US Patent & Trademark Office
Patent Public Search | Text View

United States Patent	12385072
Kind Code	B2
Date of Patent	August 12, 2025
Inventor(s)	Buck; Jordan

Bidirectional multi-enzymatic scaffolds for biosynthesizing cannabinoids

Abstract

This document relates to using bidirectional, multi-enzymatic scaffolds to biosynthesize cannabinoids in recombinant hosts.

Inventors: **Buck; Jordan (Boulder, CO)**

Applicant: **Khona Scientific Holdings, Inc.** (Lone Tree, CO)

Family ID: **1000008748536**

Assignee: **Khona Scientific Holdings, Inc. (Lone Tree, CO)**

Appl. No.: **17/962229**

Filed: **October 07, 2022**

Prior Publication Data

Document Identifier	Publication Date
US 20230265465 A1	Aug. 24, 2023

Related U.S. Application Data

continuation parent-doc US 16694417 20191125 US 11525148 child-doc US 17962229
us-provisional-application US 62836265 20190419
us-provisional-application US 62771839 20181127

Publication Classification

Int. Cl.: **C12P7/42** (20060101); **C12N15/70** (20060101); **C12N15/74** (20060101); **C12N15/81** (20060101); **C12N15/82** (20060101)

U.S. Cl.:

CPC **C12P7/42** (20130101); **C12N15/70** (20130101); **C12N15/74** (20130101); **C12N15/81** (20130101); **C12N15/8222** (20130101); C12N2330/51 (20130101); C12N2800/40 (20130101); C12Y101/01034 (20130101); C12Y101/01157 (20130101); C12Y103/01038 (20130101); C12Y103/03 (20130101); C12Y121/03007 (20150701); C12Y121/03008 (20150701); C12Y203/01009 (20130101); C12Y203/01016 (20130101); C12Y203/01206 (20150701); C12Y203/03008 (20130101); C12Y203/0301 (20130101); C12Y205/01001 (20130101); C12Y207/01036 (20130101); C12Y207/04002 (20130101); C12Y401/01033 (20130101); C12Y402/01017 (20130101); C12Y404/01026 (20150701); C12Y503/03002 (20130101); C12Y604/01002 (20130101)

Field of Classification Search

USPC: None

References Cited

U.S. PATENT DOCUMENTS

Patent No.	Issued Date	Patentee Name	U.S. Cl.	CPC
8697654	12/2013	Cheng et al.	N/A	N/A
8884100	12/2013	Page et al.	N/A	N/A
9822384	12/2016	Poulos et al.	N/A	N/A
9856460	12/2017	Dueber et al.	N/A	N/A
11525148	12/2021	Buck	N/A	C12N 15/74
2003/0143562	12/2002	Anderson et al.	N/A	N/A
2005/0204419	12/2004	Helgeson et al.	N/A	N/A
2011/0008829	12/2010	Dueber et al.	N/A	N/A
2011/0145940	12/2010	Voytas et al.	N/A	N/A
2012/0144523	12/2011	Page et al.	N/A	N/A
2013/0130347	12/2012	Delisa et al.	N/A	N/A
2013/0164808	12/2012	McAuliffe et al.	N/A	N/A
2014/0370595	12/2013	Dueber et al.	N/A	N/A
2016/0010126	12/2015	Poulos et al.	N/A	N/A
2017/0139496	12/2016	Kang et al.	N/A	N/A
2017/0166950	12/2016	Wolkowicz	N/A	N/A
2019/0055314	12/2018	Luo et al.	N/A	N/A

FOREIGN PATENT DOCUMENTS

Patent No.	Application Date	Country	CPC
2644548	12/2006	CA	N/A
103360471	12/2012	CN	N/A
106414747	12/2016	CN	N/A
WO 2014113738	12/2013	WO	N/A
WO 2015196275	12/2014	WO	N/A
WO 2016010827	12/2015	WO	N/A
WO 2017139496	12/2016	WO	N/A
WO 2018200888	12/2017	WO	N/A

OTHER PUBLICATIONS

Chichili et al., “Linkers in the structural biology of protein-protein interactions,” Protein Sci., Feb. 2013, 22(2):153-167. cited by applicant
Andre et al. (2016) Cannabissativa: The Plant of the Thousand and One Molecules, Frontiers Plant Sci., vol. 7, article 19, pp. 1-15. cited by applicant
Becker et al., “High-Efficiency Transformation of Yeast by Electroporation,” Meth. Enzymology, 1991, 194:182-187. cited by applicant
Carvalho et al., “Designing microorganisms for heterologous biosynthesis of cannabinoids,” FEMS Yeast Research, Jun. 2017, 17(4):fox037, 11 pages. cited by applicant
Chen et al., “Fusion protein linkers: Property, design and functionality,” Adv. Drug Deliv. Reviews, Oct. 2013, 65(10):1357-1369. cited by applicant
Cho et al., “Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease,” Nat. Biotechnology, Mar. 2013, 31(3):230-232. cited by applicant
Christian et al., “Targeting DNA Double-Strand Breaks with TAL Effector Nucleases,” Genetics, Oct. 1, 2010, 186(2):757-761. cited by applicant
Cong et al., “Multiplex Genome Engineering Using CRISPR/Cas Systems,” Science, Feb. 15, 2013, 339(6121):819-823. cited by applicant
Dicarlo et al., “Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems,” Nucleic Acids Research, Apr. 2013, 41(7):4336-4343. cited by applicant
Dinkel et al., “The eukaryotic linear motif resource ELM: 10 years and counting,” Nucleic Acids Research, Jan. 2014, 42(Database issue):D259-D266. cited by applicant
Dueber et al., “Synthetic protein scaffolds provide modular control over metabolic flux,” Nat. Biotechnology, Aug. 2009, 27(8):753-759. cited by applicant
Durrens et al., “Expression of the avian gag-myc oncogene in *Saccharomyces cerevisiae*,” Curr. Genetics, Jul. 1990, 18(1):7-12. cited by applicant
EP Extended Search Report in European Appln. No. 19888674.9, dated Jul. 22, 2022, 6 pages. cited by applicant
Gagne, et al., “Identification of olivetolic acid cyclase from Cannabis sativa reveals a unique catalytic route to plant polyketides,” Proc. Natl. Acad. Sci. USA, Jul. 31, 2012, 109(31):12811-12816. cited by applicant
GenBank Accession No. AAA52679.1, “3-hydroxy-3-methylglutaryl coenzyme A reductase [*Homo sapiens*],” dated Nov. 8, 1994, 2 pages. cited by applicant
GenBank Accession No. AAA62411.1, “3-hydroxy-3-methylglutaryl coenzyme A synthase [*Homo sapiens*],” dated Feb. 24, 1995, 1 page. cited by applicant
GenBank Accession No. AAA74463.1, “ATP citrate-lyase [Rattus norvegicus],” dated Aug. 18, 1995, 1 page. cited by applicant
GenBank Accession No. AAC49920.1, “isopentenyl diphosphate:dimethylallyl diphosphate isomerase [*Arabidopsis thaliana*],” dated Feb. 24, 1998, 1 page. cited by applicant
GenBank Accession No. AAC50440.1, “mevalonate pyrophosphate decarboxylase [*Homo sapiens*],” dated Apr. 16, 1996, 1 page. cited by applicant
GenBank Accession No. AAC67348.1, “mevalonate diphosphate decarboxylase [*Arabidopsis thaliana*],” dated Mar. 11, 2002, 1 page. cited by applicant
GenBank Accession No. AAD31719.1, “mevalonate kinase [*Arabidopsis thaliana*],” dated May 10, 2000, 1 page. cited by applicant
GenBank Accession No. AAF82407.1, “mevalonate kinase [*Homo sapiens*],” dated Jun. 10, 2016, 1 page. cited by applicant
GenBank Accession No. AAH06089.1, “Phosphomevalonate kinase [*Homo sapiens*],” dated Oct. 19, 2016, 2 pages. cited by applicant
GenBank Accession No. AAH10004.1, “Farnesyl diphosphate synthase (farnesyl pyrophosphate synthetase, dimethylallyltranstransferase, geranyltranstransferase) [*Homo sapiens*],” dated Jul. 15,

2006, 2 pages. cited by applicant
GenBank Accession No. AAH30985.1, "Malonyl CoA:ACP acyltransferase (mitochondrial) [*Homo sapiens*]," dated Sep. 8, 2006, 2 pages. cited by applicant
GenBank Accession No. AAH31149.1, "Crk protein [Mus musculus]," dated Oct. 7, 2003, 2 pages. cited by applicant
GenBank Accession No. AAK56081.1, "ATP citrate lyase [Mus musculus]," dated Feb. 21, 2002, 1 page. cited by applicant
GenBank Accession No. AAK79760.1, "1-deoxy-D-xylulose 5-phosphate reductoisomerase [Clostridium acetobutylicum ATCC 824]," dated Jan. 30, 2014, 1 page. cited by applicant
GenBank Accession No. AAK80036.1, "Deoxyxylulose-5-phosphate synthase [Clostridium acetobutylicum ATCC 824]," dated Jan. 30, 2014, 1 page. cited by applicant
GenBank Accession No. AAK80816.1, "Acetyl-CoA acetyltransferase [Clostridium acetobutylicum ATCC 824]," dated Jan. 30, 2014, 1 page. cited by applicant
GenBank Accession No. AAK80844.1, "Isopentenyl monophosphate kinase, IPK [Clostridium acetobutylicum ATCC 824]," dated Jan. 30, 2014, 1 page. cited by applicant
GenBank Accession No. AAK81121.1, "4-diphosphocytidyl-2-methylerithritol synthase (Sugar Nucleotide Phosphorylase family) [Clostridium acetobutylicum ATCC 824]," dated Jan. 30, 2014, 1 page. cited by applicant
GenBank Accession No. AAM61343.1, "1-deoxy-D-xylulose 5-phosphate reductoisomerase DXR [*Arabidopsis thaliana*]," dated Jan. 27, 2006, 1 page. cited by applicant
GenBank Accession No. AAM67058.1, "acetoacyl-CoA-thiolase [*Arabidopsis thaliana*]," dated Jan. 27, 2006, 1 page. cited by applicant
GenBank Accession No. AAN17431.1, "putative 3-hydroxybutyryl-CoA dehydrogenase [*Arabidopsis thaliana*], " dated Sep. 24, 2002, 2 pages. cited by applicant
GenBank Accession No. AAN81487.1, "1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase [*Escherichia coli* CFT073]," dated Jan. 31, 2014, 1 page. cited by applicant
GenBank Accession No. AAP35407.1, "isopentenyl-diphosphate delta isomerase [*Homo sapiens*]," dated May 13, 2003, 1 page. cited by applicant
GenBank Accession No. AAP86010.1, "putative enoyl-(ACP) reductase (plasmid) [Cupriavidus necator H16]," dated Jul. 25, 2016, 1 page. cited by applicant
GenBank Accession No. AAP94122.1, "acetyl-CoA carboxylase 1 [*Homo sapiens*]," dated Jul. 27, 2003, 2 pages. cited by applicant
GenBank Accession No. AAS11086.1, "acetyl-CoA carboxylase, biotin carboxylase [Treponema denticola ATCC 35405]," dated Jan. 31, 2014, 2 pages. cited by applicant
GenBank Accession No. AAS11105.1, "3-hydroxyacyl-CoA dehydrogenase, putative [Treponema denticola ATCC 35405]," dated Jan. 31, 2014, 2 pages. cited by applicant
GenBank Accession No. AAS11855.1, "4-diphosphocytidyl-2C-methyl-D-erythritol kinase [Treponema denticola ATCC 35405]," dated Jan. 31, 2014, 2 pages. cited by applicant
GenBank Accession No. AAS12424.1, "1-deoxy-D-xylulose-5-phosphate synthase [Treponema denticola ATCC 35405]," dated Jan. 31, 2014, 2 pages. cited by applicant
GenBank Accession No. AAS12810.1, "2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase [Treponema denticola ATCC 35405]," dated Jan. 31, 2014, 2 pages. cited by applicant
GenBank Accession No. AAS12811.1, "2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase [Treponema denticola ATCC 35405]," dated Jan. 31, 2014, 2 pages. cited by applicant
GenBank Accession No. AAS12860.1, "1-deoxy-D-xylulose 5-phosphate reductoisomerase [Treponema denticola ATCC 35405]," dated Jan. 31, 2014, 2 pages. cited by applicant
GenBank Accession No. AC157384.3, "Bos taurus clone CH240-61D19, Working Draft Sequence, 6 unordered pieces," dated Jul. 11, 2008, 46 pages. cited by applicant
GenBank Accession No. ACJ56139.1, "Geranyltranstransferase(Farnesyl-diphosphate synthase) [Acinetobacter baumannii AB307-0294]," dated Jan. 31, 2014, 1 page. cited by applicant

GenBank Accession No. ACJ57023.1, "Enoyl-CoA hydratase [Acinetobacter baumannii AB307-0294]," dated Jan. 31, 2014, 1 page. cited by applicant
GenBank Accession No. ACJ58210.1, "4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase [Acinetobacter baumannii AB307-0294]," dated Jan. 31, 2014, 2 pages. cited by applicant
GenBank Accession No. ACJ59227.1, "2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase [Acinetobacter baumannii AB307-0294]," dated Jan. 31, 2014, 1 page. cited by applicant
GenBank Accession No. ACS85236.1, "malonyl CoA-acyl carrier protein transacylase [Dickeya paradisiaca Ech703]," dated Oct. 25, 2017, 1 page. cited by applicant
GenBank Accession No. ADI91469.1, "enoyl-CoA hydratase [Acinetobacter oleivorans DR1]," dated Jan. 30, 2014, 1 page. cited by applicant
GenBank Accession No. AEC07908.1, "4-(cytidine 5′-phospho)-2-C-methyl-D-erithritol kinase [*Arabidopsis thaliana*]," dated Jul. 20, 2017, 2 pages. cited by applicant
GenBank Accession No. AED97354.1, "4-hydroxy-3-methylbut-2-enyl diphosphate synthase [*Arabidopsis thaliana*]," dated Jul. 20, 2017, 3 pages. cited by applicant
GenBank Accession No. AEE35849.1, "hydroxy methylglutaryl CoA reductase 1 [*Arabidopsis thaliana*]," dated Jul. 20, 2017, 4 pages. cited by applicant
GenBank Accession No. AEE83052.1, "hydroxymethylglutaryl-CoA synthase / HMG-CoA synthase / 3-hydroxy-3-methylglutaryl coenzyme A synthase [*Arabidopsis thaliana*]," dated Jul. 20, 2017, 3 pages. cited by applicant
GenBank Accession No. AEE86362.1, "4-hydroxy-3-methylbut-2-enyl diphosphate reductase [*Arabidopsis thaliana*]," dated Jul. 20, 2017, 4 pages. cited by applicant
GenBank Accession No. AFD33345.1, "acyl-activating enzyme 1 [Cannabis sativa]," dated Aug. 1, 2012, 1 page. cited by applicant
GenBank Accession No. AFD33347.1, "acyl-activating enzyme 3 [Cannabis sativa]," dated Aug. 1, 2012, 1 page. cited by applicant
GenBank Accession No. AFN42527.1, "olivetolic acid cyclase [Cannabis sativa]," dated Aug. 2, 2012, 1 page. cited by applicant
GenBank Accession No. AGO55277.1, "polyketide biosynthesis malonyl CoA-acyl carrier protein transacylase BaeC [Serratia plymuthica 4Rx13]," dated Jan. 30, 2014, 2 pages. cited by applicant
GenBank Accession No. AHM22925.1, "2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase [Nicotiana tabacum]," dated Mar. 22, 2014, 1 page. cited by applicant
GenBank Accession No. AIE72439.1, "trans-2-enoyl-CoA reductase (plasmid) [Klebsiella michiganensis]," dated Jul. 8, 2016, 2 pages. cited by applicant
GenBank Accession No. AIZ91493.1, "3-hydroxyacyl-CoA dehydrogenase [*Escherichia coli* str. K-12 substr. MG1655]," dated Dec. 15, 2014, 2 pages. cited by applicant
GenBank Accession No. ALI39443.1, "acetyl-CoA acetyltransferase [*Escherichia coli* str. K-12 substr. MG1655]," dated Dec. 16, 2015, 2 pages. cited by applicant
GenBank Accession No. AMC97367.1, "pyruvate dehydrogenase [*Escherichia coli* str. K-12 substr. MG1655]," dated Feb. 2, 2016, 2 pages. cited by applicant
GenBank Accession No. ANM65835.1, "1-deoxy-D-xylulose 5-phosphate synthase 1 [*Arabidopsis thaliana*]," dated Jul. 20, 2017, 2 pages. cited by applicant
GenBank Accession No. AUG14916.1, "dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex [*Escherichia coli* str. K-12 substr. MG1655]," dated Dec. 17, 2017, 2 pages. cited by applicant
GenBank Accession No. BAA21534.1, "N-WASP [Rattus rattus]," dated Dec. 27, 2006, 1 page. cited by applicant
GenBank Accession No. BAB00624.1, "ATP citrate-lyase [Ciona intestinalis]," dated Jul. 15, 2000, 1 page. cited by applicant
GenBank Accession No. BAB21592.1, "2-C-methyl-D-erythritol 4-phosphate cytidyltransferase [*Arabidopsis thaliana*]," dated Feb. 14, 2004, 1 page. cited by applicant

GenBank Accession No. BAC41356.1, “tetrahydrocannabinolic acid synthase precursor [Cannabis sativa],” dated Sep. 15, 2004, 1 page. cited by applicant
GenBank Accession No. BAF65033.1, “cannabidiolic acid synthase [Cannabis sativa],” dated Jun. 29, 2007, 1 page. cited by applicant
GenBank Accession No. BAG14339.1, “olivetol synthase [Cannabis sativa],” dated Jun. 20, 2009, 1 page. cited by applicant
GenBank Accession No. CAJ91294.1, “Enoyl-CoA hydratase [Cupriavidus necator H16],” dated Mar. 7, 2015, 2 pages. cited by applicant
GenBank Accession No. CAJ92510.1, “pyruvate dehydrogenase complex, dehydrogenase (E1) component [Cupriavidus necator H16],” dated Mar. 7, 2015, 2 pages. cited by applicant
GenBank Accession No. CAJ92511.1, “dihydrolipoamide acetyltransferase (E2) component of pyruvate dehydrogenase complex [Cupriavidus necator H16],” dated Mar. 7, 2015, 2 pgs. cited by applicant
GenBank Accession No. CAJ92573.1, “Acetyl-CoA acetyltransferase [Cupriavidus necator H16],” dated Mar. 7, 2015, 2 pages. cited by applicant
GenBank Accession No. CAQ66339.1, “Phosphomevalonate kinase [Lactobacillus casei BL23],” dated Feb. 27, 2015, 2 pages. cited by applicant
GenBank Accession No. CAQ66619.1, “Pyruvate dehydrogenase complex, E2 component, dihydrolipoamide acetyltransferase [Lactobacillus casei BL23],” dated Feb. 27, 2015, 2 pages. cited by applicant
GenBank Accession No. CAQ66794.1, “Mevalonate kinase [Lactobacillus casei BL23],” dated Feb. 27, 2015, 2 pages. cited by applicant
GenBank Accession No. CAQ66795.1, “Diphosphomevalonate decarboxylase [Lactobacillus casei BL23],” dated Feb. 27, 2015, 2 pages. cited by applicant
GenBank Accession No. CAQ66796.1, “Isopentenyl-diphosphate delta-isomerase (IPP isomerase) (Isopentenyl pyrophosphate isomerase) [Lactobacillus casei BL23],” dated Feb. 27, 2015, 2 pages. cited by applicant
GenBank Accession No. CAQ66932.1, “Farnesyl-diphosphate synthase [Lactobacillus casei BL23],” dated Feb. 27, 2015, 2 pages. cited by applicant
GenBank Accession No. CAQ67081.1, “Hydroxymethylglutaryl-CoA synthase [Lactobacillus casei BL23],” dated Feb. 27, 2015, 2 pages. cited by applicant
GenBank Accession No. CAQ67082.1, “Hydroxymethylglutaryl-CoA reductase [Lactobacillus casei BL23],” dated Feb. 27, 2015, 2 pages. cited by applicant
GenBank Accession No. CAQ67083.1, “Acetyl-CoA acetyltransferase (Acetoacetyl-CoA thiolase) [Lactobacillus casei BL23],” dated Feb. 27, 2015, 2 pages. cited by applicant
GenBank Accession No. CAQ67359.1, “Biotin carboxylase [Lactobacillus casei BL23],” dated Feb. 27, 2015, 2 pages. cited by applicant
GenBank Accession No. CDH63564.1, “4-hydroxy-3-methylbut-2-enyl diphosphate reductase [*Escherichia coli* PMV-1],” dated Sep. 18, 2013, 2 pages. cited by applicant
GenBank Accession No. CDH63708.1, “1-deoxy-D-xylulose 5-phosphate reductoisomerase [*Escherichia coli* PMV-1],” dated Sep. 18, 2013, 1 page. cited by applicant
GenBank Accession No. CDH63925.1, “1-deoxy-D-xylulose-5-phosphate synthase [*Escherichia coli* PMV-1],” dated Sep. 18, 2013, 1 page. cited by applicant
GenBank Accession No. CDH64802.1, “4-diphosphocytidyl-2-C-methyl-D-erythritol kinase [*Escherichia coli* PMV-1],” dated Sep. 18, 2013, 1 page. cited by applicant
GenBank Accession No. CDH66379.1, “2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase [*Escherichia coli* PMV-1],” dated Sep. 18, 2013, 2 pages. cited by applicant
GenBank Accession No. CDH66380.1, “2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase [*Escherichia coli* PMV-1],” dated Sep. 18, 2013, 1 page. cited by applicant
GenBank Accession No. DAA07148.1, “TPA: Etr1p [*Saccharomyces cerevisiae* S288C],” dated

