



US012385072B2

(12) **United States Patent**
Buck(10) **Patent No.:** US 12,385,072 B2
(45) **Date of Patent:** *Aug. 12, 2025(54) **BIDIRECTIONAL MULTI-ENZYMATIC SCAFFOLDS FOR BIOSYNTHESIZING CANNABINOIDS**(71) Applicant: **Khona Scientific Holdings, Inc.**, Lone Tree, CO (US)(72) Inventor: **Jordan Buck**, Boulder, CO (US)(73) Assignee: **Khona Scientific Holdings, Inc.**, Lone Tree, CO (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 349 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: 17/962,229

(22) Filed: Oct. 7, 2022

(65) **Prior Publication Data**

US 2023/0265465 A1 Aug. 24, 2023

Related U.S. Application Data

(63) Continuation of application No. 16/694,417, filed on Nov. 25, 2019, now Pat. No. 11,525,148.

(60) Provisional application No. 62/836,265, filed on Apr. 19, 2019, provisional application No. 62/771,839, filed on Nov. 27, 2018.

(51) **Int. Cl.**

C12P 7/42 (2006.01)
C12N 15/70 (2006.01)
C12N 15/74 (2006.01)
C12N 15/81 (2006.01)
C12N 15/82 (2006.01)

(52) **U.S. Cl.**

CPC **C12P 7/42** (2013.01); **C12N 15/70** (2013.01); **C12N 15/74** (2013.01); **C12N 15/81** (2013.01); **C12N 15/8222** (2013.01); **C12N 2330/51** (2013.01); **C12N 2800/40** (2013.01); **C12Y 101/01034** (2013.01); **C12Y 101/01157** (2013.01); **C12Y 103/01038** (2013.01); **C12Y 103/03** (2013.01); **C12Y 121/03007** (2015.07); **C12Y 121/03008** (2015.07); **C12Y 203/01009** (2013.01); **C12Y 203/01016** (2013.01); **C12Y 203/01206** (2015.07); **C12Y 203/03008** (2013.01); **C12Y 203/0301** (2013.01); **C12Y 205/01001** (2013.01); **C12Y 207/01036** (2013.01); **C12Y 207/04002** (2013.01); **C12Y 401/01033** (2013.01); **C12Y 402/01017** (2013.01); **C12Y 404/01026** (2015.07); **C12Y 503/03002** (2013.01); **C12Y 604/01002** (2013.01)

(58) **Field of Classification Search**

None

See application file for complete search history.

(56)

References Cited

U.S. PATENT DOCUMENTS

8,697,654 B2	4/2014	Cheng et al.
8,884,100 B2	11/2014	Page et al.
9,822,384 B2	11/2017	Poulos et al.
9,856,460 B2	1/2018	Dueber et al.
11,525,148 B2 *	12/2022	Buck C12N 15/74
2003/0143562 A1	7/2003	Anderson et al.
2005/0204419 A1	9/2005	Helgeson et al.
2011/0008829 A1	1/2011	Dueber et al.
2011/0145940 A1	6/2011	Voytas et al.
2012/0144523 A1	6/2012	Page et al.
2013/0130347 A1	5/2013	Delisa et al.
2013/0164808 A1	6/2013	McAuliffe et al.
2014/0370595 A1	12/2014	Dueber et al.
2016/0010126 A1	1/2016	Poulos et al.
2017/0139496 A1	5/2017	Kang et al.
2017/0166950 A1	6/2017	Wolkowicz
2019/0055314 A1	2/2019	Luo et al.

FOREIGN PATENT DOCUMENTS

CA 2644548	9/2007
CN 103360471	10/2013
CN 106414747	2/2017
WO WO 2014113738	7/2014
WO WO 2015196275	12/2015
WO WO 2016010827	1/2016
WO WO 2017139496	8/2017
WO WO 2018200888	11/2018

OTHER PUBLICATIONS

Chichili et al., "Linkers in the structural biology of protein-protein interactions," *Protein Sci.*, Feb. 2013, 22(2):153-167.Andre et al. (2016) Cannabissativa: The Plant of the Thousand and One Molecules, *Frontiers Plant Sci.*, vol. 7, article 19, pp. 1-15.Becker et al., "High-Efficiency Transformation of Yeast by Electroporation," *Meth. Enzymology*, 1991, 194:182-187.Carvalho et al., "Designing microorganisms for heterologous biosynthesis of cannabinoids," *FEMS Yeast Research*, Jun. 2017, 17(4):fox037, 11 pages.Chen et al., "Fusion protein linkers: Property, design and functionality," *Adv. Drug Deliv. Reviews*, Oct. 2013, 65(10):1357-1369.Cho et al., "Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease," *Nat. Biotechnology*, Mar. 2013, 31(3):230-232.Christian et al., "Targeting DNA Double-Strand Breaks with TAL Effector Nucleases," *Genetics*, Oct. 1, 2010, 186(2):757-761.Cong et al., "Multiplex Genome Engineering Using CRISPR/Cas Systems," *Science*, Feb. 15, 2013, 339(6121):819-823.Dicarlo et al., "Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems," *Nucleic Acids Research*, Apr. 2013, 41(7):4336-4343.

(Continued)

Primary Examiner — Christian L Fronda

(74) Attorney, Agent, or Firm — Fish & Richardson P.C.

(57) **ABSTRACT**

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

21 Claims, 55 Drawing Sheets

Specification includes a Sequence Listing.

(56)

References Cited**OTHER PUBLICATIONS**

- Dinkel et al., "The eukaryotic linear motif resource ELM: 10 years and counting," *Nucleic Acids Research*, Jan. 2014, 42(Database issue):D259-D266.
- Dueber et al., "Synthetic protein scaffolds provide modular control over metabolic flux," *Nat. Biotechnology*, Aug. 2009, 27(8):753-759.
- Durrens et al., "Expression of the avian gag-myc oncogene in *Saccharomyces cerevisiae*," *Curr. Genetics*, Jul. 1990, 18(1):7-12.
- EP Extended Search Report in European Appln. No. 19888674.9, dated Jul. 22, 2022, 6 pages.
- 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.
- GenBank Accession No. AAA52679.1, "3-hydroxy-3-methylglutaryl coenzyme A reductase [*Homo sapiens*]," dated Nov. 8, 1994, 2 pages.
- GenBank Accession No. AAA62411.1, "3-hydroxy-3-methylglutaryl coenzyme A synthase [*Homo sapiens*]," dated Feb. 24, 1995, 1 page.
- GenBank Accession No. AAC74463.1, "ATP citrate-lyase [*Rattus norvegicus*]," dated Aug. 18, 1995, 1 page.
- GenBank Accession No. AAC49920.1, "isopentenyl diphosphate:dimethylallyl diphosphate isomerase [*Arabidopsis thaliana*]," dated Feb. 24, 1998, 1 page.
- GenBank Accession No. AAC50440.1, "mevalonate pyrophosphate decarboxylase [*Homo sapiens*]," dated Apr. 16, 1996, 1 page.
- GenBank Accession No. AAC67348.1, "mevalonate diphosphate decarboxylase [*Arabidopsis thaliana*]," dated Mar. 11, 2002, 1 page.
- GenBank Accession No. AAD31719.1, "mevalonate kinase [*Arabidopsis thaliana*]," dated May 10, 2000, 1 page.
- GenBank Accession No. AAFA82407.1, "mevalonate kinase [*Homo sapiens*]," dated Jun 10, 2016, 1 page.
- GenBank Accession No. AAH06089.1, "Phosphomevalonate kinase [*Homo sapiens*]," dated Oct. 19, 2016, 2 pages.
- GenBank Accession No. AAH10004.1, "Farnesyl diphosphate synthase (farnesyl pyrophosphate synthetase, dimethylallyltransferase, geranyltransferase) [*Homo sapiens*]," dated Jul. 15, 2006, 2 pages.
- GenBank Accession No. AAH30985.1, "Malonyl CoA:ACP acyltransferase (mitochondrial) [*Homo sapiens*]," dated Sep. 8, 2006, 2 pages.
- GenBank Accession No. AAH31149.1, "Crk protein [*Mus musculus*]," dated Oct. 7, 2003, 2 pages.
- GenBank Accession No. AAK56081.1, "ATP citrate lyase [*Mus musculus*]," dated Feb. 21, 2002, 1 page.
- GenBank Accession No. AAK79760.1, "1-deoxy-D-xylulose 5-phosphate reductoisomerase [Clostridium acetobutylicum ATCC 824]," dated Jan. 30, 2014, 1 page.
- GenBank Accession No. AAK80036.1, "Deoxysxylulose-5-phosphate synthase [Clostridium acetobutylicum ATCC 824]," dated Jan. 30, 2014, 1 page.
- GenBank Accession No. AAK80816.1, "Acetyl-CoA acetyltransferase [Clostridium acetobutylicum ATCC 824]," dated Jan. 30, 2014, 1 page.
- GenBank Accession No. AAK80844.1, "Isopentenyl monophosphate kinase, IPK [Clostridium acetobutylicum ATCC 824]," dated Jan. 30, 2014, 1 page.
- GenBank Accession No. AAK81121.1, "4-diphosphocytidyl-2-methylerithritol synthase (Sugar Nucleotide Phosphorylase family) [Clostridium acetobutylicum ATCC 824]," dated Jan. 30, 2014, 1 page.
- GenBank Accession No. AAM61343.1, "1-deoxy-D-xylulose 5-phosphate reductoisomerase DXR [*Arabidopsis thaliana*]," dated Jan. 27, 2006, 1 page.
- GenBank Accession No. AAM67058.1, "acetoacyl-CoA-thiolase [*Arabidopsis thaliana*]," dated Jan. 27, 2006, 1 page.
- GenBank Accession No. AAN17431.1, "putative 3-hydroxybutyryl-CoA dehydrogenase [*Arabidopsis thaliana*]," dated Sep. 24, 2002, 2 pages.
- GenBank Accession No. AAN81487.1, "1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase [*Escherichia coli* CFT073]," dated Jan. 31, 2014, 1 page.
- GenBank Accession No. AAP35407.1, "isopentenyl-diphosphate delta isomerase [*Homo sapiens*]," dated May 13, 2003, 1 page.
- GenBank Accession No. AAP86010.1, "putative enoyl-(ACP) reductase (plasmid) [Cupriavidus necator H16]," dated Jul. 25, 2016, 1 page.
- GenBank Accession No. AAP94122.1, "acetyl-CoA carboxylase 1 [*Homo sapiens*]," dated Jul. 27, 2003, 2 pages.
- GenBank Accession No. AAS11086.1, "acetyl-CoA carboxylase, biotin carboxylase [Treponema denticola ATCC 35405]," dated Jan. 31, 2014, 2 pages.
- GenBank Accession No. AAS11105.1, "3-hydroxyacyl-CoA dehydrogenase, putative [Treponema denticola ATCC 35405]," dated Jan. 31, 2014, 2 pages.
- GenBank Accession No. AAS11855.1, "4-diphosphocytidyl-2C-methyl-D-erythritol kinase [Treponema denticola ATCC 35405]," dated Jan. 31, 2014, 2 pages.
- GenBank Accession No. AAS12424.1, "1-deoxy-D-xylulose-5-phosphate synthase [Treponema denticola ATCC 35405]," dated Jan. 31, 2014, 2 pages.
- GenBank Accession No. AAS12810.1, "2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase [Treponema denticola ATCC 35405]," dated Jan. 31, 2014, 2 pages.
- GenBank Accession No. AAS12811.1, "2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase [Treponema denticola ATCC 35405]," dated Jan. 31, 2014, 2 pages.
- GenBank Accession No. AAS12860.1, "1-deoxy-D-xylulose 5-phosphate reductoisomerase [Treponema denticola ATCC 35405]," dated Jan. 31, 2014, 2 pages.
- GenBank Accession No. AC157384.3, "Bos taurus clone CH240-61D19, Working Draft Sequence, 6 unordered pieces," dated Jul. 11, 2008, 46 pages.
- GenBank Accession No. ACJ56139.1, "Geranyltransferase(Farnesyl-diphosphate synthase) [Acinetobacter baumannii AB307-0294]," dated Jan. 31, 2014, 1 page.
- GenBank Accession No. ACJ57023.1, "Enoyl-CoA hydratase [Acinetobacter baumannii AB307-0294]," dated Jan. 31, 2014, 1 page.
- 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.
- GenBank Accession No. ACJ59227.1, "2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase [Acinetobacter baumannii AB307-0294]," dated Jan. 31, 2014, 1 page.
- GenBank Accession No. ACS85236.1, "malonyl CoA-acyl carrier protein transacylase [Dickeya paradisiaca Ech703]," dated Oct. 25, 2017, 1 page.
- GenBank Accession No. ADI91469.1, "enoyl-CoA hydratase [Acinetobacter oleivorans DR1]," dated Jan. 30, 2014, 1 page.
- GenBank Accession No. AEC07908.1, "4-(cytidine 5'-phospho)-2-C-methyl-D-erithritol kinase [*Arabidopsis thaliana*]," dated Jul. 20, 2017, 2 pages.
- GenBank Accession No. AED97354.1, "4-hydroxy-3-methylbut-2-enyl diphosphate synthase [*Arabidopsis thaliana*]," dated Jul. 20, 2017, 3 pages.
- GenBank Accession No. AEE35849.1, "hydroxy methylglutaryl CoA reductase 1 [*Arabidopsis thaliana*]," dated Jul. 20, 2017, 4 pages.
- 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.
- GenBank Accession No. AEE86362.1, "4-hydroxy-3-methylbut-2-enyl diphosphate reductase [*Arabidopsis thaliana*]," dated Jul. 20, 2017, 4 pages.
- GenBank Accession No. AFD33345.1, "acyl-activating enzyme 1 [*Cannabis sativa*]," dated Aug. 1, 2012, 1 page.
- GenBank Accession No. AFD33347.1, "acyl-activating enzyme 3 [*Cannabis sativa*]," dated Aug. 1, 2012, 1 page.
- GenBank Accession No. AFN42527.1, "olivetolic acid cyclase [*Cannabis sativa*]," dated Aug. 2, 2012, 1 page.

(56)

References Cited**OTHER PUBLICATIONS**

- GenBank Accession No. AGO55277.1, "polyketide biosynthesis malonyl CoA-acyl carrier protein transacylase BaeC [*Serratia plymuthica* 4Rx13]," dated Jan. 30, 2014, 2 pages.
- GenBank Accession No. AHM22925.1, "2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase [*Nicotiana tabacum*]," dated Mar. 22, 2014, 1 page.
- GenBank Accession No. AIE72439.1, "trans-2-enoyl-CoA reductase (plasmid) [*Klebsiella michiganensis*]," dated Jul. 8, 2016, 2 pages.
- GenBank Accession No. AIZ91493.1, "3-hydroxyacyl-CoA dehydrogenase [*Escherichia coli* str. K-12 substr. MG1655]," dated Dec. 15, 2014, 2 pages.
- GenBank Accession No. ALI39443.1, "acetyl-CoA acetyltransferase [*Escherichia coli* str. K-12 substr. MG1655]," dated Dec. 16, 2015, 2 pages.
- GenBank Accession No. AMC97367.1, "pyruvate dehydrogenase [*Escherichia coli* str. K-12 substr. MG1655]," dated Feb. 2, 2016, 2 pages.
- GenBank Accession No. ANM65835.1, "1-deoxy-D-xylulose 5-phosphate synthase 1 [*Arabidopsis thaliana*]," dated Jul. 20, 2017, 2 pages.
- GenBank Accession No. AUG14916.1, "dihydrolipoylysine-residue acetyltransferase component of pyruvate dehydrogenase complex [*Escherichia coli* str. K-12 substr. MG1655]," dated Dec. 17, 2017, 2 pages.
- GenBank Accession No. BAA21534.1, "N-WASP [*Rattus rattus*]," dated Dec. 27, 2006, 1 page.
- GenBank Accession No. BAB00624.1, "ATP citrate-lyase [*Ciona intestinalis*]," dated Jul. 15, 2000, 1 page.
- GenBank Accession No. BAB21592.1, "2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase [*Arabidopsis thaliana*]," dated Feb. 14, 2004, 1 page.
- GenBank Accession No. BAC41356.1, "tetrahydrocannabinolic acid synthase precursor [*Cannabis sativa*]," dated Sep. 15, 2004, 1 page.
- GenBank Accession No. BAF65033.1, "cannabidiolic acid synthase [*Cannabis sativa*]," dated Jun. 29, 2007, 1 page.
- GenBank Accession No. BAG14339.1, "olivetol synthase [*Cannabis sativa*]," dated Jun. 20, 2009, 1 page.
- GenBank Accession No. CAJ91294.1, "Enoyl-CoA hydratase [*Cupriavidus necator* H16]," dated Mar. 7, 2015, 2 pages.
- GenBank Accession No. CAJ92510.1, "pyruvate dehydrogenase complex, dehydrogenase (E1) component [*Cupriavidus necator* H16]," dated Mar. 7, 2015, 2 pages.
- GenBank Accession No. CAJ92511.1, "dihydrolipoamide acetyltransferase (E2) component of pyruvate dehydrogenase complex [*Cupriavidus necator* H16]," dated Mar. 7, 2015, 2 pgs.
- GenBank Accession No. CAJ92573.1, "Acetyl-CoA acetyltransferase [*Cupriavidus necator* H16]," dated Mar. 7, 2015, 2 pages.
- GenBank Accession No. CAQ66339.1, "Phosphomevalonate kinase [*Lactobacillus casei* BL23]," dated Feb. 27, 2015, 2 pages.
- GenBank Accession No. CAQ66619.1, "Pyruvate dehydrogenase complex, E2 component, dihydrolipoamide acetyltransferase [*Lactobacillus casei* BL23]," dated Feb. 27, 2015, 2 pages.
- GenBank Accession No. CAQ66794.1, "Mevalonate kinase [*Lactobacillus casei* BL23]," dated Feb. 27, 2015, 2 pages.
- GenBank Accession No. CAQ66795.1, "Diphosphomevalonate decarboxylase [*Lactobacillus casei* BL23]," dated Feb. 27, 2015, 2 pages.
- GenBank Accession No. CAQ66796.1, "Isopentenyl-diphosphate delta-isomerase (IPP isomerase) (Isopentenyl pyrophosphate isomerase) [*Lactobacillus casei* BL23]," dated Feb. 27, 2015, 2 pages.
- GenBank Accession No. CAQ66932.1, "Farnesyl-diphosphate synthase [*Lactobacillus casei* BL23]," dated Feb. 27, 2015, 2 pages.
- GenBank Accession No. CAQ67081.1, "Hydroxymethylglutaryl-CoA synthase [*Lactobacillus casei* BL23]," dated Feb. 27, 2015, 2 pages.
- GenBank Accession No. CAQ67082.1, "Hydroxymethylglutaryl-CoA reductase [*Lactobacillus casei* BL23]," dated Feb. 27, 2015, 2 pages.
- GenBank Accession No. CAQ67083.1, "Acetyl-CoA acetyltransferase (Acetoacetyl-CoA thiolase) [*Lactobacillus casei* BL23]," dated Feb. 27, 2015, 2 pages.
- GenBank Accession No. CAQ67359.1, "Biotin carboxylase [*Lactobacillus casei* BL23]," dated Feb. 27, 2015, 2 pages.
- GenBank Accession No. CDH63564.1, "4-hydroxy-3-methylbut-2-enyl diphosphate reductase [*Escherichia coli* PMV-1]," dated Sep. 18, 2013, 2 pages.
- GenBank Accession No. CDH63708.1, "1-deoxy-D-xylulose 5-phosphate reductoisomerase [*Escherichia coli* PMV-1]," dated Sep. 18, 2013, 1 page.
- GenBank Accession No. CDH63925.1, "1-deoxy-D-xylulose-5-phosphate synthase [*Escherichia coli* PMV-1]," dated Sep. 18, 2013, 1 page.
- GenBank Accession No. CDH64802.1, "4-diphosphocytidyl-2-C-methyl-D-erythritol kinase [*Escherichia coli* PMV-1]," dated Sep. 18, 2013, 1 page.
- GenBank Accession No. CDH66379.1, "2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase [*Escherichia coli* PMV-1]," dated Sep. 18, 2013, 2 pages.
- GenBank Accession No. CDH66380.1, "2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase [*Escherichia coli* PMV-1]," dated Sep. 18, 2013, 1 page.
- GenBank Accession No. DAA07148.1, "TPA: Etr1p [*Saccharomyces cerevisiae* S288C]," dated Mar. 27, 2017, 2 pages.
- GenBank Accession No. DAA07337.1, "TPA: pyruvate dehydrogenase (acetyl-transferring) subunit E1 beta [*Saccharomyces cerevisiae* S288C]," dated Mar. 27, 2017, 2 pages.
- GenBank Accession No. DAA10474.1, "TPA: dihydrolipoylysine-residue acetyltransferase [*Saccharomyces cerevisiae* S288C]," dated Mar. 27, 2017, 2 pages.
- GenBank Accession No. DAA10992.1, "TPA: [acyl-carrier-protein] S-malonyltransferase [*Saccharomyces cerevisiae* S288C]," dated Mar. 27, 2017, 2 pages.
- GenBank Accession No. EAZ63544.2, "Phosphomevalonate kinase [*Scheffersomyces stipitis* CBS 6054]," dated Jul. 11, 2011, 2 pages.
- GenBank Accession No. EDL06069.1, "syntrophin, acidic 1, isoform CRA_b [*Mus musculus*]," dated Jul. 26, 2016, 2 pages.
- GenBank Accession No. J04537.1, "*Arabidopsis thaliana* HMG-CoA reductase (HMG1) mRNA, complete cds," dated Sep. 29, 2015, 2 pages.
- GenBank Accession No. JN717233.1, "Cannabis sativa acyl-activating enzyme 1 mRNA, complete cds," dated Aug. 1, 2012, 2 pages.
- Havranek et al., "Automated design of specificity in molecular recognition," *Nat. Struct. Biology*, Jan. 2003, 10(1):45-52.
- 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.
- Hwang et al., "Efficient genome editing in zebrafish using a CRISPR-Cas system," *Nat. Biotechnology*, Mar. 2013, 31(3):227-229.
- Ito et al., "Transformation of Intact Yeast Cells Treated with Alkali Cations," *J. Bacteriology*, Jan. 1983, 153(1):163-168.
- Jiang et al., "Manipulation of GES and ERG20 for geraniol over-production in *Saccharomyces cerevisiae*," *Metab. Engineering*, May 2017, 41:57-66.
- Jiang et al., "RNA-guided editing of bacterial genomes using CRISPR-Cas systems," *Nat. Biotechnology*, Jan. 29, 2013, 31(3):233-239.
- Jinek et al., "A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity," *Science*, Aug. 17, 2012, 337(6096):816-821.
- Kim et al., "Optimization of hexanoic acid production in recombinant *Escherichia coli* by precise flux rebalancing," *Bioresour. Technology*, Jan. 2018, 247:1253-1257.
- 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.
- Makarova et al., "Evolution and classification of the CRISPR-Cas systems," *Nat. Rev. Microbiology*, Jun. 2011, 9(6):467-477.
- Mali et al., "RNA-Guided Human Genome Engineering via Cas9," *Science*, Feb. 15, 2013, 339(6121):823-826.

(56)

References Cited**OTHER PUBLICATIONS**

- PCT International Search Report and Written Opinion in International Appln. No. PCT/US2019/063029, dated Feb. 6, 2020, 9 pages.
- Proschel et al., "Engineering of metabolic pathways by artificial enzyme channels," *Front. Bioeng. Biotechnology*, Oct. 2015, 3:168, 13 pages.
- 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.
- Redden et al., (2015) The synthetic biology toolbox for tuning gene expression in yeast, *FEMS Yeast Res.*, vol. 15, pp. 1-10.
- 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.
- 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.
- Sirikanaramas et al., "The Gene Controlling Marijuana Psychoactivity," *J. Biol. Chemistry*, Jun. 9, 2004, 279(38):39767-39774.
- Song et al., "Engineering *Saccharomyces cerevisiae* for geranylgeranol overproduction by combinatorial design," *Sci. Reports*, Nov. 8, 2017, 7:14991, 11 pages.
- 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.
- Taura et al., "Cannabidiolic-acid synthase, the chemotype-determining enzyme in the fiber-type *Cannabis sativa*," *FEBS Letters*, Jun. 2007, 581(16):2929-2934.
- Taura et al., "Characterization of olivetol synthase, a polyketide synthase putatively involved in cannabinoid biosynthetic pathway," *FEBS Letters*, Jun. 2009, 583(12):2061-2066.
- Teyra et al., "Elucidation of the binding preferences of peptide recognition modules: SH3 and PDZ domains," *FEBS Letters*, Jun. 2012, 586(17):2631-2637.
- Whitaker et al., "Metabolic Pathway Flux Enhancement by Synthetic Protein Scaffolding," *Meth. Enzymology*, 2011, 497(Part A):447-468.
- 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.
- GenBank Accession No. AAS11585.1, "hydroxymethylbutenyl pyrophosphate reductase [*Treponema denticola* ATCC 35405]," dated Jan. 31, 2014, 2 pages.
- 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.
- GenBank Accession No. CAQ66617.1, "Pyruvate dehydrogenase complex, E1 component, alpha subunit [*Lactobacillus casei* BL23]," dated Feb. 27, 2015, 2 pages.

* cited by examiner

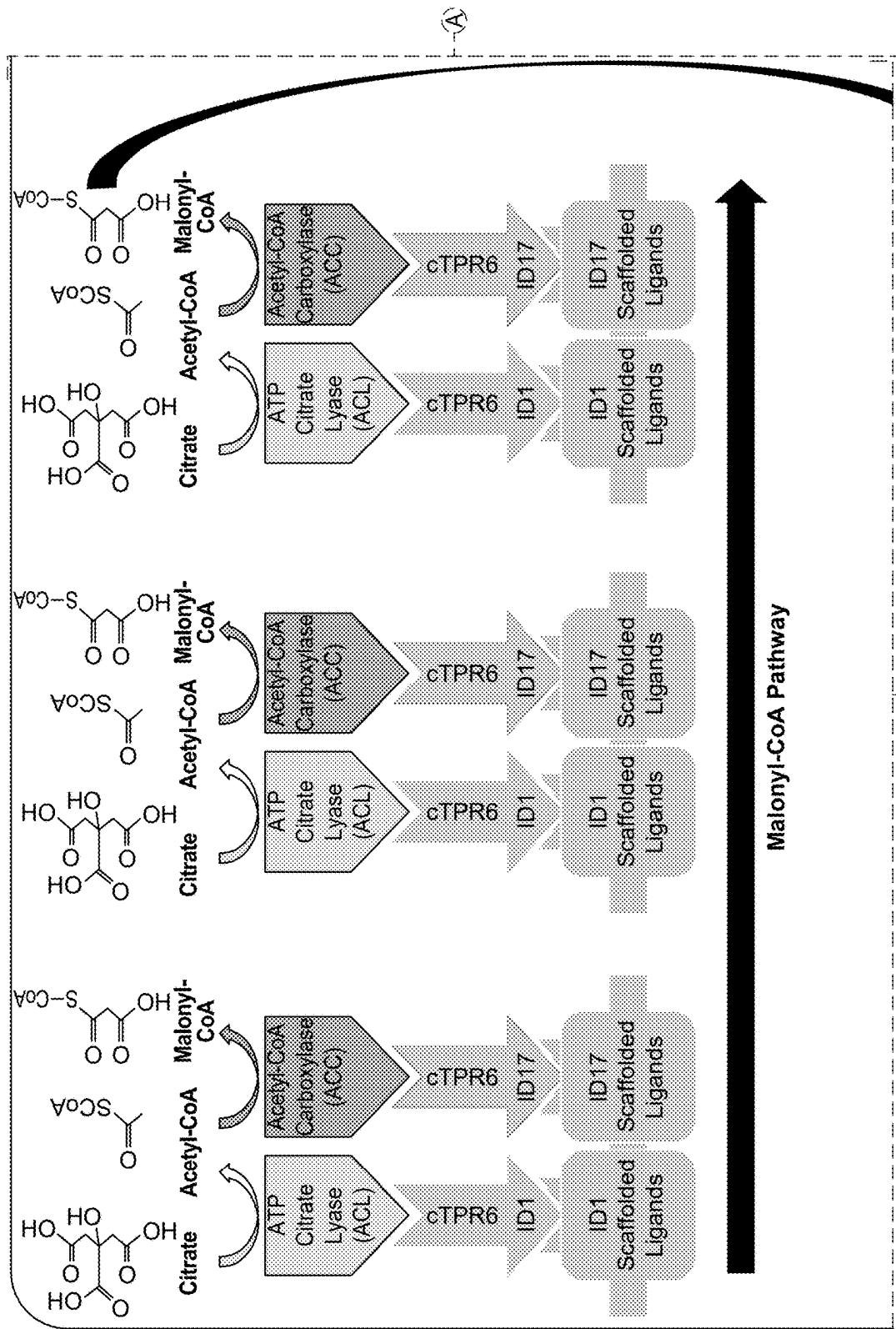


Figure 1A.
Malonyl-CoA Pathway

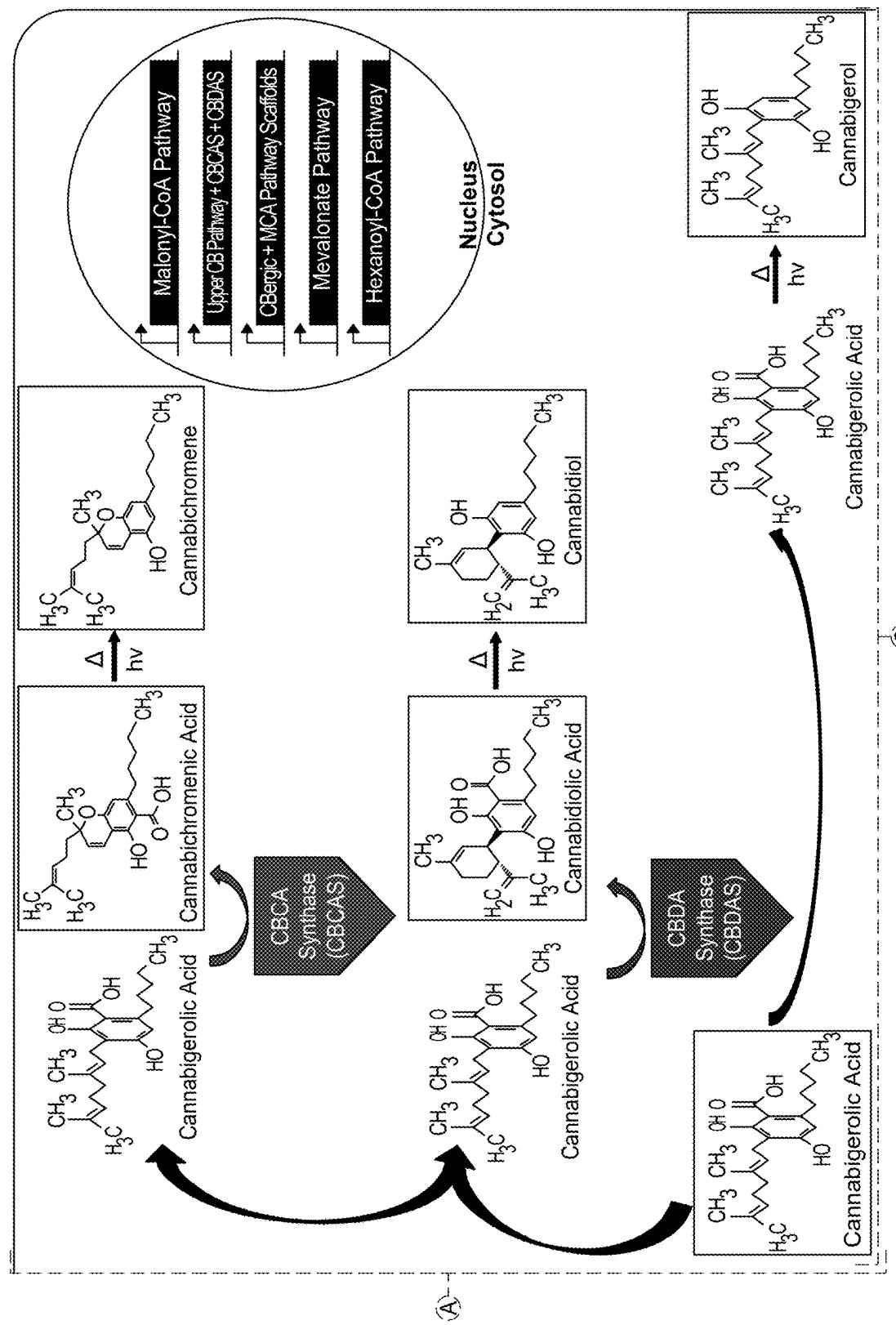


Figure 1A. (Cont'd)

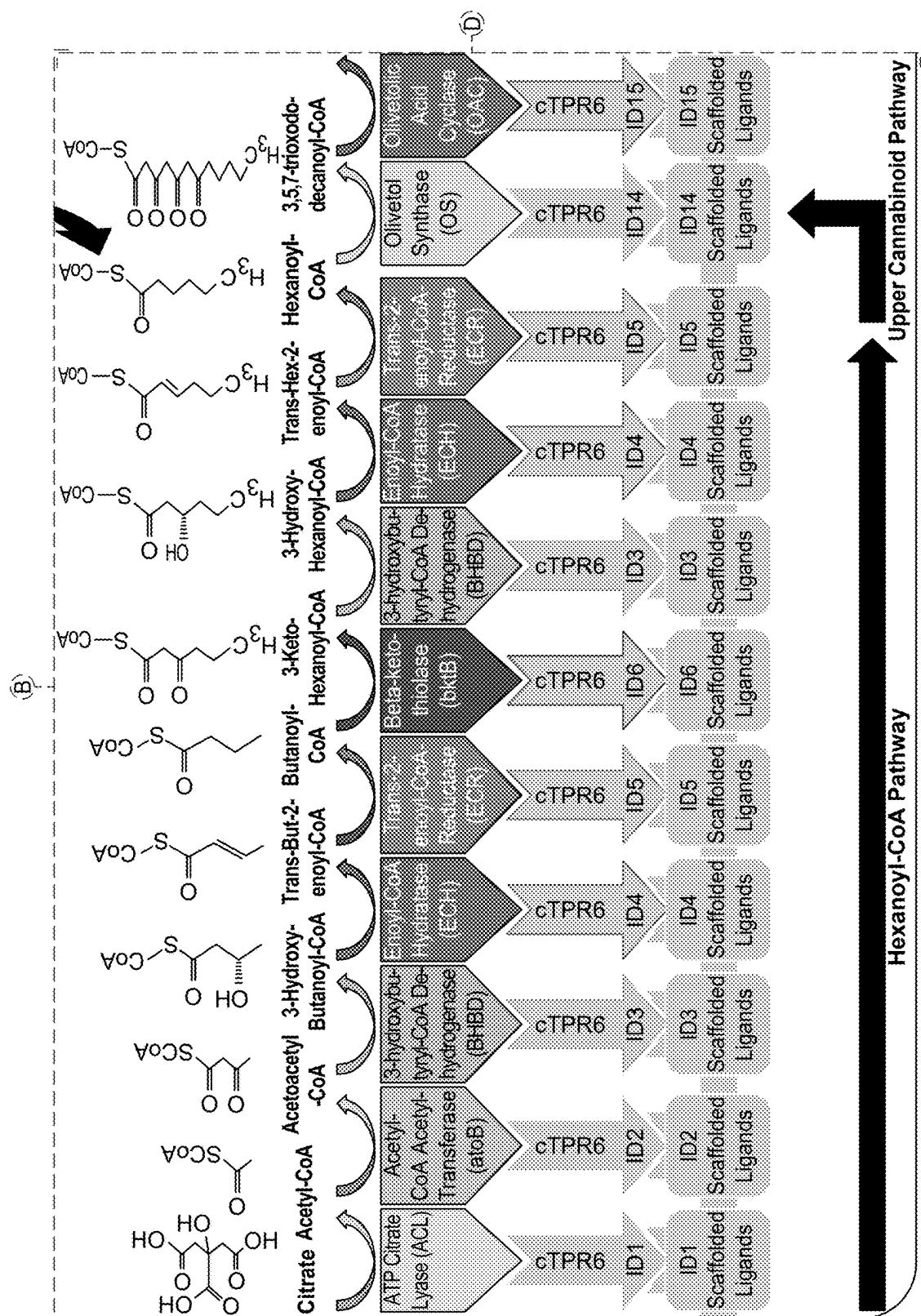


Figure 1A. (Cont'd)

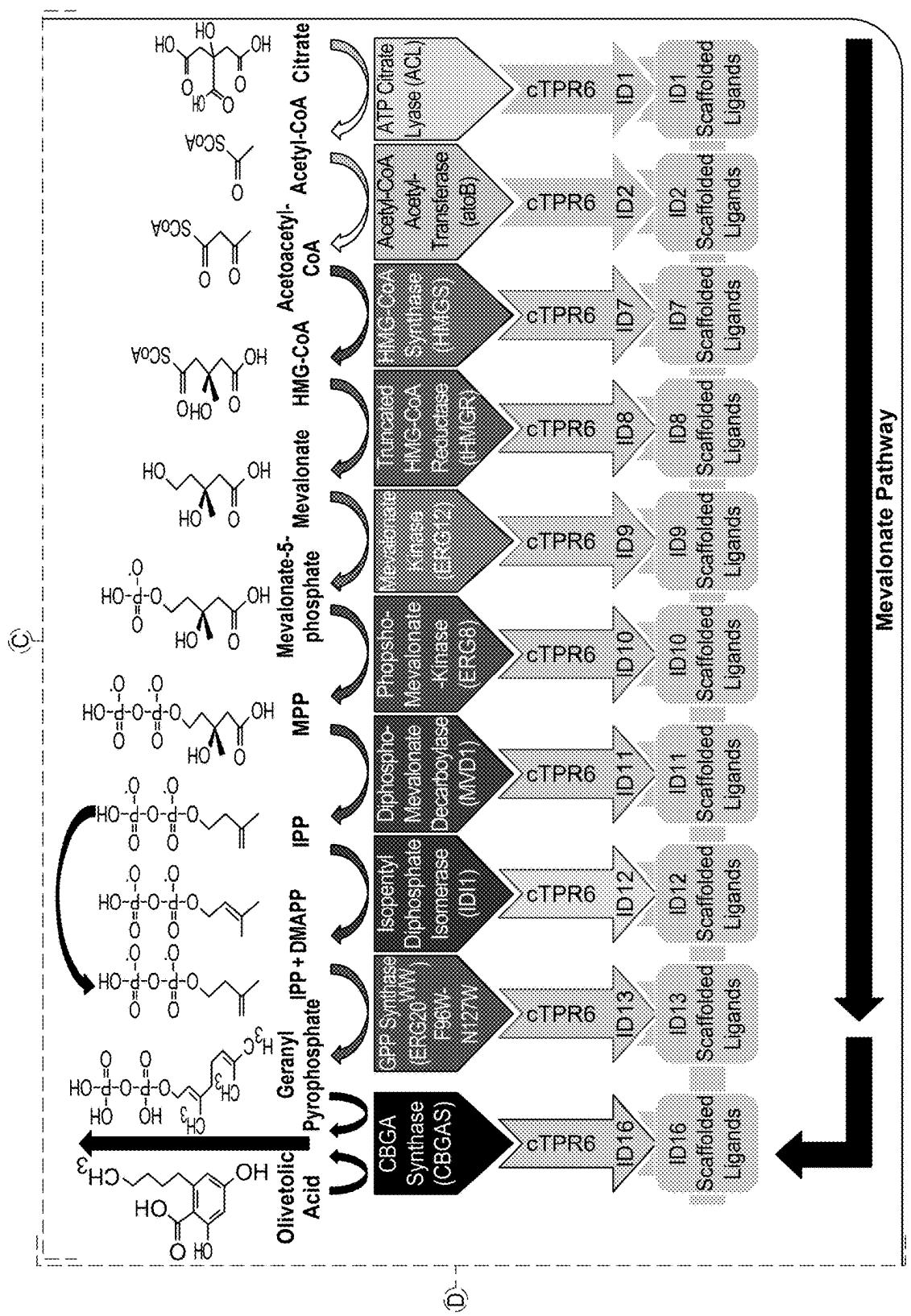


Figure 1A. (Cont'd)

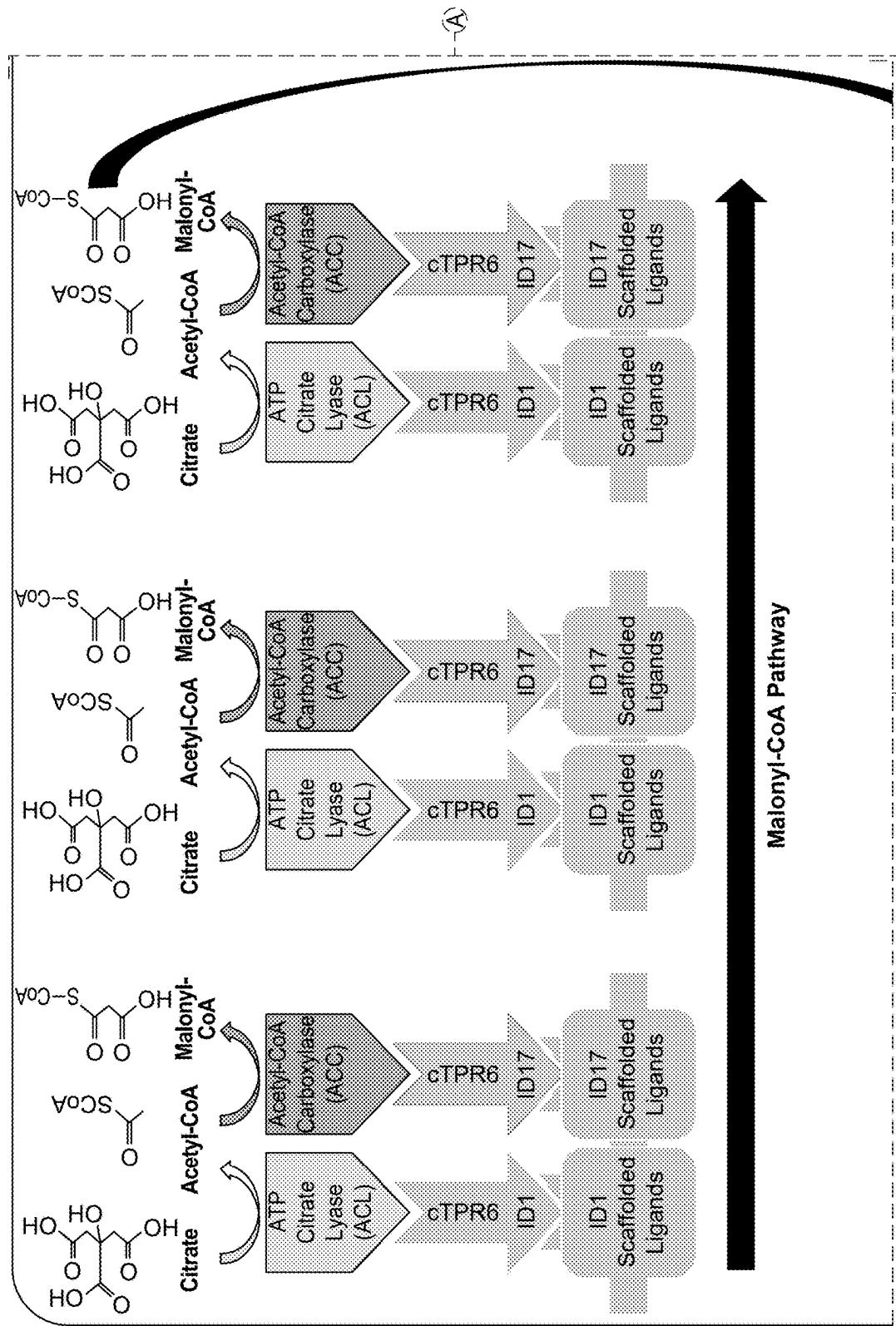


Figure 1B.

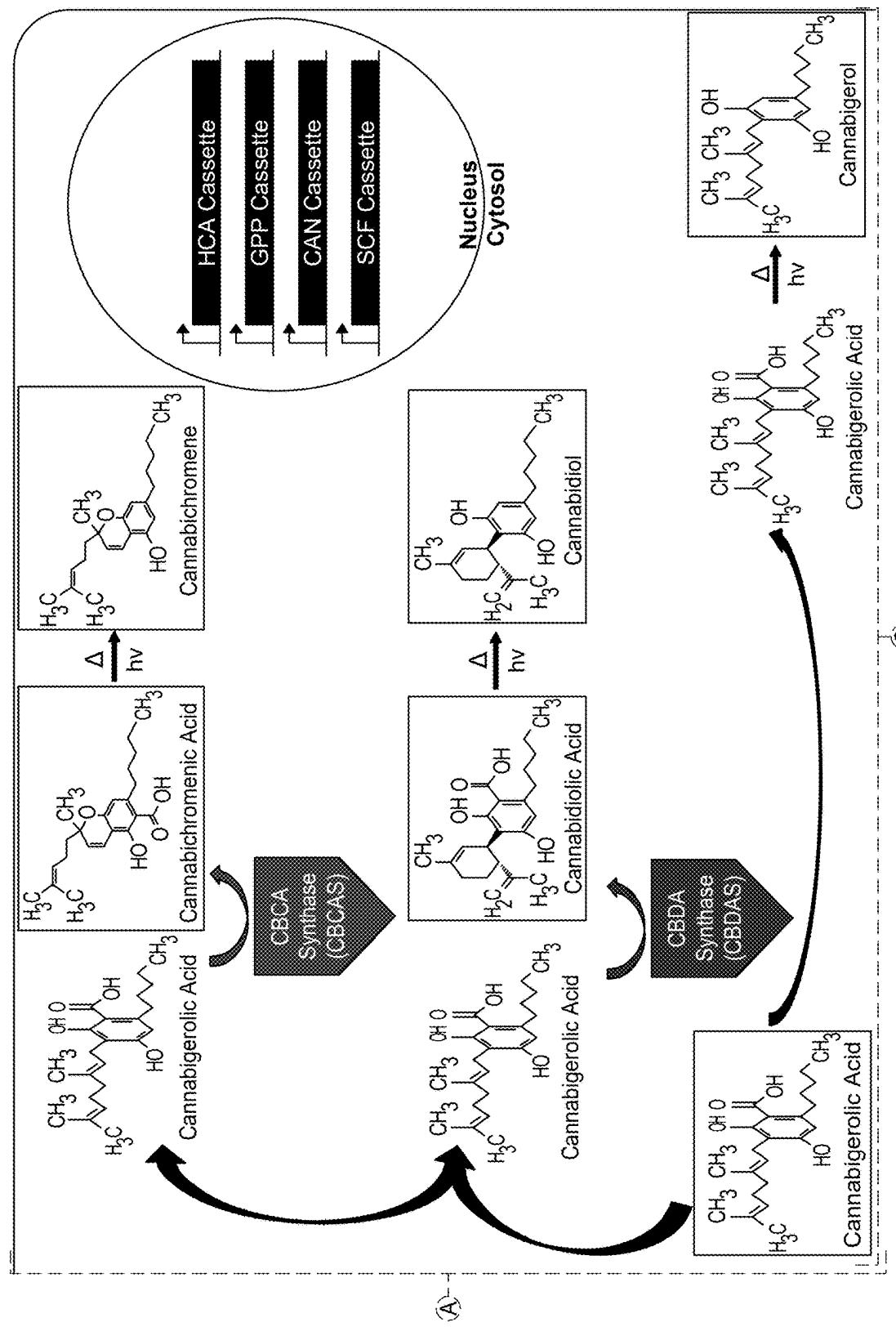


Figure 1B. (Cont'd)

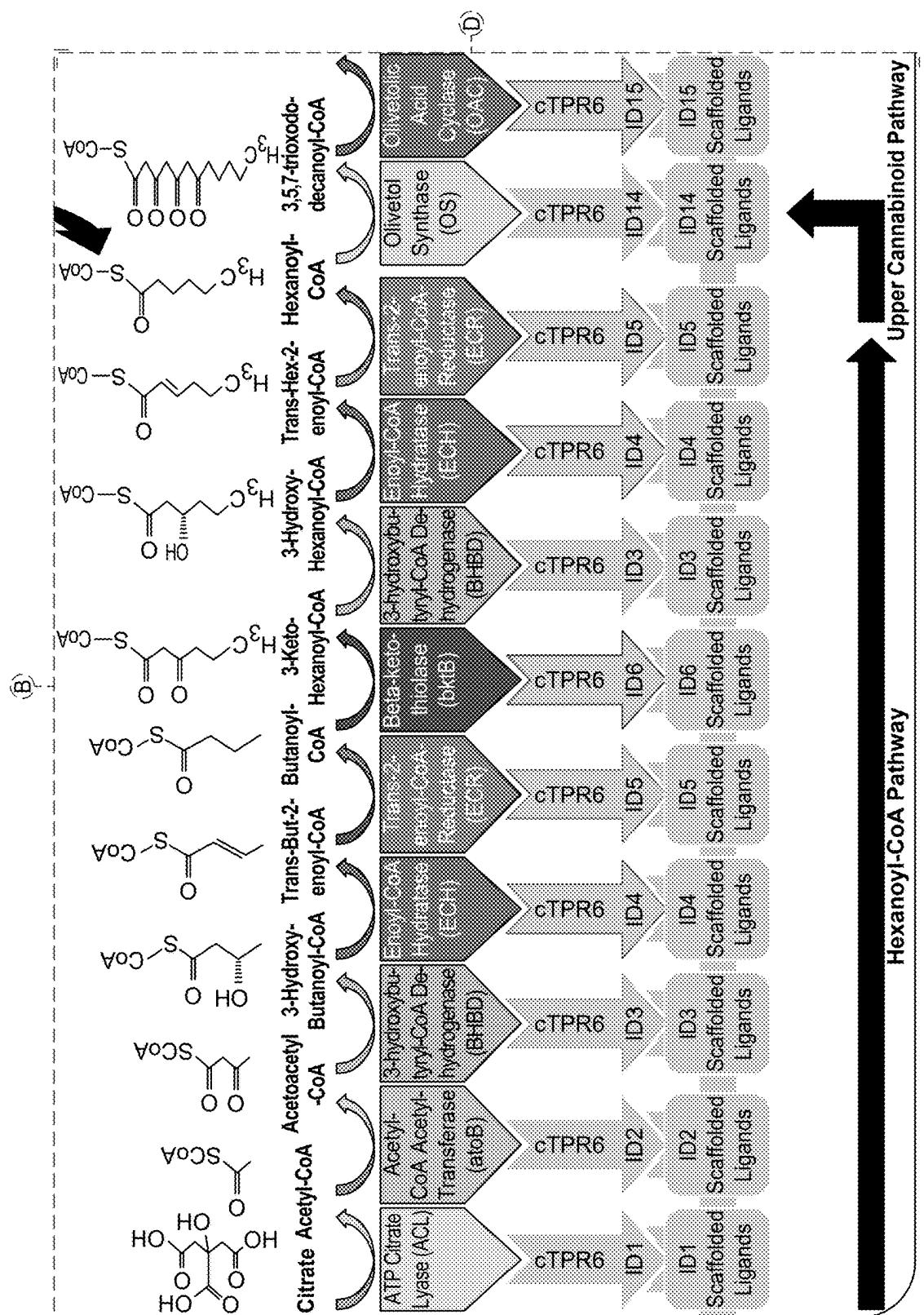


Figure 1B. (Cont'd)

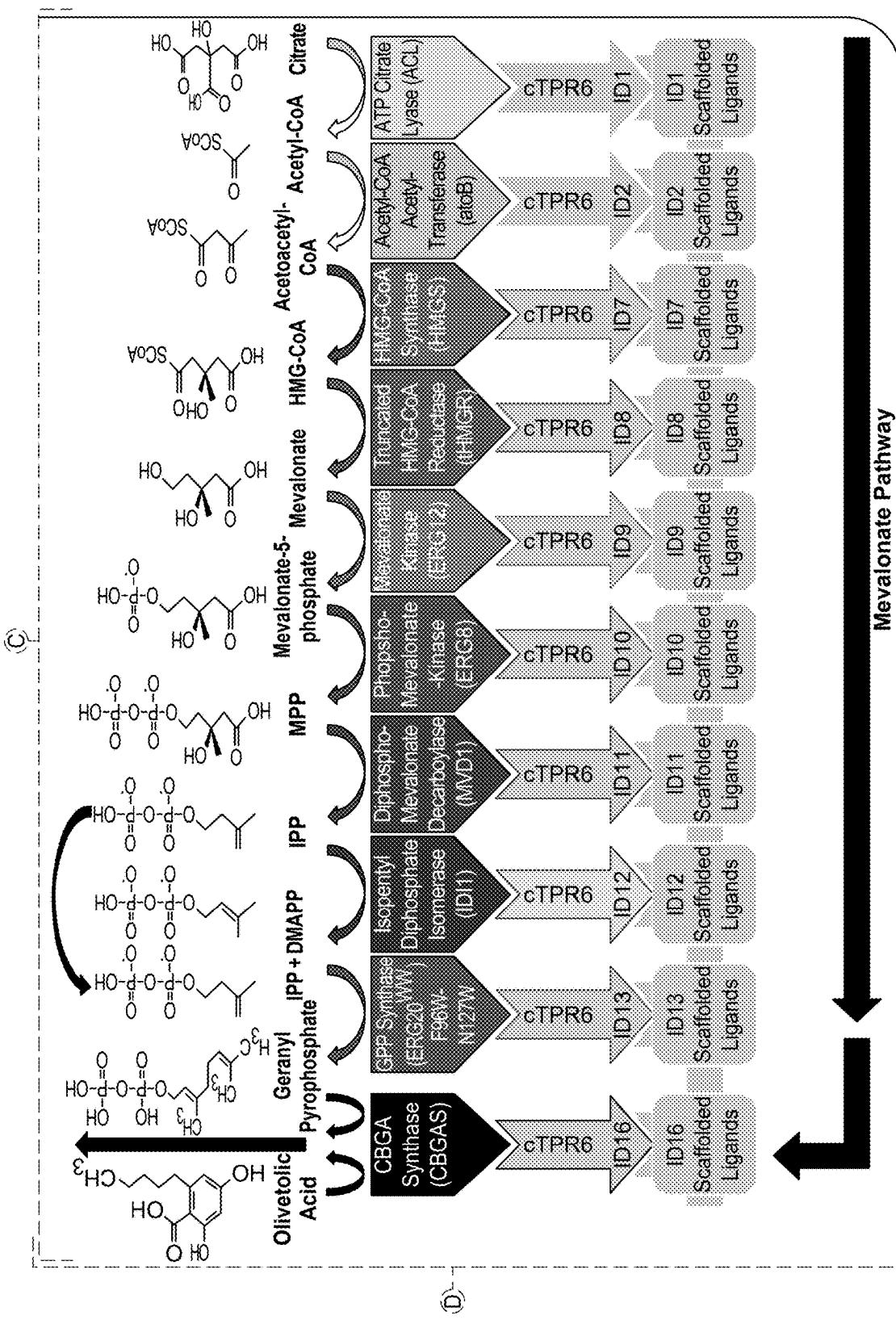


Figure 1B. (Cont'd)

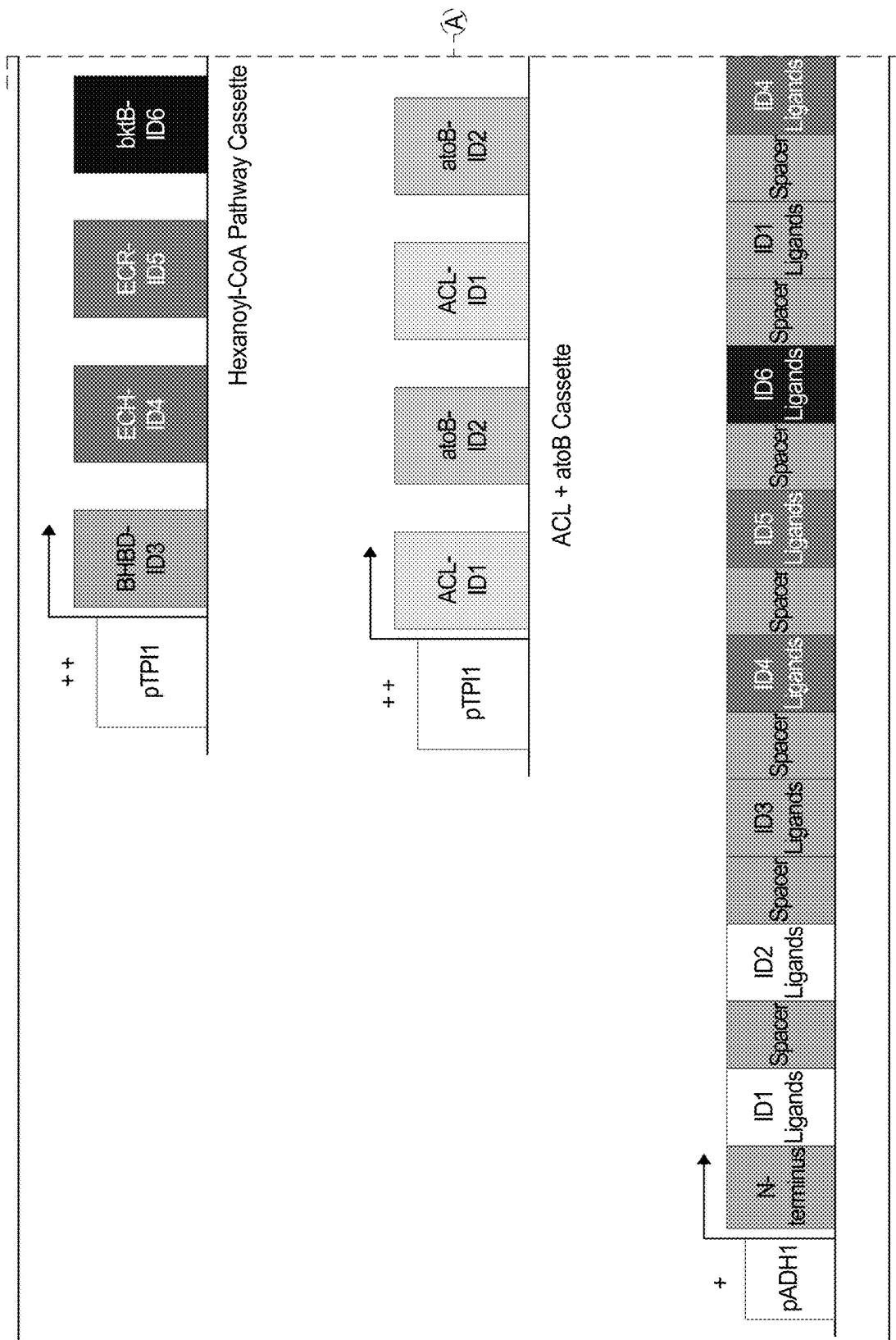


Figure 2A.

