. Site-directed mutagenesis can also be performed in vivo (e.g., transplacement “pop-in pop-out”, in vivo site-directed mutagenesis with synthetic oligonucleotides and the like, for example).

PCR-based mutagenesis can be performed using PCR with oligonucleotide primers that contain the desired mutation or mutations. The technique functions in a manner similar to standard site-directed mutagenesis, with the exception that a thermocycler and PCR conditions are used to replace replication and selection of the clones in a microorganism host. As PCR-based mutagenesis also uses a circular plasmid vector, the amplified fragment (e.g., linear nucleic acid molecule) containing the incorporated genetic modifications can be separated from the plasmid containing the template sequence after a sufficient number of rounds of thermocycler amplification, using standard electrophoretic procedures. A modification of this method uses linear amplification methods and a pair of mutagenic primers that amplify the entire plasmid. The procedure can take advantage of the *E. coli* Dam methylase system which causes DNA replicated in vivo to be sensitive to the restriction endonucleases DpnI. PCR synthesized DNA is not methylated and is therefore resistant to DpnI. This approach allows digestion of the template plasmid, leaving the genetically modified, PCR synthesized plasmids for isolating and transforming into a host bacteria for DNA repair and replication, thereby facilitating subsequent cloning and identification steps. A certain amount of randomness can be added to PCR-based sited directed mutagenesis by using partially degenerate primers.

DNA shuffling is a method which uses DNA fragments from members of a mutant library and reshuffles the fragments randomly to generate new mutant sequence combinations. The fragments are typically generated using DNaseI, followed by random annealing and re-joining using self priming PCR. The DNA overhanging ends, from annealing of random fragments, provide “primer” sequences for the PCR process. Shuffling can be applied to libraries generated by any of the above mutagenesis methods.

Error prone PCR and its derivative rolling circle error prone PCR uses increased magnesium and manganese concentrations in conjunction with limiting amounts of one or two nucleotides to reduce the fidelity of the Taq polymerase. The error rate can be as high as 2% under appropriate conditions, when the resultant mutant sequence is compared to the wild type starting sequence. After amplification, the library of mutant coding sequences must be cloned into a suitable plasmid. Although point mutations are the most common types of mutation in error prone PCR, deletions and frameshift mutations are also possible. There are a number of commercial error-prone PCR kits available, including those from Agilent and Takara Bio, U.S.A. (e.g., World Wide Web Uniform Resource Locator (URL) agilent.com and World Wide Web Uniform Resource Locator (URL) clontech.com, respectively, for example). Rolling circle error-

prone PCR is a variant of error-prone PCR in which wild-type sequence is first cloned into a plasmid, then the whole plasmid is amplified under error-prone conditions.

In contrast to site-directed or specific mutagenesis, random mutagenesis does not require any sequence information and can be accomplished by a number of widely different methods. Random mutagenesis often is used to generate mutant libraries that can be used to screen for the desired genotype or phenotype. Non-limiting examples of random mutagenesis include; chemical mutagenesis, UV-induced mutagenesis, insertion element or transposon-mediated mutagenesis, DNA shuffling, error-prone PCR mutagenesis, and the like.

Chemical mutagenesis often involves chemicals like ethyl methanesulfonate (EMS), nitrous acid, mitomycin C, N-methyl-N-nitrosourea (MNU), diepoxybutane (DEB), 1, 2, 7, 8-diepoxyoctane (DEO), methyl methane sulfonate (MMS), N-methyl- N'-nitro-N-nitrosoguanidine (MNNG), 4-nitroquinoline 1-oxide (4-NQO), 2-methyloxy-6-chloro-9(3-[ethyl-2-chloroethyl]-aminopropylamino)-acridinedihydrochloride (ICR-170), 2-amino purine (2AP), and hydroxylamine (HA), provided herein as non-limiting examples. These chemicals can cause base-pair substitutions, frameshift mutations, deletions, transversion mutations, transition mutations, incorrect replication, and the like. In some embodiments, the mutagenesis can be carried out in vivo. Sometimes the mutagenic process involves the use of the host organism's DNA replication and repair mechanisms to incorporate and replicate the mutagenized base or bases.

Another type of chemical mutagenesis involves the use of base-analogs. The use of base-analogs causes incorrect base pairing which in the following round of replication is corrected to a mismatched nucleotide when compared to the starting sequence. Base analog mutagenesis introduces a small amount of non-randomness to random mutagenesis, because specific base analogs can be chosen which can be incorporated at certain nucleotides in the starting sequence. Correction of the mispairing typically yields a known substitution. For example, bromo-deoxyuridine (BrdU) can be incorporated into DNA and replaces T in the sequence. The host DNA repair and replication machinery can sometime correct the defect, but sometimes will mispair the BrdU with a G. The next round of replication then causes a G-C transversion from the original A-T in the native sequence.

Ultra violet (UV) induced mutagenesis is caused by the formation of thymidine dimers when UV light irradiates chemical bonds between two adjacent thymine residues. Excision repair mechanism of the host organism correct the lesion in the DNA, but occasionally the lesion is incorrectly repaired typically resulting in a C to T transition.

In some embodiments, an altered activity can be found by screening cells or an organism under conditions that select for the desired change in activity. For example, certain microorganisms can be adapted to increase or decrease an activity by selecting or screening the organism in question on a media containing substances that are poorly metabolized or even toxic. An increase in the ability of an organism to grow a substance that is normally poorly metabolized may result in an increase in the growth rate on that substance, for example. A decrease in the sensitivity to a toxic substance might be manifested by growth on higher concentrations of the toxic substance, for example. Modifications obtained in this manner are not limited to alterations in promoter sequences. That is, screening microorganisms by selective pressure, as described above, can yield genetic alterations that can occur in non-promoter sequences, and sometimes also can occur in sequences that are not in the nucleotide sequence of interest, but in a related nucleotide sequences (e.g., a gene involved in a different step of the same pathway, a transport gene, and the like). Such mutants sometimes can be found by isolating variants from unique environments.

Cells or organisms with altered activities can also be isolated using genetic selection and screening of cells or organisms challenged on selective media or by identifying naturally occurring variants from unique environments. For example, 2-deoxy-D-glucose is a toxic glucose analog. Growth of yeast on this substance yields mutants that are glucose-deregulated. A number of mutants have been isolated using 2-deoxy-D-glucose including transport mutants, and mutants that ferment glucose and galactose simultaneously instead of glucose first then galactose when glucose is depleted. Similar techniques have been used to isolate mutant microorganisms that can metabolize plastics (e.g., from landfills), petrochemicals (e.g., from oil spills), and the like, either in a laboratory setting or from unique environments.

Similar methods can be used to isolate cells or organisms having existing mutations in a desired activity when the activity exists at a relatively low or nearly undetectable level in the cell or organism of choice, in some embodiments. The method generally consists of growing the cell or organism to a specific density in liquid culture, concentrating the cells, and plating the cells on

various concentrations of the substance to which an increase in metabolic activity is desired. The cells are incubated at a moderate growth temperature, for 5 to 10 days. To enhance the selection process, the plates can be stored for another 5 to 10 days at a low temperature. The low temperature sometimes can allow strains that have gained or increased an activity to continue growing while other strains are inhibited for growth at the low temperature. Following the initial selection and secondary growth at low temperature, the plates can be replica plated on higher or lower concentrations of the selection substance to further select for the desired activity.

Insertion element or transposon-mediated mutagenesis makes use of naturally occurring or modified naturally occurring mobile genetic elements. Transposons often encode accessory activities in addition to the activities necessary for transposition (e.g., movement using a transposase activity, for example). In many examples, transposon accessory activities are antibiotic resistance markers (e.g., Tn903 kan^r). Insertion elements typically only encode the activities necessary for movement of the nucleic acid sequence. Insertion element and transposon mediated mutagenesis often can occur randomly, however specific target sequences are known for some transposons. Mobile genetic elements like IS elements or Transposons (Tn) often have inverted repeats, direct repeats or both inverted and direct repeats flanking the region coding for the transposition genes. Recombination events catalyzed by the transposase cause the element to remove itself from the genome and move to a new location, leaving behind a portion of an inverted or direct repeat. Classic examples of transposons are the “mobile genetic elements” discovered in maize. Transposon mutagenesis kits are commercially available which are designed to leave behind a 5 codon insert (e.g., Mutation Generation System kit, ThermoFisher Scientific, World Wide Web Uniform Resource Locator (URL) thermofisher.com/, for example). This allows the artisan to identify the insertion site, without fully disrupting the function of most genes.

Introduction of nucleic acids into cells

Engineered cells and organisms can be prepared by altering, introducing and/or removing nucleotide sequences in the host genome or in stably maintained epigenetic nucleic acid reagents, as described herein. The nucleic acid reagents used to alter, introduce or remove nucleotide sequences in the host genome or epigenetic nucleic acids can be prepared using the methods described herein and/or available to the artisan.

Nucleic acid sequences having a desired activity can be isolated from cells of a suitable organism using lysis and nucleic acid purification procedures described in a known reference manual (e.g., Maniatis, T., E. F. Fritsch and J. Sambrook (1982) *Molecular Cloning: a Laboratory Manual*; Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.) or using commercially available cell lysis and DNA purification reagents and kits. In some embodiments, nucleic acids used to engineer cells or microorganisms can be provided for conducting methods described herein after processing of the organism containing the nucleic acid. For example, the nucleic acid of interest may be extracted, isolated, purified or amplified from a sample (e.g., from a cell(s) or organism of interest or culture containing a plurality of cells or organisms of interest, like yeast or bacteria for example). The term “isolated” as used herein refers to nucleic acid removed from its original environment (e.g., the natural environment if it is naturally occurring, or a host cell if expressed exogenously), and thus is altered "by the hand of man" from its original environment. An isolated nucleic acid generally is provided with fewer non-nucleic acid components (e.g., protein, lipid) than the amount of components present in a source sample. A composition containing isolated sample nucleic acid can be substantially isolated (e.g., about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or greater than 99% free of non-nucleic acid components). The term “purified” as used herein refers to sample nucleic acid provided that contains fewer nucleic acid species than in the sample source from which the sample nucleic acid is derived. A composition containing sample nucleic acid may be substantially purified (e.g., about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or greater than 99% free of other nucleic acid species). The term “amplified” as used herein refers to subjecting nucleic acid of a cell, organism or sample to a process that linearly or exponentially generates amplicon nucleic acids having the same or substantially the same nucleotide sequence as the nucleotide sequence of the nucleic acid in the sample, or portion thereof. As noted herein, the nucleic acids used to prepare nucleic acid reagents as described herein can be subjected to fragmentation or cleavage.

Amplification of nucleic acids is sometimes necessary when dealing with cells or organisms that are difficult to culture. Where amplification may be desired, any suitable amplification technique can be utilized. Non-limiting examples of methods for amplification of polynucleotides include, polymerase chain reaction (PCR); ligation amplification (or ligase chain reaction (LCR)); amplification methods based on the use of Q-beta replicase or template-dependent polymerase (see US Patent Publication Number US20050287592); helicase-dependent isothermal

amplification (Vincent et al., "Helicase-dependent isothermal DNA amplification". EMBO reports 5 (8): 795–800 (2004)); strand displacement amplification (SDA); thermophilic SDA nucleic acid sequence based amplification (3SR or NASBA) and transcription-associated amplification (TAA). Non-limiting examples of PCR amplification methods include standard PCR, AFLP-PCR, Allele-specific PCR, Alu-PCR, Asymmetric PCR, Colony PCR, Hot start PCR, Inverse PCR (IPCR), In situ PCR (ISH), Intersequence-specific PCR (ISSR-PCR), Long PCR, Multiplex PCR, Nested PCR, Quantitative PCR, Reverse Transcriptase PCR (RT-PCR), Real Time PCR, Single cell PCR, Solid phase PCR, combinations thereof, and the like. Reagents and hardware for conducting PCR are commercially available.

Protocols for conducting the various type of PCR listed above are readily available to the artisan. PCR conditions can be dependent upon primer sequences, target abundance, and the desired amount of amplification, and therefore, one of skill in the art may choose from a number of PCR protocols available (see, e.g., U.S. Pat. Nos. 4,683,195 and 4,683,202; and PCR Protocols: A Guide to Methods and Applications, Innis et al., eds, 1990). PCR often is carried out as an automated process with a thermostable enzyme. In this process, the temperature of the reaction mixture is cycled through a denaturing region, a primer-annealing region, and an extension reaction region automatically. Machines specifically adapted for this purpose are commercially available. A non-limiting example of a PCR protocol that may be suitable for embodiments described herein is, treating the sample at 95ºC for 5 minutes; repeating forty-five cycles of 95ºC for 1 minute, 59ºC for 1 minute, 10 seconds, and 72ºC for 1 minute 30 seconds; and then treating the sample at 72ºC for 5 minutes. Multiple cycles frequently are performed using a commercially available thermal cycler. Suitable isothermal amplification processes known and selected by the person of ordinary skill in the art also may be applied, in certain embodiments. In some embodiments, nucleic acids encoding polypeptides with a desired activity can be isolated by amplifying the desired sequence from a cell or organism having the desired activity using oligonucleotides or primers designed based on sequences described herein.

Synthetic nucleic acids, e.g., codon-optimized sequences, can be generated using a variety of methods. For example, whole-scale synthetic chemistry can be used to generate an entire sequence. Other methods include use of chemically-generated oligonucleotides in amplification methods, e.g., recursive PCR, that build an entire nucleotide sequence (see, e.g., Prodromou and Pearl (1992) *Protein Engineering* 5(8):827-829; Yehezkel et al. (2013) *Gene Synthesis:*

Methods and Protocols in Methods in Molecular Biology 852:35-47, Jean Piccoud (ed.) Springer Science and Business Media LLC).

Amplified, isolated and/or purified nucleic acids can be cloned into the recombinant DNA vectors described herein or into suitable commercially available recombinant DNA vectors. Cloning of nucleic acid sequences of interest into recombinant DNA vectors can facilitate further manipulations of the nucleic acids for preparation of nucleic acid reagents, (e.g., alteration of nucleotide sequences by mutagenesis, homologous recombination, amplification and the like, for example). Standard cloning procedures (e.g., enzymatic digestion, ligation, and the like) are known (e.g., described in Maniatis, T., E. F. Fritsch and J. Sambrook (1982) *Molecular Cloning: a Laboratory Manual*; Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).

In some embodiments, nucleic acid sequences prepared by isolation or amplification can be used, without any further modification, to add an activity to a cell or microorganism and thereby create a genetically modified or engineered cell or microorganism. In certain embodiments, nucleic acid sequences prepared by isolation or amplification can be genetically modified to alter (e.g., increase or decrease, for example) a desired activity. In some embodiments, nucleic acids, used to add an activity or composition to a cell or organism, sometimes are genetically modified to optimize the heterologous polynucleotide sequence encoding the desired activity (e.g., polypeptide or protein, for example). The term “optimize” as used herein can refer to alteration to increase or enhance expression by preferred codon usage. The term optimize can also refer to modifications to the amino acid sequence to increase the activity of a polypeptide or protein, such that the activity exhibits a higher catalytic activity as compared to the “natural” version of the polypeptide or protein.

A heterologous, recombinant, or mutagenized polynucleotide can be introduced into a nucleic acid reagent for introduction into a host cell or organism, thereby generating an engineered cell or microorganism. Standard recombinant DNA techniques (restriction enzyme digests, ligation, and the like) can be used by the artisan to combine a nucleic acid of interest into a suitable nucleic acid reagent capable of (i) being stably maintained by selection in the host cell or organism, or (ii) being integrated into the genome of the host cell or organism. Sometimes nucleic acid reagents include two replication origins to allow manipulation of the same nucleic acid reagent in bacteria before final introduction of the final product into the host cell or organism (e.g., yeast or fungus for example). Standard molecular biology and recombinant

DNA methods are known (e.g., described in Maniatis, T., E. F. Fritsch and J. Sambrook (1982) *Molecular Cloning: a Laboratory Manual*; Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).

Nucleic acid reagents can be introduced into cells or microorganisms using various techniques. Non-limiting examples of methods used to introduce heterologous nucleic acids into various cells or organisms include; transformation, transfection, transduction, electroporation, ultrasound-mediated transformation, particle bombardment and the like. In some instances the addition of carrier molecules (e.g., bis-benzimdazolyl compounds, for example, see US Patent 5595899) can increase the uptake of DNA in cells that may be difficult to transform by conventional methods. Conventional methods of transformation are known (e.g., described in Maniatis, T., E. F. Fritsch and J. Sambrook (1982) *Molecular Cloning: a Laboratory Manual*; Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).

Linear DNA transformed into a host cell can be integrated into the genome by homologous recombination. The localization of genomic integration is determined by the homologous sequence at the ends of the transformed linear DNA. Fig. 1 is a diagrammatic representation of a cassette for the addition of a gene of interest (GOI) into a host non-functional ura3 locus using the single crossover integration method. The core of the cassette contains the GOI gene with a promoter (P_{XXX}) and terminator (T_{XXX}) for controlling transcription of the GOI gene. These DNA cassettes are typically generated by overlap extension PCR assembly of the cassette elements or by PCR amplification from circular plasmids containing the entire cassette. Additionally, circular plasmids containing cassette elements may be cut within (e.g., the middle of) the *URA3* ORF to generate a linear DNA fragment used in transforming cells. A circular DNA vector containing the cassette core and an intact *URA3* gene can be linearized by endonuclease-mediated cutting the vector such that it splits the *URA3* selectable marker within (e.g., the middle of) the ORF. The resulting linear DNA contains the expression cassette of the gene of interest (GOI) positioned between the *URA3* promoter (P_{URA3}) and terminator (T_{URA3}). Parts A and B of Fig. 1 show the results of integration of one copy (A) and two copies (B) of the cassette into a Ura^- auxotrophic mutant strain. Integration of one cassette generates an added, functional GOI expression unit and may or may not provide for expression of a functional Ura3p, depending on the nature of ura3 locus and the location of the split in the *URA3* selectable marker. Integration of two or more copies of the cassette in tandem arrays generates a complete, functional *URA3* sequence by combining the 5' end of *URA3* from one copy of the

cassette and the 3' end of *URA3* from the second copy of the cassette. Additional copies may also be integrated. Transformants can be selected by growth on uracil-free media. This integration method thus favors selection of transformants containing multiple copies of the GOI.

In some embodiments, other auxotrophic or dominant selection markers can be used in place of URA3 (e.g., an auxotrophic selectable marker), with the appropriate change in selection media and selection agents. Auxotrophic selectable markers are used in strains deficient for synthesis of a required biological molecule (e.g., amino acid or nucleoside, for example). Non-limiting examples of additional auxotrophic markers include; HIS3, TRP1, LEU2, LEU2-d, and LYS2. Certain auxotrophic markers (e.g., URA3 and LYS2) allow counter selection to select for the second recombination event that pops out all but one of the direct repeats of the recombination construct. HIS3 encodes an activity involved in histidine synthesis. TRP1 encodes an activity involved in tryptophan synthesis. LEU2 encodes an activity involved in leucine synthesis. LEU2-d is a low expression version of LEU2 that selects for increased copy number (e.g., gene or plasmid copy number, for example) to allow survival on minimal media without leucine. LYS2 encodes an activity involved in lysine synthesis, and allows counter selection for recombination out of the LYS2 gene using alpha-aminoadipate (α-aminoadipate).

Dominant selectable markers are useful because they also allow use of industrial and/or prototrophic strains for genetic manipulations. Additionally, dominant selectable markers provide the advantage that rich medium can be used for plating and culture growth, and thus growth rates are markedly increased. Non-limiting examples of dominant selectable markers include; Tn903 kan^r, Cm^r, Hyg^r, CUP1, and DHFR. Tn903 kan^r encodes an activity involved in kanamycin antibiotic resistance (e.g., typically neomycin phosphotransferase II or NPTII, for example). Cm^r encodes an activity involved in chloramphenicol antibiotic resistance (e.g., typically chloramphenicol acetyl transferase, for example). Hyg^r encodes an activity involved in hygromycin resistance by phosphorylation of hygromycin B (e.g., hygromycin phosphotransferase, or HPT). CUP1 encodes an activity involved in resistance to heavy metal (e.g., copper, for example) toxicity. DHFR encodes a dihydrofolate reductase activity which confers resistance to methotrexate and sulfanilamde compounds.

Homologous recombination can also be used as a tool for mutagenesis. Homologous recombination can be used to specifically target regions of known sequence for insertion of heterologous nucleotide sequences using the host cell's natural DNA replication and repair

enzymes. Homologous recombination methods sometimes are referred to as mutagenesis, transplacement, knock-out mutagenesis or knock-in mutagenesis. Integration of a nucleic acid sequence into a host genome by a double crossover homologous recombination event inserts the entire nucleic acid reagent at the targeted location. A second homologous recombination event driven by direct repeat DNA sequences contained in the integrated nucleic acid cassette excises (e.g., "pop out" or "loop out") all but a portion of the nucleic acid reagent, leaving behind a heterologous sequence, often referred to as a "footprint" or "scar". Mutagenesis by insertion (e.g., knock in) or by leaving behind a disrupting heterologous nucleic acid (e.g., knock out) serves to disrupt or "knock out" the function of the gene or nucleic acid sequence in which insertion occurs. By combining selectable markers and/or auxotrophic markers with nucleic acid reagents designed to provide the appropriate nucleic acid target sequences, the artisan can target a selectable nucleic acid reagent to a specific genomic region, and then select for recombination events that "pop out" a portion of the inserted nucleic acid reagent.

Such methods take advantage of nucleic acid reagents that have been specifically designed with known target nucleic acid sequences at or near a nucleic acid or genomic region of interest. Popping out typically leaves a "foot print" of left over sequences that remain after the recombination event. The left over sequence can disrupt a gene and thereby reduce or eliminate expression of that gene. In some embodiments, the method can be used to insert sequences, upstream or downstream of genes that can result in an enhancement or reduction in expression of the gene. In certain embodiments, new genes can be introduced into the genome of a host cell or organism using similar homologous recombination methods. An example of a yeast recombination system using the *URA3* gene and 5-FOA is described herein.

One method for genetic modification is described by Alani et al. ("A method for gene disruption that allows repeated use of URA3 selection in the construction of multiply disrupted yeast strains", *Genetics* 116(4):541-545 August 1987). The original method uses a *URA3* gene cassette with 1000 base pairs (bp) of the same nucleotide sequence cloned in the same orientation on either side of the *URA3* cassette. Targeting sequences of about 50 bp are added to each side of the construct. The double-stranded targeting sequences are complementary to sequences in the genome of the host cell or organism. The targeting sequences allow site-specific recombination in a region of interest. A modification of the original technique replaces the two 1000 bp sequence direct repeats with two 200 bp direct repeats. The modified method also uses 50 bp targeting sequences. The modification reduces or eliminates recombination of

a second knock out into the 1000 bp repeat left behind in a first mutagenesis, therefore allowing multiply knocked out yeast. Additionally, the 200 bp sequences used in the method are uniquely designed, self-assembling sequences that leave behind identifiable footprints. The technique used to design the sequences incorporate design features such as low identity to the yeast genome, and low identity to each other. Therefore, a library of the self-assembling sequences can be generated to allow multiple knockouts in the same organism, while reducing or eliminating the potential for integration into a previous knockout.

Fig. 2 is a diagrammatic illustration of an exemplary gene knock out cassette. In this example, two slightly different cassettes are shown for use in separately disrupting each of the two *FAT1* alleles in a diploid yeast such as *Candida viswanathii*. The two cassette-containing nucleic acid segments are referred to as "Deletion 1" and "Deletion 2," respectively, in the figure. Each cassette contains a *URA3* gene including a URA3 promoter (P_{URA3}) and terminator (T_{URA3}). The complete *URA3* expression cassette provides for expression of orotidine-5'-monophosphate (OMP) dicarboxylase in a Ura^- host cell into which the cassette has integrated, and yields a prototrophic transformant that can be selected for by growth in uracil-free media. Integration into the *FAT1* locus by a first crossover event is provided for by the presence of sequences located on either side of the cassette that are homologous to sequences in the target locus (e.g., *FAT1*). Additionally, each cassette contains a repeat of the P_{URA3} sequence immediately downstream of the terminator sequence. This repeat sequence can be used in a second recombination event that results in the looping out of the *URA3* gene sequence which is facilitated by growth of the prototrophic transformants in the presence of 5-FOA yielding a ura^- auxotroph. All or a portion of the P_{URA3} sequence repeat remains in the genome and disrupts the *FAT1* gene allele such that it no longer yields a functional gene product. The heterozygous transformant can then be transformed with the second cassette (e.g., Deletion 2) and undergo the same two crossover events to yield a homozygous Ura^- cell. The two separate deletion cassette-containing fragments differ in the sequences of the target gene that they contain on each side of the *URA3* cassette which results in integration into different positions in the target gene.

The URA3 cassette makes use of the toxicity of 5-FOA in yeast carrying a functional URA3 gene. Uracil synthesis-deficient host yeast are transformed with the modified URA3 cassette, using standard yeast transformation protocols, and the transformed cells are plated on minimal media minus uracil. In some embodiments, PCR can be used to verify correct insertion into the

region of interest in the host genome, and in certain embodiments the PCR step can be omitted. Inclusion of the PCR step can reduce the number of transformants that are counter selected to "pop out" the URA3 cassette. The transformants (e.g., all or the ones determined as correct by PCR, for example) can then be counter-selected on media containing 5-FOA, which will select for recombination events looping out the URA3 cassette, thus rendering the yeast Ura$^-$ again, and resistant to 5-FOA toxicity. Targeting sequences used to direct recombination events to specific regions are presented herein. A modification of the method described above can be used to integrate genes into the chromosome in which, following recombination, a functional gene is left in the chromosome next to a, e.g., 200-bp, footprint. Such methods provide for addition of a desired nucleic acid into the host genome in combination with disruption of an endogenous nucleic acid.

Fig. 3A, Fig. 3B, and Fig 3C show diagrammatic illustrations of a knock-in gene disruption method which disrupts one target gene ("GOI1") and also adds a desired gene of interest ("GOI2") at the disrupted locus. As shown in Fig. 3A, the basic URA3 disruption cassette can be the same as that described in Fig. 2; however, there is an additional expression cassette immediately downstream of the second P_{URA3} repeat sequence. This expression cassette contains the gene of interest, GOI2, for adding to the endogenous GOI1 locus and includes a promoter (P_{XXX}) and terminator (T_{XXX}) for controlling transcription of GOI2. Immediately upstream of the first P_{URA3} sequence is a sequence of nucleotides of the GOI1 gene, and immediately downstream of the terminator (T_{XXX}) for GOI2 is another sequence of the GOI1 gene. These sequences are for use in integration of the cassettes into the GOI1 locus. Fig. 3B shows the locus after the first crossover event. These transformants are selected for growth in uracil-free media. In order to remove the *URA3* gene and regenerate an auxotrophic cell that can be further modified using the *URA3* marker method, the transformants are grown in the presence of 5-FOA to facilitate the second crossover event. The result of that event is shown in Fig. 3C which depicts the P_{URA3} sequence that remains followed by a functional GOI2 cassette.

Protein engineering methods

As described herein, one method of altering carbon flux in cells and organisms is to modify one or more activities involved in carbon processing in cells. These activities can be modified by altering one or more elements directly and/or indirectly involved in the activities. Such elements include, but are not limited to, nucleic acids (e.g., transcription regulatory elements, addition

and/or deletion of nucleic acids), peptides (e.g., signal peptides regulating protein localization in cells) and polypeptides (e.g., enzymes regulating reactions in metabolic pathways). Peptides and polypeptides can be modified in multiple ways, including, for example, alteration of the primary structure (i.e., amino acid sequence), secondary structure, post-translational chemical modification (e.g., phosphorylation, acylation, glycosylation) and processing (e.g., proteolytic cleavage). Many protein modifications can be achieved through alteration of the nucleic acid encoding the protein in a cell. Alteration of the nucleic acid coding sequence can result in alteration of the amino acid sequence which in turn can modify the intra- and inter-polypeptide interactions of the encoded protein. Such alterations can thus result in modification of the activity of the polypeptide and the activity of any metabolic processes in which it may participate.

In some embodiments of the cells, organisms, compositions and methods provided herein, a modified polypeptide can be expressed in a cell or organism by introducing a modification into nucleic acid encoding the polypeptide in the cell or organism. Modified polypeptides often have an activity different than the activity of an unmodified counterpart. A modified activity sometimes is a different transport activity, a different catalytic activity, a different substrate specificity, or a different catalytic activity and a different substrate specificity. A different activity sometimes is an activity that is higher (e.g., increased activity) or lower (e.g., decreased activity) than the activity of an unmodified counterpart polypeptide. In some embodiments, the catalytic activity of a modified polypeptide is higher or lower than the catalytic activity of the unmodified counterpart for a particular substrate. In certain embodiments, the specificity of a modified polypeptide for a particular substrate is higher or lower than the specificity of the unmodified counterpart for a particular substrate. A modified polypeptide often is active and an activity of a modified polypeptide often can be detected (e.g., substrate turnover can be detected). An activity for a particular polypeptide that is modified sometimes is referred to as a "target activity." As described herein, target activities include, but are not limited to, activities of ω-oxidation, β-oxidation, acetyl-CoA processing, carnitine/acetylcarnitine shuttle, membrane transport, fatty acid biosynthesis, acyl-CoA formation/degradation. Non-limiting examples of particular target activities include carnitine acetyltransferase, carnitine translocase, acetyl-CoA carboxylase, ATP citrate lyase, acetyl-CoA hydrolase, acetyl-CoA synthetase, thioesterase, acyl-CoA synthetase, monooxygenase, cytochrome P450 reductase, alcohol dehydrogenase, alcohol oxidase, aldehyde dehydrogenase, acyl-CoA oxidase, 3-ketoacyl-CoA thiolase, peroxisomal transporter, peroxisome biogenesis factor and multifunctional enzyme (e.g., enoyl-CoA hydratase and/or 3-

hydroxyacyl-CoA dehydrogenase) activities. In some of the embodiments provided herein, these and other activities can be modified in a cell or organism.

One or more particular modifications can be selected to generate a modified polypeptide having a target activity. Modifications often are amino acid modifications (e.g., deletion, insertion of one or more amino acids). Amino acid modifications sometimes are amino acid substitutions. Amino acid substitutions sometimes are conservative, non-limiting examples of which include substitution of an amino acid containing an acidic moiety for another amino acid containing an acidic moiety (e.g., D, E), substitution of an amino acid containing a basic moiety for another amino acid containing a basic moiety (e.g., H, K, R), substitution of an amino acid containing an aliphatic chain moiety for another amino acid containing an aliphatic chain moiety (e.g., V, L, I, A), substitution of an amino acid containing a cyclic moiety for another amino acid containing a cyclic moiety (e.g., W, F, Y), and substitution of an amino acid containing a polar moiety for another amino acid containing a polar moiety (e.g., S, T). Amino acid substitutions sometimes are non-conservative, non-limiting examples of which include substitution of an amino acid containing an acidic moiety for an amino acid containing a basic moiety, substitution of an amino acid containing a basic moiety for an amino acid containing an acidic moiety, substitution of an amino acid containing relatively small moiety (e.g., G, A) for another amino acid containing a relatively large moiety (e.g., Y, W, F, I, L), and substitution of an amino acid containing a relatively large moiety for another amino acid containing an relatively small moiety.

Particular modifications can be selected using any suitable method known in the art. In certain embodiments, a reference structure is known for a related polypeptide with a known activity, and modifications to a target polypeptide can be guided by alignment of the target polypeptide structure to the reference structure. A reference structure sometimes is a primary structure (e.g., polynucleotide or polypeptide sequence) and the primary structure of a target can be aligned to the reference structure using an alignment method known in the art. Particular amino acids in the target that align with (e.g., are identical to or homologous to) or do not align with (e.g., are not identical to or not homologous to) particular amino acids in the reference can be selected for modification. Selections can be made by inspection of an alignment or by software known in the art that identifies, scores and/or ranks amino acids for modification based on an alignment. A reference structure sometimes is a secondary structure, tertiary structure or quaternary structure, each of which are three dimensional structures pertaining to a polypeptide. A primary structure of a target polypeptide can be modeled to a secondary, tertiary or

quaternary reference structure using three-dimensional modeling software known in the art. A secondary, tertiary or quaternary structure of a target polypeptide can be compared to a secondary, tertiary or quaternary reference structure using three-dimensional comparative software known in the art. Particular structures (e.g., a particular individual amino acid; a particular group of contiguous or non-contiguous amino acids) in the target that align with or map to, or do not align with or map to, particular structures in the reference can be selected for modification. Also, particular structures in the target that are in proximity to a substrate or co-factor can be selected for modification. Selections can be made by inspection of an alignment or map or by software known in the art that identifies, scores and/or ranks amino acids and/or structures for modification based on an alignment and map. After particular amino acids and/or structures are selected for modification in a first polypeptide, amino acids and structures in a second polypeptide that align with the selected amino acids and structures in the first polypeptide may be identified.

For example, a structural model of a protein can be created based on the crystal structure of the protein using SWISS-MODEL, which has been described by Arnold *et al.* ((2006) *Bioinformatics* 22: 195-201), Guex *et al.* ((2009) *Electrophoresis* 30 Supplement 1: S162-S173) and Kiefer *et al*. ((2009) *Nucleic Acids Res.* 37 (Database issue): D387-D392). As described herein, the resulting structural model can be analyzed to identify sites in the protein that potentially participate in determining an activity of the protein. HotSpot Wizard is an example of a tool for identifying sites for engineering of substrate specificity and/or activity of enzymes using a combination of structural, functional and sequence analysis and has been described by Pavelka *et al*. ((2009) *Nucleic Acids Res*. 37 (Web Server issue): W376-W383) (see also: HotSpot Wizard 1.7 World Wide Web Uniform Resource Locator (URL) loschmidt.chemi.muni.cz/hotspotwizard/index.jsp). Identification of such sites facilitates a determination of possible amino acids to target for mutagenesis in modifying the activity of a protein (e.g., enzyme). Part of the HotSpot Wizard analysis is the identification of homologs by a BLAST search (see, e.g., Johnson *et al*. (2008) *Nucleic Acids Res*. 36 (Web Server issue): W5-W9) and their alignment using MUSCLE as described, for example, by Edgar ((2004) *BMC Bioinformatics* 5: 113 and *Nucleic Acids Res*. 32: 1792-1797). The multiple sequence alignment reveals the variety of amino acids found at each position and their relative frequency amongst all the sequences. This information can be useful in determining possible amino acid substitutions that may be made at identified sites in the protein.

In a non-limiting example, particular amino acid substitutions for a *Candida* spp. Pox5 acyl-CoA oxidase polypeptide are provided herein. For example, some substitutions were designed to modify a substrate specificity of an acyl-CoA oxidase polypeptide. As described herein, in embodiments in which the target product molecule is a six-carbon fatty acid (e.g., adipic acid) produced by β-oxidation of a longer-chain fatty acid, it is optimal to modify the activity of acyl-CoA oxidases (which can catalyze the first step in β-oxidation) in host cells or organisms such that there is little to no activity on substrates with chain lengths less than 8 carbons. Deletion of nucleic acids encoding acyl-CoA oxidases (e.g., Pox4 in *Candida viswanathii*) with relatively broad carbon-chain length specificity that are active on short-chain length substrates prevents generation of fatty acid products with fewer than eight carbon atoms (i.e., chain length shorter than C8) by peroxisomal β-oxidation. This is because the remaining acyl-CoA oxidase activity (e.g., Pox5) is specific for longer chain substrates and has low activity on substrates with carbon chain lengths less than 10. In order to increase the activity of Pox5 on substrates with a chain length of 8 carbons and thereby increase the amount of 6-carbon fatty acid target molecule products relative to 8-carbon fatty acid molecules, the Pox5 protein was subjected to engineering as described herein. Modified Pox5 proteins obtained by amino acid substitutions of the wild-type *Candida viswanathii* Pox5 protein (made via corresponding nucleotide sequence changes in the nucleic acid encoding the protein) resulted in an increased ratio of 6-carbon to 8-carbon fatty acid products in *Candida* cells expressing the modified enzymes as compared to cells expressing wild-type Pox5p (as shown by experimental results presented in the Examples herein). Using the teachings described herein, a primary structure of another acyl-CoA oxidase polypeptide can be aligned with the amino acid sequence or modeled structure of a Pox5 polypeptide and some or all amino acids of the other polypeptide that align with those selected for modification in the Pox5 polypeptide also can be selected for modification.

Additional nonlimiting examples of protein modifications that can be made in altering the carbon flux in a cell or organism include modifications to alter a substrate specificity of an acyl-CoA dehydrogenase polypeptide produced in the cell and that is involved in β-oxidation. An acyl-CoA dehydrogenase enzyme can require an NAD cofactor in carrying out a catalytic function. Sometimes a co-factor specificity of an acyl-CoA dehydrogenase is modified, and in some embodiments the modified polypeptide prefers to utilize oxygen as a co-factor.