Mar. 27, 2017, 2 pages. cited by applicant
GenBank Accession No. DAA07337.1, "TPA: pyruvate dehydrogenase (acetyl-transferring) subunit E1 beta [*Saccharomyces cerevisiae* S288C]," dated Mar. 27, 2017, 2 pages. cited by applicant
GenBank Accession No. DAA10474.1, "TPA: dihydrolipoyllysine-residue acetyltransferase [*Saccharomyces cerevisiae* S288C]," dated Mar. 27, 2017, 2 pages. cited by applicant
GenBank Accession No. DAA10992.1, "TPA: [acyl-carrier-protein] S-malonyltransferase [*Saccharomyces cerevisiae* S288C]," dated Mar. 27, 2017, 2 pages. cited by applicant
GenBank Accession No. EAZ63544.2, "Phosphomevalonate kinase [Scheffersomyces stipitis CBS 6054]," dated Jul. 11, 2011, 2 pages. cited by applicant
GenBank Accession No. EDL06069.1, "syntrophin, acidic 1, isoform CRA_b [Mus musculus]," dated Jul. 26, 2016, 2 pages. cited by applicant
GenBank Accession No. J04537.1, "*Arabidopsis thaliana* HMG-CoA reductase (HMG1) mRNA, complete cds," dated Sep. 29, 2015, 2 pages. cited by applicant
GenBank Accession No. JN717233.1, "Cannabis sativa acyl-activating enzyme 1 mRNA, complete cds," dated Aug. 1, 2012, 2 pages. cited by applicant
Havranek et al., "Automated design of specificity in molecular recognition," Nat. Struct. Biology, Jan. 2003, 10(1):45-52. cited by applicant
Horn et al., "Synthetic Protein Scaffolds Based on Peptide Motifs and Cognate Adaptor Domains for Improving Metabolic Productivity," Front. Bioeng. Biotechnology, Nov. 2015, 3:191, 7 pages. cited by applicant
Hwang et al., "Efficient genome editing in zebrafish using a CRISPR-Cas system," Nat. Biotechnology, Mar. 2013, 31(3):227-229. cited by applicant
Ito et al., "Transformation of Intact Yeast Cells Treated with Alkali Cations," J. Bacteriology, Jan. 1983, 153(1):163-168. cited by applicant
Jiang et al., "Manipulation of GES and ERG20 for geraniol overproduction in *Saccharomyces cerevisiae*," Metab. Engineering, May 2017, 41:57-66. cited by applicant
Jiang et al., "RNA-guided editing of bacterial genomes using CRISPR-Cas systems," Nat. Biotechnology, Jan. 29, 2013, 31(3):233-239. cited by applicant
Jinek et al., "A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity," Science, Aug. 17, 2012, 337(6096):816-821. cited by applicant
Kim et al., "Optimization of hexanoic acid production in recombinant *Escherichia coli* by precise flux rebalancing," Bioresour. Technology, Jan. 2018, 247:1253-1257. cited by applicant
Klein et al. (2014) Design and characterization of structured protein linkers with differing flexibilities, Prot. Eng. Design Select., vol. 27, No. 10, pp. 325-330. cited by applicant
Makarova et al., "Evolution and classification of the CRISPR-Cas systems," Nat. Rev. Microbiology, Jun. 2011, 9(6):467-477. cited by applicant
Mali et al, "RNA-Guided Human Genome Engineering via Cas9," Science, Feb. 15, 2013, 339(6121):823-826. cited by applicant
PCT International Search Report and Written Opinion in International Appln. No. PCT/US2019/063029, dated Feb. 6, 2020, 9 pages. cited by applicant
Proschel et al., "Engineering of metabolic pathways by artificial enzyme channels," Front. Bioeng. Biotechnology, Oct. 2015, 3:168, 13 pages. cited by applicant
Quintero et al. (2007) An improved system for estradiol-dependent regulation of gene expression in yeast, Microb. Cell Factor., vol. 6, No. 10, pp. 1-9. cited by applicant
Redden et al, (2015) The synthetic biology toolbox for tuning gene expression in yeast, FEMS Yest Res., vol. 15, pp. 1-10. cited by applicant
Reinke et al., "A Synthetic Coiled-Coil Interactome Provides Heterospecific Modules for Molecular Engineering," J. Am. Chem. Society, Apr. 13, 2010, 132(17):6025-6031. cited by applicant

Rodriguez et al., “ATP citrate lyase mediated cytosolic acetyl-CoA biosynthesis increases mevalonate production in *Saccharomyces cerevisiae*,” Microb. Cell Factories, Mar. 2016, 15:48, 12 pages. cited by applicant
Sirikantaramas et al., “The Gene Controlling Marijuana Psychoactivity,” J. Biol. Chemistry, Jun. 9, 2004, 279(38):39767-39774. cited by applicant
Song et al., “Engineering *Saccharomyces cerevisiae* for geranylgeraniol overproduction by combinatorial design,” Sci. Reports, Nov. 8, 2017, 7:14991, 11 pages. cited by applicant
Stout et al., “The hexanoyl-CoA precursor for cannabinoid biosynthesis is formed by an acyl-activating enzyme in Cannabis sativa trichomes,” Plant Journal, Aug. 2012, 71(3):353-365. cited by applicant
Taura et al., “Cannabidiolic-acid synthase, the chemotype-determining enzyme in the fiber-type Cannabis sativa,” FEBS Letters, Jun. 2007, 581(16):2929-2934. cited by applicant
Taura et al., “Characterization of olivetol synthase, a polyketide synthase putatively involved in cannabinoid biosynthetic pathway,” FEBS Letters, Jun. 2009, 583(12):2061-2066. cited by applicant
Teyra et al., “Elucidation of the binding preferences of peptide recognition modules: SH3 and PDZ domains,” FEBS Letters, Jun. 2012, 586(17):2631-2637. cited by applicant
Whitaker et al., “Metabolic Pathway Flux Enhancement by Synthetic Protein Scaffolding,” Meth. Enzymology, 2011, 497(Part A):447-468. cited by applicant
Zakeri et al., “Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin,” Proc. Natl. Acad. Sci. USA, Mar. 20, 2012, 109(12):E690-E697. cited by applicant
GenBank Accession No. AAS11585.1, “hydroxymethylbutenyl pyrophosphate reductase [Treponema denticola ATCC 35405],” dated Jan. 31, 2014, 2 pages. cited by applicant
GenBank Accession No. AAS11783.1, “1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase [Treponema denticola ATCC 35405],” dated Jan. 31, 2014, 2 pages. cited by applicant
GenBank Accession No. CAQ66617.1, “Pyruvate dehydrogenase complex, E1 component, alpha subunit [Lactobacillus casei BL23],” dated Feb. 27, 2015, 2 pages. cited by applicant

---

Primary Examiner: Fronda; Christian L

Attorney, Agent or Firm: Fish & Richardson P.C.

Background/Summary

CROSS-REFERENCE TO RELATED APPLICATIONS (1) This application is a continuation of U.S. application Ser. No. 16/694,417, filed Nov. 25, 2019, which claims priority to U.S. Application Serial Nos. 62/836,265, filed on Apr. 19, 2019 and 62/771,839, filed on Nov. 27, 2018. The disclosures of the prior applications are considered part of (and are incorporated by reference in) the disclosure of this application.

TECHNICAL FIELD
(1) This document relates to methods and materials for biosynthesizing cannabinoids, and more particularly to using bidirectional multi-enzymatic scaffolds to biosynthesize cannabinoids.
SEQUENCE LISTING
(2) This application contains a Sequence Listing that has been submitted electronically as an XML file named “47300-0003002_SL_ST26.XML.” The XML file, created on May 12, 2023, is 487,591 bytes in size. The material in the XML file is hereby incorporated by reference in its entirety.
BACKGROUND

(3) The emerging therapeutic potential of cannabinoids warrants industrial-scale production to meet compounding future demands. Traditional cannabinoid production efforts rely on large-scale farming of *Cannabis sativa* L. However, agricultural cannabinoid production is problematic due to issues such as uncontrollable environmental factors and scaling limitations.

SUMMARY

(4) This document is based, at least in part, on the discovery that a bidirectional, multi-enzymatic scaffold can be engineered to allow high-throughput cannabinoid production in recombinant host cells. By controlling the localization, spatial orientation, and stoichiometry of enzymes catalyzing the biosynthesis of cannabinoids and cannabinoid precursors, the multi-enzymatic scaffolds described herein allow flux-optimized cannabinoid biosynthesis in genetically-engineered host cells.

(5) In one aspect, this document features a host cell capable of producing one or more cannabinoids selected from the group consisting of cannabigerolic acid, cannabidiolic acid, and cannabichromenic acid. The host cell includes at least three different exogenous nucleic acids, wherein the first and the second exogenous nucleic acids each encode a plurality of engineered enzymes selected from the group consisting of acetyl-CoA acetyltransferase, a 3-hydroxybutyryl-CoA dehydrogenase, an enoyl-CoA hydratase, a beto-ketothiolase, a trans-enoyl-CoA reductase, an HMG-CoA synthetase, an HMG-CoA reductase, a mevalonate kinase, a phosphomevalonate kinase, a diphosphomevalonate decarboxylase, an isopentenyl-diphosphate delta isomerase, a geranyl-diphosphate synthase, an olivetol synthase, an olivetolic acid cyclase, and a CBGA synthase; wherein each of the engineered enzymes includes a heterologous interaction domain, wherein the heterologous interaction domain comprises a first and a second peptide motif, and wherein each heterologous interaction domain is different from each other; and wherein the third exogenous nucleic acid encodes a polypeptide scaffold comprising a plurality of peptide ligands, wherein each peptide ligand comprises an amino acid sequence that can bind to the first or the second peptide motif of one of the heterologous interaction domains. The plurality of engineered enzymes further can include an ATP citrate lyase and an acetyl-CoA carboxylase. The host cell further can include an exogenous nucleic acid encoding a cannabidiolic acid synthase (CBDAS) and a cannabichromenic acid synthase (CBCAS). The host cell can include an exogenous CBDAS. The host cell can include an exogneous CBCAS. The host cell can include an exogenous CBDAS and an exogenous CBCAS. The host cell can include an exogenous hexanoyl-CoA synthetase. The host cell can include at least four different exogenous nucleic acids, wherein the first, second, and fourth nucleic acids each encode a plurality of the engineered enzymes. The host cell can include at least five different exogenous nucleic acids, wherein the first, second, fourth, and fifth nucleic acid each encode a plurality of the engineered enzymes. The host cell can include at least six different exogenous nucleic acids, wherein the first, second, fourth, fifth, and sixth nucleic acids each encode a plurality of the engineered enzymes. Each exogenous nucleic acid can include a constitutive promoter operably linked to the sequence encoding the engineered enzyme or polypeptide scaffold or an inducible promoter operably linked to the sequence encoding the engineered enzyme or polypeptide scaffold. In some embodiments, the promoter is a GAL1-10 promoter. In some embodiments, a constitutive promoter used to express the polypeptide scaffold has weaker constitutive activity level than a constitutive promoter used to express the engineered enzymes. In some embodiments, a constitutive promoter is used to express the engineered enzymes and an inducible promoter is used to express the polypeptide scaffold. In some embodiments, an inducible promoter is used to express the engineered enzymes and a constitutive promoter is used to express the polypeptide scaffold.

(6) Any of the host cells can be bacterial, yeast, algae, or plant cells. A bacterial cell can be selected from the group consisting of *Escherichia coli*, *Bacillus*, *Brevibacterium*, *Streptomyces*, and *Pseudomonas* cells. A yeast cell can be selected from the group consisting of *Pichia pastoris*, *Saccharomyces cerevisiae*, *Yarrowia lipolytica*, *Kluyveromyces marxianus*, and Komagataella

phaffii cells. An algae cell can be *Dunaliella* sp., *Chlorella variabilis, Euglena mutabilis*, or *Chlamydomonas reinhardtii* cells. A plant cell can be a *Cannabis* or tobacco cell.
(7) In some embodiments, each of the engineered enzymes is of the formula: enzyme—linker.sub.1−spacer−linker.sub.2− motif.sub.1− linker.sub.3−motif.sub.2, where linkers 1, 2, and 3 can be the same or different, motif 1 and motif 2 can be the same or different, and where motif 1 and motif 2 form the heterologous interaction domain. A scaffold polypeptide can be of the formula: N-terminus—[Ligand 1−linker−Ligand 2−Spacer]n−(optionally-tagged)C-terminus, where n is the number of heterologous interaction domains, and where ligand 1 and ligand 2 bind motif 1 and motif 2, respectively, of the heterologous interaction domain. The scaffold polypeptide can be tagged with a MYC tag, FLAG tag, or HA tag. The host cell further can include a nucleic acid encoding a second polypeptide scaffold comprising a plurality of peptide ligands, wherein each peptide ligand comprises an amino acid sequence that can bind to a different motif of the heterologous interaction domain. The linker can have a flexible GS-rich sequence flanking a rigid α-helical moiety. The spacer can be the cTPR6 spacer.
(8) This document also features a method of producing one or more cannabinoids selected from the group consisting of cannabigerolic acid, cannabidiolic acid, and cannabichromenic acid. The method can include culturing any of the host cells described herein under conditions wherein the host cell produces the one or more cannabinoids. The host cells can be cultured in a culture medium supplemented with citrate, glucose, hexanoic acid, and/or other carbon source, and/or in a culture medium supplemented with malonyl-CoA. The method further can include extracting the one or more cannabinoids from the host cells.
(9) Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.
(10) Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

Description

DESCRIPTION OF DRAWINGS
(1) FIG. **1**A is a schematic of one representative embodiment of a multi-enzymatic cannabinoidergic scaffold within a cell. The multi-enzymatic scaffold includes enzymes of the hexanoyl-CoA pathway, enzymes of the upper cannabinoid pathway, and enzymes of the mevalonate pathway. The schematic also depicts a second scaffold according to one embodiment containing enzymes of the malonyl-CoA pathway and depicts a non-scaffolded cannabidiolic acid synthase (CBDAS) and a non-scaffolded cannabichromenic acid synthase (CBCAS). ID refers to enzyme-linked interaction domain; cTPR6 refers to a spacer sequence; scaffolded ligands refer to the tandem peptide ligands that form the scaffold-binding sites specific for each enzyme-linked ID. The target products cannabigerolic acid (CBGA), cannabigerol (CBG), cannabidiolic acid (CBDA), cannabidiol (CBD), cannabichromenic acid (CBCA), and cannabichromene (CBC), are boxed for emphasis. CBG can be produced by decarboxylation of CBGA, CBD can be produced by decarboxylation of CBDA, and CBC can be produced by decarboxylation of CBCA. For each decarboxylation, the ‘Δ’ symbols represent heat and the ‘hv’ symbols represent light.
(2) FIG. **1**B is a schematic of one representative embodiment of a bidirectional, multi-enzymatic

scaffold within a cell (e.g., a yeast cell). The multi-enzymatic scaffold (referred to as SCF gene cassette in the nucleus) includes enzymes of the hexanoyl-CoA pathway (referred to as HCA cassette in nucleus), enzymes of the upper cannabinoid pathway (referred to as CAN cassette in nucleus), and enzymes of the mevalonate pathway (referred to as GPP cassette in nucleus). The schematic also depicts a second scaffold according to one embodiment containing enzymes of the malonyl-CoA pathway and depicts a non-scaffolded CBDAS and a non-scaffolded CBCAS. ID refers to enzyme-linked interaction domain; cTPR6 refers to a spacer sequence; scaffolded ligands refer to the tandem peptide ligands that form the scaffold-binding sites specific for each enzyme-linked ID. The target products CBGA, CBG, CBDA, CBD, CBCA, and CBC are boxed for emphasis. CBG can be produced by decarboxylation of CBGA, CBD can be produced by decarboxylation of CBDA, and CBC can be produced by decarboxylation of CBCA. For each decarboxylation, the 'Δ' symbols represent heat and the 'hv' symbols represent light.

(3) FIG. **2**A is a schematic of gene cassettes according to one embodiment for the engineering of cannabinoidergic cells.

(4) FIG. **2**B is a schematic of gene cassettes used in Examples 2-4 for biosynthesizing cannabinoids in yeast.

(5) FIG. **3** is an example of an enzyme-scaffold complex.

(6) FIG. **4** is a schematic of one representative embodiment of a multi-enzymatic cannabinoidergic scaffold within a cell. The multi-enzymatic scaffold includes enzymes of the hexanoyl-CoA pathway, enzymes of the upper cannabinoid pathway, and enzymes of the mevalonate pathway. The schematic also depicts a second scaffold according to one embodiment containing enzymes of the malonyl-CoA pathway and depicts a non-scaffolded CBDAS and a non-scaffolded CBCAS. Pyruvate dehydrogenase (E1) and dihydrolipoyl transacetylase (E2) are substituted for ATP citrate lyase in both of the depicted scaffolds. ID refers to enzyme-linked interaction domain; cTPR6 refers to a spacer sequence; scaffolded ligands refer to the tandem peptide ligands that form the scaffold-binding sites specific for each enzyme-linked ID. The target products CBGA, CBG, CBDA, CBD, CBCA, and CBC are boxed for emphasis. CBG can be produced by decarboxylation of CBGA, CBD can be produced by decarboxylation of CBDA, and CBC can be produced by decarboxylation of CBCA. For each decarboxylation, the 'Δ' symbols represent heat and the 'hv' symbols represent light.

(7) FIG. **5** is a schematic of one representative embodiment of a multi-enzymatic cannabinoidergic scaffold within a cell. The multi-enzymatic scaffold includes enzymes of the hexanoyl-CoA pathway, enzymes of the upper cannabinoid pathway, and enzymes of the MEP (2-C-methylerythritol 4-phosphate) pathway. The schematic also depicts a second scaffold according to one embodiment containing enzymes of the malonyl-CoA pathway and depicts a non-scaffolded CBDAS and a non-scaffolded CBCAS. ID refers to enzyme-linked interaction domain; cTPR6 refers to a spacer sequence; scaffolded ligands refer to the tandem peptide ligands that form the scaffold-binding sites specific for each enzyme-linked ID. The target products CBGA, CBG, CBDA, CBD, CBCA, and CBC are boxed for emphasis. CBG can be produced by decarboxylation of CBGA, CBD can be produced by decarboxylation of CBDA, and CBC can be produced by decarboxylation of CBCA. For each decarboxylation, the 'Δ' symbols represent heat and the 'hv' symbols represent light.

(8) FIG. **6**A contains the amino acid sequence of each of the following enzymes: an ATP citrate lyase (SEQ ID NO:83), acetyl-CoA acetyltransferase (atoB) (SEQ ID NO:84), a 3-hydroxybutyryl-CoA dehydrogenase (SEQ ID NO:85), an enoyl-CoA hydratase (SEQ ID NO:86), a trans-enoyl-CoA reductase (SEQ ID NO:88), a beta-ketothiolase (bktB) (SEQ ID NO:87), an HMG-CoA synthase (SEQ ID NO:90), a truncated HMG-CoA reductase (SEQ ID NO:91), a mevalonate kinase (SEQ ID NO:92), a phosphomevalonate kinase (SEQ ID NO:93), a diphosphomevalonate decarboxylase (SEQ ID NO:94), an isopentenyl-diphosphate delta isomerase (SEQ ID NO:95), a mutant geranyl-diphosphate synthase (ERG20.sup.WW) (SEQ ID NO:96), an olivetol synthase

(SEQ ID NO:98), an olivetolic acid cyclase (SEQ ID NO:99), a CBGA synthase (SEQ ID NO:100), an acetyl-CoA carboxylase (SEQ ID NO:97), a CBDA synthase (SEQ ID NO:101), a CBCA synthase (SEQ ID NO:102), and a hexanoyl-CoA synthetase (SEQ ID NO:89).

(9) FIG. **6**B contains the amino acid sequence of engineered enzymes of the formula Enzyme−Enzyme Linker−cTPR6 Spacer−ID Linker−ID Motif #1−ID Motif Linker− ID Motif #2, where the linkers (enzyme linker, ID linker, and ID motif linker) can be the same or different, and ID motif #1 and ID motif #2 can be the same or different. The amino acid sequence of the following engineered enzymes are provided: ATP citrate lyase (ID1) (SEQ ID NO:103), an acetyl-CoA acetyltransferase (atoB) (ID2) (SEQ ID NO:104), a 3-hydroxybutyryl-CoA dehydrogenase (ID3) (SEQ ID NO:105), an enoyl-CoA hydratase (ID4) (SEQ ID NO:106), a trans-enoyl-CoA reductase (ID5) (SEQ ID NO:107), a beto-ketothiolase (bktB) (ID6) (SEQ ID NO:108), an HMG-CoA synthase (ID7) (SEQ ID NO:109), a truncated HMG-CoA reductase (ID8) (SEQ ID NO:110), a mevalonate kinase (ID9) (SEQ ID NO:111), a phosphomevalonate kinase (ID10) (SEQ ID NO:112), a diphosphomevalonate decarboxylase (ID11) (SEQ ID NO:113), an isopentenyl-diphosphate delta isomerase (ID12) (SEQ ID NO:114), a mutant geranyl-diphosphate synthase (ERG20.sup.WW) (ID13) (SEQ ID NO:115), an olivetol synthase (ID14) (SEQ ID NO:116), an olivetolic acid cyclase (ID15) (SEQ ID NO:117), a CBGA synthase (ID16) (SEQ ID NO:118), and an acetyl-CoA carboxylase (ID17) (SEQ ID NO:211).

(10) FIG. **6**C contains the amino acid sequence of a polypeptide scaffold of the formula: N-terminus− [Ligand #1− ID Motif #1 Ligand−Linker− ID Motif #2 Ligand—Scaffolded ID-binding Site Spacer]n− (Myc)3-tagged C-terminus, where n is 16 and the ID motif ligands correspond to the motifs for IDs 1-16 as shown in Table 2. See SEQ ID NO:119.

(11) FIG. **6**D contains the amino acid sequence of a polypeptide scaffold of the formula: N-terminus−[Ligand #1−ID Motif #1 Ligand−Linker−ID Motif #2 Ligand—Scaffolded ID-binding Site Spacer]n−(FLAG)3-tagged C-terminus, where n is 2 and the ID motif ligands correspond to the motifs for IDs 1 and 17 as shown in Table 2. See SEQ ID NO:120.

(12) FIG. **7** is a schematic of one representative embodiment of a scaffold with the minimal requirements for cannabigerolic acid synthesis. The scaffold contains enzymes of the upper cannabinoid pathway. In this embodiment, a non-scaffolded hexanoyl-CoA synthetase (HCS), a non-scaffolded CBDAS, and a non-scaffolded CBCAS also are used. ID refers to enzyme-linked interaction domain; cTPR6 refers to a spacer sequence; scaffolded ligands refer to the tandem peptide ligands that form the scaffold-binding sites specific for each enzyme-linked ID. The target products CBGA, CBG, CBDA, CBD, CBCA, and CBC are boxed for emphasis. CBG can be produced by decarboxylation of CBGA, CBD can be produced by decarboxylation of CBDA, and CBC can be produced by decarboxylation of CBCA. For each decarboxylation, the ‘Δ’ symbols represent heat and the ‘hv’ symbols represent light.

(13) FIG. **8** is a schematic of one representative embodiment of a bi-directional scaffold containing a HCS on the N-terminus of the scaffold, a geranyl pyrophosphate synthase (GPPS) on the C-terminus of the scaffold, and the enzymes of the upper cannabinoid pathway between the HCS and GPPS. In this embodiment, a non-scaffolded CBDAS and a non-scaffolded CBCAS also can be used. ID refers to enzyme-linked interaction domain; cTPR6 refers to a spacer sequence; scaffolded ligands refer to the tandem peptide ligands that form the scaffold-binding sites specific for each enzyme-linked ID. The target products CBGA, CBG, CBDA, CBD, CBCA, and CBC are boxed for emphasis. CBG can be produced by decarboxylation of CBGA, CBD can be produced by decarboxylation of CBDA, and CBC can be produced by decarboxylation of CBCA. For each decarboxylation, the ‘Δ’ symbols represent heat and the ‘hv’ symbols represent light.

(14) FIG. **9** is a schematic of one representative embodiment of a unidirectional scaffold containing enzymes of the upper cannabinoid pathway, shown with soluble enzymes from the precursor pathways (hexanoyl-CoA pathway, mevalonate pathway, and malonyl-CoA pathway), and soluble CBDAS and CBCAS. ID refers to enzyme-linked interaction domain; cTPR6 refers to a spacer

sequence; scaffolded ligands refer to the tandem peptide ligands that form the scaffold-binding sites specific for each enzyme-linked ID. The target products CBGA, CBG, CBDA, CBD, CBCA, and CBC are boxed for emphasis. CBG can be produced by decarboxylation of CBGA, CBD can be produced by decarboxylation of CBDA, and CBC can be produced by decarboxylation of CBCA. For each decarboxylation, the 'Δ' symbols represent heat and the 'hv' symbols represent light.

(15) FIG. **10** is a schematic of one representative embodiment of a multi-enzymatic cannabinoidergic scaffold within a cell. The multi-enzymatic scaffold includes enzymes of the malonyl-CoA (MCA) pathway, enzymes of the upper cannabinoid pathway, and enzymes of the mevalonate pathway. The schematic also depicts a separate scaffold according to one embodiment containing enzymes of the hexanoyl-CoA pathway and depicts a non-scaffolded CBDAS and a non-scaffolded CBCAS. ID refers to enzyme-linked interaction domain; cTPR6 refers to a spacer sequence; scaffolded ligands refer to the tandem peptide ligands that form the scaffold-binding sites specific for each enzyme-linked ID. The target products CBGA, CBG, CBDA, CBD, CBCA, and CBC are boxed for emphasis. CBG can be produced by decarboxylation of CBGA, CBD can be produced by decarboxylation of CBDA, and CBC can be produced by decarboxylation of CBCA. For each decarboxylation, the 'Δ' symbols represent heat and the 'hv' symbols represent light.