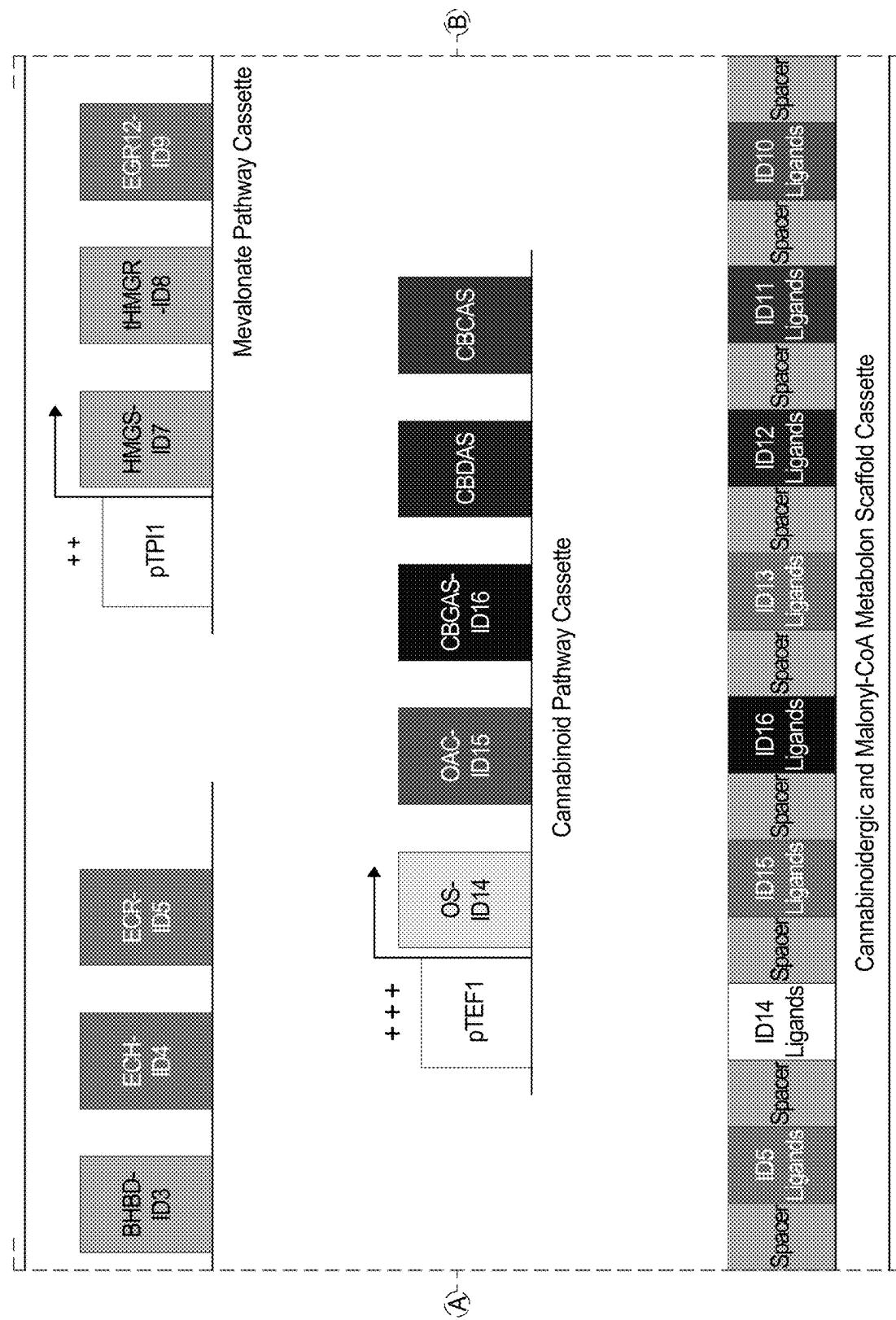


Figure 2A. (Cont'd)

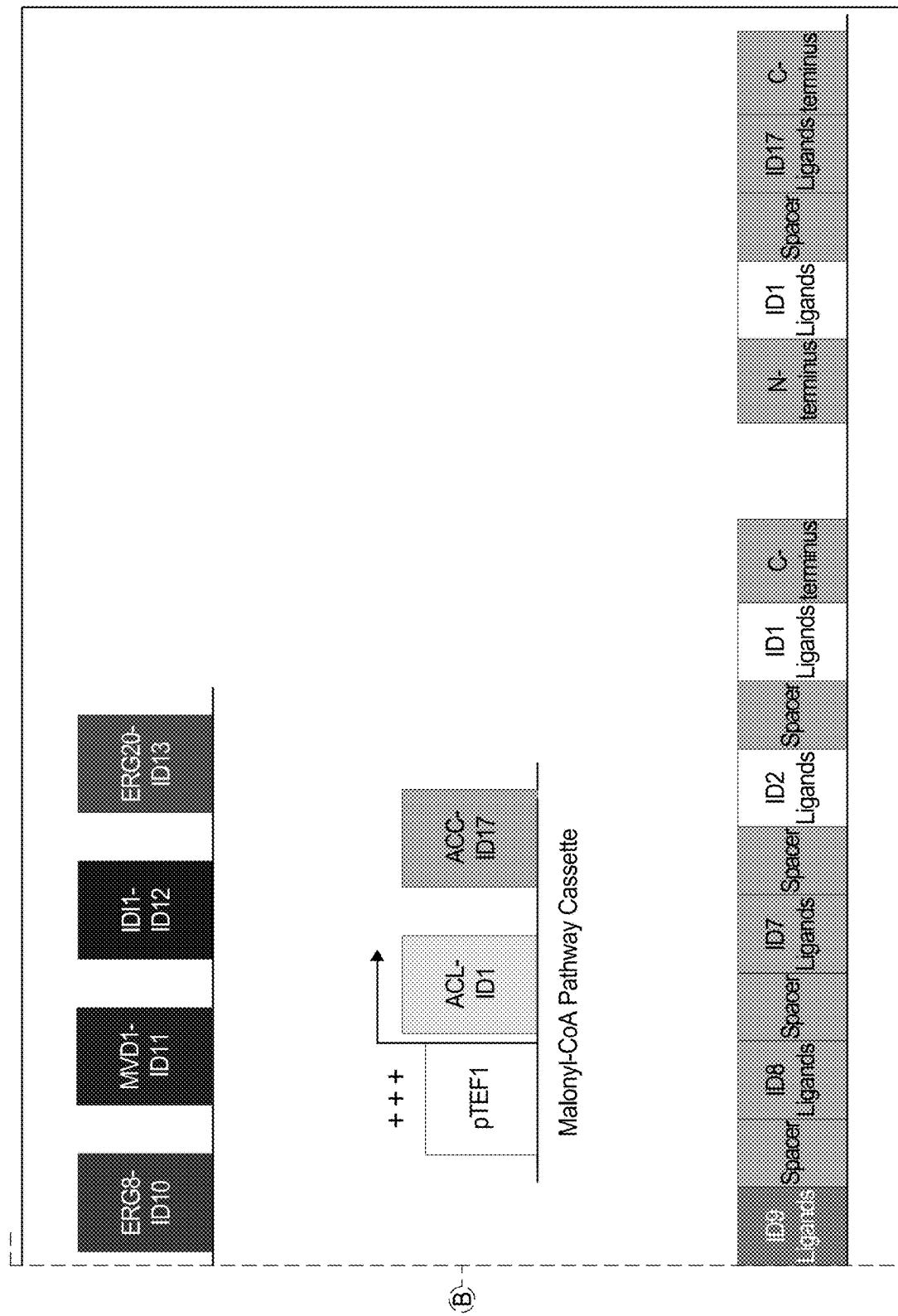


Figure 2A. (Cont'd)

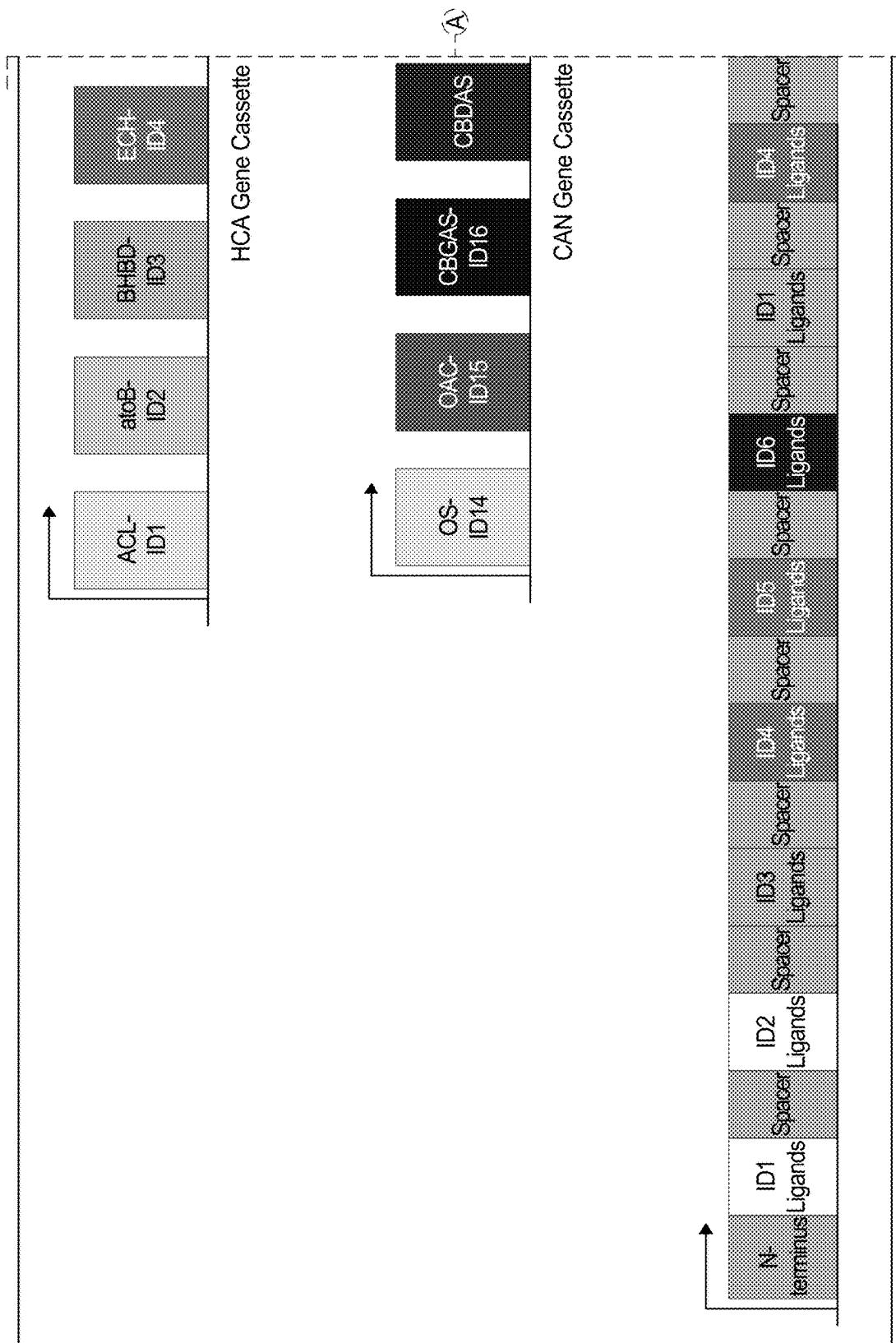


Figure 2B.

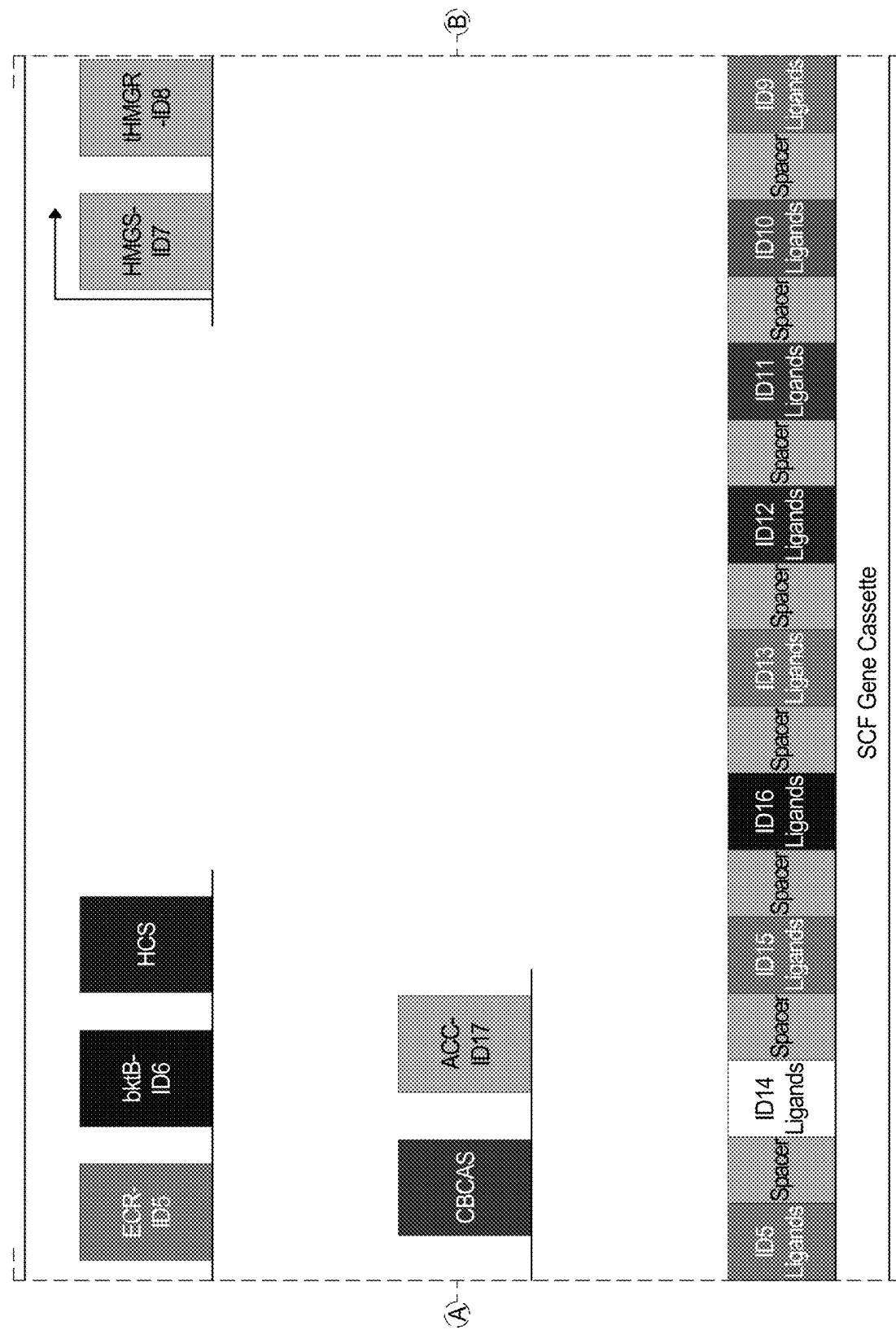


Figure 2B. (Cont'd)

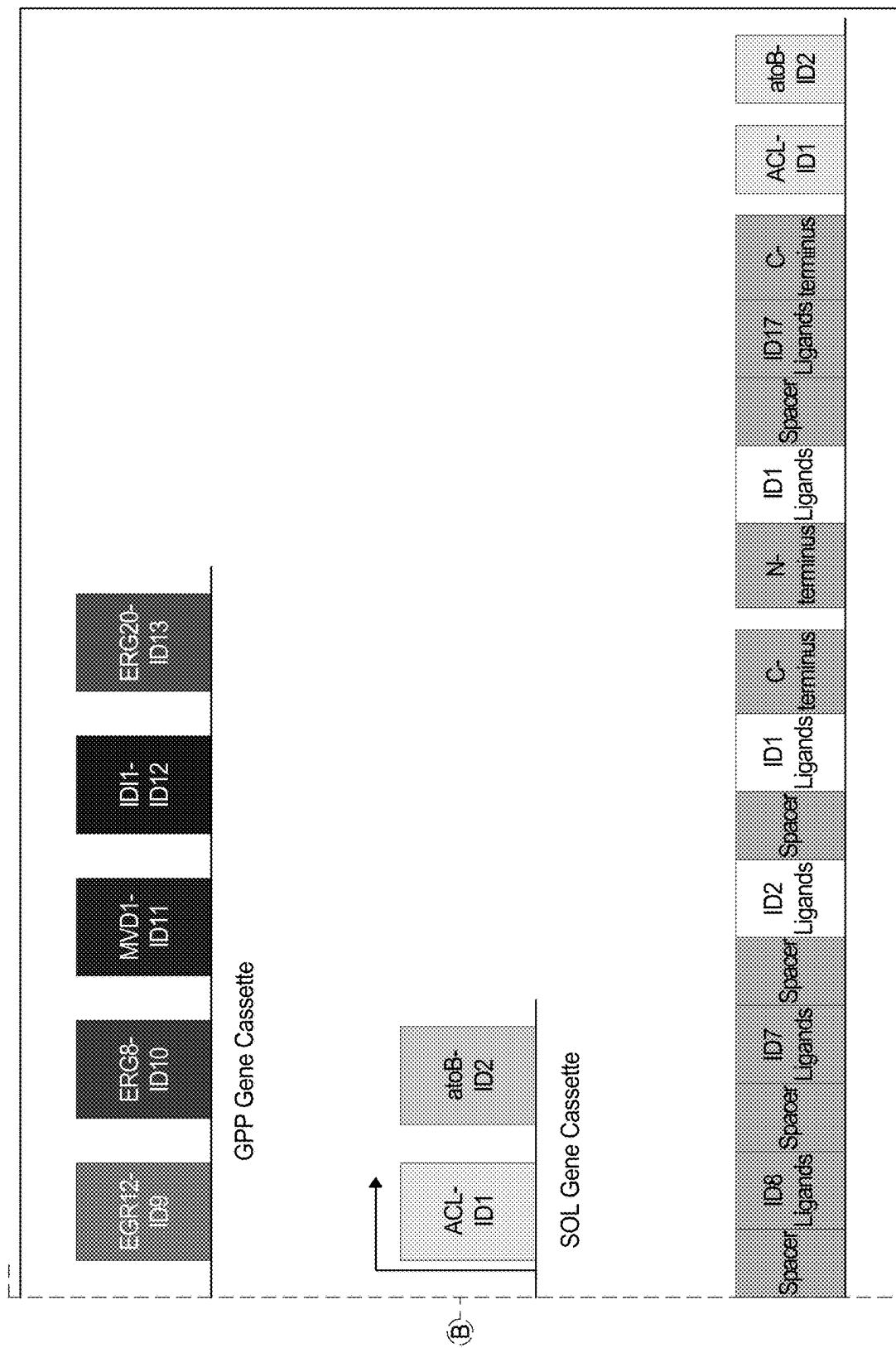


Figure 2B. (Cont'd)

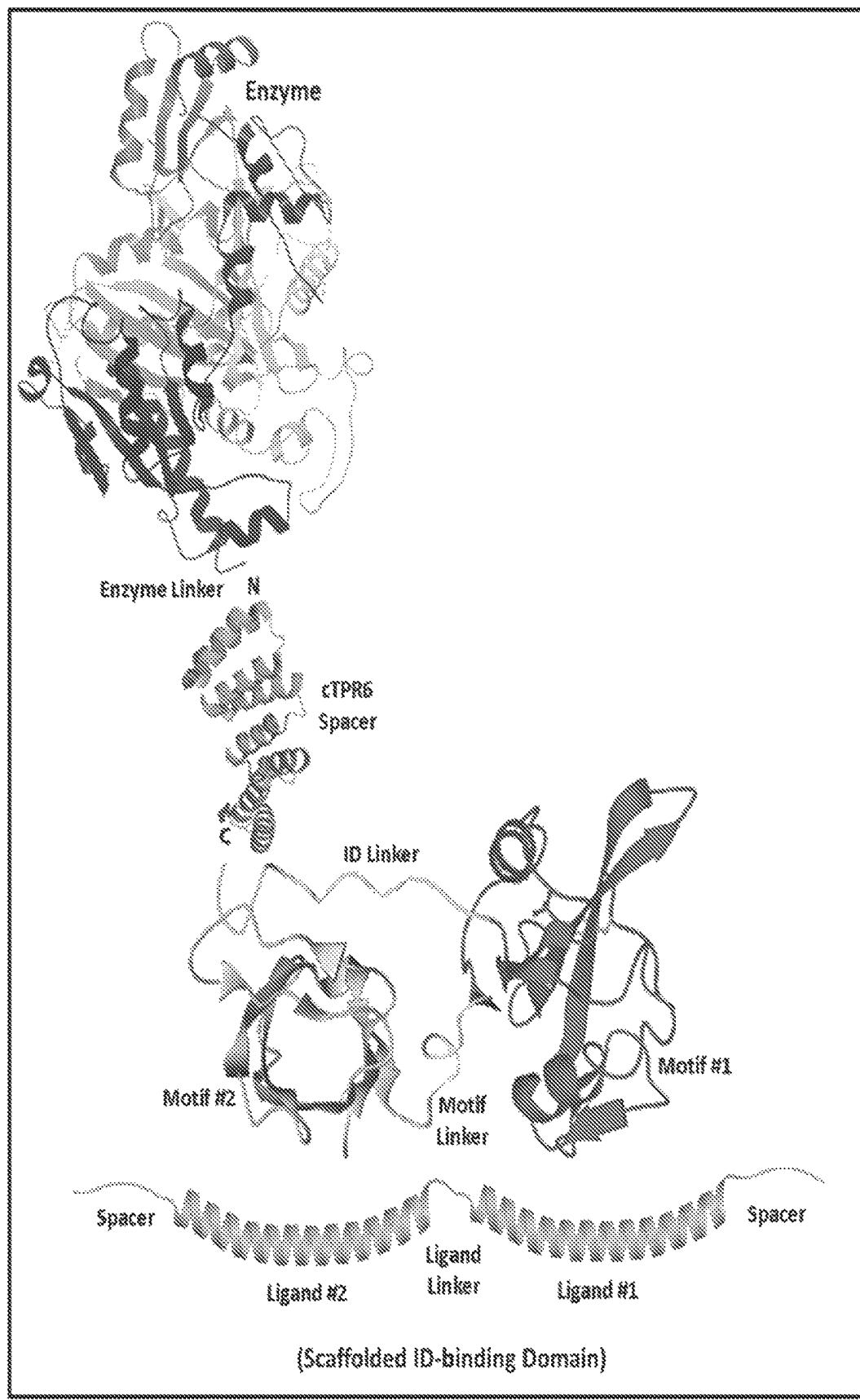


Figure 3.

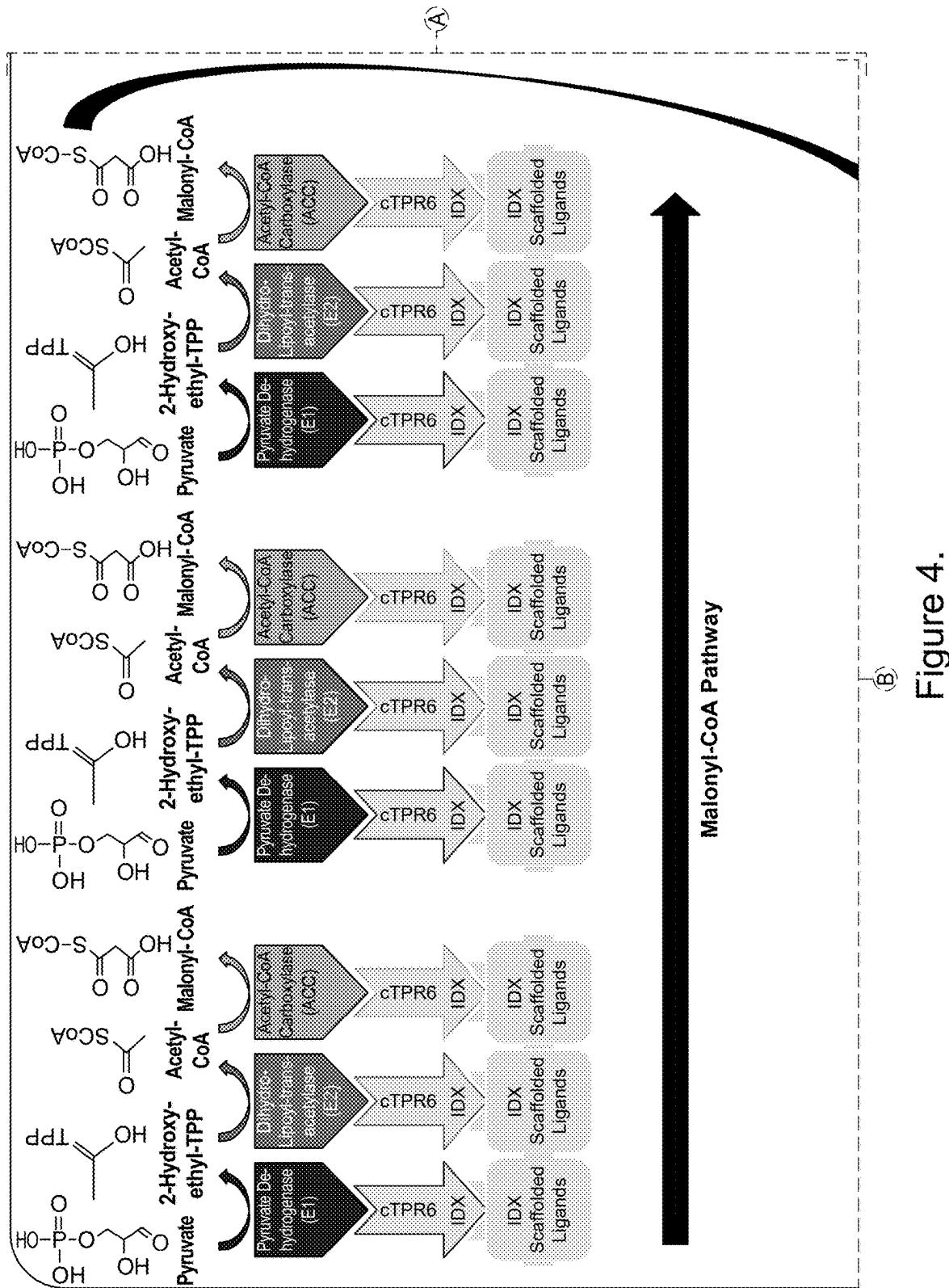
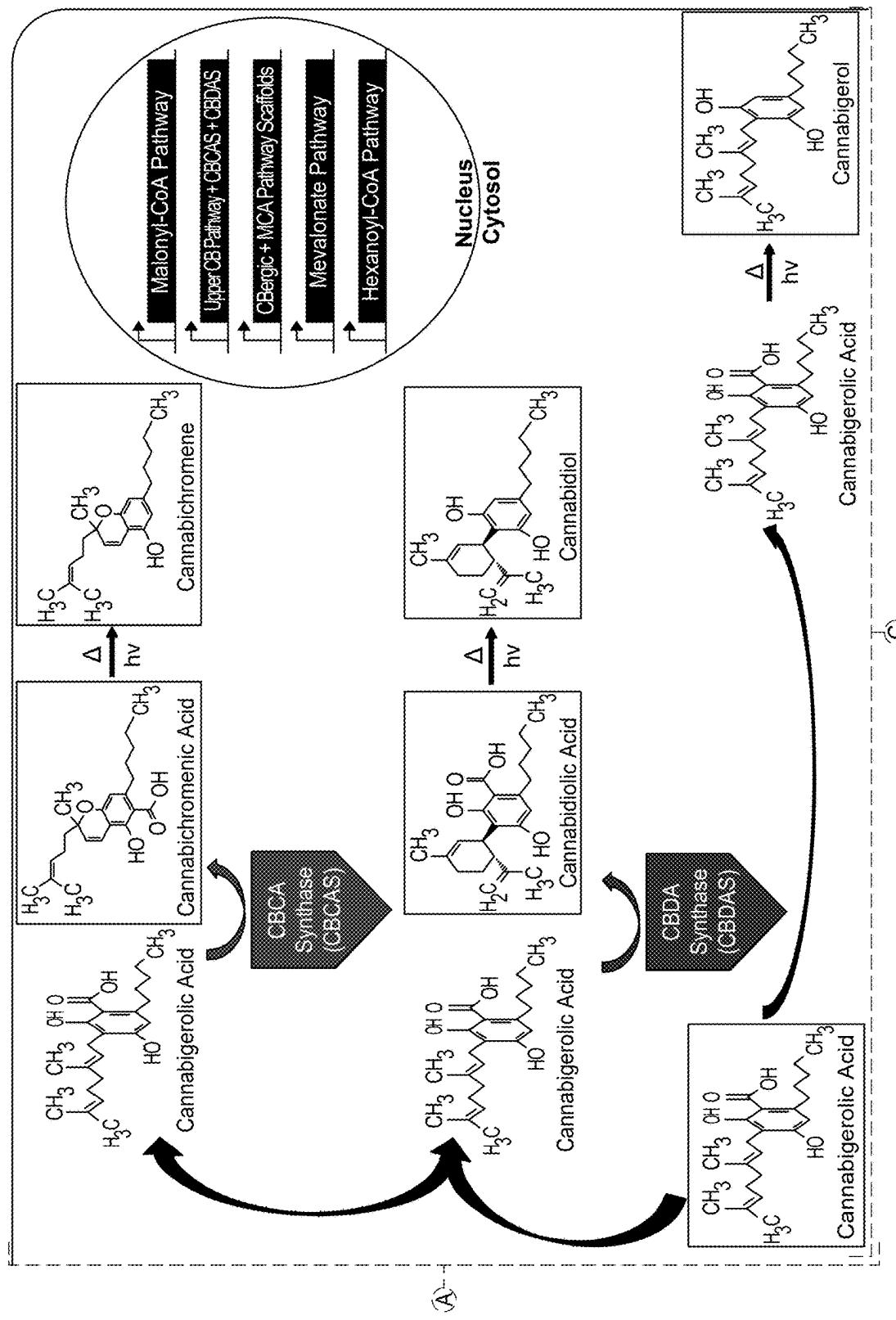


Figure 4.



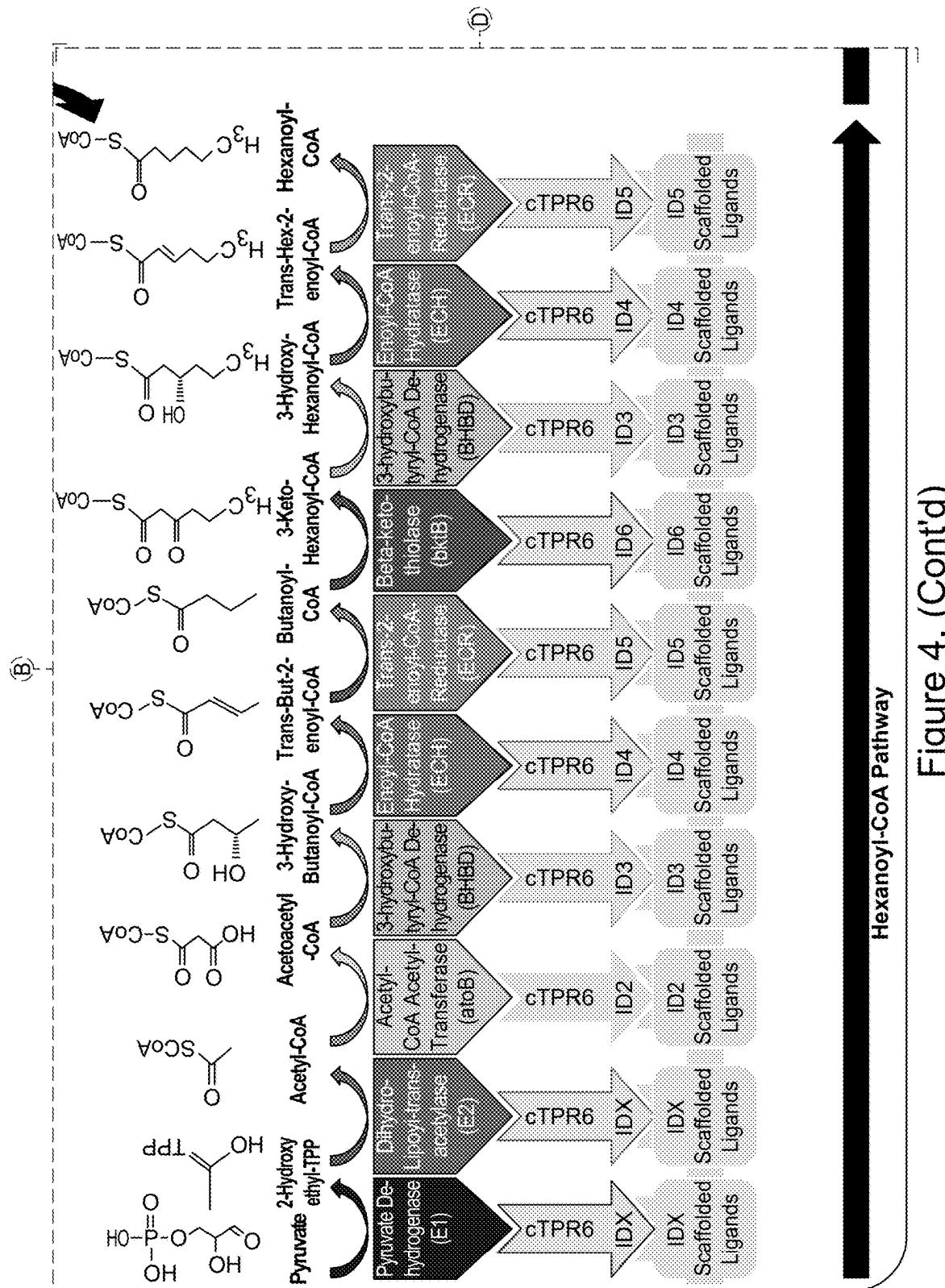
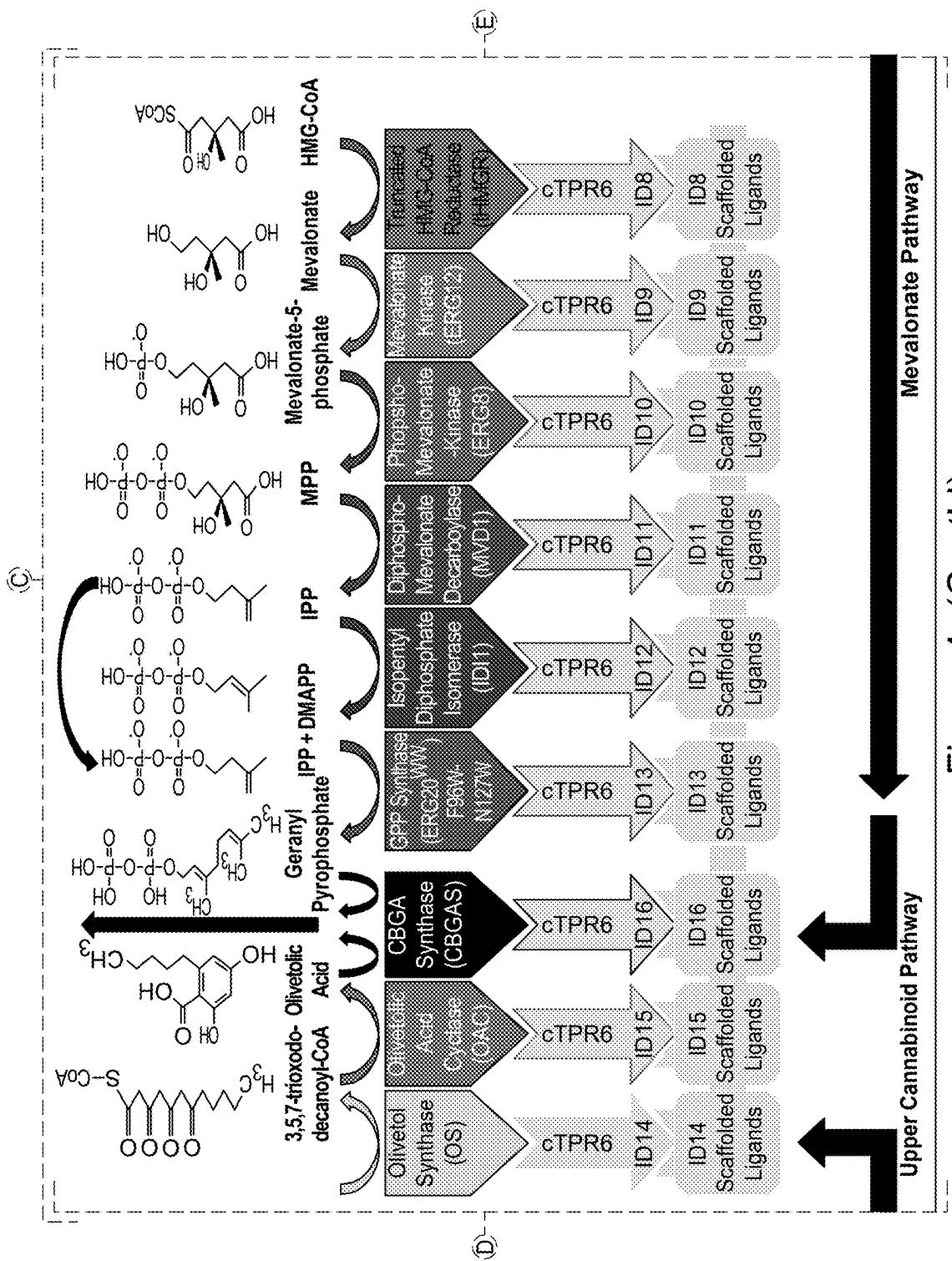


Figure 4. (Cont'd)



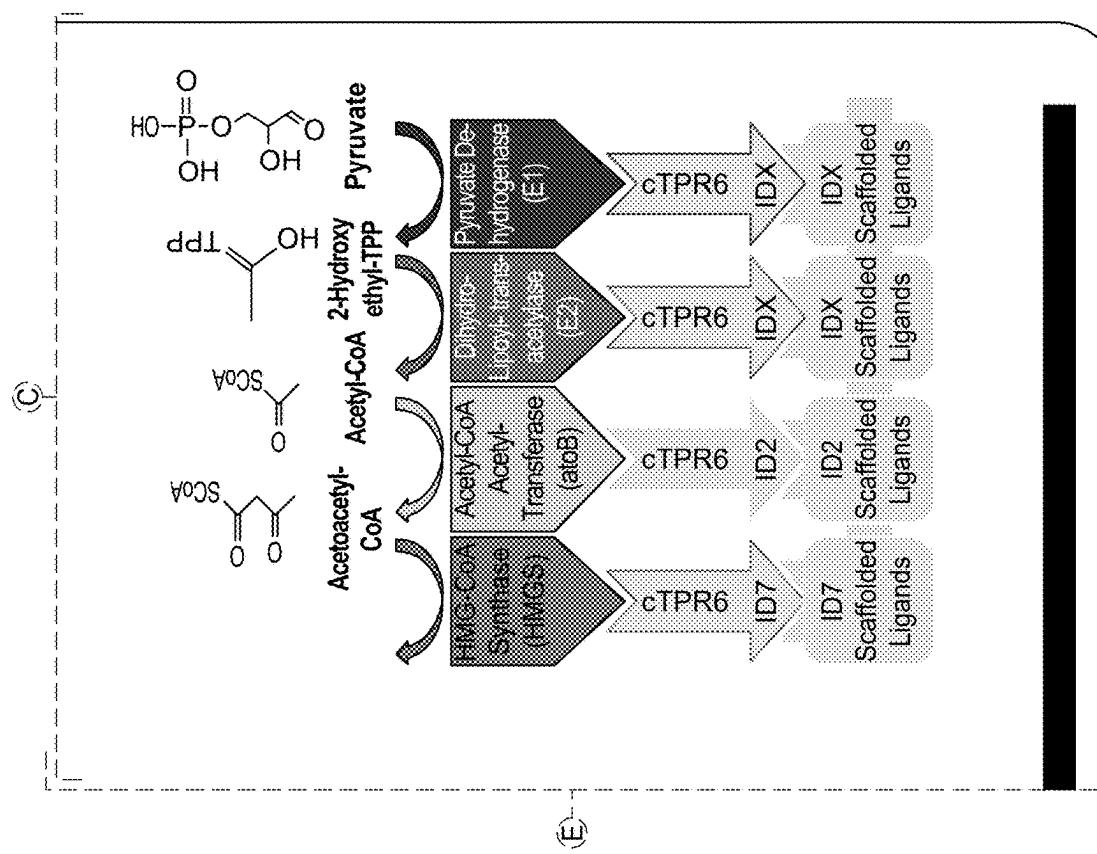


Figure 4. (Cont'd)

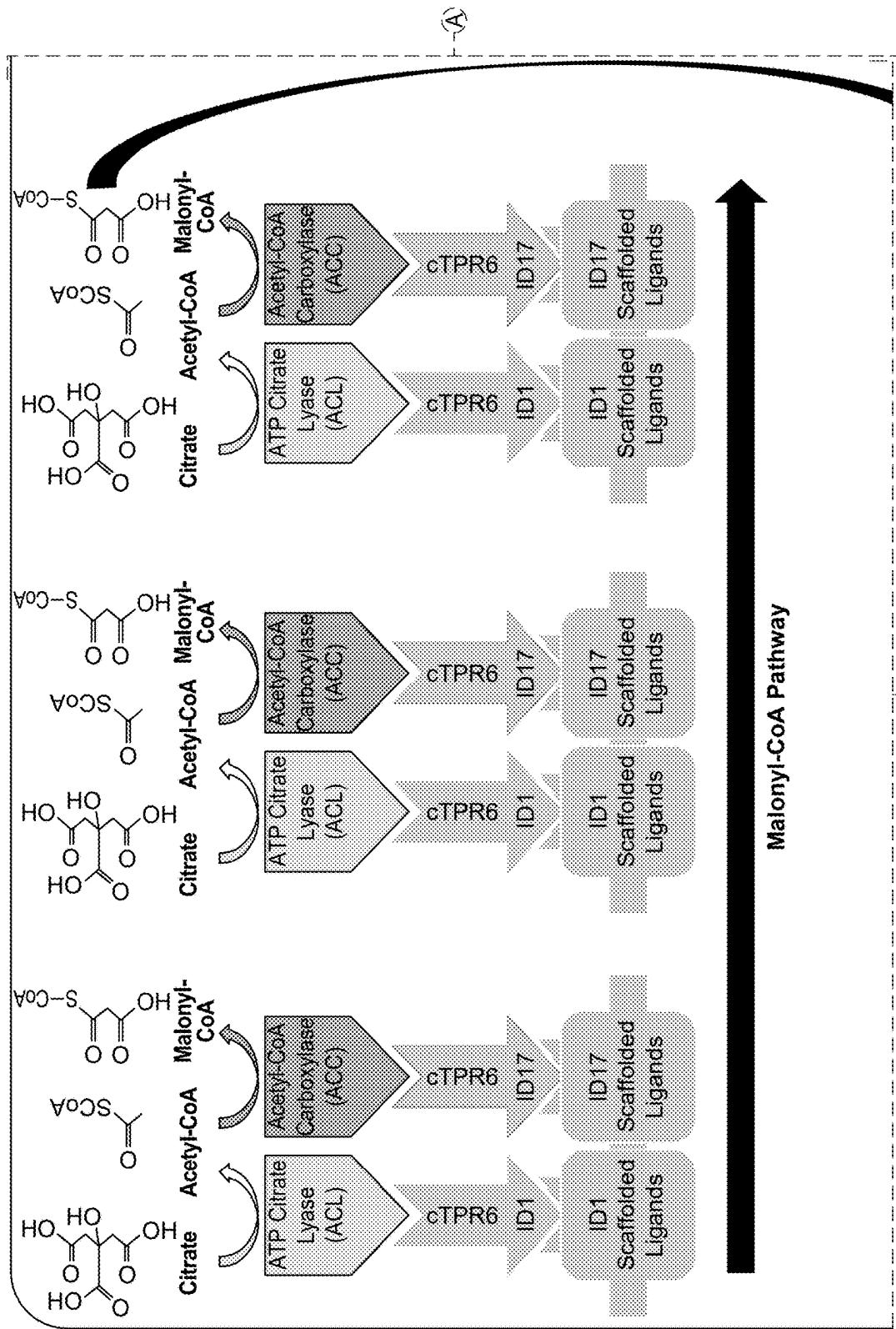


Figure 5.
Malonyl-CoA Pathway

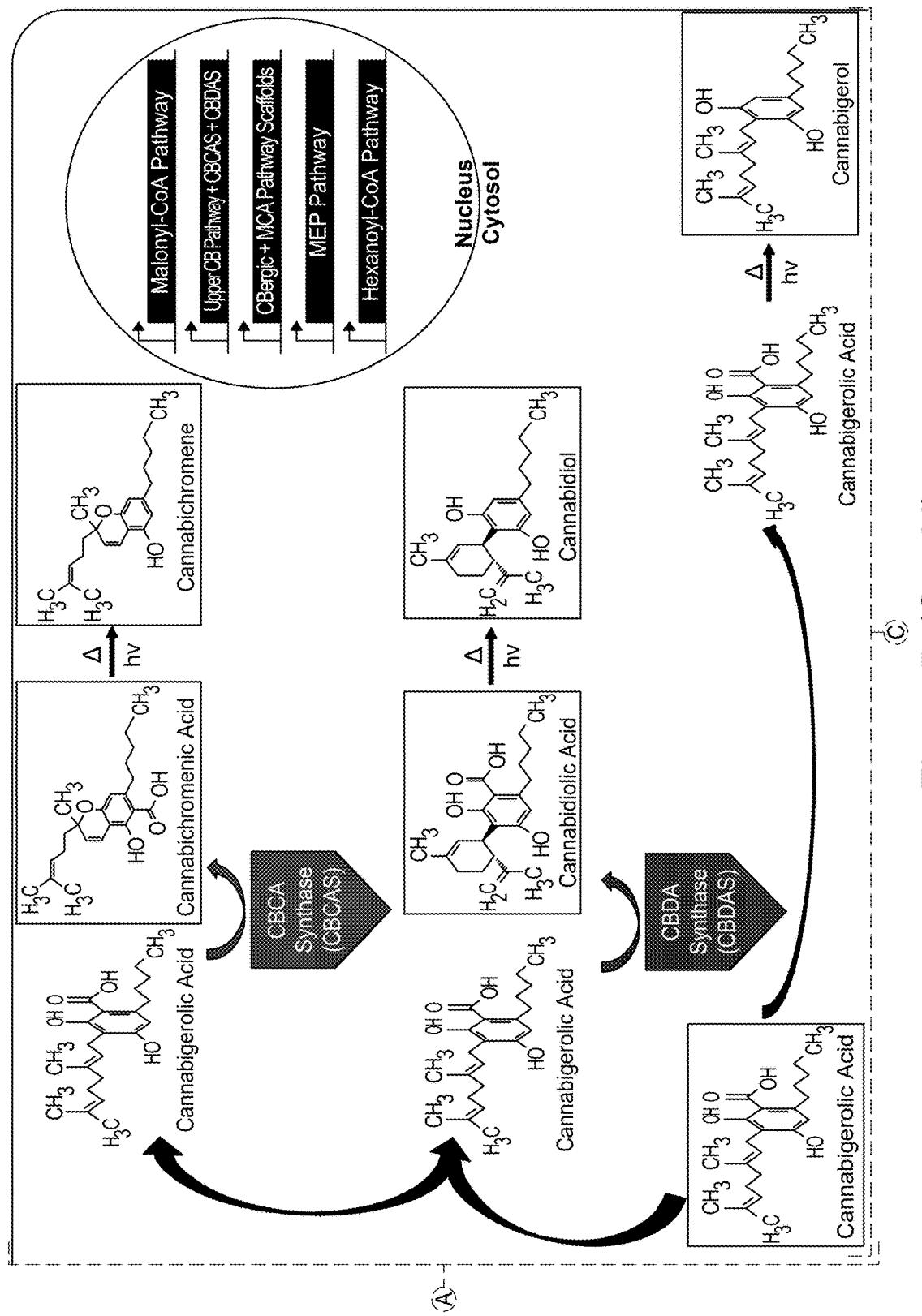


Figure 5. (Cont'd)