In another non-limiting example of a protein modification designed to modify an enzyme activity, amino acid substitutions can be made to enhance or reduce regulation of the enzyme. For

example, enzymes can be regulated in a number of ways, including, for example, covalent modification of an enzyme such as phosphorylation/dephosphorylation and acetylation/deacetylation. The activity of an enzyme can be modified by altering its ability to be activated or inhibited within a cell. In one embodiment, regulation of an enzyme by phosphorylation can be decreased or eliminated by modifying a nucleic acid encoding the enzyme to substitute codons for phosphorylatable amino acid residues (e.g., serine) with codons for non-phosphorylatable residues (e.g, alanine). Computer-assisted software programs are available for identifying potential phosphorylatable amino acid residues (see, e.g., NetPhos (World Wide Web Uniform Resource Locator (URL) cbs.dtu.dk/services/NetPhos/), NetPhosYeast (World Wide Web Uniform Resource Locator (URL) cbs.dtu.dk/services/NetPhosYeast/)). In an example described herein, an acetyl-CoA carboxylase protein (e.g., Acc1 of *Candida viswanathii*) is modified to reduce regulation of the enzyme by phosphorylation. Because the dephosphorylated state is the active state of the enzyme, the protein (and nucleic acid encoding it) was modified to eliminate one or more phosphorylatable serine residues by substituting them with alanine residues, thereby relieving the regulation by phosphorylation. For example, as described herein, a *Candida viswanathii* acetyl-CoA carboxylase endogenous, wild-type enzyme was modified to substitute alanine residues for one or more of the following serine amino acid residues: S652, S1131, S1138, S1153, S1158.

One or more activities of a modified polypeptide can be characterized using any suitable assay known in the art. A modified polypeptide can be expressed in a cell or organism other than a target organism in which a target product will be produced, for assaying activity. For example, a modified polypeptide can be expressed in a bacterium (e.g., *E. coli*), assayed and then introduced into a yeast (e.g., *Candida* spp. yeast) for production of a target molecule product.

Engineered carbon flux pathways for efficient production of target molecules

Provided herein are multiple compositions for, and methods of, modifying cells and organisms to alter carbon flux. Also provided are the modified cells and organisms generated by the methods. The modification methods can be combined in a number of ways as described herein to engineer cell- or organism-based systems for enhanced, efficient production of target molecules. Also provided herein are methods of producing target molecules, including, for

example, organic acids, polyketides and terpenes, using the modified cells or organisms provided herein.

Included in the cells, organisms, compositions and methods provided herein are modified cells and organisms in which carbon processing activities have been engineered to enhance carbon flow through cellular oxidative metabolism pathways. One advantage of such modified bioproduction systems is that they are well suited for use with lower cost, alternative carbon sources, including, for example, non-carbohydrate and non-fermentable carbon sources such as aliphatic compounds and hydrocarbons (e.g, alkanes, fatty acids and fatty alcohols). Use of such carbon sources is not only more cost-effective but can also have the added advantage of reducing the environmental impact of harmful wastes (e.g., agro-industrial by-products, waste cooking oil and waste motor oil) that can be used as feedstocks in target molecule production instead of being discarded. As also described herein, embodiments of the cell- and microbial-based systems in which carbon processing activities have been engineered to direct carbon flow through oxidative metabolism can be controlled to provide for maximal, coordinated and highly efficient target molecule production based on, for example, use of carbon source-dependent transcription regulation.

Figs. 5-11 schematically illustrate non-limiting embodiments of engineered carbon flux pathways of modified cells and organisms that can be used to produce a target molecule (e.g., adipic acid, malonyl-CoA, 3-hydroxypropionic acid, polyketide, triacetic acid lactone, terpene) from various starting carbon sources or feedstocks.

Engineered carbon recycling loop pathways for a platform target molecule production system

In order to minimize carbon loss and increase process efficiency of biological cell-based production systems, provided herein are cells and organisms (e.g., microorganisms) that have been modified to reduce, inhibit, slow and/or delay carbon flow into one or more growth and/or energy production metabolic pathways so that it is available for use in other inherent and/or engineered production processes. In doing so, carbon that would be lost to metabolic pathways uninvolved in target molecule production in an unmodified cell or organism are rescued or captured for use in target molecule production processes. As such, modified cells and organisms provided herein in some embodiments are useful as platform systems (as well as

production systems) that can be used as the basis for further engineering for enhanced production of many different desired target molecules either singly or multiply in co-production cell- and microbial-based systems.

Engineered pathways for capturing carbon atoms expelled from the β-oxidation pathway

Figs. 5 and 6 depict possible cellular modifications in exemplary embodiments of a eukaryotic (i.e., yeast in this example) platform system designed to capture carbon atoms in the cytosol by enhancing carbon flow through cellular oxidative metabolism pathways (ω-oxidation and peroxisomal β-oxidation) and reducing flow of carbon into mitochondria, and the endoplasmic reticulum and lipid particles (in the form of acyl-CoA). Although multiple, possible, cellular modifications are illustrated in Figs. 5 and 6, as described herein, some of the modifications depicted in the figures are optional enhancements of exemplary engineered systems and may or may not be included in a modified cell or organism depending on, for example, the intended use of the system (e.g., development of a particular single, or multiple, target molecule(s) production system) and the selection of variable features (e.g., host cell or organism, carbon source, regulatory controls (such as transcription control elements), culture conditions and the like) of the system. Thus, it is understood that any optional modifications set forth in the exemplary systems shown in Figs. 5 and 6 are non-limiting and may or may not be included in a particular engineered system and, if included, may be in utilized in different combinations than illustrated in the figures.

Acetyl-CoA is a central molecule in the biochemical network of a cell that can be utilized for the biosynthesis of many useful chemicals. The β-oxidation pathway produces acetyl-CoA through the oxidation of fatty acids. In yeast, β-oxidation can be localized to the peroxisomal compartment which thus can be a primary location of fatty acid-derived acetyl-CoA. Generally, the peroxisomal acetyl-CoA would be converted to acetyl-carnitine by carnitine acetyltransferase (Cat2p). The acetyl-carnitine, being smaller, can diffuse out of the peroxisome and be transported across the mitochondrial inner membrane by Crc1p (a translocase protein). Once inside the mitochondria, the acetyl-carnitine is converted back into acetyl-CoA by mitochondrial Cat2p and can be used in the TCA cycle for energy generation or the synthesis of other biomolecules (Fig. 4). To take advantage of the acetyl-CoA generated by peroxisomal β-oxidation, genetic engineering strategies may be employed to reroute the carbon

in acetyl-CoA destined for the mitochondria to the cytosol instead, thereby making it available for use in biosynthetic pathways making desired chemical products. Thus, the amount of carbon that is lost to the TCA and lipid-generating (e.g., glycerol-3-phosphate (G3P) and/or dihydroxyacetone phosphate (DHAP)) pathways at the expense of target molecule production is reduced in this platform system. In embodiments of a platform system such as, for example, systems shown in Figs. 5 and 6, carbon processing activities can be engineered to enhance carbon flow through cellular oxidative metabolism pathways, e.g., ω-oxidation and peroxisomal β-oxidation, and decrease carbon flow to mitochondria and other organelles. Such platform systems can include an acetyl group carbon recycle loop that diverts acetyl moieties generated in the breakdown of fatty acids in peroxisomal β-oxidation into cytosolic fatty acid synthesis to regenerate a fatty acid that can be subjected to another cycle of peroxisomal β-oxidation. The recycle loop is depicted in Figs. 5 and 6 by the dark, solid reaction arrows beginning with extracellular fatty acid internalization in the upper left corner of the figure, extending through ω-oxidation and into the peroxisome for β-oxidation which yields acetyl-CoA that is transported out of the peroxisome and into the cytosol (initially in the form of either acetyl-carnitine or acetate), utilized in fatty acid synthesis to generate acyl-CoA which is then hydrolyzed to free fatty acid for re-entry into the loop at the starting point of ω-oxidation.

As shown in Figs. 5 and 6, carbon flux through the ω-oxidation and β-oxidation (peroxisomal) pathways can be enhanced through one or more of multiple modifications introduced via genetic manipulation of the cell. The enhancements can begin with the cellular internalization of external carbon. A non-fermentable or alternative carbon source (e.g., fatty acids, alkanes) enters the modified cell through the plasma membrane from the extracellular medium (shown in the upper left corner of Figs. 5 and 6). In an unmodified cell, a long-chain fatty acid (either as the carbon source or generated from processing of a carbon source, e.g., alkane) would be activated (through thioesterifcation with CoA) to acyl-CoA upon cell entry by acyl-CoA synthetase (encoded, e.g., by *FAA1* and/or *FAT1*). However, in some embodiments, a gene(s) encoding cytosolic and/or membrane-bound acyl-CoA synthetase can be disrupted or deleted resulting in a decrease or elimination of cytosolic and/or membrane-bound acyl-CoA synthetase and, thus, cytosolic and/or membrane-bound acyl-CoA synthetase activity in the cytosol. Cytosolic activation of fatty acids can thus also be decreased or eliminated in such modified cells (indicated in Figs. 5 and 6 as a lightly shaded dotted line reaction arrow labeled as “faa1Δ” and “fat1Δ” in blackened ovals). Most fatty acid metabolic pathways, including lipid (e.g, triacylglycerides (TAG) and phospholipids (PL)) biosynthesis and protein acylation, require that

a free fatty acid be activated to acyl-CoA (or to acyl-ACP) prior to being metabolized. Therefore, in embodiments that include an enhancement such as a reduced or abolished acyl-CoA synthetase activity, the decreased cytosolic fatty acid activation can result in fewer internalized fatty acid carbons being lost to such pathways at the expense of target molecule-producing processes.

Free fatty acids that have entered the cell (shown as "FA" in Figs. 5 and 6), or generated from metabolism of an alkane carbon source, can then undergo oxidation to dicarboxylic acids (DCA) through ω-oxidation ("ω-ox" in Figs. 5 and 6). The availability of this oxidative process in the cell presents multiple advantages in these platform systems for target molecule production. For example, long-chain fatty acids that have not been activated to acyl-CoA do not readily cross the peroxisomal membrane; however, long-chain dicarboxylic acids are able to enter peroxisomes. Therefore, conversion of free fatty acids to DCA through ω-oxidation can be a further enhancement of carbon flow toward peroxisomal β-oxidation, particularly because there are no or limited other pathways in the cell for processing of free dicarboxylic acids. The availability of ω-oxidation-processing of free fatty acids in the modified cell is also beneficial to engineered production systems in which the target molecule (or an intermediate in target molecule production) is a dicarboxylic acid (e.g., adipic acid, suberic acid, sebacic acid, dodecanedioic acid, tetradecanedioic acid). For example, as shown in Figs. 5 and 6, under certain conditions, a dicarboxylic acid processed in β-oxidation can be converted into a shorter chain diacid which can be secreted from the cell as a target molecule upon removal of the coenzyme A carrier via hydrolysis catalyzed by peroxisomal thioesterase. Thus, the ω-oxidation pathway in a modified cell can serve as a cellular gateway for funneling internalized fatty acids into oxidative metabolism and target molecule production and away from cytosolic activation. Modification of ω-oxidation activity is another potential enhancement of these embodiments. For example, if a host cell or organism (e.g., *Candida* spp, *Yarrowia* spp, *Bacillus* spp, *Blastobotrys* spp) expresses an endogenous ω-oxidation pathway, one or more enzymes (e.g., monooxygenase, cytochrome P450 reductase, such as CPRB, and others) of the pathway can be modified (e.g., as described herein) to increase catalytic activity and/or alter substrate specificity in order to increase fatty acid processing in the pathway and/or target specific fatty acids for processing into dicarboxylic acids. If a host cell or organism does not express an endogenous ω-oxidation pathway, it can be genetically modified to express heterologous enzymes to engineer an ω-oxidation pathway in the cell or organism.

In an oxidative metabolism-enhanced platform system, dicarboxylic acids, such as those generated by ω-oxidation, can traverse the peroxisomal membrane and move into peroxisomes where they can be activated via thioesterification to a dicarboxylic acid ester (shown as DCA-CoA in Figs. 5 and 6) and enter β-oxidation ("β -ox" in Figs. 5 and 6). In each cycle of β-oxidation, fatty acids are degraded through removal of two carbons from the carbon chain which are released as acetyl-CoA. The remaining fatty acid carbon chain can reenter another cycle of oxidation as an acyl-CoA shortened by two carbons atoms. Through successive cycles, a monocarboxylic fatty acid can be completely degraded such that only acetyl-CoA (for fatty acids with an even number of carbon atoms in the chain) or propionyl-CoA (for fatty acids with an odd number of carbon atoms in the chain) remains. Through successive cycles of β-oxidation of a dicarboxylic acid, the molecule can be completely degraded such that only succinyl-CoA (for fatty diacids with an even number of carbon atoms in the chain) or malonyl-CoA (for fatty diacids with an odd number of carbon atoms in the chain) remains. Thus, in these platform systems, the enhanced movement of fatty acids toward degradation via β-oxidation can yield acetyl-CoA (which can be used in target molecule production), and, in certain instances as described herein, shorter chain diacids at the completion of the oxidative process. A short-chain diacid thus produced can be a final target molecule (or a precursor or intermediate in the production of a target molecule).

The oxidative metabolism aspect of some platform systems, such as those shown in Figs. 5 and 6, can be further enhanced through modification of β-oxidation activity. For example, one or more enzymes (e.g., acyl-CoA oxidase, ketoacyl-CoA thiolase, multifunctional enzyme hydratase and/or dehydrogenase, and others) of the pathway can be modified (e.g., as described herein) to increase catalytic activity and/or alter substrate specificity in order to increase fatty acid processing in the pathway and/or select for specific fatty diacids for processing into target dicarboxylic acids. One example of a modification of β-oxidation activity, as described herein, is alteration of the substrate specificity of one or more acyl-CoA oxidase enzymes in the pathway, such as Pox4 and/or Pox5 of *Candida* yeast strains. In so doing, the process can be optimized for the production of fatty diacids of particular carbon chain lengths. For example, by genetically modifying a host cell or microorganism to decrease or eliminate Pox4 expression and/or activity in the host (e.g., *Candida*), the amount of shorter-chain (e.g., having less than about 8-10 carbons) fatty acids or diacids resulting from β-oxidation of longer chain fatty acids can be increased. Production of fatty acids or diacids of particular lengths can also be enhanced by genetically modifying (e.g., mutagenesis of the gene coding sequence to

alter the encoded amino acid sequence) the activity of another acyl-CoA oxidase, such as Pox5, to alter the substrate specificity. For example, as described herein, some alterations of a *Candida* Pox5 amino acid sequence increase activity of the enzyme on C8 substrates and provide for a relative increase in the amount of C6 diacid (adipic acid) produced and decrease in the amount of C8 and longer diacids resulting from β-oxidation of a longer chain fatty acid. Thus, the platform system shown in Figs. 5 and 6 can also serve as a production system for diacids of particular carbon chain lengths. Additional optional modifications that can provide for enhanced carbon flux through β-oxidation (and enhanced target molecule production) in these systems include, but are not limited to, modification of β-oxidation-associated activities, such as peroxisome biogenesis and proliferation activities. For example, as described herein, the abundance and/or volume of peroxisomes in which β-oxidation occurs can be increased in host cells through genetic modification. An example of such a modification is increasing the transcription of, and/or number of copies of, one or more peroxin-encoding nucleic acids (e.g., *PEX11*) in a host cell. Amplification of such peroxin-encoding nucleic acids and/or activities can lead to an overall increased β-oxidation capacity.

One feature of the carbon recycle loops of the platform systems shown in Figs. 5 and 6 is the management and capture of acetyl-CoA generated during β-oxidation. Peroxisomal acetyl-CoA generally has two main fates: (i) conversion to acetyl-carnitine for transfer to mitochondria for use in the TCA cycle and (ii) the generation of malate in the glyoxylate cycle ("GlyOx" in Figs. 5 and 6) which is then used in gluconeogenesis or moves into mitochondria. In unmodified cells, these uses of acetyl-CoA generated in β-oxidation represent loss of carbon atoms that could be used in target molecule production. Through modifications that are a part of the platform systems shown in Figs. 5 and 6, acetyl groups can be captured either (1) as they move through the cytosol toward the mitochondria in the form of acetyl-carnitine or (2) in the form of acetate generated in peroxisomes. In capturing these acetyl group carbons, they thus can be diverted from the TCA cycle.

Capture of carbon from acetyl-carnitine

In the example platform system depicted in Fig. 5, modifications in the host cell that enhance the capture and diversion of acetyl group carbon include, but are not limited to, modification of acetyl-carnitine entry into mitochondria, and modification of conversion of cytosolic acetyl-carnitine to acetyl-CoA. As described herein, unmodified cells may contain a cytosolic carnitine

acetyltransferase activity for conversion of cytosolic acetyl-carnitine to acetyl-CoA. However, in some instances, it may not be as catalytically active and/or abundant as it is in organelles, e.g., peroxisomes and mitochondria. In a platform system provided herein and depicted in Fig. 5, one modification that can be made to enhance capture of acetyl groups in the cytosol is to increase the amount and/or activity of cytosolic carnitine acetyltransferase. As described herein, methods of achieving this include increasing the copy number of nucleic acids encoding cytosolic carnitine acetyltransferase in the cell, increasing the transcription of such nucleic acids and/or introducing nucleic acid encoding a more active cytosolic carnitine acetyltransferase enzyme into the cell (e.g., modifying an endogenous cytosolic enzyme activity by replacing it with, or adding to it, a heterologous enzyme activity). For example, in one embodiment described herein, a *Candida* mitochondrial/peroxisomal carnitine acetyltransferase (e.g, Cat2) with greater catalytic activity than an endogenous *Candida* cytoplasmic carnitine acetyltransferase (e.g., Yat1) can be recombinantly expressed cytosolically in a host cell by engineering a nucleic acid encoding the more active enzyme such that the encoded enzyme lacks a mitochondrial (and a peroxisomal) targeting sequence of amino acids (shown, as $CAT2^{cyt}$ in Fig. 5; see also, e.g., amino acid SEQ ID NO: 4 and a nucleotide sequence (SEQ ID NO: 61) encoding the amino acid sequence). Once acetyl-carnitine in transit from the peroxisomes to the mitochondria has been converted to acetyl-CoA in the cytosol by carnitine acetyltransferase activity in the cytosol (e.g., $CAT2^{cyt}$), it cannot cross the mitochondrial inner membrane and is diverted from the TCA cycle. This acetyl-CoA is now available for use in target molecule production. The amount of carnitine acetyltransferase activity in the cytosol of such a modified cell or organism can be further increased by using a strong and/or fatty acid-inducible heterologous promoter (e.g., a yeast *HDE* gene promoter) to regulate transcription of the engineered nucleic acid encoding a carnitine acetyltransferase activity.

Another modification that can enhance cytosolic capture and diversion of acetyl moieties in cells is an alteration of acetyl-carnitine uptake into mitochondria from the cytosol (shown as faded, dotted reaction arrow lines into and in the mitochondrial compartment in Fig. 5). One method of modifying mitochondrial acetyl-carnitine uptake is by altering the processing of acetyl-carnitine that occurs in the mitochondria to convert it to acetyl-CoA for use in the TCA cycle. For example, by decreasing the amount and/or activity level of the enzyme that catalyzes this processing, i.e., mitochondrial carnitine acetyltransferase, there can be a corresponding decrease in conversion of acetyl-carnitine to acetyl-CoA in the mitochondria. Without being limited or bound by theory, this can introduce a bottleneck in acetyl-carnitine processing in the

mitochondria which slows acetyl-CoA entry into the TCA cycle. If the mitochondrial carnitine acetyltransferase activity is not sufficient to handle the acetyl carbon flux coming from the peroxisome, then the cytoplasmic acetyl-carnitine concentration should build up and, in effect, acetyl-carnitine is diverted from the TCA cycle. The increased concentration of cytoplasmic acetyl-carnitine can thus be a source of substrate for carnitine acetyltransferase activity in the cytosol which converts the substrate to cytosolic acetyl-CoA for use in target molecule production. The amount and/or activity of mitochondrial carnitine acetyltransferase can be decreased in a number of ways, as described herein. For example, the number of copies of nucleic acid encoding the enzyme in a host cell can be reduced (e.g, an endogenous gene encoding the enzyme can be disrupted or deleted), the transcription of such nucleic acid can be decreased and/or nucleic acid encoding a less active mitochondrial carnitine acetyltransferase enzyme can be introduced into the cell (i.e., replacing the endogenous mitochondrial enzyme with a heterologous enzyme). For example, in one embodiment described herein, a *Candida* cytoplasmic carnitine acetyltransferase (e.g, Yat1) which is less active than an endogenous *Candida* mitochondrial carnitine acetyltransferase (e.g., Cat2) can be recombinantly expressed in a host cell mitochondria by engineering a nucleic acid encoding the less active enzyme such that the encoded enzyme includes a mitochondrial targeting sequence of amino acids (shown as "*CAT2*" in a diagonal line-hatched background in the mitochondria in Fig. 5; see also, e.g., amino acid SEQ ID NOS: 10, 11 and 12 and nucleotide SEQ ID NOS: 67, 68 and 69 encoding such amino acid sequences). The modified nucleic acid can be introduced into a host cell in which the endogenous mitochondrial carnitine acetyltransferase gene has been disrupted or deleted. Although not specifically indicated in Fig. 5, in some cells and organisms (e.g., *Candida* spp), an endogenous gene encoding a mitochondrial carnitine acetyltransferase may also encode the cell's peroxisomal carnitine acetyltransferase. For example, such a gene can encode an enzyme that includes mitochondrial and peroxisomal targeting sequences for localization to each of these areas of the cell. If the gene encoding an endogenous mitochondrial carnitine acetyltransferase in such a cell is disrupted or deleted, it may be optimal (e.g., for cell viability and/or efficient processing of peroxisomal acetyl-CoA) to introduce a heterologous nucleic acid encoding carnitine acetyltransferase that includes a peroxisomal targeting sequence into the cell.

Another method of modifying mitochondrial acetyl-carnitine uptake is by altering a transport mechanism that moves acetyl-carnitine into the mitochondrial matrix, e.g., an acetyl-carnitine translocase. A mitochondrial inner-membrane transport protein (e.g., Crc1p) may function as an

DEMANDE OU BREVET VOLUMINEUX

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

CECI EST LE TOME 1 DE 3
CONTENANT LES PAGES 1 À 261

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 3
CONTAINING PAGES 1 TO 261

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

NOM DU FICHIER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

What is claimed is:

1. A genetically modified microorganism, comprising:

one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides, wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide:

- wherein the one or more heterologous nucleic acids that encode one or more terpene biosynthesis polypeptides is chosen from terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase and valencene synthase, or
- wherein the one or more heterologous nucleic acids is a cytochrome P450 reductase and is regulated by a nucleic acid that provides for alkane induction of expression of the cytochrome P450 reductase.

2. The microorganism of claim 1, wherein the microorganism is a fungus.

3. The microorganism of claim 2, wherein the fungus is a yeast.

4. The microorganism of claim 3, wherein the yeast is chosen from *Candida* spp, *Yarrowia* spp, *Rhodotorula* spp, *Rhodosporidium* spp, *Cryptococcus* spp, *Trichosporon* spp, *Lipomyces* spp, and *Blastobotrys* spp.

5. The microorganism of claim 4, wherein the fatty acid is a saturated fatty acid or an unsaturated fatty acid.

6. The microorganism of claim 5, wherein the fatty acid is chosen from one or more of oleic acid, palmitoleic acid, erucic acid, linoleic acid, palmitic acid, caproic acid, enanthic acid, caprylic acid pelargonic acid, capric acid, undecylic acid, lauric acid, myristic acid, pentadecanoic acid, margaric acid, stearic acid arachidic acid, behenic acid, tridecylic acid, and linolenic acid.

7. The microorganism of claim 4, wherein the alkane is chosen from one or more of hexane, heptane, nonane, decane, undecane, dodecane, tridecane, tetradecane, pentadecane, hexadecane, heptadecane, and octadecane

8. The microorganism of claim 7, wherein the nucleic acid that provides for fatty acid or alkane induction of expression of a terpene biosynthesis polypeptide comprises a fatty acid response element or an alkane response element.

9. The microorganism of claim 8, wherein the fatty acid response element comprises an oleic acid response element.

10. The microorganism of claim 8, wherein the alkane response element comprises an alkane response element 1 (ARE1) sequence.

11. The microorganism of claim 10, wherein the nucleic acid that provides for fatty acid induction of expression of a terpene biosynthesis polypeptide comprises a promoter region chosen from promoter regions of genes encoding hydratase-dehydrogenase-epimerase (HDE), acyl co-A oxidase (POX), acyl co-A thiolase (POT), peroxin (PEX) and peroxisomal adenine nucleotide transporter protein (ANT1).

12. The microorganism of claim 11, wherein the microorganism is *Candida viswanathii*.

13. The microorganism of claim 11, wherein the microorganism is *Blastobotrys adeninivorans*.

14. The microorganism of claim 13, wherein the one or more heterologous nucleic acids encode phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase.

15. The microorganism of claim 14, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

16. The microorganism of claim 14, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

17. The microorganism of claim 14, 15 or 16, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encodingan acyl co-A oxidase 4 (POX4).

18. The microorganism of claim 14, 15 or 16, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encodinga *Candida*acyl co-A oxidase 4 (POX4).

19. The microorganism of claim 13, wherein the one or more heterologous nucleic acids encode valencene synthase.

20. The microorganism of claim 19, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

21. The microorganism of claim 19, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

22. The microorganism of claim 19, 20 or 21, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a terminator region of the gene encodingan acyl co-A oxidase 4 (POX4).

23. The microorganism of claim 19, 20 or 21, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a terminator region of the gene encodinga *Candida*acyl co-A oxidase 4 (POX4).

24. The microorganism of claim 13, wherein the one or more heterologous nucleic acids encode geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase.

25. The microorganism of claim 24, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

26. The microorganism of claim 24, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene

cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

27. The microorganism of claim 24, 25 or 26, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encodingan acyl co-A oxidase 4 (POX4).

28. The microorganism of claim 24, 25 or 26, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encodinga *Candida*acyl co-A oxidase 4 (POX4).

29. The microorganism of claim 28, wherein the one or more heterologous nucleic acids encoding the one or more terpene biosynthesis polypeptides are endogenously expressed in a microorganism chosen from *Cronobacter* spp, *Callitropsis* spp, *Xanthophyllomyces* spp, *Agrobacterium* spp, and *Pantoea* spp.

30. The microorganism of claim 29, wherein the amount and/or activity of aRas2 protein has been decreased.

31. The microorganism of claim 30, wherein the microorganism has been genetically modified to reduce or eliminate expression of an endogenous *RAS2* gene.

32. The microorganism of claim 31, wherein the amount and/or activity of an Faa1 protein has been decreased.

33. The microorganism of claim 32, wherein the microorganism has been genetically modified to reduce or eliminate expression of an endogenous *FAA1* gene.

34. The microorganism of claim 33, wherein the amounts and/or activities of one or more proteins chosen from acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, and isopentyl diphosphate delta isomerase,have been increased.

35. The microorganism of claim 34,wherein the microorganism has been genetically modified to increase expression of one or more endogenous genes chosen from *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and *IDI1*.

36. The microorganism of claim 35, wherein the amount and/or activity of dimethylallyltranstransferase or farnesyl diphosphate synthetase has been increased.

37. The microorganism of claim 36, wherein the microorganism has been genetically modified to increase expression of an endogenous *ERG20* gene.

38. The microorganism of claim 35, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1*is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

39. The microorganism of claim 35 or 38, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1*is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

40. The microorganism of claim 37, wherein expression of*ERG20*is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

41. The microorganism of claim 37 or 40, wherein expression of*ERG20*is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

42. A genetically modified microorganism, comprising:

one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides, and

a genetic modification that alters the expression of a polypeptide providing for transport of acetyl-carnitine in the microorganism,

wherein the expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for glucose, fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide.

43. The microorganism of claim 42, wherein the microorganism is a fungus.

44. The microorganism of claim 43, wherein the fungus is a yeast.

45. The microorganism of claim 44, wherein the yeast is chosen from *Candida* spp, *Yarrowia* spp, *Rhodotorula* spp, *Rhodosporidium* spp, *Cryptococcus* spp, *Trichosporon* spp, *Lipomyces* spp, and *Blastobotrys* spp.

46. The microorganism of claim 42, wherein the one or more heterologous nucleic acids encode one or more terpene biosynthesis polypeptides chosen from terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase, valencene synthase, and cytochrome p450 reductase.

47. The microorganism of claim 46, wherein the genetic modification reduces expression of the polypeptide providing for transport of acetyl-carnitine in the microorganism relative to a microorganism that does not have the genetic modification.

48. The microorganism of claim 47, wherein the genetic modification is a disruption, deletion or knockout of (i) a polynucleotide that encodes a polypeptide providing for transport of acetyl-carnitine, or (ii) a promoter operably linked to a polynucleotide that encodes a polypeptide providing for transport of acetyl-carnitine, whereby endogenous activity of a polypeptide providing for transport of acetyl-carnitine is reduced or abolished.

49. The microorganism of claim 47, further comprising a nucleic acid encoding a polypeptide providing for transport of acetyl-carnitine wherein expression of the nucleic acid encoding a polypeptide providing for transport of acetyl-carnitine is regulated by a promoter that provides for reduced expression relative to endogenous expression.

50. The microorganism of claim 49, wherein the genetic modification comprises replacing the promoter of an endogenous gene encoding the polypeptide providing for transport of acetyl-carnitine in the microorganism with a promoter that provides for reduced expression of the polypeptide in the microorganism relative to a microorganism that does not have the genetic modification.

51. The microorganism of claim 49 or 50, wherein the promoter that provides for reduced expression relative to endogenous expression is a promoter for a glucose-6-phosphate isomerase (G6PI) gene.

52. The microorganism of claim 49 or 50, wherein the promoter that provides for reduced expression relative to endogenous expression is a promoter for a *Candida* glucose-6-phosphate isomerase (G6PI) gene.

53. The microorganism of claim 42, wherein the polypeptide providing for transport of acetyl-carnitine in the microorganism is an acetyl-carnitine translocase (CRC).

54. The microorganism of claim 53, wherein the acetyl-carnitine translocase (CRC) is acetyl-carnitine translocase 1 (CRC1).

55. The microorganism of claim 42, wherein the microorganism is *Candida viswanathii.*

56. The microorganism of claim 42, wherein the microorganism is *Blastobotrys adeninivorans.*

57. The microorganism of claim 42, wherein the one or more heterologous nucleic acids encode phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase.

58. The microorganism of claim 57, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

59. The microorganism of claim 57, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

60. The microorganism of claim 57, 58 or 59, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encodingan acyl co-A oxidase 4 (POX4).

61. The microorganism of claim 57, 58 or 59, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encodinga *Candida*acyl co-A oxidase 4 (POX4).

62. The microorganism of claim 42, wherein the one or more heterologous nucleic acids encode geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase.

63. The microorganism of claim 62, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a hydratase-dehydrogenase-epimerase (HDE).

64. The microorganism of claim 62, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

65. The microorganism of claim 62, 63 or 64, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encodingan acyl co-A oxidase 4 (POX4).

66. The microorganism of claim 62, 63 or 64, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encodinga *Candida*acyl co-A oxidase 4 (POX4).

67. The microorganism of claim 42, wherein the one or more heterologous nucleic acids encoding the one or more terpene biosynthesis polypeptides are endogenously expressed in a microorganism chosen from *Cronobacter* spp, *Callitropsis* spp, *Xanthophyllomyces* spp, *Agrobacterium* spp, and *Pantoea* spp.

68. The microorganism of claim 42, wherein the amount and/or activity of aRas2 protein has been decreased.

69. The microorganism of claim 42, wherein the microorganism has been genetically modified to reduce or eliminate expression of an endogenous *RAS2* gene.

70. The microorganism of claim 42, wherein the amount and/or activity of an Faa1 protein has been decreased.

71. The microorganism of claim 42, wherein the microorganism has been genetically modified to reduce or eliminate expression of an endogenous *FAA1* gene.

72. The microorganism of claim 42, wherein the amounts and/or activities of one or more proteins chosen from acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, and isopentyl diphosphate delta isomerase, have been increased.

73. The microorganism of claim 42, wherein the microorganism has been genetically modified to increase expression of one or more endogenous genes chosen from *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and *IDI1*.

74. The microorganism of claim 42, wherein the amount and/or activity of dimethylallyltranstransferase or farnesyl diphosphate synthetase has been increased.

75. The microorganism of claim 42, wherein the microorganism has been genetically modified to increase expression of an endogenous *ERG20* gene.

76. The microorganism of claim 73, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1*is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

77. The microorganism of claim 73 or 76, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1*is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

78. The microorganism of claim 75, wherein expression of*ERG20*is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

79. The microorganism of claim 75 or 78, wherein expression of*ERG20*is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

80. A genetically modified *Candida viswanathii* yeast, comprising one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides, wherein expression of at least one of the heterologous nucleic acids is regulated by a nucleic acid that provides for glucose, fatty acid or alkane induction of expression of the terpene biosynthesis polypeptide and

- wherein the one or more heterologous nucleic acids encoding one or more terpene biosynthesis polypeptides is chosen from terpene synthase, phytoene synthase, geranylgeranyl diphosphate synthase, phytoene desaturase, lycopene cyclase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase, zeaxanthin glucosyltransferase and valencene synthase, or
- wherein the one or more heterologous nucleic acids is a cytochrome P450 reductase and is regulated by a nucleic acid that provides for glucose or alkane induction of expression of the cytochrome P450 reductase.

81. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the one or more heterologous nucleic acids encode phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase

82. The genetically modified *Candida viswanathii* yeast of claim 81, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

83. The genetically modified *Candida viswanathii* yeast of claim 82, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a terminator region of the gene encoding a *Candida*acyl co-A oxidase 4 (POX4).

84. The genetically modified *Candida viswanathii* yeast of claim 81, wherein expression of the heterologous nucleic acids encoding phytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase is regulated by a promoter region of the gene encoding a *Candida* glyceraldehyde-3-phosphate dehydrogenase (GPD).

85. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the one or more heterologous nucleic acids encode valencene synthase.

86. The genetically modified *Candida viswanathii* yeast of claim 85, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

87. The genetically modified *Candida viswanathii* yeast of claim 86, wherein expression of the heterologous nucleic acids valencene synthase is regulated by a terminator region of the gene encoding a *Candida*acyl co-A oxidase 4 (POX4).

88. The genetically modified *Candida viswanathii* yeast of claim 85, wherein expression of the heterologous nucleic acids encoding valencene synthase is regulated by a promoter region of the gene encoding a *Candida* glyceraldehyde-3-phosphate dehydrogenase (GPD).

89. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the one or more heterologous nucleic acids encode geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase.

90. The genetically modified *Candida viswanathii* yeast of claim 89, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

91. The genetically modified *Candida viswanathii* yeast of claim 90, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a terminator region of the gene encoding a *Candida*acyl co-A oxidase 4 (POX4).

92. The genetically modified *Candida viswanathii* yeast of claim 89, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase is regulated by a promoter region of the gene encoding a *Candida* glyceraldehyde-3-phosphate dehydrogenase (GPD).

93. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the one or more heterologous nucleic acids encode geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase,and cytochrome p450 reductase.

94. The genetically modified *Candida viswanathii* yeast of claim 93, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase,and cytochrome p450 reductase is regulated by a promoter region of the gene encoding a *Candida* hydratase-dehydrogenase-epimerase (HDE).

95. The genetically modified *Candida viswanathii* yeast of claim 94, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase,and cytochrome p450 reductase is regulated by a terminator region of the gene encoding a *Candida*acyl co-A oxidase 4 (POX4).

96. The genetically modified *Candida viswanathii* yeast of claim 93, wherein expression of the heterologous nucleic acids encoding geranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase,and cytochrome p450 reductase is regulated by a promoter region of the gene encoding a *Candida* glyceraldehyde-3-phosphate dehydrogenase (GPD).

97. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the one or more heterologous nucleic acids encoding the one or more terpene biosynthesis polypeptides are endogenously expressed in a microorganism chosen from *Cronobacter* spp, *Callitropsis* spp, *Xanthophyllomyces* spp, *Agrobacterium* spp, and *Pantoea* spp.

98. The genetically modified *Candida viswanathii* yeast of claim 80, further comprising a genetic modification that alters the expression of one or more nucleic acids encoding one or more endogenous polypeptides.

99. The genetically modified *Candida viswanathii* yeast of claim 98, wherein the one or more endogenous polypeptides are chosen from polypeptides having one or more of the following activities: acyl-CoA synthetase activity, acyl-CoA oxidase activity, ATP-binding cassette transporter activity, carnitine acetyltransferase activity, transport of acetyl-carnitine, acyl-CoA thioesterase activity, acyl-CoA hydrolase activity, aldehyde dehydrogenase activity, monooxygenase activity, or monooxgenase reductase activity.

100. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the amount and/or activity of aRas2 protein has been decreased.

101. The genetically modified *Candida viswanathii*yeast of claim 80, wherein the yeast has been genetically modified to reduce or eliminate expression of an endogenous *RAS2* gene.

102. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the amount and/or activity of an Faa1 protein has been decreased.

103. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the yeast has been genetically modified to reduce or eliminate expression of an endogenous *FAA1* gene.

104. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the amounts and/or activities of one or more proteins chosen from acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, and isopentyl diphosphate delta isomerase, have been increased.

105. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the yeast has been genetically modified to increase expression of one or more endogenous genes chosen from *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and *IDI1*.

106. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the amount and/or activity of dimethylallyltranstransferase or farnesyl diphosphate synthetase has been increased.

107. The genetically modified *Candida viswanathii* yeast of claim 80, wherein the yeast has been genetically modified to increase expression of an endogenous *ERG20* gene.