(16) FIG. **11** is a schematic of one representative embodiment of a multi-enzymatic cannabinoidergic scaffold within dual compartments of a cell, the cytosol and mitochondria/plastid.

(17) FIG. **12**A contains the nucleotide sequences encoding each of the following: an ATP citrate lyase (SEQ ID NO:121), an acetyl-CoA acetyltransferase (atoB) (SEQ ID NO:122), a 3-hydroxybutyryl-CoA dehydrogenase (SEQ ID NO:123), an enoyl-CoA hydratase (SEQ ID NO:124), a trans-enoyl-CoA reductase (SEQ ID NO:125), a beto-ketothiolase (bktB) (SEQ ID NO:126), an HMG-CoA synthase (SEQ ID NO:127), a truncated HMG-CoA reductase (SEQ ID NO:128), a mevalonate kinase (SEQ ID NO:129), a phosphomevalonate kinase (SEQ ID NO:130), a diphosphomevalonate decarboxylase (SEQ ID NO:131), an isopentenyl-diphosphate delta isomerase (SEQ ID NO:132), a geranyl-diphosphate synthase (ERG20.sup.WW) (SEQ ID NO:133), an olivetol synthase (SEQ ID NO:134), an olivetolic acid cyclase (SEQ ID NO:135), a CBGA synthase (SEQ ID NO:136), an acetyl-CoA carboxylase (SEQ ID NO:137), a CBDA synthase (SEQ ID NO:138), a CBCA synthase (SEQ ID NO:139), and a hexanoyl-CoA synthetase (SEQ ID NO:140).

(18) FIG. **12**B contains the nucleotide sequences encoding engineered enzymes of the formula: Enzyme−Enzyme Linker−cTPR6 Spacer− ID Linker− ID Motif #1− ID Motif Linker− ID Motif #2, where the Enzyme Linker, ID Linker, and ID Motif Linker can be the same or different, and where ID Motif #1 and ID Motif #2 can be the same or different. The nucleotide sequences encoding the following engineered enzymes are provided: ATP citrate lyase (ID1) (SEQ ID NO:141), an acetyl-CoA acetyltransferase (atoB) (ID2) (SEQ ID NO:142), a 3-hydroxybutyryl-CoA dehydrogenase (ID3) (SEQ ID NO:143), an enoyl-CoA hydratase (ID4) (SEQ ID NO:144), a trans-enoyl-CoA reductase (ID5) (SEQ ID NO:145), a bktB (ID6) (SEQ ID NO:146), an HMG-CoA synthase (ID7) (SEQ ID NO:147), a truncated HMG-CoA reductase (ID8) (SEQ ID NO:148), a mevalonate kinase (ID9) (SEQ ID NO:149), a phosphomevalonate kinase (ID10) (SEQ ID NO:150), a diphosphomevalonate decarboxylase (ID11) (SEQ ID NO:151), an isopentenyl-diphosphate delta isomerase (ID12) (SEQ ID NO:152), a mutant geranyl-diphosphate synthase (ERG20.sup.WW) (ID13) (SEQ ID NO:153), an olivetol synthase (ID14) (SEQ ID NO:154), an olivetolic acid cyclase (ID15) (SEQ ID NO:155), a CBGA synthase (ID16) (SEQ ID NO:156), and an acetyl-CoA carboxylase (ID17) (SEQ ID NO:157).

(19) FIG. **12**C contains the nucleotide sequence (SEQ ID NO:158) encoding a scaffold polypeptide that contains the peptide ligands corresponding to IDs 1-16 as shown in Table 2 and a triplicate myc tag on the C-terminus.

(20) FIG. **12**D contains the nucleic acid sequence (SEQ ID NO:159) encoding a scaffold polypeptide that contains the peptide ligands corresponding to IDs 1 and 17, and a triplicate FLAG

tag on the C-terminus.
(21) FIG. **13**A contains the amino acid sequence of scaffold-binding engineered enzymes and a soluble hexanoyl-CoA synthetase (HCS) (SEQ ID NO:209) encoded by the HCA gene cassette. The scaffold-binding engineered enzymes are ATP Citrate Lyase (ACL) (ACL— Enzyme Linker— cTPR6 Spacer— ID Linker— ID1) (SEQ ID NO:160); Acetyl-CoAAcetyltransferase (atoB) (atoB — Enzyme Linker—cTPR6 Spacer— ID Linker— ID2) (SEQ ID NO:161); 3-Hydroxybutyryl-CoA Dehydrogenase (BHBD) (BHBD— Enzyme Linker—cTPR6 Spacer— ID Linker— ID3) (SEQ ID NO:162); Enoyl-CoA Hydratase (ECH) (ECH— Enzyme Linker—cTPR6 Spacer— ID Linker— ID4) (SEQ ID NO:163); Trans-Enoyl-CoA Reductase (ECR) (ECR— Enzyme Linker— cTPR6 Spacer— ID Linker— ID5) (SEQ ID NO:164); and Beta-Ketothiolase (bktB) (bktB— Enzyme Linker—cTPR6 Spacer— ID Linker— ID6) (SEQ ID NO:165).
(22) FIG. **13**B contains the amino acid sequences of scaffold-binding engineered enzymes encoded by the GPP gene cassette. The scaffold-binding engineered enzymes are HMG-CoA Synthase (HMGS) (HMGS— Enzyme Linker—cTPR6 Spacer— ID Linker— ID7) (SEQ ID NO:166); truncated HMG-CoA Reductase (tHMGR) (tHMGR—Enzyme Linker—cTPR6 Spacer— ID Linker— ID8) (SEQ ID NO:167); Mevalonate Kinase (ERG12) (ERG12— Enzyme Linker— cTPR6 Spacer— ID Linker— ID9) (SEQ ID NO:168); Phosphomevalonate Kinase (ERG8) (ERG8 — Enzyme Linker—cTPR6 Spacer—ID Linker— ID10) (SEQ ID NO:169); Diphosphomevalonate Decarboxylase (MVD1) (MVD1— Enzyme Linker—cTPR6 Spacer— ID Linker— ID11) (SEQ ID NO:170); Isopentenyl-Diphosphate Delta-Isomerase (IDI1) (IDI1— Enzyme Linker—cTPR6 Spacer— ID Linker— ID12) (SEQ ID NO:171); and Geranyl-Diphosphate Synthase (ERG20.sup.WW) (ERG20.sup.WW— Enzyme Linker—cTPR6 Spacer— ID Linker— ID13) (SEQ ID NO:172).
(23) FIG. **13**C contains the amino acid sequences of scaffold-binding engineered enzymes, a soluble CBDA synthase (SEQ ID NO:173), and a soluble CBCA synthase (SEQ ID NO:174) encoded by the CAN gene cassette. The scaffold-binding engineered enzymes are Olivetol Synthase (OS) (OS— Enzyme Linker—cTPR6 Spacer— ID Linker—ID14) SEQ ID NO:175); Olivetolic Acid Cyclase (OAC) (OAC— Enzyme Linker—cTPR6 Spacer— ID Linker— ID15) (SEQ ID NO:176); CBGA Synthase (CBGAS—Enzyme Linker—cTPR6 Spacer— ID Linker— ID16) (SEQ ID NO:177); and Acetyl-CoA Carboxylase (ACC) (ACC— Enzyme Linker—cTPR6 Spacer— ID Linker— ID17) (SEQ ID NO:178).
(24) FIG. **13**D contains the amino acid sequences of the Cannabinoidergic Metabolon Scaffold (CBSCFLD)— (Myc)3 (SEQ ID NO:179) and the Malonyl-CoA Metabolon Scaffold (MCASCFLD)— (FLAG)3 (SEQ ID NO:180).
(25) FIG. **14**A contains codon-optimized nucleotide sequences (SEQ ID NOs:181-187) encoding the enzymes of FIG. **13**A.
(26) FIG. **14**B contains the codon-optimized nucleotide sequences (SEQ ID NOs:188-194) encoding the enzymes of FIG. **13**B.
(27) FIG. **14**C contains the codon-optimized nucleotide sequences (SEQ ID NOs:195-200) encoding the enzymes of FIG. **13**C.
(28) FIG. **14**D contains the codon-optimized nucleotide sequences (SEQ ID NO:201 and SEQ ID NO:202) encoding the scaffolds of FIG. **13**D.
(29) FIG. **15**A contains the nucleotide sequence of the HCA gene cassette (SEQ ID NO:203).
(30) FIG. **15**B contains the nucleotide sequence of the GPP gene cassette (SEQ ID NO:204).
(31) FIG. **15**C contains the nucleotide sequence of the CAN gene cassette (SEQ ID NO:205).
(32) FIG. **15**D contains the nucleotide sequence of the SCF gene cassette (SEQ ID NO:206).
(33) FIG. **15**E contains the nucleotide sequence of the SOL gene cassette (SEQ ID NO:207).
(34) FIG. **16** is a map of the pCCI-Brick plasmid construct.
(35) FIG. **17** is a map of a pESC-TRP (“vHCA”) vector construct. In this map, the vector contains a TRP gene allowing selection in tryptophan deficient media. Similar vectors also were made in

which the TRP gene was replaced with a LEU gene allowing selection in leucine deficient media, a HIS3 gene allowing selection in histidine deficient media, or a URA3 gene allowing selection in uracil deficient media.

(36) FIG. **18** is a graph of the proliferation curves for yCBSCF and yCBSOL cultures. Line plots depicting cell proliferation curves were fitted via nonlinear regression of cell density measurements (OD.sub.600nm) recorded in 12-hour intervals over a 48-hour incubation period for yCBSCF and yCBSOL cultures. Initial cell densities for all cultures were standardized to OD.sub.600mn=0.3. For all measures, n=3 biological replicates for yCBSCF and yCBSOL cultures. Floating data points depict means with 95% confidence intervals. Dotted lines represent 95% confidence intervals for regression curve fits.

(37) FIGS. **19**A-**19**E show a comparison of cannabinoid and precursor titers for scaffolded and soluble cannabinoid biosynthesis. Representative mass spectra of target analytes isolated from (FIG. **19**A) yCBSOL and (FIG. **19**B) yCBSCF cultures incubated for 48 hours in basal culture media. Bar plots depicting (FIG. **19**C) Total (aggregate) cannabinoid (CBGA+CBDA+CBCA+CBG+CBD+CBC) titers, (FIG. **19**D) cannabinoid precursor (OVA) titers and summated parent and decarboxylation derivative (CBGA+CBG, CBDA+CBD, and CBCA+CBC) cannabinoid titers, and (FIG. **19**E) separated parent (COO(H)) cannabinoid (CBGA, CBDA, and CBCA) and decarboxylation derivative (ACOOH) cannabinoid (CBG, CBD, and CBC) titers for 48-hour yCBSOL (left) and yCBSCF (right) cultures grown in basal culture media. For all measures, n=3 biological replicates for yCBSCF and yCBSOL cultures. CB, cannabinoid; Cannabigerolic acid, CBGA; cannabigerol, CBG; cannabidiolic acid, CBDA; cannabidiol, CBD; cannabichromenic acid, CBCA; cannabichromene, CBC; olivetolic acid, OVA. Floating asterisks indicate statistically significant (determined by Bonferroni's multiple comparisons post-hoc test; a=0.05) between-strain differences for yCBSCF versus yCBSOL cultures. Bar plots depict means with 95% confidence intervals. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

(38) FIG. **20** is a bar plot of the impact of citrate and hexanoate supplementation on scaffolded and soluble cannabinoid biosynthesis. Total cannabinoid (CBGA+CBDA+CBCA+CBG+CBD+CBC) titers are shown for yCBSOL and yCBSCF cultures incubated for 48 hours in basal, hexanoate (300 mg/L)-supplemented, and buffered (pH 6.0) citrate (300 mg/L)-supplemented culture media. Floating asterisks indicate statistically significant (determined by Bonferroni's multiple comparisons post-hoc test; a=0.05) between-strain differences for yCBSCF versus yCBSOL cultures. Lines with asterisks indicate statistically significant (determined by Bonferroni's multiple comparisons post-hoc test; a=0.05) within-strain differences for basal media total cannabinoid titers versus citrate-supplemented media total cannabinoid titers for yCBSCF cultures. Bar plots depict means with 95% confidence intervals. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

(39) FIGS. **21**A and **21**B show concentration-response parameterization of scaffolded and soluble cannabinoid biosynthesis from citrate. In FIG. **21**A, line plots are shown depicting eight-point concentration ([citrate])—response (total cannabinoid titers) curves fitted via asymmetric sigmoidal (five-parameter) logistic regression and in FIG. **21**B, bar graphs are shown depicting concentration-response parameter estimates (CB.sub.Max, the estimated maximum total cannabinoid titers and citrate EC.sub.50, the estimated citrate concentration yielding half-maximal total cannabinoid titers) for 48-hour yCBSCF and yCB.sub.SOL cultures incubated for 48 hours in culture media supplemented with 0, 10, 30, 100, 300, 1000, 3000, or 10000 mg/L buffered (pH 6.0) citrate. For all measures, n=3 biological replicates for yCBSCF and yCB.sub.SOL cultures. Floating asterisks indicate statistically significant (determined by Bonferroni's multiple comparisons post-hoc test; α=0.05) between-strain differences for yCBSCF versus yCB.sub.SOL cultures. Floating data points and bar plots depict means with 95% confidence intervals. Dotted lines represent 95% confidence intervals for regression curve fits *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

DETAILED DESCRIPTION

(40) This document provides methods and materials for producing cannabinoids in host cells or in

vitro using a bidirectional, multi-enzymatic scaffold, which can control the localization and stoichiometry of enzymes catalyzing the biosynthesis of cannabinoids and cannabinoid precursors. As described herein, one or more cannabinoids including cannabigerolic acid (CBGA), cannabidiolic acid (CBDA), cannabichromenic acid (CBCA), and tetrahydrocannabinolic acid, can be produced using a bidirectional, multi-enzymatic scaffold and one or more soluble cannabinoid synthesis enzymes, and the conjugate bases, cannabigerolate, cannabidiolate, cannabichromenate, and tetrahydrocannabinolate, respectively, and decarboxylation products, cannabigerol (CBG), cannabidiol (CBD), cannabichromene (CBC), and tetrahydrocannabinol, respectively, of these cannabinoids also can be produced, as can the tetrahydrocannabinolic acid oxidation product cannabinolic acid and its decararboxylation product cannabinol. The bidirectional, multi-enzymatic scaffold described herein results in significant increases in cannabinoid production in recombinant hosts, including total cannabinoid, CBGA, CBG, CBDA, CBD, CBCA, CBC, and olivetolic acid precursor production, as compared with cannabinoid production in recombinant hosts using the same enzymes that are not bound to a scaffold. As used herein, enzymes that are not bound to a scaffold are referred to as soluble or non-scaffolded. While one particular form of a cannabinoid or other compound may be referenced herein, it is understood that any of its neutral or ionized forms, including any salt forms thereof or decarboxylation derivatives thereof (e.g., produced in the presence of heat and light), are included unless otherwise indicated. It is understood by those skilled in the art that the specific form will depend on factors such as pH and carboxylation status.

(41) In general, enzymes described herein, which can be co-localized on one or more scaffolds and used for producing cannabinoids or cannabinoid precursors, are engineered to contain an interaction domain (ID), which can be separated from the enzyme by an amino acid spacer sequence at the N- or C-terminus of the enzyme. The ID can be composed of two or more scaffold-binding motifs. The engineered enzymes also can include one or more linkers between the enzyme, spacer, and/or ID. The engineered enzymes can bind to a scaffold, which is a polypeptide that contains unique ID-binding domains, i.e., tandem peptide ligands, as shown in FIG. **1**A and FIG. **1**B, such that the enzymes are co-localized to the scaffold. In other words, each enzyme can be engineered to contain a protein-protein interaction domain that is specific for ligand or ligands (binding site) on the scaffold such that the enzyme can be localized to a discrete location along the scaffold via non-covalent interactions. In some cases, the engineered enzymes can be chimeric enzymes. The scaffolded ligands can be separated using amino acid linkers or spacers. See, for example, Horn and Sticht, *Frontiers in Bioengineering and Biotechnology*, 2015, volume 3, article 191; Whitaker and Dueber, *Methods in Enzymology*, Chapter 19, "Metabolic Pathway Flux Enhancement by Synthetic Protein Scaffolding," Volume 497, 2011, for descriptions of IDs, binding domains, linkers and spacers. IDs also can be referred to as adaptor domains.

(42) Typically, each interaction domain consists of two tandem scaffold-binding motifs that continue/extend from the C-terminus of the engineered enzyme and that can bind to their corresponding scaffolded peptide ligands, which are constructed in tandem along the scaffold. Dual-binding of enzymes to the scaffold ensures fixed spatial orientation, increases binding specificity for each ID-scaffold interaction, and better tethers each enzyme to the scaffold, all of which can improve pathway flux by enabling substrate channeling through each enzymatic step in the scaffolded biosynthetic pathways.

(43) In some embodiments, there are more than two, e.g., three, four, five, six, seven, eight, nine, or ten, or more molecules of each enzyme localized to the scaffold. In addition, the ratio of any given enzyme in a biosynthetic pathway to any other enzyme in the biosynthetic pathway can be varied. For example, the ratio of one engineered enzyme in a pathway to a second engineered enzyme in the same pathway can be varied, e.g., from about 1:5 to about 5:1, e.g., from about 1:5 to about 2:5, from about 2:5 to about 3:5, from about 3:5 to about 5:5, from about 5:5 to about 5:3, from about 5:3 to about 5:2, or from about 5:2 to about 5:1.

(44) The peptide ligands are typically short peptide sequences, ranging in length from 3 to 50

amino acid residues. For example, a peptide ligand can be 3-10, 7-15, 10-20, 15-25, 20-30, 25-35, 30-40, 35-45, or 40-50 amino acids in length. There is a database of over 200 different motifs available on the web at elm. eu.org that can be used as described herein. See, for example, Dinkel et al., *Nucleic Acids Res.* 2014; 42(Database issue): D259—D266.

(45) An ID can be a peptide sequence ranging in length 3 to 200 amino acid residues. For example, the ID can be 3-10, 7-15, 10-20, 15-25, 20-30, 25-35, 30-40, 35-45, 40-50, 45-55, 50-60, 65-75, 70-80, 85-95, 90-100, 100-110, 105-115, 110-120, 115-125, 120-130, 125-135, 130-140, 135-145, 140-150, 135-145, 140-150, 145-155, 150-160, 165-175, 170-180, 175-185, 180-190, 185-195, or 190-200 amino acids in length. For example, an ID can be a SH2 domain, a SH3 domain, a PDZ domain, a GTPase binding domain (GBD), a leucine zipper domain, a PTB domain, an FHA domain, a WW domain, a 14-3-3 domain, a death domain, a caspase recruitment domain, a bromodomain, a chromatin organization modifier, a shadow chromo domain, an F-box domain, a HECT domain, a RING finger domain, a sterile alpha motif domain, a glycine-tyrosine-phenylalanine domain, a SNAP domain, a VHS domain, an ANK repeat, an armadillo repeat, a WD40 repeat, an MH2 domain, a calponin homology domain, a Dbl homology domain, a gelsolin homology domain, a PB1 domain, a SOCS box, an RGS domain, a Toll/IL-1 receptor domain, a tetratricopeptide repeat, a TRAF domain, a Bcl-2 homology domain, a coiled-coil domain, a bZIP domain, a fibronectin receptor domain, a FNDC domain, a SAMD domain, a WBP domain, and/or a SASH domain. See, e.g., U.S. Pat. No. 9,856,460 for a list of domains that can be uses as an ID as described herein.

(46) For example, an ID can be a “Src homology2” (SH2) or a “Src homology3” (SH3) domain. SH2 domains are highly conserved structures of approximately 100 amino acid residues that comprise two α-helices and seven β-strands. The SH2 domain can have a promiscuous or strict specificity for a 3-5 amino acid motif flanking a phosphorylated tyrosine. See, Horn and Sticht, 2015, *supra*. For example, a SH2 domain that can be used as an ID as described herein can be residues 5-122 of a mouse Ct10 regulator of kinase adaptor (Crk) protein having GenBank Accession No. AAH31149.

(47) SH3 domains are small modules of approximately 60 residues that bind proline-rich ligands, which bind to the domain surface at three shallow grooves formed by conserved aromatic residues and exhibit two different binding orientations. See, Horn and Sticht, 2015, *supra*. In some embodiments, the proline-rich ligand can have a core PXXP motif flanked by a positively charged residue. Class I PZP domains recognize ligands conforming to the consensus +XXPXXP (where + is either Arg or Lys), while Class II domains recognize PXXPX+ motifs and bind to ligands in the opposite orientation. See, Teyra, et al., *FEBS Lett.*, 2012 586(17):2631-7. Individual SH3 domains do not measurably interact with other SH3 domain family ligands within an organism, minimizing cross-talk and increasing the number of domain/ligand pairs available for simultaneous use. See, Whitaker and Dueber, 2011, *supra*. For example, a SH3 domain that can be used as an ID as described herein can be residues 134-190 of a mouse Crk protein having GenBank Accession No. AAH31149 and its peptide ligand can be

(48) TABLE-US-00001 (SEQ ID NO: 1) PPPALPPKRRR.

(49) For example, an ID can be a PDZ (PSD-95/Discs-large/ZO1) domain. PDZ domains are approximately 100 amino acid residues in length and target specific motifs at the C-terminus of the binding partner. The peptide ligand adopts a β-strand and extends an existing β-sheet within the PDZ domain upon binding. At least four different classes of ligands are known for PDZ domains exhibiting a distinct binding specificity. See, Horn and Sticht, 2015, *supra*. For example, grouped PDZ domains into two main specificity classes based on distinct ligand signatures: Class I PDZ domains recognize a (X[T/S]XϕCOOH) motif, Class II PDZ domains recognize a (XϕXϕINCOOH) motif, and Class III PDZ domains recognize a X[ED]XϕCOOH motif, where X is any residue and ϕ is a hydrophobic amino acid. See, Teyra, et al., 2012, *supra*. PDZ and SH3 domains are found throughout eukaryotic and eubacterial genomes. For example, a PDZ domain

that can be used as an ID as described herein can be residues 77-171 of a mouse α-syntrophin protein having GenBank Accession No. EDL06069 and the peptide ligand can be
(50) TABLE-US-00002 (SEQ ID NO: 208) GVKESLV.
(51) For example, an ID can be a GBD domain from a protein such as the Wiskott-Aldrich syndrome-like protein (N-WASP). Isolated GBD domains do not adopt a single, discrete structure under physiological conditions but rather exhibit multiple, loosely packed conformations in solution. The corresponding peptide ligand has been deduced from the autoinhibited form of the GBD. See, Horn and Sticht, 2015, *supra*. For example, a GBD domain that can be used as an ID described herein can include residues 196 to 274 of a rat N-WASP protein having GenBank Accession No. BAA21534, and its peptide ligand, which can be LVGALMHVMQKRSRAIHSSDEGEDQAGDEDED (SEQ ID NO:2), can be used as a peptide ligand as described herein.
(52) For example, an ID can have a leucine zipper or synthetic coiled-coil domain. A leucine zipper domain can include multiple interspersed leucine residues approximately seven amino acid residues apart. Havranek, and Harbury ((2003), *Nat. Struct. Biol.* 10, 45-52) identified new pairs of homodimers or heterodimers by altering residues between leucine zipper pairs based on computational prediction. Reinke, et al. ((2010). *J. Am. Chem. Soc.* 132, 6025-6031) identified three pairs of synthetic coiled coils that do not exhibit measurable self-association. See, Whitaker and Dueber, 2011, *supra*. One example of an ID that can be used as described herein can be ITIRAAFLEKENTALRTEIAELEKEVGRCENIVSKYETRYGPL (SEQ ID NO:3), and its peptide ligand for use as described herein can be
(53) TABLE-US-00003 (SEQ ID NO: 4) LEIRAAFLEKENTALRTRAAELRKRVGRCRNIVSKYETRYGPL.
(54) For example, an ID can be a dockerin polypeptide, which can localize to a specific cohesion polypeptide on a scaffold described herein. Cohesion-dockerin pairs are particularly useful for ex vivo applications as binding is calcium dependent. See, Whitaker and Dueber, 2011, *supra.*
(55) Combinations of IDs that have high affinity for their peptide ligands and high specificity, i.e., minimal cross-reactivity, can be used as described herein to allow for binding of multiple, different enzymes to a scaffold provided herein. For example, at least three different enzymes can be localized on a scaffold. In some embodiments, at least four different enzymes can be localized on a scaffold. In some embodiments, at least five different enzymes can be localized on a scaffold. In some embodiments, at least six different enzymes can be localized on a scaffold. In some embodiments, at least seven different enzymes can be localized on a scaffold. In some embodiments, at least eight different enzymes can be localized on a scaffold. In some embodiments, at least nine different enzymes can be localized on a scaffold. In some embodiments, at least ten different enzymes can be localized on a scaffold. In some embodiments, at least eleven different enzymes can be localized on a scaffold. In some embodiments, at least twelve different enzymes can be localized on a scaffold. In some embodiments, at least fifteen different enzymes can be localized on a scaffold. In some embodiments, at least seventeen different enzymes can be localized on a scaffold. In some embodiments, at least eighteen different enzymes can be localized on a scaffold. In some embodiments, at least twenty different enzymes can be localized on a scaffold. In some embodiments, at least twenty-one different enzymes can be localized on a scaffold.
(56) Table 1 provide exemplary combinations of heterologous IDs, i.e., IDs that are different from each other, that can be used in seventeen different engineered enzymes and Table 2 provides the corresponding exemplary combinations of peptide ligands that can be used to localize the seventeen different enzymes to one or more scaffolds. In the embodiments shown in Tables 1 and 2, each ID is composed of two tandem peptide motifs as are the corresponding peptide ligands, which interact with the tandem peptide motifs. It will be appreciated that any one of the enzymes listed in Tables 1 and 2 can be used in combination with any of the listed combinations of IDs and

corresponding peptide ligands.