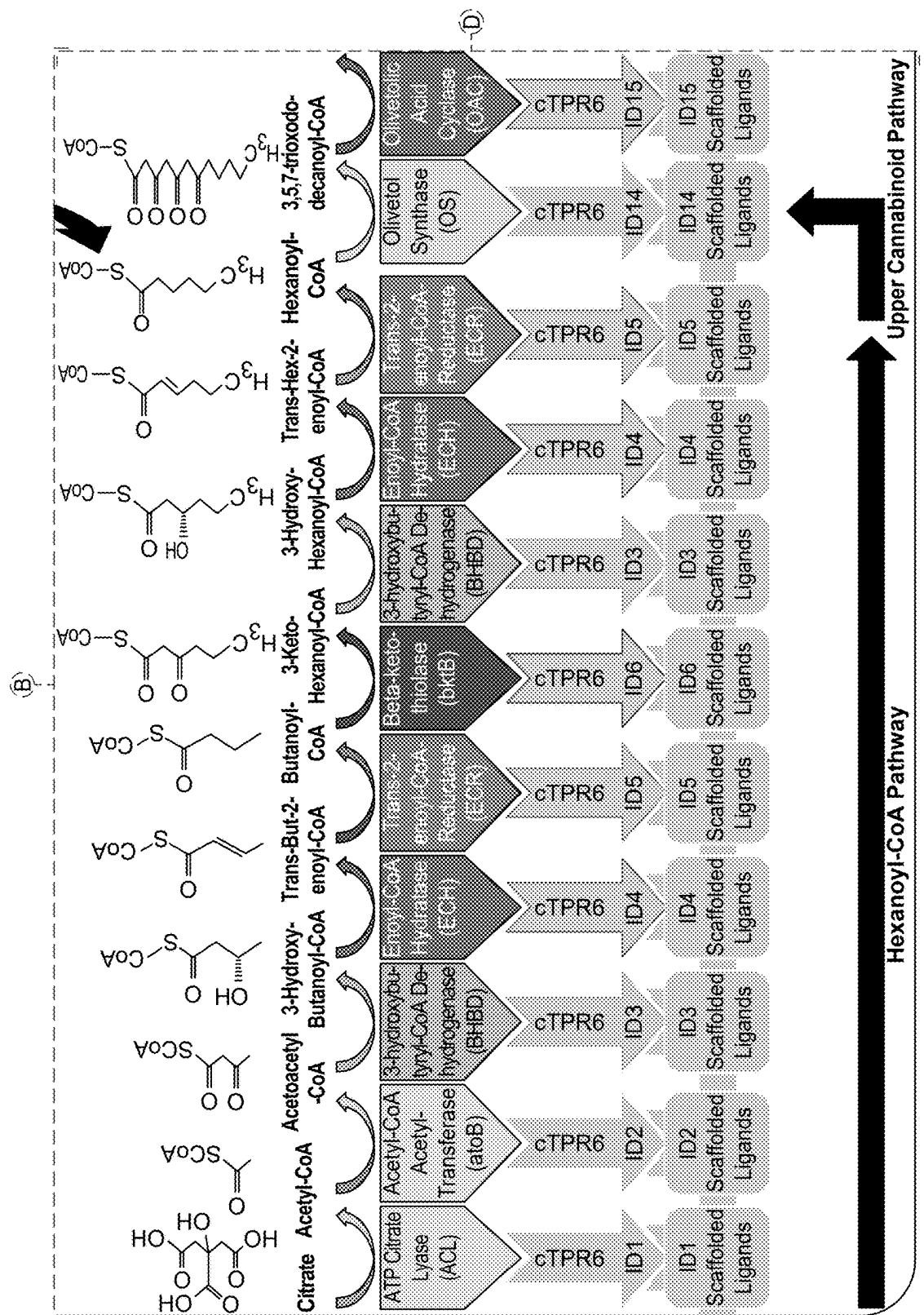


Figure 5. (Cont'd)
Hexanoyl-CoA Pathway

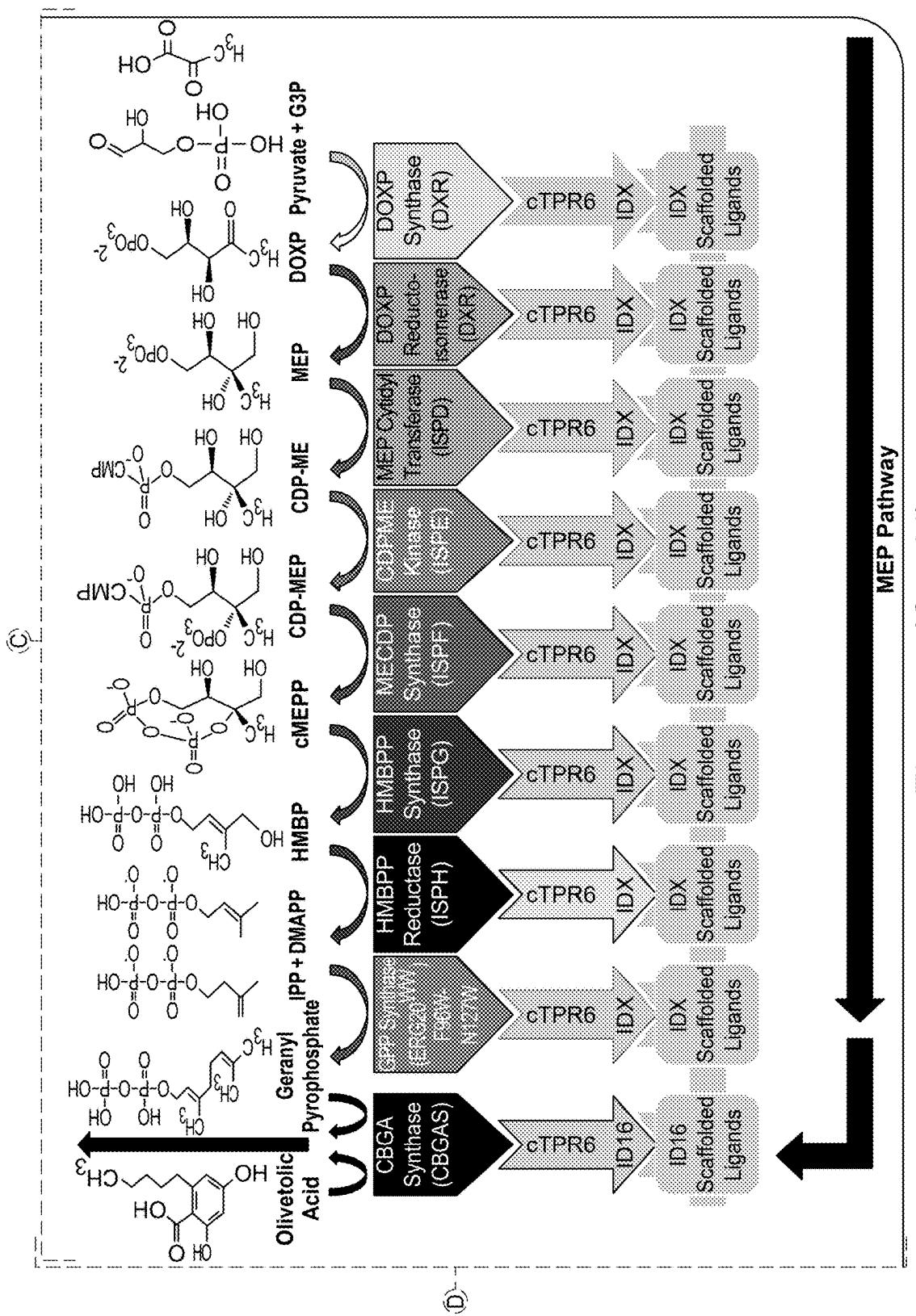


Figure 5. (Cont'd)

Figure 6A

ATP Citrate Lyase

MSAKAISEQTGKELLYKFICTTSAIQNRFKYARVTPDTDWARLLQDHWPWLLSQNLVVKPDQLIKRRGKGLGVVNLTLDGVKSWLKPRLGQEAVGKATGFLKNFIELPFVPHSQAEFYVCYATREGDYVLFHHEGGVDVGVDAKAQKLLGVDEKLNPEDIKKHLLVHAPEDKKEILASFISGLFNFYEDLYFTYLEINPLVVTKGTVVLDLAAKVDATAFYICKVKWGDIEFPFFGREAYPEEAYIADLDAKSGASLKLTLNPGRITMVAGGGASVVSNTICDLGGVNELANYGEYSGAPSEQQTYYDIAKTIISLMTRKHPDGKILIIIGGSIANFTNVAATFKGIVRAIRDYQGPLKEHEVTIFVRGGPNYQEGLRVMGEVGKTTGIPIHVFGTETHMTAIVGMALGHRIPIPQNPPTAAHTANFLNASGSTSTPAPSRTASFSESSRADEVAPAKKAKPAMPQDSVPSPRSLSQGKSTTLSRHTKAIWGMQTRAVQGMDFDYCSRDEPSVAAMVYPFTGDHKQKFYWHGKEILIPVFKNMADAMRKHPEVDVLINFASLRSAYDSTMETMNYAQIRTIAIAEGIPEALTRKLKKADQKGVTIIGPATVGGIKPGCFKIGNTGGMLDNILASKLYRPGSVAYVSRSGGMSNELNNIISRTTDGVYEGVAIGGDRYRGSTFMDHVLRYQDTPGVKMIVVLGEIGGTEEYKICRGIKEGRSLTPCWCIGTCATMFSSVEQFGHAGACANQASETAVAKNQALKAEAVFVPRSFDELGEIIQSVEVDLVANGIVVPAQEVPVPPPTVPMDSWARELGLIRKPASFMTSICDERGQELIYAGMPITEVFKEEMGIGGVGLGLWFQKRLPKYSQFIELMCLMVTAHDHPAVSGAHNTIICARAGKDLVSSLTSGLLTGDRFGGALAAAAMFSKAFDSGIIIPMEFVNKMKEGKLMIGIGHRVKSINNPDMRVQILKDYVRQHFATPLLDYALEVEKITTSKKPNLILNVNDGLIGVAFVDMRLRNCGSTREEADEYIDIGALNGIFVLRSMGFIGHYLDQKRLKQGLYRHPWDISYVLPEHMSM

Acetyl-CoA Acetyltransferase (atoB)

MKNCVIVSAVRTAIGSFNGSLASTSAIDLGATVIKAAIERAKIDSQHVDEVIMGNVLQAGLGQNPARQALLKSGLAETVCGFTVNKVGCGSLKVALAAQAIQAGQAQSIVAGGMENMSLAPYLLDAKARSGYRLGDGQVYDVILRDGLMCATHGYHMGITAENVAKEYGITREMQDELALHSQRKAAAIESGAFTAEVIPVNVVTRKKTFVFSQDEFPKANSTAEALGALRPAFDKAGTVTAGNASGINDGAAALVIMEESAALAAGLTLPLARIKSYASGGVPPALMGMGPVATQKALQLAGLQLADIDLIEANEAAFAQFLAVGKNLGFDSEKVNVNGGAIALGHPIGASGARILVTLHAMQARDKTLGLATCIGGGQQIAMVIERLN

3-Hydroxybutyryl-CoA Dehydrogenase

MKKVCVIGAGTMGSGIAQAFAAKGFEVLRDIKDEFVDRGLDFINKNLSKLVKKGKIEEATKVEILTRISGTVDLNMAADCSDLVIEAAVERMDIKKQIFADLDNICKPETILASNTSSLSITEVASATKRPDVKIGMHFFNPAPVMKLVEVIRGIATSQETFDAVKETSIAIGKDPVEVAEAPGFVNRILIPMINEAVGILAEGIASVEDIDKAMKLGANHPMGPLEGDFIGLDICLAIMDVLYSETGDSKYRPTHLLKKYVRAGWLGRKSGKGFYDYSK

Enoyl-CoA Hydratase

MELNNVILEKEGVAVVTINRPKALNALNSDTLKEMDYVIGEIENDSEVLAVILTGAGEKSFVAGADISEMKEMNTIEGRKFGILGNKVFRRELLEKPVIAAVNGFALGGGCEIAMSCDIRIASSNARFGQPEVGLGITPGFGGTQRLSRLVGMGMAKQLIFTAQNIKADEALRIGLVNKVVEPSELMTAKEIANKIVSNAPVAVKLSKQAINRGMQCDIDTALAFESEAFGECFSTEDQKDAMTAFIEKRKIEGFKNR

Trans-Enoyl-CoA Reductase

MIVKPMVRNNICLNAHPQGCKKGVEDQIEYTKKRITAEVKAGAKAPKNVLVLCNSGYGLASRITAAGFYGAATIGVSFEKAGSETKYGTPGWYNNAFDEAAKREGLYSVTIDGDAFSDEIKAQVIEEAKKKGIKFDLIVYSLASPVRTDPDTGIMHKSVLKPFKGKTFGKTVDPFTGELKEISAEPANDEEAATVKVMGGEDWERWIKQLSKEGLLEEGCITLAYSYIGPEATQALYRKGTIGKAKEHLEATAHRLNKENPSIRAFVSVNKGTVTRASAVIPVIPLYLASFVKVMKEKGHNHEGCIEQITRLYAERLYRKDGTVDEENRIRIDDWELEEDVQKAVSALMEKVTGENAESLTDLAGYRHDFLASNGFDVEGINYEAEVERFDRI

Figure 6A (continued)**Beta-Ketothiolase (bktB)**

MTREVVVSGVRTAIGTFFGSLKDVAPEL GALVVREALARAQVSGDDVGHVVFGNVIQTEPRDMYLG RVA AVNG
GVTINAPALTNR LCGS GLQAIVSAAQ TILLG DTDVAIGGGAESMSRAPYLAPAARWGARMGDAGLVDMMLGALHD
PFHRIHMGVT AENV AKEYDISRAQQDEAALES HRRASA AIAKAGYFKDQIVPVVSKGRKG DVT FDT DEHVRHDATID
MTKLRP VFVKENGTV TAGNASGLN DAAA VMMERA EERRGLKPLARL VS YGHAGVDPKAMGIGPVPA T KIALER
AGLQVSDL DVIEANE AFAAQACAVTKALGLDPAKVNPN GSGISLGHPIGATGALITVKA LHELN RVQGRY ALVTMCIG
GGQGIAIFERI

HMG-CoA Synthase

MKLSTKLCWCGIKGRLRPQKQQQLHNTNLQMTELKKQKTAEQKTRPQNVGIGI QIYIPTQCVNQSELEKFDGV SQ
GKYTIGLQTNMSFVNDR EDIYSMSLT VLSKLIKS YNIDTNKIGRLEVGTETLIDSKSVKS VLMQLFGE NT DVEGIDTL
NACYGGTNALFNSLNWIESNAWDGRDAIVVCGDIA YDKGAARPTGGAGTVAMWIGPD APIVFD SVRAS YMEHAYD
FYKP DFTSEYPYDGHFSLTCYVKALDQVYKSYSKKAISKGLVSDPAGSDALNVLYFDY NVFHPTCKLVT KSYGR
LLYNDFRANPQLFP EVDAELATRDYDES LTDK NIEKT FVN VAKPFI KERVAQS LIVPTNTGNM YTASVYAAFASLLNY
VG SDDLQGKRVGLFSYGS GLAASLYSCKIVGDVQHII KELD ITN KLA KRIT ETPKD YEA AIELRENAHLKKNFKPQGSIE
HLQSGVYYLTNI DDKFRRSYDVKK

Truncated HMG-CoA Reductase

MVA VRRKALSILAEAPV LASDR LPYKNYDYDRVFGACCEN VIGYMPLPGVIGPLVIDGTSYHIPM ATEGCLV ASAM
RGCKAINAGGGATTVLT KDG MTRGPV VRFP TLKRS GACKI WLDSEEGQNAIKAFN STSRFARLQHI QTCLAGD LLF
MRFR TTG DAMGM NMIS KGVEYSLKQMVEEY GWEDMEV SVSG NYCTD KKP AAINWIEGRG KSV VAEATIPGDVV
RKVLKSDV S ALVEL NIAKNL VG SAMAGS VGG FNHA ANL VTA VFL ALGQD PAQN VESSNCIT LMKEV DGLR ISVSM
PSIEVGTIGGGTVLE PQGAM LDLLGVRGP HATAP GTN ARQLARIVACAVLAG ELS CAALAAGH LVQSHM THNR

Mevalonate Kinase

MSLPFLTSAPGKVII FGEHS A VYNKP AVAASV SALRTYLLI SESSAPDTIELDFPD ISFNHKWSINDFNAITEDQVNSQK
LAKAQQATDGLSQELV SLLDPLLAQLSES FHYH AFCFLYMFV CLCPHAKNIFSLKSTLP IAGLGSSASISVSLALA
MAYLGG LIGS NDLEK LSENDK HVNQWA FIGE KCIHGTPSGIDNAVATYGN ALF EKD SHNGTINTNNFKFLDDF PAIP
MILTYTRIPRSTKDL VARV RVLV TEKFP EVMKPILDAMGECALQGLEIM T LSKCKGT DDEAVETNN ELYEQLLEIRIN
HGLLVSIGVSHPGLELIK NLSDD LRIGSTKLTGAGGGGCSL LRR RDITQE QIDS FKKLQDDFSYETFETD LGGTGCC
LLSAKNLNKDLKIKSLVFQ LFENKTTKQQID DLLPGNTNLPWTS

Phosphomevalonate Kinase

MSELRAFSAPGK ALLAGGYLV LDTKYE AFV VGLS ARM HA VAPY GSLQGSDKF E V RVKS KQFKD GEW LYHIS PKSG
FIPVSIGGSKNP FIEKVIANVFSYFKPNMDYCNRNLFVIDIFSDDAYHSQEDSVTEHGRN RLSF HSHRIEEVPKTGL
GSSAGGI LVT VLT TA LASFFV SDLENNV DKYREV IHNLAQV AH CQA QGKIGSGFD VAA AYGSIRY RRFPPA LISNL
PDIGSATY GS KLA HL VDEEDW NITIKSNH LPSGLT WMGDI KNGSETV KVQ KVKNWYD SHMPESLKIY TELDHANS
RFMDGLSKLDR LHETHDDYSDQIFESLERNDCTCQKYPEITEVRDAVATIRR SFRK ITKESGADIEPPVQTSL LDCQ
TLKGVL T CLIPGAGGYDAIAVITKQD VDLRA QTANDKRF SKVQWL DVTQ ADW GVR KEKD PETY LDK

Figure 6A (continued)**Diphosphomevalonate Decarboxylase**

MTVYTASVTAPVNIAITLKYWGKRDTKLNLPNTNSSISVTLSQDDLRLTSAATAPEFERDTLWLNGEPhSIDNERTQNC
LRDLRQLRKEMESKDASLPTLSQWKLHIVSENNPTAAGLASSAAAGFAALVSAIAKLYQLPQSTSEISRIARKGSGSA
CRSLFGGYVAWEMGKAEDGHDSMAVQIADSSDWQPQMVKACVLVVSDIKKDVSSQGMQLTVATSELFKERIEHVP
KRFEVMRKAIVEKDFATFAKETMMDNSNSFHATCLDSFPPIFYMNDTSKRIISWCHTINQFYGETIVAYTFDAGPNAVL
YYLAENESKLFAFIYKLFGSVPGWDKKFTTEQLEAFNHQFESSNFTARELDLELQKVARDVILTVQVGSGPQETNESLI
DAKTGLPKE

Isopentenyl-Diphosphate Delta-Isomerase

MTADNNSPMPHAGAVSSYAKLVQNQTPEDILEEFPEIPLQQRPNTRSSSETNDESGETCFSGHDEEQIKLMNENCIVL
DWDDDNAIGAGTKKVCHLMENIEKGLLHRAFSVFIFNEQGEELLQQRATEKITFPDLWTNTCCSHPLCIDDELGLKGKL
DDKIKGAITA AVR KLD H E L G I P E D E T K T R G K F H F L N R I H Y M A P S N E P W G E H E I D Y I L F Y K I N A K E N L T V N P N V N E V R D F
KWVSPNDLKTMFADPSYKFTPWFKICENYLFNWWEQLDDLSEVENDRQIHRL

Geranyl-Diphosphate Synthase (ERG20^{WW})

MEAKIDELINNDPVWSSQNESLISKPYNHILLPGKNFRNLIVQINRVMNLPKDQLAIVSQIVELLHNSSLIDDEDNA
PLRRGQTTSHLIWGVPTINTANYMYFRAMQLVSQLTKEPLYHWLITIFNEELNLHRGQQLDIYWRDFLPEIPTQE
MYLNVMVMKTTGGLFRLTLRMLMEALSPSSHGHSLVPFNFNLGIYQIRDDYLNKDFQMSSEKGFAEDITEGKLSFPIV
HALNFTKTKGQTEQHNEILRILLRTSDKDILKLIQILEFDTNSLAYTKNFINQLVNMIKNDNENKYLPLDASHSDTATN
LHDELLYIIDHLSEL

Olivetol Synthase

MNHLRAEGPASVLAIGTANPENILLQDEFPDYYFRVTKSEHMTQLKEKFRKICDKSMIRKRNCFNLNEEHLKQNPRLVE
HEMQTLDARQDMLVVVPKLGKDACA KAI KEW G QPK SKITHLIFT SASTTDMPGADYHCACKLGLSPSVKRVMMYQ
LGCYGGGTVLRIAKDIAENNKGARVLAVCCDIMACLFRGPSES DLELLVGQAIFGDGAAVIVGAEPDESVERPIFE
LVSTGQTILPNSEGTTIGGHIREAGLIFDLHKDVPMLISNNIEKCLIEAFTPIGSDWNSIFWITHPGGKAILDKVEEKLHLK
SDKFVDSRHVLSEHGNMSSSTVLFVMDELKRSLEEGKSTTGDFEWGVLF GFGPGLTVERVVVRSPVIKY

Olivetolic Acid Cyclase

MAVKHLIVLKFKDEITEAQKEEFFKTYVNLVNIIPAMKDVTQKNKEEGYTHIVEVTFESVETIQDYIIHPAHVG
FGDVYRSFWEKLLIFDYTPRK

CBGA Synthase

MGLSSVCTFSQTNYHTLLNPHNNNPKTSLLCYRHPKTPIKYSYNFNSKHCASTKSFLQNKCSSESLSIAKNSIRAAT
TNQTEPPESDNHSVATKILNFGKACWKLQRPYTIIAFTSCACGLFGKELLHNTNLISWSLMFKAFFFLVAILCIASFTTI
NQIYDLHIDRINKPDPLPLASGEISVNTAWIMSIIVALFGLIITIKMKGGPLYIFGYCFGIFGGIVYSVPPFRWKQNPSTAFL
LNFLAHII TFTF YY ASRA ALGLPFELRPSFTL AFM KSM GSAL ALI K DAS D VEG DT K FG I ST L AS KY G SR NL TL F C SG
IVLLSYVAAILAGIWIWPQAFNSNVMLLSHAILAFWLILQTRDFALTNYDPEAGR RFYEFMWKLYYA EYL VY VFI

Figure 6A (continued)

Acetyl-CoA Carboxylase

MSEESLFESSPQKMEYEITNYSERHTELPGHFIGLNTVDKLEESPLRDFVKSHGGHTVISKILIANNGIAAVKEIRSVRK
 WAYETFGDDRTVQFVAMATPEDLEANAEIYIRMAQYIEVPGGTNNNNYANVLDIVIAERADVDAWAGWGHASE
 NPLLPEKLSQSQRKRVFIGPPGNAMRSLGDKISSTIVAQSAKVPICIPWSGTGVDTVHVDEKTGLVSVDDDIYQKGCT
 SPEDGLQKAKRIGFPVMIKASEGGGGKGIRQVEREEDFIALYHQAANEIPGSPIFIMKLAGRARHLEVQLLADQYGTNI
 SLFGRDCSVQRHHQKIEEAPVTIKAETFHEMEKAAVRLGKLVGYVSAGTVEYLYSHDDGFYFLELNPRLQVEHP
 TTEMVSGVNLPAALQIAMGIPMHRISDIRTLGGMNPHSASEIDFEFKTQDATKKQRRPIPKGHCTACRITSEDPN
 FKPSGGTLHELNFRSSNVWYFSVGNNGNIHSFSDSQFGHIFAFGENRQASRKHMVALKELSRGDFRTTVEYL
 KLLTEDFEDNTITTGWLDLJTHKMTAEKPDPPLAVICGAATKAFLASEEARHKYIESLQKGQVLSKD
 LQTMFPVDF
 IHEGKRYKFTVAKSGNDRTLTFINGSKCDIILRQLSDGGLIAIGGKSHTIYWKEEVAATRLSVDSMTTLEVEN
 DPQL
 RTPSPGKLVKFLVNGEHIKGQPYAEIEMKMQMPLVSQENGIVQLLKQPGSTIVAGDIMA
 MLDDPSKVHALPFE
 GMLPDFGSPVIEGTKPAYFKSLVSTLENILKGYDNQVMNASLQQLIEVRNPKLPYSEWLHISALHSRLPAKLDEQ
 MEELVARSLRRGAVFPARQLSKLDMAVKNPEYNPDKLLGAVVPLADIAHKYSNGLEAHESIFVHFLEYYEVEKL
 FNGPNVREENIIKLRDENPKDLKVALTVLSHSKVSAKNNLILAIKHYQPLCKLSSKVSAFSTPLQHIVELESKATAK
 VALQAREILIQGALPSVKERTEQIEHILKSSVVKAVYGSNPKRSEPDNLKLDLSNYYVFDVLLQFLTHQDPV
 VTA
 AAQVYIRRAYRAYTIGDIRVHEGTVPIEWKFQLPSAAFSTFPTVSKMGMRASVSDLSYVANSQSSPLREGILM
 AVDHDDVDEILSQSLEVPRHQSSNSGPAPDRSGSSASLSNVANCVASTEGFESEEEILVRLREILDLNKQELINAS
 IRRITFMFGFKDGSPKYYTFNGPNYNEINETIRHIEPALAFQLELGRLSNFNPKIPTDNRNIHVYEAVSKTSPLDKRFF
 TRGIIRTGHIRDDISIQEYLTS
 SEANRLMSDILDNLEVTDTNSDNLHIFINFI
 AVFDISPEDVEAAFGGLERFGKRLRLR
 VSSAEIRIIKDPQTGAPVPLRALINNVSGYVI
 KTEMYTEVKNAKG
 EWFVFKSLKGPMH
 LRP
 IATPYPVKEWLQPKRY
 KAHLMGTTVYDFPELFRQASSQWK
 NFSADV
 KLTD
 DFFIS
 NELIEDENGELTE
 VEREPGAN
 AIGMVA
 FKITV
 KTP
 PEY
 PRGRQFVVVANDITFKIGSF
 GPKQ
 DEEFFNK
 VTEYARKR
 GIPRIYLA
 ANSGAR
 IMAEEIVPL
 QVA
 WND
 AANPD
 KGF
 QYLYLTSEG
 METLKK
 FD
 KENS
 VLT
 ERT
 VING
 GE
 ERF
 VIK
 TIG
 SED
 GLG
 VEC
 LRG
 GSI
 AGAT
 S
 RAY
 HD
 IT
 IT
 L
 VTC
 CSV
 GIGAYL
 VRL
 GQRA
 I
 QV
 EG
 QP
 II
 LT
 GAP
 A
 IN
 KML
 GRE
 VTS
 NL
 QL
 GG
 TQ
 IM
 YN
 NG
 V
 SH
 TA
 V
 D
 D
 LAG
 VE
 K
 I
 V
 E
 W
 M
 SY
 V
 P
 AKRNMPV
 ILET
 KDT
 WDR
 P
 D
 F
 PT
 N
 D
 E
 T
 Y
 D
 VR
 W
 M
 I
 E
 G
 R
 E
 T
 S
 G
 F
 EY
 GL
 F
 DK
 G
 S
 F
 F
 E
 T
 L
 S
 G
 W
 A
 K
 G
 V
 V
 V
 G
 R
 A
 R
 L
 G
 G
 I
 PL
 G
 V
 I
 G
 V
 E
 T
 R
 T
 V
 E
 N
 L
 I
 P
 A
 D
 P
 A
 N
 P
 S
 A
 E
 T
 L
 I
 Q
 E
 P
 G
 Q
 V
 W
 H
 P
 N
 S
 A
 F
 K
 T
 A
 Q
 A
 I
 N
 D
 F
 N
 G
 E
 Q
 L
 P
 M
 I
 L
 A
 N
 W
 R
 G
 F
 S
 G
 Q
 R
 D
 M
 F
 N
 E
 L
 V
 K
 Y
 R
 D
 D
 K
 Y
 R
 E
 L
 R
 S
 Q
 L
 S
 N
 K
 L
 A
 D
 R
 E
 R
 E
 L
 L
 P
 I
 Y
 G
 Q
 I
 S
 L
 Q
 F
 A
 D
 L
 H
 D
 R
 S
 S
 R
 M
 V
 A
 K
 G
 V
 I
 S
 K
 E
 L
 E
 W
 T
 E
 A
 R
 R
 R
 F
 F
 W
 R
 L
 R
 R
 R
 L
 N
 E
 E
 Y
 L
 I
 K
 R
 L
 S
 H
 Q
 V
 G
 E
 A
 S
 R
 L
 E
 K
 I
 A
 R
 I
 R
 S
 W
 Y
 P
 A
 S
 V
 D
 H
 E
 D
 D
 R
 Q
 V
 A
 T
 W
 I
 E
 N
 Y
 K
 T
 L
 D
 D
 K
 L
 G
 L
 K
 L
 E
 S
 F
 A
 Q
 D
 L
 A
 K
 I
 R
 S
 D
 H
 D
 N
 A
 I
 D
 G
 L
 S
 E
 V
 I
 K
 M
 L
 S
 T
 D
 D
 K
 E
 K
 L
 L
 K
 T
 L
 K

CBDA Synthase

MKCSTFSFWFVCKIIFFFSFNIQTSIANPRENFLKCFSQYIPNNATNLKLYTQNNPLYMSVLNSTIHNLRFTSDTPK
 PLVIVTPSHVSHIQGTILCSKKVGLQIRTRSGGDSEGMYSIISQVPPVIVDLRNMRSIKIDVHSQTAWVEAGATLGEVY
 YW
 NE
 K
 N
 E
 N
 L
 S
 L
 A
 A
 G
 Y
 C
 P
 T
 V
 C
 A
 G
 G
 H
 F
 G
 G
 G
 G
 Y
 G
 P
 L
 M
 R
 N
 Y
 G
 L
 A
 A
 D
 N
 I
 I
 D
 A
 H
 L
 V
 N
 V
 H
 G
 K
 V
 L
 D
 R
 K
 S
 M
 G
 E
 D
 L
 F
 W
 A
 I
 R
 G
 G
 G
 A
 E
 S
 F
 G
 I
 I
 V
 A
 W
 K
 I
 R
 L
 V
 A
 V
 P
 K
 S
 T
 M
 F
 S
 V
 K
 K
 I
 M
 E
 I
 H
 E
 L
 V
 K
 L
 V
 N
 K
 W
 Q
 N
 I
 A
 Y
 K
 Y
 D
 K
 D
 L
 L
 M
 T
 H
 F
 R
 T
 R
 N
 I
 T
 D
 N
 Q
 G
 K
 N
 K
 T
 A
 I
 H
 T
 Y
 F
 S
 S
 V
 F
 G
 G
 V
 D
 S
 L
 D
 L
 M
 N
 K
 S
 F
 P
 E
 L
 G
 I
 K
 K
 T
 D
 C
 K
 E
 L
 S
 W
 I
 D
 T
 I
 F
 Y
 S
 G
 V
 V
 N
 Y
 T
 D
 D
 F
 N
 K
 E
 I
 L
 L
 D
 R
 S
 A
 G
 K
 K
 T
 A
 F
 S
 I
 K
 L
 D
 Y
 V
 K
 K
 L
 I
 P
 E
 T
 A
 M
 V
 K
 I
 L
 E
 K
 L
 Y
 Y
 E
 E
 E
 V
 G
 V
 G
 M
 Y
 V
 L
 Y
 P
 Y
 G
 G
 I
 M
 D
 E
 I
 S
 E
 A
 I
 P
 P
 F
 P
 H
 R
 A
 G
 I
 M
 Y
 E
 L
 W
 Y
 T
 A
 T
 W
 E
 K
 Q
 E
 D
 N
 E
 K
 H
 I
 N
 W
 R
 I
 N
 Y
 N
 F
 M
 T
 P
 Y
 V
 S
 Q
 N
 P
 R
 L
 A
 Y
 L
 N
 Y
 R
 D
 L
 D
 L
 G
 K
 T
 N
 P
 E
 S
 P
 N
 N
 Y
 T
 Q
 A
 R
 I
 W
 G
 E
 K
 Y
 F
 G
 K
 N
 F
 D
 R
 L
 V
 K
 V
 K
 T
 K
 A
 D
 P
 N
 N
 F
 R
 N
 E
 Q
 S
 I
 P
 P
 L
 P
 R
 R
 H
 H

CBCA Synthase

MNCSTFSFWFVCKIIFFFSFNIQTSIANPQENFLKCFSEYIPNNPANPKFIYTQHDQLYMSVLNSTIHNLRFTSDTPK
 PLVIVTPSNVSHIQASILCSKKVGLQIRTRSGGDSEGMYSIISQVPPVIVDLRNMRSIKIDVHSQTAWVEAGATLGEVY
 YW
 W
 I
 N
 E
 M
 N
 E
 N
 L
 S
 L
 A
 A
 G
 Y
 C
 P
 T
 V
 C
 A
 G
 G
 H
 F
 G
 G
 G
 G
 Y
 G
 P
 L
 M
 R
 N
 Y
 G
 L
 A
 A
 D
 N
 I
 I
 D
 A
 H
 L
 V
 N
 V
 H
 G
 K
 V
 L
 D
 R
 K
 S
 M
 G
 E
 D
 L
 F
 W
 A
 I
 R
 G
 G
 G
 A
 E
 S
 F
 G
 I
 I
 V
 A
 W
 K
 I
 R
 L
 V
 A
 V
 P
 K
 S
 T
 M
 F
 S
 V
 K
 K
 I
 M
 E
 I
 H
 E
 L
 V
 K
 L
 V
 N
 K
 W
 Q
 N
 I
 A
 Y
 K
 Y
 D
 K
 D
 L
 L
 M
 T
 H
 F
 R
 T
 R
 N
 I
 T
 D
 N
 Q
 G
 K
 N
 K
 T
 A
 I
 H
 T
 Y
 F
 S
 S
 V
 F
 G
 G
 V
 D
 S
 L
 D
 L
 M
 N
 K
 S
 F
 P
 E
 L
 G
 I
 K
 K
 T
 D
 C
 K
 E
 L
 S
 W
 I
 D
 T
 I
 F
 Y
 S
 G
 V
 V
 N
 Y
 T
 D
 D
 F
 N
 K
 E
 I
 L
 L
 D
 R
 S
 A
 G
 K
 K
 T
 A
 F
 S
 I
 K
 L
 D
 Y
 V
 K
 K
 L
 I
 P
 E
 T
 A
 M
 V
 K
 I
 L
 E
 K
 L
 Y
 Y
 E
 E
 E
 V
 G
 V
 G
 M
 Y
 V
 L
 Y
 P
 Y
 G
 G
 I
 M
 D
 E
 I
 S
 E
 A
 I
 P
 P
 F
 P
 H
 R
 A
 G
 I
 M
 Y
 E
 L
 W
 Y
 T
 A
 T
 W
 E
 K
 Q
 E
 D
 N
 E
 K
 H
 I
 N
 W
 R
 I
 N
 Y
 N
 F
 M
 T
 P
 Y
 V
 S
 Q
 N
 P
 R
 L
 A
 Y
 L
 N
 Y
 R
 D
 L
 D
 L
 G
 K
 T
 N
 P
 E
 S
 P
 N
 N
 Y
 T
 Q
 A
 R
 I
 W
 G
 E
 K
 Y
 F
 G
 K
 N
 F
 D
 R
 L
 V
 K
 V
 K
 T
 K
 A
 D
 P
 N
 N
 F
 R
 N
 E
 Q
 S
 I
 P
 P
 L
 P
 R
 R
 H
 H

Figure 6A (continued)**Hexanoyl-CoA Synthetase**

MGKNYKSLDSVVASDFIALGITSEVAETLHGRЛАEIVNYGAATPQTWINIANHILSPDLPFLHQMLFYGCYKDFGPA
PPAWIPDPEKVKSTNLGALLEKRGKEFLGVKYKDPISSFSHFQEFSVRNPEVYWRVLMDEMKSFSKDPECILRRD
DINNPGGSEWLPGGYLNSAKNCLNVNSNKLNNDTMIVRDEGNDLPLNKLTLDQLRKRVWLVGYALEEMGLEKG
CAJAIDMPMHDAVVYLAIVLAGYVVVSIADFSAPAEISTRRLRSKAKAIFTQDHURGKKRIPLYSRVVEAKSPMAIVP
CSGSNIGAELRDGDISWDYFLERAKEFKNCEFTAREQPVDAYTNILFSSGTTGEPKAIPWTQATPLKAAADGWSHLD
IRKGDVIVWPTNLGWMMGPWLVYASLLNGASIALYNGSPLVSGFAKFVQDAKVTMLGVVPSIVRSWKSTNCVSGYD
WSTIRCFSSSGEASNVDEYLWLMGRANYKPVIEMCGTEIGGAFSAGSFLQAQSLSFFSSQCMGCTLYILDKNGYP
MPKNKPGIGELALGPVMFGASKTLLNGNHHHDVYFKGMPTLNGEVLRRHGDIFFELTSNGYYHAHGRADDTMNIGGIKI
SSIEIERVCNEVDDRFETTAIGVPPLGGGPEQLVIFFVLKDSNDTTIDLNQLRLSFNLGLQKKLNPLFKVTRVVPLSSL
PRTATNKIMRRVLRQQFSHFE

Figure 6B

ATP Citrate Lyase – Enzyme Linker – cTPR6 Spacer – ID Linker – ID1

Acetyl-CoA Acetyltransferase (atoB) – Enzyme Linker – cTPR6 Spacer – ID Linker – ID2

MKNCVIVSAVRTAIGSFNGSLASTSAIDLGATVIKAAIERAKIDSQHVDEVMGNVLQAGLGQNPARQALLKSGLAETV
CGFTVNKGCGSGLKSVLAQQAQIAGQQAQSIVAGGMENMSLAPYLLDAKARSGYRLGDGQVYDVIILRDGLMCATH
GYHMGITAENVAKEYGITREMQDELALHSQRKAAAIESGAFTAIEVPNVTRKKTFVFSQDFPKANSTAEALGAL
RPAFDKAGTVTAGNAGINDGAAALVIMEESAALAAGTPLARIKSAYASGGVPPALMGMPVPATQKALQLAGLQLA
DIDLIEANEAFAAQFLAVGKNLGFDFSEKVNVNGGAIALGHPIGASSGARILVTLLHAMQARDKTLGLATLCIGGGQGIAM
VIERLNKLSGGGGGGGGGGGGSAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPPNNAEAWYNLGNAYYKQGDY
YQKAIEYYQKALELDPPNNAEAWYNLGNAYYKQGDYQKAIEDYQKALELDPPNNLQAEAWKNLGNAYYKQGDYQKAIE
YYQKALELDPPNNASAWYNLGNAYYKQGDYQKAIEYYQKALELDPPNNAKAWYRRNGNAYYKQGDYQKAIEDYQKALE
LDPPNNRSRSAGGGGGGGGGGGGGGASSYYHHHHHHLESTSLYKKAGSGSNEVTTLENDAAFIENENAYLEKEIAR
LRKEKAALRNRLAHKKGSAGSAAGSGEFGSAEAAKEAAKAGSAGSAGSGEFGSSYYHHHHHHLESTSLYKK
GSGSQVKAELKNRVAVKLNRNEQLKNKVEELKNRNAYLKNELATLENEVARLENDVAE

3-Hydroxybutyryl-CoA Dehydrogenase – Enzyme Linker – cTPR6 Spacer – ID Linker – ID3

MKKVCVIGAGTMSGQIAQAFAAKGFEVVLRIKDEFVDRGLDFINKNLSKLVKKGKIEEATKVIELTRISGTVDLNMAACDCL
VIEAAVERMDIKKQIFADLDNICKPETILASNNTSSL SITEVASATKRPDKVIGHFFNPAPVMKLV E VIRGIATQSQETFDAVKE
SIAIGKDPVEAEAPGFVNRLIPMINEAVGILAEGIASVEDIDKAMKLG ANHPMGLELGDFIGLDICLAIMDVLYSETGDSK
YRPHTLLKKYVRAGWLGRKSGKGFYDYSKLLSGGGGGGGGGGGSAEAWYNLGNAYYKQGDYQKAIEYYQKAEL
DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKAELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKAELDPNNLQA E AW
KNLGNAYYKQGDYQKAIEYYQKAELDPNNASAWYNLGNAYYKQGDYQKAIEYYQKAELDPNNAKAWYRRRGNAYYKQ
GDYQKAIEYYQKAELDPNNRSRSAGGGGGGGGGGGGGASENLYFQGENLYFQGDSSESCWCGRKASETCGCNT
ARYCGSCFCQHKDWKHHICGQLTQAQQGSAGSAAGSFGSAAEAAAKEAAAKAGSAGSAAGSFGESGMVA SESQLK
KMVSKYKYRDLTVRETNVITLYKDLKPVLDSYVFNDGSRELMLNTGTIPVYRGNTYNIPICLWLLDTYPYNPPICFVKP
SSMTIKTGKHDANGKIYLPHEWKHPQSDLLGLIQVMIVVFGDEPPVFSRP

Figure 6B (continued)

Enoyl-CoA Hydratase – Enzyme Linker – cTPR6 Spacer – ID Linker – ID4

MELNNVILEKEGVAVVTINRPKALNALNSDTLKEMDYVIGEIENDSEVLAVILTGAGEKSFGADISEMKEMNTIEG
RKFGILGNKVFRRLLELPVIAAVNGFALGGGCEIAMSCDIRIASSNARFGQPEVGLGITPGFGGTQRQLSLRVGMGM
AKQLIFTAQNIKADEALRIGLVNKVEPSELMTAKEIANKIVSNAPVAKLSKQAINRGMQCDIDTALAFSEAFGECEF
STEDQKDAMTAFIEKRKIEGFKNRKLSSGGGGGGGGSAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPN
NAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEDYQKALELDPNQLQAEAW
KNLGNAYYKQGDYQKAIEYYQKALELDPNNASAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAKAWYRRGNAY
YKQGDYQKAIEDYQKALELDPNNRSRSAGGGGGGGGGGGASGPLGSPLTASMLASAPPQEQQKQLGERLFP
LIQAMHPTLAGKITGMLLEIDNSELLLHMLESPESLRSKVDEAVAVLQAHQAKEAAQKAGSAGSAAGSGEFGSAEAAA
KEAAAAGSAGSAAGSGEFGNSTNMSVPTDGAVTTSQIPASEQETLVRPKPLLKLSVGAQKDTYTMKEVLFYLG
QYIMTKRLYDEKQQHIVYCSNDLLGDLFGVPSFSVKEHRKIYTMIYRNLLV

Trans-Enoyl-CoA Reductase – Enzyme Linker – cTPR6 Spacer – ID Linker – ID5

MIVKPMVRNNICLNAHPQGKKGVEDQIETYTKKRITAEVKGAKAPKNVLGCSNGYGLASRITAAGFYGAATIGVS
FEKAGSETKYGTPGWYNNLAFDEAAKREGLYSVTIDGDAFSDEIKAQVIEEAKKKGIKFDLIVYSLASPVRTDPDTGIM
HKSVLKPFKGKTFGTVDPTGELKEISAEPANDEEAAATVKVMGGEDWERWIKQLSKEGLLEEGCILAYSYIGPEA
TQALYRKGTIGKAKEHLEATAHRLNKENPSIRAFSVNKGLVTRASAIPVIPLYLASFVKMKEGNHEGCIEQITRLY
AERLYRKDGTTIPVDEENRIRIDDWELEEDVQKAVSALMEKVGENAEESLTDLAGYRHDFLASNGFDVEGINYEAEVE
RFDRIKLSGGGGSGGGSGGGSAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDY
QKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEDYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEY
YQKALELDPNNASAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEDYQKAEL
DPNNRSRSAGGGGGGGGGGGGGGASSYYHHHHHHLESTSLLYKAGGSNLLATLRSTAVALLENENHVLEKEKEKL
RKEKEQLLNKLEAYKGSAGSAGSGEFGSAEAAAEEAKAGSAGSAGSGEFGSSYYHHHHHHLESTSLLYKAG
SGSKRIAYLRKKIAALKDNDANLEKDIANLENEIERLIKEIKTLENEVASHEQ

Beta-Ketothiolase (bktB) – Enzyme Linker – cTPR6 Spacer – ID Linker – ID6

MTRVVVVSGVRTAIGTFFGSSLKDVAPEL GALVREALARAQVSGDDVGHVFGNVIQTEPRDMYLGRVAAVNG
GVTINAPALTNVRLCGSGLQAIQSAAQ TILLGDTDVAIGGGAESMSRAPY LAPAARW GARMGDAGLVDMMLGALHD
PFHRIHMVT AENVAK EYDISRAQQDEA ALES HRRASAAIKAGYFKDQIVPVSKGRKG DVT FDT DEHVRHDATIDD
MTKLRP VFVKENGTVTAGN ASGLNDAAA VVM MERA EAERR GLKPLARL VS YGHAGVDPKAMIGPVPA T KIALER
AGLQVS DLDVIEANEAF AAQAC AVTK ALGL DP KVN P NGS G ISL GHPI GAT GALITVKA LHE LN RV/Q GR YAL VTM CIG
GG QGIAAIFERIKLSSGGGSGGGGSGGGSAEAWYNLGNAYYKQGDYQKAIEYYQKAELDPNNAEAWYNLGNAY
YKQGDYQKAIEYYQKAELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKAELDPNNLQAEAWKLNQGNAYYKQGD
YQKAIEYYQKAELDPNNASAWYNLGNAYYKQGDYQKAIEYYQKAELDPNNNAKAWYRRGNAYYKQGDYQKAIED
YQKAELDPNNRSRSAGGGGGGGGGGGASDV MW EYK WENT GD AEL YGPFTSQAQM QTWT VSEGYFPDGVYC
RKLDPPGGQFYNSKRIDFDLYTGSAGSAAGSGEFGS AEEAAKAGSAGSAAGSGEFGSES DSVEFNN AISYV
NIKIKTRFLDHPEIYRSFLEIHTYQKEQLHTKGRPF RGMSEEEVFTEVANLFRGQEDLLSEFGQFLPEAKR

Figure 6B (continued)

HMG-CoA Synthase – Enzyme Linker – cTPR6 Spacer – ID Linker – ID7

MKLSTKLCWCGIKGRLRPQKQQQLHNTNLQMTELKKQKTAEQKTRPQNVGIGIYIPTQCVNQSELEKFDGVSQ
GKYTIGLGGTNMSFVNDRDIYSMSLTVLSKLIKSYNIIDTNKIGRLEVGTETLIDKSKSVKSLMQLFGENTDVEGIDL
NACYGGTNALFNLSNWIESNAWDGRDAIVVCGDIAIDKGAAARPTEGGAGTVAMWIGPDAPIVFDSVRASYMEHAYD
FYKPDFTSEYPVDGHFSLTVCYKALDQVYKSYSKKAISKGLVSDPAGSDALNVLKYFDYVFHVPTCKLVTKSYGR
LLYNDFRANPQLFPEVDAELATRDYDESLDKNEKTFVNVAKFHKERVAQSLIVPTNTGNMYTASVYAAFASLLNY
VGSSDLQKGKRVGLFSYGSGLAASLYSCKIVGDVQHIIKELDITNKLAKRITETPKDYEEAIELRENAHLKKNFKPQGSIE
HLQSGVYYLTNIDDKFRRSYDVKKKLSGGGGSGGGGGSGGGGSAEAWYNLGNAYYKQGDYQKAIEYYQKALELDP
NNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNLQAAE
WKNLGNAYYKQGDYQKAIEYYQKALELDPNNASAWYNNLGNAYYKQGDYQKAIEYYQKALELDPNNAKAWYRRGN
AYYKQGDYQKAIEYYQKALELDPNNRSRSAGGGGGGGGGGGGGASLGPLPPGWEVRSTVSGRIYFVDHNRTT
QFTDPRHLHGSAAGSGEFGSAEAAAEEKAGSAGSAAGSGEFGSGAMGPLPPGWEKRTDSNGRUYFVNH
NTRITQWEDPRS

Truncated HMG-CoA Reductase – Enzyme Linker – cTPR6 Spacer – ID Linker – ID8

MVAVRRKALSILAEAPVLASDRLPYKNYDYDRVFGACCENVIGYMPVGVIGPLVIDGTSYHIMPATTEGCLVASAM
RGCKAINAGGGATTVLTKDMTRGPVVFRTPLKRSGACKIWLDSEEGQNAKKAFNSTSRFARLQHQITCLAGDLLF
MRFRTTTGDAMGMNMISKGVEYSLKQMVVEYGWEDMEVSVSGNYCTDKPAAINWIEGRGKSVAEATIPGDVV
RKVLKSDVSALVELNIAKNLVGSAMAGS VGGFNAAHANLTVAVFLALGQDPAQNVESSNCITLMKEVDGDLRISVSM
PSIEVGTTGGGTLEPGQAMLDLLGVRGPHATAPGTMARQLARIVACAVLAGELSLCAALAAGHLVQSHMTHNRKLS
GGGGGGGGGGGGSAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYY
QKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNLQAEAWKNLGNAYYKQGDYQKAIEYYQKALE
LDPNNASAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAKAWYRRGNAYYKQGDYQKAIEYYQKALELDPNNRS
RSAGGGGGGGGGGGGASSYYHHHHHHLESTSLYKKAGSEFFRRERNKMAAKCRNRRERLTDTLQAETDQLE
DEKSALQTEIANLLKEKEKLEFILAHRPACKIPDDLGFPPEEMSLEGSAGSAGSGEFGSAAEAAKEAAKAGSAGSA
AGSGEFGSSYYHHHHHHLESTSLYKKAGSGSQKVESLKQKIEELKQRKAQLKNDIANLEKEIAYAET

Mevalonate Kinase – Enzyme Linker – cTPR6 Spacer – ID Linker – ID9

Figure 6B (continued)