108. The genetically modified *Candida viswanathii* yeast of claim 105, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1* is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

109. The genetically modified *Candida viswanathii* yeast of claim 105 or 108, wherein expression of *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and/or *IDI1*is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

110. The genetically modified *Candida viswanathii*yeast of claim 107, wherein expression of*ERG20* is regulated by a promoter region of the gene encoding oleate-induced peroxisomal protein (POX18).

111. The genetically modified *Candida viswanathii* yeast of claim 107 or 110, wherein expression of*ERG20* is regulated by a terminator region of the gene encoding oleate-induced peroxisomal protein (POX18).

112. A method for producing a terpene comprising:

contacting the genetically modified microorganism of any one of claims 1 to 41 with a feedstock comprising a carbon source, and

culturing the microorganism under conditions in which the terpenes are produced from the feedstock.

113. A method for producing a terpene comprising:

contacting the genetically modified microorganism of any one of claims 42 to 79 with a feedstock comprising a carbon source, and

culturing the microorganism under conditions in which the terpenes are produced from the feedstock.

114. A method for producing a terpene comprising:

contacting the genetically modified *Candida viswanathii*yeast of any one of claims 80 to 111 with a feedstock comprising a carbon source, and

culturing the microorganism under conditions in which the terpenes are produced from the feedstock.

115. The method of claim 112, wherein the feedstock comprises one or more fatty acids.

116. The method of claim 115, wherein the feedstock comprises one or more saturated fatty acids or one or more unsaturated fatty acids; or one or more saturated fatty acids and one or more unsaturated fatty acids.

117. The method of claim 116, wherein the feedstock comprises one or more fatty acids chosen from oleic acid, palmitoleic acid, erucic acid, linoleic acid, palmitic acid, caproic acid, enanthic acid, caprylic acid pelargonic acid, capric acid, undecylic acid, lauric acid, myristic acid, pentadecanoic acid, margaric acid, stearic acid arachidic acid, behenic acid, tridecylic acid, and linolenic acid.

118. The method of claim 117 wherein the feedstock comprises oleic acid.

119. The method of claim 112, wherein the feedstock comprises one or more sugars.

120. The method of 119, wherein the feedstock comprises glucose.

121. The method of claim 112, wherein the feedstock comprises one or more alkane hydrocarbons.

122. The method of claim 121, wherein the feedstock comprises one or more alkane hydrocarbons chosen from one or more ofhexane, heptane, nonane, decane, undecane, dodecane, tridecane, tetradecane, pentadecane, hexadecane, heptadecane, and octadecane.

123. The method of claim 121, wherein the feedstock comprises one or more alkane hydrocarbons chosen from C10 to C14 alkane hydrocarbons, nonane and octadecane.

124. The method of claim 112, wherein the feedstock comprises one or more vegetable oils, derivatives thereof, or byproducts thereof.

125. The method of claim 124, wherein the feedstock comprises crude palm oil (CPO).

126. The method of claim 124, wherein the feedstock comprises palm oil fatty acid distillate (PFAD).

127. The method of claim 112, wherein the terpene is chosen from lycopene, beta carotene and astaxanthin.

128. The method of claim 127, wherein the yield or titer of the terpene is between about 0.001 g/L to about 2.0 g/L.

129. The method of claim 127, wherein the yield or titer of the terpene is between about 0.05 g/L to about 0.5 g/L.

130. The method of claim 127, wherein the yield or titer of the terpene is between about 0.001 g/L to about 0.2 g/L.

131. The method of claim 128, wherein the yield or titer of lycopene is between about 0.03 g/L to about 2.0 g/L.

132. The method of claim 129, wherein the yield or titer of lycopene is between about 0.05 g/L to about 0.5 g/L.

133. The method of claim 130, wherein the yield or titer of lycopene is between about 0.001 g/L to about 0.2 g/L.

134. The method of claim 128, wherein the yield or titer of beta carotene is between about 0.1 g/L to about 0.2 g/L.

135. The method of claim 130, wherein the yield or titer of beta carotene is between about 0.01 g/L to about 0.1 g/L.

136. The method of claim 130, wherein the yield or titer of beta carotene is between about 0.001 g/L to about 0.01 g/L.

137. The method of claim 128, wherein the yield or titer of astaxanthin is between about 0.005 g/L to about 0.01 g/L.

138. The method of claim 130, wherein the yield or titer of astaxanthin is between about 0.001 g/L to about 0.005 g/L.

139. The method of claim 112, further comprising isolating the terpene.

140. A genetically modified yeast comprising one or more heterologous nucleic acids encodingphytoene synthase, geranylgeranyl diphosphate synthase and phytoene desaturase, wherein the amount and/or activity of a Ras2 protein has been decreased.

141. A genetically modified yeast comprising one or more heterologous nucleic acids encodinggeranylgeranyl diphosphate synthase, phytoene desaturase and bifunctional lycopene cyclase/phytoene synthase, wherein the amount and/or activity of a Ras2 protein has been decreased.

142. A genetically modified yeast comprising one or more heterologous nucleic acids encodinggeranylgeranyl diphosphate synthase, phytoene desaturase, bifunctional lycopene cyclase/phytoene synthase, β-carotene ketolase, β-carotene hydroxylase, astaxanthin synthase,and cytochrome p450 reductase, wherein the amount and/or activity of a Ras2 protein has been decreased.

143. The genetically modified yeast of claim 142, wherein the yeast has been genetically modified to reduce or eliminate expression of an endogenous *RAS2* gene.

144. The genetically modified yeast of claim 142, wherein the amount and/or activity of an Faa1 protein has been decreased.

145. The genetically modified yeast of claim 142, wherein the yeast has been genetically modified to reduce or eliminate expression of an endogenous *FAA1* gene.

146. The genetically modified yeast of claim 142, wherein the amounts and/or activities of one or more proteins chosen from acetyl-CoA C-acetyltransferase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase, diphosphomevalonate decarboxylase, and isopentyl diphosphate delta isomerase,have been increased.

147. The genetically modified yeast of claim 142, wherein the yeast has been genetically modified to increase expression of one or more endogenous genes chosen from *ERG10*, *ERG13*, *HMG1*, *ERG12*, *ERG8*, *MVD1*, and *IDI1*.

148. The genetically modified yeast of claim 142, wherein the amount and/or activity of dimethylallyltranstransferase or farnesyl diphosphate synthetase has been increased.

149. The genetically modified yeast of claim 142, wherein the yeast has been genetically modified to increase expression of an endogenous *ERG20* gene.

150. The genetically modified yeast of claim 142, wherein the yeast is *Candida viswanathii*.

151. The genetically modified yeast of claim 142, wherein the yeast is *Blastobotrys adeninivorans*.

152. Use of the genetically modified yeast of any one of claims 140 to 151 for the production of a terpene.

153. Use of the genetically modified yeast of any one of claims 140 and 143-151 for the production of lycopene.

154. Use of the genetically modified yeast of claim 141 and 143-151 for the production of beta carotene.

155. Use of the genetically modified yeast of any one of claims 142-151 for the production of astaxanthin.

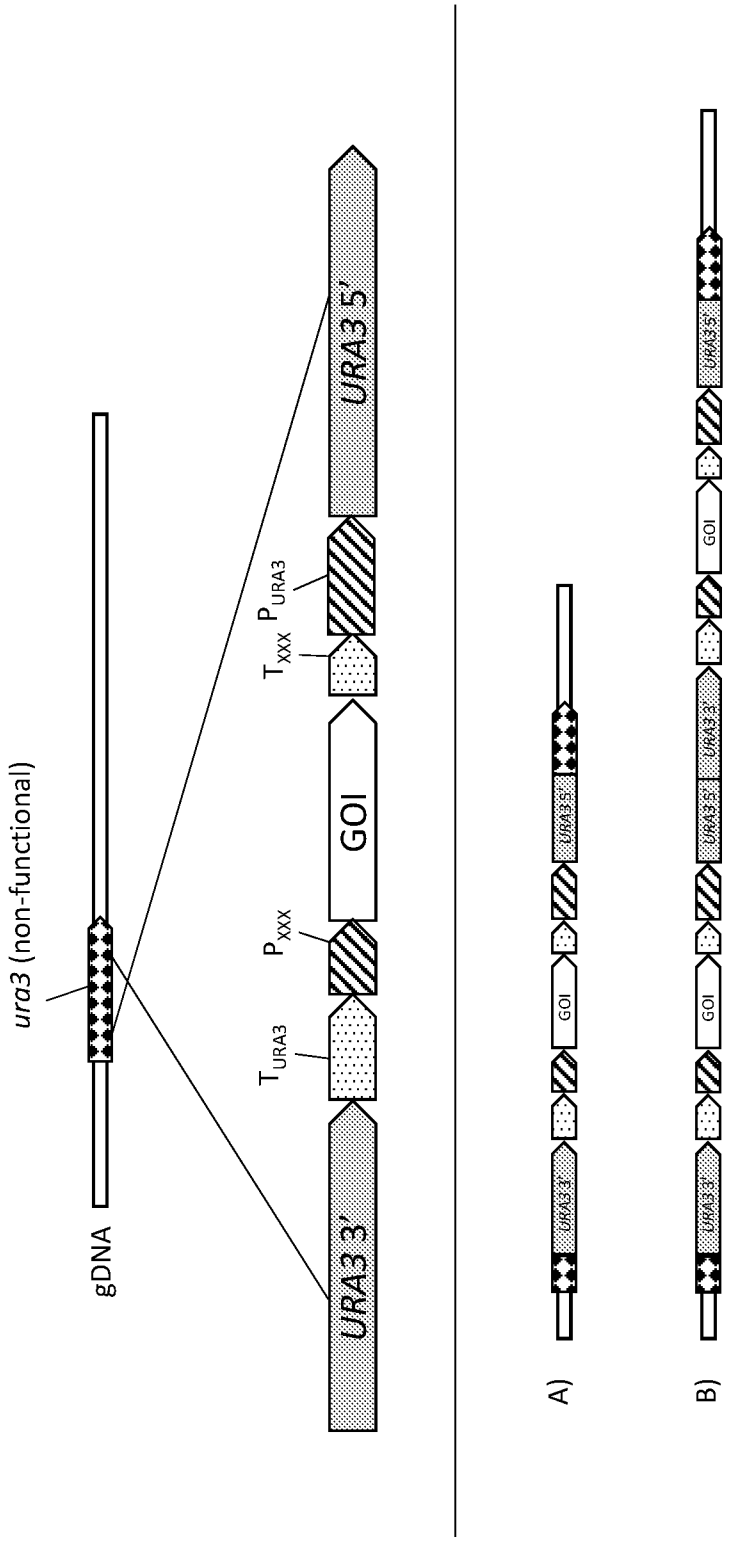

ura3 (non-functional)
gDNA
URA3 3'
T URA3
P XXX
GOI
T XXX
P URA3
URA3 5'
A)
B)
URA3 3'
GOI
URA3 5'

FIG. 1

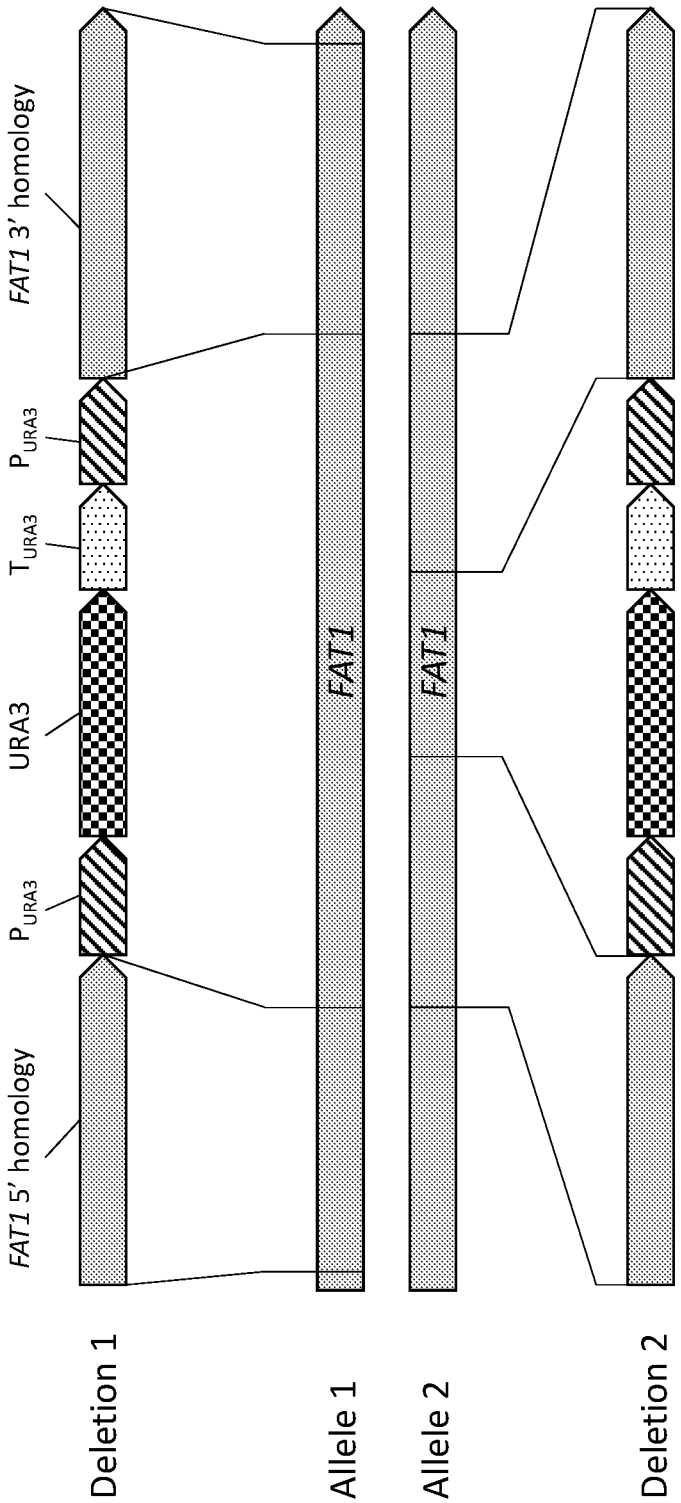

FAT1 5' homology
P URA3
URA3
T URA3
P URA3
FAT1 3' homology
FAT1
FAT1
Deletion 1
Allele 1
Allele 2
Deletion 2

FIG. 2

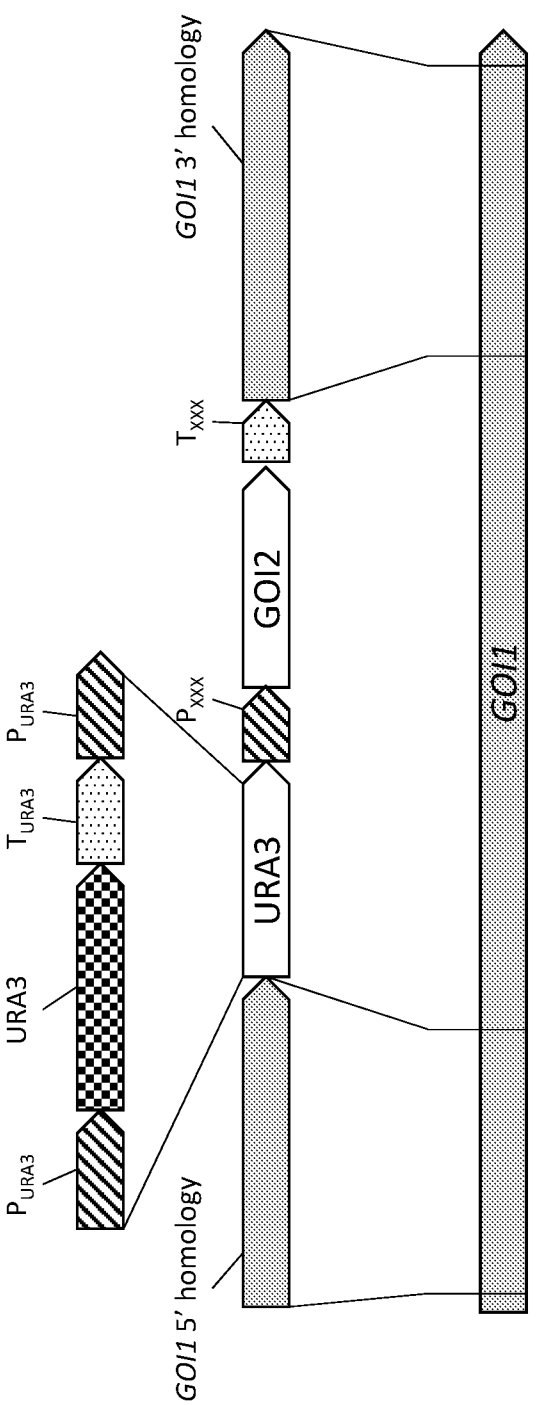


FIG. 3A

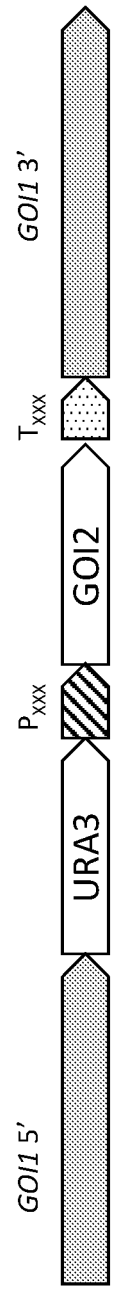


FIG. 3B

FIG. 3C

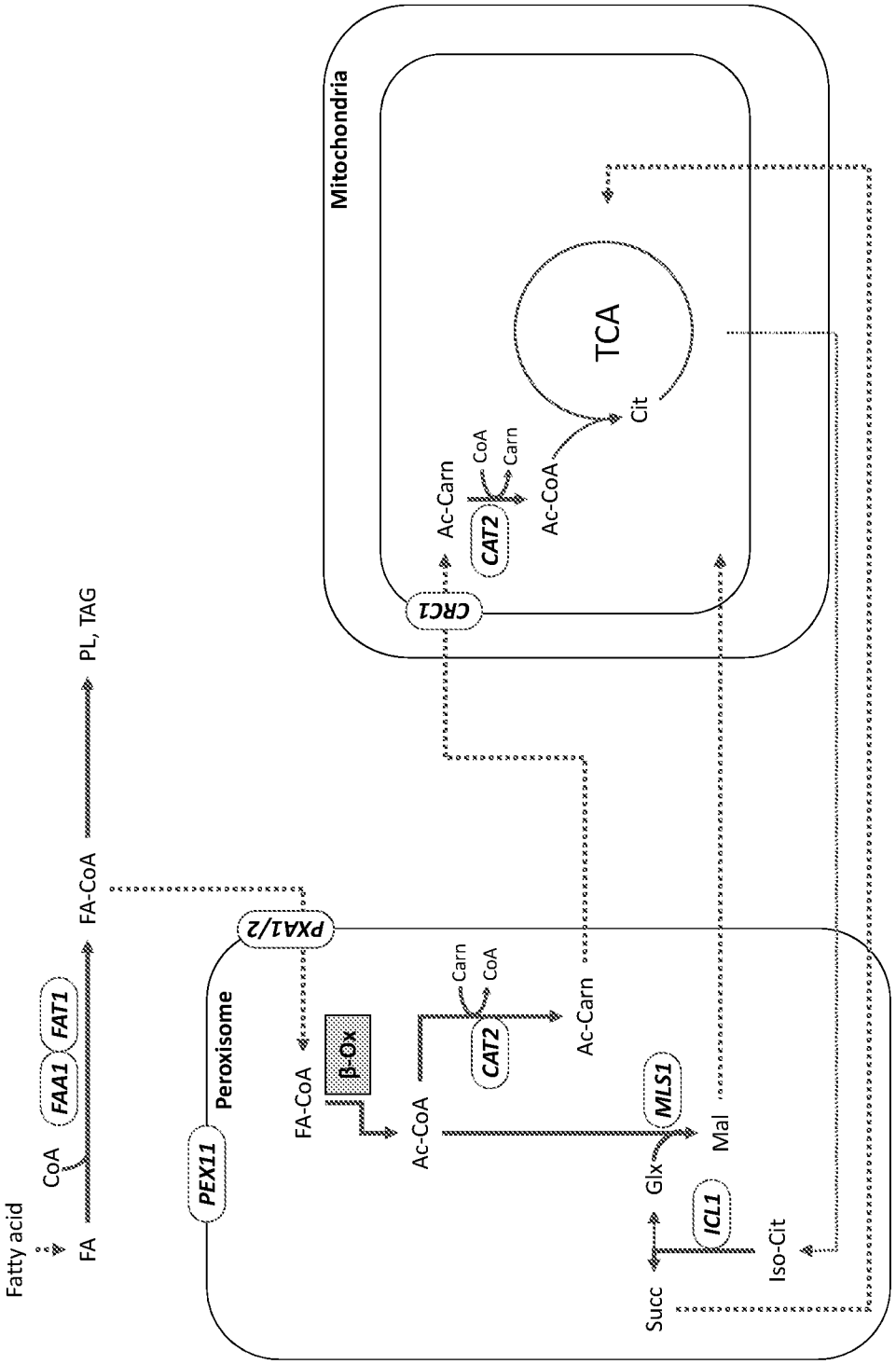

Fatty acid
FA
CoA
FAA1
FAT1
FA-CoA
PL, TAG
PEX11
Peroxisome
PXA1/2
FA-CoA
β-Ox
Ac-CoA
Carn
CoA
CAT2
Ac-Carn
MLS1
Mal
Glx
ICL1
Iso-Cit
Succ
Mitochondria
CRC1
Ac-Carn
CoA
Carn
CAT2
Ac-CoA
TCA
Cit

FIG. 4

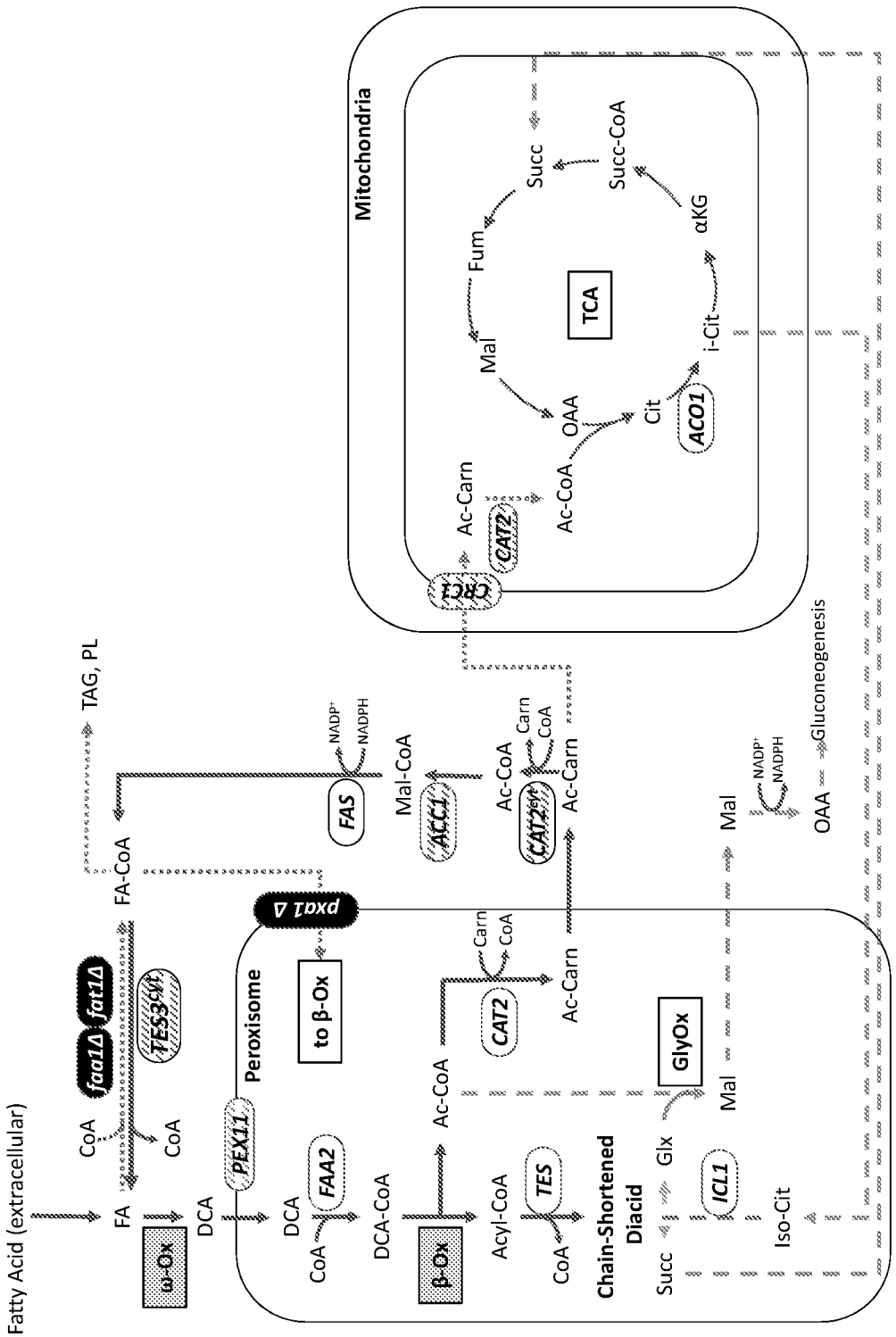

Fatty Acid (extracellular)
FA
ω-Ox
DCA
CoA
faa1Δ fat1Δ
TES3OE
FA-CoA
TAG, PL
PEX11
Peroxisome
pxa1 Δ
DCA
CoA
FAA2
DCA-CoA
β-Ox
Acyl-CoA
TES
CoA
Chain-Shortened Diacid
Succ
Glx
ICL1
Iso-Cit
to β-Ox
Ac-CoA
CAT2
Carn
CoA
Ac-Carn
GlyOx
Mal
FAS
NADP+
NADPH
Mal-CoA
ACC1
Ac-CoA
CAT2ox
Carn
CoA
Ac-Carn
Mal
NADP+
NADPH
OAA
Gluconeogenesis
Mitochondria
CRC1
Ac-Carn
CAT2
Ac-CoA
OAA
Cit
ACO1
i-Cit
αKG
Succ-CoA
Succ
Fum
Mal
TCA

FIG. 5

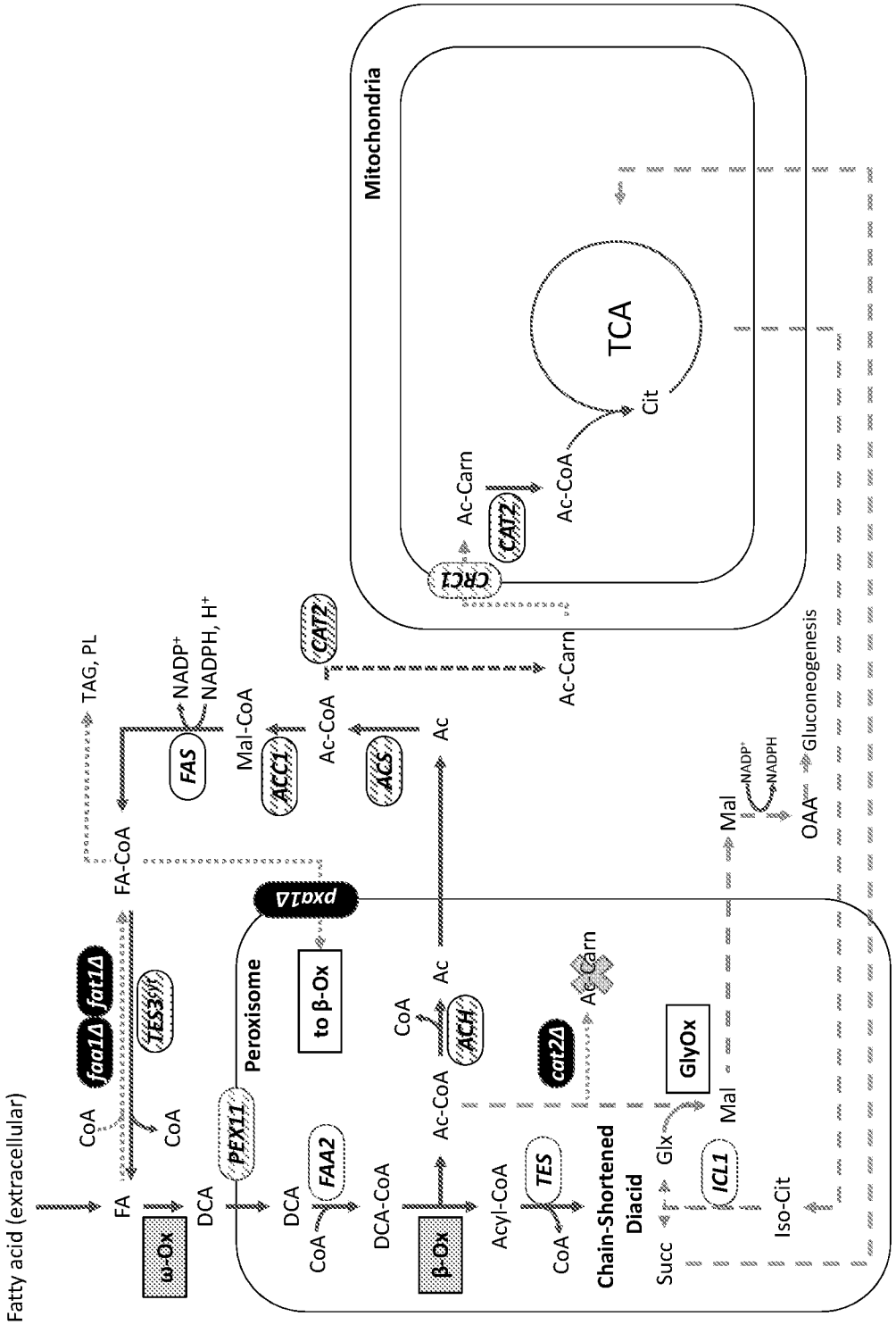

Fatty acid (extracellular)
FA
ω-Ox
DCA
CoA
faa1Δ
fat1Δ
TES3(?)
FA-CoA
TAG, PL
NADP+
NADPH, H+
FAS
Mal-CoA
ACC1
Ac-CoA
ACS
Ac
CAT2
Ac-Carn
Mitochondria
CRC1
Ac-Carn
CAT2
Ac-CoA
TCA
Cit
Peroxisome
PEX11
DCA
CoA
FAA2
DCA-CoA
β-Ox
pxa1Δ
to β-Ox
CoA
Ac
ACH
Ac-CoA
Acyl-CoA
TES
CoA
cat2Δ
Ac-Carn
Chain-Shortened Diacid
Succ
Glx
ICL1
Iso-Cit
GlyOx
Mal
Mal
NADP+
NADPH
OAA
Gluconeogenesis

FIG. 6

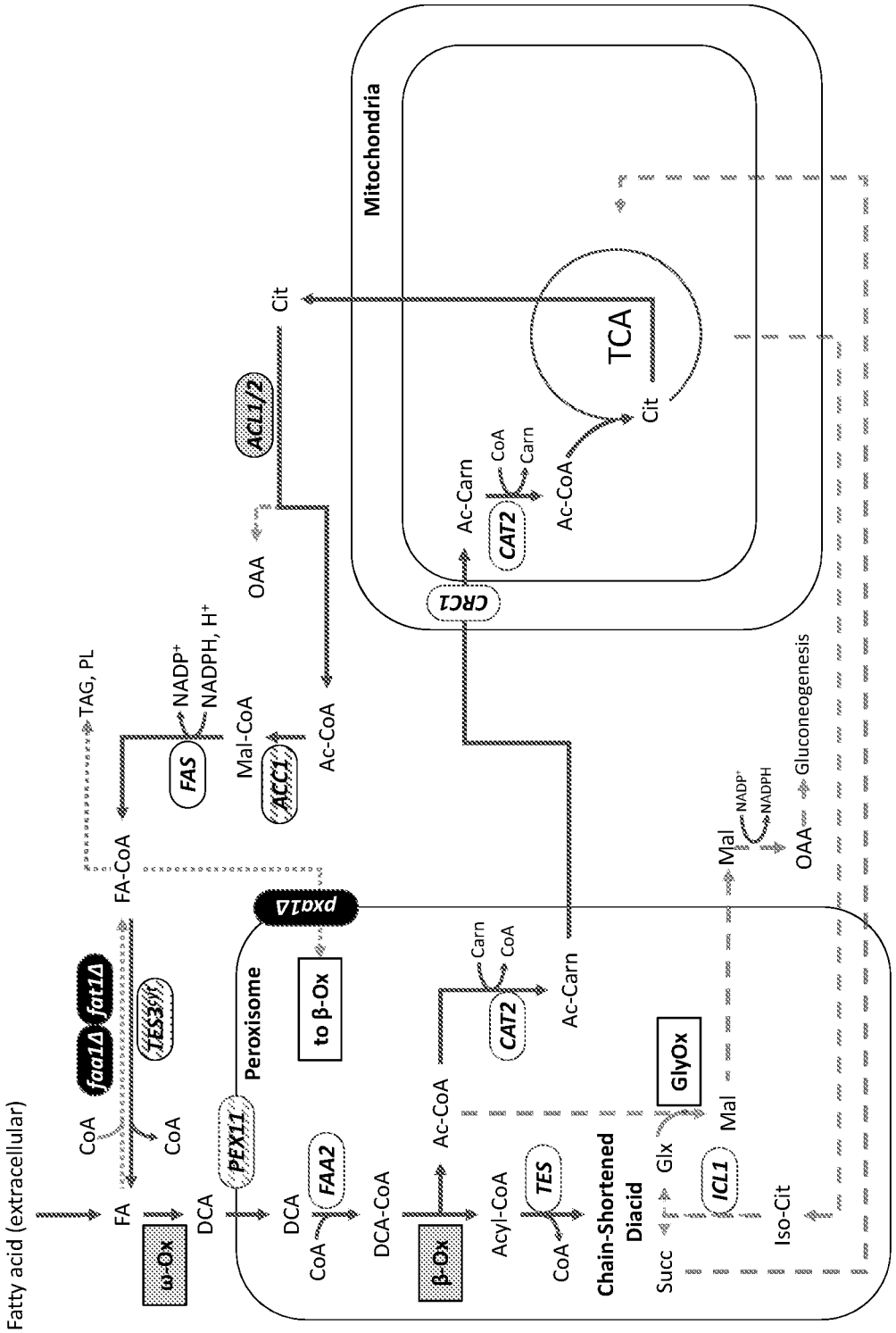

Fatty acid (extracellular)
FA
ω-Ox
DCA
CoA
faa1Δ fat1Δ
TES3(ff)
FA-CoA
TAG, PL
NADP+
NADPH, H+
FAS
Mal-CoA
ACC1
Ac-CoA
OAA
ACL1/2
Cit
Mitochondria
TCA
Cit
Ac-Carn
CoA
Carn
CAT2
Ac-CoA
CRC1
Peroxisome
PEX11
pxa1Δ
DCA
CoA
FAA2
DCA-CoA
to β-Ox
Ac-CoA
β-Ox
Acyl-CoA
TES
CoA
Carn
CoA
CAT2
Ac-Carn
Chain-Shortened Diacid
Succ
Glx
ICL1
Iso-Cit
GlyOx
Mal
Mal
NADP+
NADPH
OAA
Gluconeogenesis