(57) TABLE-US-00004 TABLE 1 Interaction Domain Motif Sequences in Engineered Enzymes ID ID Motif ID Motif #1 ID Motif ID Motif #2 Enzyme # #1 Amino Acid Sequence #2 Amino Acid Sequence ATP Citrate Lyase 1 SYNZIP1 SYYHHHHHHLESTSLYKKAGSG SYNZIP2 SYYHHHHHHLESTSLYKKAGSGS SNLVAQLENEVASLENENETLK ARNAYLRKKIARLKKDNLQLERD KKNLHKKDLIAYLEKEIANLRK EQNLEKIIANLRDEIARLENEVASH KIEE ((SEQ ID NO: 5)) EQ (SEQ ID NO: 6) Acetyl-CoA 2 SYNZIP3 SYYHHHHHHLESTSLYKKAGSG SYNZIP4 SYYHHHHHHLESTSLYKKAGSGS Acetyltransferase SNEVTTLENDAAFIENENAYLE QKVAELKNRVAVKLNRNEQLKNK (atoB) KEIARLRKEKAALRNRLAHKK VEELKNRNAYLKNELATLENEVA (SEQ ID NO: 7) RLENDVAE (SEQ ID NO: 8) 3-hydroxybutyryl- 3 MYND ENLYFQGENLYFQGDSSESCWN UEV MAVSESQLKKMVSKYKYRDLTVR CoA Dehydrogenase CGRKASETCSGCNTARYCGSFC ETVNVITLYKDLKPVLDSYVFNDG QHKDWEKHHHICGQTLQAQQ SSRELMNLTGTIPVPYRGNTYNIPI (SEQ ID NO: 9) CLWLLDTYPYNPPICFVKPTSSMTI KTGKHVDANGKIYLPYLHEWKHP QSDLLGLIQVMIVVFGDEPPVFSRP (SEQ ID NO: 10) Enoyl-CoA 4 PABP GPLGSPLTASMLASAPPQEQKQ MDM2 NTNMSVPTDGAVTTSQIPASEQET Hydratase MLGERLFPLIQAMHPTLAGKITG LVRPKPLLLKLLKSVGAQKDTYT MLLEIDNSELLHMLESPESLRSK MKEVLFYLGQYIMTKRLYDEKQQ VDEAVAVLQAHQAKEAAQKA HIVYCSNDLLGDLFGVPSFSVKEH (SEQ ID NO: 11) RKIYTMIYRNLVV (SEQ ID NO: 12) Trans-Enoyl-CoA 5 SYNZIP10 SYYHHHHHHLESTSLYKKAGSG SYNZIP22 SYYHHHHHHLESTSLYKKAGSGS Reductase SNLLATLRSTAAVLENENHVLE KRIAYLRKKIAALKKDNANLEKDI KEKEKLRKEKEQLLNKLEAYK ANLENEIERLIKEIKTLENEVASHE (SEQ ID NO: 13) Q (SEQ ID NO: 14) Beta-ketothiolase 6 GYF DVMWEYKWENTGDAELYGPFT PAH ESDSVEFNNAISYVNKIKTRFLDHP (bktB) SAQMQTWVSEGYFPDGVYCRK EIYRSFLEILHTYQKEQLHTKGRPF LDPPGGQFYNSKRIDFDLYT RGMSEEEVFTEVANLFRGQEDLLS (SEQ ID NO: 15) EFGQFLPEAKR (SEQ ID NO: 16) HMG-COA Synthase 7 WW1A LGPLPPGWEVRSTVSGRIYFVD WW1B GAMGPLPPGWEKRTDSNGRVYFV HNNRTTQFTDPRLH (SEQ ID NHNTRITQWEDPRS (SEQ ID NO: 17) NO: 18) HMG-COA 8 FOS SYYHHHHHHLESTSLYKKAGSE SYNZIP9 SYYHHHHHHLESTSLYKKAGSGS Reductase FFRRERNKMAAAKCRNRRRELT QKVESLKQKIEELKQRKAQLKNDI DTLQAETDQLEDEKSALQTEIA ANLEKEIAYAET NLLKEKEKLEFILAAHRPACKIP (SEQ ID NO: 20) DDLGFPEEMSLE (SEQ ID NO: 19) Mevalonate Kinase 9 VHS1 MEPAMEPETLEARINRATNPLN VHS2 GAMGSMAEAEGESLESWLNKATN KELDWASINGFCEQLNEDFEGP PSNRQEDWEYIIGFCDQINKELEGP PLATRLLAHKIQSPQEWEAIQAL QIAVRLLAHKIQSPQEWEALQALT TVLETCMKSCGKRFHDEVGKFR VLEACMKNCGRRFHNEVGKFRFL FLNELIKVVSPKYLGSRTSEKVK NELIKVVSPKYLGDRVSEKVKTKV NKILELLYSWTVGLPEEVKIAEA IELLYSWTMALPEEAKIKDAYHML YQMLKKQGIVKS (SEQ ID KRQGIVQSDPPIPVDRTLIPSPPPRP NO: 21) KN (SEQ ID NO: 22) Phosphomevalonate 10 SYNZIP13 SYYHHHHHHLESTSLYKKAGSG SYNZIP15 SYYHHHHHHLESTSLYKKAGSGSF Kinase SQKVEELKNKIAELENRNAVKK ENVTHEFILATLENENAKLRRLEA NRVAHLKQEIAYLKDELAAHEF KLERELARLRNEVAWL (SEQ ID E (SEQ ID NO: 23) NO: 24) Diphospho- 11 MATH AMADLEQKVLEMEASTYDGVFI SKP1 ASIKLQSSDGEIFEVDVEIAKQSVTI mevalonate WKISDFPRKRQEAVAGRIPAIFS KTMLEDLGMDDEGDDDPVPLPNV Decarboxylase PAFYTSRYGYKMCLRIYLNGDG NAAILKKVIQWCTHHKDDPPPPED

TGRGTHLSLFFVVMKGPNDALL DENKEKRTDDIPVWDQEFLKVDQ
RWPFNQKVTLMLLDQNNREHV GTLFELILAANYLDIKGLLDVTCKT
IDAFRPDVTSSSFQRPVNDMNIA VANMIKGKTPEEIRKTFNIKNDFTE
SGCPLFCPVSKMEAKNSYVRDD EEEAQVRKENQWC (SEQ ID AIFIKAIVDLTGL
(SEQ ID NO: 26) NO: 25) Isopentenyl- 12 SYNZIP5 SYYHHHHHHLESTSLYKKAGSG
SYNZIP6 SYYHHHHHHLESTSLYKKAGSGS Diphosphate Delta-
SNTVKELKNYIQELEERNAELK QKVAQLKNRVAYKLKENAKLENI Isomerase
NLKEHLKFAKAELEFELAAHKF VARLENDNANLEKDIANLEKDIAN E (SEQ ID
NO: 27) LERDVAR (SEQ ID NO: 28) Geranyl-Diphosphate 13 PDZ1
LCTMKKGPSGYGFNLHSDKSKP PDZ2 SSGALIYTVELKRYGGPLGITISGTE Synthase
GQFIRSVDPDSPAEASGLRAQDR EPFDPIIISSLTKGGLAERTGAIHIG
IVEVNGVCMEGKQHGDVVSAIR DRILAINSSSLKGKPLSEAIHLLQM
AGGDETKLLVVDRE (SEQ ID AGETVTLKIKKQTDAQPASS (SEQ NO: 29) ID
NO: 30) Olivetol Synthase 14 SH2A GNNLETYEWYNKSISRDKAEKL SH2B
GSHPWFFGKIPRAKAEEMLSKQRH LLDTGKEGAFMVRDSRTPGTYT
DGAFLIRESESAPGDFSLSVKFGND VSVFTKAIISENPCIKHYHIKET
VQHFKVLRDGAGKYFLWVVKFNS NDSPKRYYVAEKYVFDSIPLLIQ
LNELVDYHRSTSVSRNQQIFLRDIE YHQYNGGGLVTRLRYPVCG QVPQQPT (SEQ
ID NO: 32) (SEQ ID NO: 31) Olivetolic Acid 15 PTB1
GQDRSEATLIKRFKGEGVRYKA PTB2 GSHMGSQFWVTSQKTEASERCGL Cyclase
KLIGIDEVSAARGDKLCQDSMM QGSYILRVEAEKLTLLTLGAQSQIL
KLKGVVAGARSKGEHKQKIFLT EPLLFWPYTLLRRYGRDKVMFSFE
ISFGGIKIFDEKTGALQHHHAVH AGRRCPSGPGTFTFQTSQGNDIFQ
EISYIAKDITDHRAFGYVCGKEG AVEAAIQQQKAQGKVGQAQDILR
NHRFVAIKTAQAAEPVILDLRDL LEHHHHHH (SEQ ID NO: 210)
FQLIYELKQREELEKKA (SEQ ID NO: 33) CBGA Synthase 16 SH3A
AEYVRALFDFNGNDEEDLPFKK SH3B LIKHMRAEALFDFTGNSKLELNFK
GDILRIRDKPEEQWWNAEDSEG AGDVIFLLSRINKDWLEGTVRGAT
KRGMIPVPYVEKY (SEQ ID GIFPLSFVKILK NO: 34) (SEQ ID NO: 35) Acetyl-
CoA 17 FAT GSHMRLGAQSIQPTANLDRTDD PEX GAMATPGSENVLPREPLIATAVKF
Carboxylase LVYLNVMELVRAVLELKNELA LQNSRVRQSPLATRRAFLKKKGLT
QLPPEGYVVVVKNVGLTLRKLI DEEIDMAFQQSGTAADEPSSLW
GSVDDLLPSLPSSSRTEIEGTQK (SEQ ID NO: 37) LLNKDLAELINKMRLAQQNAVTS
LSEECKRQMLTASHTLAVDAKN LLDAVDQAKVLANLAHPPAE (SEQ ID NO: 36)
(58) TABLE-US-00005 TABLE 2 Tandem Peptide Ligand Sequences in Scaffold ID
Motif #1 ID Motif #2 ID ID Motif Scaffolded Ligand ID Motif Scaffolded Ligand
Enzyme # #1 Amino Acid Sequence #2 Amino Acid Sequence ATP Citrate Lyase 1
SYNZIP1 SYYHHHHHHLESTSLYKKAGS SYNZIP2 SYYHHHHHHLESTSLYKKAGSGS
GSARNAYLRKKIARLKKDNLQ NLVAQLENEVASLENENETLKKK
LERDEQNLEKIIANLRDEIARLE NLHKKDLIAYLEKEIANLRKKIEE NEVASHEQ (SEQ
ID NO: 6) (SEQ ID NO: 5) Acetyl-CoA 2 SYNZIP3 SYYHHHHHHLESTSLYKKAGS
SYNZIP4 SYYHHHHHHLESTSLYKKAGSGS Acetyltransferase
GSQKVAELKNRVAVKLNRNEQ NEVTTLENDAAFIENENAYLEKEI (atoB)
LKNKVEELKNRNAYLKNELAT ARLRKEKAALRNRLAHKK (SEQ
LENEVARLENDVAE (SEQ ID ID NO: 7) NO: 8) 3-hydroxybutyryl- 3 MYND
RPPTISNPPPLISSAKHPSV UEV NFLQSRPEPTAPPEESFRSG (SEQ CoA Dehydrogenase
(SEQ ID NO: 38) ID NO: 39) Enoyl-CoA Hydratase 4 PABP
SKGTGLNPNAKVWQEIAPGN MDM2 PDGGTTFEHLWSSLEPDSTY (SEQ (SEQ ID
NO: 40) ID NO: 41) Trans-Enoyl-CoA 5 SYNZIP10 SYYHHHHHHLESTSLYKKAGS
SYNZIP22 SYYHHHHHHLESTSLYKKAGSGS Reductase GSKRIAYLRKKIAALKKDNAN

NLLATLRSTAAVLENENHVLEKEK LEKDIANLENEIERLIKEIKTLE
EKLRKEKEQLLNKLEAYK (SEQ ID NEVASHEQ (SEQ ID NO: 14) NO: 13) Beta-
Ketothiolase 6 GYF PATSQHPPPPPGHRSQAPSH PAH ELNSLLILLEAAEYLERRDR
(SEQ (bktB) (SEQ ID NO: 42) ID NO: 43) HMG-COA Synthase 7 WW1A
FQMPADTPPPAYLPPEDPMT WW1B ERESNEEPPPPYEDPYWGNG (SEQ (SEQ ID
NO: 44) ID NO: 45) HMG-COA Reductase 8 FOS SYYHHHHHHLESTSLYKKAGS
SYNZIP9 SYYHHHHHHLESTSLYKKAGSEFF GSQKVESLKQKIEELKQRKAQL
RRERNKMAAAKCRNRRRELTDTL KNDIANLEKEIAYAET (SEQ
QAETDQLEDEKSALQTEIANLLKE ID NO: 20) KEKLEFILAAHRPACKIPDDLGFPE
EMSLE (SEQ ID NO: 19) Mevalonate Kinase 9 VHS1 VSSTKLVSFHDDSDEDLLHI
VHS2 AAATPISTFHDDSDEDLLHV (SEQ (SEQ ID NO: 46) ID NO: 47)
Phosphomevalonate 10 SYNZIP13 SYYHHHHHHLESTSLYKKAGS SYNZIP15
SYYHHHHHHLESTSLYKKAGSGS Kinase GSFENVTHEFILATLENENAKL
QKVEELKNKIAELENRNAVKKNR RRLEAKLERELARLRNEVAWL
VAHLKQEIAYLKDELAAHEFE (SEQ ID NO: 24) (SEQ ID NO: 23)
Diphosphomevalonate 11 MATH HDDSLPHPQQATDDSGHESD SKP1
GSPNAGSVEQTPKKPGLRRR (SEQ Decarboxylase (SEQ ID NO: 48) ID NO: 49)
Isopentenyl- 12 SYNZIP5 SYYHHHHHHLESTSLYKKAGS SYNZIP6
SYYHHHHHHLESTSLYKKAGSGS Diphosphate Delta- GSQKVAQLKNRVAYKLKENA
NTVKELKNYIQELEERNAELKNLK Isomerase KLENIVARLENDNANLEKDIAN
EHLKFAKAELEFELAAHKFE (SEQ LEKDIANLERDVAR (SEQ ID ID NO: 27)
NO: 28) Geranyl-Diphosphate 13 PDZ1 TDEEREETEEEVYLLNSTTL PDZ2
DGNVSGTQRLDSATVRTYSC (SEQ Synthase (SEQ ID NO: 50) ID NO: 51)
Olivetol Synthase 14 SH2A ALVDDAADYEPPPSNNEEAL SH2B
RELFDDPSYVNVQNLDKARQ (SEQ (SEQ ID NO: 52) ID NO: 53) Olivetolic Acid
15 PTB1 KNTKSMNFDNPVYRKTTEEE PTB2 RSLPSTWIENKLYGMSDPNW (SEQ Cyclase
(SEQ ID NO: 54) ID NO: 55) CBGA Synthase 16 SH3A
VVDNSPPPALPPKKRQSAPS SH3B TORSKPQPAVPPRPSADLIL (SEQ (SEQ ID NO:
56) ID NO: 57) Acetyl-CoA 17 FAT SATRELDELMASLSDFKIQG PEX
DLALSENWAQEFLAAGDAVD Carboxylase (SEQ ID NO: 58) (SEQ ID NO: 59)
(59) The spacers or linkers connecting an enzyme and ID, as well as a binding domain on a scaffold, can be peptide sequences ranging in length from 6 to 250 amino acid residues. The term "spacer" typically refers to a longer and more structurally-rigid peptide sequence and the term "linker" typically refers to a shorter and more structurally-flexible peptide sequence. In embodiments in which both terms are used, linker typically refers to a sequence that is about 3 to about 50 amino acids in length and spacer typically refers to a sequence that is longer (e.g., about 36 to about 250 amino acids in length). For example, a linker can be 6-15, 10-20, 15-25, 20-30, 25-35, 30-40, 35-45, or 40-50 amino acids in length. A spacer can be, for example, 36-40, 40-50, 45-55, 50-60, 55-65, 60-70, 65-75, 70-80, 75-85, 90-100, 95-105, 100-110, 105-115, 110-120, 115-125, 120-130, 125-135, 130-140, 135-145, 140-150, 145-155, 150-160, 165-175, 170-180, 175-185, 180-190, 185-195, 190-200, 195-205, 200-210, 205-215, 210-220, 215-225, 220-230, 225-235, 230-240, 235-245, or 240-250 amino acids in length. See, for example, Chen, et al., *Adv Drug Deliv Rev.* 2013 65(10): 1357-1369. In either case, the linker/spacer can be a series of small and/or hydrophilic and/or other amino acid residues that can adapt flexible and/or rigid structures. For example, the linker can be a series of glycine residues, a series of alanine residues, a series of serine residues, or a series of alternating glycine and serine (or threonine) residues such as (G-S).sub.8 (SEQ ID NO:60), (G-S).sub.10 (SEQ ID NO:61), or (G-S).sub.15 (SEQ ID NO:62), or contain mainly glycine residues such as (GGGGS).sub.3 (SEQ ID NO:63) or (GGGGS).sub.4 (SEQ ID NO:64), or contain any other series of canonical or non-canonical amino acid residues or combinations thereof. In some embodiments, a linker can include glutamic acid, alanine, and lysine

residues such as (EAAAK).sub.2 (SEQ ID NO:65), (EAAAK).sub.3 (SEQ ID NO:66), or (EAAAK).sub.4 (SEQ ID NO:67). See, Horn and Sticht, 2015, *supra*. In some embodiments, a linker can be a combination of glycine, alanine, proline and methionine residues, such as AAAGGM (SEQ ID NO:68), AAAGGMPPAAAGGM (SEQ ID NO:69), AAAGGM (SEQ ID NO:70), or PPAAAGGMM (SEQ ID NO:71). See, e.g., U.S. Pat. No. 9,856,460.

(60) Based on amino acid composition, linkers or spacers can be either structured or intrinsically unstructured. For example, in some embodiments, a spacer can have a sequence that adopts a more structurally-rigid α-helical conformation and a linker can have a GS-rich peptide sequence that is more structurally-flexible. For example, in some embodiments, a linker can include flexible GS-rich sequences flanking one or more rigid α-helical moieties, e.g., GS-rich sequences flanking duplicate, triplicate, or quadruplicate α-helical moieties. For example, in some embodiments, a linker or spacer can have the sequence GSAGSAAGSGEF (SEQ ID NO:72), KLSGGGGSGGGGSGGGGS (SEQ ID NO:73), GSAGSAAGSGEFGSAEAAAKEAAAKAGSAGSAAGSGEFGS (SEQ ID NO:74), GSAGSAAGSGEFAEAAAKEAAAKAGSAGSAAGSGEF (SEQ ID NO:75), or GSAGSAAGSGEFGSAEAAAKEAAAKEAAAKEAAAKAGSAGSAAGSGEFGS (SEQ ID NO:76).

(61) In some embodiments, the ligands on the scaffold can be separated by linkers that are 20-50 amino acid residues in length (e.g., 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, or 50 amino acid residues in length). In some embodiments, the IDs engineered at the C-terminus or N-terminus of each scaffolded enzyme can contain a linker (e.g., a flexible linker) of 15 to (e.g., 20) amino acid residues in length flanking a spacer of 15 to 50 (e.g. 36) amino acid residues. In some embodiments, the ID can be separated from the enzyme by a spacer sequence such as the cTPR6 spacer, which includes sextuplicate rigid α-helical moieties and can have the sequence:

(62) TABLE-US-00006 (SEQ ID NO: 77)
AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQ
GDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEDYQKAL
ELDPNNLQAEAWKNLGNAYYKQGDYQKAIEYYQKALELDPNNASAWYN
LGNAYYKQGDYQKAIEYYQKALELDPNNAKAWYRRGNAYYKQGDYQKA
IEDYQKALELDPNNRSRSA.

(63) In some embodiments, the engineered enzyme can be of a formula: enzyme −linker.sub.1−spacer−linker.sub.2−motif.sub.1−linker.sub.3−motif.sub.2, where linkers 1, 2, and 3 can be the same or different, and motif 1 and motif 2 can be the same or different. In some embodiments, linker 1 can be referred to as the enzyme linker, i.e., it connects the enzyme to the spacer such as cTPR6 spacer, and can include flexible GS-rich moieties flanking a rigid α-helical moiety such as KLSGGGGSGGGGSGGGGS (SEQ ID NO:73). In some embodiments, linker 2 can be referred to as the ID linker and can include, for example, flexible GS-rich moieties flanking a rigid α-helical moiety such as GGGGSGGGGSGGGGAS (SEQ ID NO:78). In some embodiments, linker 3 can be referred to as the motif linker and can include flexible GS-rich moieties flanking a rigid α-helical moiety such as

(64) TABLE-US-00007 (SEQ ID NO: 74)
GSAGSAAGSGEFGSAEAAAKEAAAKAGSAGSAAGSGEFGS.

Table 1 provides non-limiting examples of motifs 1 and motifs 2, which are used together to form heterologous IDs. FIG. **3** contains a schematic of an exemplary engineered enzyme of this formula complexed with a scaffold. FIG. **6**B and FIGS. **13**A-C contain the amino acid sequence of an ATP citrate lyase, atoB, a 3-hydroxybutyryl-CoA dehydrogenase, an enoyl-CoA hydratase, a trans-enoyl-CoA reductase, a beto-ketothiolase (bktB), an HMG-CoA synthase, a truncated HMG-CoA reductase, a mevalonate kinase, a phosphomevalonate kinase, a diphosphomevalonate decarboxylase, an isopentenyl-diphosphate delta isomerase, a geranyl-diphosphate synthase

(ERG20.sup.WW), an olivetol synthase, an olivetolic acid cyclase, a CBGA synthase, and an acetyl-CoA carboxylase according to this formula. In some embodiments, linkers 1 and 2 can be (G.sub.4S).sub.3, the spacer can be the cTPR6 sequence, and linker 3 can be (GS).sub.8.