Phosphomevalonate Kinase – Enzyme Linker – cTPR6 Spacer – ID Linker – ID10

MSELRAFSAPGKALLAGGYLVLDTKYEAFVVGSLARMHAVAHPYGSLOQGSDKFEVRVSKQFKDGEWLYHISPKSG
FIPVSIGGSKNPFIKEVIANVFSYFKPNMDDYCRNRLVIDIFSDDAYHSQEDSVTEHRGNRRLSFHSHRIEVPKTGL
GSSAGG_[j2]LVTVLTTALASFFVSDLENNVDKYREVIHNLAQVAHCQAQKGIGSGFDVAAAAGSIRYRRFPALISNL
PDIGSATYGSKLAHLVDEEDWNITIKSNHLPGLTLWMGDIKNGSETVKLVQKVKNWYDSHMPESLKIYTELHDANS
RFMDGLSKLDRLHETHDDYSQDFESLERNDCTCQKYPEITEVRDAVATIRRSPRKITKESGADIEPPVQTSLDDCQ
TLKGVLTCLIPGAGGYDAIAVITKQDVDLRAQTANDKRFSKVQWLDTQADWGVRKEDPETYLDKLLSGGGGG
GGSGGGGSAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDP
NNAEAWYNLGNAYYKQGDYQKAIEDYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNA
WYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEDYQKALELDPNNSRSRAGGG
SGGGGGGGGASSYYHHHHHHLESTSLYKKAGSGSQKVEELNKIAELENNAVKKNRVAHLKQEIAYLKDELAAH
EFEFGSAGSAAGSGEFGSAEAAAKEAAAKAGSAGSAAGSGEFGSSYYHHHHHHLESTSLYKKAGSGSFENVTHEFIL
ATLENENAKLRLLEAKLERELARLRNEVAWL

Diphosphomevalonate Decarboxylase – Enzyme Linker – cTPR6 Spacer – ID Linker – ID11

MTVYTASVTAPVNIAITLKYWGKRDTKLNLPTNSSISVTLSQDDLRTLTSATAPEFERDTLWLNGEPEHSIDNERTQNC
LRDLRQLRKEMESKDASLPTLSQWKLHIVSENNFTAAGGLASSAAGFAALVSIAKLYQLPQSTSEISRIARKGSGSA
CRSLFGGYVAWEMGKAEDGHDSMAVQIADSSDWPMQKACVLVSDIKDVSSTQGMQLTVATSELFKERIEHVVP
KRFEVMRKAIVEKDFATFAKETMMDSNSFHATCLDSFPPIFYMNDTSKRIISWCHTINQFYGETIVAYTFADGPNAVL
YYLAENESKLFAFIYKLFGSVPGWDKKFTTEQLEAFNHQFESSNFTARELDLELQKDVARVILTQVGSGPQETNESLI
DAKTGLPKEKLSGGGGGGGGGGGGGGSAEAWYNLGNAYYKQGDYQKAIEYYQKALEELDPNNAEAWYNLGNAYYK
QGDYQKAIEYYQKALEELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALEELDPNNAEAWYNLGNAYYKQGDYQ
KAIEYYQKALEELDPNNASA WYNLGNAYYKQGDYQKAIEYYQKALEELDPNNAKAWYRRGNAYYKQGDYQKAIEYYQ
KALEELDPNNRSRSAGGGGGGGGGGGGGGGASAMADELEQKVLEMEASTYDGVFIWKSIDFPRKRQEA VAGRIPAIFS
PAFYTSRYGYKCMCIRIYLNQDGTRGRTHLSLFFFVMKGPNDA LLRWPFNQKVTLMLLDQNNNRHVIDAFRPDVTTSS
SFQRPVNDMNIASGCPLFCPVSKMEA KNSYVRDDAIFAKIADLTGLGSAGSAA GSGEFGSAEAAAKEAAKAGSAG
SAAGSGEFGSASIKLQSSDGEIFEV DVEIAQSVTIKT MLEDLGMDDEGGDDDPVPLPNVNAIILKVIQWC THHKDPP
PPEDDENKEKRTDDIPVWDQEFLKVDQGTLFELILA NYLDIKG LLDVTCKTVANMIKGTKPEEIRKTFNIKNDFTEE
EEAQVRKENQWC

Isopentenyl-Diphosphate Delta-Isomerase – Enzyme Linker – cTPR6 Spacer – ID Linker – ID12

MTADNNNSMPHGAVSSYAKLVQNQTPEDILEEFPEIPIPLQQRPNTRSSETSNDESGETCSGHDEEQIKLMNENCIVL
DWDDNAIGAGTKVKCHLMENIEKGLLHRAFSVFIFNEQGELLLQQRATEKITFPDLWTNTCCSHPLCIDDELGLKGKL
DDKIKGAIITAAVRKLDHELGIPEDETKTRGKFHFLNRHYMAPSNEPWGEHEIDYILFYKINAKENLTVNPVNVEVRDF
KWVSPNDLKTMFADPSYKFTPWFKIICENYLFNWWEQLDDLSEVENDRQIHRMLKLSGGGSGGGGSGGGGSAE
AWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGN
AYYKQGDYQKAIEDYQKALELDPNNLQAEAWKNLGNAYYKQGDYQKAIEYYQKALELDPNNASAWYNLGNAYYKQ
GDYQKAIEYYQKALELDPNNAKAWYRRGNAYYKQGDYQKAIEDYQKALELDPNNRSRSAGGGGSGGGGSGGGGA
SSYYHHHHHLESTSLYKKAGSGSNTVKELKNYIQELEERNAELKNLKEHLKFAKAELEFELAHHKFEGSAGSAAGS
GEFGSAAAAKEAAAKAGSAGSAAGSGEFGSSYYHHHHHLESTSLYKKAGSGSQKVQLKNRVAYKLKENAKLE
NIVARLENDNANLEKDIANLEKDIANLERDVAR

Figure 6B (continued)**Geranyl-Diphosphate Synthase (ERG20^{WW}) – Enzyme Linker – cTPR6 Spacer – ID Linker – ID13**

MEAKIDELINNDPVWSSQNESLISKPYNHILLKPGKNFRLNLIVQINRVMNLPKDQLAIVSQIVELLHNSSLIDDEDNA
PLRRGQTTSHLIWGVPISTINTANYMYFRAMQLVSQLTTKEPLYHWLITIFNEELINLHRGQGLDIYWRDFLPEIPTQE
MYLNVMVMKNTGGFLRLTLRLMEALSPSSHGHGHSVPFIINLLGIYQIRDDYLNLKDFQMSSEKGFAEDITEGKLSFPIV
HALNFTKTKGQTEQHNEILRILLRTSDKDILKLIQILEFDTNSLAYTKNFINQLVNMIKNDNENKYLPLASHSDTATN
LHDELLYIIDHLSKLSGGGGSGGGGSGGGSAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLG
NAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAKAWYRRGNAYYKQGDYQK
AIEDYQKALELDPNNRSRSAGGGGSGGGGGASLCTMKKGPGSGYGFNLHSDKSKPGFIRSVDPDSPAES
GLRAQDRIVEVNGVCMEGKQHGDRVSAIRAGGDETKLVVDRREGSAGSAAGSGEFGSAAAAKEAAKAGSAGSA
AGSGEFGSSSGAIYTVELKRYGGPLGITISGTEEPFDPIISSLTKGGLAERTGAIHIGDRILAINSSLKGKPLSEAIHLL
QMAGETVTLKIKKQTDQPASS

Olivetol Synthase – Enzyme Linker – cTPR6 Spacer – ID Linker – ID14

MNHLRAEGPASVLAIGTANPENILLQDEFDPYYFRVTKEHMTQLKEKFRKICDKSMIRKRNCFLNEEHLKQNPRLV
HEMQTLDARQDMLVVEPKLKGKDACAIAKEWGQPKSKITHLIFTSTTDMPGADYHCAKLLGLSPSVKRVMMYQ
LGCYGGGTVLRIAKDIAENNKGARVLAVCCDIMACLFRGPSESDELLVGQAIQGDAAVIVGAEPDESVERPIPE
LVSTGQTILPNSEGTTGGIREAGLIFDLHKDVPMLISNNIEKCLIEAFTPIGSDWNSIFWITHPGGKAILDKVEEKLHLK
SDKFVDSRHLVSEHGNMSSSTVLFVMDCLRKSLEEGKSTTGDFEWGVLFGFPGPLTVERVVRSPVIKYKLSGG
GGSGGGGGSGGGSAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKA
LELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAKAWYRRGNAYYKQGDYQKAIEYYQKALELDP
NNASAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAKAWYRRGNAYYKQGDYQKAIEYYQKALELDPNNRSRSA
GGGGSGGGGGGGASGNLNLEYEWINKSISRDKAKEKLLDTGKEGAFMVRDSDRTPGTVSVFTKAIISENPCIK
HYHIKETNDSPKRYYYAEKYVFDSIPLIQYHQYNGGLVTRLRYPVCGGSAGSAAGSGEFGSAAAAKEAAKAG
SAGSAAGSGEFGSGSHPWFFGKIPRAKEEMLSQQRHDGAFLIRESESAPGDFSLSVKGNDVQHFKVLRDGAGK
YFLWVKFNSLNELVYHRSTSRSQQFLRDLIEQVPPQQPT

Olivetolic Acid Cyclase – Enzyme Linker – cTPR6 Spacer – ID Linker – ID15

MAVKHLIVLKKDEITEAQKEEFFKTYVNVLNIPAMKDVYWGKDVTKNKEEGYTHIVEVTFESVETIQDYIIHPAHVG
FGDVYRSPWEKLLIFDYTPRKLSGGGGGGGGGGSAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNA
EAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWKN
LGNAYYKQGDYQKAIEYYQKALELDPNNAKAWYRRGNAYYKQGDYQKAIEYYQKALELDPNNAKAWYRRGNAYYK
QGDYQKAIEYYQKALELDPNNAKAWYRRGNAYYKQGDYQKAIEYYQKALELDPNNAKAWYRRGNAYYK
RGDKLCQDSMMKLGVAGARSKGEHKQKIFLTISFGGIKIFDEKTGALQHHHAVHEISYIAKDTDHRAFGYVCGKE
GNHRFVAIKTAQAAEPVILDRLFLQIYELKQREELEKAGSAGSAAGSGEFGSAAAAKEAAKAGSAGSAAGSG
EFGSGSHMGSQFWVTSQKTEASERCGLQGSYILRVEAEKLTLTLGAQSQILEPLLFWPYTLLRRYGRDKVMFSFE
AGRRCPSGPGBTFTFQTSQGNDIFQAVEAAIQQQKAQGKVGQAQDILRLEHHHHHH

Figure 6B (continued)

CBGA Synthase – Enzyme Linker – cTPR6 Spacer – ID Linker – ID16

MGLSSVCTFSQTNYHTLLNPHNNNPKTSLLCYRHPKTPIKYSYNFPSKHCSTKSFHQNKCSESLSIAKNSIRAA
TNQTEPPESDNHSVATKILNFGKACWKLQRPYTIIAFTSCACGLFGKELLHNTNLISWSLMFKAFFFLVAILCIASFTTTI
NQIYDLHIDRINKPDPLLASGEISVNTAWIMSIIVALFGLIITIKMKGGLYIFGYZGFGIVVSVPPFRWKQNPSTAFL
LNFLAHIIITNFTFYASRAALGLPFEIIRPSFTLLAFMSMGSAALIKDASDVEGDTKFGISTLASKYGRSRNLTLCFCSG
IVLLSYVAIALAGIWPQAFNSNVMLLSHAIJAFWLILQTRDFALTNYDPEAGRPFYEFMWKLYYAEYLVVFIKLSGGG
GSGGGGGGGGSAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKAL
ELDPNNAEAWYNLGNAYYKQGDYQKAIEDYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPN
NASAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAKAWYRRGNAYYKQGDYQKAIEDYQKALELDPNRRSAG
GGGGGGGGGGGGGSAEYVRALFDNGNDEEDLPFKKGDIILRIRDKPEEQWWNAEDSEGKRGMPVYVEKYGS
AGSAAGSGEFGSAAAKAAGSAGSAAGSGEFGSLIKHMRAEALFDFTGNSKLELNFKAGDVIFLSSRINKDW
LEGTVRGATGIFPLSFVKILK

Acetyl-CoA Carboxylase – Enzyme Linker – cTPR6 Spacer – ID Linker – ID17

MSEESLFFESSPQKMEYEITNYSERHTELPGHFIGLNTVDKLEESPLRDFVKSHGGHTVISKILIANNGIAVKEIRSVRK
WAYETFGDDRTVQFVAMATPEDLEANAEMYRMAQYIEPGGTNNNNYANVDLIVDIAERADVDAWAGWGHASE
NPLLPEKLSQSRSRKVIFIGPPGNAMRSLSGDKISSTIVAQSAKVPICPWSGTGVDTVHVDEKTGLVSVDDDIYQKGCT
SPEDGLQKAKRIGFPVMIKASEGGGGKIRQVEREEDFIALYHQAANEIPGSPIFIMKLAGRARHLEVQLLAQYGTNI
SLFRDCSVQRRHQKIIEEAPVTIKAETFHEMekaavRLGKLVGYVSAGTVELYSHDDGFYFLENPRLQVEHP
TTEMVGVLPAALQIAMGIPMHRISDIRTLGGMNPHSASEIDFEFKTQDATKKQRRPIPKGHCTACRITSEDPDNG
FKPSGGTLHELNFRSSNVWGYFSVGNNGNIHSFSDSQFGHIFAFGENRQASRKHMVALKELSIRGDFRTTVEYLI
KLLETEDDFEDNTITTGWLDLTHKMTAEKPDP TLAVICGAATKAFLASEEARTHKEYIESLQKGQVLSKDLLQTMFPDF
IHEGKRYKFTVAKSGNDRYTLFINGSKCDILRQLSDGGLLIAIGGKSHTIWKEEVATRLSVDSMTTLEVENDPTQL
RTPSPGKLVKFLVENGHEIIKGQPYAEIEVMKMQMPLVSQENGIVQLLKQPGSTIVAGDIMAINTLDDPSKVHALPFE
GMLPDFGSPVIEGTPKPAYKFKSLVSTLENILKGYDQNQVIMNASLQQLIEVRNPKLPYSEWKLHISALHSRLPAKLD EQ
MEEELVARSLRRGAVFPARQLSKLIDMAVKNPEYNPDKLLGAVVPLADIAHKYNSGNEAHEHSIFTVFHLEYYEVEKL
FNGPNVREENIIKLDRDENPKDLKDVALVLSHSKVSAKNNLILAIKHYQPLCKLSSKVSIAFSTPLQHIVELESKATAK
VALQAREIILQGALPSVKERTEQIEHLKSSVVKAVYGSNNPKRSEPDNLNKLDLIDSNYVFDVLLQFLTHQDPVTTAA
AAQVYIRRAYRAYTGDIRDVHEGTVPIEWKFQLPSSAFTFPTVSKGMNRRAVSVDLSYVANSQSSPLREGILM
AVDHDDVDEILSQSLEVPRHQSSSSNGPAPDRGSSASLSNVANCVASTEGFESEEEILVRLREILDLNKQELINAS
IRRITFMFGFKDGSPYKYYTFNGPNNENETIRHIEPALAFQLELGRLSNFNIKPIFTDNRNHIVYEAVSKTSPLDKRFF
TRGIIRTGHIRDDISIQEYLTSEANRLMSDILDNLEVTDTNSSDLNHIFINFIAVFDISPEDVEAAFGGFLERFGKRLRLR
VSSAEIRIIKDPQTGAPVPLRALINNSGYVIKTEMYTEVKNAKGEWVFKSLGKPGSMHLRPIATPYPVKEWLQPKRY
KAHLMGTTVYDFPELFRQASSQWKNFSA DVKLTDFFISNELIEDENGELTEVEREPGANAIGMVAFKITVKTPPEY
PRGRQFVVVANDITFKIGSFGPQEDEFFNKVTEYARKRGPRIYLAANGARIGMAEIVPLFQVAWDAANPDKG
QYLYLTSEGMETLKKFDKENSVLTERTVINGEERFVIKIIGSEDGLGVECLRGSGLIAGATSRAYHDITLTVTCRSV
GIGAYLVRLGQRAIQVEGQPIILTGA PAINKMLGREVYTSNLQLGGTQIMYNNGVSHLTAVDDL A GVEKIVEWMSYVP
AKRNMPVPILETKDTWDRPVDFPTNDETYDVRWMIEGRETESGFEYGLFDKGSSFTLSGWAKGVVGRARLGGI
PLGVIGVETRTVENLIPADPANPNSAETLICQEPGQVWHPNSAFKTAQAI DFNNGEQLPMMILANWRGFGSGQRD
FNEVLKYGSFIVDALVDYKQPIIYIPPTGELRGGSWV/DPTINADQM E MYADV NARAGVLEPQGMVGIFRREKLL
DTMNRLDDKYRELRSQSLNSKSLAPEVHQQISQKQADRE RELLPIYGQISLQFADLHDRSSRMVAKGVISKELEWTEA
RRFFFWRLLRRRLNEEYLIKRLSHQVGEASRLEKARI RSWYPASVHD EDDRQVATWIEENYKTLDDKLKGLKLESFA
QDLAKKIRSDHDNAIDGLSEVIMLSTDDEKLLKTLKLSGGGGGGGGGGGGSAEAWYLNGLNAYYKQGDYQKAI EDYQKA
AIEYYQKALEDPNNAEAWYLNGLNAYYKQGDYQKAI EYYQKALEDPNNAEAWYLNGLNAYYKQGDYQKAI EDYQK
LEDPNNAEAWYLNGLNAYYKQGDYQKAI EDYQKALEDPNNAEAWYLNGLNAYYKQGDYQKAI EYYQKALEDP
NNAKAWYRRGNAYYKQGDYQKAI EDYQKALEDPNNSRSR SAGGGGGGGGGGGGGASGSHMRILGAQSIQPTAN
LDRTDDLVLYNVMELVRAVLELKNEAQLPPEGYV/W/KNVGLTRKLIGSVDDLLPSLPSSSRTEI EG TQKLLN KDLA
ELINKMRLAQQNATSLSEECKRQMLTASHTLAVDAKNLLDAV DQAKYLANLIAHPPAEGSAGSAAGSGEFGSAA
AKEAAKAGSAGSAAGSGEFGSGAMATPGSENVLPREPLIATAVKFLQNSRVQSPLATRR AFLKKGLTDEEIDM
AFQQSGTAADEPSSLW

Figure 6C

Cannabinoidergic Metabolon Scaffold – (Myc)₃

MGSAGSAAGSGEFGSAGSAAGSGEFGSAGSAAGSGEFSYYHHHHHHLESTSPLYKKAGSGSARNAYLRKKIA
RLKDNQLQLERDEQNLEKIIANLRDEIARLENENAVSHEQGSAGSAAGSGEFAAAKEAAAKAGSAGSAAGSG
EFSYYHHHHHHLESTSPLYKKAGSGSNLVQLENENAVSLENENETLKKNLHKKDLIAYLEKEIANLRKKIEEGSA
GSAAGSGEFGSAEEAAKEAAKEAAKEAAAKAGSAGSAAGSGEFGSSYYHHHHHHLESTSPLYKKAGSGSQ
KVAELKNRAVAKLNRNEQLKNKVEELKNRNAYLKNELATLENEVARLENDVAEGSAGSAAGSGEFAAAKEA
AAKAGSAGSAAGSGEFSYYHHHHHHLESTSPLYKKAGSGSNEVTTLENDAAFIENENAYLEKEIARLRKEKAALR
NRLAHKKGSAGSAAGSGEFGSAEEAAKEAAKEAAKEAAKEAAAKAGSAGSAAGSGEFGSRPPTISNPPPLISSAK
HPSVGSAGSAAGSGEFAAAKEAAKEAAAKAGSAGSAAGSGEFGNLFQRPEPTAPPEESFRSGGSAGSAAGSGE
FGSAEEAAKEAAKEAAKEAAKEAAAKAGSAGSAAGSGEFGSSKGTLNPNAKVWQEIAPGNGSAGSAAGSGE
AEAAKEAAAKAGSAGSAAGSGEFPDGTTFEHLWSSLEPDSTYGSAGSAAGSGEFGSAEEAAKEAAKEA
AAKEAAAKAGSAGSAAGSGEFGSSYYHHHHHHLESTSPLYKKAGSGSKRIYLRKKIAALKKDNNLEKDIANLE
NEIERLIKEIKTLENEVASHEQGSAGSAAGSGEFAAAKEAAAKAGSAGSAAGSGEFGSSYYHHHHHHLESTSPLY
KKAGSGSNLLATLRSTAATLENENHVLEKEKEKLKEKEQLLNKLEAYKGAGSAAGSGEFGSAEEAAKEAA
KEAAKEAAAKAGSAGSAAGSGEFGSPATSQHPPPPPGRSQAPSHGSAGSAAGSGEFGSAEEAAKEAAAKAG
SAGSAAGSGEFLNSLLLEAAEAYLERRDRGSAGSAAGSGEFGSAEEAAKEAAKEAAAKAGSAGSAAGSGE
AGSGEFGSRPPTISNPPPLISSAKHPSVGSAGSAAGSGEFGSAEEAAKEAAKEAAAKAGSAGSAAGSGE
TAPPEESFRSGGSAGSAAGSGEFGSAEEAAKEAAKEAAKEAAAKAGSAGSAAGSGEFGSSKGTLNPNA
KVWQEIAPGNGSAGSAAGSGEFAAAKEAAKEAAAKAGSAGSAAGSGEFPDGTTFEHLWSSLEPDSTYGSAGS
AAGSGEFGSAEEAAKEAAKEAAKEAAAKAGSAGSAAGSGEFGSSYYHHHHHHLESTSPLYKKAGSGSKRIA
YLRKKIAALKDNNLEKDIANLENEIERLIKEIKTLENEVASHEQGSAGSAAGSGEFAAAKEAAAKAGSAGS
AAGSGEFSYYHHHHHHLESTSPLYKKAGSGSNLLATLRSTAATLENENHVLEKEKEKLKEKEQLLNKLEAYKG
SAGSAAGSGEFGSAEEAAKEAAKEAAKEAAAKAGSAGSAAGSGEFGSAVDDADYEPPPSNNEALGSA
GSAAGSGEFAAAKEAAAKAGSAGSAAGSGEFLRELFDPSYVNQNLKARQGSAGSAAGSGEFGSAEEAA
AKEAAKEAAKEAAAKAGSAGSAAGSGEFGSKNTSMNFDPVYRKTTEEGSAGSAAGSGEFGSAEEAAKEAA
AAAKAGSAGSAAGSGEFRSLPSTWIENKLGYMSDPNWGSAGSAAGSGEFGSAEEAAKEAAKEAAKEAA
KAGSAGSAAGSGEFGSVVDNSPPPAPPKKRQSAPSGSAGSAAGSGEFGSAEEAAKEAAAKAGSAGSAAGSGE
FTQRSKPQPAVPPRPSADLILGSAGSAAGSGEFGSAEEAAKEAAKEAAKEAAAKAGSAGSAAGSGEFGGST
DEEREETEEEVYLLNSTTLLGSAGSAAGSGEFGSAEEAAKEAAKEAAAKAGSAGSAAGSGEFGSSYYHHHHHH
LESTSPLYKKAGSGSNTVKELNYIQELEERNAELKNLKEHLK
FAKAELEFELAAHKFEGSAGSAAGSGEFGSAEEAAKEAAKEAAKEAAKEAAAKAGSAGSAAGSGEFGSHDDSLP
HPQQATDDSGHESDGSAGSAAGSGEFGSAEEAAKEAAKEAAAKAGSAGSAAGSGEFGSPNAGSVEQTPKKPGLRR
GSAGSAAGSGEFGSAEEAAKEAAKEAAKEAAAKAGSAGSAAGSGEFGSSYYHHHHHHLESTSPLYKKAGS
GSFENVTHEFILATLENENAKLRRLEAKLRELARLNEAVLGSAGSAAGSGEFGSAEEAAKEAAAKAGSAGSA
AGSGEFSYYHHHHHHLESTSPLYKKAGSGSQKVEELKNKIAELENRNAVKKNRVAHLKQEIALKDELAAHEFE
GSAGSAAGSGEFGSAEEAAKEAAKEAAKEAAKEAAAKAGSAGSAAGSGEFGSVSSTKLVSFHDDSDEDLLHIGS
AGSAAGSGEFGSAEEAAKEAAKEAAAKAGSAGSAAGSGEFAAATPISTFHDDSDEDLLHVGAGSAAGSGEFGSAEEAA
AKEAAKEAAKEAAAKAGSAGSAAGSGEFGSSYYHHHHHHLESTSPLYKKAGSGSQKVESLKQKIEELKQRK
AQLKNDIANLEKEIAYAETGSAGSAAGSGEFGSAEEAAKEAAKEAAAKAGSAGSAAGSGEFGSSYYHHHHHH
LESTSPLYKKAGSEFFRERNRKMAAKCRNRRRELDTLQAETDQLEDEKSALQTEIANLLKEKEKLEFILA
HRPACKIPDDL
GFPEEMSLEGSAGSAAGSGEFGSAEEAAKEAAKEAAKEAAAKAGSAGSAAGSGEFGSFQMPADTPPPAY
LPPEPDPMTGSAGSAAGSGEFGSAEEAAKEAAAKAGSAGSAAGSGEFGFERESNEPPPPYEDPYWGNGGSAGSA
AGSGEFGSAEEAAKEAAKEAAKEAAAKAGSAGSAAGSGEFGSSYYHHHHHHLESTSPLYKKAGSGSQKVA
ELKNRAVAKLNRNEQLKNKVEELKNRNAYLKNELATLENEVARLENDVAEGSAGSAAGSGEFGAEEAAKEAAAK

Figure 6C (continued)

AGSAGSAAGSGEFSYYHHHHHLESTS LYKKAGSGSNEVTTLENDAAF IENENAYLEKEIARLRKEKAALRNRL
AHKKSYYHHHHHLESTS LYKKAGSGSARNAYLRKIA RLKKDNLQLERDEQNLEKIIANLRDEIARLENEVASH
EQGSAGSAAGSGEFAAAAKEAAAKAGSAGSAAGSGEFSYYHHHHHLESTS LYKKAGSGSNLVAQLEN EV
ASLENENETLKKKNLHKKDLIAYLEKEIANLRKKIEEGSAGSAAGSGEFGSAEAAAKEAAAKEAAAKAG
SAGSAAGSGEFGSEQKLISEEDLEQKLISEEDLG SAGSAAGSGEFGSAGSAAGSGEFGSAGSA
AGSGEF

Figure 6D**Malonyl-CoA Metabolon Scaffold ~ (FLAG)₃**

MGSAGSAAGSGEFGSAGSAAGSGEFGSAGSAAGSGEFSYYHHHHHLESTS LYKKAGSGSARNAYLRKKIA
RLKKDNLQLERDEQNLEKIIANLRDEIARLENEVASH EQGSAGSAAGSGEFAAAAKEAAAKAGSAGSAAGSG
EFSYYHHHHHLESTS LYKKAGSGSNLVAQLEN EVASLENENETLKKKNLHKKDLIAYLEKEIANLRKKIEEGS
GSAAGSGEFGSAEAAAKEAAAKEAAAKEAAAKAGSAGSAAGSGEFGSSATRELD ELMASLSDFKIQGGSAGS
AAGSGEFAAAAKEAAAKAGSAGSAAGSGEFDLASENWAQEFLAAGDAVDGSAGSAAGSGEFGSAEAAAK
EAAAKEAAAKEAAAKAGSAGSAAGSGEFGSDYKDDDDKDYKDDDDKDYKDDDDKG SAGSAAGSGEFGSAG
SAAGSGEFGSAGSAAGSGEF

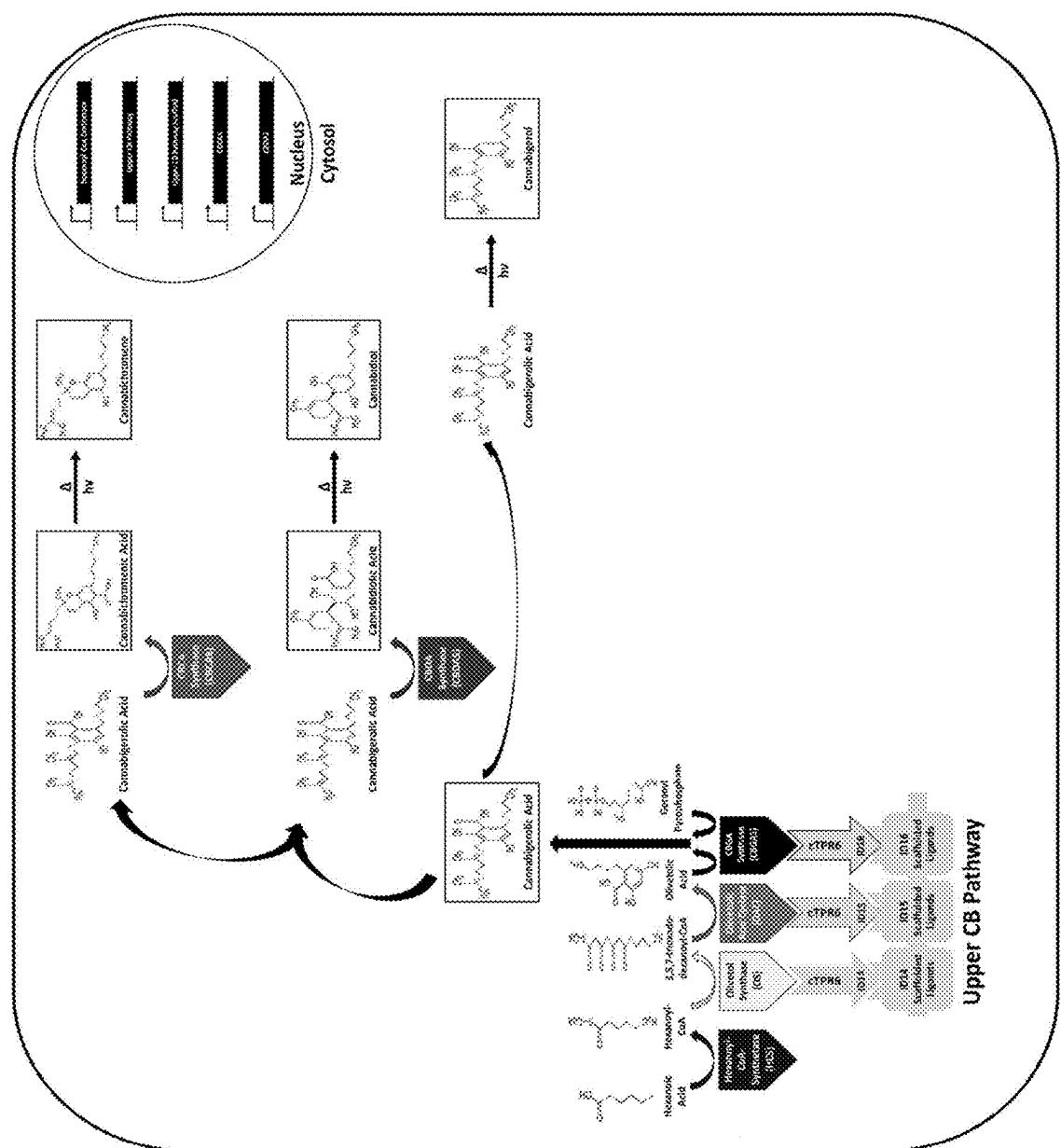
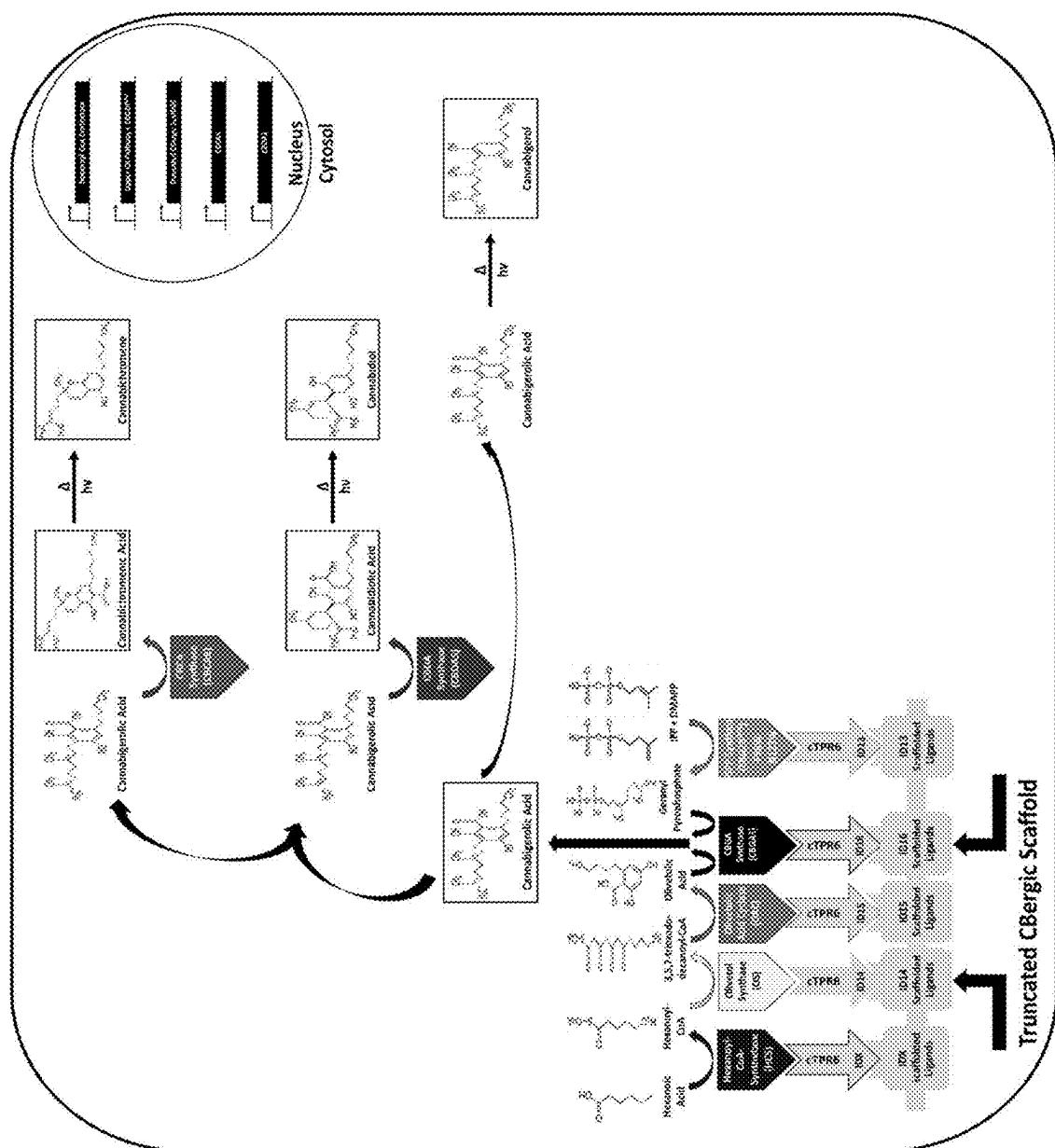


Figure 7.



४८

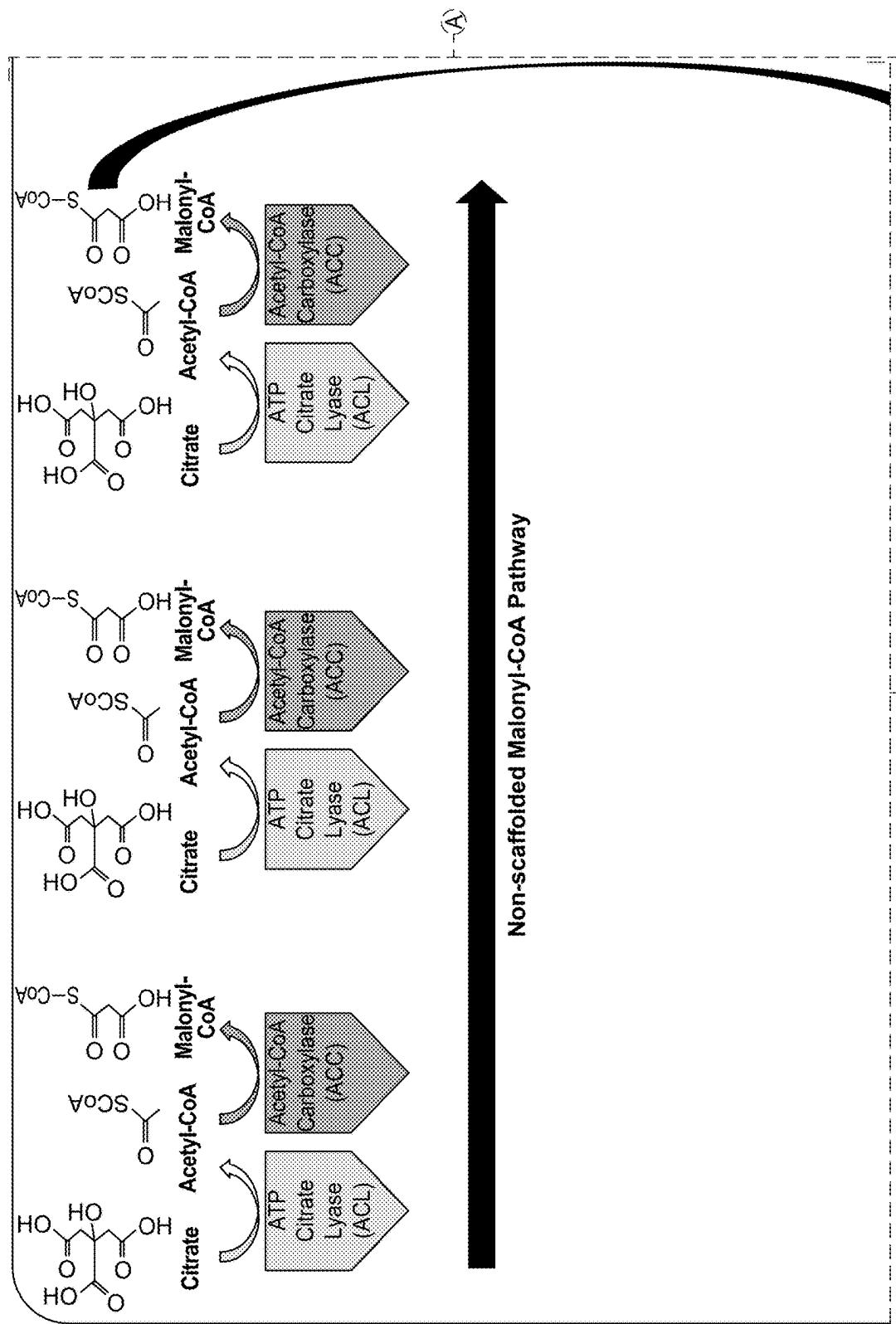


Figure 9.

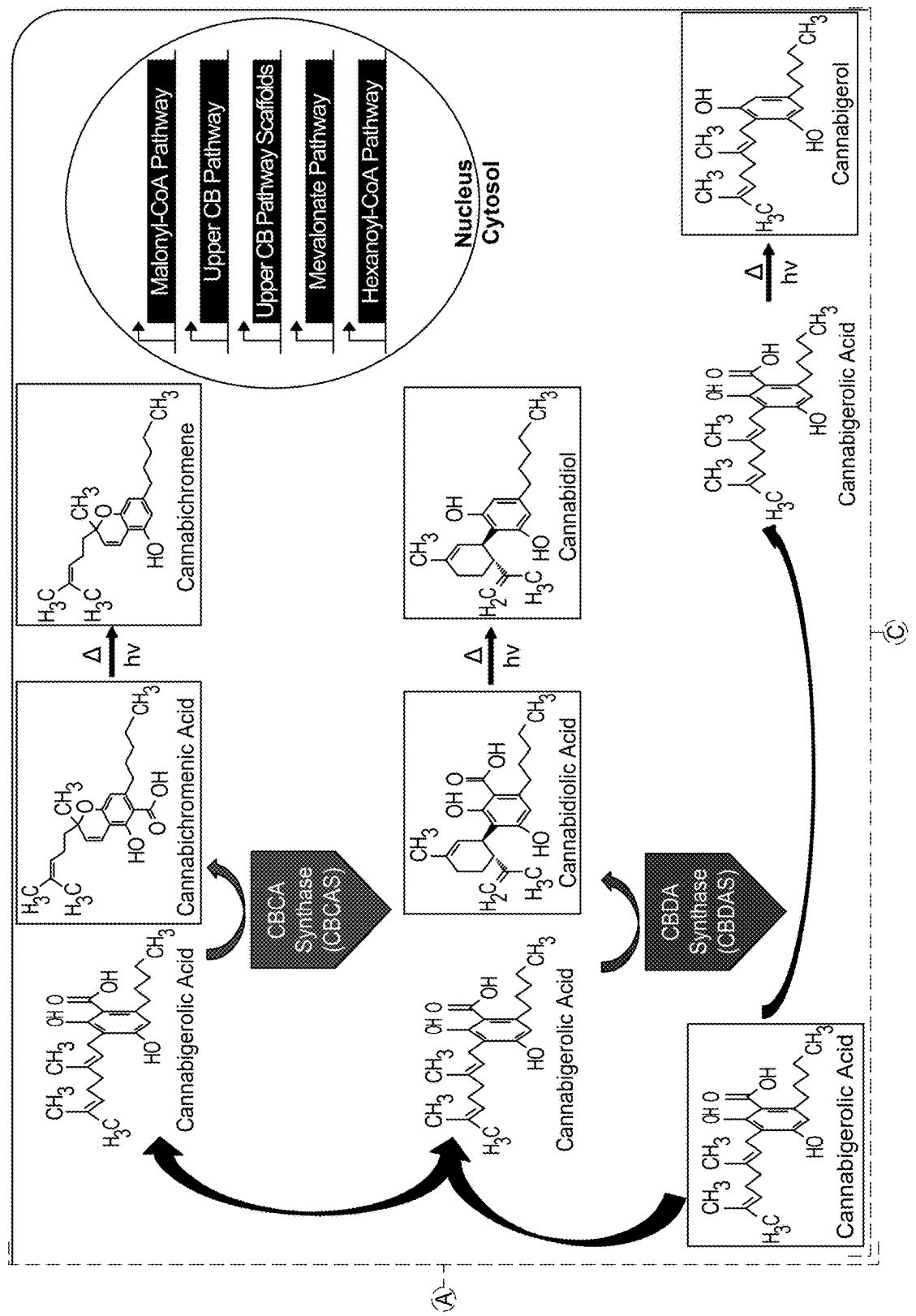


Figure 9. (Cont'd)

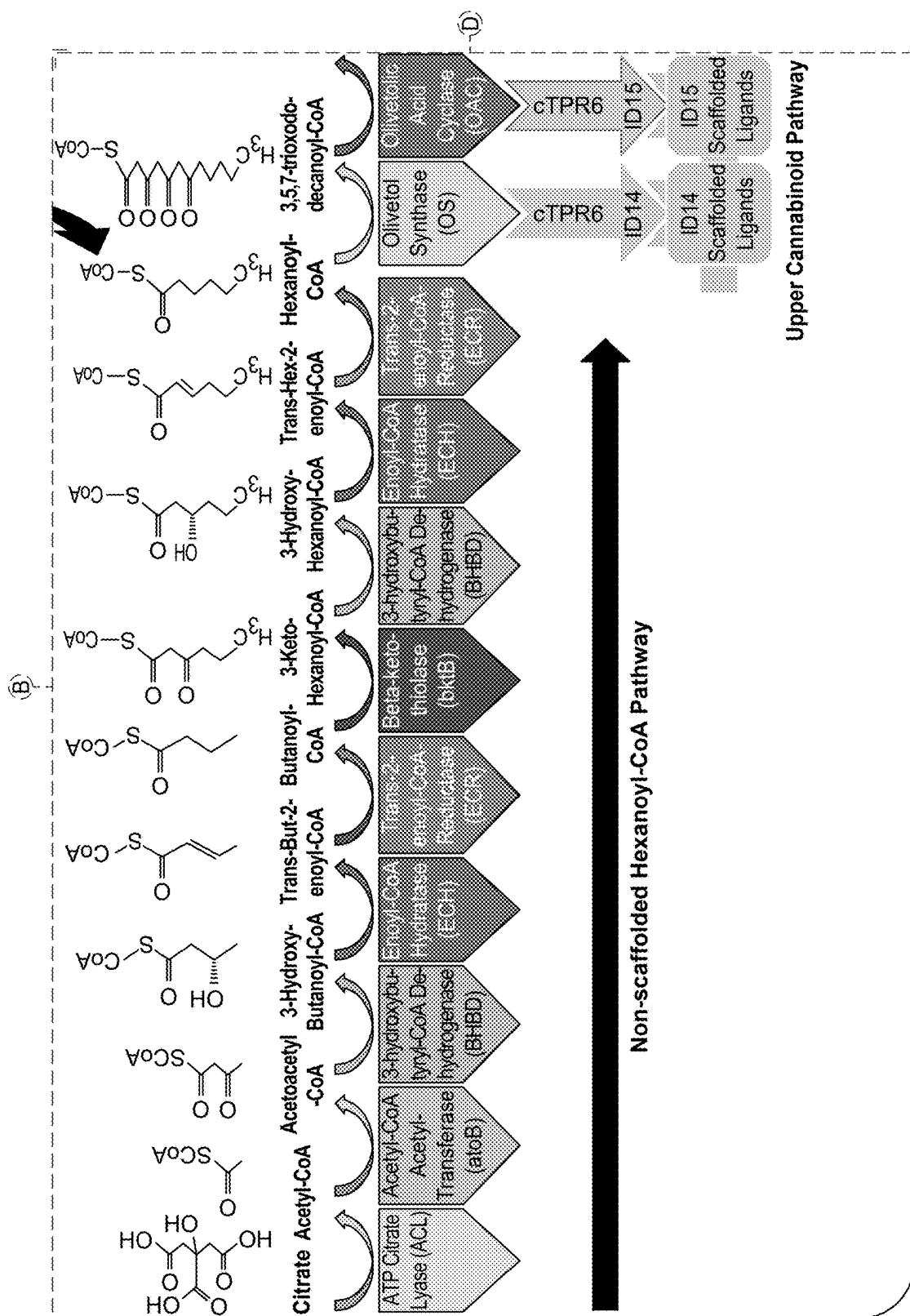


Figure 9. (Cont'd)

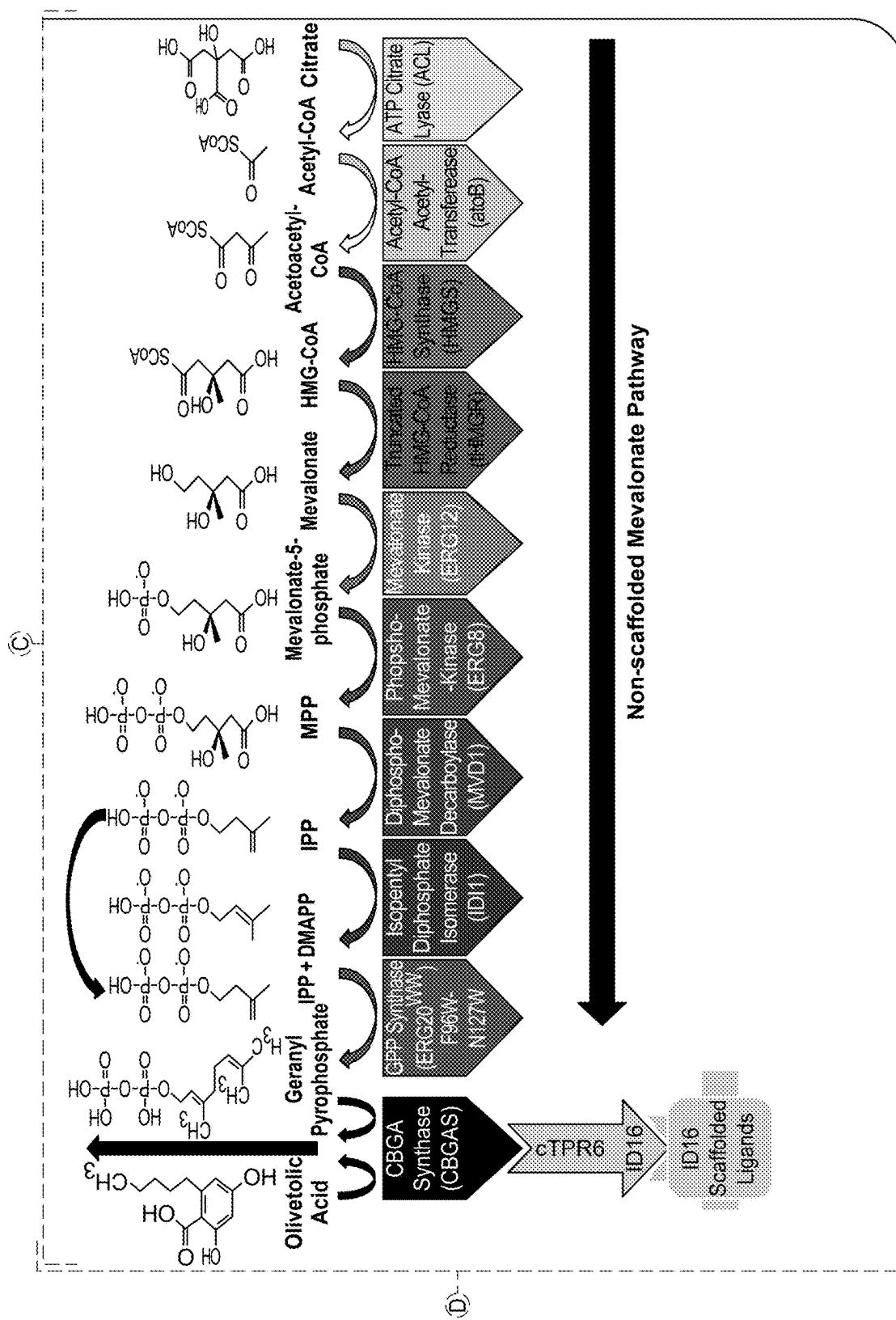


Figure 9. (Cont'd)

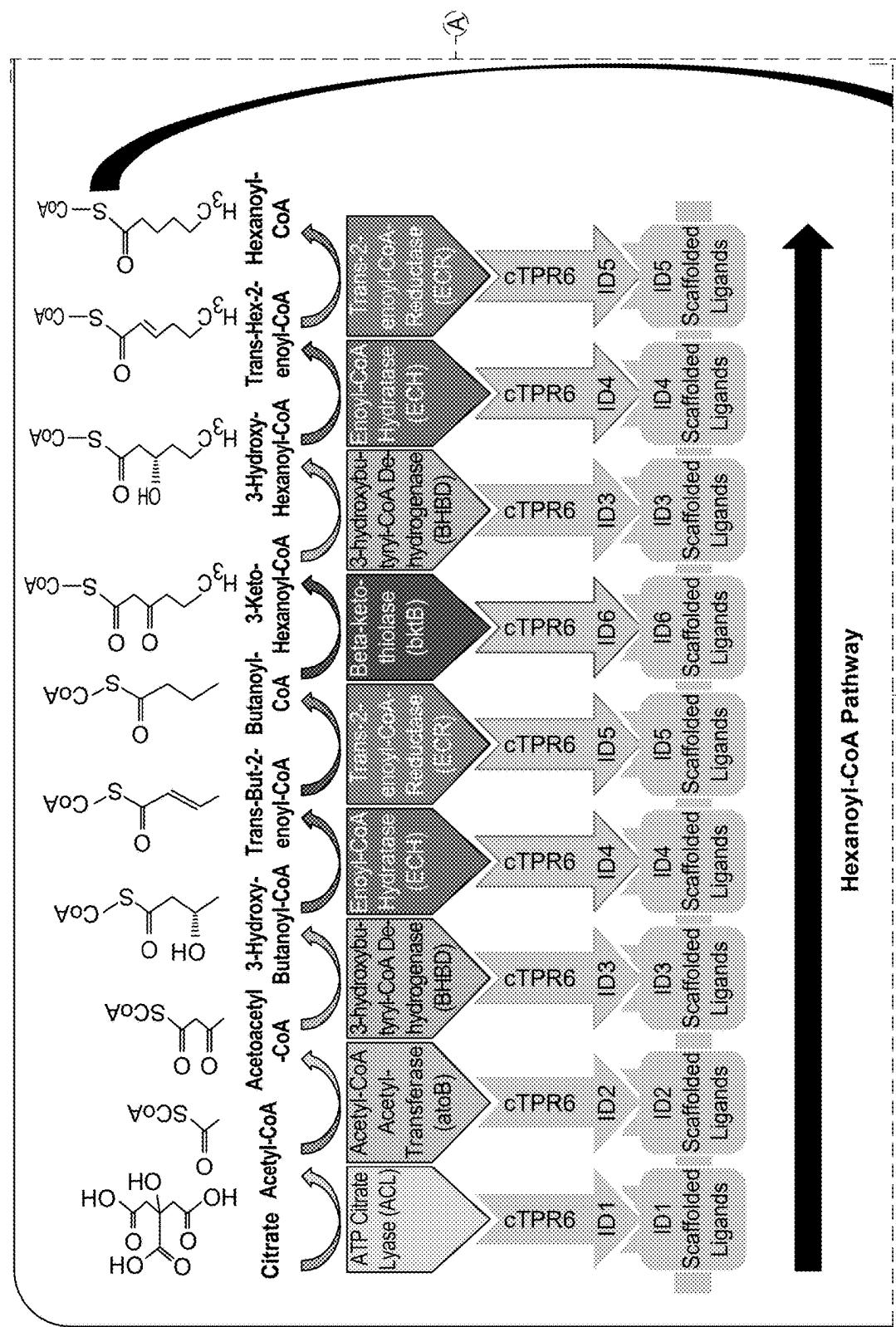


Figure 10.