FIG. 7

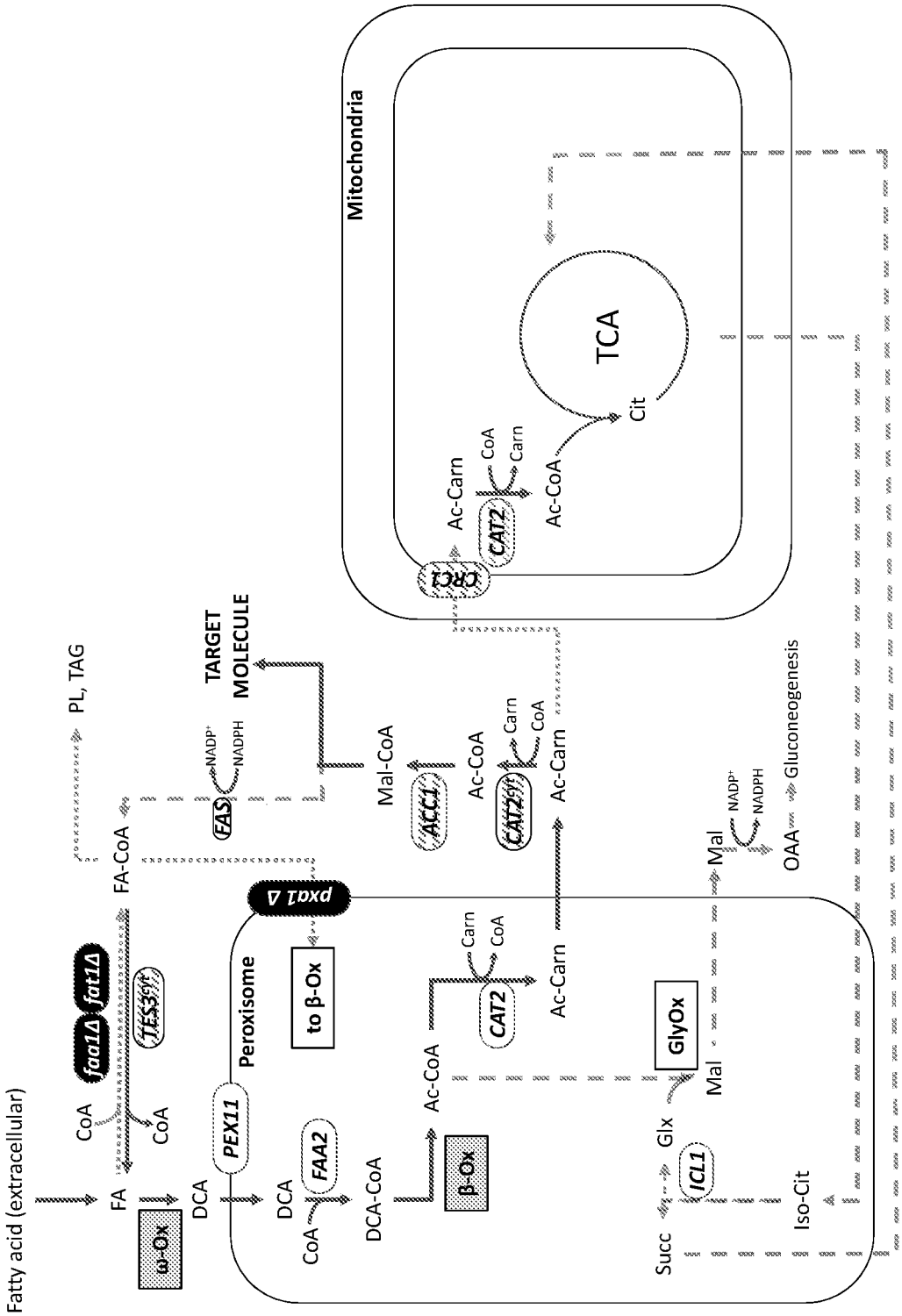

Fatty acid (extracellular)
FA
ω-Ox
DCA
CoA
faa1Δ fat1Δ
TES3(t)
FA-CoA
PL, TAG
PEX11
Peroxisome
pxa1 Δ
DCA
CoA
FAA2
DCA-CoA
to β-Ox
β-Ox
Ac-CoA
Carn
CoA
CAT2
Ac-Carn
FAS
NADP+
NADPH
TARGET MOLECULE
Mal-CoA
ACC1
Ac-CoA
CAT2(t)
Carn
CoA
Ac-Carn
Mitochondria
CRC1
Ac-Carn
CAT2
CoA
Carn
Ac-CoA
TCA
Cit
Mal
NADP+
NADPH
OAA
Gluconeogenesis
GlyOx
Glx
Succ
ICL1
Iso-Cit

FIG. 8

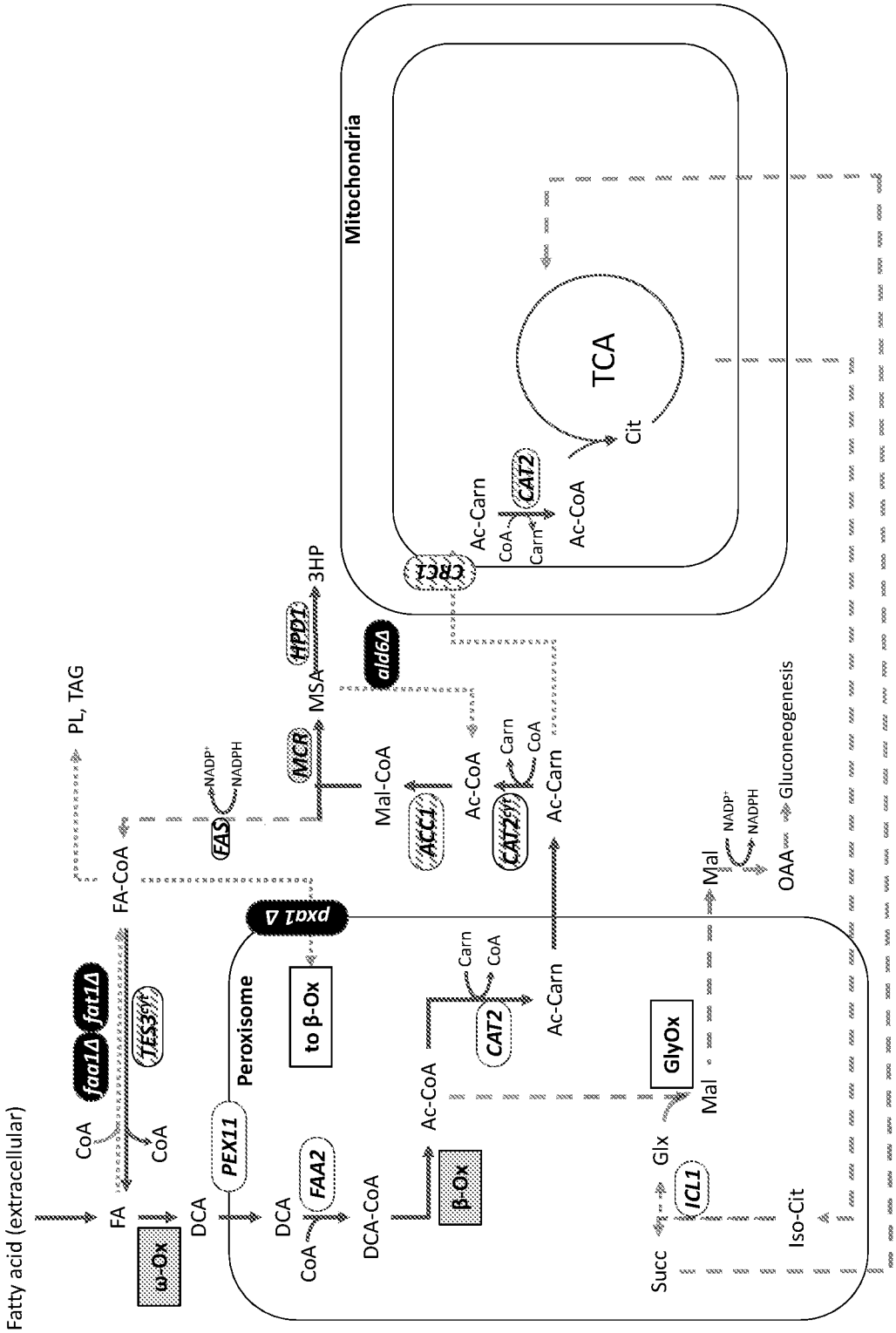

Fatty acid (extracellular)
FA
ω-Ox
DCA
CoA
faa1Δ fat1Δ
TES3(t)
PEX11
Peroxisome
pxa1 Δ
DCA
CoA
FAA2
DCA-CoA
to β-Ox
β-Ox
Ac-CoA
Carn
CoA
CAT2
Ac-Carn
GlyOx
Mal
Succ
Glx
ICL1
Iso-Cit
NADP+
NADPH
Gluconeogenesis
OAA
FA-CoA
PL, TAG
FAS
MCR
MSA
HPD1
3HP
ald6Δ
Mal-CoA
ACC1
Ac-CoA
CAT2(t)
Carn
CoA
Ac-Carn
Mitochondria
CRC1
Ac-Carn
CoA
Carn
CAT2
Ac-CoA
TCA
Cit

FIG. 9

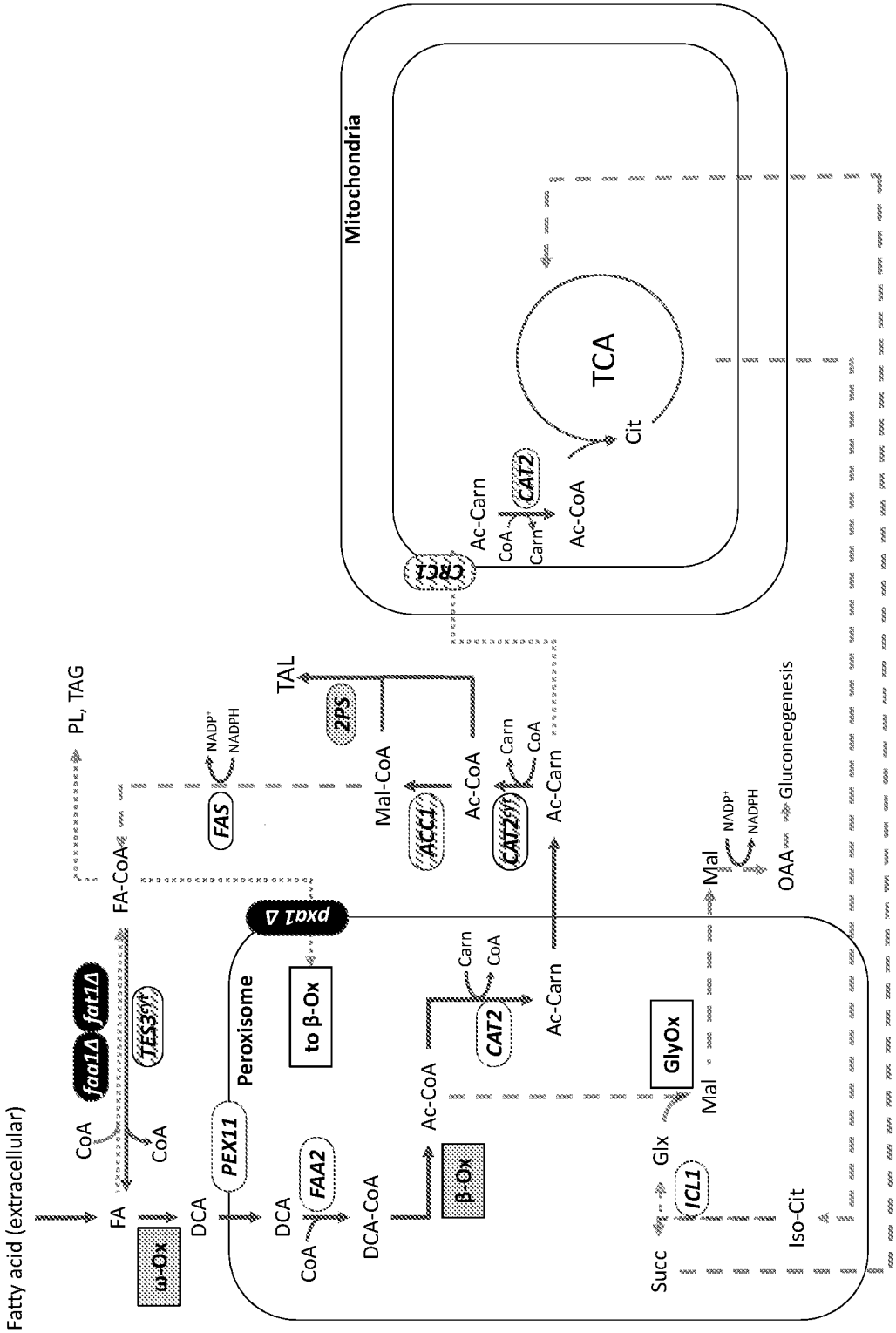

Fatty acid (extracellular)
FA
ω-Ox
DCA
CoA
faa1Δ fat1Δ
TES3(t)
CoA
FA-CoA
PEX11
Peroxisome
pxa1 Δ
DCA
CoA
FAA2
DCA-CoA
to β-Ox
β-Ox
Ac-CoA
Carn
CoA
CAT2
Ac-Carn
GlyOx
Glx
Succ
ICL1
Iso-Cit
Mal
NADP+
NADPH
OAA
Gluconeogenesis
PL, TAG
FAS
NADP+
NADPH
TAL
2PS
Mal-CoA
ACC1
Ac-CoA
CAT2(t)
Carn
CoA
Ac-Carn
Mitochondria
CRC1
Ac-Carn
CoA
Carn
CAT2
Ac-CoA
TCA
Cit

FIG. 10

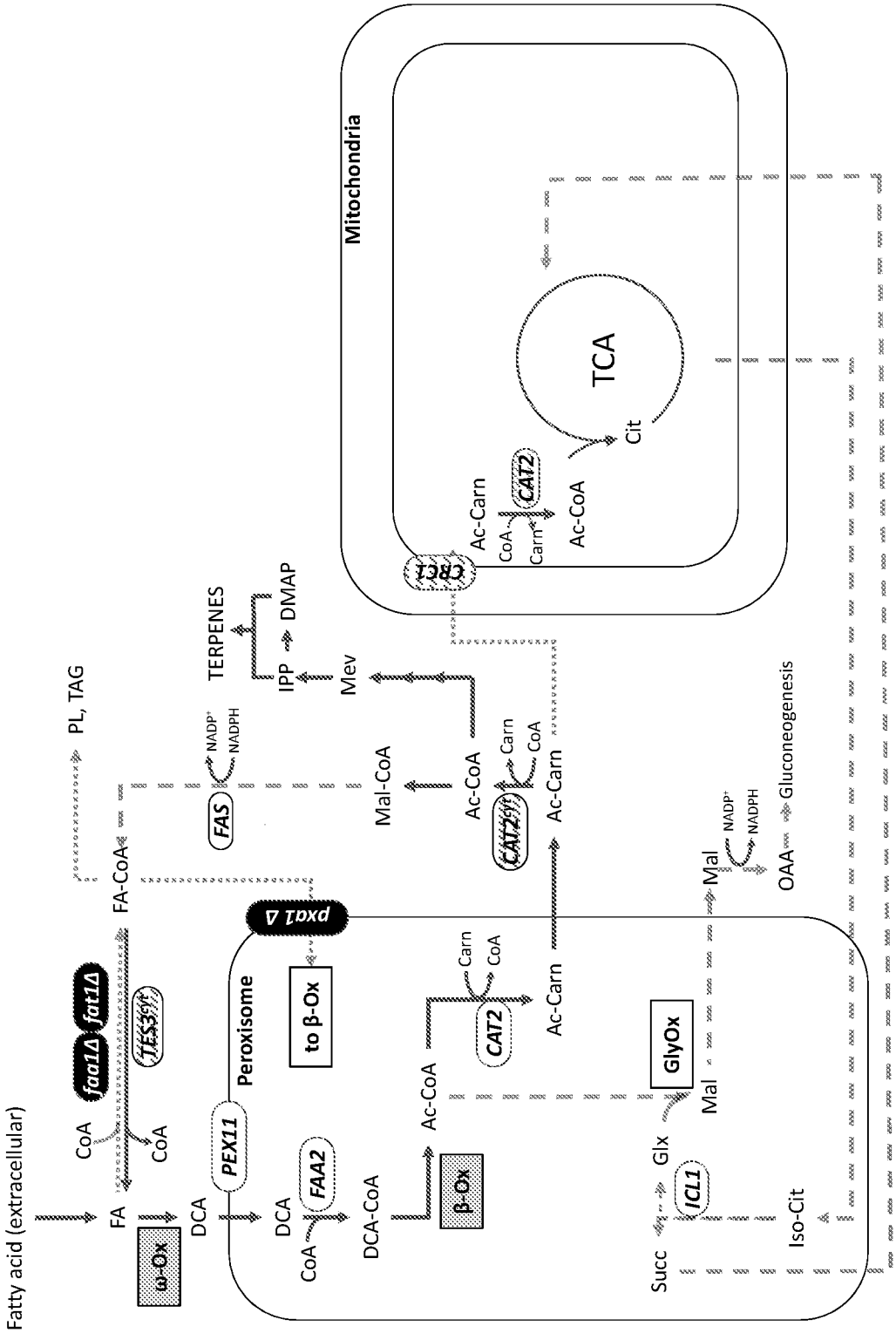

Fatty acid (extracellular)
FA
ω-Ox
DCA
CoA
faa1Δ fat1Δ
TES3(v)
CoA
PEX11
Peroxisome
pxa1 Δ
DCA
CoA
FAA2
DCA-CoA
to β-Ox
β-Ox
Ac-CoA
Carn
CoA
CAT2
Ac-Carn
FA-CoA
PL, TAG
NADP+
NADPH
FAS
TERPENES
IPP
DMAP
Mev
Mal-CoA
Ac-CoA
Carn
CoA
CAT2(v)
Ac-Carn
Succ
Glx
ICL1
Iso-Cit
Mal
GlyOx
Mal
NADP+
NADPH
OAA
Gluconeogenesis
Mitochondria
CRC1
Ac-Carn
CoA
Carn
CAT2
Ac-CoA
TCA
Cit

FIG. 11

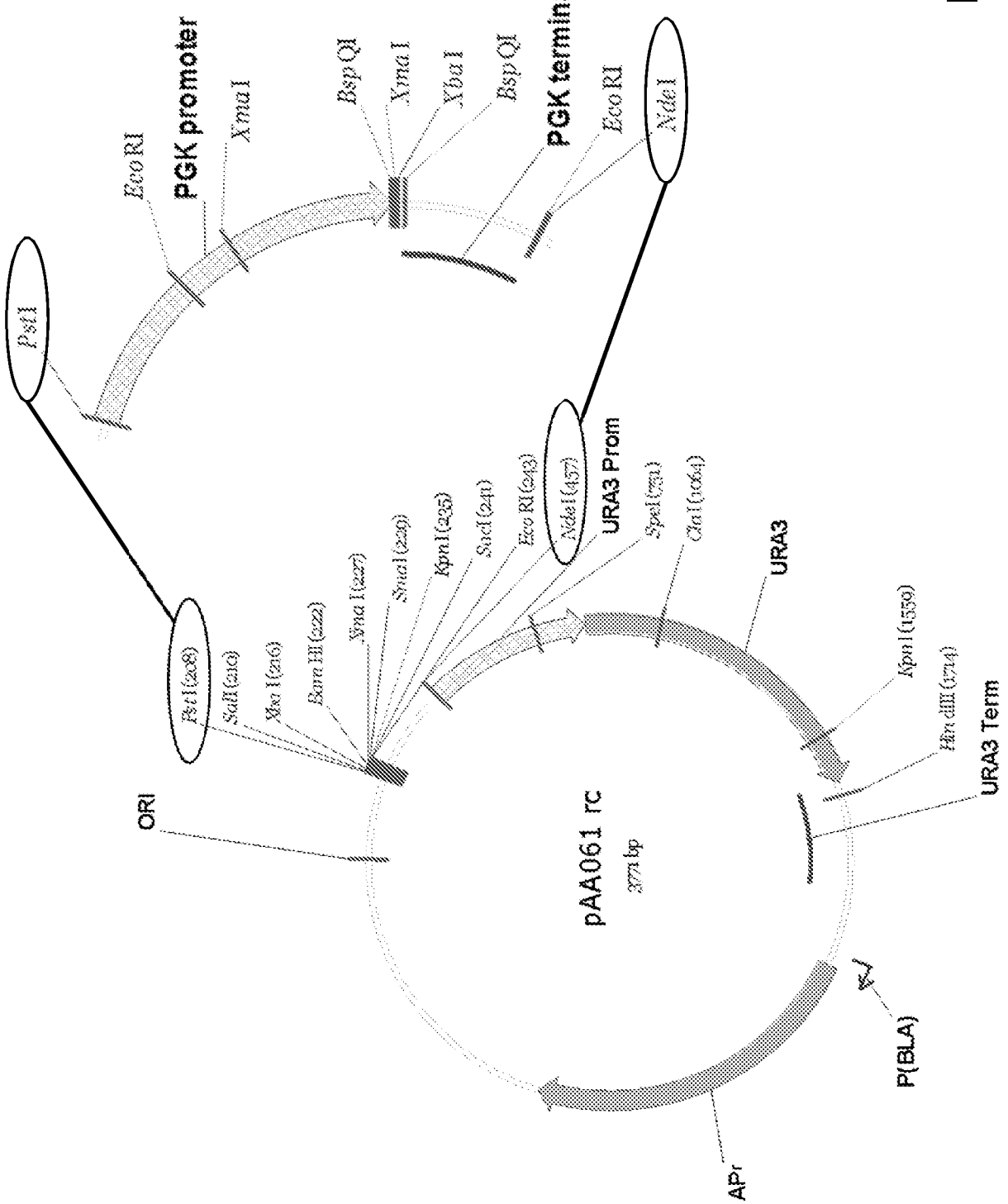

PGK promoter
PGK terminator
URA3 Prom
URA3
URA3 Term
ORI
pAA061 rc
3771 bp
P(BLA)
APr
Pst I
Eco RI
Xma I
Bsp QI
Xba I
Nde I
Sal I (210)
Xho I (216)
Bam HI (222)
Sma I (227)
Kpn I (235)
Sac I (241)
Eco RI (243)
Nde I (457)
Spe I (931)
Cla I (1064)
Kpn I (1539)
Hin dIII (1714)

FIG. 12

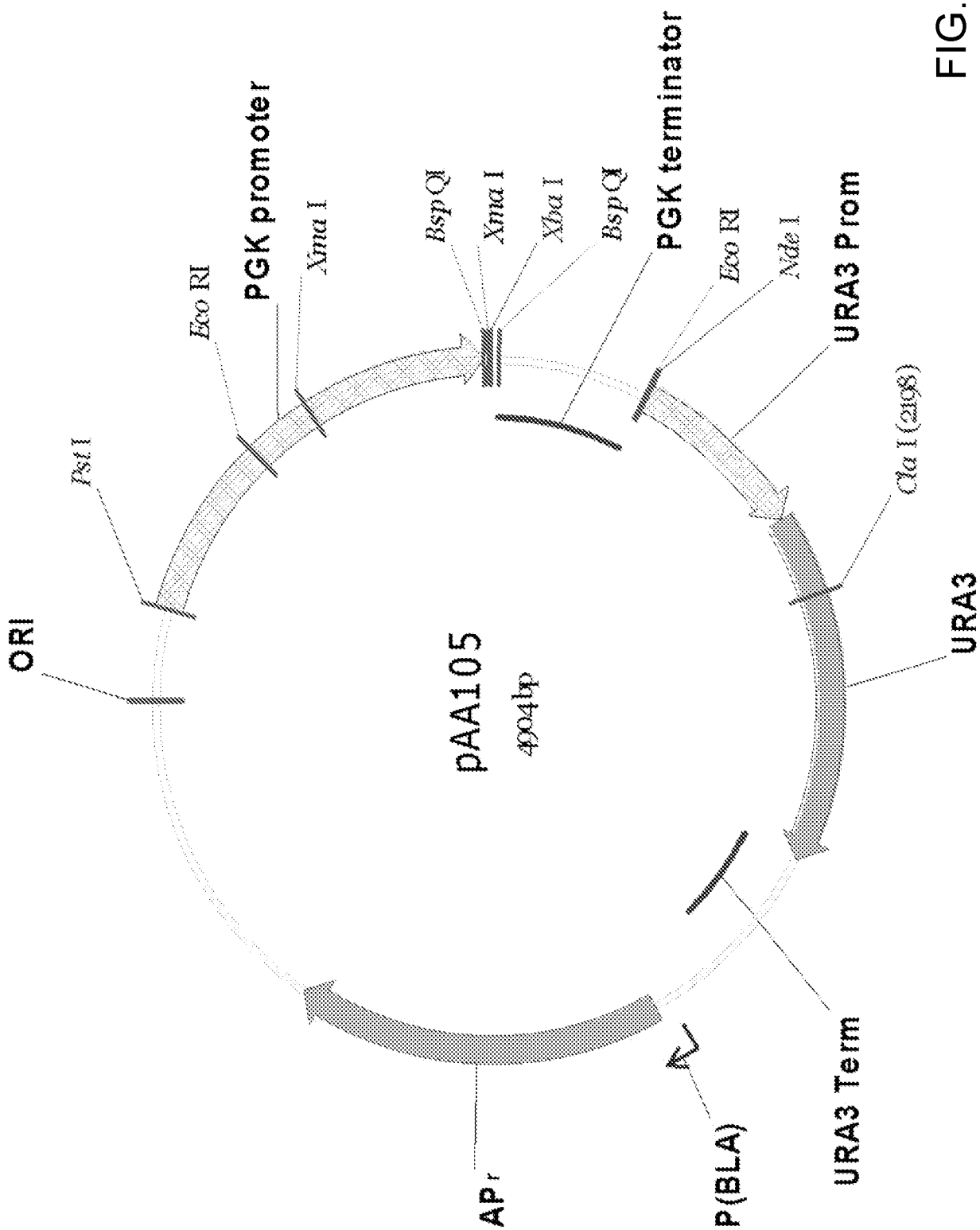

pAA105
4904bp
ORI
Pst I
Eco RI
PGK promoter
Xma I
Bsp QI
Xma I
Xba I
Bsp QI
PGK terminator
Eco RI
Nde I
URA3 Prom
Cla I (2198)
URA3
URA3 Term
P(BLA)
AP r

FIG. 13

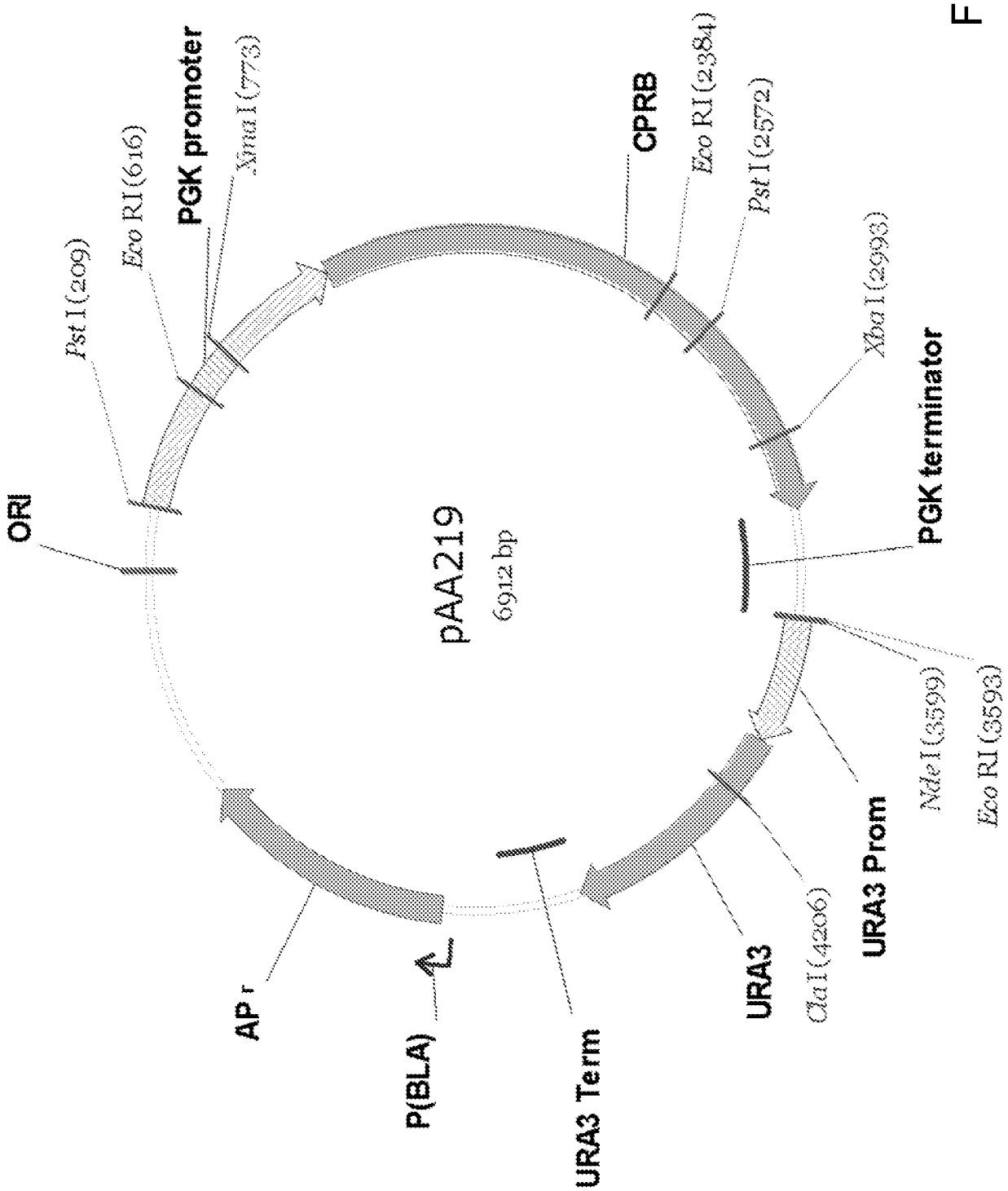

pAA219
6912 bp
ORI
Pst I (209)
Eco RI (616)
PGK promoter
Xma I (773)
CPRB
Eco RI (2384)
Pst I (2572)
Xba I (2993)
PGK terminator
Nde I (3599)
Eco RI (3593)
URA3 Prom
Cla I (4206)
URA3
URA3 Term
P(BLA)
AP r

FIG. 14

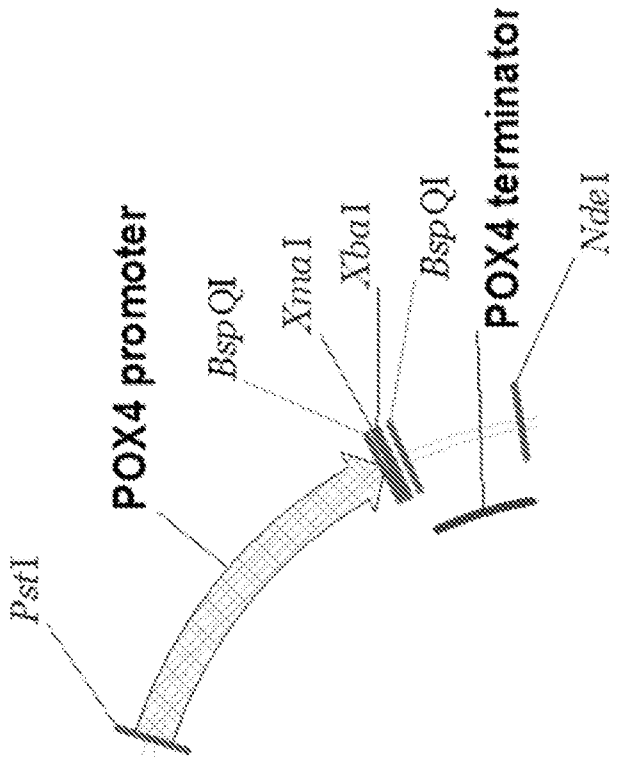

POX4 promoter
POX4 terminator
PstI
BspQI
XmaI
XbaI
BspQI
NdeI

FIG. 15

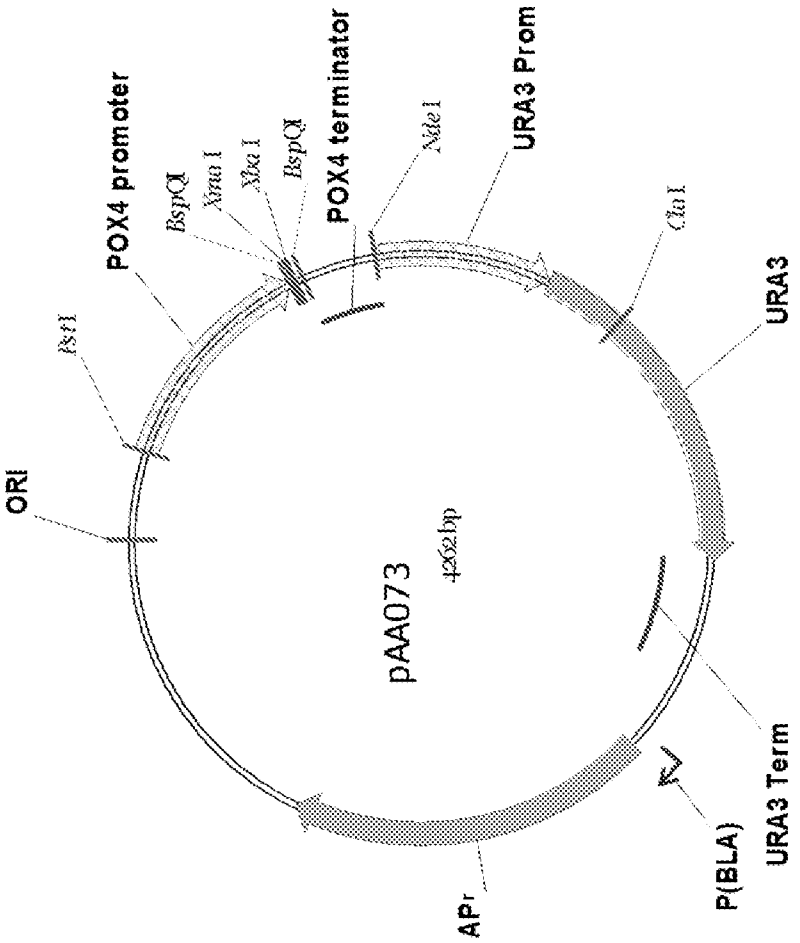

POX4 promoter
BspQI
Xma I
Xba I
BspQI
POX4 terminator
Nde I
URA3 Prom
Cla I
URA3
BstI
ORI
pAA073
4262bp
URA3 Term
P(BLA)
APr

FIG. 16

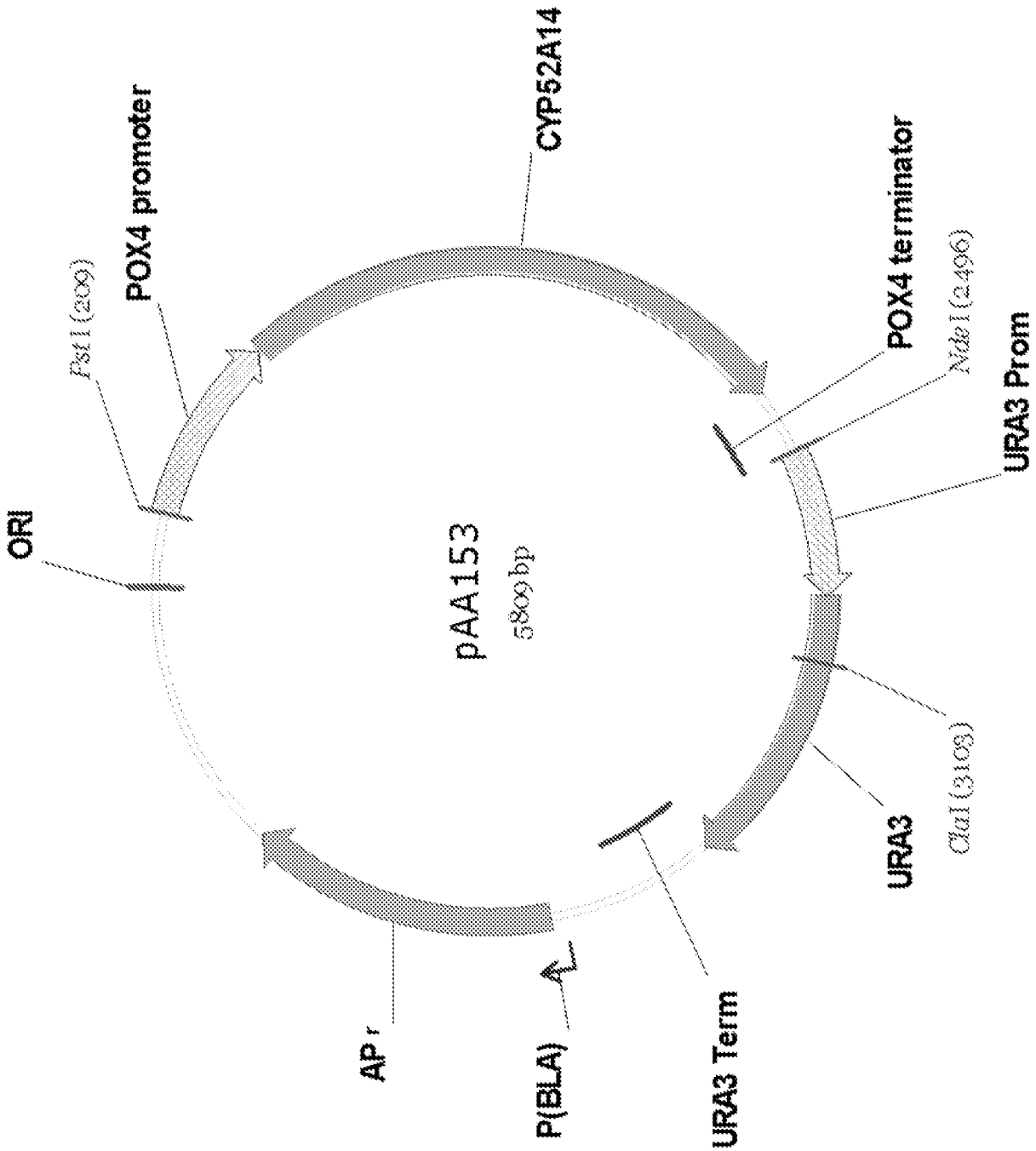

pAA153
5809 bp
CYP52A14
POX4 terminator
URA3 Prom
URA3
URA3 Term
P(BLA)
AP r
ORI
POX4 promoter
Pst I (209)
Nde I (2496)
Cla I (3103)