(65) In some embodiments, a scaffold can be of a formula: N-terminus−[Ligand #1− linker−Ligand #2−Spacer]n−(optionally-tagged)C-terminus, where n is the number of interaction domains. The linker can be referred to as a scaffolded ligand linker and can be used to connect and separate paired motif-binding ligands that recruit/localize each enzyme to its scaffold-binding site. Such a linker can include flexible GS-rich moieties flanking a rigid α-helical moiety and have a sequence such as GSAGSAAGSGEFAEAAAKEAAAKAGSAGSAAGSGEF (SEQ ID NO:75). The spacer can be referred to as a scaffolded ID-binding site spacer and can be used to connect and separate the scaffold-binding sites (composed of the paired motif binding ligands) for each enzyme. Such a spacer can include flexible GS-rich moieties flanking a rigid α-helical moiety and have a sequence such as GSAGSAAGSGEFGSAEAAAKEAAAKEAAAKEAAAKAGSAGSAAGSGEFGS (SEQ ID NO:76). The N-terminus can include a flexible GS-rich sequence to help stabilize and solubilize the scaffold. For example, the N-terminus can have the sequence GSAGSAAGSGEFGSAGSAAGSGEFGSAGSAAGSGEF (SEQ ID NO:79). The C-terminus can include a flexible GS rich sequence flanking a rigid α-helical moiety to stabilize and solubilize the scaffold and can be optionally tagged (e.g., with a MYC tag, a FLAG tag, or other tag described below) to ease purification or detection of the scaffold. For example, a C-terminal sequence with a triplicate MYC tag can have the sequence GSAGSAAGSGEFGSAEAAAKEAAAKEAAAKEAAAKAGSAGSAAGSGEFGSEQK LISEEDLEQKLISEEDLEQKLISEEDLGSAGSAAGSGEFGSAGSAAGSGEFGSAGS AAGSGEF (SEQ ID NO:80). For example, a C-terminal sequence with a triplicate FLAG tag can have the sequence GSAGSAAGSGEFGSAEAAAKEAAAKEAAAKEAAAKAGSAGSAAGSGEFGSDYK DDDDKDYKDDDDKDYKDDDDKGSAGSAAGSGEFGSAGSAAGSGEFGSAGSAA GSGEF (SEQ ID NO:81). FIG. **6**C and FIG. **13**D each contain an example of a scaffold polypeptide of this formula that contains the peptide ligands corresponding to IDS 1-16 as shown in Table 2, and a triplicate MYC tag on the C-terminus. For example, FIG. **13**D contains an example of a scaffold polypeptide (see SCF gene cassette of FIG. **2**B) containing a triplicate MYC tag. FIG. **6**D and FIG. **13**D each contain an example of a scaffold polypeptide that contains the peptide ligands corresponding to IDs 1 and 17 as shown in Table 2 and a triplicate FLAG tag on the C-terminus. Accordingly, the amino acid sequence of a scaffold can depend on the sequence of the peptide ligands that can bind to the selected ID motif of the enzymes.

(66) In some embodiments, any one of the enzymes can be engineered to include an N-terminal or C-terminal linker motif that allows covalent (isopeptide) bonding to the scaffold. See, for example, the SpyTag and SpyCatcher system described by Zakeri, et al., *Proc. Natl. Acad. Sci.*, 2012 109 (12) E690-E697.

(67) In some embodiments involving multi-enzymatic scaffolds described herein, the first engineered enzyme of a biosynthetic pathway can produce a first product that can be a substrate for the second engineered enzyme of the biosynthetic pathway, the second engineered enzyme of the biosynthetic pathway can produce a second product that can be a substrate for the third engineered enzyme of the biosynthetic pathway, and so forth. In some cases, the second engineered enzyme can be immobilized on the scaffold such that it is positioned adjacent to or very close to the first engineered enzyme. The third engineered enzyme can be immobilized on the scaffold such that it is positioned adjacent or very close the second engineered enzyme. In this way, the effective concentration of the first product can be high, and the second engineered enzyme can act efficiently on the first product, the third engineered enzyme can act efficiently on the second product, and so forth.

(68) As shown in FIGS. **1**A and **1**B, one example of a multi-enzymatic scaffold contains enzymes

of the hexanoyl-CoA pathway on the N-terminus of the scaffold, enzymes of the mevalonate pathway on the C-terminus of the scaffold, and enzymes of the upper cannabinoid pathway in between. Within any of the pathways, the enzymes can be from a single source, i.e., from one species or genera, or can be from multiple sources, i.e., different species or genera. Nucleic acids encoding the enzymes described herein have been identified from various organisms and are readily available in publicly available databases such as GenBank or EMBL (see below).

(69) A fully-assembled multi-enzymatic scaffold provided herein can adopt stoichiometry and a spatial arrangement that can help maximize pathway flux and minimize accumulation of pathway intermediates and by-products. Such scaffolds can facilitate substrate channeling both within and between cannabinoid and cannabinoid precursor pathways. Specifically, this scaffolding system can facilitate unidirectional flux through each of the primary cannabinoid precursor pathways, and converging near the midpoint of the scaffold. The hexanoyl-CoA/olivetolic acid (OVA) pathway can begin at the N-terminus of the scaffold, and the mevalonate or MEP pathway can begin at the C-terminus of the scaffold. The enzyme catalyzing the rate-limiting/committed step in cannabinoid biosynthesis, a CBGA synthase, can be localized at the intersection of these precursor pathways near the scaffold midpoint.

(70) By this design, the two primary precursors for cannabinoid biosynthesis, hexanoyl-CoA/olivetolic acid and geranyl pyrophosphate, can be bi-directionally delivered to a CBGA synthase at this intersection. The CBGA synthase can catalyze biosynthesis of CBGA, the primary cannabinoid from which all other cannabinoids are derived. Substrate channeling within and between the scaffolded pathways can accelerate the kinetics of the composite pathway in accordance with the law of mass action.

(71) In the embodiment shown in FIGS. **1**A and **1**B, the N-terminal hexanoyl-CoA pathway can include an ATP citrate lyase (ACL) (also can be referred to as an ATP citrate synthase), an acetyl-CoA acetyltransferase (atoB), two 3-hydroxy-acyl-CoA dehydrogenases (BHBDs), two enoyl-CoA hydratases (ECHs), a beta-ketothiolase (bktB), and two trans-2-enoyl-CoA-reductases (ECRs).

(72) In the hexanoyl-CoA pathway shown in FIGS. **1**A and **1**B, citrate, from cellular metabolism and/or supplemented in the growth medium, can be used as a substrate for ACL-catalyzed acetyl-CoA synthesis. ACL is classified under EC 2.3.3.8. Acetyl-CoA can be used as a substrate for atoB-catalyzed acetoacetyl-CoA synthesis. atoB is classified under EC 2.3.1.9. Acetoacetyl-CoA can serve as the substrate for BHBD-catalyzed 3-hydroxybutanoyl-CoA synthesis. BHBD is classified under EC 1.1.1.157. 3-hydroxybutanoyl-CoA can serve as the substrate for ECH-catalyzed trans-but-2-enoyl-CoA synthesis. ECH is classified under EC 4.2.1.17. Trans-but-2-enoyl-CoA can serve as the substrate for ECR-catalyzed butanoyl-CoA synthesis. ECR is classified under EC 1.3.8.1. Butanoyl-CoA can serve as the substrate for bktB-catalyzed 3-keto-hexanoyl-CoA synthesis. bktB is classified under EC 2.3.1.9. The bktB catalyzing the production of 3-ketohexanoyl CoA from butanoyl-CoA can be the same as, or different from, the atoB used to catalyze the production of acetoacetyl-CoA from acetyl-CoA. 3-ketohexanoyl-CoA is the substrate for BHBD-catalyzed 3-hydroxyhexanoyl-CoA synthesis. BHBD is classified under EC 1.1.1.157. The BHBD catalyzing the production of 3-hydroxyhexanoyl-CoA can be the same as, or different from, the BHBD used to catalyze the production of 3-hydroxybutanoyl-CoA. 3-hydroxyhexanoyl-CoA can be the substrate for ECH-catalyzed trans-hex-2-enoyl-CoA synthesis. ECH is classified under 4.2.1.17. The ECH catalyzing the production of trans-hex-2-enoyl-CoA can be the same as, or different from, the ECH used to catalyze the production of trans-but-2-enoyl-CoA. Trans-hex-2-enoyl-CoA can be the substrate for ECR-catalyzed hexanoyl-CoA synthesis. ECR is classified under EC 1.3.1.38 or EC 1.3.1.44. The ECR catalyzing the production of hexanoyl-CoA can be the same as, or different from, the ECR used to catalyze the production of butanoyl-CoA

(73) In some embodiments, a hexanoyl-CoA synthetase (HCS) enzyme can be substituted for the scaffolded enzymes of the hexanoyl-CoA pathway or can be included in a soluble form in addition to the scaffolded enzymes of the hexanoyl-CoA pathway, and in some embodiments, hexanoic acid

can be added to the growth media as a substrate for HCS-catalyzed hexanoyl-CoA production. The HCS can be included on the scaffold, N-terminal to the upper cannabinoid pathway in FIGS. **1**A and **1**B, and/or it can be non-scaffolded (soluble).

(74) In the embodiment shown in FIGS. **1**A and **1**B, the C-terminal mevalonate pathway can include an ACL, an atoB, a hydroxymethylglutaryl-CoA, an HMG-CoA synthase (HMGS), an HMG-CoA reductase (HMGR), a mevalonate kinase (ERG12), a phosphomevalonate kinase (ERGS), a diphospho mevalonate decarboxylase (MVD1), an isopentyl diphosphate isomerase (IDI1), and a mutant GPP synthase (mGPPS). In the mevalonate pathway shown in FIGS. **1**A and **1**B, citrate from cellular metabolism and/or supplemented in the growth medium, can be used as a substrate for ACL-catalyzed acetyl-CoA synthesis. ACL is classified under EC 2.3.3. Acetyl-CoA can be used as a substrate for bktB-catalyzed acetoacetyl-CoA synthesis. bktB is classified under EC 2.3.1.9. Acetoacetyl-CoA can be the substrate for HMGS-catalyzed HMG-CoA synthesis. HMG-CoA can be the substrate for HMGR catalyzed mevalonate synthesis. HMGR is classified under EC 1.1.1.88 or 1.1.1.34. Mevalonate can be the substrate for mevalonate kinase-catalyzed mevalonate-5 phosphate synthesis. Mevalonate kinase is classified under EC 2.7.1.36. Mevalonate-5-phosphate can be the substrate for phosphomevalonate kinase-catalyzed mevalonate pyrophosphate synthesis. Phosphomevalonate kinase is classified under EC 2.7.4.2. Mevalonate pyrophosphate can be the substrate for diphosphomevalonate decarboxylase-catalyzed isopentyl pyrophosphate synthesis. Diphosphomevalonate decarboxylase is classified under EC 4.1.1.33. Isopentyl pyrophosphate can be the substrate for isopentyl diphosphate isomerase-catalyzed dimethylallyl pyrophosphate synthesis. Isopentyl diphosphate isomerase is classified under EC. 5.3.3.2. Dimethylallyl pyrophosphate can be the substrate for geranyl pyrophosphate synthase (GPPS)-catalyzed geranyl pyrophosphate synthesis. GPPS is classified under EC 2.5.1.1.

(75) As acetyl-CoA can be the initial substrate for the hexanoyl-CoA, mevalonate/geranyl pyrophosphate, and malonyl-CoA cannabinoid precursor biosynthetic pathways, the inclusion of ACL at both the N-terminus and C-terminus of the multi-enzymatic scaffold in FIGS. **1**A and **1**B can directly couple the scaffolded pathways to cellular metabolism via ACL-catalyzed production of acetyl-CoA from citric acid cycle-derived citrate. The citrate also can be supplemented into the culture medium (e.g., as buffered citrate). In some embodiments, the ACL enzyme is included only at the N-terminus of the scaffold. In some embodiments, the ACL enzyme is included only at the C-terminus of the scaffold. In some embodiments, the ACL enzyme is included in soluble form.

(76) In some embodiments, the 2-C-methylerythritol 4-phosphate (MEP) pathway, which also can produce geranyl pyrophosphate, can be substituted for the scaffolded mevalonate pathway at the C-terminus of the scaffold or can be included in a soluble form in addition to the scaffolded mevalonate pathway. For example, as shown in FIG. **5**, the C-terminus of the scaffold can include a 1-deoxy-D-xylulose-5-phosphate (DOXP) synthase, a DOXP reductoisomerase, a MEP cytidyl transferase, a 4-diphosphocytidyl-2-C-methylerythritol (CDPME) kinase, a 2-C-methyl-D-erythritol 2,4-cyclodiphosphate (MECDP) synthase, a 4-hydroxy-3-methyl-but-2-enyl pyrophosphate (HIVIBPP) synthase, a HMBPP reductase, and a GPPS. Pyruvate and glyceraldehyde-3-phosphate (G3P) can be used as substrates for DOXP-synthase-catalyzed DOXP synthesis. DOXP is classified under EC 2.2.1.7. DOXP can be the substrate for DOXP reductoisomerase (DXR)-catalyzed MEP synthesis. DXR is classified under EC 1.1.1.267. MEP can be the substrate for 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (ISPD)-catalyzed 4-diphosphocytidyl-2-C-methylerythritol (CDP-ME) synthesis. ISPD is classified under EC 2.7.7.60. CDP-ME can be the substrate for 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (ISPE)-catalyzed 4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate (CDP-MEP) synthesis. ISPE is classified under EC 2.7.1.148. CDP-MEP can be the substrate for 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (ISPF)-catalyzed 2-C-methyl-D-erythritol 2,4-cyclodiphosphate (cMEPP) synthesis. ISPF is classified under EC 4.6.1.12. cMEPP can be the substrate for HMB-PP synthase (ISPG)-catalyzed (E)-4-Hydroxy-3-methyl-but-2-enyl pyrophosphate (HMBPP)

synthesis. ISPG is classified under EC 1.17.7.1. HMBPP can be the substrate for 4-hydroxy-3-methylbut-2-enyl diphosphate reductase (ISPH)-catalyzed isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) synthesis. ISPH is classified under EC 1.17.1.2. IPP and DMAPP can be substrates for GPPS-catalyzed geranyl pyrophosphate synthesis. GPPS is classified under EC 2.5.1.1.

(77) In some embodiments, the mevalonate pathway can be substituted for the scaffolded MEP pathway at the C-terminus of the scaffold or can be included in a soluble form in addition to the scaffolded MEP pathway.

(78) In the embodiment shown in FIG. **1**A and FIG. **1**B, a second multi-enzymatic scaffold can be co-expressed to enhance cytosolic titers of malonyl-CoA, another secondary substrate which can be used in cannabinoid biosynthesis. Such a scaffold can include an ATP citrate lyase (ACL) and acetyl-CoA carboxylase (ACC) in tandem. In some embodiments, the ACL and ACC are paired in duplicate or triplicate along the scaffold. If the ACL and ACC are paired in duplicate or triplicate, the two or three ACLs on the scaffold can be the same or different, and the two or three ACCs can be the same or different. In any of the embodiments, malonyl-CoA can be supplemented into the growth media instead of, or in addition to, being supplied by a scaffolded malonyl-CoA pathway.

(79) In any of the embodiments in which an ACL enzyme is used, a pyruvate dehydrogenase (E1) and a dihydrolipoyl transacetylase (E2) can be substituted for the ACL. For example, as shown in FIG. **4**, a pyruvate dehydrogenase (E1) and a dihydrolipoyl transacetylase (E2) can be substituted upstream of scaffolded mevalonate, hexanoyl-CoA, and malonyl-CoA pathways. Using both a pyruvate dehydrogenase (E1) and a dihydrolipoyl transacetylase can allow acetyl-CoA to be produced using pyruvate rather than citrate as the primary substrate. In such embodiments, pyruvate also can be supplemented in the growth media. Pyruvate dehydrogenases and dihydrolipoyl transacetylases are constituents of the multi-enzyme pyruvate dehydrogenase complex that catalyze acetyl-CoA production from pyruvate. E1 and E2 are found in bacteria and eukaryotes.

(80) As shown in FIG. **1**A and FIG. **1**B, the co-scaffolded upper cannabinoid pathway can include an olivetol synthase (OS), an olivetolic acid cyclase (OAC), and an aromatic prenyl-transferase (APT) such as a CBGA synthase (CBGAS). The upper cannabinoid pathway can begin using hexanoyl-CoA and three malonyl CoAs as the substrate for olivetol synthase-catalyzed 3,5,7-trioxododecanoyl-CoA synthesis. Olivetol synthase is classified under EC 2.3.1.206. 3,5,7-trioxododecanoyl-CoA can be used as a substrate for OAC-catalyzed olivetolic acid synthesis. OAC is classified under EC 4.4.1.26.

(81) At the flux intersection of the converging N-terminal hexanoyl-CoA/upper cannabinoid and C-terminal mevalonate/MEP pathways (near the scaffold midpoint), an APT such as CBGAS can use olivetolic acid from the hexanoyl-CoA/upper cannabinoid pathways and geranyl pyrophosphate from the mevalonate or MEP pathway as substrates for cannabigerolate synthesis. A suitable APT is classified under EC 2.5.1.102.

(82) In some embodiments, enzymes in the upper cannabinoid pathway can be scaffolded with a hexanoyl-CoA synthetase (HCS) to biosynthesize cannabigerolate. In some embodiments, a soluble HCS can be used with scaffolded enzymes of the upper cannabinoid pathway to biosynthesize cannabigerolate as shown in FIG. **7**. Suitable enzymes for the upper cannabinoid pathway are described above.

(83) In some embodiments, a minimal bidirectional scaffold, such as the one depicted in FIG. **8**, can be used in which HCS is on the N-terminus of the scaffold, a GPPS is on the C-terminus of the scaffold, and enzymes in the upper cannabinoid pathway are scaffolded between the HCS and GPPS.

(84) In some embodiments, such as the embodiment shown in FIG. **9**, the enzymes in the upper cannabinoid pathway can be scaffolded, while the enzymes in the hexanoyl-CoA pathway, enzymes in the mevalonate pathway, and enzymes in the malonyl-CoA pathway can be soluble. In some

embodiments, the enzymes in the upper cannabinoid pathway can be scaffolded, while the enzymes in the hexanoyl-CoA pathway, enzymes in the MEP pathway, and enzymes in the malonyl-CoA pathway can be soluble. In such embodiments, HCS can be substituted for the soluble forms of the enzymes of the hexanoyl-CoA pathway. Suitable enzymes for each of these pathways are described above.

(85) In some embodiments, the enzymes in the upper cannabinoid pathway can be scaffolded, while a hexanoyl-CoA synthase, enzymes in the mevalonate or MEP pathway, and enzymes in the malonyl-CoA pathway can be soluble. Suitable enzymes for each of these pathways are described above.

(86) In some embodiments, a HCS can be scaffolded N-terminally relative to the scaffolded enzymes in the upper cannabinoid pathway, while enzymes in the mevalonate or MEP pathway, and enzymes in the malonyl-CoA pathway can be soluble. Suitable enzymes for each of these pathways are described above.

(87) In some embodiments, the enzymes in the upper cannabinoid pathway can be scaffolded, while the enzymes in the hexanoyl-CoA pathway or a hexanoyl-CoA synthase and enzymes in the mevalonate or MEP pathways can be soluble. In some embodiments, the enzymes in the hexanoyl-CoA pathway or a hexanoyl-CoA synthase can be scaffolded N-terminal to the enzymes in the upper cannabinoid pathway, and enzymes in the mevalonate or MEP pathways can be soluble. In such embodiments, malonyl-CoA can be supplemented. Suitable enzymes for each of these pathways are described above.

(88) In some embodiments, such as the embodiment shown in FIG. **10**, a bi-directional scaffold can include enzymes of the malonyl-CoA (MCA) pathway on the N-terminus of the scaffold, enzymes of the mevalonate pathway on the C-terminus of the scaffold, and enzymes in the upper cannabinoid pathway in between. In some embodiments, a bi-directional scaffold can include enzymes of the malonyl-CoA pathway on the N-terminus of the scaffold, enzymes of the MEP pathway on the C-terminus of the scaffold, and enzymes in the upper cannabinoid pathway in between. In such embodiments, enzymes of the hexanoyl-CoA pathway can be on a separate scaffold or can be soluble. In some embodiments, HCS can be substituted for scaffolded or soluble enzymes of the hexanoyl-CoA pathway.

(89) In some embodiments, each of the pathways are on separate scaffolds. For example, in one embodiment, enzymes of the upper cannabinoid pathway can be on one scaffold, enzymes of the mevalonate or MEP pathway can be localized on one scaffold, enzymes of the hexanoyl-CoA pathway can be localized on one scaffold, and enzymes of the malonyl-CoA pathway can be localized on another scaffold.

(90) Cannabigerolic acid biosynthesized in any of the embodiments described herein can be isolated and/or can be used as a substrate for synthesis of other secondary and tertiary cannabinoids using downstream cannabinoid synthases. In order to generate a more diverse profile of cannabinoids, the downstream cannabinoid synthases typically are not scaffolded, as scaffolding would favor production of the terminal cannabinoid. In some embodiments, however, one or more of the downstream cannabinoid synthases can be included on a scaffold described herein.

(91) For example, one or more of cannabidiolic acid synthase (CBDAS), cannabichromenic acid synthase (CBCAS), tetrahydrocannabinolic acid synthase (THCAS), or other cannabinoid synthases can be used to produce additional cannabigerolate-derived cannabinoids. For example, a CBDAS; a CBCAS; a THCAS; a CBDAS and a CBCAS; a CBDAS and a THCAS; a CBCAS and a THCAS; or a CBDAS, CBCAS, and THCAS can be used to produce additional cannabigerolate-derived cannabinoids such as one or more of cannabiodiolic acid, cannabichromenic acid, and delta-9 tetrahydrocannabinolic acid. CBDAS is classified under EC 1.21.3.8 and can catalyze the synthesis of cannabidiolic acid from cannabigerolic acid. CBCAS is classified under EC 1.3.3— and can catalyze the synthesis of cannabichromenic acid from cannabigerolic acid. THCAS is classified under EC 1.21.3.7 and can catalyze the synthesis of delta-9 tetrahydrocannabinolic acid

from cannabigerolic acid.

(92) Host Cells for Producing Cannabinoids

(93) Cannabinoids can be produced in host cells or in vitro using a multi-enzymatic scaffold as described herein. Suitable host cells include any microorganism, eukaryotic or prokaryotic, such as bacteria (e.g., *Escherichia coli, Bacillus, Brevibacterium, Streptomyces*, or *Pseudomonas*), yeast (e.g., *Pichia pastoris, Saccharomyces cerevisiae, Yarrowia lipolytica, Kluyveromyces marxianus*, or Komagataella phaffii) and other fungi (e.g., *Neurospora crassa*), and green algae (e.g., *Dunaliella* sp., *Chlorella variabilis, Euglena mutabilis*, or *Chlamydomonas reinhardtii*), as well as plant cells (e.g., tobacco, *Cannabis*, or other photosynthetic plant cells) that can be maintained in culture or, in the case of plant cells such as those from tobacco or *cannabis* plants, can be engineered in culture and cultivated as intact transgenic plants. Such host cells or plant may or may not naturally produce cannabinoids.

(94) A host cell can be modified to contain one or more exogenous nucleic acids that encode a scaffold as described herein and one or more exogenous nucleic acids that encode the engineered enzymes. The term “nucleic acid” as used herein encompasses both RNA and DNA, including cDNA, genomic DNA, and synthetic (e.g., chemically synthesized) DNA. The nucleic acid can be double-stranded or single-stranded. Where single-stranded, the nucleic acid can be the sense strand or the antisense strand. In addition, nucleic acid can be circular or linear.