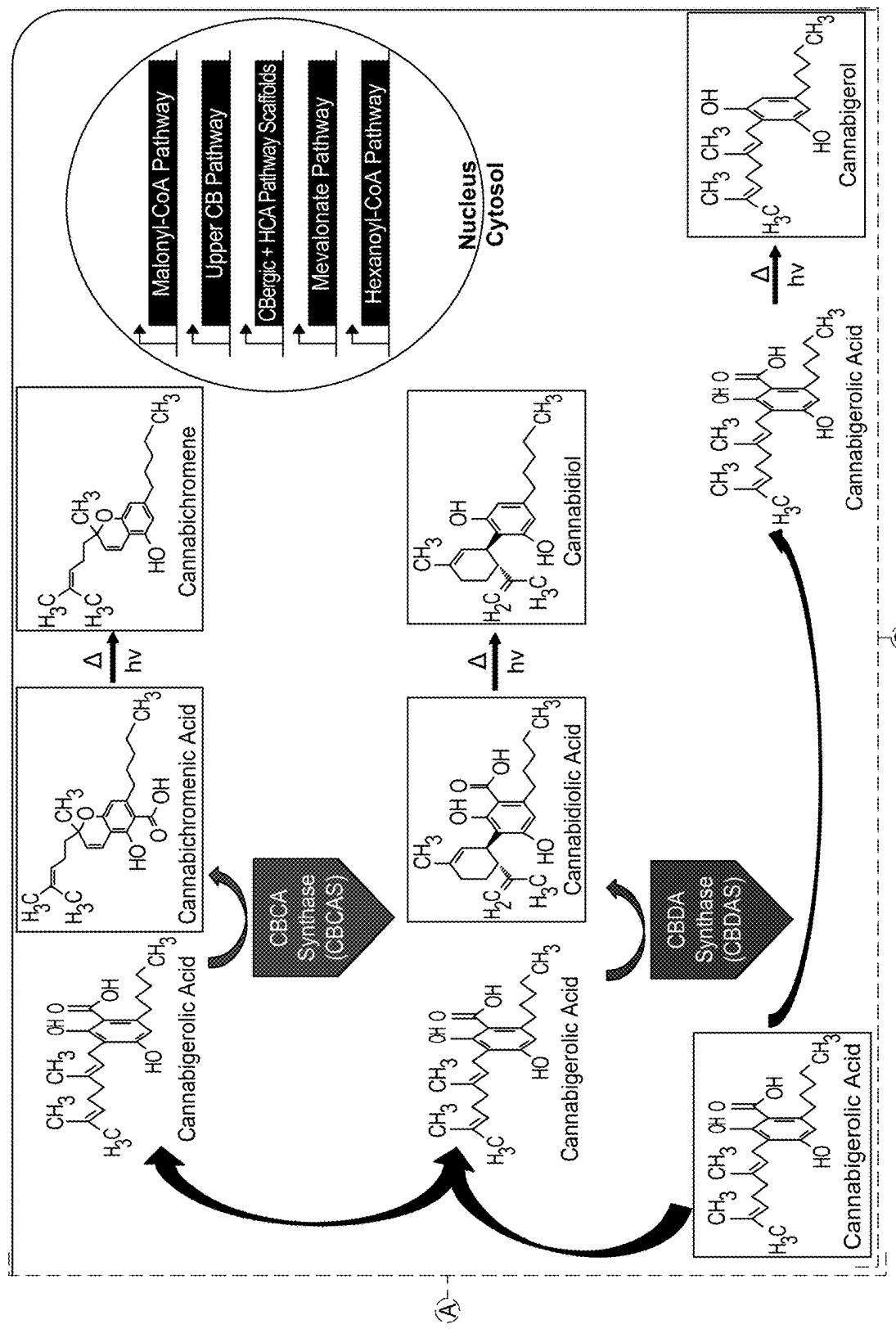


Figure 10. (Cont'd)

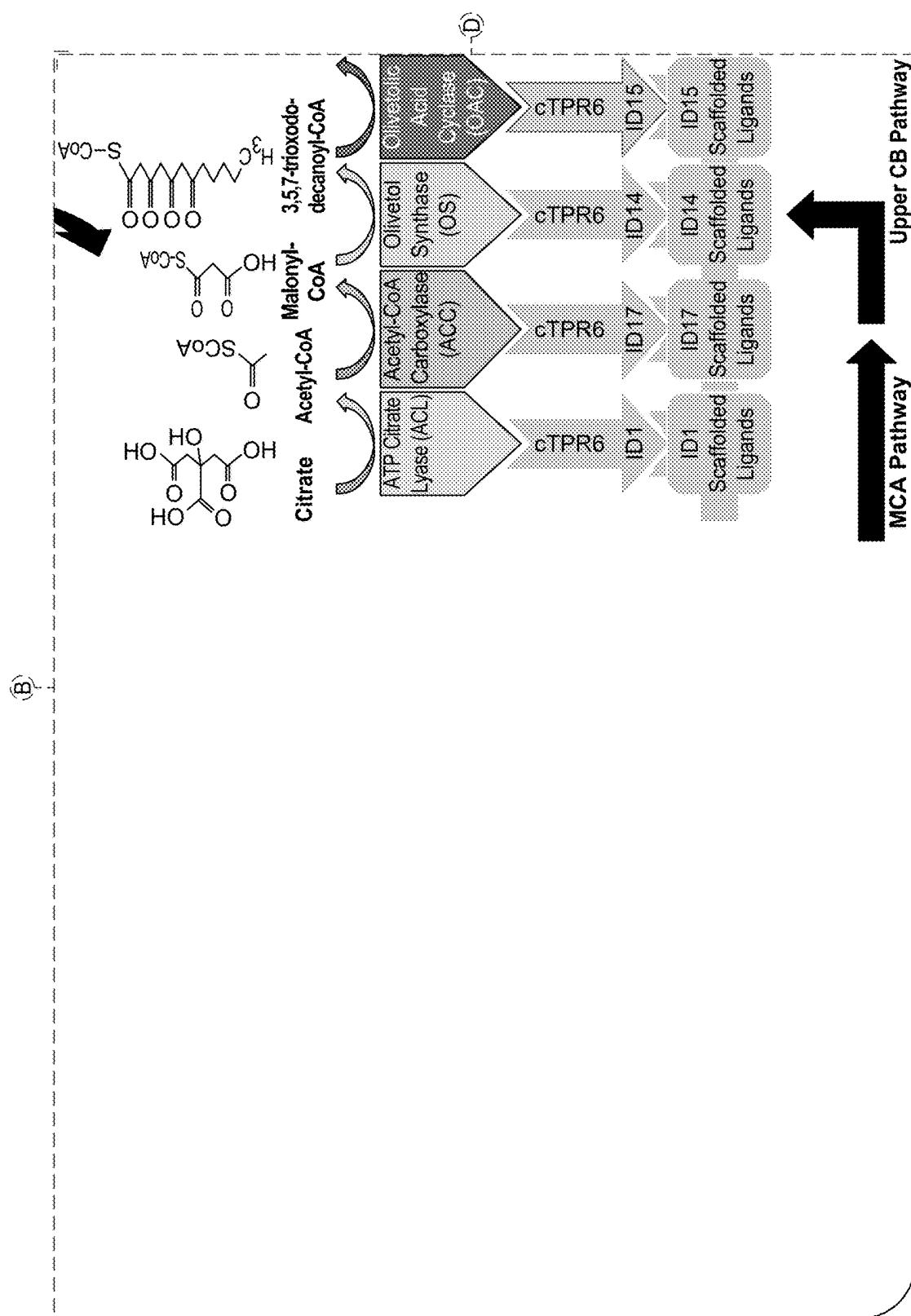


Figure 10. (Cont'd)

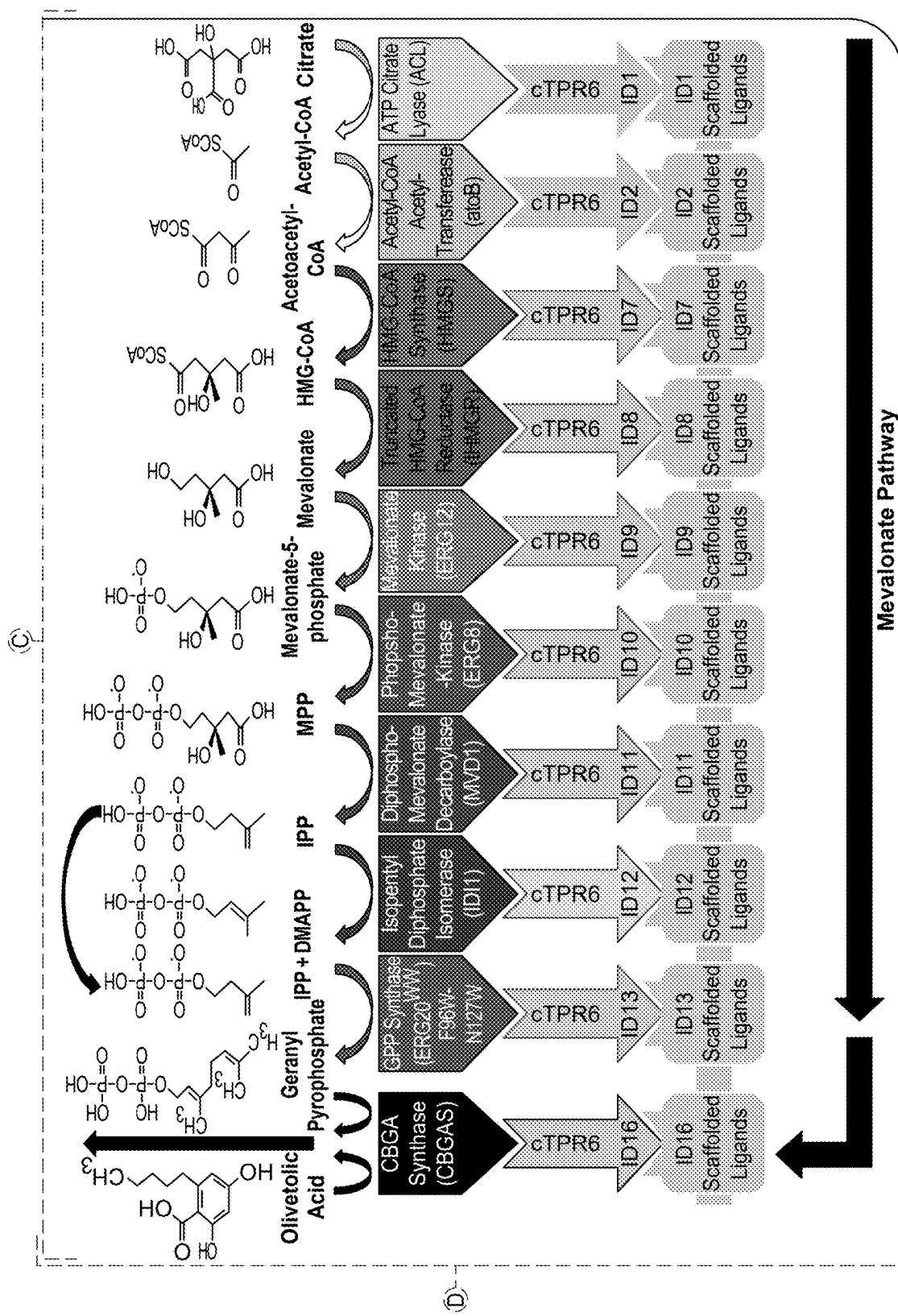


Figure 10. (Cont'd)

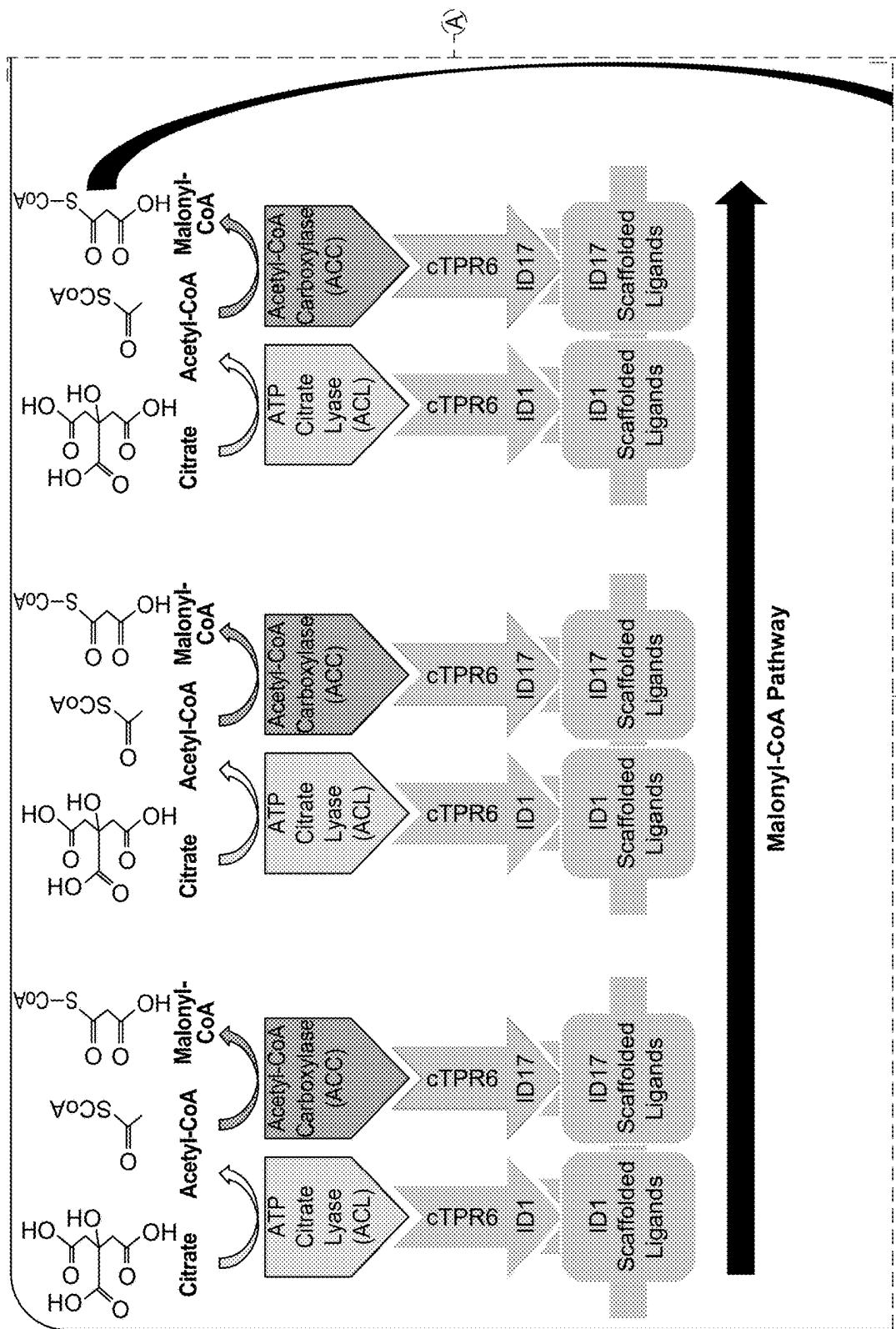


Figure 11.
Malonyl-CoA Pathway

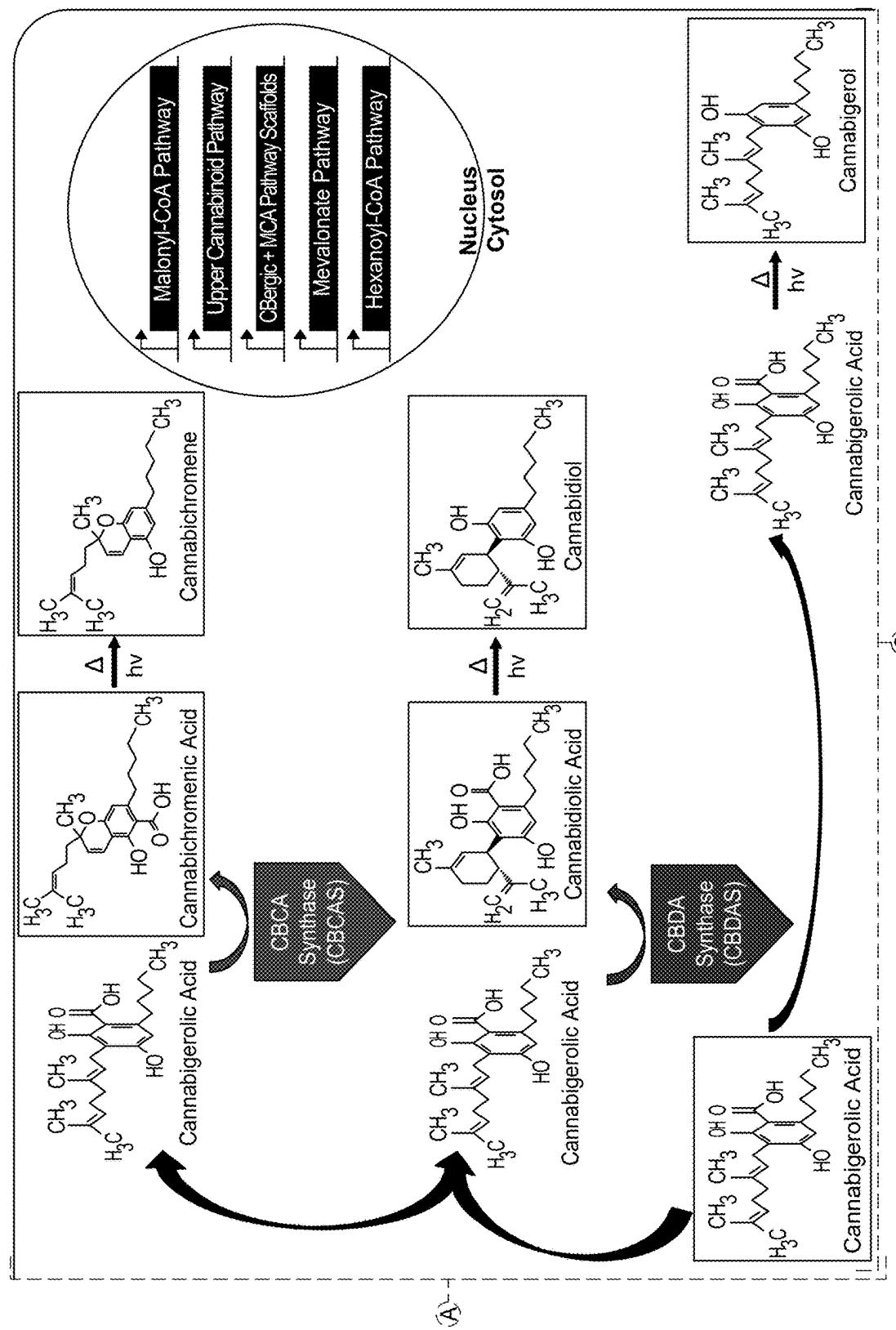


Figure 11. (Cont'd)

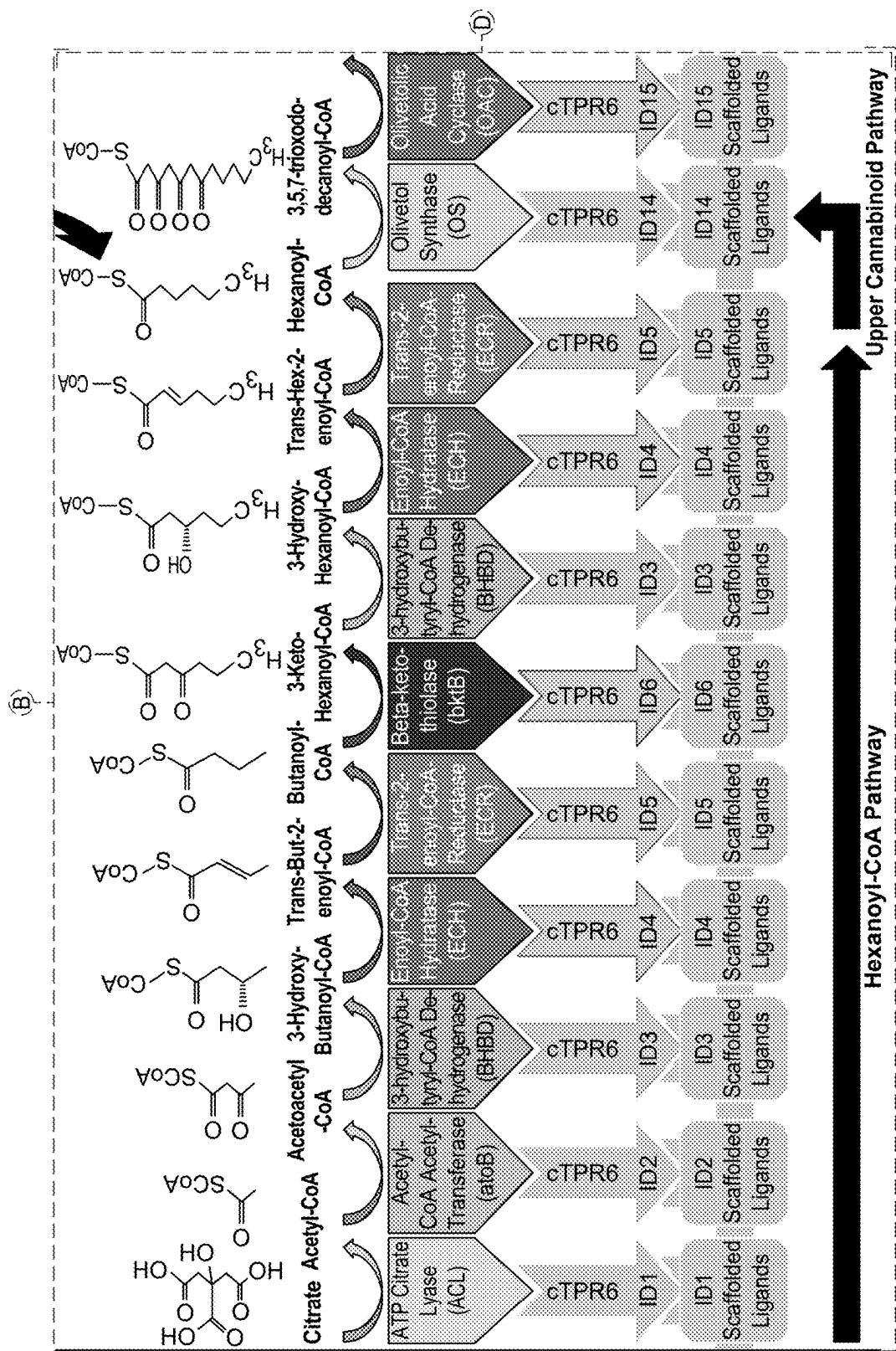


Figure 11. (Cont'd)

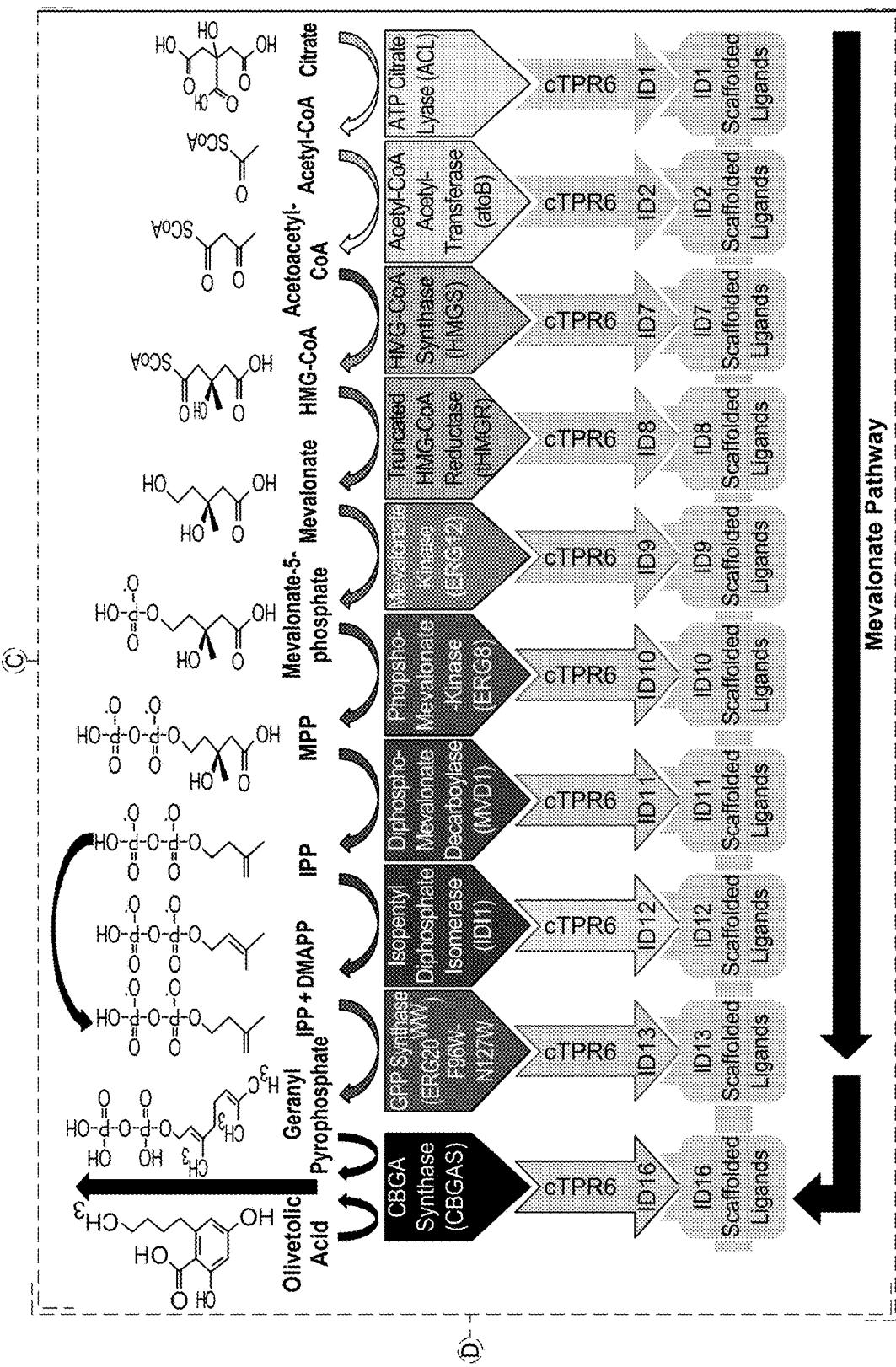


Figure 11. (Cont'd)

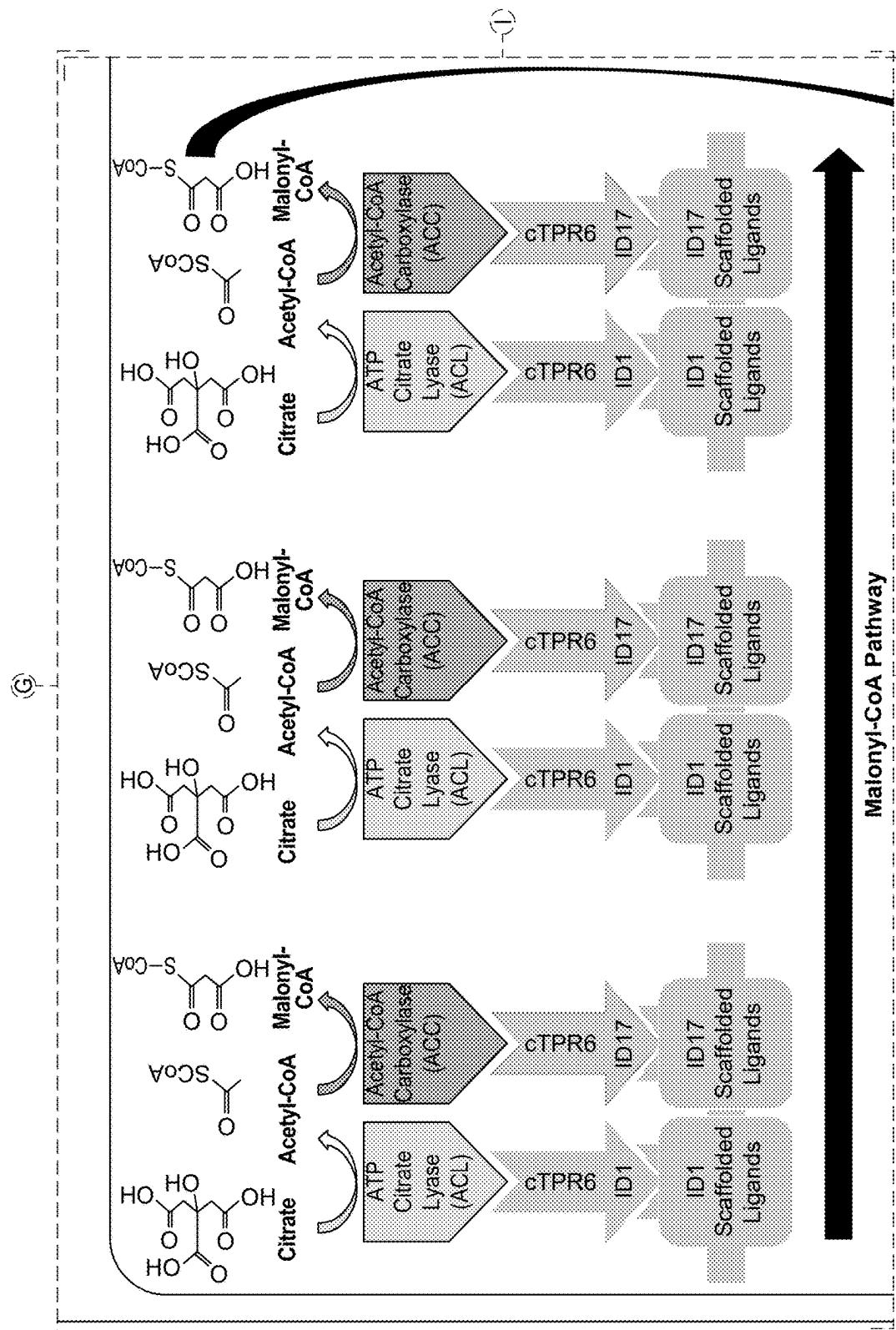


Figure 11. (Cont'd)
Malonyl-CoA Pathway

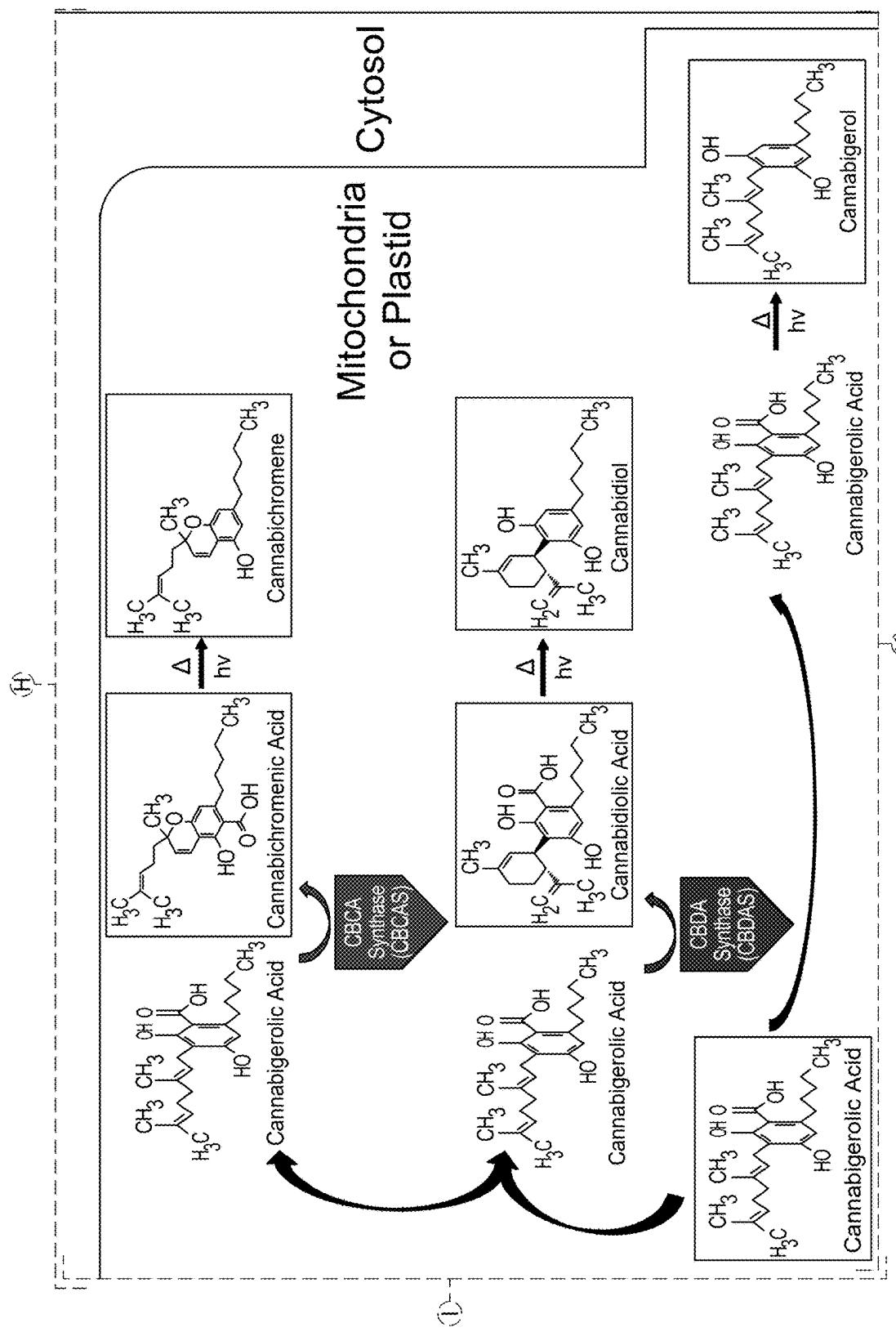


Figure 11. (Cont'd)

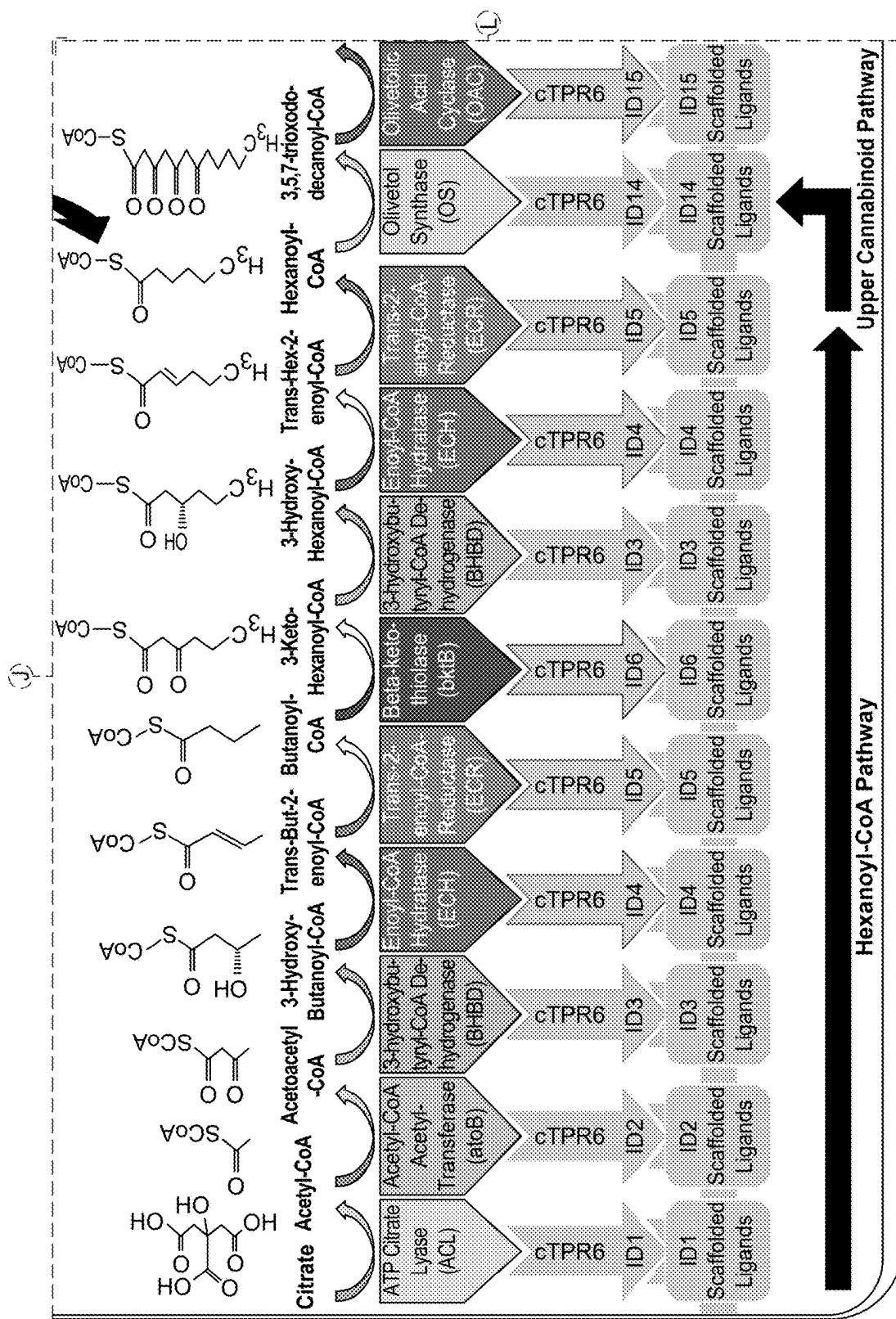


Figure 11. (Cont'd)

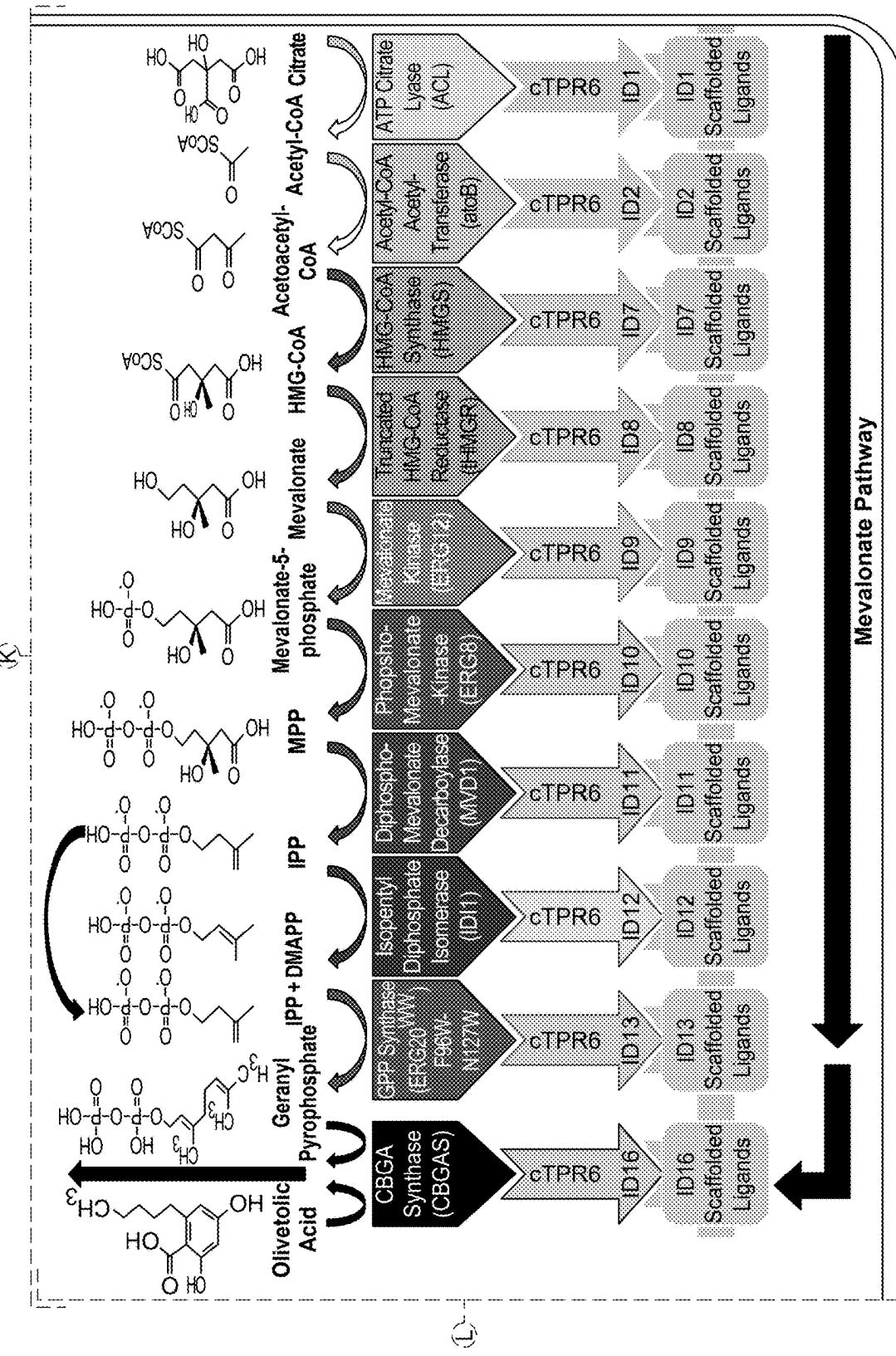


Figure 11. (Cont'd)

**BIDIRECTIONAL MULTI-ENZYMATIC
SCAFFOLDS FOR BIOSYNTHESIZING
CANNABINOID**

**CROSS-REFERENCE TO RELATED
APPLICATIONS**

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

This document relates to methods and materials for biosynthesizing cannabinoids, and more particularly to using bidirectional multi-enzymatic scaffolds to biosynthesize cannabinoids.

SEQUENCE LISTING

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

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

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.

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 interac-

tion 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 exogenous 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.

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.

In some embodiments, each of the engineered enzymes is of the formula: enzyme-linker₁-spacer-linker₂-motif₁-linker₃-motif₂, 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.

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.

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.

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

DESCRIPTION OF DRAWINGS

FIG. 1A 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.

FIG. 1B 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.

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

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

FIG. 3 is an example of an enzyme-scaffold complex.

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.

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.

FIG. 6A 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^{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).

FIG. 6B 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^{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).

FIG. 6C 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.

FIG. 6D 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.

FIG. 7 is a schematic of one representative embodiment of a scaffold with the minimal requirements for cannabinoids-gerolic 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.

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. 5 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.

10 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 15 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.

20 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 25 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 30 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.

35 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 40 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.

45 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.

FIG. 12A 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^{WW}) (SEQ ID NO:133), 50 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).

55 FIG. 12B 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^{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).

FIG. 12C 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.

FIG. 12D 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.

FIG. 13A 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-CoA Acetyltransferase (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).

FIG. 13B 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^{WW}) (ERG20^{WW}—Enzyme Linker—cTPR6 Spacer— ID Linker— ID13) (SEQ ID NO:172).

FIG. 13C 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).

FIG. 13D contains the amino acid sequences of the Cannabinoidergic Metabolon Scaffold (CBSCFLD)—(Myc)3 (SEQ ID NO:179) and the Malonyl-CoA Metabolon Scaffold (MCASCLD)—(FLAG)3 (SEQ ID NO:180).

FIG. 14A contains codon-optimized nucleotide sequences (SEQ ID NOS:181-187) encoding the enzymes of FIG. 13A.

FIG. 14B contains the codon-optimized nucleotide sequences (SEQ ID NOS:188-194) encoding the enzymes of FIG. 13B.

FIG. 14C contains the codon-optimized nucleotide sequences (SEQ ID NOS:195-200) encoding the enzymes of FIG. 13C.

FIG. 14D contains the codon-optimized nucleotide sequences (SEQ ID NO:201 and SEQ ID NO:202) encoding the scaffolds of FIG. 13D.

FIG. 15A contains the nucleotide sequence of the HCA gene cassette (SEQ ID NO:203).

FIG. 15B contains the nucleotide sequence of the GPP gene cassette (SEQ ID NO:204).

FIG. 15C contains the nucleotide sequence of the CAN gene cassette (SEQ ID NO:205).

FIG. 15D contains the nucleotide sequence of the SCF gene cassette (SEQ ID NO:206).

FIG. 15E contains the nucleotide sequence of the SOL gene cassette (SEQ ID NO:207).

FIG. 16 is a map of the pCCI-Brick plasmid construct.

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.

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_{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 $OD_{600nm}=0.3$. 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.

Figs. 19A-19E show a comparison of cannabinoid and precursor titers for scaffolded and soluble cannabinoid biosynthesis. Representative mass spectra of target analytes isolated from (FIG. 19A) yCBSOL and (FIG. 19B) yCBSCF cultures incubated for 48 hours in basal culture media. Bar plots depicting (FIG. 19C) Total (aggregate) cannabinoid (CBGA+CBDA+CBCA+CBG+CBD+CBC) titers, (FIG. 19D) cannabinoid precursor (OVA) titers and summated parent and decarboxylation derivative (CBGA+CBG,

CBDA+CBD, and CBCA+CBC) cannabinoid titers, and (FIG. 19E) 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; $\alpha=0.05$) between-strain differences for yCBSCF versus yCBSOL cultures. Bar plots depict means with 95% confidence intervals. * $p<0.05$; ** $p<0.01$; *** $p<0.001$; **** $p<0.0001$.

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; $\alpha=0.05$) between-strain differences for yCBSCF versus yCBSOL cultures. Lines with asterisks indicate statistically significant (determined by Bonferroni's multiple comparisons post-hoc test; $\alpha=0.05$) 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. * $p<0.05$; ** $p<0.01$; *** $p<0.001$; **** $p<0.0001$.

FIGS. 21A and 21B show concentration-response parameterization of scaffolded and soluble cannabinoid biosynthesis from citrate. In FIG. 21A, 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. 21B, bar graphs are shown depicting concentration-response parameter estimates (CB_{Max} , the estimated maximum total cannabinoid titers and citrate EC₅₀, the estimated citrate concentration yielding half-maximal total cannabinoid titers) for 48-hour yCBSCF and yCB_{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_{SOL} cultures. Floating asterisks indicate statistically significant (determined by Bonferroni's multiple comparisons post-hoc test; $\alpha=0.05$) between-strain differences for yCBSCF versus yCB_{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

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 decarboxylation 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.

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. 1A and FIG. 1B, 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.

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.

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

11

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.

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.

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 used as an ID as described herein.

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.

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

12

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

5

(SEQ ID NO: 1)
PPPALPPPKRRR.

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]X ϕ COOH) motif, Class II PDZ domains recognize a (X ϕ X ϕ INCOOH) motif, and Class III PDZ domains recognize a X[ED]X ϕ COOH 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 -syntrphin protein having GenBank Accession No. EDL06069 and the peptide ligand can be

30

(SEQ ID NO: 208)
GVKESLV.

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.

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 ITIRAAFLEKENTALRTEIAEL-EKEVGRCENIVSKYETRYGPL (SEQ ID NO:3), and its peptide ligand for use as described herein can be

(SEQ ID NO: 4)
LEIRAAFLEKENTALRTRAELRKRVGRCRNIVSKYETRYGPL.

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

13

useful for ex vivo applications as binding is calcium dependent. See, Whitaker and Dueber, 2011, *supra*.

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

14

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.

Table 1 provide exemplary combinations of heterologous 10 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 15 Tables 1 and 2 can be used in combination with any of the listed combinations of IDs and corresponding peptide ligands.