FIG. 17

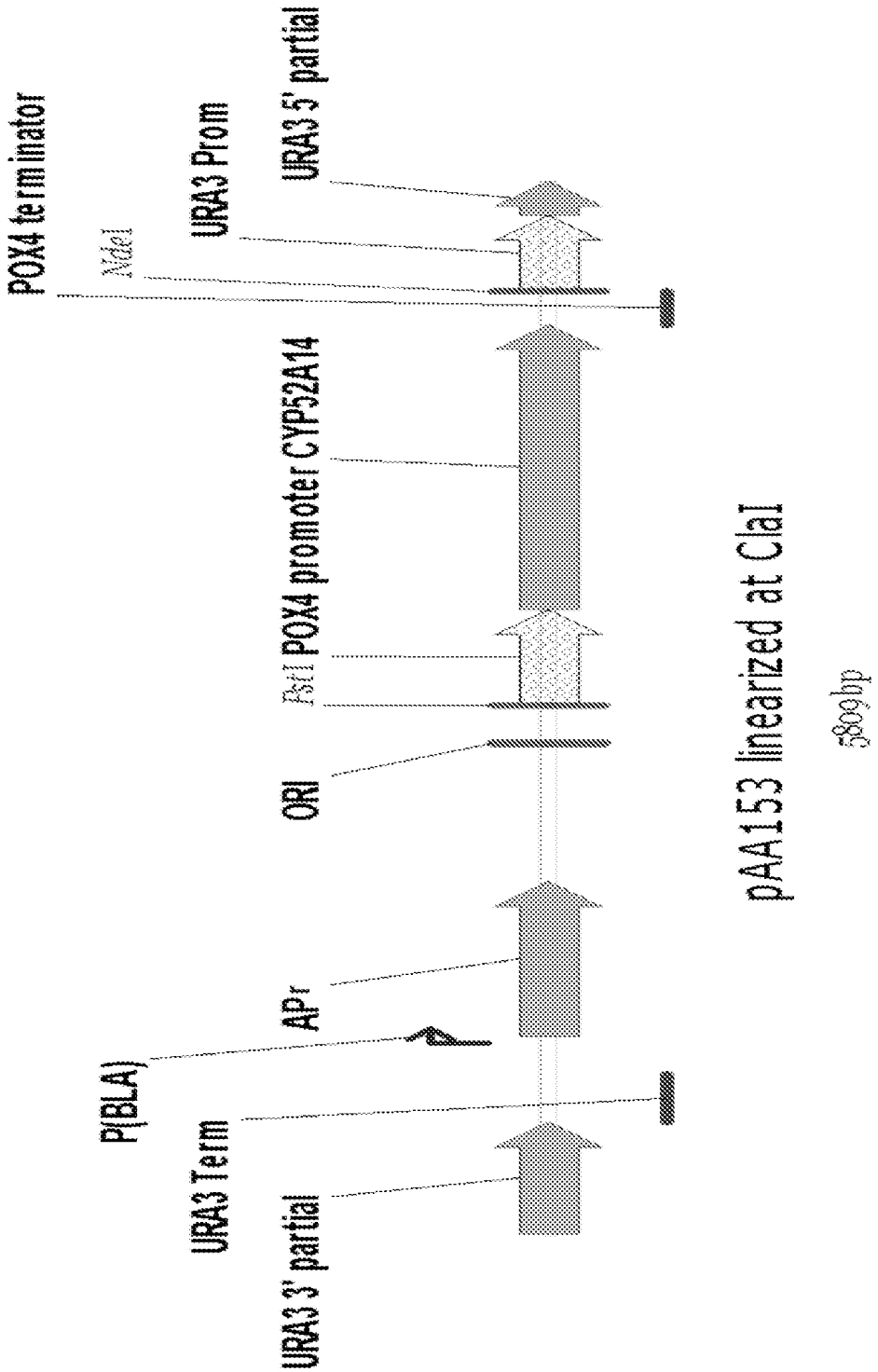

POX4 terminator
NdeI
URA3 Prom
URA3 5' partial
BstI POX4 promoter CYP52A14
ORI
APr
P(BLA)
URA3 Term
URA3 3' partial
pAA153 linearized at ClaI
5809bp

FIG. 18

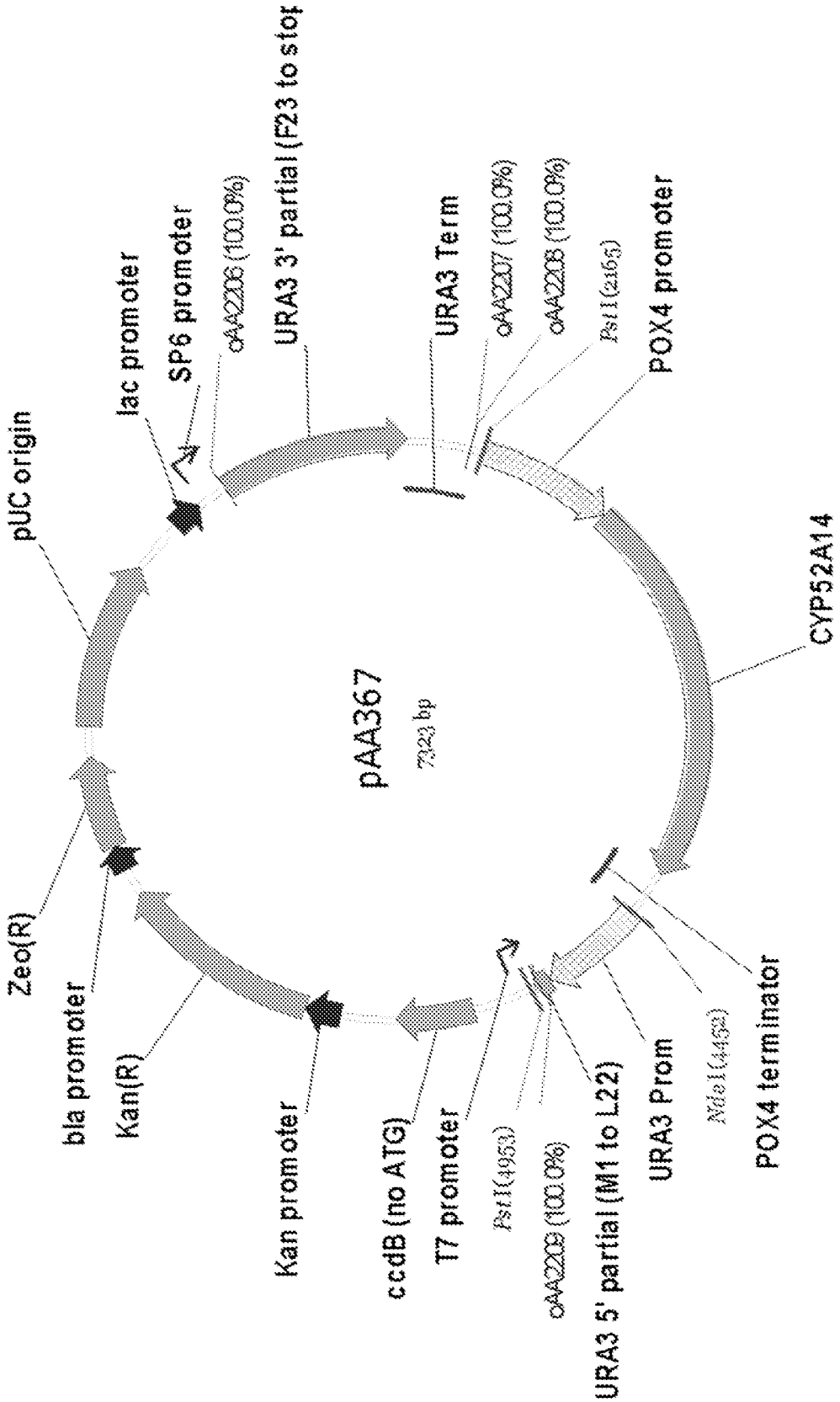

pAA367
7323 bp
pUC origin
lac promoter
SP6 promoter
oAA2206 (100.0%)
URA3 3' partial (F23 to stop
URA3 Term
oAA2207 (100.0%)
oAA2208 (100.0%)
PstI (2165)
POX4 promoter
CYP52A14
POX4 terminator
MfeI (4462)
URA3 Prom
URA3 5' partial (M1 to L22)
oAA2209 (100.0%)
PstI (4953)
T7 promoter
ccdB (no ATG)
Kan promoter
Kan(R)
bla promoter
Zeo(R)

FIG. 19

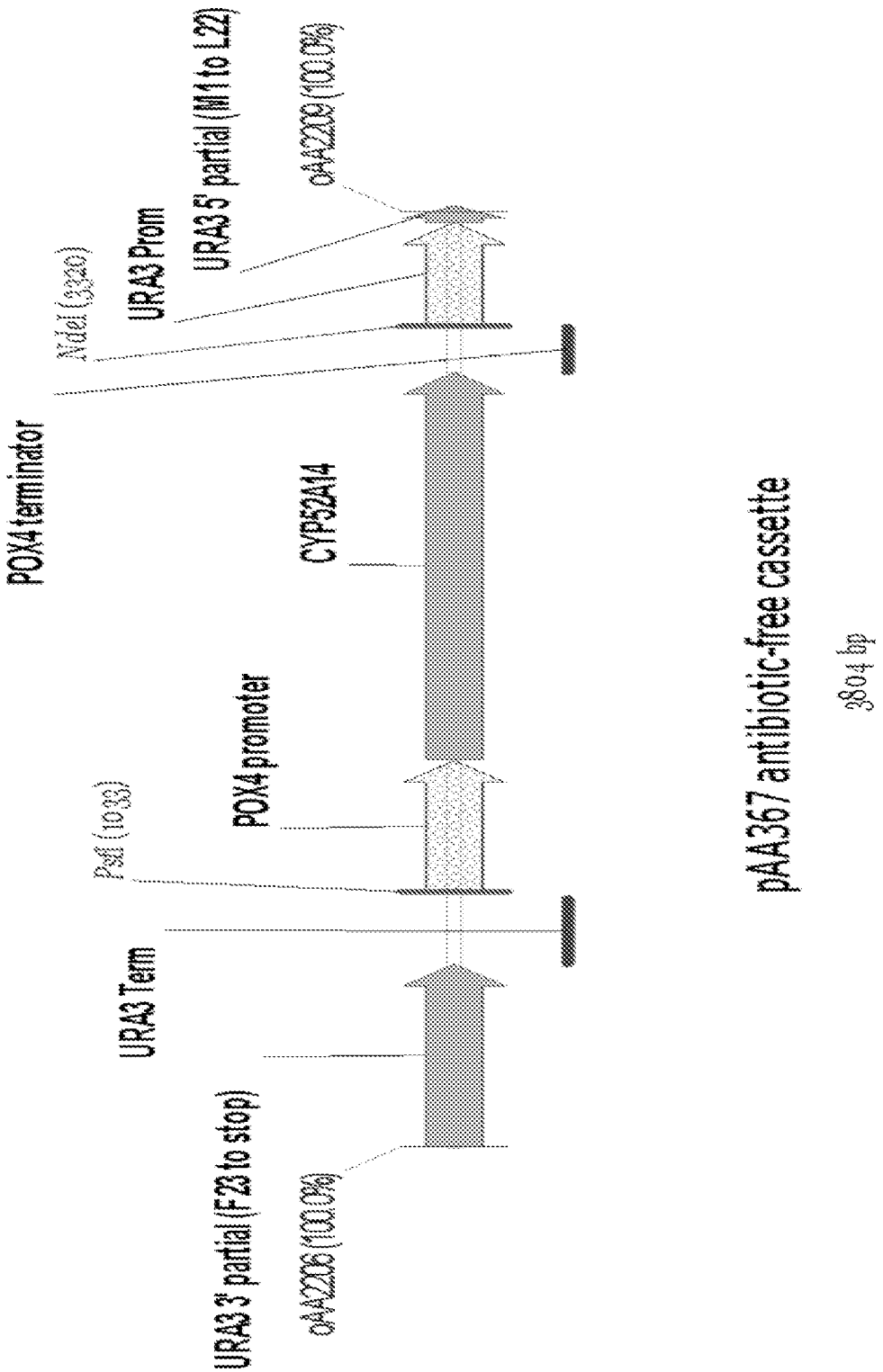

pAA367 antibiotic-free cassette
3804 bp
POX4 terminator
NdeI (3320)
URA3 Prom
URA3 5' partial (M1 to L22)
pAA2209 (100.0%)
CYP52A14
POX4 promoter
PsiI (1033)
URA3 Term
URA3 3' partial (F23 to stop)
pAA2206 (100.0%)

FIG. 20

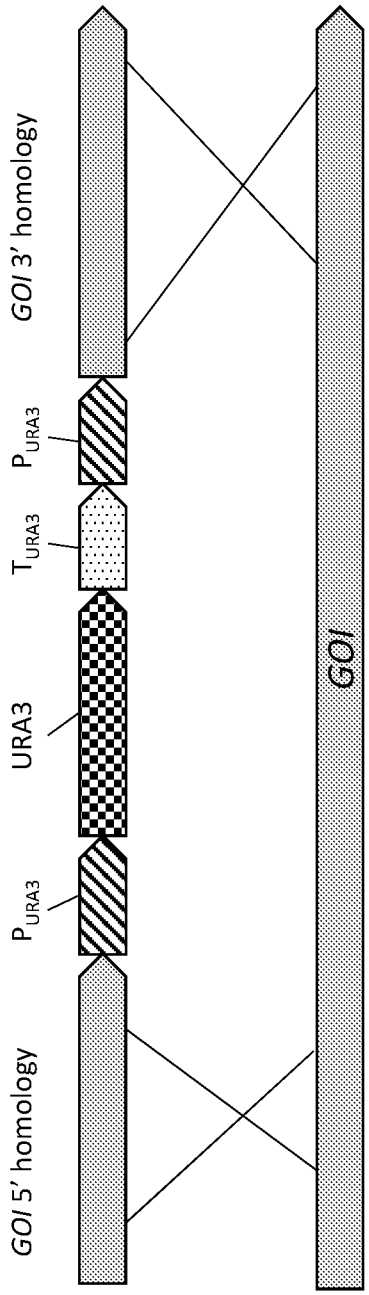


FIG. 21A

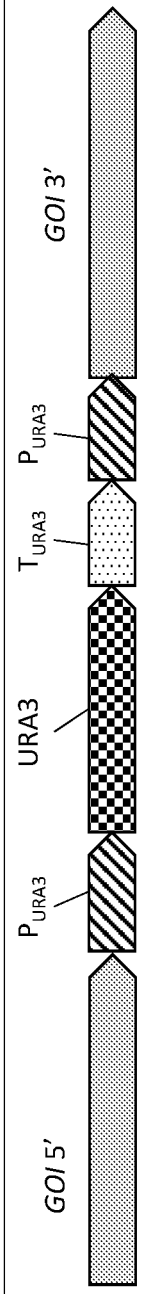


FIG. 21B

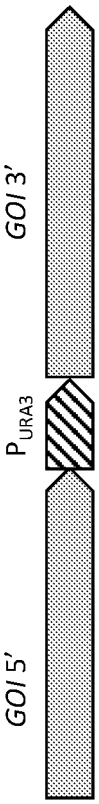


FIG. 21C

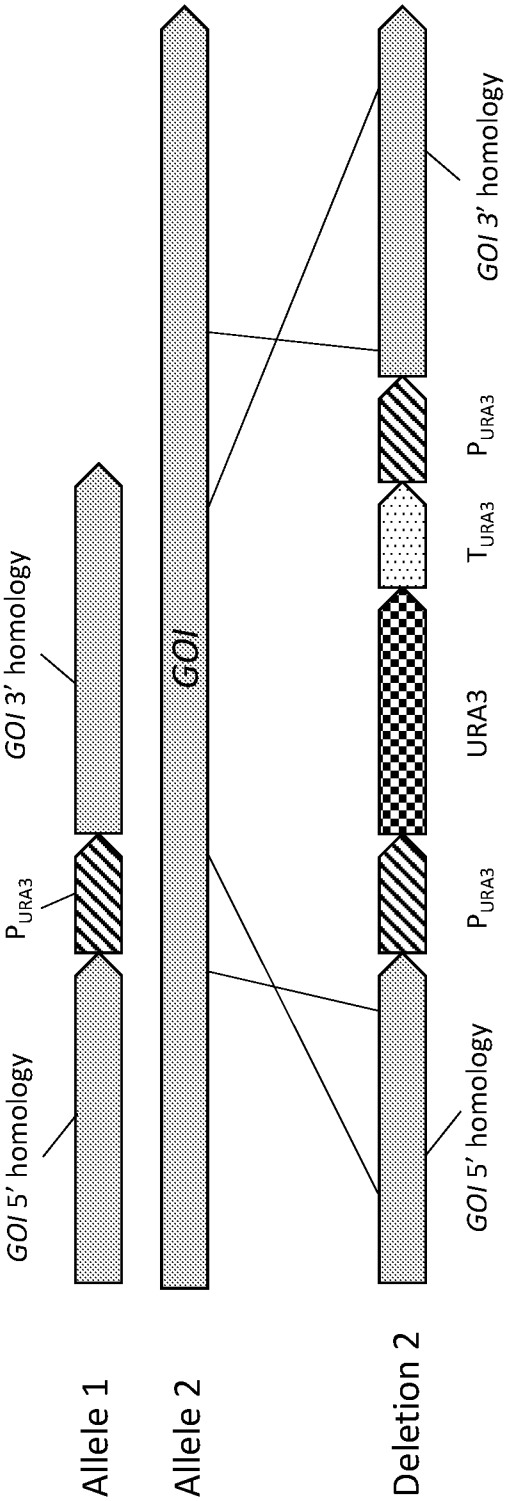

GOI 5' homology
P URA3
GOI 3' homology
GOI
GOI 5' homology
P URA3
URA3
T URA3
P URA3
GOI 3' homology
Allele 1
Allele 2
Deletion 2

FIG. 22

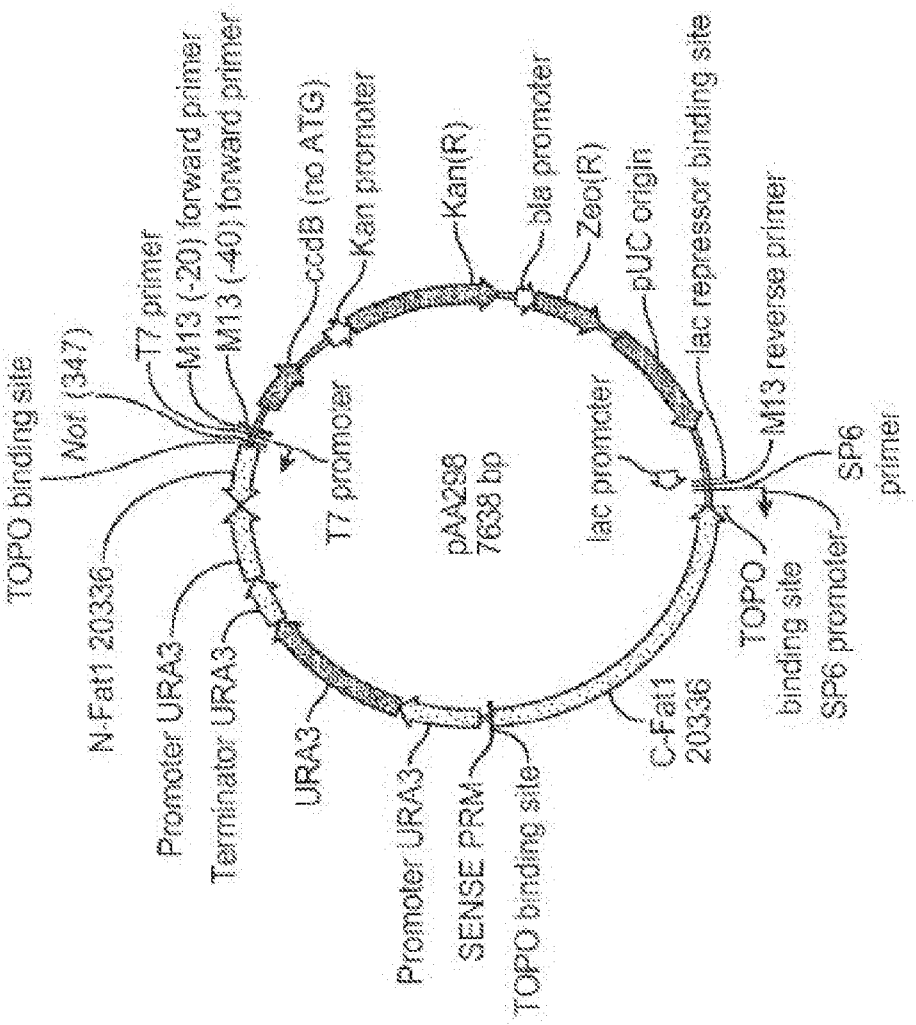

TOPO binding site
Not (347)
T7 primer
M13 (-20) forward primer
M13 (-40) forward primer
ccdB (no ATG)
Kan promoter
Kan(R)
bla promoter
Zeo(R)
pUC origin
lac repressor binding site
M13 reverse primer
SP6 primer
TOPO binding site
SP6 promoter
C-Fat1 20336
lac promoter
T7 promoter
pAA293 7638 bp
N-Fat1 20336
Promoter URA3
Terminator URA3
URA3
Promoter URA3
SENSE PRM
TOPO binding site

FIG. 23


pAA1519
6480 bp
pUC origin
lac promoter
SP6 promoter
oAA2377 (100.0%)
Cv CAT2 3' homology
Promoter URA3
Terminator URA3
Cv URA3
Promoter URA3
Cv CAT2 5' homology
oAA2372 (100.0%)
T7 promoter
ccdB (no ATG)
Kan promoter
Kan(R)
bla promoter
Zeo(R)
pAA1520
7038 bp
pUC origin
lac promoter
SP6 promoter
oAA2372 (100.0%)
Cv CAT2 5' homology
Promoter URA3
Terminator URA3
Cv URA3
Promoter URA3
Cv CAT2 3' homology
oAA2377 (100.0%)
T7 promoter
ccdB (no ATG)
Kan promoter
Kan(R)
bla promoter
Zeo(R)

FIG. 24

FIG. 25A

```
                                1                                                 50
CvCAT2 from pAA426       (1) MFNFKLSQQVLKNSTKSIMPILKKPFSTSHAKGDLFKYQSQLPKLPVPTL
      CvCAT2(-mts)       (1) ------------------MPILKKPFSTSHAKGDLFKYQSQLPKLPVPTL
  CvCAT2(-mts-pts)       (1) ------------------MPILKKPFSTSHAKGDLFKYQSQLPKLPVPTL
      CvCAT2(-pts)       (1) MFNFKLSQQVLKNSTKSIMPILKKPFSTSHAKGDLFKYQSQLPKLPVPTL
```

FIG. 25B

```
                             601                        628
CvCAT2 from pAA426     (601) KWYLVDSANEMKDVLTKGLLTDAKPKL-
      CvCAT2(-mts)     (583) KWYLVDSANEMKDVLTKGLLTDAKPKL-
  CvCAT2(-mts-pts)     (583) KWYLVDSANEMKDVLTKGLLTDAK----
      CvCAT2(-pts)     (601) KWYLVDSANEMKDVLTKGLLTDAK----
```

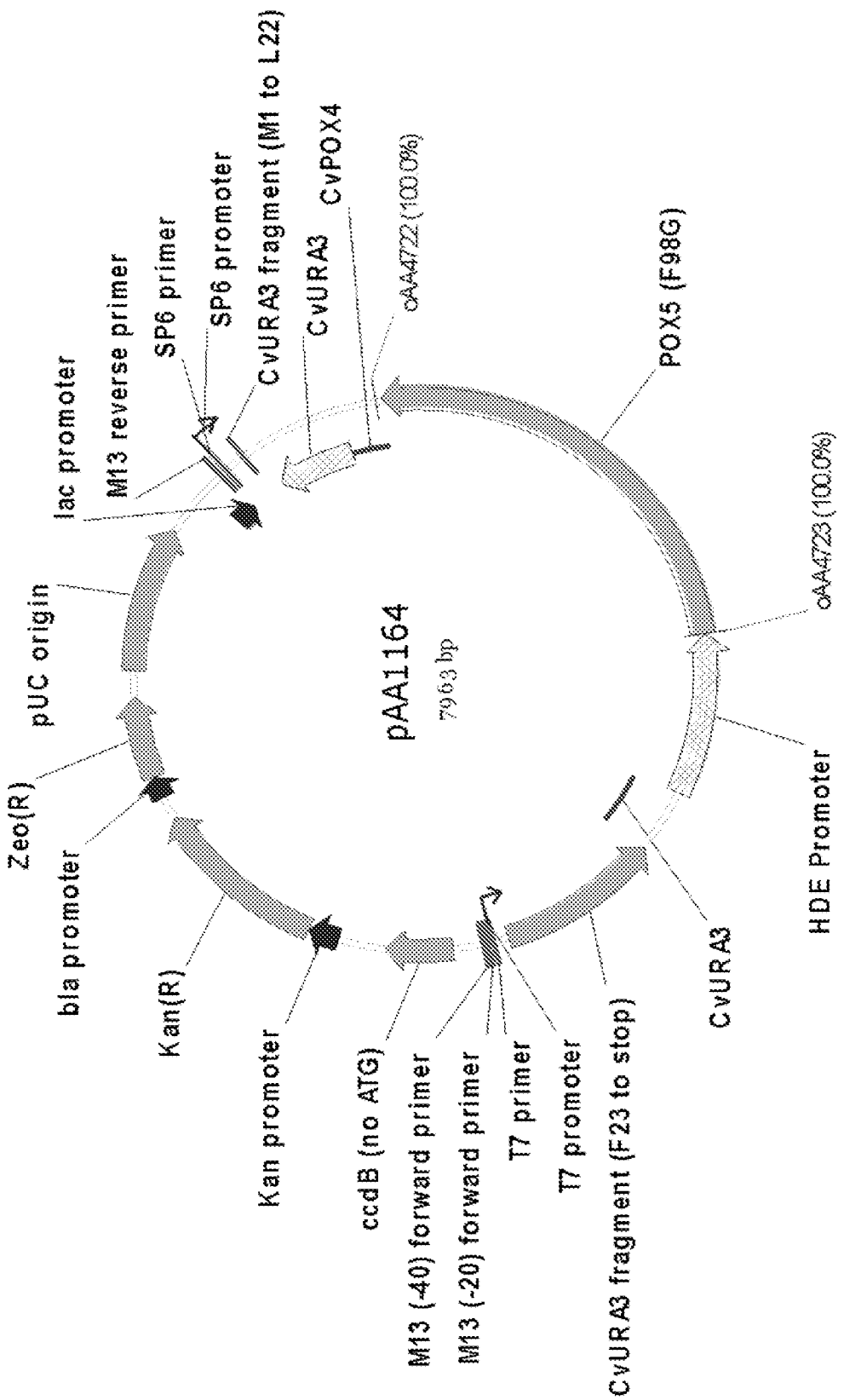

pAA1164
7963 bp
lac promoter
M13 reverse primer
SP6 primer
SP6 promoter
CvURA3 fragment (M1 to L22)
CvURA3
CvPOX4
oAA4722 (100.0%)
POX5 (F98G)
oAA4723 (100.0%)
HDE Promoter
CvURA3
CvURA3 fragment (F23 to stop)
T7 promoter
T7 primer
M13 (-20) forward primer
M13 (-40) forward primer
ccdB (no ATG)
Kan promoter
Kan(R)
bla promoter
Zeo(R)
pUC origin

FIG. 26

```
                1                                                  50
CvYat1p     (1) ----------------------------------------MSTYQFQETLEKLPIPD
CvYat1p+CAT2mts (1) MFNFKLSQQVLKNSTKSIMPILKKP-FSTSHAKGSTYQFQETLEKLPIPD
CvYAT1p+CIT1mts (1) ------MSALRSFQRSSNVAKST--LKNSVRTYSTYQFQETLEKLPIPD
CvYat1p+COX4mts (1) ----MLSRTTLRVARQQTRLLSTSRILFNSKTDQSTYQFQETLEKLPIPD
```

FIG. 27

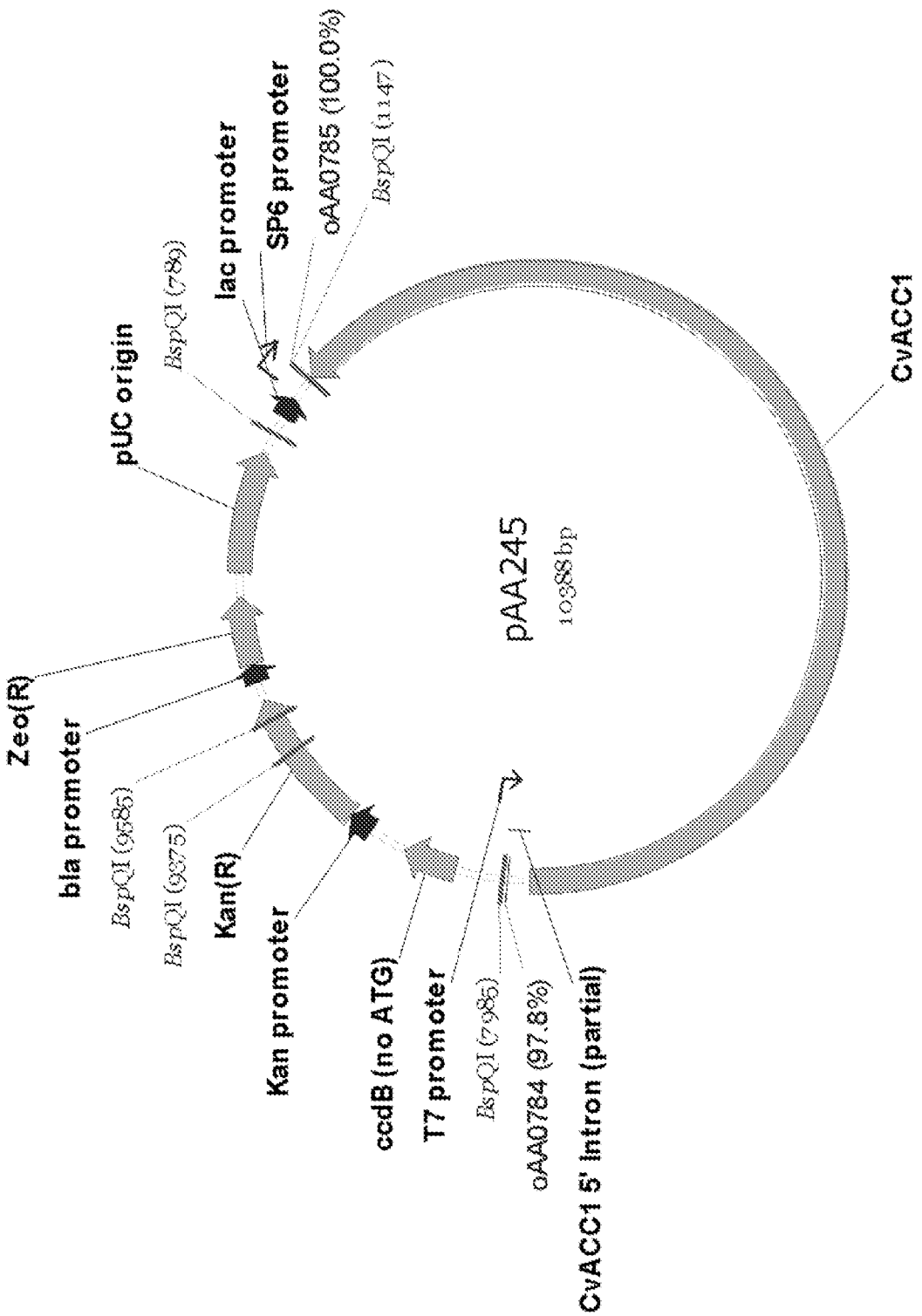

pAA245
10388 bp
CvACC1
oAA0785 (100.0%)
BspQI (1147)
SP6 promoter
lac promoter
BspQI (789)
pUC origin
Zeo(R)
bla promoter
BspQI (9585)
BspQI (9375)
Kan(R)
Kan promoter
ccdB (no ATG)
T7 promoter
BspQI (7985)
oAA0784 (97.8%)
CvACC1 5' intron (partial)

FIG. 28

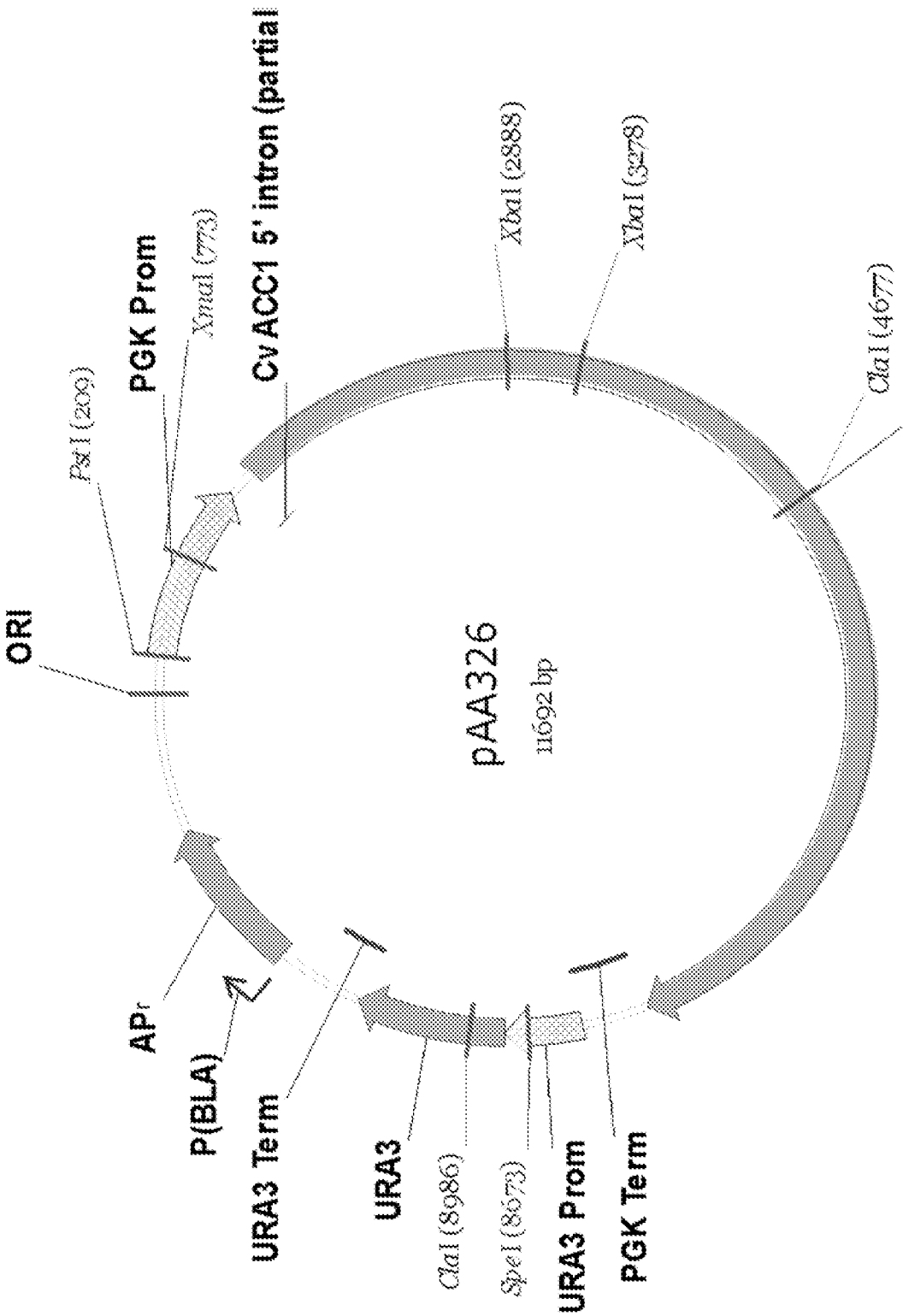

pAA326
11692 bp
ORI
PGK Prom
Cv ACC1 5' intron (partial
Cv ACC1
APr
P(BLA)
URA3 Term
URA3
URA3 Prom
PGK Term
Pst I (209)
Xma1 (773)
Xba1 (2888)
Xba1 (3278)
Cla1 (4677)
Cla1 (8986)
Spe1 (8673)