(95) The term “exogenous” as used herein with reference to nucleic acid and a particular host cell refers to any nucleic acid that does not originate from that particular host cell as found in nature. Thus, non-naturally-occurring nucleic acid is considered to be exogenous to a host cell once introduced into the host cell. It is important to note that non-naturally-occurring nucleic acid can contain nucleic acid sequences or fragments of nucleic acid sequences that are found in nature provided the nucleic acid as a whole does not exist in nature. For example, a nucleic acid molecule containing a genomic DNA sequence within an expression vector is non-naturally-occurring nucleic acid, and thus is exogenous to a host cell once introduced into the host cell, since that nucleic acid molecule as a whole (genomic DNA plus vector DNA) does not exist in nature. Thus, any vector, autonomously replicating plasmid, or virus (e.g., retrovirus, adenovirus, or herpes virus) that as a whole does not exist in nature is considered to be non-naturally-occurring nucleic acid. It follows that genomic DNA fragments produced by PCR or restriction endonuclease treatment as well as cDNAs are considered to be non-naturally-occurring nucleic acid since they exist as separate molecules not found in nature. It also follows that any nucleic acid containing a promoter sequence and polypeptide-encoding sequence (e.g., cDNA or genomic DNA) in an arrangement not found in nature is non-naturally-occurring nucleic acid.

(96) A nucleic acid that is naturally-occurring can be exogenous to a particular cell. For example, an entire chromosome isolated from a cell of organism X is an exogenous nucleic acid with respect to a cell of organism Y once that chromosome is introduced into Y's cell.

(97) It is noted that a host cell can be given an exogenous nucleic acid molecule that encodes a polypeptide having an enzymatic activity that catalyzes the production of a compound not normally produced by that host cell. Alternatively, or additionally, a host cell can be given an exogenous nucleic acid molecule that encodes a polypeptide having an enzymatic activity that catalyzes the production of a compound that is normally produced by that host cell. In this case, the recombinant host cell can produce more of the compound, or can produce the compound more efficiently, than a similar host cell not having the genetic modification.

(98) An enzyme having a particular enzymatic activity can be a polypeptide that is either naturally-occurring or non-naturally-occurring. A naturally-occurring polypeptide is any polypeptide having an amino acid sequence as found in nature, including wild-type and polymorphic polypeptides. Such naturally-occurring polypeptides can be obtained from any species including, without limitation, animal (e.g., mammalian), plant, fungal, and bacterial species. A non-naturally-occurring polypeptide is any polypeptide having an amino acid sequence that is not found in nature. Thus, a

non-naturally-occurring polypeptide can be a mutated version of a naturally-occurring polypeptide, or an engineered polypeptide such as the engineered enzymes described herein that contain IDs. For example, a non-naturally-occurring polypeptide having geranyl pyrophosphate synthase activity can be a mutated version of a naturally-occurring polypeptide having geranyl pyrophosphate synthase activity. For example, the GPPS encoded by Erg20 may include a substitution of a tryptophan for phenylalanine at position 96 and a substitution of a tryptophan for asparagine at position 127 (referred to as Erg20.sup.WW). Erg20.sup.WW favors production of geranyl pyrophosphate over farnesyl pyrophosphate. See, Jiang, et al., *Metab Eng.* 2017, 41:57-66. For example, a truncated HMGR (tHMGR) such as an N-terminally truncated HMGR that includes the catalytic domain but not the transmembrane or regulatory domains of HMGR can be used. For example, the HMGR from *A. thaliana* (GenBank Accession No. J04537) or a HMGR from *S. cerevisiae* (which contains only residues 646-1025) can be truncated to remove the transmembrane and/or regulatory domains and used in a scaffold described herein to remove a bottleneck in the mevalonate pathway. HMGR catalyzes the rate-limiting step in the mevalonate pathway (see, e.g., Song et al., 2017, *Scientific reports*, doi:10.1038/s41598-017-15005-4). For example, the nucleic acid encoding an atoB from *S. cerevisiae* can be modified to contain a synthetic 5′ UTR (such as the synthetic 5′ UTR sequence: 5′-cggcacccctacaaacagaaggaatataaa-3′ (SEQ ID NO:82)) and can be used in the scaffold as it alters atoB expression to facilitate flux-rebalancing in favor of production of acetoacetyl-CoA over the reverse reaction product butyryl-CoA (see Kim et al., 2018, *Bioresour Technol*, doi: 10.1016/j.biortech. 2017.10.014). A polypeptide can be mutated by, for example, sequence additions, deletions, substitutions, or combinations thereof.

(99) Any of the enzymes described herein that can be used to produce one or more cannabinoids can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of the corresponding wild-type enzyme. It will be appreciated that the sequence identity can be determined on the basis of the mature enzyme (e.g., with any signal sequence removed).

(100) For example, an ACL can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of a *Homo sapiens* ACL (see SEQ ID NO:83, FIG. **6**A), or an ACL from *Rattus norvegicus, Mus musculus*, or *Ciona intestinalis*, e.g., GenBank Accession Nos. AAA74463, AAK56081, and BAB00624, respectively.

(101) For example, an acetyl-CoA acetyltransferase (atoB) can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of an *Escherichia coli* atoB (see SEQ ID NO:84, FIG. **6**A), or an atoB from *Cupriavidus necator, Clostridium acetobutylicum*, or *Arabidopsis thaliana*, e.g., GenBank Accession Nos. CAJ92573, AAK80816, and AAM67058, respectively. In some embodiments, a malonyl-CoA acyl carrier protein transacylase from *Saccharomyces cerevisiae, Homo sapiens, Serratia plymuthica*, or *Dickeya paradisiaca* can be substituted for atoB, e.g., GenBank Accession Nos. DAA10992, AAH30985, AG055277, and ACS85236, respectively.

(102) For example, a 3-hydroxy-butyryl-CoA dehydrogenase (BHBD) can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of a *Clostridium acetobutylicum* BHBD (see SEQ ID NO:85, FIG. **6**A), or a BHBD from *Escherichia coli, Treponema denticola*, or *Arabidopsis thaliana*, e.g., GenBank Accession Nos. AIZ91493, AAS11105, and AAN17431, respectively.

(103) For example, an enoyl-CoA hydratase (ECH) can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of a *Clostridium acetobutylicum* ECH (see SEQ ID NO:86, FIG. **6**A), or an ECH from *Acinetobacter oleivorans, Cupriavidus necator*, or *Acinetobacter baumannii*, e.g., GenBank Accession Nos. ADI91469, CAJ91294, and ACJ57023, respectively.

(104) For example, a beta-ketothiolase (bktB) can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of a

Cupriavidus necator bktB (see SEQ ID NO:87, FIG. **6**A), or a bktB from *Escherichia coli, Lactobacillus casei,* or *Clostridium acetobutylicum,* e.g., GenBank Accession Nos. ALI39443, CAQ67083, and AAK80816, respectively.

(105) For example, a trans-2-enoyl-CoA-reductase (ECR) can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of a *Treponema denticola* ECR (see SEQ ID NO:88, FIG. **6**A), or an ECR from *Cupriavidus necator, Saccharomyces cerevisiae,* or *Klebsiella michiganensis,* e.g., GenBank Accession Nos. AAP86010, DAA07148, and AIE72439, respectively.

(106) For example, a hexanoyl-CoA synthetase (HCS), which is a type of acyl-activating enzyme (AAE), can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of a *C. sativa* AAE1 (see SEQ ID NO:89, FIG. **6**A, GenBank Accession No. AFD33345) or *C. sativa* AAE3 (GenBank Accession No. AFD33347). The *C. sativa* AAE1 and AAE3 each can use hexanoate as a substrate. See, Stout, et al., *Plant* 1, 71(3): 353-365 (2012). In some embodiments, the AAE encoded by CsAAE1 can be used. See, GenBank Accession No. JN717233 for the coding sequence. In some embodiments, the AAE encoded by CsAAE3 can be used. See, GenBank Accession No. JN717233 for the coding sequence. In some embodiments, both CsAAE1 and CsAAE3 can be used.

(107) For example, an HMG-CoA synthase (HMGS) can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of a *S. cerevisiae* HMGS (see SEQ ID NO:90, FIG. **6**A), or an HMGS from *Arabidopsis thaliana, Lactobacillus casei,* or *Homo sapiens,* e.g., GenBank Accession Nos. AEE83052, CAQ67081, and AAA62411, respectively.

(108) For example, an HMG-CoA reductase (HMGR), N-terminally truncated or canonical, can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of a *S. cerevisiae* HMGS (see SEQ ID NO:91, FIG. **6**A), or an HMGR from *Arabidopsis thaliana, Lactobacillus casei,* or *Homo sapiens,* e.g., GenBank Accession Nos. AEE35849, CAQ67082, and AAA52679, respectively.

(109) For example, a mevalonate kinase can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of a *S. cerevisiae* mevalonate kinase (see SEQ ID NO:92, FIG. **6**A), or a mevalonate kinase from *Arabidopsis thaliana, Lactobacillus casei,* or *Homo sapiens,* e.g., GenBank Accession Nos. AAD31719, CAQ66794, and AAF82407, respectively.

(110) For example, a phosphomevalonate kinase can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of a *S. cerevisiae* phosphomevalonate kinase (see SEQ ID NO:93, FIG. **6**A), or a mevalonate kinase from *Scheffersomyces stipitis, Lactobacillus casei,* or *Homo sapiens,* e.g., GenBank Accession Nos. EAZ63544, CAQ66339, and AAH06089, respectively.

(111) For example, a diphosphomevalonate decarboxylase can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of a *S. cerevisiae* diphosphomevalonate decarboxylase (see SEQ ID NO:94, FIG. **6**A), or a diphosphomevalonate decarboxylase from *Arabidopsis thaliana, Lactobacillus casei,* or *Homo sapiens,* e.g., GenBank Accession Nos. AAC67348, CAQ66795, and AAC50440, respectively.

(112) For example, an isopentyl diphosphate isomerase can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of a *S. cerevisiae* isopentyl diphosphate isomerase (see SEQ ID NO:95, FIG. **6**A), or an isopentyl diphosphate isomerase from *Arabidopsis thaliana, Lactobacillus casei,* or *Homo sapiens,* e.g., GenBank Accession Nos. AAC49920, CAQ66796, and AAP35407, respectively.

(113) For example, a geranyl pyrophosphate synthase (GPPS) (also known as a geranyl-diphospate synthase) can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of the *S. cerevisiae* GPS or a GPPS from

Acinetobacter baumannii, Lactobacillus casei, or *Homo sapiens*, e.g., GenBank Accession Nos. ACJ56139, CAQ66932, and AAH10004, respectively. In some embodiments, a mutant GPPS can be used. For example, the GPPS encoded by Erg20 may include a substitution of a tryptophan for phenylalanine at position 96 and a substitution of a tryptophan for asparagine at position 127 (referred to as Erg20.sup.WW) (see SEQ ID NO:96, FIG. **6**A). Erg20.sup.WW favors production of geranyl pyrophosphate over farnesyl pyrophosphate. See, Jiang, et al., *Metab Eng.* 2017 41:57-66. In some cases, substituting a glutamic acid for lysine at position 179 of Erg20 (Erg20.sup.K179E) can be used to produce a GPPS that favors production of geranyl pyrophosphate. See, WO2016010827A1.

(114) For example, a DOXP synthase can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of an *Escherichia coli, Clostridium acetobutylicum, Treponema denticola*, or *Arabidopsis thaliana* DOXP synthase, e.g., GenBank Accession Nos. CDH63925, AAK80036, AAS12424, and ANM65835, respectively.

(115) For example, a DOXP reductoisomerase can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of an *Escherichia coli, Clostridium acetobutylicum, Treponema denticola*, or *Arabidopsis thaliana* DOXP reductoisomerase, e.g., GenBank Accession Nos. CDH63708, AAK79760, AAS12860, and AAM61343, respectively.

(116) For example, a MEP cytidyl transferase can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of an *Escherichia coli, Clostridium acetobutylicum, Treponema denticola*, or *Arabidopsis thaliana* MEP cytidyl transferase, e.g., GenBank Accession Nos. CDH66380, AAK81121, AAS12810, and BAB21592, respectively.

(117) For example, a CDPME kinase can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of an *Escherichia coli, Clostridium acetobutylicum, Treponema denticola*, or *Arabidopsis thaliana* CDPME kinase, e.g., GenBank Accession Nos. CDH64802, AAK80844, AAS11855, and AEC07908, respectively.

(118) For example, a MECDP synthase can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of an *Escherichia coli, Nicotiana tabacum, Treponema denticola*, or *Acinetobacter baumannii* MECDP synthase, e.g., GenBank Accession Nos. CDH66379, AHM22925, AAS12811, and ACJ59227, respectively.

(119) For example, an HMBPP synthase can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of an *Escherichia coli, Acinetobacter baumannii, Treponema denticola*, or *Arabidopsis thaliana* HMBPP synthase, e.g., GenBank Accession Nos. AAN81487, ACJ58210, AAS11783, and AED97354, respectively.

(120) For example, an HMBPP reductase can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of an *Escherichia coli, Acinetobacter baumannii, Treponema denticola*, or *Arabidopsis thaliana* HMBPP reductase, e.g., GenBank Accession Nos. CDH63564, ACJ57384, AAS11585, and AEE86362, respectively.

(121) For example, an acetyl-CoA carboxylase (ACC) can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of a *S. cerevisiae* acetyl-CoA carboxylase (see SEQ ID NO:97, FIG. **6**A), or an acetyl-CoA carboxylase from *Homo sapiens, Treponema denticola*, or *Cupriavidus necator*, e.g., GenBank Accession Nos. AAP94122, AAS11086, and CAQ67359, respectively.

(122) For example, a pyruvate dehydrogenase (E1) and dihydrolipoyl transacetylase (E2) can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of a *Saccharomyces cerevisiae, Escherichia coli, Clostridium acetobutylicum*, or *Cupriavidus necator* E1 and E2, e.g., GenBank Accession Nos. DAA07337,

AMC97367, CAQ66617, and CAJ92510 for E1, and DAA10474, AUG14916, CAQ66619, and CAJ92511 for E2, respectively.

(123) For example, an olivetol synthase (OS) can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of an OS from *C. sativa* set forth in SEQ ID NO:98 (FIG. **6**A) or the OS from *C. sativa* having GenBank Accession No. BAG14339. See, for example, Taura, et al., *FEBS Letters* 583 (2009) 2061-2066.

(124) For example, an olivetolic acid cyclase (OAC) can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of an OAC from *C. sativa* set forth in SEQ ID NO:99 (FIG. **6**A) or the OAC from *C. sativa* having GenBank Accession No. AFN42527. See, for example, Gagne, et al., *Proc. Natl. Acad. Sci. USA,* 2012 109 (31) 12811-12816.

(125) For example, a CBGAS can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of an aromatic prenyl-transferase (APT) from *Cannabis sativa* such as the CBGAS set forth in SEQ ID NO:100 (FIG. **6**A). See, for example, U.S. Patent Publication No. 20120144523A1 and U.S. Pat. No. 8,884,100B2. In some embodiments, a soluble APT from *Streptomyces* (e.g., NphB) can be used. See, for example, Carvalho et al., *FEMS Yeast Research,* 17, 2017, fox037.

(126) For example, a cannabidiolic acid synthase (CBDAS) can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of a CBDAS from *C. sativa* set forth in SEQ ID NO:101 (FIG. **6**A) or the amino acid sequence of a CBDAS from *C. sativa* having GenBank Accession No. BAF65033. See, for example, Taura, et al., *FEBS Lett.* 581 (16), 2929-2934 (2007).

(127) For example, a cannabichromenic acid synthase (CBCAS) can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of a CBCAS from *C. sativa* set forth in SEQ ID NO:102 (FIG. **6**A) or the amino acid sequence of a CBCAS from *C. sativa* as set forth in SEQ ID NO:2 of WO 2015/196275 A1. SEQ ID NO:2 of WO 2015/196275 A1 includes an N-terminal 28 amino acid signal peptide. All or a portion of the signal peptide can be removed from the sequence. The CBDAS from *C. indica* or *C. ruderalis* also can be used. In some embodiments, an *Escherichia coli* or yeast optimized nucleic acid sequence encoding a *C. sativa* CBCAS as set forth in SEQ ID NOs: 8 and 9, respectively, of WO 2015/196275 A1 can be used.

(128) For example, a tetrahydrocannabinolic acid synthase (THCAS) can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) to the amino acid sequence of a THCAS from *C. sativa* having GenBank Accession No. BAC41356. See, for example, Sirikantaramas, et al., *J. Biol. Chem.* 279 (38), 39767-39774 (2004).

(129) The percent identity (homology) between two amino acid sequences can be determined as follows. First, the amino acid sequences are aligned using the BLAST 2 Sequences (B12seq) program from the stand-alone version of BLASTZ containing BLASTP version 2.0.14. This stand-alone version of BLASTZ can be obtained from Fish & Richardson's web site (e.g., www.fr.com/blast/) or the U.S. government's National Center for Biotechnology Information web site (www.ncbi.nlm.nih.gov). Instructions explaining how to use the Bl2seq program can be found in the readme file accompanying BLASTZ. Bl2seq performs a comparison between two amino acid sequences using the BLASTP algorithm. To compare two amino acid sequences, the options of Bl2seq are set as follows: -i is set to a file containing the first amino acid sequence to be compared (e.g., C: \seq1.txt); -j is set to a file containing the second amino acid sequence to be compared (e.g., C: \seq2.txt); -p is set to blastp; -o is set to any desired file name (e.g., C: \output.txt); and all other options are left at their default setting. For example, the following command can be used to generate an output file containing a comparison between two amino acid sequences: C: \B12seq-i c: \seq1.txt-j c: \seq2.txt-p blastp-o c: \output.txt. If the two compared sequences share homology (identity), then the designated output file will present those regions of homology as aligned

sequences. If the two compared sequences do not share homology (identity), then the designated output file will not present aligned sequences. Similar procedures can be following for nucleic acid sequences except that blastn is used.

(130) Once aligned, the number of matches is determined by counting the number of positions where an identical amino acid residue is presented in both sequences. The percent identity (homology) is determined by dividing the number of matches by the length of the full-length polypeptide amino acid sequence followed by multiplying the resulting value by 100. It is noted that the percent identity (homology) value is rounded to the nearest tenth. For example, 78.11, 78.12, 78.13, and 78.14 is rounded down to 78.1, while 78.15, 78.16, 78.17, 78.18, and 78.19 is rounded up to 78.2. It also is noted that the length value will always be an integer.

(131) It will be appreciated that a number of nucleic acids can encode a polypeptide having a particular amino acid sequence. The degeneracy of the genetic code is well known to the art; i.e., for many amino acids, there is more than one nucleotide triplet that serves as the codon for the amino acid. For example, codons in the coding sequence for a given enzyme can be modified such that optimal expression in a particular species (e.g., bacteria or fungus) can be attained, using appropriate codon bias tables for that species. For example, the nucleotide sequences set forth in FIG. **12**A are the nucleic acid sequences encoding an ATP citrate lyase, an atoB, a 3-hydroxbutyryl-CoA dehydrogenase, an enoyl-CoA hydratase, a beto-ketothiolase (bktB), a trans-enoyl-CoA reductase, an HMG-CoA synthase, an HMG-CoA reductase, a mevalonate kinase, a phosphomevalonate kinase, a diphosphomevalonate decarboxylase, an isopentenyl-diphosphate delta isomerase, a geranyl-diphosphate synthase (ERG20.sup.WW), an olivetol synthase, an olivetolic acid cyclase, a CBGA synthase, a CBDA synthase, a CBCA synthase, an acetyl-CoA carboxylase, and a hexanoyl-CoA synthetase. The nucleic acid sequences for the ATP citrate lyase, atoB, 3-hydroxybutyryl-CoA dehydrogenase, enoyl-CoA hydratase, trans-enoyl-CoA reductase, bktB, olivetol synthase, olivetolic acid cyclase, CBGA synthase, CBDA synthase, and CBCA synthase have been codon optimized for expression in yeast. FIGS. **14**A-**14**C contain codon optimized (for expression in yeast) nucleic acid sequences encoding the engineered enzymes of FIGS. **13**A-**13**C.

(132) In addition to sequence similarity, it will be appreciated that enzymes and scaffolds with structural and/or functional similarity to the enzymes and scaffolds described herein are also encompassed within the scope of the document.

(133) This document provides recombinant host cells that can be used to produce one or more cannabinoids as described herein. For example, an individual host cell can contain exogenous nucleic acid such that the scaffold polypeptide and each of the enzymes to be immobilized on the scaffold are expressed. It is important to note that such host cells can contain any number and/or combination of exogenous nucleic acid molecules. For example, a particular host cell can contain an exogenous nucleic acid encoding the scaffold, and additional exogenous nucleic acids encoding the enzymes of the malonyl-CoA pathway, enzymes of the hexanoyl-CoA pathway or encoding a HCS, and enzymes of the mevalonate or MEP pathway. A single exogenous nucleic acid can encode one enzyme or more than one enzyme (e.g., one or more copies of from one to ten (or more) enzymes, from one to eight, from one to seven, from one to six, from one to five, from one to four, or from two to three enzymes). Thus, the number of different exogenous nucleic acids needed to produce the engineered enzymes to be localized on the scaffold will depend on the design of the scaffold and/or the particular embodiment. FIG. **2**A and FIG. **2**B each provide a non-limiting schematic of suitable gene cassettes for expressing the scaffolds and enzymes. FIG. **12**C provides the nucleic acid sequence encoding a scaffold polypeptide containing the peptide ligands corresponding to IDs 1-16 as shown in Table 2 and a triplicate MYC tag. See also FIG. **14**D for the codon-optimized nucleic acid sequence encoding the scaffold polypeptide of FIG. **13**D. FIG. **12**D provides the nucleic acid sequence encoding a scaffold polypeptide that contains the peptide ligands corresponding to IDs 1 and 17, and a triplicate FLAG tag. See also FIG. **14**D.