TABLE 1

Interaction Domain Motif Sequences in Engineered Enzymes					
Enzyme	ID #	ID Motif #1	ID Motif #1 Amino Acid Sequence	ID Motif #2	ID Motif #2 Amino Acid Sequence
ATP Citrate Lyase	1	SYNZIP1	SYYHHHHHHLESTSLSYKKAGSG SNLVAQLENEVASLENENETLK KKNLHKKDLIAYLEKEIANLRK KIEE ((SEQ ID NO: 5))	SYNZIP2	SYYHHHHHHLESTSLSYKKAGSGS ARNAYLKKIARLKDDNLQLERD EQNLEKIIANLRDEIARLENEVASH EQ (SEQ ID NO: 6)
Acetyl-CoA Acetyltransferase (atoB)	2	SYNZIP3	SYYHHHHHHLESTSLSYKKAGSG SNEVTTLENDAAFIENENAYLE KEIARLRKEKAALRNRLAHKK (SEQ ID NO: 7)	SYNZIP4	SYYHHHHHHLESTSLSYKKAGSGS QKVAELKNRVAVKLNRNEQLKNK VEELKNRNAYLKNELATLNEVVA RLENDVAE (SEQ ID NO: 8)
3-hydroxybutyryl-CoA Dehydrogenase	3	MYND	ENLYFQGENLYFGDSSESCWN CGRKASETCGCGNTARYCGSF QHDKWEKHHICGQTLQAQQ (SEQ ID NO: 9)	UEV	MAVSESQLKKMVKSYKYRDLTWR ETVNVIITYKDLKPVLDSDYVFNDG SSRELMNLTGTTIPVYRGNTYNIPI CLWLLDTYPYNPICFVKPTSSMTI KTGKHVDANGKIYLPLYHEWKHP QSDLGLIQVMIVVFGDEPPVFSRP (SEQ ID NO: 10)
Enoyl-CoA Hydratase	4	PAPP	GPLGSPLTASMLASAPPQEQQ MLGERLFPPLIQAMHPTLAGKITG MLLEIDNSELLHMLESPESLRSK VIDEAVAVLQAHQAKEAAQKA (SEQ ID NO: 11)	MDM2	NTNMSVPTDGAVTTSQIPASEQET LVRPKPLLLKLLKSVGAQKDYT MKEVLFLGQYIMTKRLYDEKQQ HIVYCSNDLLGDLFGVPSFSVKEH RKIYTMIFYRNLLV (SEQ ID NO: 12)
Trans-Enoyl-CoA Reductase	5	SYNZIP10	SYYHHHHHHLESTSLSYKKAGSG SNLATLRLSTAVALENENHVLE KEKEKLRLKEKEQLLNKLEAKY (SEQ ID NO: 13)	SYNZIP22	SYYHHHHHHLESTSLSYKKAGSGS KRIAYLRLKIAALKDNANLEKDI ANLENEIERLIKEIKTLENEVASHE Q (SEQ ID NO: 14)
Beta-ketothiolase (bktB)	6	GYF	DVMWEYKWEINTGDAELYGPFT SAQMQTWSEGYPPDGVCRK LDPGGQFYNSKRIDFDLYT (SEQ ID NO: 15)	PAH	ESDSVEFNNAISYVNKIKTRFLDHP EIYRSFLEILHTYQKEQLHTKGRPF RGMSEEEVFTEVANLFRQEDLLS EFGQFLPEAKR (SEQ ID NO: 16)
HMG-COA Synthase	7	WW1A	LGPLPPGWEVRSTVSGRIYFVD HNRRRTQFTDPRHLH (SEQ ID NO: 17)	WW1B	GAMGPLPPGWEKRTDSNGRKYFV NHNTRITQWEDPRS (SEQ ID NO: 18)
HMG-COA Reductase	8	FOS	SYYHHHHHHLESTSLSYKKAGSE FFRRERNKMAAAKCRNRREL DTLQAETDQLEDEKSALQTEIA NLLKEKEKLEPILAAHRPACKIP DDLGFPEEMSLE (SEQ ID NO: 19)	SYNZIP9	SYYHHHHHHLESTSLSYKKAGSGS QKVESLKQKIEELKQRKAQLKNDI ANLEKEIAYAET (SEQ ID NO: 20)

TABLE 1-continued

Enzyme	Interaction Domain Motif Sequences in Engineered Enzymes					
	ID #	ID Motif #1	ID Motif #1 Amino Acid Sequence	ID Motif #2	ID Motif #2	Amino Acid Sequence
Mevalonate Kinase	9	VHS1	MEPAMEPETLEARINRATNPLN KELDWASINGFCQEQLNEFEGP PLATRLLAHKIQSPEWEAIQAL TVLETCKMKSCGKRPHDEVGKFR FLNLELIKVVSPKYLGSRTSEKVK NKILELLYSWTVGLPEEVKIAEA YQMLKKQGIVKS (SEQ ID NO: 21)	VHS2	GAMGMSMAEAGESLESWLNKATN PSNRQEDWEYIIGFCDQINKELEGP QIAVRLLAHKIQSQPEWEALQALT VLEACMKNCGRRFHNEVGKFRFL NELIKVVSPKYLGDRVSKEVKTGV IELLYSWTMALPPEAKIKDAYHML KRQGIVQSDPPPIPVRTLIPSPPPRP KN (SEQ ID NO: 22)	
Phosphomevalonate Kinase	10	SYNZIP13	SYYHHHHHHLESTS L YKKAGSGS SQKVEELKNKIAELEN R NAVKK N R V A H L K Q E I A Y L K D E A H E F E (SEQ ID NO: 23)	SYNZIP15	SYYHHHHHHLESTS L YKKAGSGSF ENVTHEFILATLENENAKLRRLEA KLERELARLRNEVAWL (SEQ ID NO: 24)	
Diphospho-mevalonate Decarboxylase	11	MATH	AMADLEQKVLEMEASTYDGVFI WKISDFPRKRQEAVAGRIPAIFS PAFYTSRYGYKMCLRIYLNGDG TGRGTHLSLFFFVVMKGPNNDALL RWPFNQKVTLMLLDQNNREHV IDAPRPDVTS SSSFQRPVNDMNIA SGCPLFCPVSKM EAKNSYVRDD AIFIKAIVDLTGL (SEQ ID NO: 25)	SKP1	ASIKLQSSDGEIFEV D V E I A K Q S V T I KTMLEDLGMDDEGDDP VPLPNV NAAILKKVIQWC THHKDDPPP PED DENKEKRTTDI P V W D Q E F L K V D Q GTLFELILA ANYLDI KG L LD V T C K T VANMIKGK T PEEIRKTFNIK N D F T E EEEAQVRKENQWC (SEQ ID NO: 26)	
Isopentenyl-Diphosphate Delta-Isomerase	12	SYNZIP5	SYYHHHHHHLESTS L YKKAGSGS SNTVKELKNYI Q E E R N A E L K NLKEHLKFKA E LE F E L A H K F E (SEQ ID NO: 27)	SYNZIP6	SYYHHHHHHLESTS L YKKAGSGS QKVAQLKNRVAYKL KENAKLEN I VARLENDNANLEKDIANLEKDIAN LERDVAR (SEQ ID NO: 28)	
Geranyl-Diphosphate Synthase	13	PDZ1	LCTMKKGPGSGYGFNLHSDKSKP GQFIRSVDDPSA E ASGLRAQDR IVEVNGVCMEGKQHGDVVSAIR AGGDET KLLVV DRE (SEQ ID NO: 29)	PDZ2	SSGALIYTVELKRYGGPLGITISGTE EPFDPIIISSSLTKGGLAERTGAIHIG DRILAINSSSLKGKPLSEAIHLLQM AGETVTLKIKKQTDAQPASS (SEQ ID NO: 30)	
Olivetol Synthase	14	SH2A	GNNLETYEWYNKSISRDKA EKL LLDTGKEGAFMVRD SRT PGT YT VSVPTKAI ISENPCIKHYHI KET NDSPKRYVVAEKYVFDSIPLL IQ YHQYNGGLVTRLRYPVCG (SEQ ID NO: 31)	SH2B	GSHPWFFGKIPRKA E E M L S K Q R H DGAFLIRESESAPGDFSL SVKFGND VQHPKVL RD GAGKYFLWVVKFNS LNELVDYHRSTSVSRNQQI F LR DIE QVPQQPT (SEQ ID NO: 32)	
Olivetolic Acid Cyclase	15	PTB1	GQDRSEATLKR KFG EGV RYKA KLIGIDEVSAARGDKL CQDSMM KLKG VVAGARS KGE HKQ KI FLT ISFGGIKIFDEKTGALQHHHA VH EISYI AKDITDH R AFGYVG CGKEG NHRV AIA KTAQAA E P V I L D L R D L FQLIYELKQREELEKKA (SEQ ID NO: 33)	PTB2	GSHMGSQFWVTSQKTEASER CGL QGSYI LRVEAEKLTLLTGQ S Q I L EPPLFWPYTLLR RYGRD KV MFS F E AGRRCPSGP GTFTF QT S Q G N D I F Q AVEAAIQQQKAQGKVGQ A Q D I L R LEHHHHHH (SEQ ID NO: 210)	
CBGA Synthase	16	SH3A	A E Y V R A L F D F N G N D E E D L P F K K G D I L R I R D K P E E Q W N A E D S E G K R G M I P V P V Y V E K Y (SEQ ID NO: 34)	SH3B	L I K H M R A E A L F D F T G N S K L E L N F K A G D V I F L L S R I N K D W L E G T V R G A T G I F P L S F V K I L K (SEQ ID NO: 35)	
Acetyl-CoA Carboxylase	17	FAT	GSHMRLGAQSIQPTANL DRT DD L V Y L N V M E L V R A V L E L K N E L A Q L P P E G Y V V V V K N V G L T L R K L I G S V D D L L P S L P S S R T E I E G T Q K L L N K D L A E L I N K M R L A Q Q N A V T S L S E E C K R Q M L T A S H T L A V D A K N L L D A V D Q A K V L A N L A H P P A E (SEQ ID NO: 36)	PEX	G A M A T P G S E N V L P R E P L I A T A V K F L Q N S R V R Q S P L A T R A F L K K G L T D E E I D M A F Q Q S G T A A D E P S S L W (SEQ ID NO: 37)	

TABLE 2

Tandem Peptide Ligand Sequences in Scaffold					
Enzyme	ID #	ID Motif #1	ID Motif #2	Scaffolded Ligand Amino Acid Sequence	Scaffolded Ligand Amino Acid Sequence
ATP Citrate Lyase	1	SYNZIP1	SYNZIP2	SYYHHHHHLESTSLYKKAGS GSARNAYLRKKIARLKKDNLQ LERDEQNLEKIIIANLRDEIARLE NEVASHEQ (SEQ ID NO: 6)	SYYHHHHHLESTSLYKKAGSGS NLVAQLENEVASLENENETLKKK NLHKDKLIAYLEKEIANLRKKIEE (SEQ ID NO: 5)
Acetyl-CoA Transferase (atoB)	2	SYNZIP3	SYNZIP4	SYYHHHHHLESTSLYKKAGS GSQKVAELKNRVAVKLNKRNEQ LKNKVEELKNRNAVYLKNELAT LENEVARLENDVAE (SEQ ID NO: 8)	SYYHHHHHLESTSLYKKAGSGS NEVTTLENDAAFIENENAYLEKEI ARLRKEKAALRNRLAHKK (SEQ ID NO: 7)
3-hydroxybutyryl-CoA Dehydrogenase	3	MYND	UEV	RPPTISNPPLISSAKHPSV (SEQ ID NO: 38)	NFLQSRPEPTAPPEESFRSG (SEQ ID NO: 39)
Enoyl-CoA Hydratase	4	PABP	MDM2	SKGTGLNPNAKVWQEIAPGN (SEQ ID NO: 40)	PDGGTTFEHLWSSLEPDSTY (SEQ ID NO: 41)
Trans-Enoyl-CoA Reductase	5	SYNZIP10	SYNZIP22	SYYHHHHHLESTSLYKKAGS GSKRIAYLRKKIAALKDNAN LEKDIANLENEIERLKEIKTLE NEVASHEQ (SEQ ID NO: 14)	SYYHHHHHLESTSLYKKAGSGS NLLATLRTSTAALLENENHVLEKEK EKLRLKEKEQQLNKLEAYK (SEQ ID NO: 13)
Beta-Ketothiolase (bktB)	6	GYF	PAH	PATSQHPPPPGHRSQAPSH (SEQ ID NO: 42)	ELNSLLILAEAYLERRDR (SEQ ID NO: 43)
HMG-COA Synthase	7	WW1A	WW1B	FQMPADTPPPAYLPFPEDPMT (SEQ ID NO: 44)	ERESNEEPPPPYEDPYWGNG (SEQ ID NO: 45)
HMG-COA Reductase	8	FOS	SYNZIP9	SYYHHHHHLESTSLYKKAGS GSQKVESLKQKIEELKQRKAQL KNDIANLEKEIAYAET (SEQ ID NO: 20)	SYYHHHHHLESTSLYKKAGSEFF RRERNKMAAKCRNRRRELTDTL QAETDQLEDEKSALQTEIANLLKE KEKLEFILAAHRCPACKIPDDLGFP EMSLE (SEQ ID NO: 19)
Mevalonate Kinase	9	VHS1	VHS2	VSSTKLVSFHDDSDEDLLHI (SEQ ID NO: 46)	AAATPISTFHDDSDEDLLHV (SEQ ID NO: 47)
Phosphomevalonate Kinase	10	SYNZIP13	SYNZIP15	SYYHHHHHLESTSLYKKAGS GSFENVTLEFILATLENENAKL RRLEAKLRELARLNEVAWL (SEQ ID NO: 24)	SYYHHHHHLESTSLYKKAGSGS QKVEELKNKIAELENRNAVKKNR VAHLKQEIAYLKDELAAHEFE (SEQ ID NO: 23)
Diphosphomevalonate Decarboxylase	11	MATH	SKP1	HDDSLPHPQQATDDSGHESD (SEQ ID NO: 48)	GSPNAGSVEQTPKKPGLRR (SEQ ID NO: 49)
Isopentenyl-Diphosphate Delta-Isomerase	12	SYNZIP5	SYNZIP6	SYYHHHHHLESTSLYKKAGS GSQKVAQLKNRVAVALKENA KLENIVARLENDNANLEKDIAN LEKDIANLERDVAR (SEQ ID NO: 28)	SYYHHHHHLESTSLYKKAGSGS NTVKELKNYIQELEERNAELKNLK EHLKFAKAELEFELAAHKFE (SEQ ID NO: 27)
Geranyl-Diphosphate Synthase	13	PDZ1	PDZ2	TDEEREETEEEVYLLNSTTL (SEQ ID NO: 50)	DGNVSGTQRQLDSATVRTYSC (SEQ ID NO: 51)
Olivetol Synthase	14	SH2A	SH2B	ALVDDAADYEPPPSNNEEAL (SEQ ID NO: 52)	RELFDDPSYVNQNLDKARQ (SEQ ID NO: 53)
Olivetolic Acid Cyclase	15	PTB1	PTB2	KNTKSMNFDNPVYRKTEE (SEQ ID NO: 54)	RSLPSTWIENKLYGMSDPNW (SEQ ID NO: 55)
CBGA Synthase	16	SH3A	SH3B	VVDNSPPPPALPPKKRQSAPS (SEQ ID NO: 56)	TORSKPQPAVPPRPSADLIL (SEQ ID NO: 57)
Acetyl-CoA Carboxylase	17	FAT	PEX	SATRELDEMASLSDFKIQC (SEQ ID NO: 58)	DLALSENWAQEFLAAGDAVD (SEQ ID NO: 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)₈ (SEQ ID NO:60), (G-S)₁₀ (SEQ ID NO:61), or (G-S)₁₅ (SEQ ID NO:62), or contain mainly glycine residues such as (GGGGS)₃ (SEQ ID NO:63) or (GGGGS)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)₂ (SEQ ID NO:65), (EAAAK)₃ (SEQ ID NO:66), or (EAAAK)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), AAAGGMPPAAGGM (SEQ ID NO:69), AAAGGM (SEQ ID NO:70), or PAAAGGMM (SEQ ID NO:71). See, e.g., U.S. Pat. No. 9,856,460.

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), KLSGGGGSGGGGS (SEQ ID NO:73), GSAGSAAGSGEFGSAAAKEAAAK-AGSAGSAAGSGEFGS (SEQ ID NO:74), GSAGSAAGSGFAEAAAKEAAAK-AGSAGSAAGSGEF (SEQ ID NO:75), or GSAGSAAGSGEFG-SAEAAAKEAAAKEAAAK-AGSAGSAAGSGEFGS (SEQ ID NO:76).

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:

(SEQ ID NO: 77)
 AEAWYNLGNAYYKQGDYQKAI EYYQKALEELDPNNAEAWYNLGNAYYKQGDYQKAI EYYQKALEELDPNNLQAEAWKNLGNAYYKQGDYQKAI EYYQKALEELDPNNNASAWYNLGNAYYKQGDYQKAI EYYQKALEELDPNNAKAWYRRGNAYYKQGDYQKAI EYYQKALEELDPNNRSRSA.

10 In some embodiments, the engineered enzyme can be of a formula: enzyme-linker₁-spacer-linker₂-motif₁-linker₃-motif₂, 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 KLSGGGGSGGGGS (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 GGGGSGGGSGGGAS (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

(SEQ ID NO: 74)
 GSAGSAAGSGEFGSAAAKEAAAKAGSAGSAAGSGEFGS.

30 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. 6B and FIGS. 13A-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^{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 (G₄S)₃, the spacer can be the cTPR6 sequence, and linker 3 can be (GS)₈.

40 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 GSAGSAAGSGEFAEAAAKEAAAK-AGSAGSAAGSGEF (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 GSAGSAAGSGEFG-SAEAAAKEAAAKEAAAK-AGSAGSAAGSGEFGS (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 GSAGSAAGSGEFG-SAEAAKEAAAKEAAKEAAAK-

AGSAGSAAGSGEFGSEQK LISEEDLEQKLISEED-LEQKLISEEDLGSAGSAAGSGEFGSAGSAAGSGEFGSAGS AAGSGEF (SEQ ID NO:80). For example, a C-terminal sequence with a triplicate FLAG tag can have the sequence GSAGSAAGSGEFG-

SAEAAKEAAAKEAAKEAAAK-

AGSAGSAAGSGEFGSDYK

DDDDKDYKDDDDKDYKDDDDKGSAGSAAGSGEF GSAGSAAGSGEFGSAGSAA GSGEF (SEQ ID NO:81). FIG. 6C and FIG. 13D 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. 13D contains an example of a scaffold polypeptide (see SCF gene cassette of FIG. 2B) containing a triplicate MYC tag. FIG. 6D and FIG. 13D 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.

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.

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.

As shown in FIGS. 1A and 1B, 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).

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.

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.

In the embodiment shown in FIGS. 1A and 1B, 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).

In the hexanoyl-CoA pathway shown in FIGS. 1A and 1B, 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-keto hexanoyl-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

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. 1A and 1B, and/or it can be non-scaffolded (soluble).

In the embodiment shown in FIGS. 1A and 1B, 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. 1A and 1B, 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.

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. 1A and 1B 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.

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 cytidyltransferase (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.

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.

In the embodiment shown in FIG. 1A and FIG. 1B, 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.

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.

As shown in FIG. 1A and FIG. 1B, 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 (CB-GAS). 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.

25

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.

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.

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.

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.

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.

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.

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.

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.

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

26

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.

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.

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.

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 cannabidiolic 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.

Host Cells for Producing Cannabinoids

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.

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.

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.

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.

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.

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^{WW}). Erg20^{WW} favors production of geranyl pyrophosphate over farnesyl

pyrophosphate. See, Jiang, et al., *Metab Eng.* 2017, 41:57-66. For example, a truncated HMGR (HMGR) 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'-cgccacccctacaacagaaggataaaa-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.

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).

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. 6A), or an ACL from *Rattus norvegicus*, *Mus musculus*, or *Ciona intestinalis*, e.g., GenBank Accession Nos. AAA74463, AAK56081, and BAB00624, respectively.

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. 6A), 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.

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. 6A), or a BHBD from *Escherichia coli*, *Treponema denticola*, or *Arabidopsis thaliana*, e.g., GenBank Accession Nos. AIZ91493, AAS11105, and AAN17431, respectively.

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. 6A), or an ECH from *Acinetobacter oleivorans*, *Cupriavidus necator*, or *Acinetobacter baumannii*, e.g., GenBank Accession Nos. ADI91469, CAJ91294, and ACJ57023, respectively.

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. 6A), or a bktB from *Escherichia coli*, *Lactobacillus casei*, or *Clostridium acetobutylicum*, e.g., GenBank Accession Nos. ALI39443, CAQ67083, and AAK80816, respectively.

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. 6A), or an ECR from *Cupriavidus necator*, *Saccharomyces cerevisiae*, or *Klebsiella michiganensis*, e.g., GenBank Accession Nos. AAP86010, DAA07148, and AIE72439, respectively.

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. 6A, 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.

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. 6A), or an HMGS from *Arabidopsis thaliana*, *Lactobacillus casei*, or *Homo sapiens*, e.g., GenBank Accession Nos. AEE83052, CAQ67081, and AAA62411, respectively.

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. 6A), or an HMGR from *Arabidopsis thaliana*, *Lactobacillus casei*, or *Homo sapiens*, e.g., GenBank Accession Nos. AEE35849, CAQ67082, and AAA52679, respectively.

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. 6A), or a mevalonate kinase from *Arabidopsis thaliana*, *Lactobacillus casei*, or *Homo sapiens*, e.g., GenBank Accession Nos. AAD31719, CAQ66794, and AAF82407, respectively.

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. 6A), or a mevalonate kinase from *Scheffersomyces stipitis*, *Lactobacillus casei*, or *Homo sapiens*, e.g., GenBank Accession Nos. EAZ63544, CAQ66339, and AAH06089, respectively.

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. 6A), or a diphosphomevalonate decarboxylase from *Arabidopsis thaliana*,

Lactobacillus casei, or *Homo sapiens*, e.g., GenBank Accession Nos. AAC67348, CAQ66795, and AAC50440, respectively.

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. 6A), or an isopentyl diphosphate isomerase from *Arabidopsis thaliana*, *Lactobacillus casei*, or *Homo sapiens*, e.g., GenBank Accession Nos. AAC49920, CAQ66796, and AAP35407, respectively.

For example, a geranyl pyrophosphate synthase (GPPS) (also known as a geranyl-diphosphate 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^{WW}) (see SEQ ID NO:96, FIG. 6A). Erg20^{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^{K179E}) can be used to produce a GPPS that favors production of geranyl pyrophosphate. See, WO2016010827A1.

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.

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.

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.

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.

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.

For example, an HMBPP synthase can have at least 70% sequence identity (e.g., at least 75%, 80%, 85%, 90%, 95%,

31

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.

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.

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. 6A), or an acetyl-CoA carboxylase from *Homo sapiens*, *Treponema denticola*, or *Cupriavidus necator*, e.g., GenBank Accession Nos. AAP94122, AAS11086, and CAQ67359, respectively.

For example, a pyruvate dehydrogenase (E1) and dihydrolipooyl 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.

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. 6A) or the OS from *C. sativa* having GenBank Accession No. BAG14339. See, for example, Taura, et al., *FEBS Letters* 583 (2009) 2061-2066.

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. 6A) 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.

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. 6A). 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.

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. 6A) 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).

For example, a cannabichromenic acid synthase (CB-CAS) 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

32

forth in SEQ ID NO:102 (FIG. 6A) 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.

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).

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 B12seq program can be found in the readme file accompanying BLASTZ. B12seq performs a comparison between two amino acid sequences using the BLASTP algorithm. To compare two amino acid sequences, the options of B12seq 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 followed for nucleic acid sequences except that blastn is used.

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.

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. 12A are

the nucleic acid sequences encoding an ATP citrate lyase, an atoB, a 3-hydroxybutyryl-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^{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. 14A-14C contain codon optimized (for expression in yeast) nucleic acid sequences encoding the engineered enzymes of FIGS. 13A-13C.

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.

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. 2A and FIG. 2B each provide a non-limiting schematic of suitable gene cassettes for expressing the scaffolds and enzymes. FIG. 12C 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. 14D for the codon-optimized nucleic acid sequence encoding the scaffold polypeptide of FIG. 13D. FIG. 12D 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. 14D.

In some embodiments, multiple nucleic acids encoding polypeptides (e.g., the nucleic acids of a gene cassette such as in FIG. 2A or FIG. 2B) 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.

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.

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 biostatic delivery are common methods for introducing nucleic acid into bacteria and yeast cells. See, e.g., Ito et al., *J. Bacteriol.* 153:163-168 (1983); Durrens et al., *Curr. Genet.* 18:7-12 (1990); and Becker and Guarente, *Methods in Enzymology* 194:182-187 (1991).

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.

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).

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-1 α (pTEF1), the pro-

motor 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. 2A, 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. 2A), the pTEF1 promoter has the strongest activity (+++ in FIG. 2A), and the activity of the pTPI1 promoter is between the other two (++) in FIG. 2A). In some embodiments, the Gal 1-10 promoter (e.g., from *S. cerevisiae*) can be used. See, e.g., FIG. 17.

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. 1B.

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.

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 or 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.

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. 6C and FIG. 6D, a scaffold can include a myc tag (e.g., (Myc)₃ tag) or a FLAG tag (FLAG)₃ tag at the C-terminus.

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.

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).

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 FokI 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.

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

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.

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.

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.

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.

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.

Producing Cannabinoids Using a Recombinant Host

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.

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 confluence 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.

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.

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).

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

Enzymatic Constructs

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 sca-

fold. 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.

5 Scaffolding Constructs

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.

10 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 15 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 20 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 25 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/ 30 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. Assessment of Cannabinoideric Potential by Transient 35 Transfection

Competent yeast and/or green algae cells are transiently transfected with plasmids encoding various permutations of the scaffold and enzymes. To establish baseline cannabinoideric capacity, cells first undergo transient transfection with 40 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 cannabinoideric 45 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 50 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 55 blot analysis of each co-immunoprecipitated enzyme.

Engineering of Stable Cannabinoideric Cell Lines

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, 60 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 65 improvement in cannabinoideric capacity conferred by multi-enzymatic scaffolding for stably engineered cannabinoideric cell lines, cannabinoid biosynthesis will be com-

pared among cells that are stimulated for enzyme but not scaffold expression and cells that are stimulated for enzyme and scaffold expression.

Validation of Multi-enzymatic Scaffolding

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.

Aerobic Fed-batch Fermentation

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).

Cannabinoid Extraction, Isolation, and Analytical Characterization

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 rotovapitated. 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.

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

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. 2B. The cassettes collectively encode all scaffold-binding engineered enzymes and the polypeptide scaffolds to which the engineered enzymes can bind.

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. 13A. The GPP gene

cassette encoded scaffold-binding engineered enzymes for scaffolded geranyl pyrophosphate (GPP) biosynthesis, namely HMGS, tHMGR, ERG12, ERGS, MVD1, IDI1, and ERG20^{WW}. See, FIG. 13B. 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. 13C. The SCF gene cassette encoded the polypeptide scaffolds for bidirectional scaffolded cannabinoid biosynthesis and scaffolded malonyl-CoA biosynthesis, namely the cannabinoidergic metabolon scaffold (CBSFC) 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. 13D. 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. 13A for the amino acids sequences of the ACL and atoB engineered enzymes.

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. 14A-14D 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. 15A-15E 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 5' SpeI/XbaI 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).

TABLE 3

HCA Gene Cassette				
Gene ID	Cassette Position	pCCI-Brick #1 ID	Yeast Vector	Yeast Vector ID
ACL	1	pHCA	pESC-TRP	vHCA
atoB	2			
BHBD	3			
ECH	4			
ECR	5			

TABLE 3-continued

bktB	6			
HCS	7			
MVA Gene Cassette				
Gene ID	Cassette Position	pCCI-Brick #2 ID	Yeast Vector	Yeast Vector ID
HMGS	1	pGPP	pESC-LEU	vGPP
tHMGR	2			
ERG12	3			
ERG8	4			
MVD1	5			
IDI1	6			
ERG20 ^{WW}	7			
CAN Gene Cassette				
Gene ID	Cassette Position	pCCI-Brick #3 ID	Yeast Vector	Yeast Vector ID
OS	1	pCAN	pESC-HIS	vCAN
OAC	2			
CBGAS	3			
CBDAS	4			
CBCAS	5			
ACC	6			
SCFLD Gene Cassette				
Gene ID	Cassette Position	pCCI-Brick #4 ID	Yeast Vector	Yeast Vector ID
CBSCF	1	pSCF	pESC-URA #1	vSCF
MCASCF	2			
ACL	3			
atoB	4			
NSCFLD Gene Cassette				
Gene ID	Cassette Position	pCCI-Brick #5 ID	Yeast Vector	Yeast Vector ID
ACL	1	pSOL	pESC-URA #2	vSOL
atoB	2			

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. 13A-13D, the codon-optimized nucleotide sequence fragments comprising each synthetic gene cassette are detailed in FIGS. 14A-14D, 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. 15A-15E, 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

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.

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_{600nm}=0.3 (stationary phase) and grown to OD_{600nm}=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 2x with 50 mL chilled nuclease-free water and 1x with 50 mL chilled electroporation buffer (1M sorbitol/1 mM CaCl₂). 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 1x 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.

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_{Parent}) were selected for by resuspension and incubation in tryptophan-, leucine-, and histidine-deficient YNB media.

The yCB_{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_{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_{SCF}) were selected for by resuspension and incubation in tryptophan-, leucine-, histidine-, and uracil-deficient YNB media. The second of the split yCB_{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 (desig-

45

nated yCB_{SOL}) were also selected for by resuspension and incubation in tryptophan-, leucine-, histidine-, and uracil-deficient YNB media.

To quantify the improvement in cannabinoidergic capacity conferred by multi-enzymatic scaffolding, cannabinoid titers were compared between triplicate yCB_{SOL} and yCB_{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_{SOL} and yCB_{SCF} , each culture was initially diluted to $OD_{600nm}=0.3$, and OD_{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_{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.

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. 19A-19E, 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, yCB_{SCF} 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

To evaluate the impacts of culture media supplementation with citrate and hexanoate precursors, cannabinoid titers were compared between triplicate yCB_{SOL} and yCB_{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 ($F_{1,4}=457.5$,

46

$p<0.0001$) 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, yCB_{SCF} but not $yCBSOL$ cultures exhibited increased total cannabinoid titers when cultured in media supplemented with 300 mg/L citrate ($p<0.0001$). Neither yCB_{SCF} 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 yCB_{SCF} and $yCBSOL$ cultures. Moreover, relative to $yCBSOL$ cultures, yCB_{SCF} 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 ($p<0.0001$).

To delineate concentration-response relationships for the supplementation of culture media with citrate, cannabinoid titers were compared between triplicate yCB_{SOL} and yCB_{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_{Max}) and citrate EC_{50} for cannabinoid biosynthesis in yCB_{SOL} and yCB_{SCF} cultures. Concentration-response curves, CB_{Max} estimates, and citrate EC_{50} estimates are depicted in FIGS. 21A and 21B. 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_{Max} and citrate EC_{50}). Compared to yCB_{SOL} cultures, yCB_{SCF} cultures exhibited markedly increased CB_{Max} ($p<0.0001$) and citrate EC_{50} ($p<0.001$) estimates.

OTHER EMBODIMENTS

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.

SEQUENCE LISTING

```

Sequence total quantity: 212
SEQ ID NO: 1      moltype = AA length = 11
FEATURE
source
1..11
mol_type = protein
organism = synthetic construct
SEQUENCE: 1
PPPALPPKRR R

SEQ ID NO: 2      moltype = AA length = 32
FEATURE
source
1..32
mol_type = protein
organism = synthetic construct
SEQUENCE: 2
LVGALMHVMQ KRSRAIHSSD EGEDQAGDED ED

SEQ ID NO: 3      moltype = AA length = 43
FEATURE
source
1..43

```

11

32

-continued

```

mol_type = protein
organism = synthetic construct
SEQUENCE: 3
ITIRAAFLEK ENTALRTEIA ELEKEVGRCE NIVSKYETRY GPL 43

SEQ ID NO: 4      moltype = AA length = 43
FEATURE          Location/Qualifiers
source           1..43
mol_type = protein
organism = synthetic construct

SEQUENCE: 4
LEIRAAFLEK ENTALRTRAA ELRKRVGRCR NIVSKYETRY GPL 43

SEQ ID NO: 5      moltype = AA length = 70
FEATURE          Location/Qualifiers
source           1..70
mol_type = protein
organism = synthetic construct

SEQUENCE: 5
SYYHHHHHHHL ESTSLYKKAG SGSNLVAQLE NEVASLENEN ETLKKNLHK KDLIAYLEKE 60
IANLRKKIEE      70

SEQ ID NO: 6      moltype = AA length = 73
FEATURE          Location/Qualifiers
source           1..73
mol_type = protein
organism = synthetic construct

SEQUENCE: 6
SYYHHHHHHHL ESTSLYKKAG SGSNARNAYLR KKIARLKKDN LQLERDEQNL EKIIANLRDE 60
IARLENEVAS HEQ      73

SEQ ID NO: 7      moltype = AA length = 65
FEATURE          Location/Qualifiers
source           1..65
mol_type = protein
organism = synthetic construct

SEQUENCE: 7
SYYHHHHHHHL ESTSLYKKAG SGSNEVTITLE NDAAFIENEN AYLEKEIARL RKEKAALRNR 60
LAHKK            65

SEQ ID NO: 8      moltype = AA length = 77
FEATURE          Location/Qualifiers
source           1..77
mol_type = protein
organism = synthetic construct

SEQUENCE: 8
SYYHHHHHHHL ESTSLYKKAG SGSSQKVAELK NRVAVKLNRN EQLKNKVEEL KNRNAYLKNE 60
LATLENEVAR LENDVAE      77

SEQ ID NO: 9      moltype = AA length = 64
FEATURE          Location/Qualifiers
source           1..64
mol_type = protein
organism = synthetic construct

SEQUENCE: 9
ENLYFQGENL YFQGDSSESC WNCGRKASET CSGCNTARYC GSFCQHKDWE KHHHICGQTL 60
QAQQ            64

SEQ ID NO: 10     moltype = AA length = 145
FEATURE          Location/Qualifiers
source           1..145
mol_type = protein
organism = synthetic construct

SEQUENCE: 10
MAVSESQQLKK MAVSKYKYRDL TVRETNVIT LYKDLKPVLQ SYVFNDGSSR ELMNLTGTIP 60
VPYRGNTYNI PICLWLLDTY PYNPPICFKV PTSSMTIKTG KHVDANGKIY LPYLHEWKHP 120
QSDLLGLIQLV MIVVFGDEPP VFSRP      145

SEQ ID NO: 11     moltype = AA length = 88
FEATURE          Location/Qualifiers
source           1..88
mol_type = protein
organism = synthetic construct

SEQUENCE: 11
GPLGSPLTAS MLASAPPQEQ KQMLGERLFP LIQAMHPTLA GKITGMLLEI DNSELLHMLE 60
SPESLRSKVD EAVAVLQAHQ AKEAAQKA      88

SEQ ID NO: 12     moltype = AA length = 107
FEATURE          Location/Qualifiers

```

-continued

source 1..107
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 12
 NTMMSVPTDG AVTTSQIPAS EQETLVRPKP LLLKLLKSVG AQKDTYTMKE VLFYLGQYIM 60
 TKRLYDEKQQ HIVYCSNDLL GDLFGVPSFS VKEHRKIYTM IYRNLVV 107

SEQ ID NO: 13 moltype = AA length = 65
 FEATURE Location/Qualifiers
 source 1..65
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 13
 SYYHHHHHHHL ESTSLYKKAG SGSNLLLATLR STAAVLENEN HVLEKEKEKL RKEKEQLINK 60
 LEAYK 65

SEQ ID NO: 14 moltype = AA length = 73
 FEATURE Location/Qualifiers
 source 1..73
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 14
 SYYHHHHHHHL ESTSLYKKAG SGSKRIAYLR KKIAALKKDAN ANLEKDIANL ENEIERLIKE 60
 IKTLENEVAS HEQ 73

SEQ ID NO: 15 moltype = AA length = 62
 FEATURE Location/Qualifiers
 source 1..62
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 15
 DVMWEYKWEN TGDAELYGPF TSAQMOTWVS EGYFPDGVYC RKLDPPGGQF YNSKRIDFDL 60
 YT 62

SEQ ID NO: 16 moltype = AA length = 85
 FEATURE Location/Qualifiers
 source 1..85
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 16
 ESDSVEFNNA ISYVNKIKTR FLDHPEIYRS FLEILHTYQK EQLHTKGRPF RGMSEEVFT 60
 EVANLFRGQE DLLSEFGQFL PEAKR 85

SEQ ID NO: 17 moltype = AA length = 36
 FEATURE Location/Qualifiers
 source 1..36
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 17
 LGPLPPGWEV RSTVSGRIYF VDHNNRTTQF TDPR LH 36

SEQ ID NO: 18 moltype = AA length = 37
 FEATURE Location/Qualifiers
 source 1..37
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 18
 GAMGPLPPGW EKRTDSNGRV YFVNHNTRIT QWEDPRS 37

SEQ ID NO: 19 moltype = AA length = 101
 FEATURE Location/Qualifiers
 source 1..101
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 19
 SYYHHHHHHHL ESTSLYKKAG SEFFRRERNK MAAAKCRNRR RELTDTLQAE TDQLEDEKSA 60
 LQTEIANLLK EKEKLEFILA AHRPACKIPD DLGFPEEMSL E 101

SEQ ID NO: 20 moltype = AA length = 59
 FEATURE Location/Qualifiers
 source 1..59
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 20
 SYYHHHHHHHL ESTSLYKKAG SGSKQVKESLK QKIEELKQRK AQLKNDIANL EKEIAYAET 59

SEQ ID NO: 21 moltype = AA length = 147
 FEATURE Location/Qualifiers
 source 1..147

-continued

```

mol_type = protein
organism = synthetic construct

SEQUENCE: 21
MEPAMEPETL EARINRATNP LNKELDWASI NGFCEQLNED FEGPPLATRL LAHKIQSPQE 60
WEAIQALTVL ETCMKSCGKR FHDEVGKFRF LNELIKVNSP KYLGSRTSEK VKNKILELLY 120
SWTVEGLPPEEV KIAEAYQMLK KQGIVKS 147

SEQ ID NO: 22      moltype = AA length = 171
FEATURE           Location/Qualifiers
source            1..171
mol_type = protein
organism = synthetic construct

SEQUENCE: 22
GAMGSMAAAE GESLESWLNK ATNPSNRQED WEIIGFCDO INKELEGPOI AVRLLAHKIQ 60
SPOEWEALQA LTVLEACMKN CGRRFHNEVG KFRFLNELIK VVSPKVYLGDR VSEKVTKVI 120
ELLYSWTMAL PEEAKIKDAY HMLKRQGIVQ SDPPPIPVDRT LIPSPPPRPK N 171

SEQ ID NO: 23      moltype = AA length = 67
FEATURE           Location/Qualifiers
source            1..67
mol_type = protein
organism = synthetic construct

SEQUENCE: 23
SYYHHHHHHHL ESTSLYKKAG SGSSQKVEELK NKIAELENRN AVKKNRVAHL KQEIAYLKDE 60
LAAHEFE                                     67

SEQ ID NO: 24      moltype = AA length = 64
FEATURE           Location/Qualifiers
source            1..64
mol_type = protein
organism = synthetic construct

SEQUENCE: 24
SYYHHHHHHHL ESTSLYKKAG SGSSFENVTHE FILATLENEN AKLRRLEAKL ERELARLRNE 60
VAWL                                         64

SEQ ID NO: 25      moltype = AA length = 168
FEATURE           Location/Qualifiers
source            1..168
mol_type = protein
organism = synthetic construct

SEQUENCE: 25
AMADLEQKVL EMEASTYDGV FIWKISDFPR KRQEAVAGRI PAIFSPAFTY SRYGYKMCLR 60
IYLNQDGDTGR GTHLSLFFVV MKGPNDALLR WPFNQKVTLN LLDQNNREHV IDAFRPDVTS 120
SSFQRPVNMD NIASGCPFLC PVSKMEAKNS YVRDDAIFIK AIVDLTGL 168

SEQ ID NO: 26      moltype = AA length = 159
FEATURE           Location/Qualifiers
source            1..159
mol_type = protein
organism = synthetic construct

SEQUENCE: 26
ASIKLQSSDG EIFEVDVIEIA KQSVTIKMTL EDLGMDDDEG DDPVPLPNVN AAILKKVIQW 60
CTHHKDDPPP PEDDENKEKR TDDIPVWDQE FLKVDQGTLF ELILAANYLD IKGLLDVTCK 120
TVANMIKGKT PEEIRKTFNI KNDTEEEEQ QVRKENQWC 159

SEQ ID NO: 27      moltype = AA length = 67
FEATURE           Location/Qualifiers
source            1..67
mol_type = protein
organism = synthetic construct

SEQUENCE: 27
SYYHHHHHHHL ESTSLYKKAG SGSSNTVKELK NYIQELEERN AELKNLKEHL KFAKAELEFE 60
LAAHKFE                                     67

SEQ ID NO: 28      moltype = AA length = 77
FEATURE           Location/Qualifiers
source            1..77
mol_type = protein
organism = synthetic construct

SEQUENCE: 28
SYYHHHHHHHL ESTSLYKKAG SGSSQKVAQLK NRVAYKLKEN AKLENIVARL ENDNANLEKD 60
IANLEKDIAN LERDVAR 77

SEQ ID NO: 29      moltype = AA length = 81
FEATURE           Location/Qualifiers
source            1..81
mol_type = protein
organism = synthetic construct

SEQUENCE: 29

```

-continued

LCTMKKGPSG YGFNLHSDKS KPGQFIRSV PDSPAEASGL RAQDRIVEVN GVCMEGKQHG 60
DVVS AIRAGG DETKLLVVDR E 81

SEQ ID NO: 30 moltype = AA length = 97
FEATURE Location/Qualifiers
source 1..97
mol_type = protein
organism = synthetic construct

SEQUENCE: 30 SSGAIYTVE LKRYGGPLGI TISGTEEPFD PIISSLTKG GLAERTGAIH IGDRILAINS 60
SSLKGKPLSE AIHLLQMAGE TVTLKIKKQT DAQPASS 97

SEQ ID NO: 31 moltype = AA length = 109
FEATURE Location/Qualifiers
source 1..109
mol_type = protein
organism = synthetic construct

SEQUENCE: 31 GNNTLEYEWY NKSISRDKAEL KLLDTGKEG AFMVRDSRTP GTYTVSVFTK AIISENPCIK 60
HYHIKETNDS PKRYYVAEKY VFDSIPLLIQ YHQYNGGGLV TRLRYPVCG 109

SEQ ID NO: 32 moltype = AA length = 104
FEATURE Location/Qualifiers
source 1..104
mol_type = protein
organism = synthetic construct

SEQUENCE: 32 GSHPWFFGKI PRAKAEEMLS KQRHDGAFLI RESESAPGDF SLSVKFGNDV QHFVLRDGA 60
GKYFLWVVKF NSLNELVVDYH RSTSVSRNQQ IFLRDIEQVP QQPT 104

SEQ ID NO: 33 moltype = AA length = 152
FEATURE Location/Qualifiers
source 1..152
mol_type = protein
organism = synthetic construct

SEQUENCE: 33 GQDRSEATLI KRFKGEGVRY KAKLIGIDEV SAARGDKLCQ DSMMKLGVV AGARSKGEHK 60
QKIFLTISFG GIKIFDEKTG ALQHHHAVHE ISYIAKDTID HRAFGYVCGK EGNHRFVAIK 120
TAQAAEPPVIL DLRDLFQLIY ELKQREELEK KA 152

SEQ ID NO: 34 moltype = AA length = 57
FEATURE Location/Qualifiers
source 1..57
mol_type = protein
organism = synthetic construct

SEQUENCE: 34 AEYVRALFDF NGNDEEDLPF KKGDILRIRD KPEEQWWNAE DSEGKRGMP VPYVEKY 57

SEQ ID NO: 35 moltype = AA length = 60
FEATURE Location/Qualifiers
source 1..60
mol_type = protein
organism = synthetic construct

SEQUENCE: 35 LIKHMRAEAL FDFTGNSKLE LNFKAGDVIF LLSRINKDWL EGTVRGATGI FPLSFVKILK 60

SEQ ID NO: 36 moltype = AA length = 153
FEATURE Location/Qualifiers
source 1..153
mol_type = protein
organism = synthetic construct

SEQUENCE: 36 GSHMRLGAQS IQPTANLDRT DDLVLYNVME LVRAVLELKN ELAQLPPEGY VVVVKNVGLT 60
LRKLIGSVDD LLPSLPLSSSR TEIECTQKLL NKDLAELINK MRLAQONAVT SLSEECKRQM 120
LTASHTLAVD AKNLLDAVDQ AKVLANLAHP PAE 153

SEQ ID NO: 37 moltype = AA length = 70
FEATURE Location/Qualifiers
source 1..70
mol_type = protein
organism = synthetic construct

SEQUENCE: 37 GAMATPGSEN VLPREPLIAT AVKFLQNSRV RQSPLATRRA FLKKKGLTDE EIDMAFQQSG 60
TAADEPSSLW 70

SEQ ID NO: 38 moltype = AA length = 20
FEATURE Location/Qualifiers
source 1..20
mol_type = protein

-continued

SEQUENCE: 38	organism = synthetic construct	
RPPTISNPPP LISSAKHPSV		20
SEQ ID NO: 39	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
SEQUENCE: 39	organism = synthetic construct	
NFLQSRPEPT APPEESFRSG		20
SEQ ID NO: 40	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
SEQUENCE: 40	organism = synthetic construct	
SKGTGLNPNA KVVQEIAPGN		20
SEQ ID NO: 41	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
SEQUENCE: 41	organism = synthetic construct	
PDGGTTFEHL WSSLEPDSTY		20
SEQ ID NO: 42	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
SEQUENCE: 42	organism = synthetic construct	
PATSQHPPP PGHRSQAPSH		20
SEQ ID NO: 43	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
SEQUENCE: 43	organism = synthetic construct	
ELNSLLLILE AAEYLERRDR		20
SEQ ID NO: 44	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
SEQUENCE: 44	organism = synthetic construct	
FQMPADTPPP AYLPPEDPMT		20
SEQ ID NO: 45	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
SEQUENCE: 45	organism = synthetic construct	
ERESNEEPPP PYEDPYWGNG		20
SEQ ID NO: 46	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
SEQUENCE: 46	organism = synthetic construct	
VSSSTKLVSFH DDSDEDLLHI		20
SEQ ID NO: 47	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
SEQUENCE: 47	organism = synthetic construct	
AAATPISTFH DDSDEDLLHV		20
SEQ ID NO: 48	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	

-continued

```

mol_type = protein
organism = synthetic construct
SEQUENCE: 48
HDDSLPHPQQ ATDDSGHESD                                         20

SEQ ID NO: 49          moltype = AA length = 20
FEATURE
source
1..20
mol_type = protein
organism = synthetic construct
SEQUENCE: 49
GSPNAGSVEQ TPKKPGLRRR                                         20

SEQ ID NO: 50          moltype = AA length = 20
FEATURE
source
1..20
mol_type = protein
organism = synthetic construct
SEQUENCE: 50
TDEEREETEE EVYLLNSTTL                                         20

SEQ ID NO: 51          moltype = AA length = 20
FEATURE
source
1..20
mol_type = protein
organism = synthetic construct
SEQUENCE: 51
DGNVSGTQRL DSATVRTYSC                                         20

SEQ ID NO: 52          moltype = AA length = 20
FEATURE
source
1..20
mol_type = protein
organism = synthetic construct
SEQUENCE: 52
ALVDDAADYE PPPSNNEEAL                                         20

SEQ ID NO: 53          moltype = AA length = 20
FEATURE
source
1..20
mol_type = protein
organism = synthetic construct
SEQUENCE: 53
RELFDDPSYV NVQNLDKARQ                                         20

SEQ ID NO: 54          moltype = AA length = 20
FEATURE
source
1..20
mol_type = protein
organism = synthetic construct
SEQUENCE: 54
KNTKSMNFDN PVYRKTTEEE                                         20

SEQ ID NO: 55          moltype = AA length = 20
FEATURE
source
1..20
mol_type = protein
organism = synthetic construct
SEQUENCE: 55
RSLPSTWIEN KLYGMSDPNW                                         20

SEQ ID NO: 56          moltype = AA length = 20
FEATURE
source
1..20
mol_type = protein
organism = synthetic construct
SEQUENCE: 56
VVDNSPPPALL PPKKRQSAPS                                         20

SEQ ID NO: 57          moltype = AA length = 20
FEATURE
source
1..20
mol_type = protein
organism = synthetic construct
SEQUENCE: 57
TQRSKPQPAV PPRPSADLIL                                         20

SEQ ID NO: 58          moltype = AA length = 20
FEATURE

```

-continued

source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 58		
SATRELDELM ASLSDFKIQG		20
SEQ ID NO: 59	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 59		
DLALSENWAQ EFLAAGDAVD		20
SEQ ID NO: 60	moltype = AA length = 16	
FEATURE	Location/Qualifiers	
source	1..16	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 60		
GSGSGSGSGS GSGSGS		16
SEQ ID NO: 61	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 61		
GSGSGSGSGS GSGSGSGSGS		20
SEQ ID NO: 62	moltype = AA length = 30	
FEATURE	Location/Qualifiers	
source	1..30	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 62		
GSGSGSGSGS GSGSGSGSGS		30
SEQ ID NO: 63	moltype = AA length = 15	
FEATURE	Location/Qualifiers	
source	1..15	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 63		
GGGGSGGGGS GGGGS		15
SEQ ID NO: 64	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 64		
GGGGSGGGGS GGGGS		20
SEQ ID NO: 65	moltype = AA length = 10	
FEATURE	Location/Qualifiers	
source	1..10	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 65		
EAAAKEAAAK		10
SEQ ID NO: 66	moltype = AA length = 15	
FEATURE	Location/Qualifiers	
source	1..15	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 66		
EAAAKEAAAK EAAAK		15
SEQ ID NO: 67	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 67		
EAAAKEAAAK EAAAKEAAAK		20
SEQ ID NO: 68	moltype = AA length = 6	

-continued

FEATURE source	Location/Qualifiers 1..6 mol_type = protein organism = synthetic construct	
SEQUENCE: 68 AAAGGM		6
SEQ ID NO: 69 FEATURE source	moltype = AA length = 14 Location/Qualifiers 1..14 mol_type = protein organism = synthetic construct	
SEQUENCE: 69 AAAGGMPAA AGGM		14
SEQ ID NO: 70 FEATURE source	moltype = AA length = 6 Location/Qualifiers 1..6 mol_type = protein organism = synthetic construct	
SEQUENCE: 70 AAAGGM		6
SEQ ID NO: 71 FEATURE source	moltype = AA length = 9 Location/Qualifiers 1..9 mol_type = protein organism = synthetic construct	
SEQUENCE: 71 PPAAGGMM		9
SEQ ID NO: 72 FEATURE source	moltype = AA length = 12 Location/Qualifiers 1..12 mol_type = protein organism = synthetic construct	
SEQUENCE: 72 GSAGSAAGSG EF		12
SEQ ID NO: 73 FEATURE source	moltype = AA length = 18 Location/Qualifiers 1..18 mol_type = protein organism = synthetic construct	
SEQUENCE: 73 KLSGGGGSGG GGSGGGGS		18
SEQ ID NO: 74 FEATURE source	moltype = AA length = 40 Location/Qualifiers 1..40 mol_type = protein organism = synthetic construct	
SEQUENCE: 74 GSAGSAAGSG EFGSAEEAK EAAAKAGSAG SAAGSGEFGS		40
SEQ ID NO: 75 FEATURE source	moltype = AA length = 36 Location/Qualifiers 1..36 mol_type = protein organism = synthetic construct	
SEQUENCE: 75 GSAGSAAGSG EFAEAAAKEA AAKAGSAGSA AGSGEF		36
SEQ ID NO: 76 FEATURE source	moltype = AA length = 50 Location/Qualifiers 1..50 mol_type = protein organism = synthetic construct	
SEQUENCE: 76 GSAGSAAGSG EFGSAEEAK EAAAKEAAAK EAAAKAGSAG SAAGSGEFGS		50
SEQ ID NO: 77 FEATURE source	moltype = AA length = 211 Location/Qualifiers 1..211 mol_type = protein organism = synthetic construct	
SEQUENCE: 77 AEAWYNLGNA YYKQGDYQKA IEYYQKALEL DPNNAEAWYN LGNAYYKQGD YQKAIEYYQK 60 AELDLPNNAE AWYNLGNA YYKQGDYQKAIE DYQKALELDP NNQAEAWKN LGNAYYKQGD 120		

-continued

YQKAIIEYYQK ALELDPNNAS AWYNLGNAYY KQGDYQKAIE YYQKALELDP NNAKAWYRRG NAYYKQGDYQ KAIEDYQKAL ELDPPNNSRSR A	180 211
SEQ ID NO: 78 FEATURE source SEQUENCE: 78 GGGGSGGGGS GGGGAS	moltype = AA length = 16 Location/Qualifiers 1..16 mol_type = protein organism = synthetic construct 16
SEQ ID NO: 79 FEATURE source SEQUENCE: 79 GSAGSAAGSG EFGSAGSAAG SGEFGSAGSA AGSGEF	moltype = AA length = 36 Location/Qualifiers 1..36 mol_type = protein organism = synthetic construct 36
SEQ ID NO: 80 FEATURE source SEQUENCE: 80 GSAGSAAGSG EFGSAEAAA AK EAAAKAEEAAK EAAAAGSAG SAAGSGEFGS EQKLISEEDL EQKLISEEDL EQKLISEEDL GSAGSAAGSG EFGSAGSAAG SGEFGSAGSA AGSGEF	moltype = AA length = 116 Location/Qualifiers 1..116 mol_type = protein organism = synthetic construct 60 116
SEQ ID NO: 81 FEATURE source SEQUENCE: 81 GSAGSAAGSG EFGSAEAAA AK EAAAKAEEAAK EAAAAGSAG SAAGSGEFGS DYKDDDDKDY KDDDKDVKD DDDKGSAGSA AGSGEFGSAG SAAGSGEFGS AGSAAGSGEF	moltype = AA length = 110 Location/Qualifiers 1..110 mol_type = protein organism = synthetic construct 60 110
SEQ ID NO: 82 FEATURE source SEQUENCE: 82 cggcacccct acaaacagaa ggaatataaa	moltype = DNA length = 30 Location/Qualifiers 1..30 mol_type = other DNA organism = synthetic construct 30
SEQ ID NO: 83 FEATURE source SEQUENCE: 83 MSAKAISEQT GKELLYKFIC TTSAIQNRFK YARVTPDTW ARLLQDHFWL LSQNLVVKPD QLIKRRGKLG LVGVNLTLDG VKSWLKPRLG QEATVGKATG FLKNFLIEPF VPHSQAEFY VCIYATREGD YLVFFHHEGGV DVGDVDAKAQ KLLVGVDEKL NPEDIKKHLL VHAPEDKKEI LASFISGLFN FYEDLYFTYI EINPLVVTKD GVYVLIDLAAK VDATADYICK VKGWDIEFPP PFGREAYPEE AYIADLDAKS GASLKLTLIN PKGRGIWTMVA GGGASVYSD TICDLGGVNE LANYGEYSGA PSEQQTYDYA KTILSLMTRE KHPDGKILII GGSIANFTNV AATFKGIVRA IRDYQGPLKE HEVTITFVRG GPNYQEGLRV MGEVGKTTG I PIHVFGTEH MTAIVGMALG HRPIPQNQPKT AAHTANFLLN ASGSTSTPAP SRTASFSESR ADEVAPAKKA KPAMPQDSVP SPRSLQGKST TLFSRHTKAI VWGMQTRAVQ GMLDFDYZCS RDEPSVAAMV YPFTGDHKQK FYWHGHKEILY PVFKNMADAM RKHPEVDVLI NFASLRSAYD STMETMNYAQ IRTIAIIAEG IPEALTRKLI KKADQKGVTI IGPATVGGIK PGCFKIGNTG GMLDNILASK LYRPGSVAYV SRSGGMSNEL NNIISRTTDG VYEGVAIGGD RYPGSTFMHD VLRYQDTPGV KMIVVLGEIG GTEYKICRG IKGRLTKP VWCWICGTCAT MFSSEVQPGH AGACANQASE TAVAKNQALK EAGVFVPRSF DELGEIQSV YEDLVANGVI VPAQEVPPT VPMDYSWARE LGLIRKPASF MTSICDERGQ ELIYAGMPIT EVFKEEMGIG GVLGLLWQK RLPKYSQOFI EMCLMVTADH GPAVSGAHNT IICARAGKDL VSSLTSGLLT IGDRFGGALD AAAKMFSAF DSGIIPMEFV NKMKKEGKLI MGIGHRVKSI NNPDMDRVQIL KDYVRQHFPA TPLLGYALEV EKITTSKPN LILNVVDGLIG VAFVDMLRNC GSFTREEADE YIDIGALNGI FVLGRSMGFI GHYLDQKRLK QGLYRHPWDD ISYVLPEHMS M	60 120 180 240 300 360 420 480 540 600 660 720 780 840 900 960 1020 1080 1101
SEQ ID NO: 84 FEATURE source SEQUENCE: 84 MKNCVIVSAV RTAIGSFNGS LASTSAIDLG ATVIKAAIER AKIDSQHVDE VIMGNVLQAG LGQNPARQAL LKSGLAEATVC GFTVNKVCGS GLKSVVALAAQ AIQAGQAQSI VAGGMENMSL	moltype = AA length = 394 Location/Qualifiers 1..394 mol_type = protein organism = Escherichia coli 60 120