FIG. 29

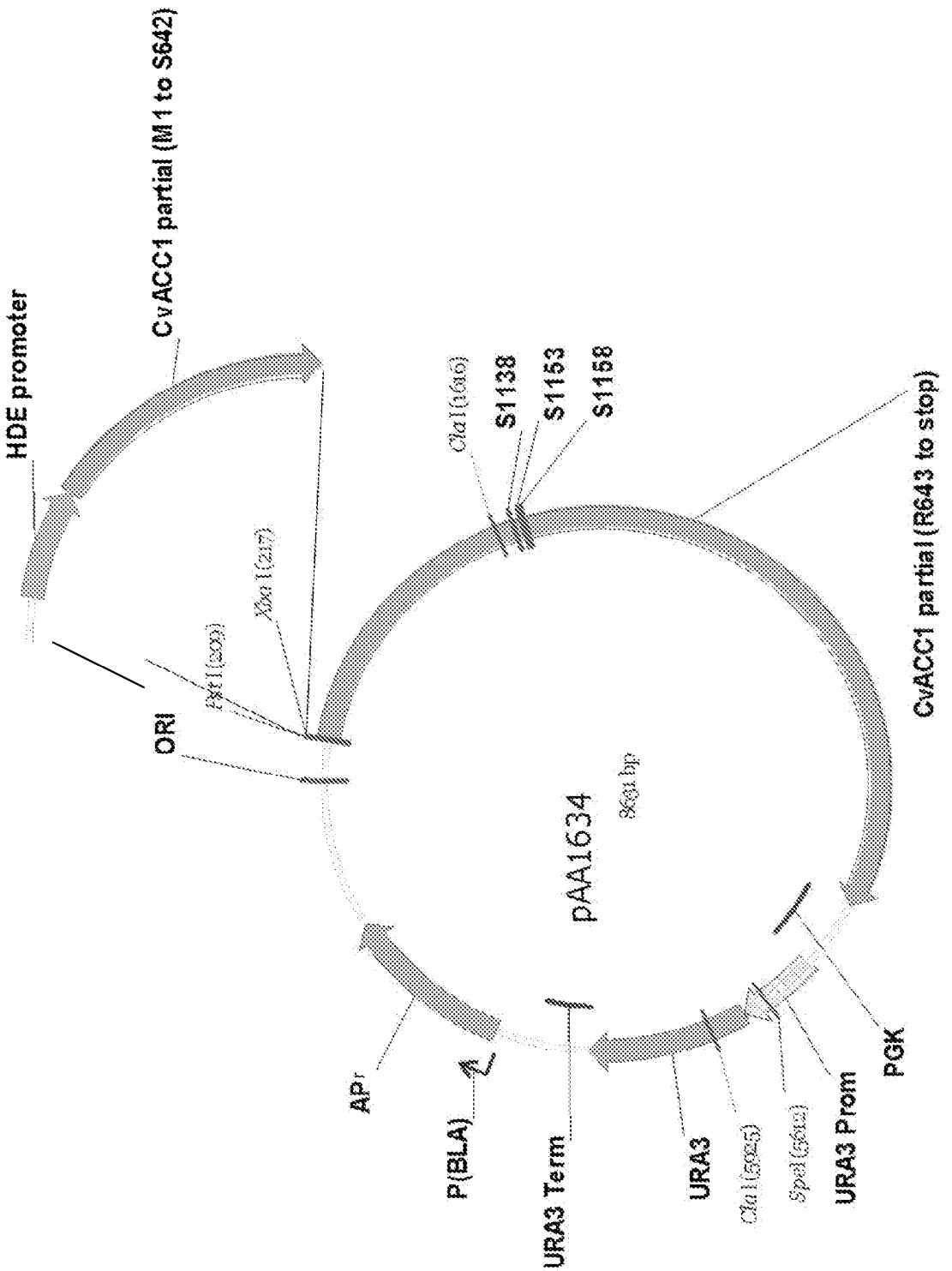

HDE promoter
CvACC1 partial (M1 to S642)
ORI
EstI (209)
XbaI (217)
ClaI (1616)
S1138
S1153
S1158
CvACC1 partial (R643 to stop)
pAA1634
8691 bp
APr
P(BLA)
URA3 Term
URA3
ClaI (6925)
SpeI (6932)
URA3 Prom
PGK

FIG. 30

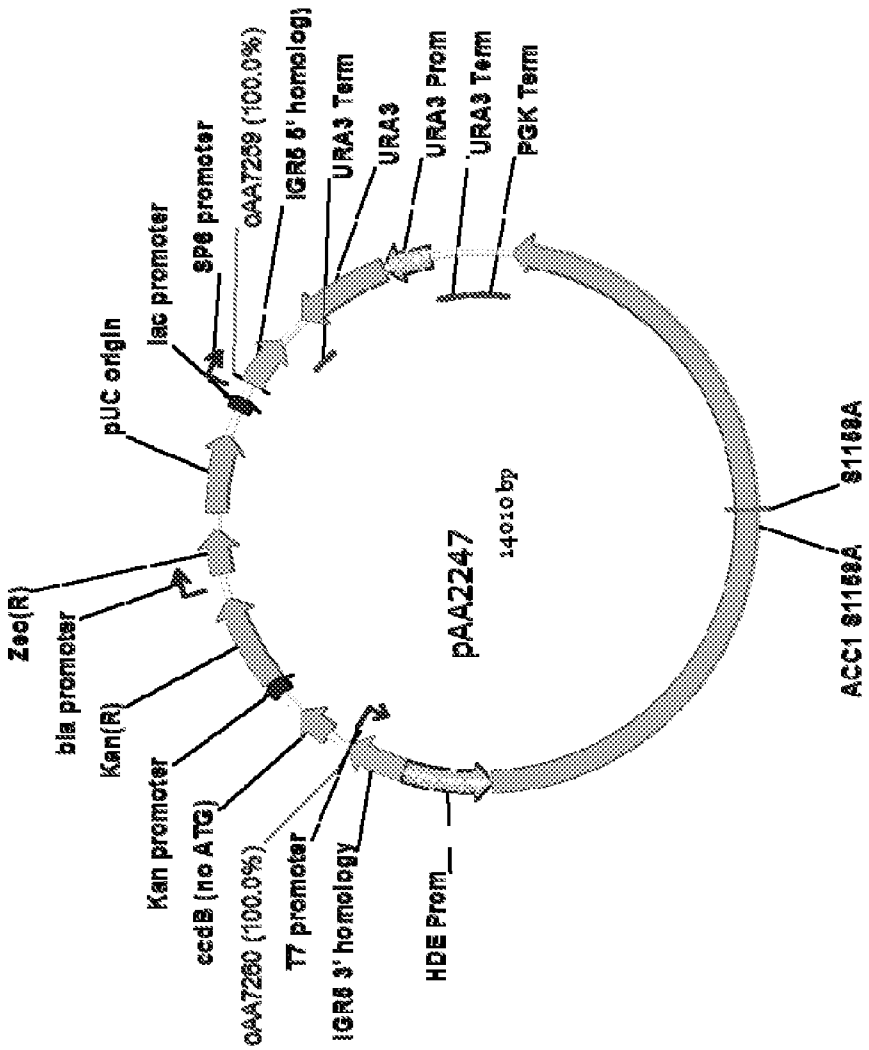

pAA2247
14010 bp
pUC origin
lac promoter
SP6 promoter
oAA7239 (100.0%)
iGR5 5' homology
URA3 Term
URA3
URA3 Prom
URA3 Term
PGK Term
S1168A
ACC1 S1168A
HDE Prom
iGR5 3' homology
T7 promoter
oAA7260 (100.0%)
ccdB (no ATG)
Kan promoter
Kan(R)
bla promoter
Zeo(R)

FIG. 31

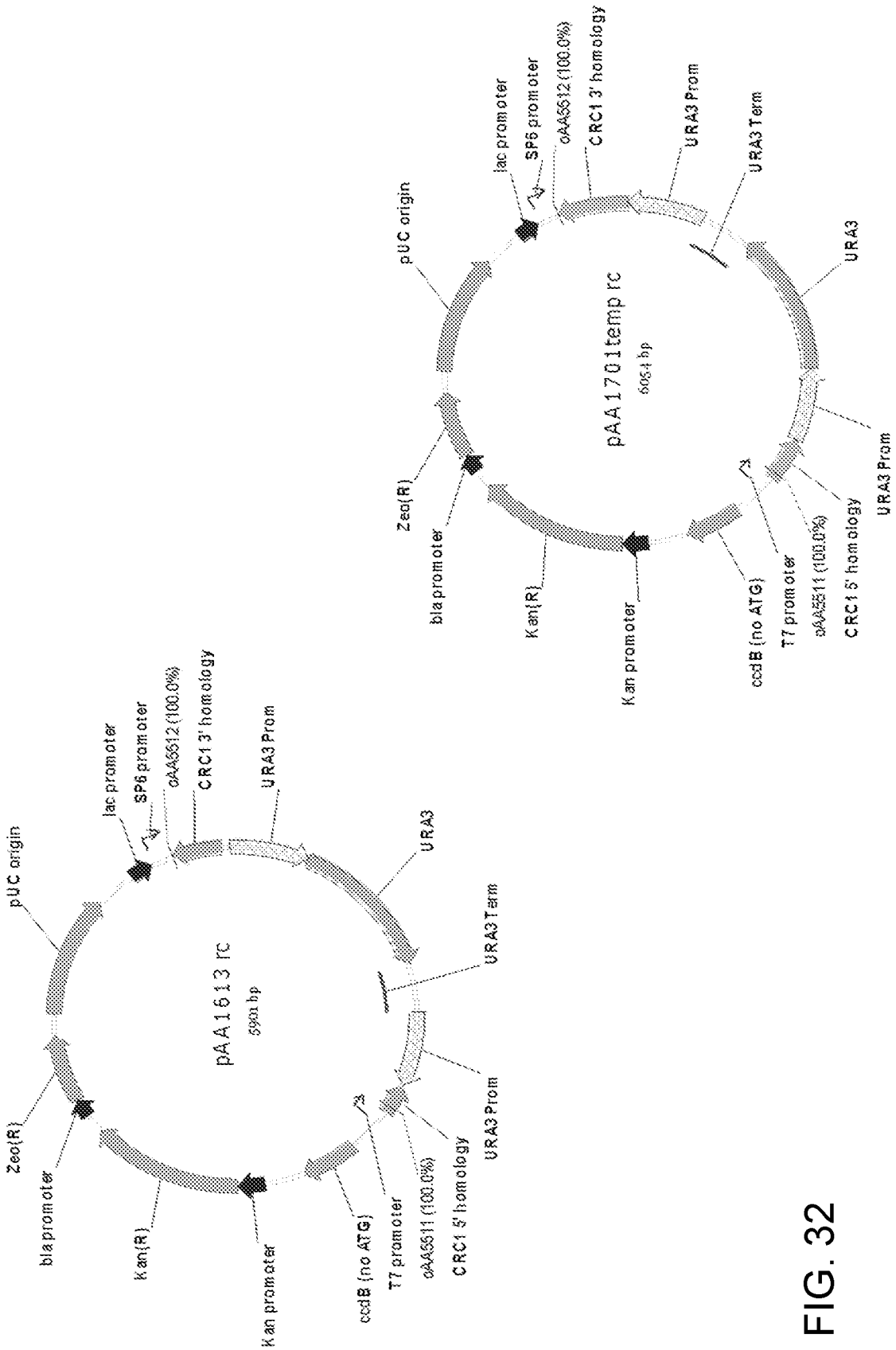

pAA1701temp rc
6054 bp
pUC origin
lac promoter
SP6 promoter
oAA5612 (100.0%)
CRC1 3' homology
URA3 Prom
URA3 Term
URA3
URA3 Prom
CRC1 5' homology
oAA5611 (100.0%)
T7 promoter
ccdB (no ATG)
Kan promoter
Kan(R)
bla promoter
Zeo(R)
pAA1613 rc
5900 bp
pUC origin
lac promoter
SP6 promoter
oAA5612 (100.0%)
CRC1 3' homology
URA3 Prom
URA3
URA3 Term
URA3 Prom
CRC1 5' homology
oAA5611 (100.0%)
T7 promoter
ccdB (no ATG)
Kan promoter
Kan(R)
bla promoter
Zeo(R)

FIG. 32

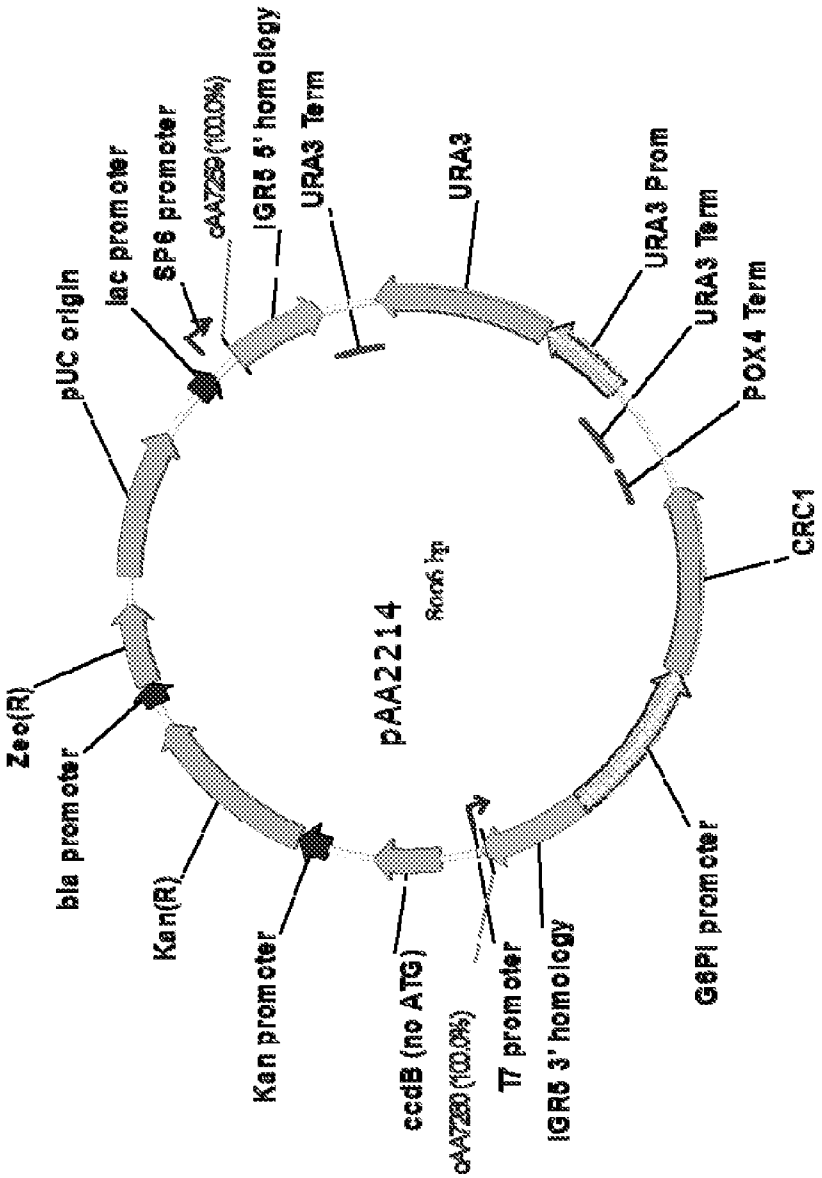

pAA2214
Bsmb bp
pUC origin
lac promoter
SP6 promoter
cAAZ289 (100.0%)
IGR5 5' homology
URA3 Term
URA3
URA3 Prom
URA3 Term
POX4 Term
CRC1
G6PI promoter
IGR5 3' homology
T7 promoter
cAAZ290 (100.0%)
ccdB (no ATG)
Kan promoter
Kan(R)
bla promoter
Zeo(R)

FIG. 33

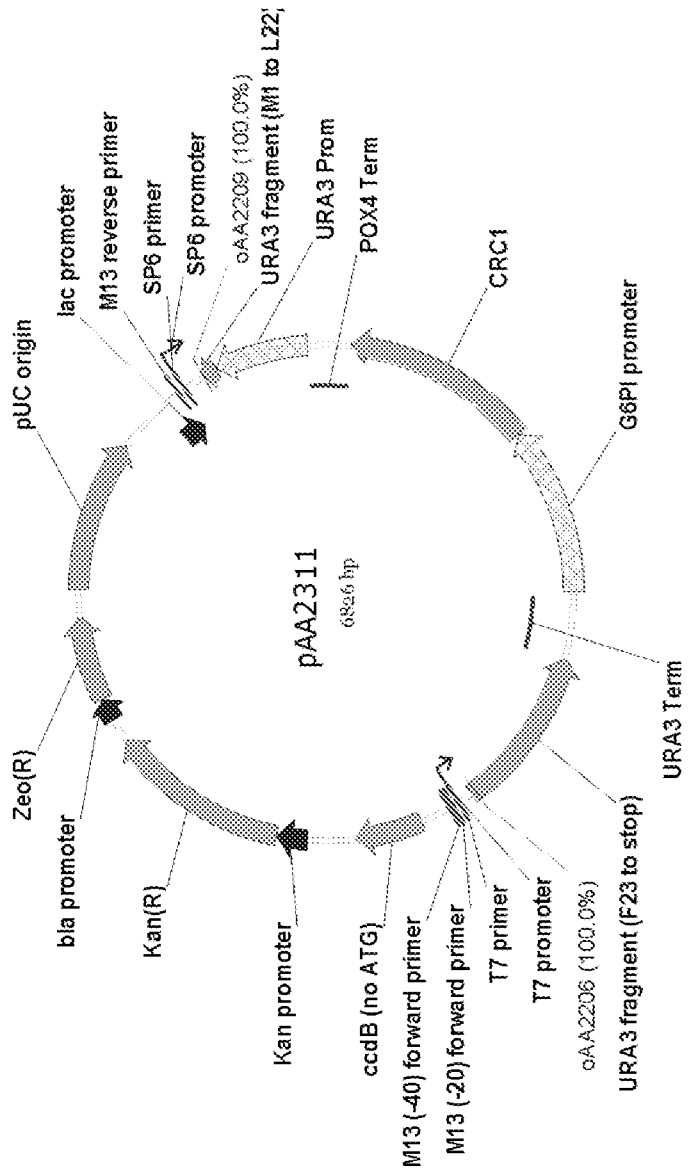

pAA2311
6826 bp
pUC origin
lac promoter
M13 reverse primer
SP6 primer
SP6 promoter
oAA2209 (100.0%)
URA3 fragment (M1 to L22
URA3 Prom
POX4 Term
CRC1
G6PI promoter
URA3 Term
URA3 fragment (F23 to stop)
oAA2206 (100.0%)
T7 promoter
T7 primer
M13 (-20) forward primer
M13 (-40) forward primer
ccdB (no ATG)
Kan promoter
Kan(R)
bla promoter
Zeo(R)

FIG. 34

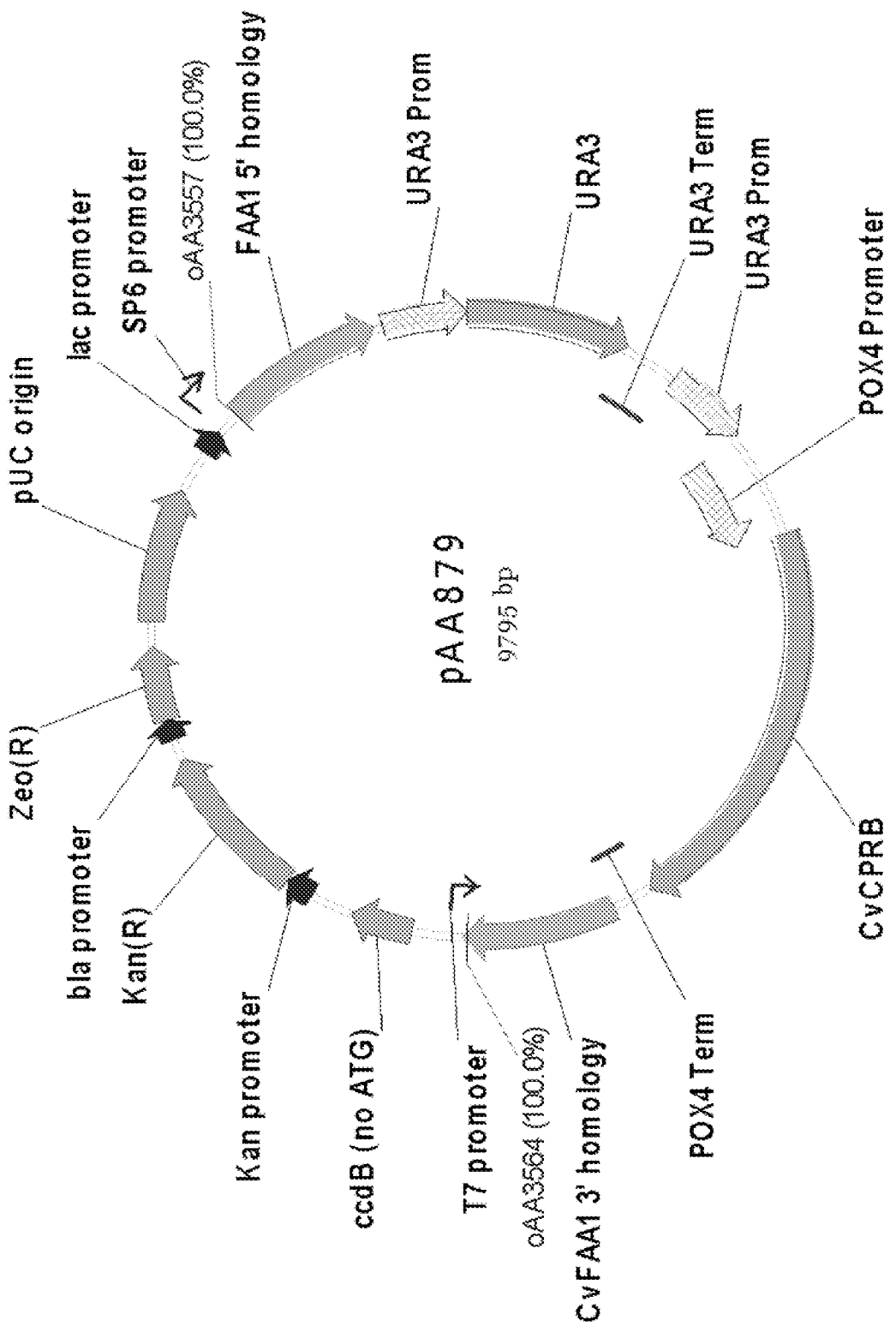

pAA879
9795 bp
pUC origin
lac promoter
SP6 promoter
oAA3557 (100.0%)
FAA1 5' homology
URA3 Prom
URA3
URA3 Term
URA3 Prom
POX4 Promoter
CvCPRB
POX4 Term
CvFAA1 3' homology
oAA3564 (100.0%)
T7 promoter
ccdB (no ATG)
Kan promoter
Kan(R)
bla promoter
Zeo(R)

FIG. 35

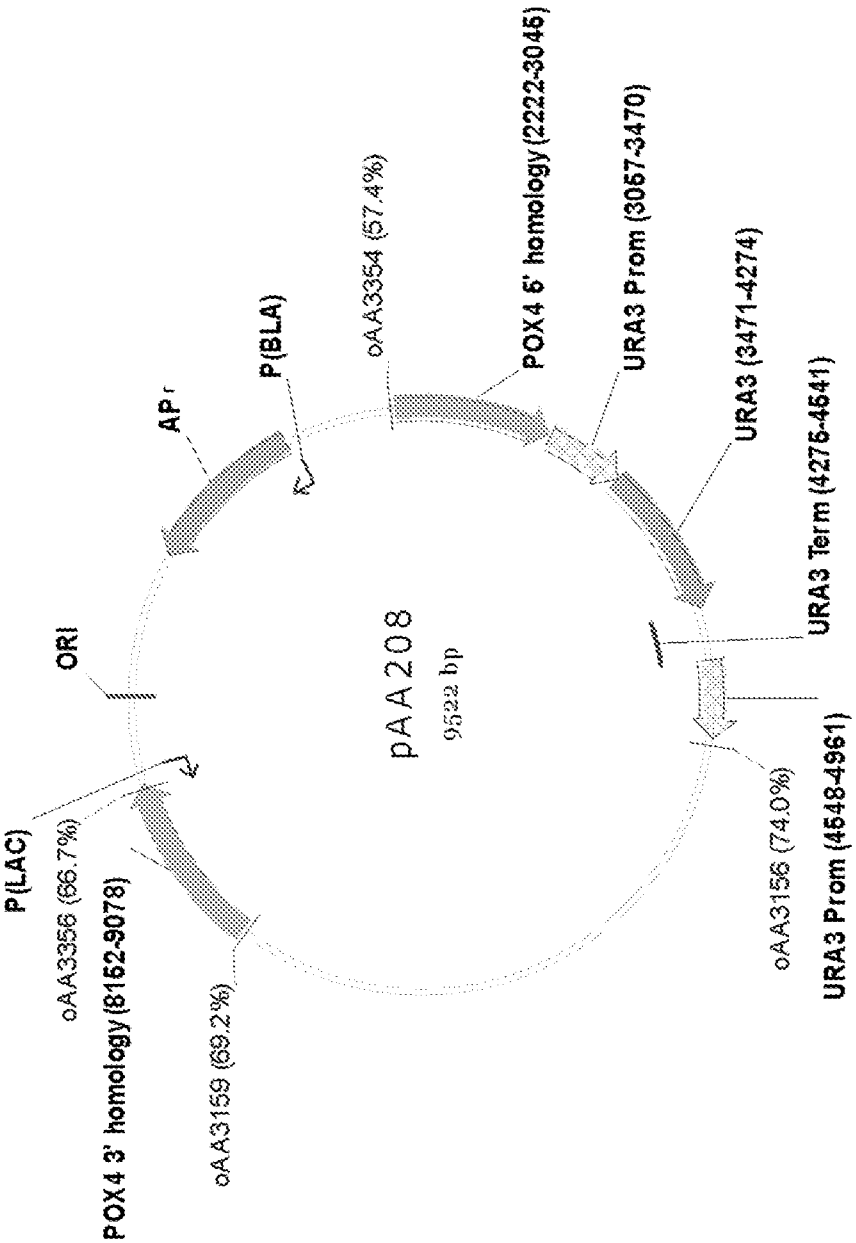

pAA208
9522 bp
ORI
AP r
P(BLA)
oAA3354 (57.4%)
POX4 5' homology (2222-3045)
URA3 Prom (3067-3470)
URA3 (3471-4274)
URA3 Term (4276-4641)
URA3 Prom (4648-4961)
oAA3156 (74.0%)
oAA3159 (69.2%)
POX4 3' homology (8152-9078)
oAA3356 (66.7%)
P(LAC)

FIG. 36

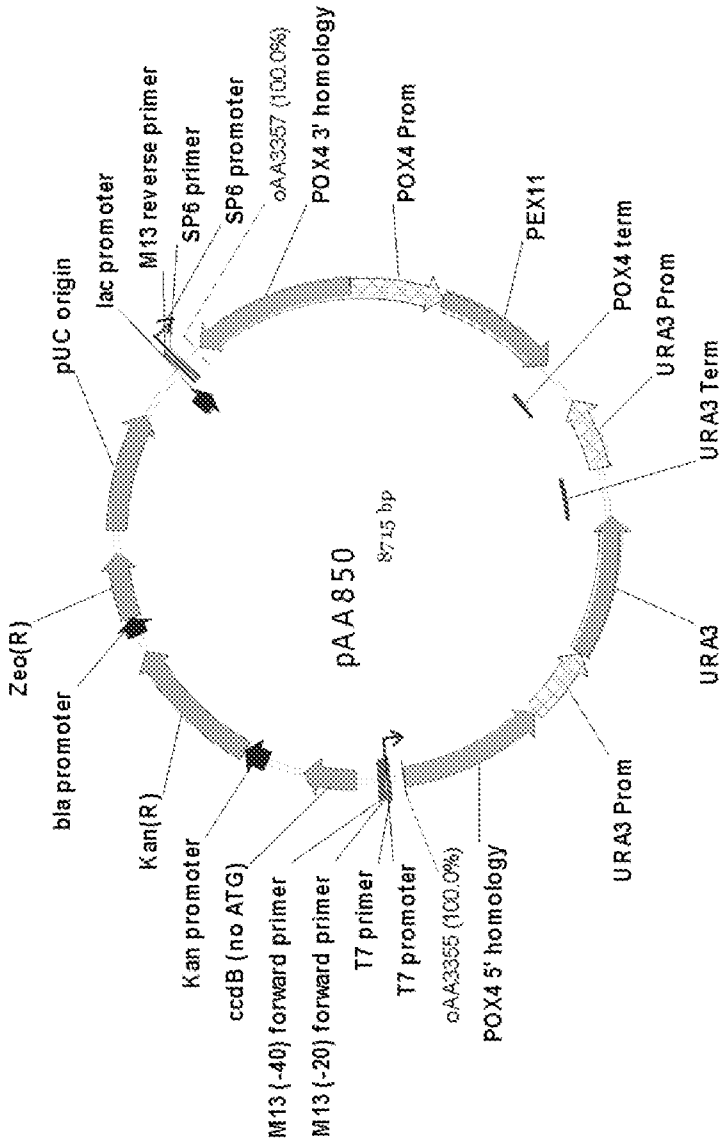

pAA850
8745 bp
pUC origin
lac promoter
M13 reverse primer
SP6 primer
SP6 promoter
oAA3357 (100.0%)
POX4 3' homology
POX4 Prom
PEX11
POX4 term
URA3 Prom
URA3 Term
URA3
URA3 Prom
POX4 5' homology
oAA3355 (100.0%)
T7 promoter
T7 primer
M13 (-20) forward primer
M13 (-40) forward primer
ccdB (no ATG)
Kan promoter
Kan(R)
bla promoter
Zeo(R)

FIG. 37

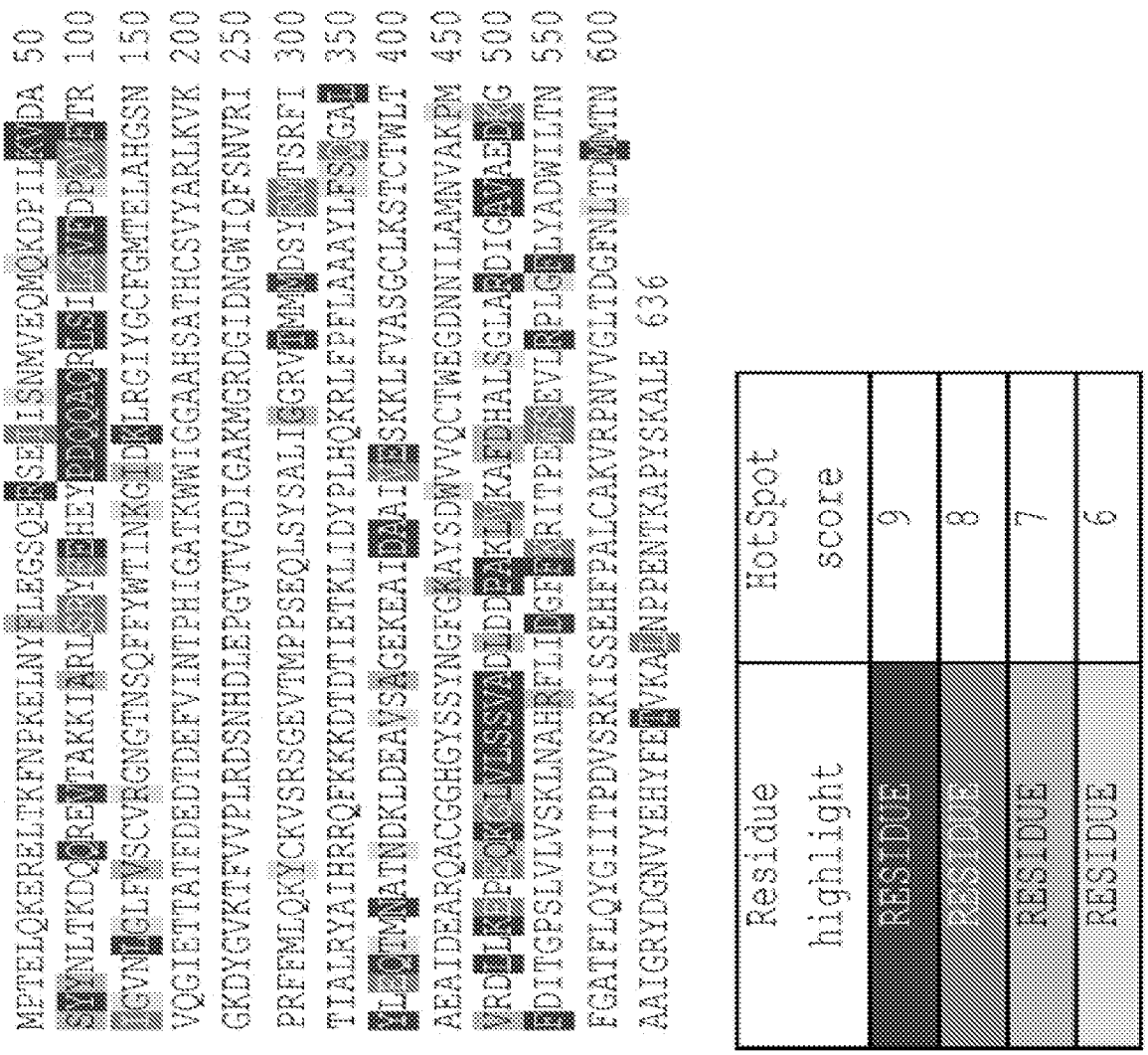


FIG. 38

Oligo A	oAA540
Oligo D	oAA541
Oligo C	F98G fwd
Oligo B	F98G rev
Oligo C	W429F fwd
Oligo B	W429F rev

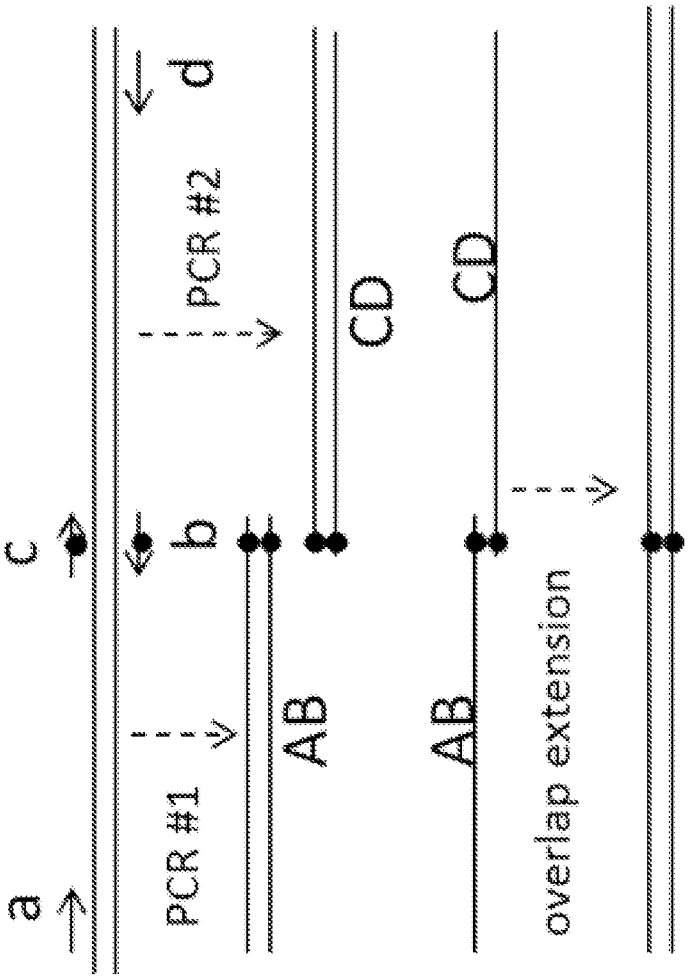


FIG. 39

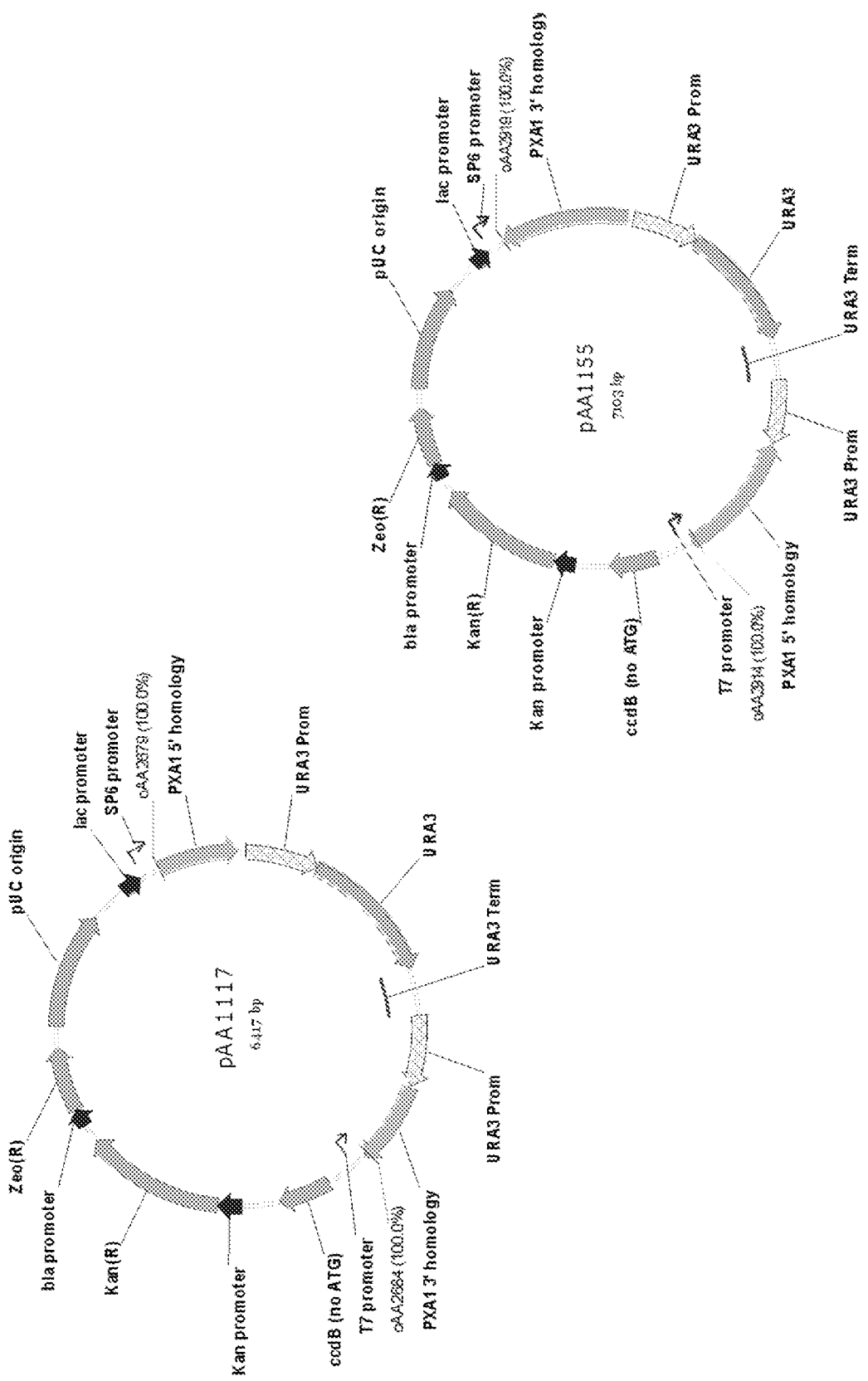

pAA1155
7103 bp
pUC origin
lac promoter
SP6 promoter
PXA1 3' homology
URA3 Prom
URA3
URA3 Term
URA3 Prom
PXA1 5' homology
T7 promoter
ccdB (no ATG)
Kan promoter
Kan(R)
bla promoter
Zeo(R)
pAA1117
6417 bp
pUC origin
lac promoter
SP6 promoter
PXA1 5' homology
URA3 Prom
URA3
URA3 Term
URA3 Prom
PXA1 3' homology
T7 promoter
ccdB (no ATG)
Kan promoter
Kan(R)
bla promoter
Zeo(R)