(134) In some embodiments, multiple nucleic acids encoding polypeptides (e.g., the nucleic acids of a gene cassette such as in FIG. **2**A or FIG. **2**B) can be linked together using a nucleic acid sequence encoding a self-cleaving peptide. During translation of the transcripts, the growing polypeptide can be cleaved at the 2A peptide with translation continuing through to the next polypeptide. When designing a vector to express the polypeptides as a polycistronic unit, the nucleic acid encoding the polypeptides and the self-cleaving peptide (e.g., a 2A peptide) can be designed such that they are in translational frame with each other. Examples of 2A peptides that can be used as described herein include, without limitation, a 2A peptide of foot-and-mouth disease virus (FMDV), a 2A peptide of equine rhinitis A virus (ERAVO), a 2A peptide of Thosea asigna virus (TaV), or a 2A peptide of porcine teschovirus-1 (PTV-1) or porcine teschovirus-2 (PTV-2). The 2A peptides from PTV-1 and PTV-2 are referred to as P2A peptides. See, e.g., SEQ ID NO:212 for a codon-optimized nucleotide sequence (for *S. cerevisiae*) encoding a P2A peptide.
(135) Further, the cells described herein can contain a single copy or multiple copies (e.g., about 5, 10, 20, 35, 50, 75, 100 or 150 copies), of a particular exogenous nucleic acid molecule. Again, the cells described herein can contain more than one particular exogenous nucleic acid molecule and/or copies thereof. For example, a particular cell can contain about 50 copies of exogenous nucleic acid molecule X as well as about 75 copies of exogenous nucleic acid molecule Y.
(136) Any method can be used to introduce an exogenous nucleic acid molecule into a host cell. In fact, many methods for introducing nucleic acid into host cells such as bacteria and yeast are well known to those skilled in the art. For example, heat shock, lipofection, electroporation, nucleofection, conjugation, fusion of protoplasts, and biolistic delivery are common methods for introducing nucleic acid into bacteria and yeast cells. See, e.g., Ito et al., *J. Bacterol.* 153:163-168 (1983); Durrens et al., *Curr. Genet.* 18:7-12 (1990); and Becker and Guarente, *Methods in Enzymology* 194:182-187 (1991).
(137) An exogenous nucleic acid molecule contained within a particular host cell can be maintained within that host cell in any form. For example, exogenous nucleic acid molecules can be integrated into the genome of the microorganism or maintained in an episomal state. In other words, a microorganism can be a stable or transient transformant. Again, a microorganism described herein can contain a single copy, or multiple copies (e.g., about 5, 10, 20, 35, 50, 75, 100 or 150 copies), of a particular exogenous nucleic acid molecule as described herein.
(138) Suitable nucleic acid constructs for expressing the engineered enzymes and scaffolds include, for example, CRISPR plasmids, baculovirus vectors, bacteriophage vectors, plasmids, phagemids, cosmids, fosmids, bacterial artificial chromosomes, viral vectors (for example, viral vectors based on vaccinia virus, poliovirus, adenovirus, adeno-associated virus, SV40, herpes simplex virus, and the like), P1-based artificial chromosomes, yeast plasmids, yeast artificial chromosomes, and other vectors. Typically such constructs include a regulatory element that promotes the expression of a nucleic acid sequence that encodes a polypeptide. Typically, regulatory elements are DNA sequences that regulate the expression of other DNA sequences at the level of transcription. Thus, regulatory elements include, without limitation, promoters, enhancers, and the like. Any type of promoter can be used to express an amino acid sequence from an exogenous nucleic acid molecule. Examples of promoters include, without limitation, constitutive promoters, tissue-specific promoters, and inducible or repressible promoters that are responsive or unresponsive to a particular stimulus (e.g., light, oxygen, chemical concentration, sound, and the like).
(139) In some embodiments, endogenous yeast promoters with varying constitutive activity levels can be used to express the engineered enzymes and/or scaffolds. To maintain an excess of enzymes relative to scaffold molecules, the scaffolds can be expressed under control of the weakest promoter. For example, one or more of the following yeast promoters can be used: the promoter from the gene encoding transcriptional elongation factor EF-1α (pTEF1), the promoter from the gene encoding phosphoglycerate kinase (PGK1), the promoter from the gene encoding triose phosphate isomerase (pTPI1), the promoter from the gene encoding a hexose transporter (pHXT7),

HXT7, the promoter from the gene encoding pyruvate kinase 1 (pPYK1), the promoter from the gene encoding alcohol dehydrogenase 1 (pADH1), or the promoter from the gene encoding triphosphate dehydrogenase (pTDH3). For example, in the embodiment shown in FIG. **2**A, the pTPI1 promoter can be used to express enzymes of the upper hexanoyl-CoA (HCA), enzymes of the lower HCA pathway, enzymes of the upper mevalonate (MVA) pathway, enzymes of the lower MVA pathway, and enzymes of the lower cannabinoid (CB) pathway, while the pTEF1 promoter can be used to express enzymes of the upper CB pathway, the atoB enzyme, and the enzymes of the malonyl-CoA pathway, and the pADH1 promoter can be used to express the scaffold. Of these promoters, the pADH1 promoter has the weakest activity (+ in FIG. **2**A), the pTEF1 promoter has the strongest activity (+++ in FIG. **2**A), and the activity of the pTPI1 promoter is between the other two (++ in FIG. **2**A). In some embodiments, the Gal 1-10 promoter (e.g., from *S. cerevisiae*) can be used. See, e.g., FIG. **17**.

(140) A nucleic acid construct also can include a selectable marker, e.g., for an antibiotic such as neomycin resistance, ampicillin resistance, tetracycline resistance, chloramphenicol resistance, or kanamycin resistance). In some embodiments, a nutritional marker gene that confers prototrophy for an essential nutrient such as tryptophan (TRP1), uracil (URA3), histidine (HIS3), leucine (LEU2), lysine (LYS2), or methionine can be included on a nucleic acid construct. See, e.g., FIG. **17**. As shown in Example 3, four different auxotrophic markers were used to sequentially select for transformed cells containing the desired combinations of nucleic acids encoding the enzymes and scaffold. For example, yeast cells transformed with a vector containing a TRP gene and the nucleic acids encoding enzymes of the hexanoyl-CoA pathway were grown in tryptophan deficient media. The transformed cells that grew in the tryptophan deficient media were selected and further transformed with a vector containing a LEU gene and nucleic acid encoding enzymes of the mevalonate pathway. The resulting transformed cells were grown on media lacking tryptophan and leucine, and the cells that grew in the media lacking tryptophan and leucine were transformed with a vector containing a HIS gene and nucleic acids encoding enzymes of the upper cannabinoid pathway. The resulting transformed cells were grown on media lacking tryptophan, leucine, and histidine, and the cells that grew in the media lacking tryptophan, leucine, and histidine were transformed with a vector containing a URA3 gene and a nucleic acid encoding a scaffold. The resulting transformed cells were grown on media lacking tryptophan, leucine, histidine, and uracil. Cells that grew in media lacking tryptophan, leucine, histidine, and uracil contained the desired combination of enzymes and scaffold as shown in FIG. **1**B.

(141) In some embodiments, the encoded enzymes (e.g., one or more enzymes from the cannabinoid biosynthesis pathway, mevalonate pathway, MEP pathway, hexanoyl-CoA pathway, or a hexanoyl-CoA synthetase) and/or the scaffold can include a targeting sequence that can be used to direct the enzymes or scaffold to one of several different intracellular compartments, including, for example, the endoplasmic reticulum (ER), mitochondria, plastids (such as chloroplasts), the vacuole, the Golgi apparatus, or protein storage vesicles (PSV). For example, a mitochondrial or plastidial targeting sequence can be used to facilitate mitochondrial or plastidial compartmentalization of cannabinoid/cannabinoid precursor biosynthesis such that the encoded enzymes and scaffold are expressed in the mitochondria or plastids of the host cell.

(142) In some embodiments, cannabinoid/cannabinoid precursor biosynthesis can be performed in two compartments by co-expressing one or more engineered enzymes and a scaffold in both the cytosolic compartment and either the plastids or mitochondria of the host cell. See, for example, FIG. **11**. It will be appreciated that while FIG. **11** depicts a scaffold containing enzymes of the hexanoyl-CoA pathway, enzymes of the upper cannabinoid pathway, and enzymes of the mevalonate pathway, dual-compartment engineering can be performed with any of the scaffolds and enzymes described herein. For example, dual-compartment engineering can be performed in two compartments by co-expressing a scaffold and enzymes of the hexanoyl-CoA pathway, enzymes of the upper cannabinoid pathway, and enzymes of the MEP pathway in both the cytosolic

compartment and either the plastids of mitochondria of the host cell. Dual-compartment engineering also can be achieved by engineering separate haploid yeast strains for cytosolic and mitochondrial/plastidial cannabinoid biosynthesis, and then mating these two haploid strains to produce a diploid lineage that is heterozygous for cytosolic and mitochondrial/plastidial cannabinoid biosynthesis.

(143) In some embodiments, the engineered enzymes and/or scaffolds also contain a tag that can be used for purification of the recombinant protein (e.g., c-myc, FLAG, polyhistidine (e.g., hexahistidine), hemagglutinin (HA), glutathione-S-transferase (GST), or maltose binding protein (MBP)) or as a detectable marker (e.g., luciferase, green fluorescent protein (GFP), or chloramphenicol acetyl transferase (CAT)). For example, in the embodiment shown in FIG. **6**C and FIG. **6**D, a scaffold can include a myc tag (e.g., (Myc).sub.3 tag) or a FLAG tag (FLAG).sub.3 tag at the C-terminus.

(144) In some embodiments, a host cell can be engineered to increase acetyl-CoA availability for cannabinoid and cannabinoid precursor biosynthesis. For example, the mitochondrial enzyme isocitrate dehydrogenase-1 (IDH1) can be placed under transient micro-RNA-mediated inducible repression. Since mitochondrial IDH1 is primarily responsible for depletion of the cellular citrate pool, micro-RNA-mediated repression of IDH1 can increase the availability and cytosolic shuttling of citrate for production of acetyl-CoA by ATP citrate lyase. The resulting increase in acetyl-CoA bioavailability can further enhance downstream hexanoyl-CoA and geranyl pyrophosphate titers by improving initial substrate availability for the hexanoyl-CoA and mevalonate pathways. The combinatorial metabolic engineering of acetyl-CoA can mitigate issues related to the siphoning of acetyl-CoA away from the endogenous metabolism of the host cells.

(145) In some embodiments, one or more conventional and/or contemporary gene editing techniques can be used to produce recombinant hosts. For example, clustered, regularly interspaced, short palindromic repeat (CRISPR) technology can be used to modify expression of an endogenous nucleic acid. The CRISPR/Cas system includes components of a prokaryotic adaptive immune system that is functionally analogous to eukaryotic RNA interference, using RNA base pairing to direct DNA or RNA cleavage. The Cas9 protein functions as an endonuclease, and CRISPR RNA (crRNA) and trans-activating RNA (tracrRNA) sequences complex with the Cas9 enzyme and direct it to a target DNA sequence (Makarova et al., *Nat Rev Microbiol* 9(6):467-477, 2011). The modification of a single targeting RNA can be sufficient to alter the nucleotide target of a Cas protein. In some cases, crRNA and tracrRNA can be engineered as a single cr/tracrRNA hybrid (also referred to as a “guide RNA” or “gRNA”) to direct Cas9 cleavage activity (Jinek et al., *Science,* 337(6096):816-821, 2012). The CRISPR/Cas system can be used in a variety of prokaryotic and eukaryotic organisms (see, e.g., Jiang et al., *Nat Biotechnol,* 31(3):233-239, 2013; Dicarlo et al., *Nucleic Acids Res,* doi:10.1093/nar/gkt135, 2013; Cong et al., *Science,* 339(6121):819-823, 2013; Mali et al., *Science,* 339(6121):823-826, 2013; Cho et al., *Nat Biotechnol,* 31(3):230-232, 2013; and Hwang et al., *Nat Biotechnol,* 31(3):227-229, 2013).

(146) Another gene-editing technique can include a sequence-specific nuclease created by fusing transcription activator-like effectors (TALEs) to, for example, the catalytic domain of the Fold endonuclease. Both native and custom TALE-nuclease (“TALEN”) fusions direct DNA double-strand breaks to specific, targeted sites. See, for example, Christian, et al., *Genetics* 186: 757-761 (2010) and U.S. Patent Publication No. 20110145940.

(147) Other suitable gene insertion techniques include the use of retroviral vectors and biolistic particle gene delivery systems (colloquially known as “gene guns”).

(148) Methods of identifying and/or selecting host cells that contain exogenous nucleic acid or a modified endogenous nucleic acid are well known to those skilled in the art. Such methods include, without limitation, the introduction and expression of a negative selection marker such as an antibiotic resistance gene, PCR, and nucleic acid hybridization techniques such as Northern and Southern analyses. In some cases, immunohistochemistry and biochemical techniques can be used

to determine if a microorganism contains a particular nucleic acid by detecting the expression of the encoded enzymatic polypeptide encoded by that particular nucleic acid molecule. For example, an antibody having specificity for an encoded enzyme can be used to determine whether or not a particular cell contains that encoded enzyme. Further, biochemical techniques can be used to determine if a cell contains a particular nucleic acid molecule encoding an enzymatic polypeptide by detecting an organic product produced as a result of the expression of the enzymatic polypeptide.

(149) This document also provides isolated nucleic acids molecules. The term “isolated” as used herein with reference to nucleic acid refers to a naturally-occurring nucleic acid that is not immediately contiguous with both of the sequences with which it is immediately contiguous (one on the 5′ end and one on the 3′ end) in the naturally-occurring genome of the organism from which it is derived. For example, an isolated nucleic acid can be, without limitation, a recombinant DNA molecule of any length, provided one of the nucleic acid sequences normally found immediately flanking that recombinant DNA molecule in a naturally-occurring genome is removed or absent. Thus, an isolated nucleic acid includes, without limitation, a recombinant DNA that exists as a separate molecule (e.g., a cDNA or a genomic DNA fragment produced by PCR or restriction endonuclease treatment) independent of other sequences as well as recombinant DNA that is incorporated into a vector, an autonomously replicating plasmid, a virus (e.g., a retrovirus, adenovirus, or herpes virus), or into the genomic DNA of a prokaryote or eukaryote. In addition, an isolated nucleic acid can include a recombinant DNA molecule that is part of a hybrid or fusion nucleic acid sequence.

(150) The term “isolated” as used herein with reference to nucleic acid also includes any non-naturally-occurring nucleic acid since non-naturally-occurring nucleic acid sequences are not found in nature and do not have immediately contiguous sequences in a naturally-occurring genome. For example, non-naturally-occurring nucleic acid such as an engineered nucleic acid is considered to be isolated nucleic acid. Engineered nucleic acid can be made using common molecular cloning or chemical nucleic acid synthesis techniques. Isolated non-naturally-occurring nucleic acid can be independent of other sequences, or incorporated into a vector, an autonomously replicating plasmid, a virus (e.g., a retrovirus, adenovirus, or herpes virus), or the genomic DNA of a prokaryote or eukaryote. In addition, a non-naturally-occurring nucleic acid can include a nucleic acid molecule that is part of a hybrid or fusion nucleic acid sequence.

(151) It will be apparent to those of skill in the art that a nucleic acid existing among hundreds to millions of other nucleic acid molecules within, for example, cDNA or genomic libraries, or gel slices containing a genomic DNA restriction digest is not to be considered an isolated nucleic acid.

(152) In some embodiments, the production of one or more cannabinoids can be performed in vitro using the scaffold and immobilized enzymes described herein, using a lysate (e.g., a buffered cell lysate) from a recombinant host cell as a source of the scaffold and enzymes, using a plurality of lysates from different host cells as the source of the scaffold and enzymes, or using an acellular reaction buffer such as a synthetic reaction buffer. For example, following co-immunoprecipitation of C-terminal Myc/Flag-tagged enzyme-bound scaffolds, scaffold-enzyme complexes can be maintained in a citrate-supplemented and/or glucose-supplemented (or other carbon source-supplemented) reaction buffer which allows in-vitro scaffolded cannabinoid biosynthesis.

(153) Producing Cannabinoids Using a Recombinant Host

(154) Typically, one or more cannabinoids can be produced by providing a recombinant host such as a recombinant microorganism and culturing the microorganism with a culture medium. In general, the culture media and/or culture conditions can be such that the microorganisms grow to an adequate density and produce cannabinoids efficiently. For example, the microorganisms can be subjected to aerobic batch fermentation. In some embodiments, one or more precursors (e.g., citrate, glucose, hexanoic acid, and/or other carbon source and/or malonyl-CoA) are supplemented in the culture medium. In some embodiments, about 30 mg/L to about 10,000 mg/L (e.g., about 100

mg/L to about 5,000 mg/L, about 200 mg/L to about 4,000 mg/L, about 300 mg/L to about 3,000 mg/L, or about 350 mg/L to about 1,000 mg/L) of buffered citrate, pH 6.0 can be added to the culture medium.

(155) For large-scale production processes, any method can be used such as those described elsewhere (Manual of Industrial Microbiology and Biotechnology, 2nd Edition, Editors: A. L. Demain and J. E. Davies, ASM Press; and Principles of Fermentation Technology, P. F. Stanbury and A. Whitaker, Pergamon). Briefly, a large vessel (e.g., a 100 gallon, 200 gallon, 500 gallon, or higher volume vessel) containing an appropriate culture medium is inoculated with a particular microorganism. After inoculation, the microorganism is incubated to allow biomass to be produced. Once a desired biomass or cellular confluency is attained, a portion or all of the broth containing the microorganisms can be transferred to a second vessel. This second vessel can be any size. For example, the second vessel can be larger, smaller, or the same size as the first vessel. Typically, the second vessel is larger than the first such that additional culture medium can be added to the broth from the first vessel. In addition, the culture medium within this second vessel can be the same as, or different from, that used in the first vessel. This system can expand to include an array consisting of any number of individual vessels.

(156) Once transferred, the microorganisms can be incubated to allow for the production of one or more cannabinoids. Once produced, any method can be used to isolate cannabinoids. For example, common separation techniques can be used to remove the biomass from the broth, and common isolation procedures (e.g., extraction such as non-polar extraction with hexane followed by ethyl-acetate), high-performance liquid chromatography (e.g., HPLC with a diode array detector (HPLC-DAD)), gas chromatography-flame ionization detection (GC-FID), or ion-exchange procedures) can be used to obtain the cannabinoids from the biomass.

(157) A host cell described herein can produce one or more cannabinoids at a concentration of at least about 10 mg per L (e.g., at least about 15 mg/L 25 mg/L, 50 mg/L, 75 mg/L, 100 mg/L, 150 mg/L, 200 mg/L, 250 mg/L or more). For example, in some embodiments, total cannabinoids (total of CBG, CBGA, CBD, CBDA, CBC, and CBCA) can be produced at a concentration of at least about 10 mg/L, 15 mg/L, 20 mg/L, 40 mg/L, 60 mg/L, 80 mg/L, or 100 mg/L or more. For example, in some embodiments, total cannabinoids (total of CBG, CBGA, CBD, CBDA, CBC, and CBCA) can be produced at a concentration from about 10 mg/L to about 500 mg/L (e.g., 20 mg/L to 450 mg/L, 40 mg/L to 380 mg/L, 60 mg/L to 280 mg/L, 60 mg/L to 250 mg/L, 60 mg/L to 150 mg/L, 80 mg/L to 400 mg/L, 80 mg/L to 300 mg/L, 80 mg/L to 250 mg/L, 80 mg/L to 200 mg/L, 80 mg/L to 175 mg/L, 90 mg/L to 400 mg/L, 90 mg/L to 300 mg/L, 90 mg/L to 250 mg/L, or 90 mg/L to 150 mg/L). In some embodiments, one or more individual cannabinoids (e.g., one or more of CBG, CBGA, CBD, CBDA, CBC, and CBCA) can be produced at concentrations of at least about 1 mg/L, 2 mg/L, 5 mg/L, 10 mg/L, 15 mg/L, 20 mg/L, 25 mg/L, 30 mg/L, 35 mg/L, 40 mg/L, 45 mg/L, 50 mg/L, 55 mg/L, 60 mg/L, 65 mg/L, 70 mg/L, 75 mg/L, 80 mg/L, 85 mg/L, 90 mg/L, 95 mg/L, 100 mg/L or more. For example, in some embodiments, one or more individual cannabinoids can be produced at a concentration from about 1 mg/L to about 100 mg/L (e.g., 2 to 90 mg/L, 2 to 80 mg/L, 2 to 70 mg/L, 2 to 60 mg/L, 2 to 50 mg/L, 2 to 40 mg/L, 2 to 30 mg/L, 2 to 20 mg/L, 2 to 15 mg/L, 3 to 90 mg/L, 3 to 80 mg/L, 3 to 70 mg/L, 3 to 60 mg/L, 3 to 50 mg/L, 3 to 40 mg/L, 3 to 30 mg/L, 3 to 20 mg/L, 3 to 15 mg/L, 4 to 90 mg/L, 4 to 80 mg/L, 4 to 70 mg/L, 4 to 60 mg/L, 4 to 50 mg/L, 4 to 40 mg/L, 4 to 30 mg/L, 4 to 20 mg/L, or 4 to 15 mg/L).

(158) The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

EXAMPLES

Example 1—General Methods

(159) Enzymatic Constructs

(160) Each enzyme construct is designed to include an interaction domain (ID) which is comprised of two tandem N-terminal or C-terminal ligand-binding motifs which are separated from the given

enzyme and from one another by an amino acid sequence containing flexible GS-rich linkers flanking a rigid α-helical spacer sequence. The motifs comprising the ID of each enzyme specifically bind tandem peptide ligands which form ID-binding sites at discrete locations along a synthetic intracellular polypeptide scaffold. Expression of each enzyme is controlled by a constitutive or inducible promoter. The nucleic acid encoding the enzyme can be codon optimized, e.g., for expression in yeast.

(161) Scaffolding Constructs

(162) ID-binding sites containing tandem peptide ligands that are specific for the tandem scaffold-binding motifs, which comprise the ID of each enzyme, are inserted at discrete positions along an intracellular polypeptide scaffold.

(163) The tandem ligands which comprise each scaffolded ID-binding site are separated from one another by a 36 amino acid residue sequence containing flexible GS-rich linkers flanking a rigid α-helical spacer sequence, while the scaffolded ID-binding sites themselves are separated from one another by a 50 amino acid residue sequence (or any other number of amino acid residues) containing flexible GS-rich linkers flanking a rigid α-helical spacer sequence. Specifically, the scaffold binding sites for each enzyme in the hexanoyl-CoA pathway are positioned (in order of catalysis) proximally to ATP citrate lyase and acetyl-CoA acetyltransferase at the N-terminus of the primary scaffold. Scaffold binding sites for each enzyme in the upper cannabinoid pathway are positioned proximally to (immediately downstream of) the binding sites for the hexanoyl-CoA pathway enzymes. The scaffold binding sites for each enzyme in the mevalonate (or MEP) pathway are positioned (in order of catalysis) proximally to ATP citrate lyase and acetyl-CoA acetyltransferase at the C-terminus of the primary scaffold. The enzyme catalyzing the rate-limiting/committed step in cannabinoid biosynthesis (CBGA synthase, the final enzymatic step in the upper cannabinoid pathway) is located at the intersection of the converging cannabinoid precursor pathways near the scaffold midpoint.

(164) Assessment of Cannabinoidergic Potential by Transient Transfection

(165) Competent yeast and/or green algae cells are transiently transfected with plasmids encoding various permutations of the scaffold and enzymes. To establish baseline cannabinoidergic capacity, cells first undergo transient transfection with the enzymes required for cannabinoid biosynthesis (but not the scaffolds), and biosynthesized cannabinoids are extracted, isolated, and quantified as described below (see “Cannabinoid Extraction, Isolation, and Analytical Characterization”). To measure the improvement in cannabinoidergic capacity conferred by multi-enzymatic scaffolding, a subset of the aforementioned cells is co-transfected with plasmids encoding one or more of the multi-enzymatic scaffolds described herein, and biosynthesized cannabinoids are extracted, isolated, and quantified. The presence of the plasmid DNA is confirmed by PCR, functional gene expression is confirmed by qRT-PCR, protein/polypeptide production is confirmed by Western blotting, and scaffolding of each enzyme is confirmed by co-immunoprecipitation of C-terminal myc/flag-tagged scaffolds followed by Western blot analysis of each co-immunoprecipitated enzyme.

(166) Engineering of Stable Cannabinoidergic Cell Lines

(167) The constructs can be integrated into the genome of host cells such yeast, green algae, or other suitable hosts via stable transfection. Gene integration is confirmed by PCR, functional gene expression is confirmed by qRT-PCR, and protein/polypeptide production is confirmed by Western blotting. Gene expression/protein synthesis is confirmed by comparing both qRT-PCR and Western blot results among samples with and without genetic engineering. To assess the improvement in cannabinoidergic capacity conferred by multi-enzymatic scaffolding for stably engineered cannabinoidergic cell lines, cannabinoid biosynthesis will be compared among cells that are stimulated for enzyme but not scaffold expression and cells that are stimulated for enzyme and scaffold expression.