-continued

APYLLDAKAR	SGYRLGDGQV	YDVILRDGLM	CATHGYHMGI	TAENVAKEYG	ITREMQDELA	180
HSQRKAAAAA	IESTGAFTAEI	VPVNVTTRKK	TFVFSQDEFP	KANSTAAELG	ALRPAFDKAG	240
TVTAGNASGI	NDGAAALVIM	EESAALAAGL	TPLARIKSYA	SGGVPPALMG	MGPVPATQKA	300
LQLAGLQLAD	IDLIEANEAF	AAQFLAVGKN	LGFDFSEKVN	NNGAIALGHP	IGASGARILV	360
TLLHAMQARD	KTLGLATLCI	GGGQGIAMVI	ERLN			394

SEQ ID NO:	85	moltype = AA	length = 282
FEATURE		Location/Qualifiers	
source		1..282	
		mol_type = protein	
		organism = Clostridium acetobutylicum	

SEQUENCE:	85					
MMKVCVIGAG	TMGSGIAQAF	AAKGFEVVLR	DIKDEFVDRG	LDFINKNL SK	LVKKGKIEEA	60
TKVEILTRIS	GTVDLNMAAD	CDLVIEAAVE	RMDIKKQIFA	DLDNICKPET	ILASNNTSLS	120
ITEVASATKR	PDKVIGHMFF	NPAPVMKLV E	VIRGIATSQE	TFDAVKETSI	AIGKDPVVEA	180
EAPGFVNRI	LIPMINEAVG	ILAEGIASVE	DIDKAMKLGA	NHPMGPLELG	DFIGLDICLA	240
IMDVLYSETG	DSKYRPHTL	KKYVRAGWL G	RKSGKGFYDY	SK		282

SEQ ID NO:	86	moltype = AA	length = 261
FEATURE		Location/Qualifiers	
source		1..261	
		mol_type = protein	
		organism = Clostridium acetobutylicum	

SEQUENCE:	86					
MELNNVILEK	EGKVAVVTIN	RPKNALNLS	DTLKEMDYVI	GEIENDSEVL	AVILTGAGEK	60
SFVAGADISE	MKEMNTIEGR	KFGILGNKVF	RRLELLEKP V	IAAVNGFALG	GGCEIAMSCD	120
IRIASSNARF	QOPEVGLGIT	PGFGGTQRQLS	R LVGMGMAKQ	LIFTAQNIKA	DEALRIGLVN	180
KVVEPSELMN	TAKEIANKIV	SNAPVAKL S	KQAINRGMQC	DIDTALAFES	EAFCGCFSTE	240
DQKDAMTAFI	EKRKIEGFKN	R				261

SEQ ID NO:	87	moltype = AA	length = 394
FEATURE		Location/Qualifiers	
source		1..394	
		mol_type = protein	
		organism = Cupriavidus necator	

SEQUENCE:	87					
MTRVVVVSG	VRTAIGTFGG	SLKDVA PAEL	GALVVREALA	RAQVSGDDVG	HVVFGNVVIQT	60
EPRDMYLGRV	AAVNGGVtin	APALT VNRLC	GSGLQAI VSA	AQTILLGTD	VAIGGAE SM	120
SRAPY LAPAA	RNGARMGDAG	LVDML GLAH	DPFHRIAHMG V	TAENVAKEYD	ISRAQQDEAA	180
LESHRRASAA	I KAGYFKDQI	VPVVS KGRKG	DVT FDT DEHV	RHDAT IDDM T	KLRPVFVKEN	240
GTVTAGNASG	LNDAAA AVVM	MERAEEARRG	LKPLAR LVS Y	GHAGVDPKAM	GIGPVPA TKI	300
ALERAGLQVS	DLDVIEANEA	FAAACACAVTK	ALGLDPAKVN	PNGSGISLGH	PIGATGALIT	360
VKALHELN RV	QGRYALVTMC	IGGGQGIAAI	FERI			394

SEQ ID NO:	88	moltype = AA	length = 397
FEATURE		Location/Qualifiers	
source		1..397	
		mol_type = protein	
		organism = Treponema denticola	

SEQUENCE:	88					
MIVKPMVRNN	ICLNAHPQGC	KKGVEDQIEY	TKKRITA EVK	AGAKAPKN VL	V LGCSNGYGL	60
ASRITA AFGY	GAATIGVSFE	KAGSETK YGT	PGWYNNL AF D	EEAKRE GLYS	VTIDGDA FSD	120
EIKAQVIEEA	KKKG IKFDLI	VYSLASP VRT	DPDTGIMHKS	VLK PFGKT FT	GKTVD PFT GE	180
LEBISAEPAN	DEEEAAATV KV	MGGEDWERWI	KOL SKEGL LE	EGCITLAYS Y	IGPEATQ ALY	240
RKG TIGAKE	HLEATAH RLN	KENPSKLV TRA	S AVIPV IPLY	LASLF KMKE	300	
KGNHEG CIEQ	I TRLYA ER L	RKDGTIPVDE	ENRIR DDD W	LEEDV QKAVS	ALMEKV TGEN	360
AESLT DLAGY	RHD FLAS NGF	D VEGINY EA E	VER FDRI			397

SEQ ID NO:	89	moltype = AA	length = 720
FEATURE		Location/Qualifiers	
source		1..720	
		mol_type = protein	
		organism = Cannabis sativa	

SEQUENCE:	89					
MGKN YKSLDS	VV ASDFIALG	IT SEVA ET LH	GRLAE IVC NY	GAAT PQTW IN	IANH ILSPDL	60
PFLSLHQMLFY	CGYKDFGPAP	PAWIPDPEK V	KSTNL GALLE	KRGKEFLGV K	YKDP ISS FSH	120
FQEFS VRNPE	VYWR TLMDE	MKISPSK DPE	CILRRDDIN	PGGSEWL PGG	YLNSAK NCLN	180
VNSNKKLNDT	MIVWR DEGND	DLPLNK TL D	QLRK RVWLV G	YALEEMGLEK	GCAIAID MPM	240
HVD AVV IYLA	I VLAG YV VVS	IADSF SAPEI	STR LRL SKAK	AIF TQDH IR	GKKR IPLY SR	300
VVEAKSPMAI	VI PCGS NI G	AEL RDG DI SW	DYFL ERA KEF	KNCEFTARE Q	PV DAY TNIL F	360
SSGTT GEPKA	I PWT QAT PLK	AAAD GWS HLD	IRKG DVIV WP	TNL GWMM GPW	L VYAS LLNG A	420
STAL YNG SPL	VSG FAKE FV QD	AKV TML GV VP	SIV RSW KST IN	CV SGY DW STI	RCF SS GEA S	480
NVDE YL WL MG	R ANY KPV IEM	CGG TEIG GAF	SAGSFL QAO S	LSS FSS QC MG	CT LYILD KNG	540
YMPK PN KPG I	GE LA GLGP VM	G ASK TLL NGN	HHD VY FK GM P	T LN GEV LR RH	G DIFE LTS NG	600
YYHA HG GRAD D	TM NI GG IKIS	SIE IER VCNE	V DDRV FETTA	I GVP PLGG GP	E QLV I FF VL K	660
DSND T TIDL N	QL RL SF NL GL	QK KLN PL FK V	TR VVPL SSS LP	R TAT NK IM RR	V LR QQ F SH F E	720

SEQ ID NO:	90	moltype = AA	length = 491
FEATURE		Location/Qualifiers	

-continued

source 1..491
 mol_type = protein
 organism = *Saccharomyces cerevisiae*

SEQUENCE: 90
 MKLSTKLCWC GIKGRLRPQK QQQLHNTNLQ MTELKKQKTA EQKTRPQNVG IKGIQIYIPT 60
 QCVNQSELEK FDGVSQGKYT IGLGQTNMSF VNDREDIYSM SLTVLSKLK SYNIDTNKIG 120
 RLEVGTETLI DKSKVKSVL MQLFGENTDV EGIDTLNACY GGTNALFNSL NWIESNAWDG 180
 RDAIVVCGDI AIYDKGAARP TGGAGTVAMW IGPDAPIVFD SVRASYMEHA YDFYKPDFTS 240
 EYPYVDGHFS LTCYVKALDQ VYKSYSSKKAI SKGLVSDPAG SDALNVLKYF DYNVFHVPTC 300
 KLVTKSYGRL LYNDFRANPQ LFPEVDAELN TRDYDESLTL KNIKTFVN V AKPFHKERVA 360
 QSLIVPTNTG NMYTASVYAA FASLLNYVGGS DDLGKGRVGL FSYGSGLAAS LYSCKIVGDV 420
 QHIKELDIT NKLAKRITET PKDYEAAIEL RENAHLKNF KPQGSIEHLQ SGVYYLTNID 480
 DKPRRSYDVK K 491

SEQ ID NO: 91 moltype = AA length = 381
 FEATURE Location/Qualifiers
 source 1..381
 mol_type = protein
 organism = *Saccharomyces cerevisiae*

SEQUENCE: 91
 MVAVRRKALS ILAEAPVLAS DRPLPKNYDY DRVFGACCEN VIGYMPPLPVG VIGPLVIDGT 60
 SYHIPMATTE GCLVASAMRG CKAINAGGGA TTVLTKDGMT RGPVVRFPPTL KRSGACKIWL 120
 DSEEGQNAIK KAFNSTSRFA RLQHJQTCLA GDLLFMRFR TTGDAMGMNM ISKGVEYSLK 180
 QMVEEYGWED MEVVSVSGNY CTDKKPAAIN WIEGRGKSVV AEATIPGDVV RKVLKSDVSA 240
 LVELNIAKNL VGSAMAGSVG GFNAHAANLV TAVFLALQGD PAQNVESSNC ITLMKEVDGD 300
 LRISVSMPSI EVGTIGGGTV LEPQGAML DL LGVRGPHTA PGTNARQLAR IVACAVLAGE 360
 LSLCAALAAG HLVQSHMTHN R 381

SEQ ID NO: 92 moltype = AA length = 443
 FEATURE Location/Qualifiers
 source 1..443
 mol_type = protein
 organism = *Saccharomyces cerevisiae*

SEQUENCE: 92
 MSLPFLTSAP GKVIIFGEHS AVYNKPAVAA SVSALRTYLL ISESSAPDTI ELDFPDISFN 60
 HKWSINDFNA TEDQVNSQK LAKAQQATDG LSQELVSLLD PLLAQLSESF HYHAFCFLY 120
 MFVCLCPHAK NIKEFLSKSTL PIGALGSSA SISVSLALAM AYLGGGLIGSN DLEKLSENDK 180
 HIVNQWAFIG EKCIHGTPSG IDNAVATYGN ALLFEKDSHN GTIHDAMNNFK LDDFFPAIPMI 240
 LTYTRIPRST KDLVARVRL VTEKPEVMK PILDAECA LQGLEIMTKL SKCKGTDDEA 300
 VETNNELYEQ LLELIRINHG LLVSIGVSHP GLELIKNLSD DLRIGSTKLT GAGGGGCSLT 360
 LLRDRITQEQ IDSFKKKLQD DFSYTFETD LGGTGCCLLS AKNLNKLKI KSLVQLFEN 420
 KTTTKQQIDD LLLPGNTNLP WTS 443

SEQ ID NO: 93 moltype = AA length = 452
 FEATURE Location/Qualifiers
 source 1..452
 mol_type = protein
 organism = *Saccharomyces cerevisiae*

SEQUENCE: 93
 MSELRAFSAP GKALLAGGYL VLDTKYEAFF VGLSARMHAV AHPYGSLQGS DKFEVRVSK 60
 QPKDGEWLYH ISPKSGFIPV SIGGSKNPFI EKVIANVFSY FKPNMDDYCN RNLFVIDIFS 120
 DDAYHSQEDS VTEHGRNRRL SFHSHRIEEV PKTGLGSSAG GLVTVLTTAL ASFFVSDLEN 180
 NVDKYREVIH NLAQVAHCQA QGKIGSGFDV AAAAYGSIRY RRFPPALISN LPDIGSATYG 240
 SKLHLVDEE DWNITIKNSH LPSGLTLWNG DIKNGSETVK LVQKVKNWYD SHMPESLK 300
 TELDHANSRF MDGLSKLDRL HETHRDYSDQ IFESLERNDC TCQKYPEITE VRDAVATIRR 360
 SFRKITKESG ADIEPPVQTS LLDDCQTLKG VLTCIPGAG GYDAIAVITK QDVLRAQTA 420
 NDKRFSKVQW LDVTQADWGV RKEKDPETYL DK 452

SEQ ID NO: 94 moltype = AA length = 396
 FEATURE Location/Qualifiers
 source 1..396
 mol_type = protein
 organism = *Saccharomyces cerevisiae*

SEQUENCE: 94
 MTVYTASVTA PVNIATLKYW GKRDTKLNLP TNSSISVTLS QDDLRTLTS AATAPEFERDT 60
 LWLNGEPHSI DNERTQNCLR DLRQLRKEME SKDASLPTLS QWKLHIVSEN NFPTAACGLAS 120
 SAAGFAALVS AIAKLYQLPQ STSEISRIAR KGSGSACRSL FGGYVAWEMG KAEDGHDSMA 180
 VQIADSDWP QMKACVLVVS DIKKDSTSQT GMQLTVATSE LFKERIEHVV PKRFEVMRKA 240
 IVEKDFATFA KETMMDNSF HATCLDSFPF IFYMMNDTSKR IIISWCHTINQ FYGETIVAYT 300
 FDAGGPNAVLY YLAENESKLF AFIYKLFGSV PGWDKKFTTE QLEAFNHQFE SSNFTARELD 360
 LELQKDVARV ILTQVGSGPQ ETNESLIDAK TGLPKE 396

SEQ ID NO: 95 moltype = AA length = 288
 FEATURE Location/Qualifiers
 source 1..288
 mol_type = protein
 organism = *Saccharomyces cerevisiae*

SEQUENCE: 95
 MTADNNNSMPH GAVSSYAKLV QNQTPEDILE EFPEIIPLQQ RPNTRSSETS NDESGETCFS 60

-continued

GHDEEQIKLM	NENCIVLDWD	DNAIGAGTKK	VCHLMEENIEK	GLLHRAFSVF	IFNEQGELL	120
QQRATEKIF	PDLWTNTCCS	HPLCIDDELG	LKGKLDDKIK	GAITAAVRKL	DHELGIPED	180
TKTRGKFHFL	NRIHYMAPSN	EPWGEHEIDY	ILFYKINAKE	NLTVPNPVNE	VRDPKWVSPN	240
DLKTMFADPS	YKFTPWFKII	CENYLFNWW	QLDDLSSEVEN	DRQIHRML		288

SEQ ID NO: 96	moltype = AA	length = 335				
FEATURE	Location/Qualifiers					
source	1..335					
	mol_type = protein					
	organism = <i>Saccharomyces cerevisiae</i>					
SEQUENCE: 96						
MEAKIDELIN	NDPVWSSQNE	SLISKPYNH	LLKPGKNFRL	NLIVQINRVM	NLPKDQLAIV	60
SQIVELLHNS	SLIIDDIE	APLRGGQTT	HLIWGVPI	NTANMYMFR	MQLVSQLTT	120
EPLYHWLITI	FNEELINLHR	GQGLDIWRD	FLPEIIPTOE	YLNLMVMNKT	GGLFRLLRL	180
MEALSPSSH	GHSLVFINL	LGIYQIRDD	SEKGFAD	EGKLSFPV	IVH	240
ALNFTKTKGQ	TEQHNEILRI	LLLRTSDKD	KLKLIQILEF	DTNSLAYTKN	FINQLVNMIK	300
NDNENKYL	PD LASHSDTATN	LHDELLYIID	HLSEL			335

SEQ ID NO: 97	moltype = AA	length = 2233						
FEATURE	Location/Qualifiers							
source	1..2233							
	mol_type = protein							
	organism = <i>Saccharomyces cerevisiae</i>							
SEQUENCE: 97								
MSEESLFE	SS POKMYEITH	YSERHTELPG	HFIGLNTVDK	LEESPRLDFV	KSHGGHTVIS	60		
KILIANNGIA	AVKEVRSVRK	WAYETFGDDR	TQVFVAMATP	EDLEANA	EYI RMADQYIEVP	120		
GGTNNNNNYAN	VDLIVDIAEE	ADVDAWAGW	GHASENPLLP	EKLSQS	KRKV IFIGPPGNAM	180		
RSLGDKISST	IVAQSAKVPC	IPWSGTGVT	VHVD EKTGLV	SVDDDIYQKG	CCTSPEDGLQ	240		
KAKRIGFPV	IKASEGGGG	GIROVEREED	FIALYHQAN	EIPGSP	FIM KLAGRARHLE	300		
VQLLAQYGT	NISLPGRCDS	VQRRHQKII	EAPVTIAKAE	TFHEMEKA	AV RLGKLVGYVS	360		
AGTVEYL	YSH DDGKFYF	LEPL	TEMVSGVNL	AAQLQIAMGI	PMHRSIDIRT	420		
LYGMNPHAS	EIDFEFKTQD	ATKKQRRPIP	KGHCTACRIT	SEDPNDGFKP	SGGTLHELN	480		
RSSSNVWGYF	SVGNNNGNIHS	FSDSQFGHIF	AFGENRQASR	KHMVVALKE	SIRGDPRTTV	540		
EYLIKLLTE	DFEDNTITTG	WLDDLT	TAEKPDPTLA	VICGAATKAF	LASEEARHKY	600		
IESTLQKGOV	SKDLQTMF	VDFIHEGKEY	KFTVAKSND	RTYTLFINGSK	CDIILRQLSD	660		
GGLLIAIGGK	SHTIYKKEEV	AATRLSVDSM	TTLLEVENDP	TQLRTPSPGK	LVKFLVENE	720		
HIIKGQPYAE	IEVMKMQMPL	VSQENGIVQL	LKQPGSTIV	GDIMA	IMTLD DPSKVHALP	780		
FEGLMPDFGS	PVIEGTKPAY	KFKVLVSTLE	NILKGYDNQV	IMNASLQQLI	EVLRNPKLPY	840		
SEWKLHISAL	HSRLPAKLD	QMEELVARS	RRGAVFPARQ	LSKLIDMAVK	NPEYNPDKLL	900		
GAVVEPLADI	AHKYSNGLEA	HEHSIFVHPL	EYYEVEKLF	NGPNVREENI	ILKLRDENPK	960		
DLDKVALT	SHSKVSAKNN	LILAIKHYQ	PLCCLKSSKVS	AIFSTPLQHI	VELESKATAK	1020		
VALQARELLI	QGALPSVKER	TEQIEHILK	SVVKVAYGNS	NPKRSEPDLN	ILKLDLDSNY	1080		
VVFVDVLLQFL	THQDPVUTAA	AAQVYVIRR	RAYTIGDIRV	HEGVTPVIVE	WKFQLPSAAF	1140		
STFPTVSKM	GMNRAVSVD	LSYVANSQSS	PLREGILMAV	DHLDVDEIL	SQSLEVI	PRH	1200	
QSSSNGPAPD	RSGSSASLSN	VANCVASTE	GFESEEEILV	RLREILDLNK	QELINASIRR	1260		
ITFMFGPKDG	SYPKYTF	PNYNNENETR	HIEPALAFQ	ELGRLSNFNI	KPIFTDNRNT	1320		
HVYEA	VSKTS	QALPSVKER	TEQIEHILK	SVVKVAYGNS	NPKRSEPDLN	1380		
DLNHIFINFI	AVFDISPEDV	EAAFGGFLER	FGKRLRLR	SSAEIRIIK	DPQTGAPVPL	1440		
RALINNVSGY	VIKTEMYTEV	KNAKGEWVFK	SLKGPGSMHL	RPIATPYPVK	EWLQPKRYKA	1500		
HLMGTTVYD	FPELFQFQASS	SQWNKFSADV	KLTDDFFISN	ELIEDENGEL	TEVEREPGAN	1560		
AIGMVAPK	ITVTPPEYPRGE	QFVVVANDIT	FKIGSFGPOE	DEFENKVTEY	ARKRGIPRIY	1620		
LAANSGARIG	MAEEIVPLFQ	VAWNDAA	PLNPD	KGPQYLYLTS	EGMETLKKFD	KENSVLTER	1680	
VINGEERFVI	KTIIGSEDGL	GVECLRGSGL	IAGATSRAYH	DIIFTITLVTC	RSVGIVGAYLV	1740		
RIGQRAIQVE	QOPII	LTGAP	AINKMLGREV	YTSNLQLGGT	QIMYNNNGVSH	LTAVIDLAGV	1800	
EKIVEWMSYV	PAKRN	LPK	LETKD	WTWDRP	VDFTPTND	YDVRWMIEGR	1860	
FDKGSFFETL	SGWAKGVVVG	RARLG	GIPLIG	VIGVETR	TRVNE	NLIPADPANP	NSAFTLIQEP	1920
GQVWHPSAF	KTAQAI	NDFN	NGEQLPMMIL	ANWRGFSGGQ	RDFMFNEV	LKY GSFIVDALVD	1980	
YKQPII	YIP	PTGELRGGSW	VVVDPTINAD	QMBEMYAD	VNA RAGVLE	FGQM VGIKEPREKL	2040	
LDTMNRLDDK	YRELRSQLSN	KSLAPEVHQ	ISKQLADRER	ELLPIYGQIS	LQFADLHDRS	2100		
SRMVAKGVIS	KELETEAR	FFF	WRLRRL	NEEYKLR	LS HQVGEASRL	KIARIRSWYP	2160	
ASVDHEDDRQ	VATWIEENYK	TL	DDKLKG	GLK LESFAQDL	KIRSDHDNAI	DGLSEVI	KML	2220
STDDKEKL	LK TLK						2233	

SEQ ID NO: 98	moltype = AA	length = 385								
FEATURE	Location/Qualifiers									
source	1..385									
	mol_type = protein									
	organism = <i>Cannabis sativa</i>									
SEQUENCE: 98										
MNHLRAEGPA	SVLAIGTANP	ENILLQDEFP	DYYFRVTKS	HMTQLKEKFR	KICDKSMIR	60				
RNCFLNEEHL	KQNPRV	LEHE	MQTL	DARQDM	LVEV	VPKL	GK	120		
IPTSA	TTDM	PGADYHCAKL	LGLSPSVKRV	MMYOLCGYGG	GTVLRIAKDI	AENNKGARVL	180			
AVCCDIMACL	FRGPSE	SDLE	LLVGQAIFGD	GAAAVIVGAE	PDES	VGERPI	FEVSTGQ	240		
LPNSEGT	IGG	HIREAGLIFD	LHKDV	PM	LIS	NNIEKCLIEA	FTPIG	ISDW	N	300
KAILDKVEEK	LHLKSDKFVD	SRHVLSEHGN	MSSSTVLFV	DLRKRSLEE	GKSTT	GDG	GFE	360		
WGVLFGFGPG	LTVERVV	VRS	VPIKY					385		

SEQ ID NO: 99	moltype = AA	length = 101
FEATURE	Location/Qualifiers	

-continued

source 1..101
 mol_type = protein
 organism = Cannabis sativa

SEQUENCE: 99
 MAVKHLIVLK FKDEITEAQK EEFFKTYVNL VNIIPAMKDV YWGKDVTQKN KEEGYTHIVE 60
 VTFESVETIQ DYIIHPAHVG FGDVYRSFWE KLLIFDYTPR K 101

SEQ ID NO: 100 moltype = AA length = 395
 FEATURE Location/Qualifiers
 source 1..395
 mol_type = protein
 organism = Cannabis sativa

SEQUENCE: 100
 MGLOSSVCTFS FQTNYHTLLN PHNNNPKTSI LCYRHPKTP I KYSYNNFP SK HCSTKSFHQ 60
 NKCSSELSIA KNSIRAAATTN QTEPPESDNH SVATKILNFG KACWKLQR PY TIIAFTSCAC 120
 GLFGKELLHN TNLISWSLMF KAFFFVLAIL CIASFTTIN QIYDLHIDRI NKPDLPLASG 180
 EISVNTAWIM SIIVALFGLI ITIKMKGGL YIFGYCFCGIF GGIVYSVPPR RWKQNPSTAF 240
 LLNLFHLAII SIFTFYASRA ALGLPFLERP SFTFLLA FMK SMGSALALIK DASDVEGDTK 300
 FGISTLASKY GSRNLTLFCG GIVLSSYVA ILAGIIWPOA FNSNVMLSH AILAFWLILQ 360
 TRDFALTNYD PEAGRPFYEF MWKLYYYAEYL VYVF 395

SEQ ID NO: 101 moltype = AA length = 544
 FEATURE Location/Qualifiers
 source 1..544
 mol_type = protein
 organism = Cannabis sativa

SEQUENCE: 101
 MKCSTFSFWF VCKIIFFFFS FNIQTSIANP RENFLKCFSQ YIPNNNATNLK LVYTQNNPLY 60
 MSVLNSTIHN LRFTSDTTPK PLVIVTPSHV SHIQGTILCS KKVGQLQIRTR SGGHDSEGMS 120
 YISQVPFVIV DLRNMRSIKI DVHSQTAWVE AGATLGEVYY WVNEKNENLS LAAGYCPTVC 180
 AGGHFGGGGY CPLMRNYGLA ADNIIIDAHLV NVHGKVLDRK SMGEDLFWAL RGGAESFGI 240
 IVAWKIRLVA VPKSTMFSVK KIMEIHELVK LVNWKWQNIAY KYDKDLLMT HFITRNITDN 300
 QGKNTKTAIHT YFSSVFLGGV DSLVLDLMNK SFPELGKKT CRQLSWIDTI IFYSGVVNYD 360
 TDNFNKEILL DRSAQGQNAF KIKLVDVKKP IPESVVFQIL EKLYEEDI GA MYALYPYGG 420
 IMDEISESAI PPFFHRAGILY ELWYICSWEK QEDNEKHLNW IRNIYNFMPY YVSKNPRLAY 480
 LNYRDLDIGI NDPKPNPNYT QARIWGEKY GKNFDRLVK KTLVDPNNFF RNEQSIPPLP 540
 RRRH 544

SEQ ID NO: 102 moltype = AA length = 545
 FEATURE Location/Qualifiers
 source 1..545
 mol_type = protein
 organism = Cannabis sativa

SEQUENCE: 102
 MNCASTFSFWF VCKIIFFFLS FNIQTSIANP QENFLKCFS E YIPNNNAPNPK FIYTOHDQLY 60
 MSVLNSTIHN LRFTSDTTPK PLVIVTPSNV SHIQQASILCS KKVGQLQIRTR SGGHDAEGLS 120
 YISQVPFAIV DLRNMRHTVKV DIHSQTAWVE AGATLGEVYY WINEMNENFS FPGGYCPTVG 180
 VGGHFSGGGY GALMRNYGLA ADNIIIDAHLV NVDGKVLDRK SMGEDLFWAI RGGENENFGI 240
 IAACKIKLVV VPSKATIFSV KKNMEIHGLV KLFNKWQNIAY KYDKDLMLT THFRTRNITD 300
 NHGKNKTTVH GFYSSVFLGGV DSLVLDLMNK SFPELGKKT DCKELSWIDT TIFYSGVVNY 360
 NTANFKKEIL LDRSAGKTKA FSIKLDYVKK LIPE TAMVKI LEKLYEEEVG VGMYVLYPYG 420
 GIIMDEISESAI IPFPFHRRAGIM YELWYTTATWE KQBDNEKHIN WVRSVNFTT PYVSONPRLA 480
 YLNRYRDLIGL KTNPESPNNY TQARIWGEKY FGKNFNRLVK VTKKADPNNF FRNEQSIPPL 540
 PPRHH 545

SEQ ID NO: 103 moltype = AA length = 1529
 FEATURE Location/Qualifiers
 source 1..1529
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 103
 MSAKAISEQT GKELLYKFIC TTSAIQNRPK YARVTPDTW ARLLQDHFWL LSQNLVVKPD 60
 QLIKRRGKLG LVGVNLTL DG VKSQQNPKRLG QEATVGKATG FLKNFLIEPF VPHSQAEEFY 120
 VCIVYATREGD YVLFVHHEGGV DVGDVDAKQ KLLVGVDEKL NPEDIKKHLL VHAPEDKKEI 180
 LASFISGLFN FYEDLYFTYL EINPLVVTKD GVYVLDLA AK VDATADYICK VKWGDIEFPP 240
 PGREGREYEE AYIADLDAKS GASLKLTLIN PKGRIWTMVA GGGASVYSD TICDLGGVNE 300
 LANYGESGA PSEQQTYDYA KTILSLMTRE KHPDGKILII GGSIANFTNV AATFKGIVRA 360
 IRDYQGPLKE HEVTIFVRG GP NYQEGLKR MGEVGKTTGI PIHVFGETH MTAIVGMALG 420
 HRIPPNQPTP AAHTANFLN AASGTSTPAP SRTASFESR ADEVA PAKKA KPAMPQDSVP 480
 SPRSLQGKST TLFSRHTKAI VWGMQTRAVQ GMLDFDYVC S RDEPSVAAMV YPFTGDHKQK 540
 FWYGHKEILI PVFKNMADAM RKHPEVDVLI NFASLRSAYD STMETMNYAQ IRTIAIIAEG 600
 IPEALTRKLI KKADQKGVTI IGPATVGGIK PGCFKIGNTG GMLDNILASK LYRPGSVAYV 660
 SRSGGMSNEL NNIISRTTDG VYEGVAIGGD RYFPGSTFMHD VLRYQDTPGV KMIVVLGEIG 720
 GTTEYKICRG IKGRLTKP VWCWICGTCAT MFSSEVQPGH AGACANQASE TAVAKNQALK 780
 EAGVFVPRSF DELGEIIQSV YEDLVANGVI VPAQEVPPTP VPMDYSHARE LGLIRKPASF 840
 MTSICDERGQ ELIYAGMPIT EVFKEMGIG GVLGLLWFQK RLPKYSQOFI EMCLMVTADH 900
 GPAVSGAHNT IICARAGKDL VSSLTSGLLT IGDRFGGALD AA AKMFSKAF DSGIIPMEFV 960
 NKMKEKGKLI MGIGHRVKSI NNPMDRMVQIL KDYVRQHFPA TPLL DYALEV EKITTSKPN 1020
 LILNVDGLIG VAFVDMRLNC GSFTREEADE YIDIGALNGI FVLGRSMGFI GHYLDQKRLK 1080

-continued

QGLYRHPWDD ISYVLPEHMS MKLSGGGGSG GGGSGGGGSA EAWYNLGNAY YKQGDYQKAI 1140
 EYYQKALELD PNNAEAWYNL GNAYYKQGDY QKAIEYYQKA LELEDPNNAEA WYNLGNAYYK 1200
 QGDYQKAIED YQKALELDPN NLQAEAWKNL GNAYYKQGDY QKAIEYYQKA LELEDPNNASA 1260
 WYNLGNAYYK QGDYQKAIEY YQKALELDPN NAKAWYRRGN AYYKQGDYQK AIEDYQKALE 1320
 LDPNNRSRSA CGGGSGGGGS GGGGASSYYH HHHHHLESTS LYKKAGSGSN LVAOLENEVA 1380
 SLENENETLK KKNLHKKDLI AYLEKEIANL RKKIEEGSAG SAAGSGEFGS AEAAAKEAAA 1440
 KAGSAGSAAAG SGEFGSSYYH HHHHHLESTS LYKKAGSGSA RNAYLRKKIA RLKKDNLQLE 1500
 RDEQNLEKII ANLRDEIARL ENEVASHEQ 1529

SEQ ID NO: 104 moltype = AA length = 821
 FEATURE Location/Qualifiers
 source 1..821
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 104
 MKNCVIVSAV RTAIGSFNGS LASTSAIDLG ATVIKAAIER AKIDSQHVDE VIMGNVLQAG 60
 LGONPARQAL LKSGLAETVC GFTVNLKVCGS GLKSVALAAQ AIQAGQAQSI VAGGMENMSL 120
 APYLLDDAKAR SGYRLGDGVQ YDVILRKDM CATHGYHMGY TAENVAKEYG ITREMQDELA 180
 LHSQRKAAAA IESGAFTAEI VPVNVTTRK TFVFSQDEFPP KANSTAEALG ALRPAPFDKAG 240
 TVTAGNASGI NDGAAALVIM EESAALAAGL TPLARIKSYA SGGVPPALMG MGPPATQKA 300
 LQLAGLQLAD IDLIEANEAF AAQFLAVGKLN LGFDSEKVNNG NGGAIALGHP IGASGARILV 360
 TLLHAMQARD KTLGLATLCI GGGQGIAMVI ERLNKLSGGG GSAAEWYNLG 420
 NAYYKQGDYQ KAIYEYQKAL ELDPNNAEAW YNLGNAYYKQ GDYQKAIEYY QKALELDPNN 480
 AEAWYNLGNA YYKQGDYQKA IEDYQKALEL DPNNLQAEAW KNLGNAKYKQ GDYQKAIEYY 540
 QKALELDPNN ASAAYNLGNA YYKQGDYQKA IEYYQKALEL DPNNNAKAWYR RGNAYYKQGD 600
 YQKAIEDYQK ALELDPNNRS RSAGGGGGASS YYHHHHHLE STSLYKKAGS 660
 GSNEVTTLEN DAAFIENENA YLEKEIARLR KEKAALRNRL AHKKGSAGSA AGSGEFGSAB 720
 AAAKEAAAKA GSAGSAAGSG EFGSSYYHHH HHHLLESTSLY KKAGSGSQKV AELKNRAVAK 780
 LNRNEQLKNK VEELKNRNAY LKNELATLEN EVARLENDVA E 821

SEQ ID NO: 105 moltype = AA length = 776
 FEATURE Location/Qualifiers
 source 1..776
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 105
 MKKVCVIGAG TMGSGIAQAF AAKGFEEVVL RIKDEFVDRG LDFINKNLSK LVKKGKIEEA 60
 TKVEILTRIS GTVDLNMAAD CDLVIAAVE RMDIKKQIFA DLDNICKPET ILASNTSSL 120
 ITEVASATKR PDKVIGHMHFF NPAPVPMKLV IRVGIATSQE TFDAVKBTSSI AIGKDPVEVA 180
 EAPGFVNVRN LIPMINEAVG ILAEGIASVE DIDKAMKLGA NHPMGPLELG DFIGLDICLA 240
 IMDVLYSETG DSKYRPHTLK KKYVRAGLWL RKGSKGFYDY SKKLSSGGGS GGGGSGGGGS 300
 AEAWYNLGNA YYKQGDYQKA IEYYQKALEL DPNNNAEAWYN LGNAYYKQGD YQKAIEYYQK 360
 ALELDPNNAE AWYNLGNAYY KQGDYQKAIE DYQKALELDP NNQAEAWKN LGNAYYKQGD 420
 YQKAIEYQK ALELDPNNAS AWYNLGNAYY KQGDYQKAIE YYQKALELDP NNNAKAWYRRG 480
 NAYYKQGDYQ KAIEDYQKAL ELDPNNRNSRS AGGGGSGGGG SGGGGASENL YFQGENLYFQ 540
 GDSSESCWCN GRKASETCG CNTARYCASF COHQKDWEKH HICGQTLQAOQ QGSAGSAAGS 600
 GEFGSAEAAA KEAAAKAGSA GSAAGSGEFG SMAVSESOLK KMVSKYKYRD LTVRETVINI 660
 TLYKDLKPVL DSYVFNDGSS RELMLNTGTI PVPYRGNTYN IPICLWLLDT YPYNPICFV 720
 KPTSSMTIKT GKHDANGKI YLPYLHEWKH PQSDLGLI V MIIVVFGDEP PVFSRP 776

SEQ ID NO: 106 moltype = AA length = 741
 FEATURE Location/Qualifiers
 source 1..741
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 106
 MELNNVILEK EGKVAVVTIN RPKALNALNS DTLKEMDYVI GEIENDSEVL AVILTGAGEK 60
 SFVAGADISE MKEMNTIEGR KFGFLGNKVF RRLELLEKPV IA AVNGFALG GGCEIAMSCD 120
 IRIASSNARF QGPVEGLGIT PGFGGTQRLL RLVMGMQAKQ LIFTAQNIKA DEALRIGLVN 180
 KVVEPSELNMN TAKEIANKIV SNAPAVAKLNS KQAINRGMOC DIDTALAFES EAFGBCFSTE 240
 DQKDAMTAIFI EKRKIEGFKN RKLSSGGGGSG GGGGGGGGS EA WYNLGNAY YKQGDYQKAI 300
 EYYQKALELD PNNAEAWYNL GNAYYKQGDY QKAIEYYQKA LELEDPNNAEA WYNLGNAYYK 360
 QGDYQKAIED YQKALELDPN NLQAEAWKNL GNAYYKQGDY QKAIEYYQKA LELEDPNNASA 420
 WYNLGNAYYK QGDYQKAIE YQKALELDPN NAKAWYRRGN AYYKQGDYQK AIEDYQKALE 480
 LDPNNRSRSA GGGGSGGGGS GGGGASGPLG SPLTASMLAS APPQEOKQML GERLFPLIQA 540
 MHPTLAGKIT CMLLEIDNSE LLHMLLESPE LRSKVDDEAVA VLQAHQAKEA AQKAGSAGSA 600
 AGSGEFGSAA AAAKEAAAKA GSAGSAAGSG EFGSNTNMVS PTDGAVTTSQ IPASEQETLV 660
 RPKPLLLKLL KSVGAKDTY TMKEVLFYLG QYIMTKRLYD EKQOHIVYCS NDLLGDLFGV 720
 PSFSVKEHRK IYTMIRRNLV V 741

SEQ ID NO: 107 moltype = AA length = 820
 FEATURE Location/Qualifiers
 source 1..820
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 107
 MIVKPMVRNN ICLNAHPQGC KKGVEDQIEY TKKRITAEVK AGAKAPKNVL VLGCNSNYGL 60
 ASRITAAGFY GAATIGVSFE KAGSETKYGT PGWYNNLAFD EAAKREGLYS VTIDGDAFSD 120
 EIIKAQVIEEA KKKGKIKFDLI VYSLASPVRT DPDTGIMHKSL KGTVDPTGE 180

-continued

LKEISAE PAN DEEEAATVKV MGGEDWERWI KOLSK EGLLE EGCITL AY SY IGPEATQAL Y 240
 RKG TIGKAKE HLEATAHRLN KENPSIRAFV SVNKG L VTRA SAVIP VPL Y LASLFKVMKE 300
 KGNHEG CIEQ ITRLYAERLY RKDG TIPVDE ENRIRID DW LEEDVQKA V ALMEKV TGEN 360
 AESL TDLAGY RHDFLAS NGF DVEGIN YEA E VERFDRIKLS GGGGSGGGGS GGGGSAEAWY 420
 NLGNAYYKQG DY QKAI EYYQ KALEDPN NA EA WYNLGNAY YKQGDYQKAI EYYQKALELD 480
 PNNAEAWYNL GNAYYKQGDY QKAI EYYQ KALEDPN NA EA WYNLGNAY YKQGDYQKAI 540
 EYYQKALELD PNNAEAWYNL GNAYYKQGDY QKAI EYYQ KALEDPN NA EA WYNLGNAYK 600
 QGDYQKAI ED YQKALELD PN NRSRSAGGGG SGGGSGGGG ASSYYHHHH HLESTSLYKK 660
 AGSGSNLLAT LRSTA AVLEN ENHVLEKEKE KLRKEKEQ L NKLEAYKGS A GSAAGS GEFG 720
 SAEAAAKAEEA AKAGSAGS A GSGEFGSSYY HHHHHHLEST SLYKKAGSGS KRIAYLK KI 780
 AALKD NNL EKDIAN LENE IERLIKEIKT LENEVASHEQ 820

SEQ ID NO: 108 moltype = AA length = 826
 FEATURE Location/Qualifiers
 source 1..826
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 108
 MTREVVVVSG VRTAIGTFGG SLKDVA PAEL GALVVREALA RAQVSGDDVG HVVFGNVIQT 60
 EPRDMYLGRV AAVNGGTIN APALT VNRLC GSGLQAIVSA AQTI LLGDTD VAIGGGAESM 120
 SRAPY LAPA P RWGARM GDAG LV DMDL GLAH DP FRIH MG V TAENVA KEYD ISRAQ D EAA 180
 LESHRRASAA I KAGYF K DQI VPVMSKGRKG DVT FDT D EH V RDAT I DDMT KLRPV FVKEN 240
 GTVTAGNASG LNDAAA VVM MERA EAERRG LKP LARL VSY CHAGVDPKAM GIGPVPA TKI 300
 ALERAGLQVS DLDVIEANEA FAAQACAVTK ALGLDPAKVN PNGSGISLGH PIGATGALIT 360
 VIKALHELN RV QGRYALV TMC IGGGGGIAAI FERIKL SGGG GS GGGGSGGG GS AEWY NLG 420
 NAYYKQGDYQ KAI EYYQ KAL ELDPN NA EA EW YNLGNAYYQ KGDYQKAI EYYQ KALELD PNN 480
 AEAWY NLGNA YYKQGDYQKA IEDYQKALEL DP NNLQAEAB KNLGNAYYQ KGDYQKAI EYY 540
 QKALELD PNN ASAWY NLGNA YYKQGDYQKA IEYYQKALEL DP NNKA WYR RG NAYYQGD 600
 YQKAIEDYQK ALELD PNN RS AAGGGGSGG GGGGGGASD VMWEYK WENT GDAELYGPFT 660
 SAQM QTWSE GYFPDGVYCR KLDPGQGQFY NSKRIDPDLY TG SAGSAAGS GEF GSAEE AAA 720
 KEAAAKAGSA GSAAGS GEFG SEDS VEFN FN AISYV NK IKT RFLDH PEIYR SFLEIL HTYQ 780
 KEQLHTKGRP PRGMSEEEVF TEVANLFRQG EDL LSEFGQF LPEAKR 826

SEQ ID NO: 109 moltype = AA length = 849
 FEATURE Location/Qualifiers
 source 1..849
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 109
 MKLSTKLCWC GIKGRLRPQK QQQLHNTNLQ MTELKKQKTA EQKTRPQNVG IKGIQIYIPT 60
 QCVNQSE EK FDGV SQGK YT IGLG QT TNMSF VNDREDIYSM SLTV LSKLIK SYNIDTNKIG 120
 RLEVGTETL DKS KVS KVS MQL QGENTD V EGID TLNACY GG TNL AFNSL NWIESNAWDG 180
 RDAIVVCGDI AIYDKAARP LTCAGT VAMM IGP DAPIVFD SVRAS YMEHA YDFYKPDFTS 240
 EYPYV DGHFS LTCYV KALDQ VYKSY SKKAI SKGLVSDPAG SDALN VLKYF DYNVHFVPTC 300
 KLVTKS YGR L YNDFRANPQ LFPEV DAE LA TRDYDES LTD KNIEKTFV NV AKPFHK ERVA 360
 QSLIVP NTG NM YTAS VY FA SFLS VY VGS DDLQG KRVGL FSYG SGLA AS LYSCKIVGDV 420
 QHI KKE LDIT KNLAKRIT ET PKDYEAAI E RENAH LKKNF K P QG SIEHL Q SGVY YLTNID 480
 DKFRRS YDVK KKL SGGGGSG GGGSGGGGS EAWY NLGNAY YKQGDYQKAI EYYQKALELD 540
 PNNAEAWYNL GNAYYKQGDY QKAI EYYQKA LEELDPN NA EA WYNLGNAYK QGDYQKAI ED 600
 YQKALELD PNN NLQAEAWK NL GNAYYKQGDY QKAI EYYQKA LEELDPN NA EA WYNLGNAYK 660
 QGDYQKAI EY YQKALELD PNN NAKAWY RRGN AYYKQGDYQK AI EYD YQKALE LDPN NR RSA 720
 GGGGSGGGGS GGGGAS LGPL PPGW EVRSTV SGRYFVDHN NRTT QFT DPR LHGSAGSAAG 780
 SGEFGSAEEA AKEAAKAGS AGSAGS GEFG GSGAMGPLPP GWEKRTDSNG RVYFVNHNTR 840
 ITQWEDPRS 849

SEQ ID NO: 110 moltype = AA length = 826
 FEATURE Location/Qualifiers
 source 1..826
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 110
 MVA VRRK ALS ILAEAPV LAS DRLPYK NYD Y DRV FGACC EN VIGYML PVG VIGPLV IDGT 60
 SYHIP MATTE GCLV ASAM RG CKAINAGGG A TT VLT KDG MTT RGP VVR FPTL KRSGACK IWL 120
 DSEEG QNAIK KAFN STS RFA RLQH QT CLA GDL LFMR FR TT GDAM GM MN ISKG VEY SLK 180
 QMVEEY GWED MEV VSVSG NY CTDK KP AAIN WIE GRG KSVV AEAT IPGD VV RKVL KSDV SA 240
 LVEL NIKA NL VGS MAGS VG GFNA HAA NL TAVFLA LQD PAQN VES SNC ITLM KEV DGD 300
 L RIS VSMPSI E VGT IGG GTV LEPQGAM L LGVR GP HATA PG TNAR QLAR IVAC AVL AGE 360
 LSLCA ALAAG HLV QSHM THN RL KLSGGGGSG GGGSGGGGS EAWY NLGNAY YKQGDYQKAI 420
 EYYQKALELD PNNAEAWYNL GNAYYKQGDY QKAI EYYQKA LEELDPN NA EA WYNLGNAYK 480
 QGDYQKAI ED YQKALELD PNN NLQAEAWK NL GNAYYKQGDY QKAI EYYQKA LEELDPN NA EA 540
 WYNLGNAYK QGDYQKAI EY YQKALELD PNN NAKAWY RRGN AYYKQGDYQK AI EYD YQKALE 600
 LDPN NR RSA GGGGSGGGGS GGGGASSYYH HHHHHLESTS LYKKAGSEFF RRERNKMAA 660
 KCRN RR RELT DTLQ AETD QL EDEK SAL QTE IAN LLKEKEK LEFILA AH RP ACKI PDDL GF 720
 PEEMSLEG SA GSAAGS GEFG SAEAAK AEEA AKAGSAGS A GSGEFGSSYY HHHHHLEST 780
 SL YKKAGSGS QK VESL KQK I EEL QRK AQL KNDIAN LEKE IAYA ET 826

SEQ ID NO: 111 moltype = AA length = 1046
 FEATURE Location/Qualifiers
 source 1..1046

-continued

	mol_type = protein organism = synthetic construct
SEQUENCE: 111	
MSLPFLTSAP GKVIIFGEHS AVYNKPAVAA SVSALRTYLL ISESSAPDTI ELDFPDISFN	60
HKWSINDFNA ITEDQVNSLQD LAKAQATDG LSQELVSLID PLLAQLSESF HYHAAFCFLY	120
MFVCLCPHAK NIKFSLKSTL PIGAGLGSSA SISVSLALAM ATGGGLIGSN DLEKLSENDK	180
HIVNQWAFIG EKCIHGTPSG IDNAVATYGN ALLFEKDSHN GTINTNNFKF LDDFFAIPMI	240
LTYTRIPRST KDLVARVRVL VTEKFPEVMK PILDAVGECMA LQGLEIMTKL SKCKGTDDEA	300
VETNNELYEQ LLELIRINHQ LLVSGIVSHP GLELIKNLSD DLRIGSTKL GAGGGGCSLT	360
LLRRDITQEQ IDSFKKKLQD DFSYETFETD LGGTGCGCLLS AKNLNDKLI KSLVFQLFEN	420
KTTTKQQQIDD LLLPGNTNLNP WTSKLSSGGG SGGGGSGGGG SAEAWNLGH AYYKQGDYQK	480
AIEYYQKALE LDPPNNAEAWY NLGNAYYKQG DYQKAIEYYQ KALELDPNNA EAWYNLGNAY	540
YKQGDYQKAI EDYQKALELP PNNAKAWEWK NLGNAYYKQG DYQKAIEYYQ KALELDPNNA	600
SAWYNLGNAY YKQGDYQKAI EYYQKALELP PNNAKAWEWK GNAYYKQGDY QKAIEDYQKA	660
LELDPPNRSR SAGGGGSGGG GSAGGGGASME PAMEPETLEA RINRATNPMLN KELDWASING	720
FCEQLNEDFE GPPLATRLLA HKIQSPQEW E AIQALTIVLET CMKSCGKRHF DEVGKFRFLN	780
ELIKVSPKY LGSRITSEKVI NKILELLYSW TVGLPVEEVKI AEAYQMLKKQ GIVKSGSAGS	840
AAGSGEFGSAA EAAKEAAAK AGSAGSAAGS GEFGSGAMGS MAAEAEGESLE SWLNKATNPS	900
NRQEDWBIIY GFCDQJINKEI EGPOIAVRLI AHKIQSPQEW E ALQALTVLE ACMKNCGRRF	960
HNEVGKFRFL NELIKVVSPK YLGDRVSEKV KTKVIELLYS WTMALPEEAK IKDAYHMLKR	1020
QGIVQSDPPI PVDRTLIPSP PPRPKN	1046
SEQ ID NO: 112	moltype = AA length = 868
FEATURE	Location/Qualifiers
source	1..868
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 112	
MSELRAFSAP GKALLAGGYL VLDTKYEAFF VGLSARMHAV AHPYGSLQGS DKFEVRVKS	60
QFKDGEWLYH ISPKSGFIPV SIGGSKNPFI EKVIANVFSY FKPNMDDYCN RNLFVIDIFS	120
DDAYHSQEDS VTEHGRNRL SFHSHRIEEV PKTGLGSSAG GLVTVLTTAL ASFFFVSDLEN	180
NVDKYREVIH NLAQVAHCQA QGKIGSGFDV AAAAYGSIRY RRFPPALISN LPDIGSATYG	240
SKLAHLVDEE DWNITKSNL PLSGLTLWNG DIKNGSETVK LVQVKVNWYD SHMPESLKIY	300
TELDHANSRTE MDGLSKLDRL HETHRDDSQ IFESLERND TCQKYPEITE VRDAVATIRR	360
SFRKITKESG ADIEPPVQTS LLDCQTLKG VLTCILPAGG GYDAIAVITK QDVDLRAQTA	420
NDKRFSKVQW LDVTQADWGV RKEKDPTYL DKKLSSGGGS GGGGSGGGGS AEAWYNLGN	480
YYKQGDYQKA EYYQKALEL DPNNAAEAWY LGNAYYKQGD YQKAIEYYQK ALELDPNNAE	540
AWYNLGNAYY KQGDYQKAI DYQKALELDP NNLOQAEAWK LGNAYYKQGD YQKAIEYYQK	600
ALELDPNNAS AWYNLGNAYY KQGDYQKAIE YYQKALELDP NNAKAWYRNG NAYYKQGDYQ	660
KAIEDYQKAL ELDPPNRSRS AGGGGSGGGG SGGGGASSYY HHHHHHHLEST SLYKKAGSGS	720
QKVEELKNKI AELENRNAVK KNRVAHLKQE IAYLKDELAA HEFEGSAGSA AGSGEFGSAB	780
AAAKEAAAKA GSAGSAAGSG EFGSSYYHHH HHHLESTSLY KKAGSGSFEN VTHEFILATL	840
ENENAKLRLR EAKLERELAR LRNEVAWL	868
SEQ ID NO: 113	moltype = AA length = 1008
FEATURE	Location/Qualifiers
source	1..1008
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 113	
MTVYTASVTA PVNIATLKYW GKRDTKLNLP TNSSISVTLQ QDQLRTLTSQ ATAPEFERDT	60
LWLNGEPHSI DNERTQNCLP DLRLQRKEME SKDASLPTLS QWKLHVSENFPTAAGLAS	120
SAAGFAALVS AIAKLYQLPQ STSEISRIAR KGSGSACRSL FGGYVAWEMG KAEGDGHDSMA	180
VQIADFSWDP QMKACVLVVS DIKKDVSSTQ GMQLTVATSE LFKERIEHVV PKRFEVMRKA	240
IVEKDFATFA KETMMDTSNF HATCLDSDP IFYMNNTSKQ IISWCHTINQ FYGETIVAYT	300
FDAGPNVLY YLAENESKLF AFYIKLFGSV PGWDKKFTTE QLEAFNHQFE SSNFTAREL	360
LELQKDVARL ILTQVGSGPQ ETNESLIDAK TGLPKEKLSL GGGGGGGGG GGGSAEAWY	420
LGNAYYKQGD YQKAIEYYQK ALELDPNNAE AWYNLGNAYY KQGDYQKAIE YYQKALELDP	480
NNAAEAWYNAE NAYYKQGDYQ KAIEDYQKAL ELDPPNQLQAE AWKNLGNAYY KQGDYQKAIE	540
YYQKALELDP NNASAWYNLG NAYYKQGDYQ KAIEDYQKAL ELDPPNNAKAW YRRGNAYYKQ	600
GDYQKAIEDY QKALELDPNN RSRSAGGGGS GGGGGGGGG SAMADLEQKV LEMEASTYDG	660
VFIWKISDFP RKRQEAVAGR IPAPIFSPAFY TSRYGYKMCI RIYLNGDGTG RGTHLSLFFF	720
VMKGPNDALL RWPFNQKVTL MLLDQNNRHEV VIDAFRPDV SSSFQRPVND MNIAISGCPFL	780
CPVSKMEAKN SYVRDDAIFI KAIVDLTGLG SAGSAAGSGE FGSAEAAKE AAAKAGSAGS	840
AAGSGEFGSAA SKLQLSSDGE IFEVDVIEAK QSVTIKTMLE DLGMDEGGD DPVPLPVNA	900
AIIKKVIIQWC THHKDPPP EDDENKEKRT DDIPVWDQEF LKVDQGTLFE LILAANYLDI	960
KGLLDVTCKT VANMIKGKTP EEIRKTFNPK NDFTEEEAAQ VRKENQWC	1008
SEQ ID NO: 114	moltype = AA length = 717
FEATURE	Location/Qualifiers
source	1..717
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 114	
MTADNNNSMPH GAVSSYAKLV QNQTPEDEILE EFPEIPLQQ RPNTRSSETS NDESGETCFS	60
GHDEEQIKLM NENCIVLDWD DNAIGAGTKK VCHLMENIEK GLLHRAFSVF IFNEQGELL	120
QORATEKTF PDLWTNTCCS HPLCIDDELG LKGKLDDKIK GAITAAVRKL DHELGIPED	180
TKTRGKFHFL NRHYMAPSN EPWGEHEIDY ILFYKINAKE NLTVNPVNNE VRDFKWVSPN	240
DLKTMFADPS YKFTPWFKII CENYLFNWWE QLDDLSEVEN DRQIHRMLKL SGGGGGGG	300