FIG. 40

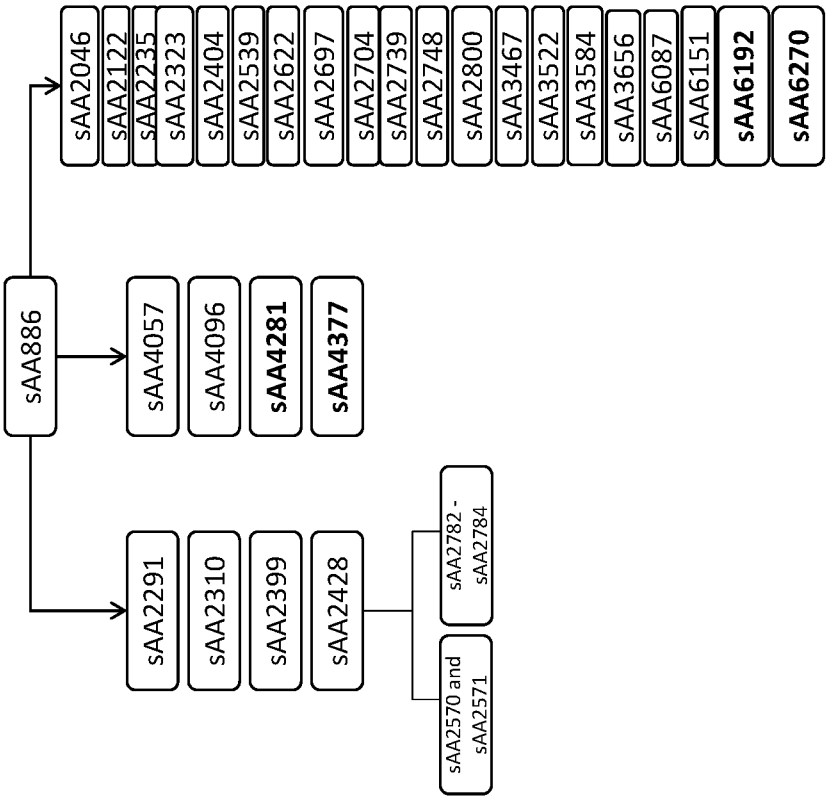

sAA886
sAA2046
sAA2122
sAA2235
sAA2323
sAA2404
sAA2539
sAA2622
sAA2697
sAA2704
sAA2739
sAA2748
sAA2800
sAA3467
sAA3522
sAA3584
sAA3656
sAA6087
sAA6151
sAA6192
sAA6270
sAA4057
sAA4096
sAA4281
sAA4377
sAA2291
sAA2310
sAA2399
sAA2428
sAA2782 - sAA2784
sAA2570 and sAA2571

FIG. 41

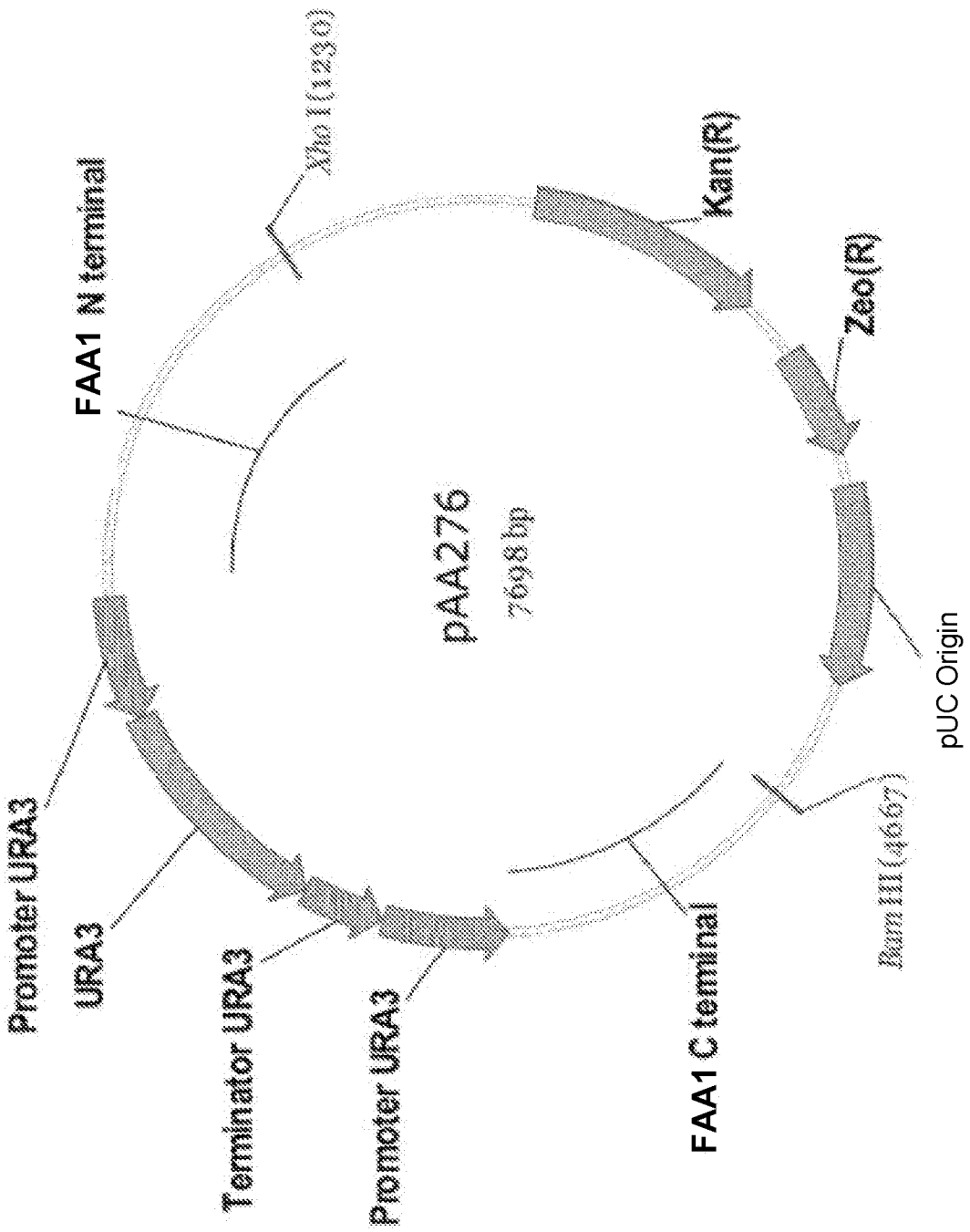

pAA276
7698 bp
Xho I (1230)
Kan(R)
Zeo(R)
pUC Origin
BamHI (4667)
FAA1 C terminal
Promoter URA3
Terminator URA3
URA3
Promoter URA3
FAA1 N terminal

FIG. 42

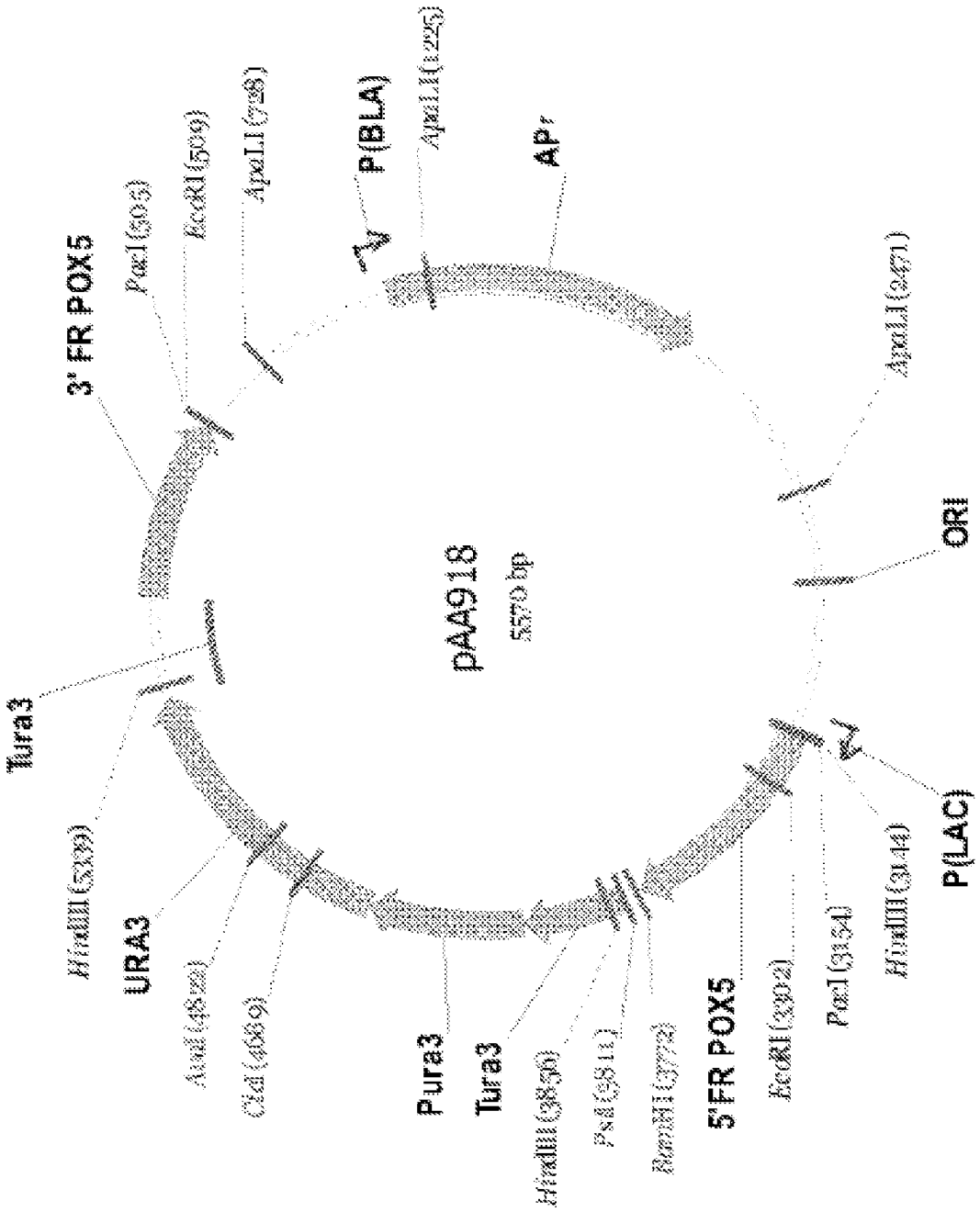

pAA918
5570 bp
3' FR POX5
Pac1(503)
EcoR1(509)
ApaLI(738)
P(BLA)
ApaLI(1225)
AP r
ApaLI(3471)
ORI
P(LAC)
HindIII(3144)
Pac1(3154)
EcoR1(3902)
5'FR POX5
BamHI(3772)
Pst1(3811)
HindIII(3850)
Tura3
Pura3
URA3
Tura3
HindIII(4329)
Aval(4824)
Cla1(4869)

FIG. 43

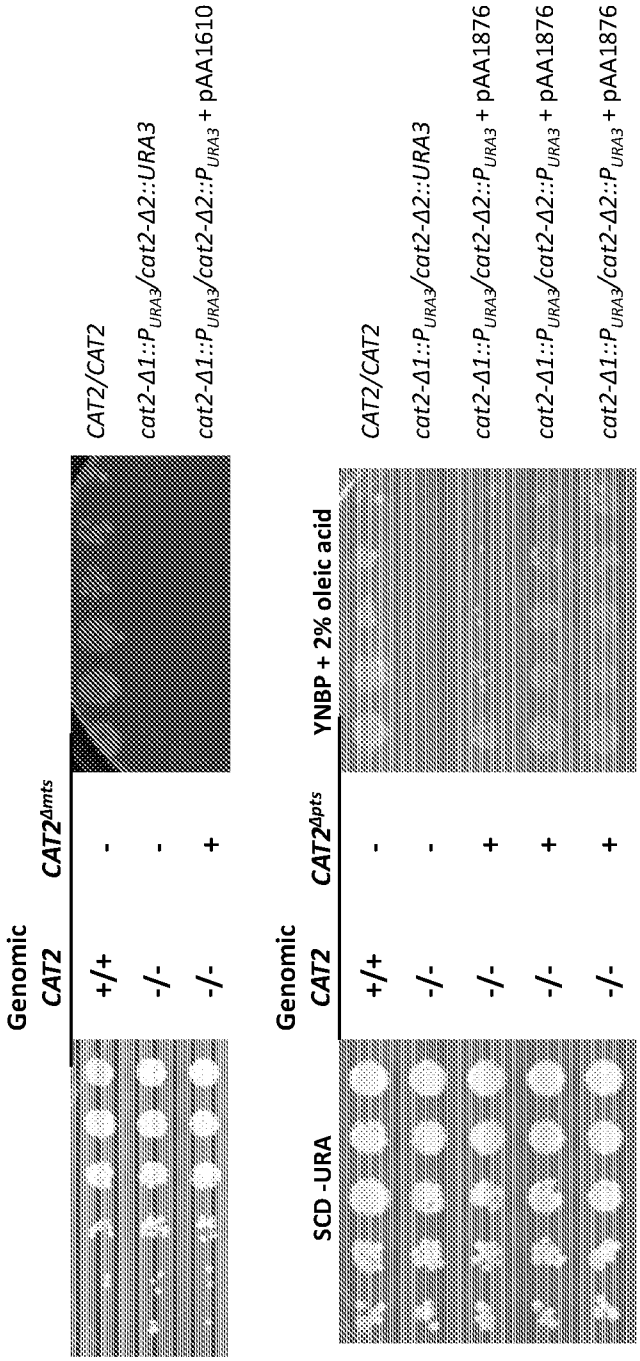

Genomic
CAT2
CAT2Δmts
+/+
-
-/-
-
-/-
+
CAT2/CAT2
cat2-Δ1::PURA3/cat2-Δ2::URA3
cat2-Δ1::PURA3/cat2-Δ2::PURA3 + pAA1610
SCD -URA
YNBP + 2% oleic acid
CAT2/CAT2
cat2-Δ1::PURA3/cat2-Δ2::URA3
cat2-Δ1::PURA3/cat2-Δ2::PURA3 + pAA1876
cat2-Δ1::PURA3/cat2-Δ2::PURA3 + pAA1876
cat2-Δ1::PURA3/cat2-Δ2::PURA3 + pAA1876

FIG. 44

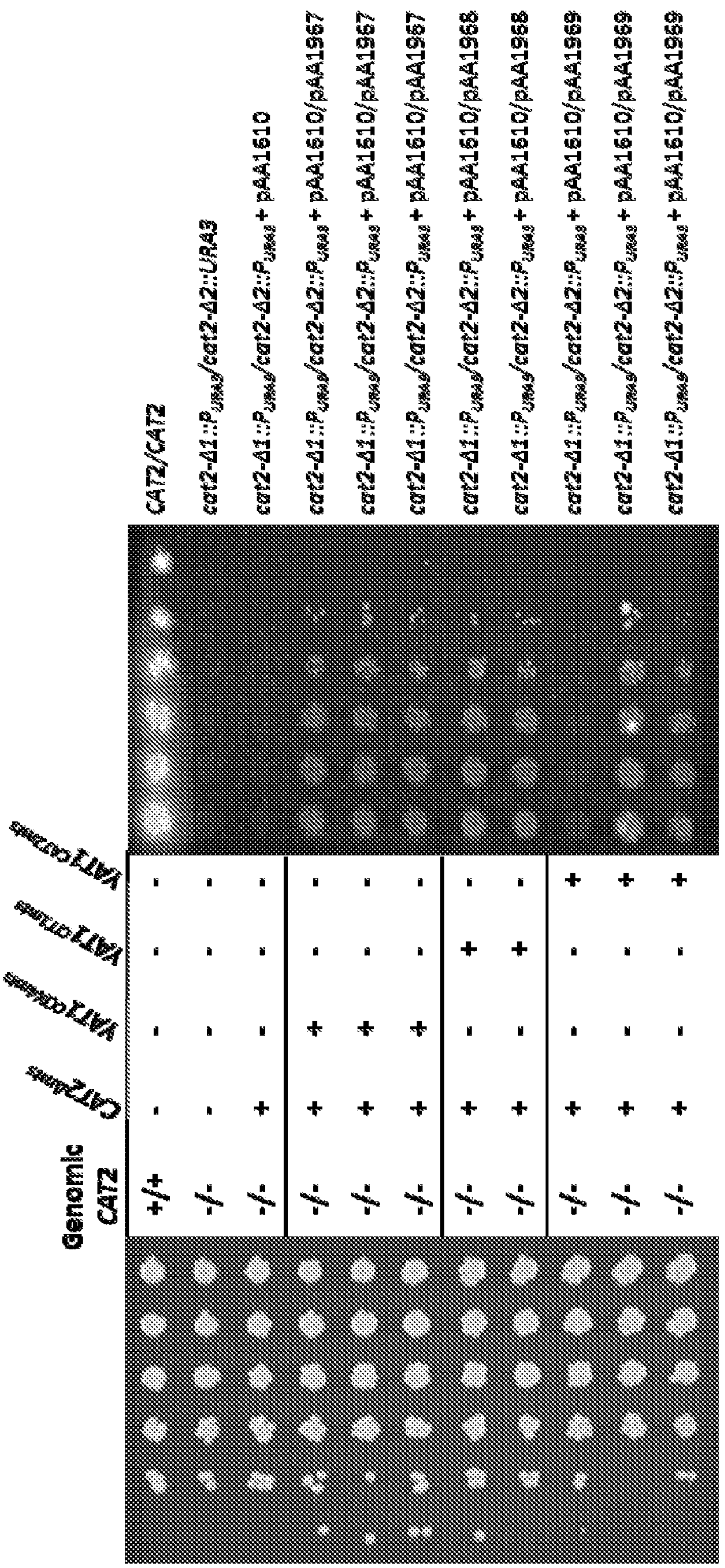

SCD -URA
Genomic CAT2
CAT2Δmts
YAT1CtCAT2mts
YAT1CtYAT1mts
YAT1CtYAT2mts
YNBP + 2% oleic acid
CAT2/CAT2
cat2-Δ1::PURA3/cat2-Δ2::URA3
cat2-Δ1::PURA3/cat2-Δ2::PURA3 + pAA1610
cat2-Δ1::PURA3/cat2-Δ2::PURA3 + pAA1610/pAA1967
cat2-Δ1::PURA3/cat2-Δ2::PURA3 + pAA1610/pAA1967
cat2-Δ1::PURA3/cat2-Δ2::PURA3 + pAA1610/pAA1967
cat2-Δ1::PURA3/cat2-Δ2::PURA3 + pAA1610/pAA1968
cat2-Δ1::PURA3/cat2-Δ2::PURA3 + pAA1610/pAA1968
cat2-Δ1::PURA3/cat2-Δ2::PURA3 + pAA1610/pAA1969
cat2-Δ1::PURA3/cat2-Δ2::PURA3 + pAA1610/pAA1969
cat2-Δ1::PURA3/cat2-Δ2::PURA3 + pAA1610/pAA1969

FIG. 45

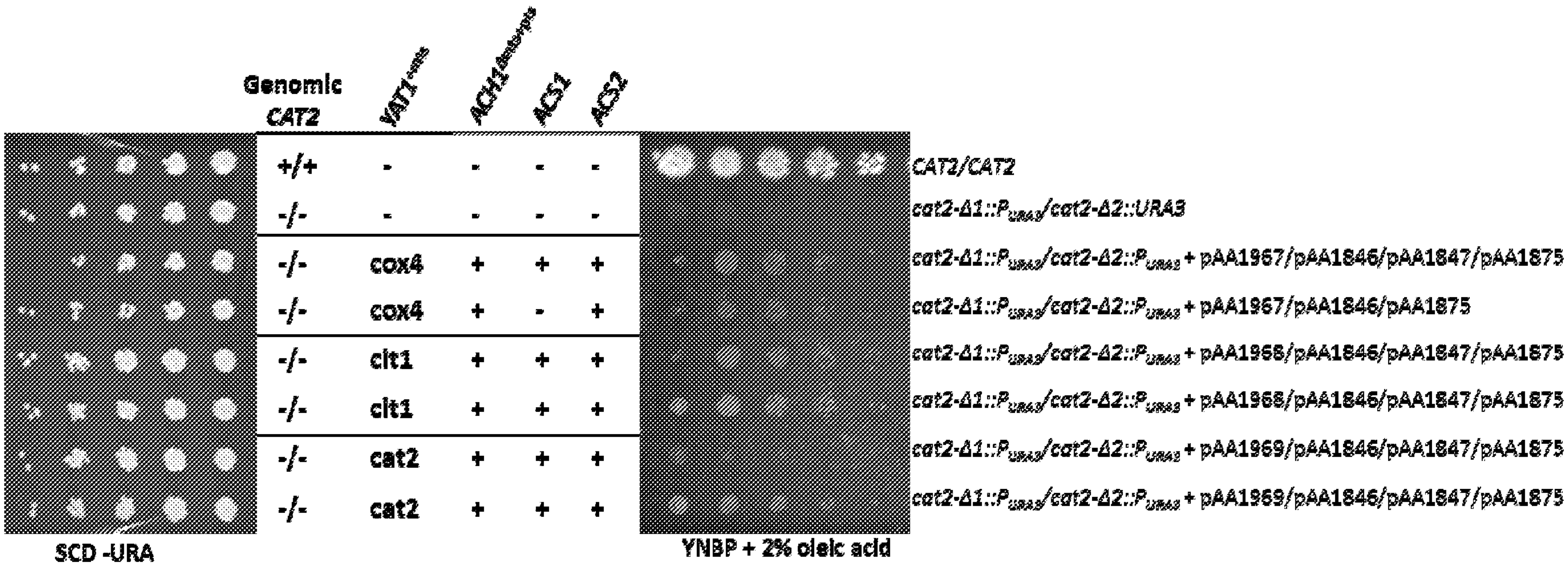

Genomic CAT2
VAT1
ACH1
ACS1
ACS2
+/+ - - - -
-/- - - - -
-/- cox4 + + +
-/- cox4 + - +
-/- cit1 + + +
-/- cit1 + + +
-/- cat2 + + +
-/- cat2 + + +
SCD -URA
YNBP + 2% oleic acid
CAT2/CAT2
cat2-Δ1::P_{URA3}/cat2-Δ2::URA3
cat2-Δ1::P_{URA3}/cat2-Δ2::P_{URA3} + pAA1967/pAA1846/pAA1847/pAA1875
cat2-Δ1::P_{URA3}/cat2-Δ2::P_{URA3} + pAA1967/pAA1846/pAA1875
cat2-Δ1::P_{URA3}/cat2-Δ2::P_{URA3} + pAA1968/pAA1846/pAA1847/pAA1875
cat2-Δ1::P_{URA3}/cat2-Δ2::P_{URA3} + pAA1968/pAA1846/pAA1847/pAA1875
cat2-Δ1::P_{URA3}/cat2-Δ2::P_{URA3} + pAA1969/pAA1846/pAA1847/pAA1875
cat2-Δ1::P_{URA3}/cat2-Δ2::P_{URA3} + pAA1969/pAA1846/pAA1847/pAA1875

FIG. 46

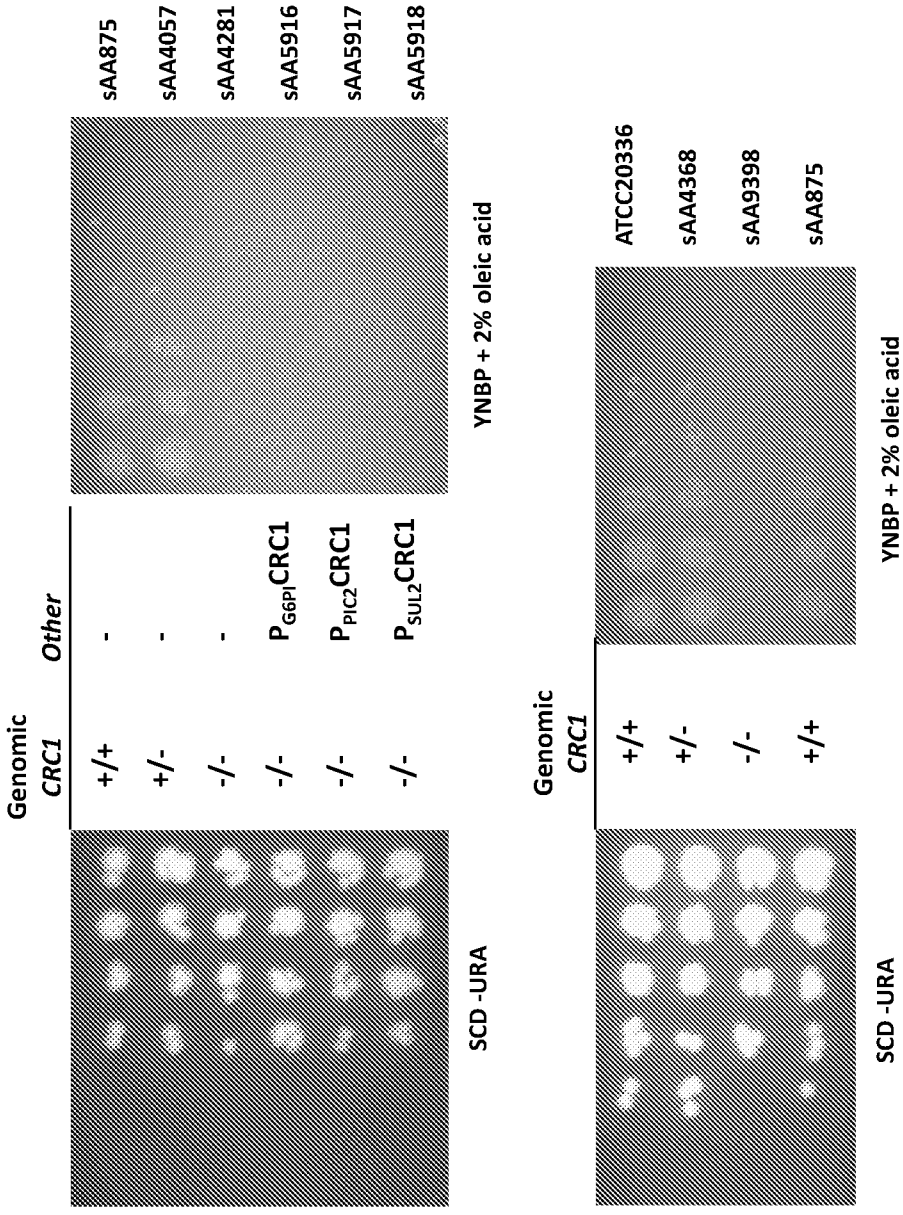

Genomic CRC1
Other
+/+
+/-
-/-
-/-
-/-
-/-
-
-
-
P G6PI CRC1
P PIC2 CRC1
P SUL2 CRC1
sAA875
sAA4057
sAA4281
sAA5916
sAA5917
sAA5918
SCD -URA
YNBP + 2% oleic acid
Genomic CRC1
+/+
+/-
-/-
+/+
ATCC20336
sAA4368
sAA9398
sAA875
SCD -URA
YNBP + 2% oleic acid