(168) Validation of Multi-enzymatic Scaffolding

(169) To verify successful multi-enzymatic scaffolding in both transiently transfected and stably engineered cells, a myc-tag (or other immunoprecipitable tag) is inserted at the N-terminal or C-terminal of the polypeptide scaffold(s). Scaffolded enzymes are selectively co-immunoprecipitated by affinity chromatography using anti-myc affinity beads. Western blots are performed to detect and quantify each co-immunoprecipitated enzyme.
(170) Aerobic Fed-batch Fermentation
(171) Stably engineered cannabinoidergic yeast, green algae, or other host cells are grown in bioreactors (or any other vessel) via aerobic batch fermentation (or any other culture technique).
(172) Cannabinoid Extraction, Isolation, and Analytical Characterization
(173) Following sufficient elicitation of cannabinoid biosynthesis, engineered yeast/green algae cells are pelleted by centrifugation and washed with TBS. The supernatant (liquid culture media) is decanted and collected. Following washing with TBS, pelleted cells are resuspended in NaOH adjusted ethanol and lysed by iterative freeze-thawing and ultrasonication. Biosynthesized cannabinoid fermentates are then harvested from both lysates and supernatants via triplicate nonpolar extractions using hexane followed by ethyl-acetate. The resulting organic fractions are pooled and roto-evaporated. High-performance liquid chromatography with a diode array detector (HPLC-DAD) or gas chromatography-flame ionization detection (GC-FID) is then applied for quantitative and qualitative measurement of biosynthesized cannabinoids.
(174) In the following examples, each 48-hour culture was lysed/homogenized by ultrasonication. Ultrasonicated samples were then subjected to triplicate liquid-liquid extractions with ethyl acetate (one volumetric equivalent of ethyl acetate per extraction). Following separation, the ethyl acetate fractions collected from each sample were pooled, and the pooled samples were centrifugally filtered. Ethyl acetate was then removed from each sample in a vacuum oven, and the residual samples were resuspended in 10 mL methanol for analytical characterization. Analytical characterization of all samples was conducted by a licensed, independent, third-party analytical testing facility (Precision Plant Molecules, Denver, CO). HPLC-DAD was utilized for quantitative and qualitative measurement of each parent and derivative cannabinoid as well as the cannabinoid precursor OVA.
Example 2—Synthetic Gene Cassette Assembly/Synthesis, Plasmid Preparation, and Polycistronic Vector Construction
(175) Five synthetic gene cassettes (entitled HCA, GPP, CAN, SCF, and SOL) were constructed for biosynthesizing cannabinoids in heterologous cells or acellular reaction buffers. See, FIG. **2**B. The cassettes collectively encode all scaffold-binding engineered enzymes and the polypeptide scaffolds to which the engineered enzymes can bind.
(176) The HCA gene cassette encoded scaffold-binding engineered enzymes for scaffolded hexanoyl-CoA biosynthesis, namely ACL, atoB, BHBD, ECH, ECR, and bktB, and encoded a soluble HCS for additional hexanoyl-CoA production from hexanoate-supplemented culture media or acellular reaction buffer. See, FIG. **13**A. The GPP gene cassette encoded scaffold-binding engineered enzymes for scaffolded geranyl pyrophosphate (GPP) biosynthesis, namely HMGS, tHMGR, ERG12, ERGS, MVD1, IDI1, and ERG20.sup.WW. See, FIG. **13**B. The CAN gene cassette encoded scaffold-binding engineered enzymes for scaffolded OAC, malonyl-CoA, and CBGA biosynthesis, namely OS and OAC, ACC, and CBGAS, respectively, as well all enzymes for soluble (non-scaffolded) CBDA and CBCA biosynthesis, namely CBDAS and CBCAS, respectively. See, FIG. **13**C. The SCF gene cassette encoded the polypeptide scaffolds for bidirectional scaffolded cannabinoid biosynthesis and scaffolded malonyl-CoA biosynthesis, namely the cannabinoidergic metabolon scaffold (CBSCF) and the malonyl-CoA metabolon scaffold (MCASCF), respectively, as well as additional copies of both ACL and atoB to enhance acetyl-CoA biosynthesis from supplemental and/or endogenous citrate and acetoacetyl-CoA biosynthesis from acetyl-CoA, respectively. See, FIG. **13**D. The SOL gene cassette lacked the polypeptide scaffolds for bidirectional scaffolded cannabinoid biosynthesis and scaffolded

malonyl-CoA biosynthesis (i.e., it was used for soluble cannabinoid biosynthesis) but, analogous to the SCF gene cassette, encoded additional copies of ACL and atoB to enhance acetyl-CoA biosynthesis from supplemental and/or endogenous citrate and acetoacetyl-CoA biosynthesis from acetyl-CoA. See FIG. **13**A for the amino acids sequences of the ACL and atoB engineered enzymes.

(177) Gene cassettes were assembled/synthesized using self-cleaving 2A peptides (P2As) to link multiple codon-optimized (for *S. cerevisiae*) gene sequences assigned to each cassette. To improve P2A cleavage, a GSG linker (comprised of a single serine residue flanked by single glycine residues) was inserted at the interface between each constituent gene sequence and the P2A linker sequence to which it was fused (of the format: gene cassette sequence 1— SG— P2A linker—gene cassette sequence 2— GSG—P2A linker—gene cassette sequence 3— GSG— P2A linker-) and so forth. See, FIGS. **14**A-**14**D for codon-optimized nucleic acid sequences encoding the engineered enzymes and scaffolds. Following assembly, each synthetic gene cassette was inserted into a pCCI-Brick plasmid, resulting in plasmids entitled pHCA, pGPP, pCAN, pSCF, and pSOL as described in Table 3. See, FIGS. **15**A-**15**E for the complete gene cassette inserted into the plasmids. Each of these plasmids then were used to amplify each synthetic gene cassette via standard plasmid prep. Plasmid DNA encoding each complete synthetic gene cassette was cloned into the SpeI/XhoI cloning site of polycistronic yeast auxotrophic selection vectors, resulting in vectors entitled vHCA, vGPP, vCAN, vSCF, and vSOL as described in Table 3, to allow iterative antibiotic/auxotrophic selection of only those cells that were transformants of one or more such polycistronic vector(s).

(178) TABLE-US-00008 TABLE 3 HCA Gene Cassette Cassette pCCI-Brick Yeast Yeast Gene ID Position #1 ID Vector Vector ID ACL 1 pHCA pESC-TRP vHCA atoB 2 BHBD 3 ECH 4 ECR 5 bktB 6 HCS 7 MVA Gene Cassette Cassette pCCI-Brick Yeast Yeast Gene ID Position #2 ID Vector Vector ID HMGS 1 pGPP pESC-LEU vGPP tHMGR 2 ERG12 3 ERG8 4 MVD1 5 IDI1 6 ERG20.sup.WW 7 CAN Gene Cassette Cassette pCCI-Brick Yeast Yeast Gene ID Position #3 ID Vector Vector ID OS 1 pCAN pESC-HIS vCAN OAC 2 CBGAS 3 CBDAS 4 CBCAS 5 ACC 6 SCFLD Gene Cassette Cassette pCCI-Brick Yeast Yeast Gene ID Position #4 ID Vector Vector ID CBSCF 1 pSCF pESC-URA #1 vSCF MCASCF 2 ACL 3 atoB 4 NSCFLD Gene Cassette Cassette pCCI-Brick Yeast Yeast Gene ID Position #5 ID Vector Vector ID ACL 1 pSOL pESC-URA #2 vSOL atoB 2

(179) The genes assigned to each synthetic gene cassette as well as the plasmids and vectors into which each synthetic gene cassette was inserted are listed in Table 3, the amino acid sequences encoded by each synthetic gene cassette are provided in FIGS. **13**A-**13**D, the codon-optimized nucleotide sequence fragments comprising each synthetic gene cassette are detailed in FIGS. **14**A-**14**D, the complete nucleotide sequences of each fully-assembled synthetic gene cassette (the complete insert sequences for each plasmid and expression vector) are provided in FIGS. **15**A-**15**E, a general map of pCCI-Brick plasmids is shown in FIG. **16**, and a general map of a polycistronic yeast auxotrophic selection vector is shown in FIG. **17**.

Example 3—Engineering of Cannabinoidergic Cells

(180) To engineer a novel heterologous pathway for the biosynthesis of cannabinoids from citrate, and to evaluate the impacts of bidirectional multi-enzymatic scaffolding thereon, competent *S. cerevisiae* cells were sequentially/iteratively transformed with, and auxotrophically selected for, expression of vHCA, vGPP, vCAN, and either vSCF (for scaffolded cannabinoid biosynthesis) or vSOL (for non-scaffolded/soluble cannabinoid biosynthesis) constructs.

(181) All vector transformation and auxotrophic selection procedures were conducted as follows. An aliquot of an overnight *S. cerevisiae* culture was inoculated into 100 mL YPD media (10 g/L yeast nitrogen base, 20 g/L peptone, and 20 g/L D-(+)-glucose) to OD.sub.600mn=0.3 (stationary phase) and grown to OD.sub.600mn=1.6 in an orbital shaker at 30° C. and 225 RPM. Cells then were harvested by centrifugation at 3000×g for 3 minutes followed by aspiration of media. The

harvested cell pellet was next washed 2× with 50 mL chilled nuclease-free water and 1× with 50 mL chilled electroporation buffer (1M sorbitol/1 mM CaCl.sub.2). Washed cells were conditioned by incubation for 30 minutes in 20 mL 0.1M LiAc/10 mM DTT in an orbital shaker at 30° C. and 225 RPM, harvested, washed 1× with 50 mL electroporation buffer, harvested, and resuspended in 100 μL electroporation buffer. The resuspended cells were transformed with a quantity of vector containing 3 μg of the target DNA insert (calculated using the vector-insert ratio for each vector) by electroporation at 2.5 kV and 25 g. To the electroporated cell suspension was then added 8 mL of YPD media containing 1M sorbitol, and the resulting suspension was incubated for one hour in an orbital shaker at 30° C. and 225 RPM. To isolate target transformants by auxotrophic selection, cells were harvested, resuspended in the appropriate yeast nitrogen base (YNB) dropout (selection) media as subsequently described for each iterative transformation step, transferred to a baffled culture flask, and incubated overnight in an orbital shaker at 30° C. and 225 RPM. The transformation and selection protocols were utilized sequentially for each assigned vector.

(182) Applying the aforementioned approach, an initial culture of electrocompetent *S. cerevisiae* cells was first transformed with vHCA, which encodes scaffold-binding engineered enzymes required for biosynthesis of HCA from citrate. Cells transformed with vHCA (designated yHCA) were selected for by resuspension and incubation in tryptophan-deficient YNB media. Selected yHCA cells (i.e., cells that grew in tryptophan-deficient YNB media) were next transformed with vGPP, which encodes scaffold-binding engineered enzymes required for biosynthesis of GPP from citrate. Cells co-transformed with vHCA and vGPP (designated yHCAGPP) were selected for by resuspension and incubation in tryptophan- and leucine-deficient YNB media. Selected yHCAGPP cells (i.e., cells that grew in tryptophan- and leucine-deficient YNB media) were then transformed with vCAN, which encodes scaffold-binding engineered enzymes required for biosynthesis of malonyl-CoA from citrate, olivetol from HCA and malonyl-CoA, OVA (olivetoic acid) from olivetol, and CBGA from OVA and GPP as well as soluble enzymes required for biosynthesis of CBDA and CBCA from CBGA). Cells co-transformed with vHCA, vGPP, and vCAN (designated yCB.sub.Parent) were selected for by resuspension and incubation in tryptophan-, leucine-, and histidine-deficient YNB media.

(183) The yCB.sub.Parent culture containing cells that grew in tryptophan-, leucine-, and histidine-deficient YNB media then was split into two separate cultures. The first of the split yCB.sub.Parent cultures was transformed with vSCF, which encodes CBSCF (cannabinoidergic metabolon scaffold) and MCASCF (malonyl-CoA metabolon scaffold) as well as additional copies of ACL and atoB. Cells co-transformed with vHCA, vGPP, vCAN, and vSCF (designated yCB.sub.SCF) were selected for by resuspension and incubation in tryptophan-, leucine-, histidine-, and uracil-deficient YNB media. The second of the split yCB.sub.Parent cultures was transformed with vSOL, which encodes additional copies of ACL and atoB but lacks both CB SCF and MCASCF. Cells co-transformed with vHCA, vGPP, vCAN, and vSOL (designated yCB.sub.SOL) were also selected for by resuspension and incubation in tryptophan-, leucine-, histidine-, and uracil-deficient YNB media.

(184) To quantify the improvement in cannabinoidergic capacity conferred by multi-enzymatic scaffolding, cannabinoid titers were compared between triplicate yCB.sub.SOL and yCB.sub.SCF cultures grown in 100 mL YPD media for 48 hours at 30° C. and 400 RPM in an incubator-shaker. To compare the proliferation rates of yCB.sub.SOL and yCB.sub.SCF, each culture was initially diluted to OD.sub.600mn=0.3, and OD.sub.600 nm measurements were recorded in 12-hour intervals thereafter. Proliferation curves are depicted in FIG. **18**. The extra sum-of-squares F-test indicated that the proliferation curves of yCB.sub.SCF and yCBSOL cultures did not significantly differ for any parameter over the 48-hour incubation period, indicating that scaffolding does not impact cellular proliferation.

(185) Total cannabinoid titers, parent (carboxylated) cannabinoid (CBGA, CBDA, and CBCA) titers, derivative (decarboxylated) cannabinoid (CBG, CBD, and CBC) titers, and cannabinoid

precursor (OVA) titers were measured. As shown in FIGS. **19**A-**19**E, mixed ANOVA detected main effects of strain ($F_{1,4}=943.8$; $p<0.0001$) and analyte (cannabinoid and cannabinoid precursor) titers ($F_{10,40}=216.4$; $p<0.0001$) and a significant strain x analyte interaction ($F_{10,40}=131.4$; $p<0.0001$). Relative to yCBSOL cultures, yCBSCF cultures exhibited increased total cannabinoid ($p<0.0001$), OVA precursor ($p<0.0001$), CBG(A) ($p<0.0001$), CBD(A) ($p<0.0001$), CBC(A) ($p<0.0001$), CBGA ($p<0.0001$), CBDA ($p<0.0001$), CBCA ($p<0.0001$), CBG ($p<0.0001$), CBD ($p<0.01$), and CBC ($p<0.001$) titers.

Example 4—Impacts of Citrate and Hexanoate Supplementation on Scaffolded and Soluble Cannabinoid Biosynthesis

(186) To evaluate the impacts of culture media supplementation with citrate and hexanoate precursors, cannabinoid titers were compared between triplicate yCB.sub.SOL and yCB.sub.SCF cultures grown in 100 mL YPD media containing 300 mg/L of either buffered citrate (pH 6.0) or hexanoate for 48 hours at 30° C. and 400 RPM in an orbital shaker. All cultures were initially diluted to $OD_{600nm}=0.3$. Cannabinoid titers for cultures grown in YPD media, citrate-supplemented YPD media, and hexanoate-supplemented YPD media were assessed and analyzed by ANOVA. As shown in FIG. **20**, mixed ANOVA detected main effects of strain ($F_{1,4}=457.5$; $p<0.0001$) and culture media supplementation ($F_{2,8=3}$ 12.5; $p<0.0001$) and a significant strain x culture media supplementation interaction ($F_{2,8}=289.6$; $p<0.0001$). Compared to basal media cultures, yCBSCF but not yCBSOL cultures exhibited increased total cannabinoid titers when cultured in media supplemented with 300 mg/L citrate ($p<0.0001$). Neither yCBSCF nor yCBSOL cultures differed in total cannabinoid titers relative to basal media when cultured in media supplemented with 300 mg/L hexanoate. For all measures, n=3 biological replicates for yCBSCF and yCBSOL cultures. Moreover, relative to yCBSOL cultures, yCBSCF cultures exhibited increased total cannabinoid titers when cultured in basal media ($p<0.0001$, data also reported in FIG. **19**) as well as media supplemented with 300 mg/L citrate ($p<0.0001$) and hexanoate ($p<0.0001$).

(187) To delineate concentration-response relationships for the supplementation of culture media with citrate, cannabinoid titers were compared between triplicate yCB.sub.SOL and yCB.sub.SCF cultures grown in 100 mL YPD media containing 0, 10, 30, 100, 300, 1000, 3000, and 10000 mg/L buffered citrate (pH 6.0) for 48 hours at 30° C. and 400 RPM in an orbital shaker. All cultures were initially diluted to $OD_{600nm}=0.3$. Following quantification, asymmetric sigmoidal (five-parameter) logistic regressions were computed to fit concentration-response curves, from which were derived estimates of the maximal cannabinoid titer (CB.sub.Max) and citrate EC.sub.50 for cannabinoid biosynthesis in yCB.sub.SOL and yCB.sub.SCF cultures. Concentration-response curves, CB.sub.Max estimates, and citrate EC.sub.50 estimates are depicted in FIGS. **21**A and **21**B. Mixed ANOVA detected main effects of strain ($F_{1,8}=69.9$; $p<0.0001$) and parameter ($F_{1,8}=66.7$; $p<0.0001$) and a significant strain x parameter interaction ($F_{1,8}=5.3$; $p<0.05$) for concentration-response parameter estimates (CB.sub.Max and citrate EC.sub.50). Compared to yCB.sub.SOL cultures, yCB.sub.SCF cultures exhibited markedly increased CB.sub.Max ($p<0.0001$) and citrate EC.sub.50 ($p<0.001$) estimates.

OTHER EMBODIMENTS

(188) It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

Claims

1. A host cell capable of producing one or more cannabinoids, said host cell comprising: (a) a first exogenous nucleic acid encoding a first polypeptide having CBGA synthase activity and

comprising a first heterologous interaction domain, (b) a second exogenous nucleic acid encoding a second polypeptide having olivetolic acid cyclase activity and comprising a second heterologous interaction domain, (c) a third exogenous nucleic acid encoding a third polypeptide having olivetol synthase activity and comprising a third heterologous interaction domain, (d) a fourth exogenous nucleic acid encoding a fourth polypeptide having trans-2-enoyl-CoA reductase activity and comprising a fourth heterologous interaction domain, (e) a fifth exogenous nucleic acid encoding a fifth polypeptide having enoyl-CoA hydratase activity and comprising a fifth heterologous interaction domain, (f) a sixth exogenous nucleic acid encoding a sixth polypeptide having 3-hydroxybutyryl-CoA dehydrogenase activity and comprising a sixth heterologous interaction domain, (g) a seventh exogenous nucleic acid encoding a seventh polypeptide having acetyl-CoA acetyltransferase activity and comprising a seventh heterologous interaction domain, (h) an eighth exogenous nucleic acid encoding an eighth polypeptide having ATP citrate lyase activity and comprising an eighth heterologous interaction domain, (i) a ninth exogenous nucleic acid encoding a ninth polypeptide having geranyl pyrophosphate synthase activity and comprising a ninth heterologous interaction domain, (j) a tenth exogenous nucleic acid encoding a tenth polypeptide having isopentyl-diphosphate isomerase activity and comprising a tenth heterologous interaction domain, (k) an eleventh exogenous nucleic acid encoding an eleventh polypeptide having diphospho-mevalonate decarboxylase activity and comprising an eleventh heterologous interaction domain, (l) a twelfth exogenous nucleic acid encoding a twelfth polypeptide having phosphomevalonate kinase activity and comprising a twelfth heterologous interaction domain, (m) a thirteenth exogenous nucleic acid encoding a thirteenth polypeptide having mevalonate kinase activity and comprising a thirteenth heterologous interaction domain, (n) a fourteenth exogenous nucleic acid encoding a fourteenth polypeptide having HMG-CoA reductase activity and comprising a fourteenth heterologous interaction domain, (o) a fifteenth exogenous nucleic acid encoding a fifteenth polypeptide having HMG-CoA synthase activity and comprising a fifteenth heterologous interaction domain, and (p) a sixteenth exogenous nucleic acid encoding a polypeptide scaffold comprising a peptide ligand for each of said first to fifteenth heterologous interaction domains, wherein each of said first to fifteenth heterologous interaction domains is different, wherein each peptide ligand for each of said first to fifteenth heterologous interaction domains is different, wherein said polypeptide scaffold comprises, in an order extending in a first direction away from said peptide ligand for said first heterologous interaction domain, (1) said peptide ligand for said second heterologous interaction domain, (2) said peptide ligand for said third heterologous interaction domain, (3) said peptide ligand for said fourth heterologous interaction domain, (4) said peptide ligand for said fifth heterologous interaction domain, (5) said peptide ligand for said sixth heterologous interaction domain, (6) said peptide ligand for said seventh heterologous interaction domain, and (7) said peptide ligand for said eighth heterologous interaction domain, and wherein said polypeptide scaffold comprises, in an order extending in the other direction away from said peptide ligand for said first heterologous interaction domain, (1) said peptide ligand for said ninth heterologous interaction domain, (2) said peptide ligand for said tenth heterologous interaction domain, (3) said peptide ligand for said eleventh heterologous interaction domain, (4) said peptide ligand for said twelfth heterologous interaction domain, (5) said peptide ligand for said thirteenth heterologous interaction domain, (6) said peptide ligand for said fourteenth heterologous interaction domain, (7) said peptide ligand for said fifteenth heterologous interaction domain, (8) said peptide ligand for said seventh heterologous interaction domain, and (9) said peptide ligand for said eighth heterologous interaction domain.

2. The host cell of claim 1, further comprising (q) a seventeenth exogenous nucleic acid encoding an acetyl-CoA carboxylase and comprising a seventeenth heterologous interaction domain, and (r) an eighteenth exogenous nucleic acid encoding a polypeptide scaffold comprising a peptide ligand for each of said eighth and seventeenth heterologous interaction domains.

3. The host cell of claim 1, wherein said host cell further comprises an exogenous nucleic acid

encoding a cannabidiolic acid synthase and a cannabichromenic acid synthase.
4. The host cell of claim 1, wherein said host cell further comprises an exogenous cannabidiolic acid synthase.
5. The host cell of claim 1, wherein said host cell further comprises an exogenous cannabichromenic acid synthase.
6. The host cell of claim 1, wherein said host cell is a bacterial or a yeast host cell.
7. The host cell of claim 6, wherein said bacterial cell is selected from the group consisting of *Escherichia coli, Bacillus, Brevibacterium, Streptomyces,* and *Pseudomonas* cells.
8. The host cell of claim 6, wherein said yeast cell is selected from the group consisting of *Pichia pastoris, Saccharomyces cerevisiae, Yarrowia lipolytica, Kluyveromyces marxianus,* and Komagataella phaffii cells.
9. The host cell of claim 1, wherein said host cell is an algae or a plant cell.
10. The host cell of claim 9, wherein said algae is *Dunaliella* sp., *Chlorella variabilis, Euglena mutabilis,* or *Chlamydomonas reinhardtii* cells.
11. The host cell of claim 9, wherein said plant cell is a *Cannabis* or tobacco cell.
12. The host cell of claim 1, wherein each of said polypeptides is of the formula: enzyme-linker1-spacer-linker2-motifi-linker3-motif2, wherein linker1, linker2, and linker3 are the same or different, wherein motif1 and motif2 are the same or different, and wherein motif1 and motif2 form said heterologous interaction domain.
13. The host cell of claim 12, wherein said scaffold polypeptide comprises a linker between each adjacent peptide ligand.
14. The host cell of claim 13, wherein said scaffold polypeptide is tagged with a MYC tag, FLAG tag, or HA tag.
15. The host cell of claim 12, wherein said linker is a flexible GS-rich sequence flanking a rigid α-helical moiety.
16. The host cell of claim 12, wherein said spacer is the cTPR6 spacer.
17. The host cell of claim 1, wherein a constitutive promoter is operably linked to one or more of said exogenous nucleic acids encoding said polypeptides or to said sixteenth exogenous nucleic acid encoding said polypeptide scaffold.
18. The host cell of claim 1, wherein a first constitutive promoter is operably linked to one or more of said exogenous nucleic acids encoding said polypeptides and a second constitutive promoter is operably linked to said sixteenth exogenous nucleic acid encoding said polypeptide scaffold.
19. The host cell of claim 18, wherein said constitutive promoter used to express said polypeptide scaffold has weaker constitutive activity level than said constitutive promoter used to express said polypeptides.
20. The host cell of claim 1, wherein each said exogenous nucleic acid comprises an inducible promoter operably linked to the sequence encoding said polypeptide or said polypeptide scaffold.
21. The host cell of claim 20, wherein said promoter is the GAL1-10 promoter.