FIG. 47

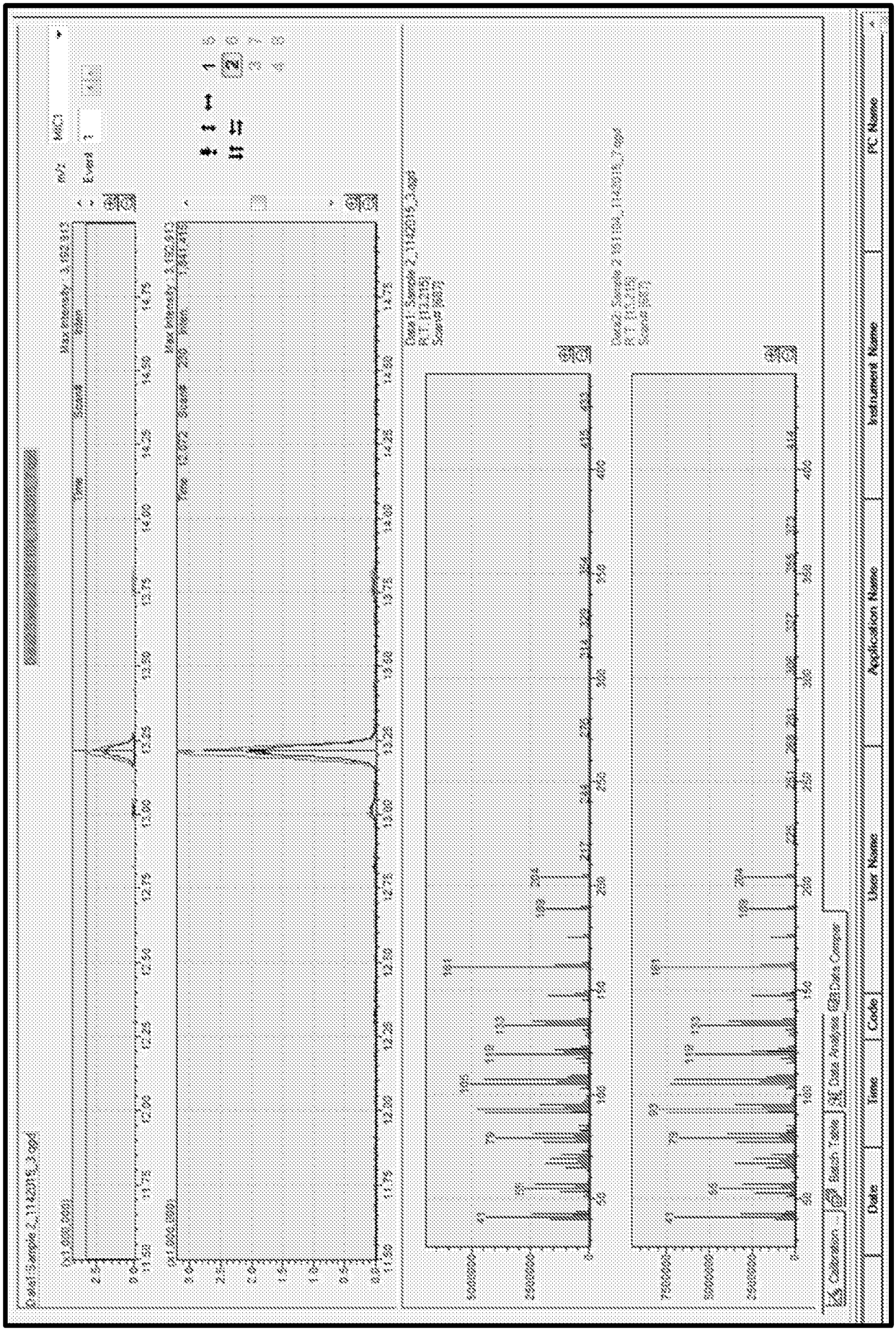


Fig. 48A

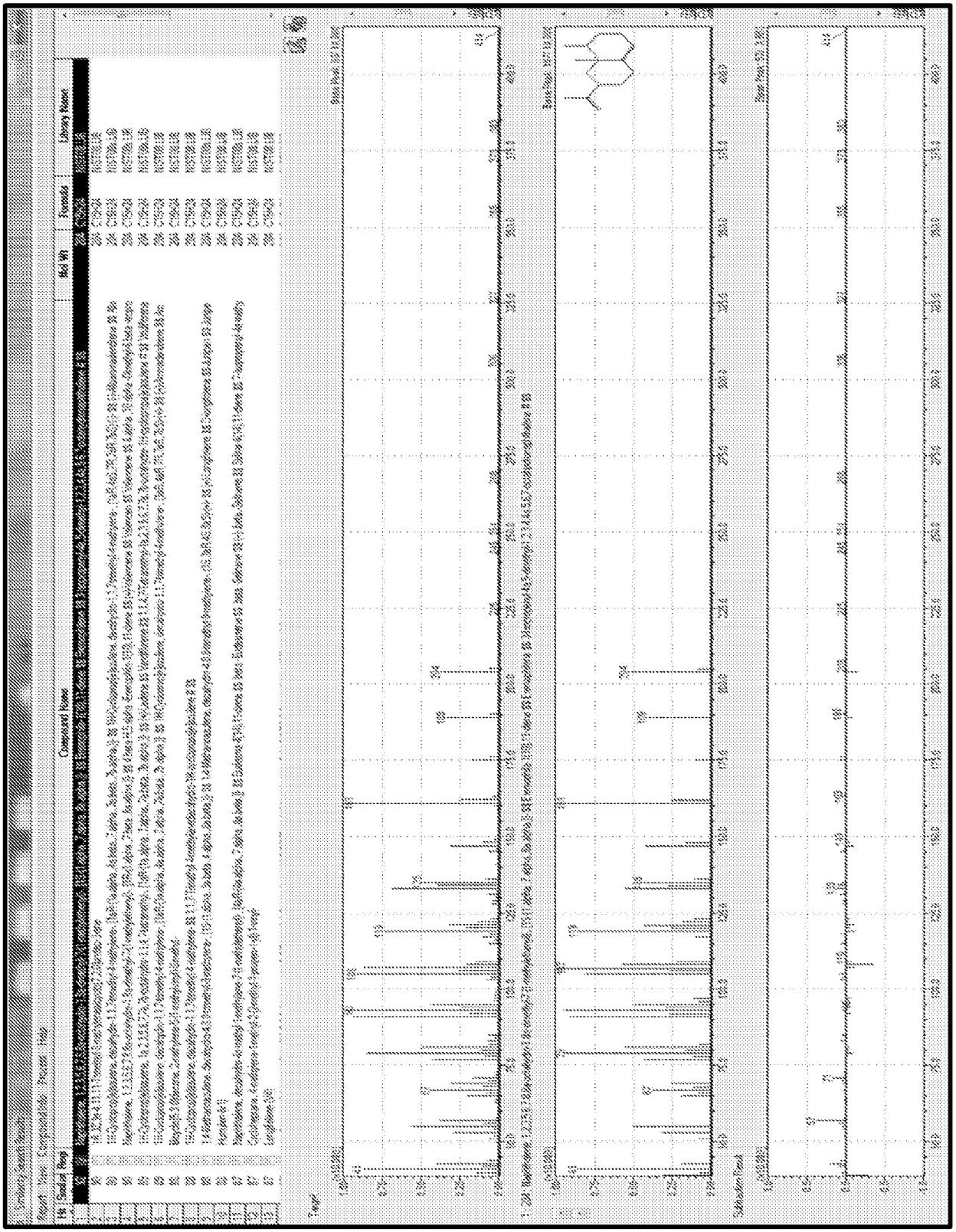


Fig. 48B

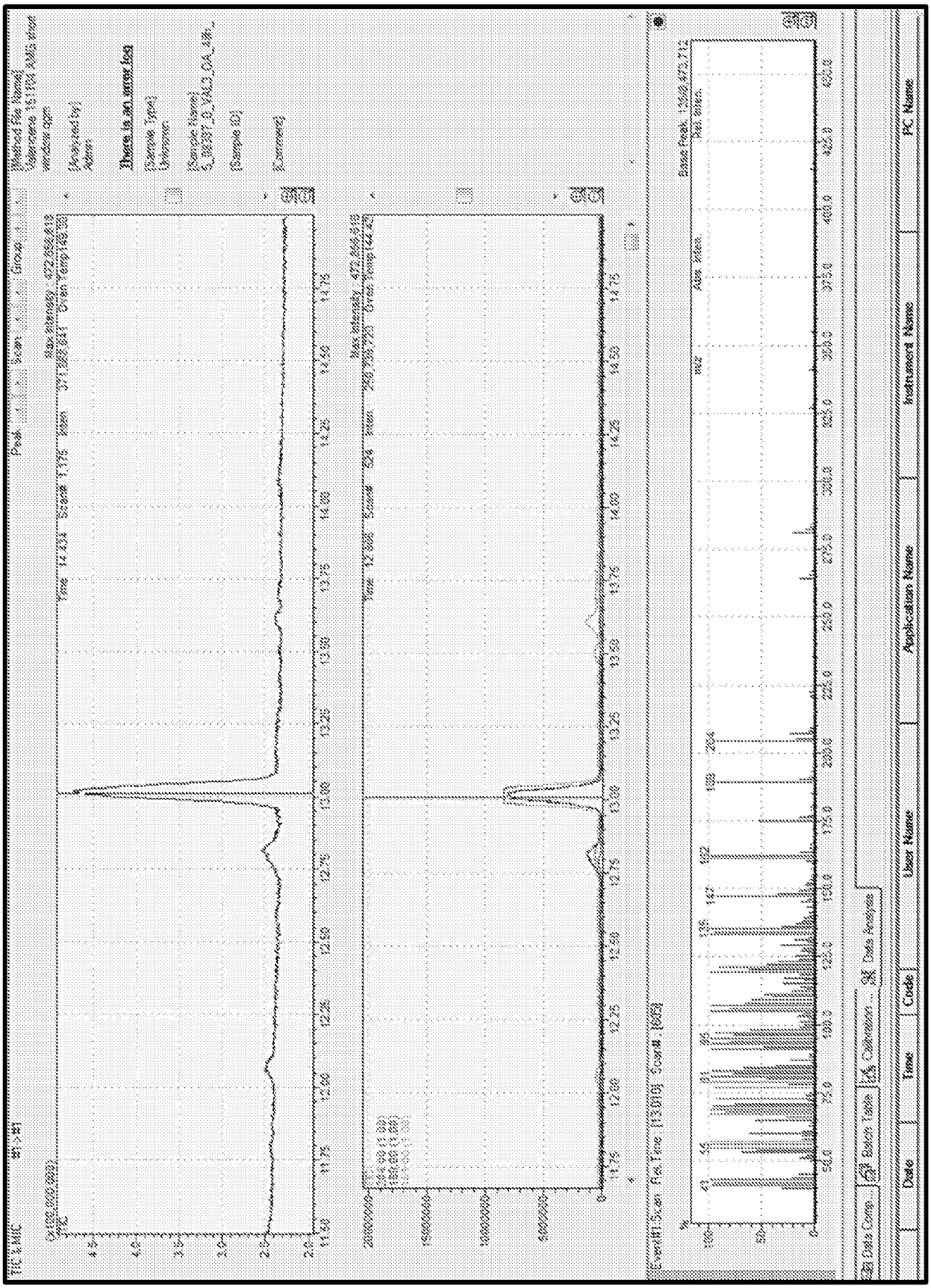


Fig. 49A

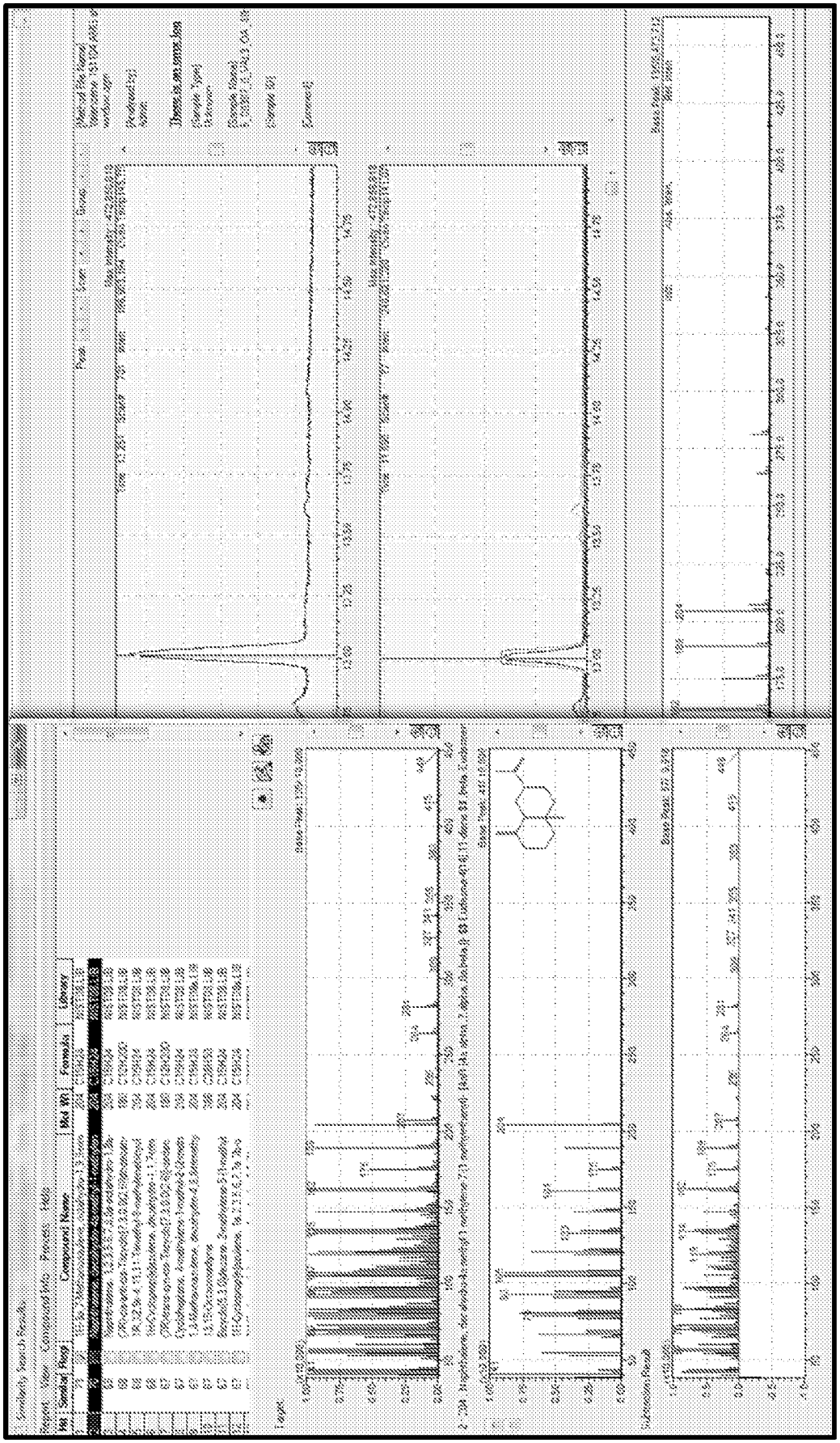


Fig. 49B

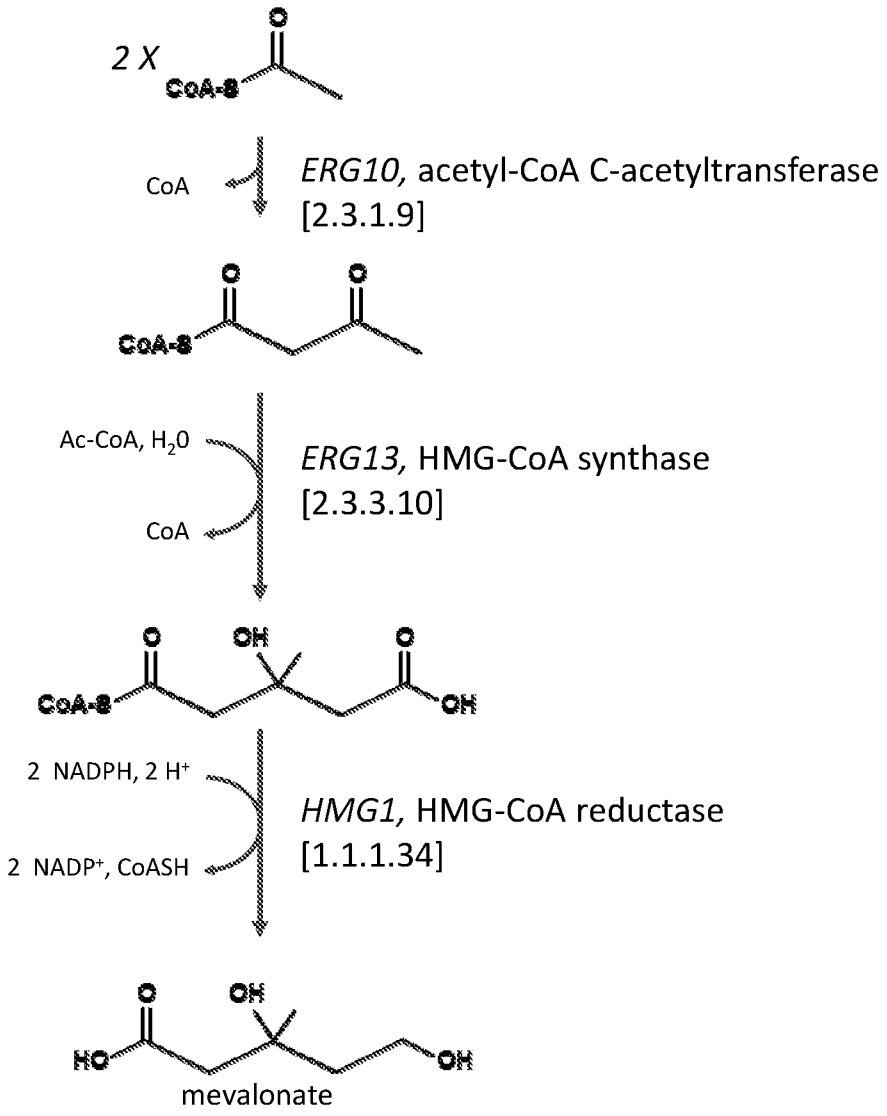

2 X
CoA-S
CoA
ERG10, acetyl-CoA C-acetyltransferase
[2.3.1.9]
CoA-S
Ac-CoA, H2O
CoA
ERG13, HMG-CoA synthase
[2.3.3.10]
CoA-S
OH
OH
2 NADPH, 2 H+
2 NADP+, CoASH
HMG1, HMG-CoA reductase
[1.1.1.34]
HO
OH
OH
mevalonate

Fig. 50

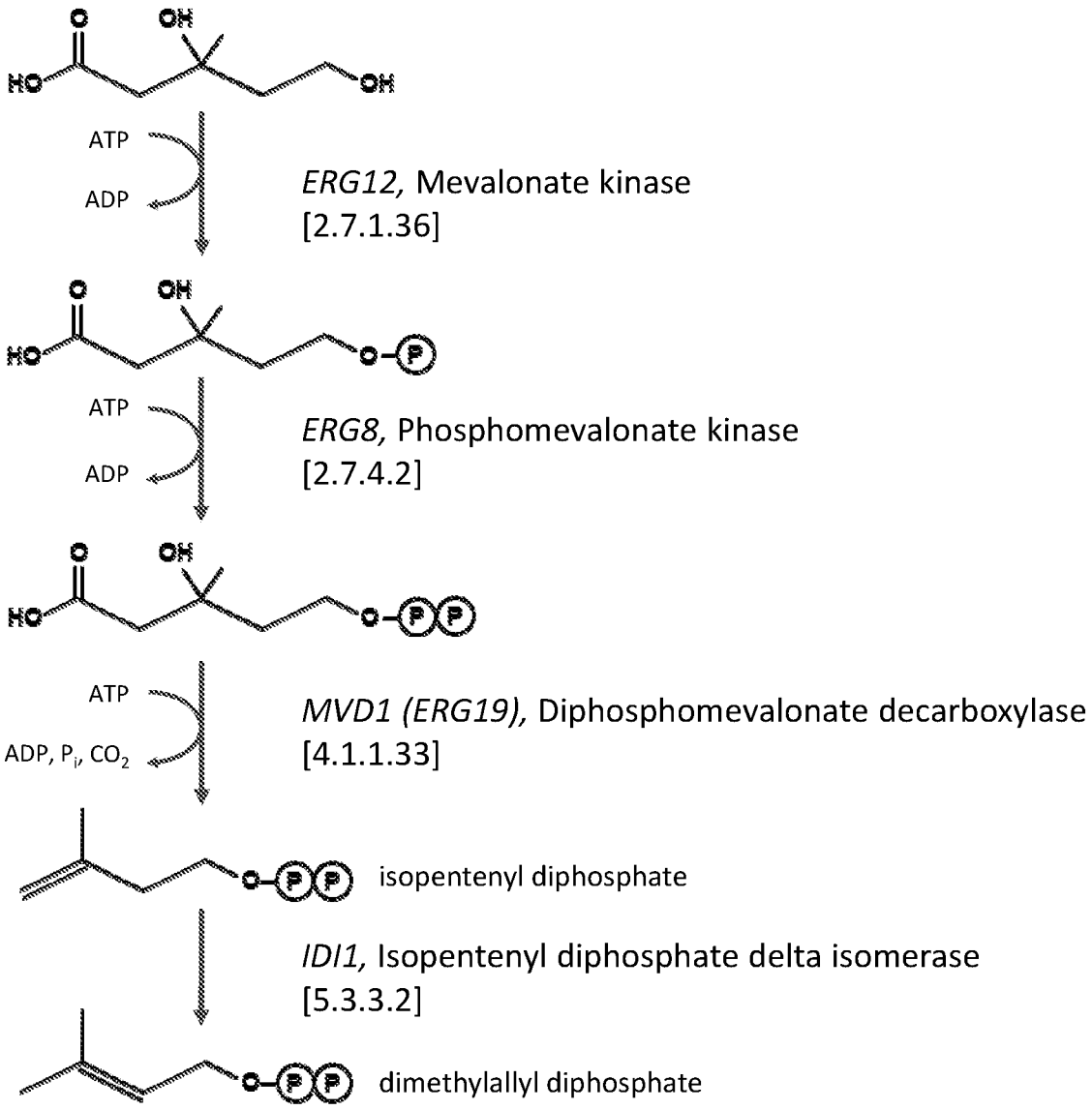

ATP
ADP
ERG12, Mevalonate kinase
[2.7.1.36]
ATP
ADP
ERG8, Phosphomevalonate kinase
[2.7.4.2]
ATP
ADP, P_i, CO_2
MVD1 (ERG19), Diphosphomevalonate decarboxylase
[4.1.1.33]
isopentenyl diphosphate
IDI1, Isopentenyl diphosphate delta isomerase
[5.3.3.2]
dimethylallyl diphosphate

Fig. 51

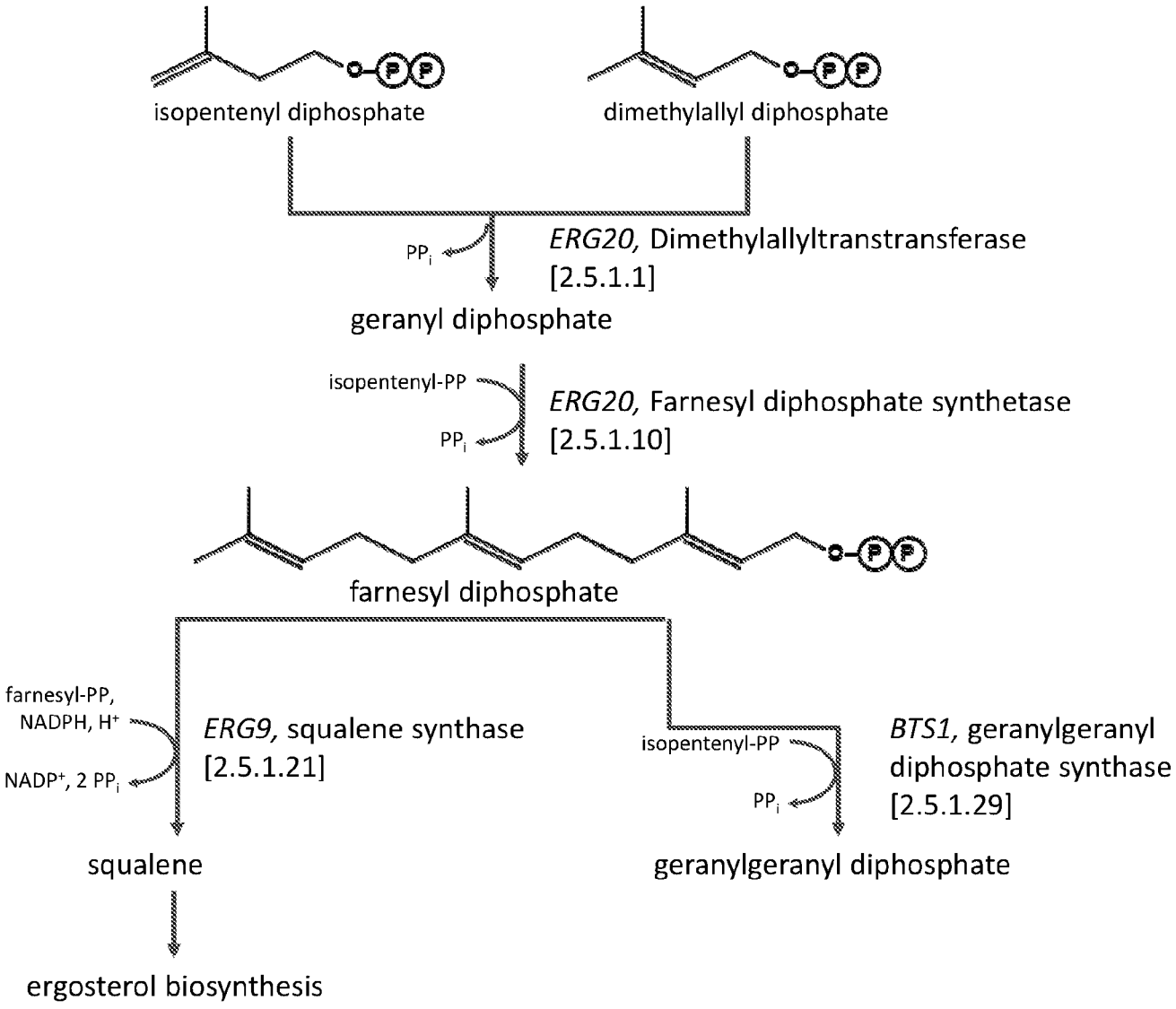

isopentenyl diphosphate
dimethylallyl diphosphate
PPi
ERG20, Dimethylallyltranstransferase
[2.5.1.1]
geranyl diphosphate
isopentenyl-PP
PPi
ERG20, Farnesyl diphosphate synthetase
[2.5.1.10]
farnesyl diphosphate
farnesyl-PP,
NADPH, H+
NADP+, 2 PPi
ERG9, squalene synthase
[2.5.1.21]
isopentenyl-PP
PPi
BTS1, geranylgeranyl
diphosphate synthase
[2.5.1.29]
squalene
geranylgeranyl diphosphate
ergosterol biosynthesis

Fig. 52

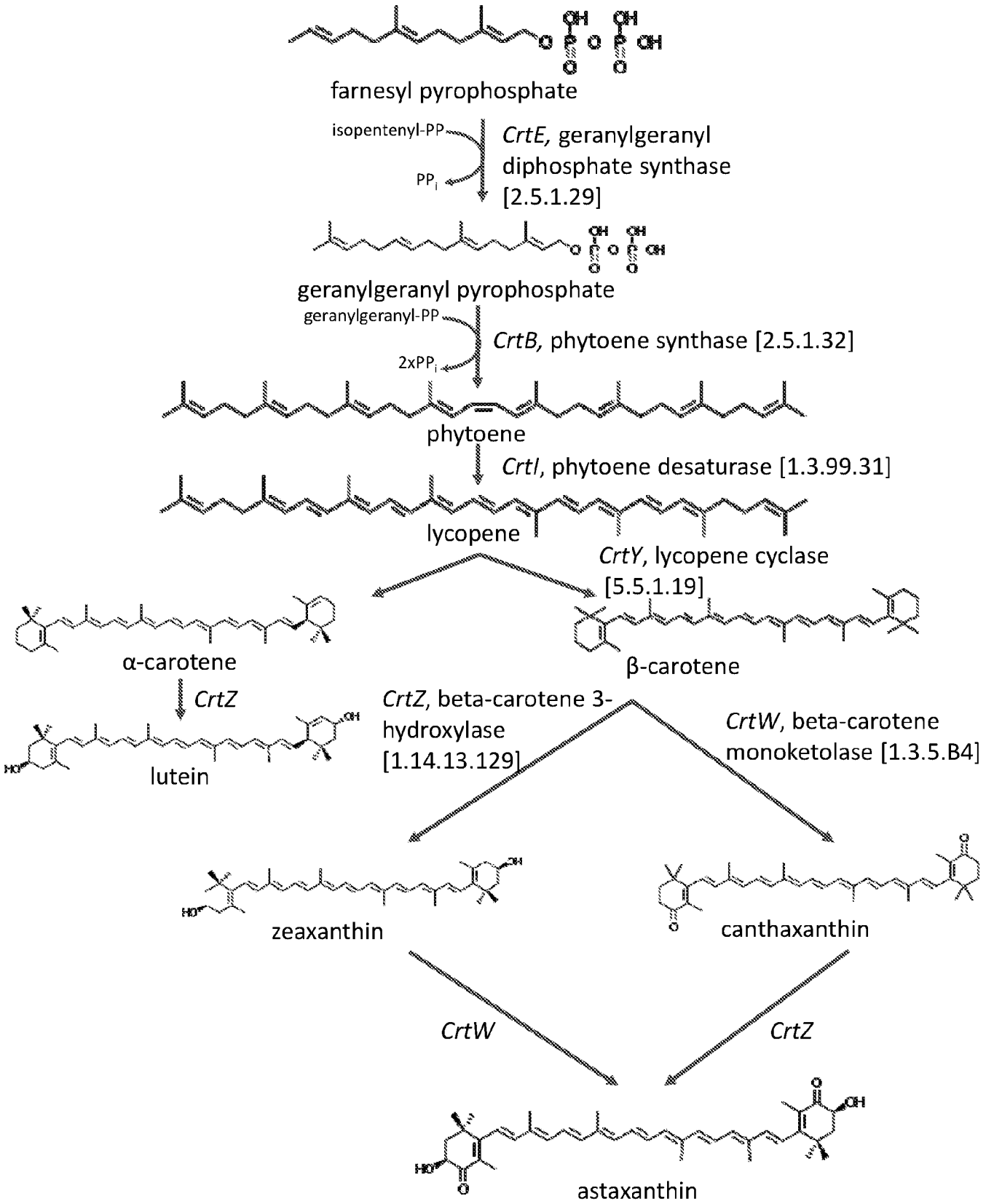

farnesyl pyrophosphate
isopentenyl-PP
PPi
CrtE, geranylgeranyl diphosphate synthase [2.5.1.29]
geranylgeranyl pyrophosphate
geranylgeranyl-PP
2xPPi
CrtB, phytoene synthase [2.5.1.32]
phytoene
CrtI, phytoene desaturase [1.3.99.31]
lycopene
CrtY, lycopene cyclase [5.5.1.19]
α-carotene
β-carotene
CrtZ
lutein
CrtZ, beta-carotene 3-hydroxylase [1.14.13.129]
CrtW, beta-carotene monoketolase [1.3.5.B4]
zeaxanthin
canthaxanthin
CrtW
CrtZ
astaxanthin

Fig. 53

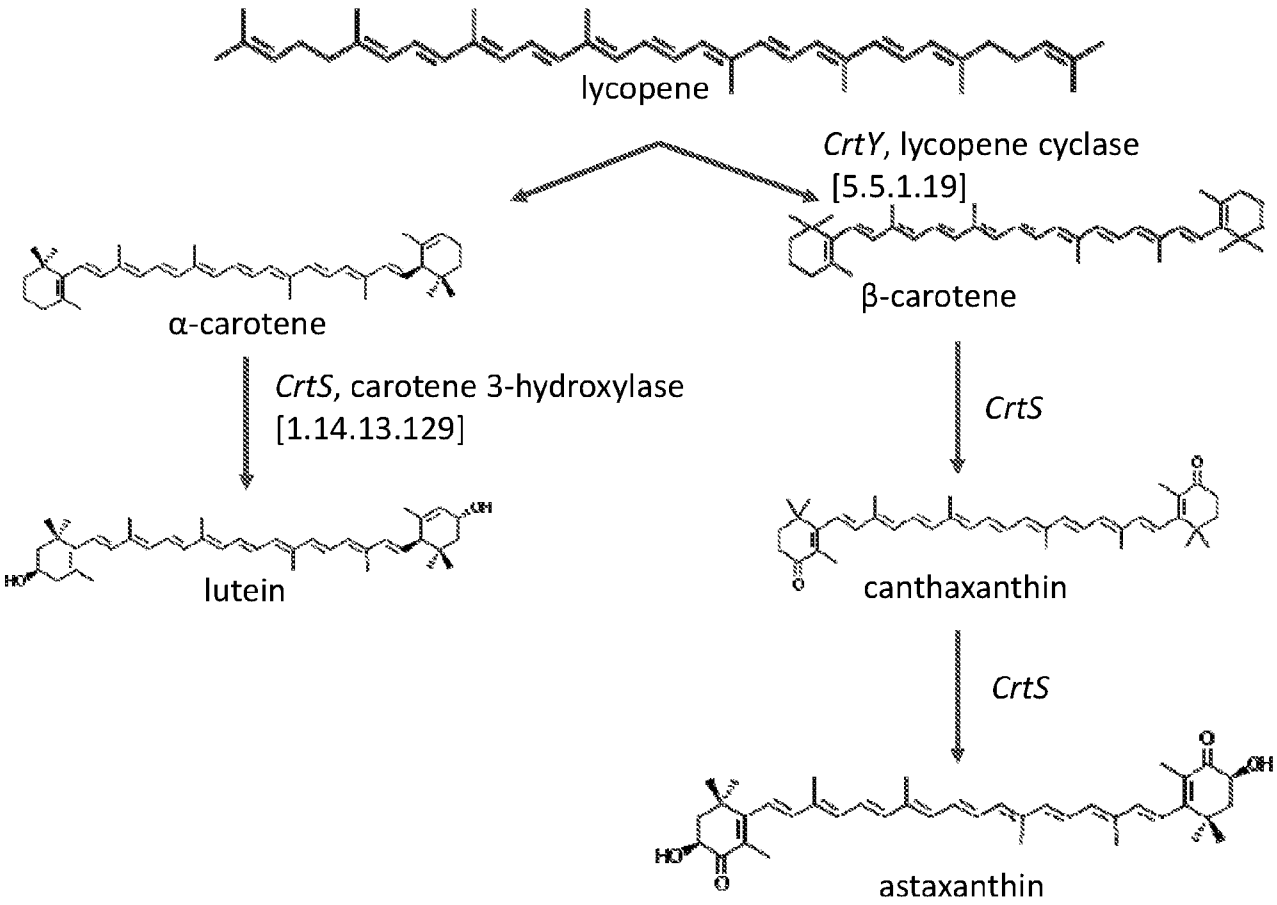
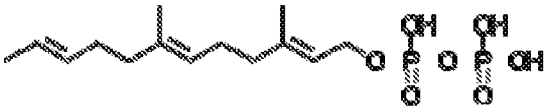

lycopene
CrtY, lycopene cyclase
[5.5.1.19]
α-carotene
β-carotene
CrtS, carotene 3-hydroxylase
[1.14.13.129]
CrtS
lutein
canthaxanthin
CrtS
astaxanthin

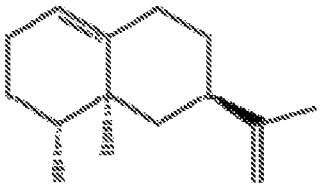
Fig. 54



farnesyl pyrophosphate

PP_i

Tps1, valencene synthase
[4.2.3.73]



valencene

Fig. 55

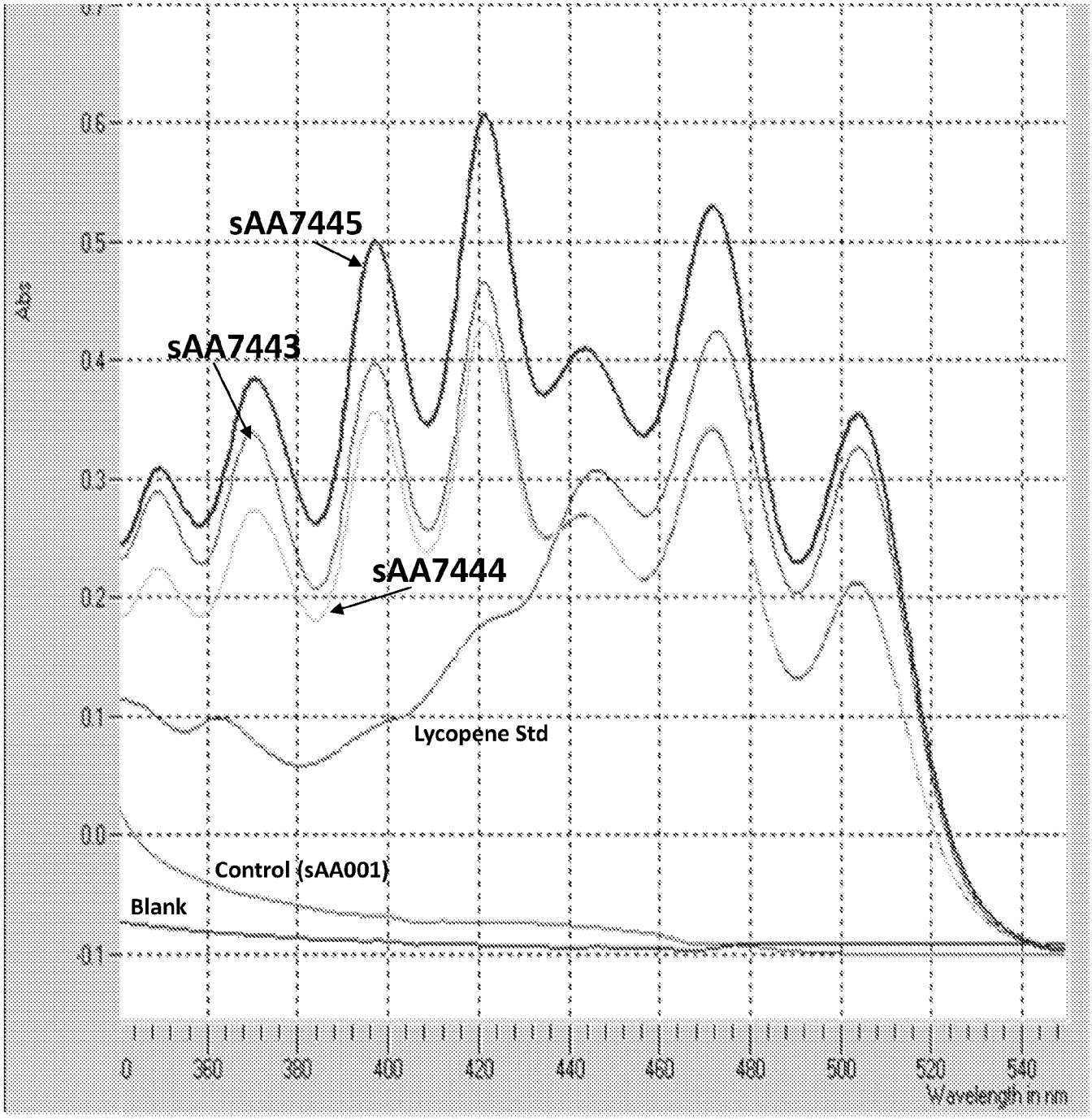

sAA7445
sAA7443
sAA7444
Lycopene Std
Control (sAA001)
Blank
AU
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
0
360
380
400
420
440
460
480
500
520
540
Wavelength in nm

Fig. 56

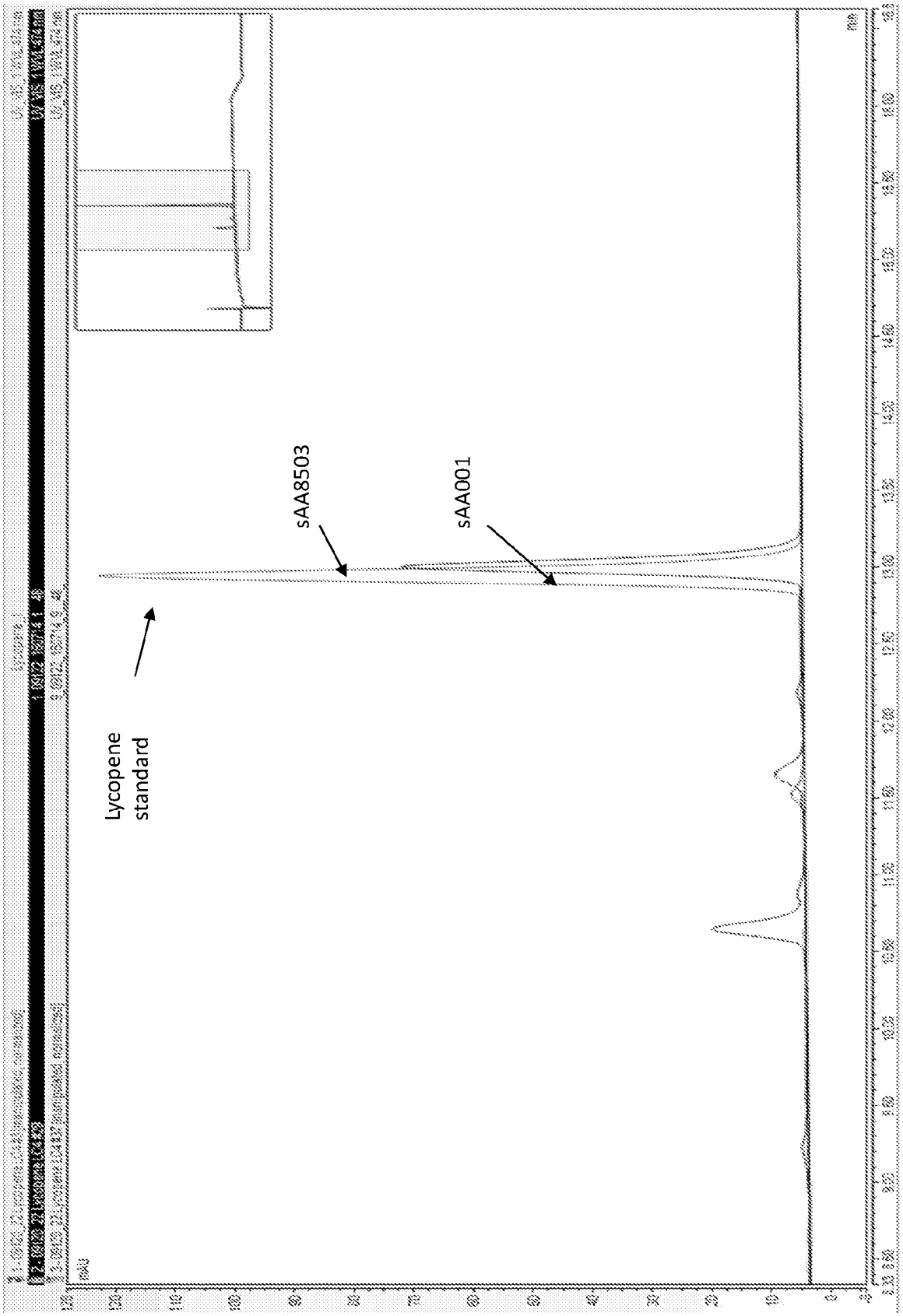
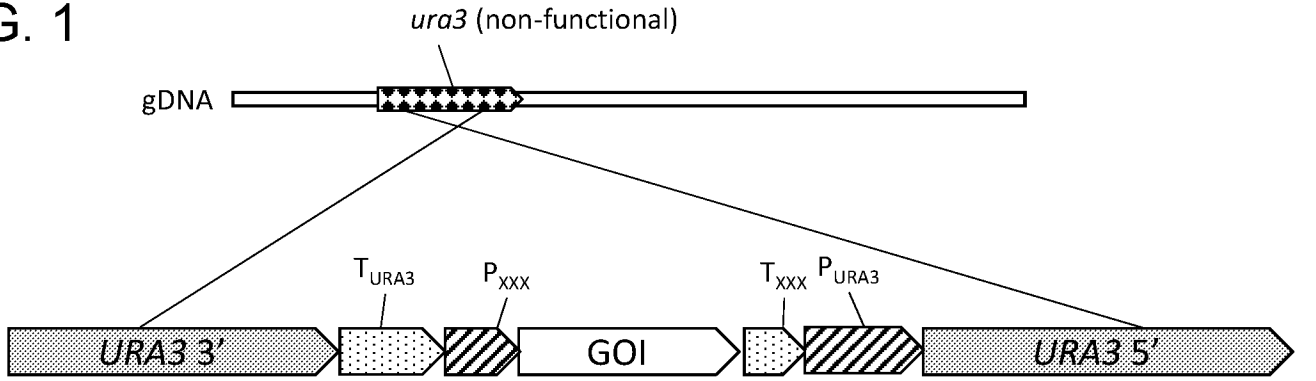
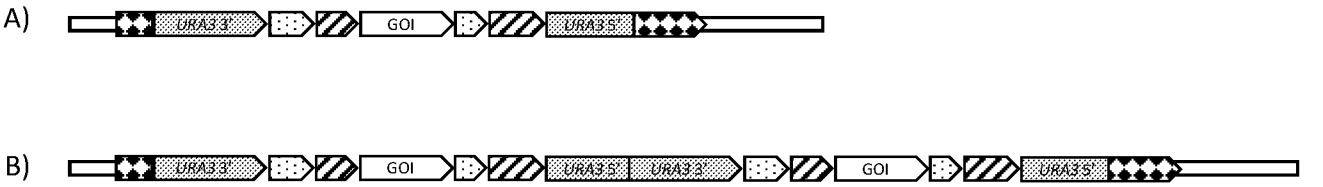


FIG. 57

FIG. 1


ura3 (non-functional)
gDNA
T URA3
P XXX
T XXX
P URA3
URA3 3'
GOI
URA3 5'


A)
URA3 3'
GOI
URA3 5'
B)
URA3 3'
GOI
URA3 5'
URA3 3'
GOI
URA3 5'