-continued

SGGGGSAEAW	YNLGNAYYKQ	GDYQKAIEYY	QKALELPNPN	AEAWYNLGN	YYKQGDYQKA	360
IEYYQKALEL	DPNNAAEAWYN	LGNAYYKQGD	YQKAIEDYQK	ALELPNNLQ	AEANKNLGN	420
YYKQGDYQKA	IEYYQKALEL	DPNNASAWYN	LGNAYYKQGD	YQKAIEYYQK	ALELPNNAK	480
AWYRRGNAYY	KQGDYQKAIE	DYQKALELDP	NNRSRSAGGG	GSGGGGSAGGG	GASSYYHHHH	540
HHELESTSLYK	KAGSGSNTVK	ELKNYIQLE	ERNAELKNLK	EHLKFAKAL	EFEELAAHKFE	600
GSAGSAAGSG	EFGSAEAAA	EAAAAGKSAG	SAAGSGEFGS	SYYHHHHHHL	ESTSLYKKAG	660
SGSQKVQQLK	NRVAYKLKEN	AKLENIVARL	ENDNANLEKD	IANLEKDIAN	LERDVAR	717

SEQ ID NO: 115	moltype = AA	length = 798				
FEATURE	Location/Qualifiers					
source	1..798					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 115						
MEEAKIDELIN	NDPVWSSQNE	SLISKPYNH	LLKPGKNFRL	NLIVQINRVM	NLPKDQLAIV	60
SQIVELLHNS	SLLIDDIEDN	APLRRGQTT	HLIWGPVSTI	NTANYMYFRA	MQLVSQLTTK	120
EPLYHWLITI	FNEELINLHR	GQGDIYWRD	FPEIIPTOQE	MLNMYMMNKT	GGFLFRLLRL	180
MEALSPSSH	GHSLVPPFINL	LGIIYQIRDD	YLNLDKFQMS	SEKGFAEDIT	EGKLSFPIVH	240
ALNFTKTKQQ	TEQHNEILR	LLRLRTSDKDI	KLKLQILEF	DTNSLAYTKN	FINOLVNMIK	300
NDNENKYLPD	LASHSDTATN	LHDELLYIID	HLSELKLSGG	GGSGGGGGGG	GGSAEAWYNL	360
GNAYYKQGDY	QKAIEYYQKA	LELDPNNAEA	WYNLGNAYYK	QGDYQKAIEY	YQKALELPN	420
NAEAWYNLGN	AYYKQGDYQK	AIEDYQKALE	LDPNNLQAEA	WKNLGNAYYK	QGDYQKAIEY	480
YQKALELPN	NASAWYNLGN	AYYKQGDYQK	AIEYYQKALE	LDPNNAKAWY	RRGNAYYKQG	540
DYQKAIEDYQ	KALELPNNR	SRSAGGGGSG	GGGSGGGGAS	LCTMKKGPSG	YGFNLHSOKS	600
KPQFQFIRSV	PDSPABASGL	RAQDRIVEVN	GVCMEGKQHG	DVVSAIRAGG	DETKLVVDR	660
EGSAGSAAGS	GEFGSAEAAA	KEAAAKAGSA	GSAAGSGEF	SSSGAIYTV	ELKRYGGPLG	720
ITISGTEEPF	DPIIISSLTK	GGLAERTGAI	HIGDRILAIN	SSSLKGKPLS	EAIHLLQMAG	780
ETVTLKIKKQ	TDAQPASS					798

SEQ ID NO: 116	moltype = AA	length = 883				
FEATURE	Location/Qualifiers					
source	1..883					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 116						
MNHLRAEGPA	SVLAIGTANP	ENILLQDEF	DYYFRVTKSE	HMTQLKEKFR	KICDKSMIRK	60
RNCFLNEEHL	KONPRLVHE	MQTLNDARQDM	LVVEVPKLGK	DACAKAIKEW	GQPKSKITHL	120
IFTSASTTDM	PGADYHCAKL	LGLSPSVER	MMYQLGCYGG	GTVLRIAKDI	AENNKGARVL	180
AVCCDIMACL	FRGPSESDL	LLVGQAIFGD	GAAAVIVGAE	PDESVERPRI	FELVSTGQT	240
LPNSEGTIGG	HIREAGLIFD	LHKDVPM LIS	NNIEKCLIEA	FTP IGDWN	S IFWTHPGG	300
KAILDKVEEK	LHLKSDKFVD	SRHVLEHGN	MSSSTVLFVM	DELRKRSLEE	GKSTTGDGFE	360
WGVLFGPGP	LTVERVVVR	VPIKQVLSR	GGGGGGGGG	GGSAEAWYNL	GNAYYKQGDY	420
QKAIEYYQKA	LELDPNNAEA	WYNLGNAYYK	QGDYQKAIEY	YQKALELPN	NAEAWYNLGN	480
AYYKQGDYQK	AIEDYQKALE	LDPNNLQAEA	WKNLGNAYYK	QGDYQKAIEY	YQKALELPN	540
NASAWYNLGN	AYYKQGDYQK	AIEYYQKALE	LDPNNAKAWY	RRGNAYYKQG	DYQKAIEDYQ	600
KALELELPNNR	SRSAGGGGSG	GGGSGGGGAS	GNNLETYEWY	NKSISRDKAE	KLLLDTGKEG	660
AFMVRDSRTP	TYTVSFTK	AIISENCPIK	YHYHCKETND	PKRYVVAEKY	VFD SPLITLIQ	720
YHQYNGGVL	TRLRYPVC	SAGSAAGSGE	FGSAEAAA	AAAKAGSAGS	AAGSGEF GSG	780
SHPWFFGKIP	RAKAEEMLSK	QRHDGAFLIR	ESESAPGDFS	LSVKFGNDVQ	HFKVLRDAG	840
KYFLWWVKFN	SLNELVVDYHR	STS VSRNQ	QPT			883

SEQ ID NO: 117	moltype = AA	length = 665				
FEATURE	Location/Qualifiers					
source	1..665					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 117						
MAVKHLIVLK	FKDEITEAQK	EEFFKTYVNL	VNIIPAMKDV	YWGDVDTQKN	KEEGYTHIVE	60
VTFESVETIQ	DYIIHPAVG	FGDVYRSFW	KLLIFDYTPR	KKLSSGGGG	GGGSGGGSA	120
EWAYNLLGNAY	YKQGDYQKA	EYYQKALELD	PNNAEAWYNL	GNAYYKQGDY	QKAIEYYQKA	180
LELDPNNAEA	WYNLGNAYYK	QGDYQKAIED	YQKALELPN	NLQAEAWKL	GNAYYKQGDY	240
QKAIEYYQKA	LELDPNNASA	WYNLGNAYYK	QGDYQKAIEY	YQKALELPN	NAKAWYRGN	300
AYYKQGDYQK	AIEDYQKALE	LLNPNRSRQK	GGGGSSGGGG	GGGGASQDR	SEATLIKRFK	360
GEGVRYKAKL	IGIDEVSAAR	GDKLCQDSMM	KLKGVVAGAR	SKGEHKQKIF	LTI SFGGIKI	420
FDEKTGALQH	HHAVHEISYI	AKDITDHRAF	GYVCGKEGNH	RFVIAKTAQA	AEPVILDRLD	480
LFLOLIYELKQ	REELEKKAGS	AGSAAGSGEF	GSAAEAAA	AAKAGSAGS	AGSGEF GSGS	540
HMGSQFWVTS	QKTEASERC	LQGSYIILRVE	AEKLTLTLLG	AQSQILEPLL	FWPYTLR	600
GRDKVUMFSFE	AGRRCPSGPG	TFTFQTSQGN	DIFQAVEAAI	QQQKAQGKVG	QAQDILRLEH	660
HHHHH						665

SEQ ID NO: 118	moltype = AA	length = 797				
FEATURE	Location/Qualifiers					
source	1..797					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 118						
MGLSSVCTFS	FQTNYHTLN	PHNNNPKTS	LCYRHPKTP	KYSYNNFP	SK HCSTKSFH	60
NKCSESLIA	KNSIRAATTN	QTEPPESDNH	SVATKILNFG	KACWKLQRPY	TIIAFTSCAC	120
GLFGKELLHN	TNLISWSLMF	KAFFFLVAIL	CIASFTTTIN	QIYDLHIDR	NKPDLPLASG	180

-continued

EISVNTAWIM	SIIVALFGGLI	ITIKMKGGLP	YIFGYCFGIF	GGIVYSVPPF	RWKQNPSTAF	240
LLNPLFLHIIIT	NFTFYASRA	ALGLPFLFELRP	SFTFLAFLMF	SMGSALALIK	DASDVEGDTK	300
FGISTLASKY	GSRNLTLCFS	GIVLLSVYAA	ILAGIWIWPQA	FNSNVMLLSH	AILAFWLILQ	360
TRDFALTNYD	PEAGRFRFYEF	MWKLYAYEAL	VVFYFIKLSGG	GGSGGGGSGG	GGSAEEAWYNL	420
GNAYYKQGDY	QKAIEYYQKA	LELDPNNAEA	WYNLGNAYYK	QGDYQKAIEY	YQALELDPN	480
NAEAWYLNGLN	AYYKQGDYQK	AIEDYQKALE	LDPNNQLQAEA	WKNLGNAYYK	QGDYQKAIEY	540
YQALELDPN	NASAWNYLGN	AIIYQKQGDYQK	AIEYYQKALE	LDPNNAKAWY	RGRNAYYKQG	600
DYQKAIEDYQ	KALELDPNIR	SRSAGGGGSG	GGGSGGGGAS	AEYVRALFDF	NGNDEEDLDPF	660
KKGDLILRIRD	KPEEQWNNAE	DSEGKRGMIP	VPYVEKYGSA	GSAAKSGEFG	SABAAKEA	720
AKAGSAGSA	GSGEGPSL	HMRAEALFDF	TGNSKLELNF	KAGDVIFLLS	RINKDWLEG	780
VRGATGIFPL	SVFKILK					797

```
SEQ ID NO: 119      moltype = AA  length = 3620
FEATURE          Location/Qualifiers
source           1..3620
                  mol_type = protein
                  organism = synthetic construct
```

SEQUENCE : 119
 MGSAAGSAGS GEFGSAGSAA GSGEFGSAGS AAGSGEFSYY HHHHHHLEST SLYKKAGSGS 60
 ARNALYRKKI ARLKKDNLNQL ERDEQNLEKI IANLRDEIAR LENEVASHEQ GSAGSAAGSG 120
 EFAEEAAKEA AAKAGSAGSA AGSGEFSSYY HHHHHLESTS LYKKAGSGSN LVAQLENEVA 180
 SLENENETLK KKNLHKKDLI AYLEKEIANL RKKIEEGSG SAAGSGEFGS AEEAAKEAAA 240
 KEAAAKEAAA KAGSAGSAA SGEGFSSYY HHHHHLESTS LYKKAGSGSQ KVAAELKNRVA 300
 VKLNRRNEQL NKVEELKNR AYLKNELATL ENEVARLEND VAEGSAGSAA GSGEFAEAAA 360
 KEAAAKAGSA GSAGSGEFGS YYHHHHHHHLS TSLYKKAGS GSNEVTTLEN DAAPIENENA 420
 YLEKEIARLR KEKAALRNRL AHKKGSAGSA AGSGEFGSAE AAAKEAAKE AAAKEAAAKA 480
 GSAGSAAGSG EFGSRPPTIS NPPPLISSAK HPSVGAGSA AGSGEFAEAA KEAAAKAGS 540
 AGSAAGSGEF NFLQSRPTE APPESFRSG GSAGSAAGSG EFGSAAEAAK EAAKEAAAK 600
 EAAAKAGSAG SAAGSGEFGS SKGTGLNPNA KVWQEIAPGN GSAGSAAGSG EFAAAEAAA 660
 AAAGKSAGSA AGSGEPFDGG TTFEHLWSLL EPDSTYGSAG SAAGSGEFGS EAAAEAAA 720
 KEAAAKEAAA KAGSAGSAA SGEGFSSYY HHHHHLESTS LYKKAGSGSK RIAYLRKKJ 780
 ALKKDNANL KDIANLENET ERLIKEIKTL ENEVASHEQ GSAGSAAGSF FAEAAKEEA 840
 AKAGSAGSAA SGEGFSSYY HHHHHLESTS YKKAGSGNL LATLRLSTA AVLENENHVLEK 900
 EKEKLRKEKE QLNLNKLEAYK GSAGSAAGSG EFGSAAEAAK AAAKEAAAK EAAAKAGSAG 960
 SAAGSGEFGS PATSQHPPP PGHRSQAPSH GSAGSAAGSG EFAAAAKEA AAKAGSAGSA 1020
 AGSGEFLNS LLILLEAAYE LERRDRGSSAG SAAGSGEFGS AAAAAKEAAA KEAAKEAAA 1080
 KAGSAGSAA SGEGFSRPT ISNPPLLISSAK AHKHPVGAG SAAGSGEFAE AAAKEAAAKA 1140
 GSAGSAAGSG EFNFSOLRPE PTAPPEESFR SGGSAGSAA SGEGFSEAAA KEAAKEAAA 1200
 AKEAAAKAGS AGSAAGSGEF GSSKGTLNP NAKVWQEIAEP GNGSAGSAAG SGFEFAEAAA 1260
 EAAAKAGSAG SAAGSGEFPD GGTTFEHLWS SLEPDSTYGS AGSAAGSGEF GSAAEAAKEA 1320
 AKEAAAKEAA AAKAGSAGSA AGSGEFSSYY YHHHHHHHLS TSLYKKAGSG SKRIAYLRKK 1380
 IAALKDDNAN LEKDIANLEN EIERLIKEKTL TLENEVASHE QGSAGSAAGS GEFAEAAAKE 1440
 AAAKAGSAGS AAGSGEFSYY HHHHHHLESTS SLYKKAGSGS NLLATLRLSTA AVLENENHVL 1500
 EKEKEKLRLKE KEQLLNKLEA YKGSAGSAA SGEGFSEAAA AAAKEAAKEA AKEAAAKAGS 1560
 AGSAAGSGEF GSALVDDAAD YEPPEPSNNEE ALGSAGSAG SGEGFSEAAA EAAAKAGSAG 1620
 SAAGSGEFPRE LFDDPSYVNVN QNLDDKARQGS AGSAAGSGEF GSAAEAAKEA AKEAAKEEA 1680
 AAKAGSAGSA AGSGEGSKN TKSMMFDNPV YRKTTEEEGS AGSAAGSGEF AAAKEAAKEAA 1740
 KAGSAGSAA SGFERSLPST WIENKLYGMS DPNWGSAGSA AGSGEFGSAE AAAKEAAAKE 1800
 AAKAEAAAKA GSAGSAAGSG EFGSVVDINS PPALPKKRQ SAPSGSAGSA AGSGEFAEAAA 1860
 AKEAAAKAGS AGSAAGSGEF TORSKQPQAV PPRPSADLIL GSAGSAAGSG EFGSAAEAAA 1920
 EAAKEAAKEA EAAAKAGSAG SAAGSGEFGS TDREERETEE VELYNNSTTL GSAGSAAGSG 1980
 EFAEAAKEA AAKAGSAGSA AGSGEFDGNV SGTQRLLSAT VRTYSCGSAG SAAGSGEFGS 2040
 AEAEEAKAEEA KEAAAKEAAA KAGSAGSAA SGEGFSSYYH HHHHHHLESTS LYKKAGSGSQ 2100
 KVAQLNRVA YKLKENAKL NIVARLENND ANLEKDIANL EKDIANLERD VARGSGASAA 2160
 GSGEFAEAAA KEAAAKAGSA GSAGSAAGSGEFS YYHHHHHHHLS TSLYKKAGS GSNTVKELKN 2220
 YIQELEERNA ELKNLKEHLK FAKAELEFEL AAHKFEGSAG SAAGSGEFGS AAAAAKEAAA 2280
 KEAAAKEAAA KAGSAGSAA SGEGFHSDDS LPHPOQATDD SGHESDGSAQ SAAGSGEFAE 2340
 AAAKEAAAKA GSAGSAAGSG EFGSPNAGSV EQTPKPGPL RRGSAGSAAG SGEGFSEAAA 2400
 KEAAAKEAAA KEAAAKAAGS AGSAAGSGEF GSSYYHHHHH HLESTSLYKK AGSGSFENVT 2460
 HEFILATLEN ENAKLRLREA KLERELARLR NEAVAWLGSAG SAAGSGEFAE AAAKEAAAKA 2520
 GSAGSAAGSG EFSYHHHHHH HLESTSLYKK AGSGSQKVEE LKNKIAELEN RNAVKKNRVA 2580
 HLKQEIAYL DELAAHEFEG SAGSAAGSGE FGSAEAAAKE AAAKEAAKE AAAKAGSAGS 2640
 AGSGEGPSV SSKTLVFSHD DSDEDDLHII SAGSAAGSGE FAEAAKEAA AKAGSAGSA 2700
 GSGEFAAATP ISTFHDSDDE DLLHVGSGS AAGSGEFGSA AAAKEEEAAK AAAKEAAAK 2760
 AGSAGSAAGS GEGFSSYYH HHHHHLESTS YKKAGSGSQ VESLKQKIEE LKQRKAQLKN 2820
 DIANLEIKEI YAETGSAGSA AGSGEFAEAA KEAAAKAGS AGSAAGSGEF SYYHHHHHHH 2880
 ESTSLYKKAG SEFFRRERKN MAAAKCRNRR RELTDTLQAE TDQLEDEKA LQTEIANLLK 2940
 EKEKLEPILA AHRPACKIPD DLGPFPEMIL EGSGAGSAA GEFGSAAEAAA KEAAKEAAA 3000
 KEAAAKAGSA GSAGSGEFG SFQMPADTP PAYLPPEDPM TGSAGSAAGS GEFAEAAAKE 3060
 AAAKAGSAGS AAGSGEFEIRE SNEEPPPYE DPYWGNGGS AAGSGEFG SAEAAAKEAAA 3120
 KEAAAKEAAA AKAGSAGSAA GSGEFGSSYY HHHHHHLESTS SLYKKAGSGS QVAAELKNRV 3180
 AVKLNRNEQL KNKVEELKNR NAYLKNELATL ENEVARLEN DVAEGSAGSA AGSGEFAEAAA 3240
 KEAAAKAGS AGSAAGSGEF SYYHHHHHHL ESTSLYKKAG GSNEVTTLE NDAAFIENEN 3300
 AYLEKEIARL RKEKAALRNRL LAHKKSYYH HHHHHLESTS YKKAGSGSAR NAYLRRKKIAR 3360
 LKKDNLQLER DEQNLEKIA NLRDEIARLE NEVASHEQGS AGSAAGSGEF AAAAKEAAA 3420
 KAGSAGSAG SGEFSSYYH HHHLESTSLY KKAGSGSNLV AQLENEVASL ENENETLKKK 3480
 NLHKKDLLIAY LEKEIANLRK KIEEGSAGSA AGSGEFGSAE AAAKEAAKE AAAKEAAAKA 3540
 GSAGSAAGSG EFGSQKLIS EEDLEQKLIS EEDLEQKLIS EEDLGSGSA AGSGEFGSAG 3600
 SAAGSGEFGS AGSAAGSGEF

-continued

SEQ ID NO: 120 moltype = AA length = 452
 FEATURE Location/Qualifiers
 source 1..452
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 120
 MGSAGSAAAGS GEFGSAGSAA GSGEFGSAGS AAGSGEFSYY HHHHHLEST SLYKKAGSGS 60
 ARNAYLRRKI ARLKKDNQLQ ERDEQNLEK IANLRDEIAR LENEVASHEQ GSAGSAAGSG 120
 EFAEEAAKEA AAKAGSAGSA AGSGEFSSYY HHHHHLESTS LYKKAGSGSN LVAQLENVA 180
 SLENENETLK KKNLHKKDLI AYLEKEIANL RKKIEEGSAG SAAGSGEFGS AEEAAKEEAAA 240
 KEAAAKEEAAGA KAGSAGSAAAG SGEFGSSATR ELDELMASL DFKIQGGSAG SAAGSGEFAE 300
 AAAEAAAKAAKA GSAGSAGSAAAG EFDLALSENW AQEFLAAGDA VDGSAGSAAAG SGEFGSAAEA 360
 AKEEAAKEEAA KEEAAAKAGS AGSAAGSGEF GSDYKDDDK DYKDDDDKDY KDDDDKGSAG 420
 SAAGSGEFGS AGSAAGSGEF GSAGSAAGSG EF 452

SEQ ID NO: 121 moltype = DNA length = 3303
 FEATURE Location/Qualifiers
 source 1..3303
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 121
 atgtctgcta aagcttattc tgaacaaact ggttaaagaat tggttataa attttatttg 60
 actacttctt ctatccaaaa tagattttaa tatgcttagat ttactccca tactgattgg 120
 gcttagattt tgcaagatca tccatgggtt ttgttccaaa atttgggtt taaaaggat 180
 caattgttta aaagaaggagg taaatgggtt ttgggttggg tttatgtac ttggatggt 240
 gttaaatctt ggttgaacc aagattgggtt caagaagacta ctgttggtaa agctactgg 300
 ttttggaaaa atttttgtat tgaaccattt gttccacattt ctcagaactt gagaattttat 360
 gttttgattt atgctactt agaaagggtat tatgtttttt ttcatcgat aggtgggttt 420
 gtatgggtt atgttgcattc taaaatggcc ttatgggttgg tttgggttgg tggaaatgg 480
 aatccagaaat atattaaaaaa acatttgtt gttcatgtc cagaagataaa aaaagaaatt 540
 ttggcttctt ttatcccttgg ttgtttaat ttatgttggg attttgtt tacttatttg 600
 gaaatataatc cattgggtt tactaaatggg ggttggattt tttttggattt ggctgtcaa 660
 gtgtatgtca ctgtgttataa ttgttggaaa gttaaatggg tgatgttattt atttccacca 720
 ccattttggta gagaagctt tccagaagaa gcttatattt ctgatgggaa tgcttaatct 780
 ggtgttcttgg tggaaatttgc ttgttggat cccaaaggtaa gaattttggc tatgttgg 840
 ggtgttgggtt ctgttgcattt ttatcccttgc actattttgtt atttgggttgg tggaaatgg 900
 ttggcaattt atgttgcattt ttgttggatcc ccatctgaa acaaaactt tgatgttgc 960
 aaaacttattt tttttttttt gacttagagaa aaacatcccag atggtaaaat ttttggattt 1020
 ggtgttggat tttgttggat ttgttggat ttgttggat ttgttggat ttgttggat 1080
 attagagat atcaagggtcc atggaaagaa catgaagtta ctatgggtt tagaaagggat 1140
 ggttccaaat atcaagaagggg tttggagatc atttgggttgg tttgggttgg tggaaatgg 1200
 ccaatttcatg tttttggatc tgaaacttcat atgactgtca ttgttggat ttgttgggtt 1260
 catagaccaa ttccaaatca accaccaact gctgttcata ctgtcaattt ttttggat 1320
 gttttttttt ctatccatc tccatgttca ttctttttt tttttttt tttttttt 1380
 attagagat atcaagggtcc atggaaagaa catgaagtta ctatgggtt tagaaagggat 1440
 ggttccaaat atcaagaagggg tttggagatc atttgggttgg tttgggttgg tggaaatgg 1500
 ctttgcattt atttttttt gttttttttt tttttttt tttttttt tttttttt 1560
 gttttgggtt tgcaaaacttgc agctgttca ggtatgttgg attttggat tttttttt 1620
 agagatgtac catctgttgc tttttttt tttttttt tttttttt tttttttt 1680
 tttttttt gtcataaaaaa aattttttt ccagggttttta aaaaatggc tgatgttgc 1740
 agaaaacatc cagaagggtt tttttttt aattttttt ctttggatgc tgctttttt 1800
 tttttttt tttttttt tttttttt tttttttt tttttttt tttttttt 1860
 atttttttt tttttttt tttttttt tttttttt tttttttt tttttttt 1920
 atttttttt tttttttt tttttttt tttttttt tttttttt tttttttt 1980
 tttttttt tttttttt tttttttt tttttttt tttttttt tttttttt 2040
 tttttttt tttttttt tttttttt tttttttt tttttttt tttttttt 2100
 tttttttt tttttttt tttttttt tttttttt tttttttt tttttttt 2160
 gttttttt tttttttt tttttttt tttttttt tttttttt tttttttt 2220
 gttttttt tttttttt tttttttt tttttttt tttttttt tttttttt 2280
 gttttttt tttttttt tttttttt tttttttt tttttttt tttttttt 2340
 gttttttt tttttttt tttttttt tttttttt tttttttt tttttttt 2400
 tttttttt tttttttt tttttttt tttttttt tttttttt tttttttt 2460
 gttttttt tttttttt tttttttt tttttttt tttttttt tttttttt 2520
 gttttttt tttttttt tttttttt tttttttt tttttttt tttttttt 2580
 gttttttt tttttttt tttttttt tttttttt tttttttt tttttttt 2640
 gttttttt tttttttt tttttttt tttttttt tttttttt tttttttt 2700
 gttttttt tttttttt tttttttt tttttttt tttttttt tttttttt 2760
 gttttttt tttttttt tttttttt tttttttt tttttttt tttttttt 2820
 gttttttt tttttttt tttttttt tttttttt tttttttt tttttttt 2880
 gttttttt tttttttt tttttttt tttttttt tttttttt tttttttt 2940
 aaaaatccatc aatatttttgc tcaatttttgc aaaaatggtgc tttttttt tttttttt 3000
 aatccatc aatatttttgc tcaatttttgc aaaaatggtgc tttttttt tttttttt 3060
 atttttttt tttttttt tttttttt tttttttt tttttttt tttttttt 3120
 tttttttt tttttttt tttttttt tttttttt tttttttt tttttttt 3180
 tttttttt tttttttt tttttttt tttttttt tttttttt tttttttt 3240
 caagggtttt atagacatcc atgggatgtat tttttttt tttttttt 3300
 atgttgcattt tttttttt tttttttt tttttttt tttttttt tttttttt 3360

-continued

-continued

gtttctaa	ttactgctgc	ttttggtat	ggtgctgta	ctatggtgt	ttctttgaa	240
aaagctggt	ctgaaactaa	atatggta	ccaggtgg	ataataattt	ggctttgtat	300
gaagctgtc	aaaggagaagg	ttttgtatct	gttactat	atggggatgc	ttttttgt	360
gaaatattaag	ctcaagtta	tgaagaagct	aaaaaaaaag	tgataaaatt	tgatgttatt	420
gtttatctt	ttggcttccc	agttttaact	gatccagata	ctggattat	gcataaaatct	480
gttttggaaac	catttggtaa	aacttttaact	ggtaaaactg	ttgatccatt	tactggtaa	540
ttgaaagaaa	tttcgtcgt	accagctaat	gatgaagaag	ctgtgtctca	tgtttaagg	600
atgggggtgg	aaagattggga	aaagatggatt	aaaaactatgt	ctaaagaagg	ttttgggg	660
gaagggtgtt	ttacttggc	ttatctttt	atgggttccag	agactctca	agtttggat	720
agaaaaggtt	ctatggtaa	agctaaagaa	catttggaaag	ctactgtca	tagattgaat	780
aaagaaaaatc	catctttag	agcttttttt	tctgttataa	aagggttgg	tactagatc	840
tctgtctgtt	ttcccgat	tccattgtat	ttgggttctt	tgtttaagg	tatgaaagaa	900
aaaggataatc	atgaagggtt	tattgtaccaa	attactatgt	tgtatgtca	aaaggatgtat	960
agaaaagatg	gtactatccc	agttgtatggaa	gaaaatagaa	ttagaattga	tgatgtggaa	1020
ttggaaagaa	atgttcaaaa	agctgtttt	gctttgtatgg	aaaaagttac	tgggtggaaaat	1080
gtgtttatctt	tgactgtt	ggctgggtt	agacatgatt	ttttggctt	taatgggttt	1140
gtatgtttaat	ttqaqctqaa	qttqaaat	tttgcataat	tttgcataat	t	1191

SEQ ID NO: 126 moltype = DNA length = 1182
FEATURE Location/Qualifiers
source 1..1182
mol_type = other DNA
organism = synthetic construct

SEQ ID NO: 127 moltype = DNA length = 1476
FEATURE Location/Qualifiers
source 1..1476
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 127

atgaaactct	caactaaaact	ttgttggtgt	ggtattaaag	gaagacttag	gccgaaagg	60
caacaacaat	tacacaatac	aaactgc当地	atgactgaa	taaaaaaaa	aaaagaccgt	120
gaacaaaaaa	ccgacactca	aatgtcggt	attaaaggta	tccaaattta	catacccaact	180
caatgtgtca	accatctga	gctagagaaa	tttgcgttgc	tttctcaagg	taataacaca	240
attgggtctgg	gc当地aaaccaa	catgtcttt	gtcaatgaca	gagaagatata	ctactcgatg	300
tccctaactg	ttttgtctaa	gttgc当地aa	agttacaaca	tc当地acccaa	caaattgtt	360
agattagaaag	teggtaactga	aactctgatt	gacaactc当地	agtc当地gtca	gtctgtctgg	420
atgc当地attg	tttgc当地aa	cactgc当地	gaaggatatt	acacgtttaa	tgccgttac	480
gggtggtacca	acgcgttgc	caactcttgc	aactgtattt	aatctaaacgc	atggatgtt	540
agagacgc当地	tttgc当地gtt	cggtgatatt	gccatcttacg	ataagggtgc	cgcaagacca	600
accgggtgtt	ccggtaactgt	tgctatgttgc	atcggttctg	atgc当地ccaa	tgatatttgc	660
tctgttaaagg	cttcttactgg	ggaaacacgc当地	tc当地atgtt	acaacttgc当地	tttaccaggc	720
gaatactctt	acgtcgatgg	tc当地tttgc当地	ttaacttgc当地	acgtcaaggc当地	tcttgc当地aa	780
gtttacaaga	gttattccaa	gaaggctatt	tcttaagggtt	tggtagc当地	tcccgttgc当地	840
tc当地ggatgtt	tgaacgtt	tttgc当地tttgc当地	gactacaacg	tttccatgt	tccaaacctgt	900
aaatgtgttca	caaatacata	cggtgatatta	ctatataacg	atttcagacg	caatccctca	960
tttgttccca	aaagtgc当地	cgcaataggat	actcgc当地att	atgacgaaatc	ttaaccaggc当地	1020
agaacacattt	aaaaaaaaactt	tgttaatgtt	gctaaaggccat	tccaaaaga	gagagttgc当地	1080
caatcttgc当地	ttgttccaa	aaacacagg	aacatgtaca	ccgcatctgt	ttatgc当地cc	1140
tttgc当地atctc	tattaaactt	tggttgatct	gacgacttac	aaggcaagcg	tggttggtt当地	1200
ttttcttactg	gttcccggtt	agctgc当地atct	ctatatttctt	gcaaaaattgt	tggtgc当地gtc当地	1260
caacatattt	tcaaggaaat	agatattact	aacaatttag	ccaagagaatc	caccggaaactt	1320
ccaaaggattt	acgaagctgc当地	ctcgatggat	agagaaaaatg	ccccatttgc当地	gaagaacttgc当地	1380
aaacctcaag	gttccatgg	gcatggatc当地	agtggatgtt	actacttgc当地	caacatcgatc当地	1440
gacaaaattta	gaagatctt	cgatgttaaa	aaataaa			1476

SEQ ID NO: 128 moltype = DNA length = 1143
FEATURE Location/Qualifiers

-continued

-continued

```

tttagaaaaa taactaaaga atctgggcg gatatcgAAC ctcccgtaca aactagCTTA 1140
ttggatgatt gccagacTTT aaaaggAGTT tttaCTTGt taatacCTGG tgctggTTG 1200
tatgacgcCA ttgcAGTgtat tactaAGCA gatgttGATC tttagggCTCA aaccgCTAAT 1260
gacaaaAGAT ttctAAAGGT tcaatggCTG gatgtAACTC aggctgACTG ggggttAGG 1320
aaagaaaaAG atccggAAAC ttatCTTgtat aaataa 1356

```

```

SEQ ID NO: 131      moltype = DNA length = 1191
FEATURE          Location/Qualifiers
source           1..1191
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 131
atgaccgttt acacgcATC cgttaccgcA cccgtcaaca tcgcaacCCt taagtatTTG 60
ggaaaaAGGG acacgAAggtt gaatTCGCC accaattCTG ccatatCAGt gactttatG 120
caagatgacc tcagaACGTT gacCTCTGCG gctactGCAC ctgagTTGA acgcgacACT 180
ttgtggTTAAttggAAACC acacagcATC gacaatGAAa gaactcaAAA ttgtctGCgC 240
gacctacGCC AattaAGAAg ggaatGGAA tCGAGGGACt ctcattGCC cacattATC 300
caatggAAAC tccacatTTG ctccatTTGt aacttCCtA cagcAGCTGG tttagCTCC 360
tccgCTGTCG gtttGCTGc attgttCTtC gcaattGCTA agttatACA atttACACAG 420
tcaactTCAG aataatCTAG aataGCAAGA aaggGGTCTG gttcAGCTG tagatCttG 480
tttggcGGat acgtggCTG ggaaatGGGA aaagCTGAGG atggTCatGA ttccatGGCA 540
gtacAAatTCG cagacgCTC tGactggCCt catGAAAGG ctGtGtGtCt agttGTCAGC 600
gtatattAAAAGGAGTGTAG ttccatCTAG ggtatGCAAT tGaccGtGGC aacCCGCAA 660
ctatttaAAAG aagaATTGA acatGTCGA ccaaAGAGAT ttGAAGTCAt gCGtaAAAGCC 720
atttgtAAAG aagattTCGc caccTTGCA aaggAAACAA tGatGGATC caactCTTC 780
catGCCACAT tttccatTCAC atatttCTACt tGAATGACAC tTCCAAAGGCGT 840
atcatAGTtGtGccACAC cattaATCAg ttttacGGAG aacaATCAGt tGcataCACG 900
tttgatGCGGt GtCCAAATGc tGtGtGtGtC tacttagCTG aaaaGAGtC gaaACTCTT 960
gcatttatCt ataaattGTT tggctCTGt tctggatGGG acaAGAAatt tactactGAG 1020
cagCTGAGGt CTTCAACCA tcaatTTGAA tcatCTAACT ttactGCAcG tGAATTGGAT 1080
cttgAGGTGc AaaggAGATGt tgccAGAGtG atTTAACTC aagtCGGTc aggCCCAcAA 1140
gaaaacaAAAC aatCTTGTat tgacGCAAAG actGGtCTAC caaAGGAATA a 1191

```

```

SEQ ID NO: 132      moltype = DNA length = 867
FEATURE          Location/Qualifiers
source           1..867
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 132
atgactgocG acaacaATAG tatGCCCAT ggtgcAGTat ctatTTACGc caaattAGT 60
caaaaccaAA caccTGAAGA catTTGGAA gagTTTCTGt aaatttTCC attacaACAA 120
agacacATAA cccgATCTAG tGAGACGTCt aatGACGAAAC gCGGAGAAAC atGTTTTCT 180
ggTCatATGtGtGAGGAGTATG taaGTTATGt gtatGTTTt ggattGGGAC 240
gataatGota ttggTGCcGG tACCAAGAAA gtttGTCatT taatGAAAAA tattGAAAG 300
ggTTTactAC atcGTGcATT ctccGTCttt atttCAATG aacaAGGTGA attactTTA 360
caacAAAGAG ccactGAAAG aataATCtCt cctGATCttt ggactAAACAC atGTCGCTC 420
catCCactAT tGATTGATGA cgaattAGGT tGAGGGGt agttagAGCA taAGATTAAg 480
ggcgcTTAtta ctgcGGCGGT gagaAAActA gatCAtGAAT tagGtATCC agaAGAtGAA 540
actaAGAcAA gggGtaAGTT tcaattttTA aacAGAtCC attACAtGGC accAAAGCaAT 600
gaaccatGGG tGAGACATGA aattGATTACt atCtatttTA atAAGAtCtAA CGCTAAAGAA 660
aacttgAActG tcaacccAAAG cgtcaatGAA gttAGAGACT tcaaAtGGGT ttcaACAAAT 720
gattGAAAAt tGATGTTGc tgaccAAAGT tacaAGTTA cgcCTGGTt taAGATTt 780
tgcgagaAtt acttattCAA ctggTGGGAG caattAGATG acCTTCTGA agtGgAAAt 840
gacaggcAAAt ttcAtGAtAAt gctataAA 867

```

```

SEQ ID NO: 133      moltype = DNA length = 1059
FEATURE          Location/Qualifiers
source           1..1059
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 133
atggCTTCAG AAAAGGAAAt aagaAGAGAA agattCTtG ACgttACtCC AAgttAGtT 60
gaagaATTGt acgttAGTTt ttGtGtGCTtA aagaAGCCTG CGATTGGTAt 120
gtcactCTT taaACTACAA tACTCCAGGT ggttaATGtA atAGGTTGt gAGtGtAGtT 180
gataCTTATG CTATCTTGTc TAACAAACo gttGAACAtA tagGTCAAGA AGAAtACGAA 240
aaggTCGtCA ttttGGGTTG tGtGtGtGAA ttGtGtGCAg cataCTTTt ggTTGCGGAt 300
gacatGATGGt atAGTCTAt aacaAGAAGA ggtcaACCCt gctGGtACAA agtTCCAGAA 360
gttGGTgAAA tGAGCCATTAA tGtGtGtGAA tGtGtGtGtA aataCTtCtA taaATTGTT 420
aagtCacATT tcaAGAACGt AAGTACTACt atCgAtAtTA CCgAAtTt CCACGAGtT 480
acttTCCAA AAGAATTGGG tcaattGATGt gattGAtAA CTGACtGtA AGATAAAGtT 540
gacttGtCAA AGTTTCCtT gAAGAAACAt tCAttCAtGt TCACtTTGt AACTGtCTt 600
tactCCtTtCt tttGCGtAGt CGCCtGGtCt AtGtACGtAGt CTGtGtAtACt tgatGAAA 660
gacttGAGC AAGCAAGAGA tGTTTGTAtA CTTTGGGtG AAtACTtCtA AAtCCAAGAt 720
gactactTAGt ACTGTTGCGt TACTCCAGtA CAAAtAGtGA AAtCgGTAC AGAttAtCtA 780
gacaAtAAAGt GCAgTTGGGT tAtTAACAAg GCTTGGAtA tagCATtGtC CGAAcAAAGA 840
aagacttTGGt AtGAAAActA CGGtAAAGAGt GACTCAGtGt CGAAGCtAA gTGTAAAGAA 900
atTTTAAtG AtTTGAAAGt tGAAcAAAttG tAccAtGAtA ACgAAGAAAtCt CATGtCtAA 960
gacttAAAGt CAAAGAttG tcaAGtGtAtA GAAtCAAGAGt GTTTAAAGC CGACGTTtG 1020
acagCTTtCt tGAtAAAGAt tcaAAAGAGA 1059

```

-continued

-continued

gaagatctgg aggccaacgc agaatatacgtatggccg atcaatacat tgaagtgc 360
gggtgacta ataataacaa ctacgctaac gtagacttgc tcgtagacat cccgaaaaga 420
gcagacgtat acggcgatg ggctggctgg ggtcacgcct ccggaaatcc actatggc 480
gaaaatttg ccacgttcaa gagaaaggatc atctttattt ggcctccagg taacccatg 540
aggcttttag gtgataaaat ctccttaccat atttgcttcaaaagtgctaa agtcccattgt 600
atccatggt ctggtacccgg tggacacc gttcacgtgg acgagaaaac cggtctggc 660
tctgtcgact atgacatc tcaaaaagggt tggtgacttctccatgaa tggtttaca 720
aaggccaaac gttatggttt ctctgtcatg attaaggcat cgaagggtgg tggtgttaa 780
ggtatcajac aagtgtacaa tgaaggaaat ttcatcgatc tataccacca ggcaccaac 840
gaaattccac gtcacccat tttcatcgatc aagttggccg gttagacgcgc tcacttggaa 900
gttcaactcg tagcagatc gtaacggatca aatatttctt tggctcgatg agactgttcc 960
gttcagagac gtcatcaaaa aattatcgaa gaagcaccat tataattgc caaggtgtaa 1020
acatttcacg agatggaaaaa ggctggccgtt agactggggaa aactatgtgg ttatgtctt 1080
gcccgttacc tggagttatc atatttccat gatgtggaa aatttactt tttagaattt 1140
aaccacaaat tacaactcgatc gatccaaacaa acggaaatgg ttcctcgatg taatccat 1200
gcacgttcaat tacaatccgt tttgttgc tctatggatc gaataatgtga catttgcact 1260
tttatggta tgaatcttca ttctggcttca gaaatcgatc tggatcaaa aactcaatgg 1320
gcccaccaaga aacaaagaag acctatttca aagggttcatc gtaccgttgc tcgtatcaca 1380
tcagaatgcg caaacgtatgg attcaacca tccgggtgtt ctggatcgatc actaaacttc 1440
cggttcttcctt ctaatgtttt gggttacttcc tccgtggatc acaatggtaa tatttactt 1500
ttttcgactt ctcgttccg ccatattttt gctttttgtt aaaaatgacca agtcccggtt 1560
aaacacatgg ttgttgcctt gaaggaaatttgg tccatttaggg gtgatccatg aactactgtt 1620
gaatacttgc tcaaaactttt ggaaacttgc gatggcgagg ataacactat taccacccgtt 1680
tggttggactt atttgttacatc tccataaaatgg acggcgattdtcc aactcttgc 1740
gttcatgttgc gtggcgatca aaaggcttcc ttgcatctg aagaaggcccccccaaaatggat 1800
atcgaatctt tacaacaaaggg acaatgttca tctaaagacc tactgcacaaat tatgttcccc 1860
gtagattttt tccatgggg taaaagatc aagttcacccg tagctaaatc cggtaatgc 1920
cggttacatcattttatccatgg ttttttttttgcataccatcactcgatcactatctgtt 1980
gggtgttctt tgattggatcatc ggccgtttaaa tccgcataccatcattttggaa aagaaggat 2040
gtgtcttacaa gattatccgt tgactctatcacttgc tggatgttgc aaaaatgttca 2100
accaggatgc gtactccatc ccctggtaaa ttggtaatatttgc ttttttttttgc 2160
cacatttcac aggggccaaacc atatgcacaaatggatc tggaaatgc aatggcttcc 2220
gtttctcaag aaaaatggatcatc tggccatgg ttttttttttgc ttttttttttgc 2280
gggtatcatc tggccatggatcatc ggccgtttaaa tccgcataccatcattttggaa aagaaggat 2340
ttttaaggttgc tgctggccaga ttttttttttgc ttttttttttgc 2400
aaatctcaatc ttttttttttgc ttttttttttgc 2460
attatggacttccatgg ttttttttttgc ttttttttttgc 2520
tcgaatgttgc aacttgcatc ttttttttttgc ttttttttttgc 2580
caaataatggaaatggatcatc ttttttttttgc ttttttttttgc 2640
tttaagttatc ttttttttttgc ttttttttttgc 2700
ggccggccgtcg tggaaaccatc ggccgttatttgc ttttttttttgc 2760
catgaacatc ttttttttttgc ttttttttttgc 2820
aatggtccaaatggatcatc ttttttttttgc ttttttttttgc 2880
gatcttagatc aacttgcgttgc aacttgcgttgc ttttttttttgc 2940
ctgtatcttgc ttttttttttgc ttttttttttgc 3000
gccccatcttgc ttttttttttgc ttttttttttgc 3060
gtcgcttcatc aaccaatggatcatc ttttttttttgc ttttttttttgc 3120
aactgaacaaatggatcatc ttttttttttgc ttttttttttgc 3180
aatccaaatggatcatc ttttttttttgc ttttttttttgc 3240
gctctgatcatc aacttgcgttgc aacttgcgttgc ttttttttttgc 3300
gtgtcttcatc ttttttttttgc ttttttttttgc 3360
cacaaggatgttgc ttttttttttgc ttttttttttgc 3420
tccacccatc ttttttttttgc ttttttttttgc 3480
ttgtcataatc ttttttttttgc ttttttttttgc 3540
gatcatttgc ttttttttttgc ttttttttttgc 3600
caatcttgc ttttttttttgc ttttttttttgc 3660
gttgcataatc ttttttttttgc ttttttttttgc 3720
agggttggatcatc ttttttttttgc ttttttttttgc 3780
atccatggatcatc ttttttttttgc ttttttttttgc 3840
ccaaatttgc ttttttttttgc ttttttttttgc 3900
gaataggatcatc ttttttttttgc ttttttttttgc 3960
catgttcatc ttttttttttgc ttttttttttgc 4020
attatggatcatc ttttttttttgc ttttttttttgc 4080
gctaacatggatcatc ttttttttttgc ttttttttttgc 4140
gatgttggatcatc ttttttttttgc ttttttttttgc 4200
gaaggccgttgc ttttttttttgc ttttttttttgc 4260
tcttttgc ttttttttttgc ttttttttttgc 4320
cggtccatggatcatc ttttttttttgc ttttttttttgc 4380
aagaacgcataatggatcatc ttttttttttgc ttttttttttgc 4440
agggttggatcatc ttttttttttgc ttttttttttgc 4500
agacatcttgc ttttttttttgc ttttttttttgc 4560
caacttgc ttttttttttgc ttttttttttgc 4620
tcccaatggatcatc ttttttttttgc ttttttttttgc 4680
gggttggatcatc ttttttttttgc ttttttttttgc 4740
gcttttttttgc ttttttttttgc ttttttttttgc 4800
caatcttgc ttttttttttgc ttttttttttgc 4860
gacaaatcttgc ttttttttttgc ttttttttttgc 4920
tttttttgc ttttttttttgc ttttttttttgc 4980
gttgcataatggatcatc ttttttttttgc ttttttttttgc 5040

-continued

tttataaaacg gtgaagaaaat atttgtcata aagacaatta ttgggtctga agatgggta 5100
ggtgtcaat gtctacgtgg atctgggttta attgtgggtg caacgtcaag ggcttaccac 5160
gatatttcctt ccatcacccctt agtcaactgt agatccgtcg gtatccgtgc ttatgggtt 5220
cggttgggtc aaagagctat tcagggtcga ggcgcggcaa ttatgtttaac tgggtctct 5280
gcaatcaaca aaatgtggg tagaagaat tataacttta acttacaatt ggggtgtact 5340
caaatacatgt ataacaacgg tggttcacat ttgactgttgg acgtgggtta 5400
gagaagattt ttgaatggat gtcttatgtt ccagcoaagg gtaatatgcc agttccatc 5460
ttggaaacta aagacacatg ggatagacca ttgttgcattca ctccaaactaa tgatggaaact 5520
tacatgttata gatggatgtt gatgggtcgt gagactgttgg gatggatttg atatgggtt 5580
tttgataaag ggtctttctt tgaaacttgg tcaggatggg ccaagggtgt tgcgttgg 5640
agagccgcgtc ttgggtgtat tccactgggt gttattgggtt ttgaacaagg aactgtcgag 5700
aacttgatttct ctgcgtatcc agctaattcca aatagtgttgg aaacattaaat tcaaaacgt 5760
ggtcaagttt ggcataccaaat ctcgccttc aagaactgtc aagctatcaa tgacttttaac 5820
aacgtgtaaat aatttgcataat gatgttggt gccaacttgg gatgggttcc ttgggttca 5880
cgtgtatgt tcaacgaaat cttgaagtat ggttgcgttta ttgttgcgcg attgggttggat 5940
tacaaaaacaaat caattttat ctatatccca cctaccgggtg aactaaggagg tggtcattgg 6000
tttgggtgtcc atccaaactat caacgcgtac caaatggaaa ttgttgcggca cgtcaacgc 6060
agagttgtgtt ttttggaaacc acaaggatgtt gttggatatac atgttccgtag aaaaaaaaaatgg 6120
ctggacaccaat tgaacagatggatgacaaatgacatcatacagaaat tgagatctca attatccaa 6180
aagagtttgg ctccagaagat acatcagcaat atatccaaagg aatttagctga tcgtgagaga 6240
gaactattgc caatttacggca aacaatcgtt ctcaattttt ctgtatggca cgtatgggtt 6300
tcacgtatgg tggccaagggtt tggttattttt aaggaactgg aatggaccggca ggcacgtcg 6360
tttttttttttggagatgttgg aagaaggatgg aacaaggaaat attgttattttt aaggttgcgt 6420
catcaggtagt ggcgaaggatcc aagatttagaa aagatcgcggaa gaatttagatc ttgggttcc 6480
gttttcgtgg accatgaaatgatggcaat gtcgcacat gatggatggaa aaaaactacaaa 6540
acttttggacg ataaaactaaatgggttggaaat ttagatgttgc ttcgttcaaga cttatgtt 6600
aagatcagaaat gcgccatcatc acaatgttcatc gatggattat ctgtatggatc caagatgtt 6660
tctaccgtatc ataaagaaaaat attgttggaaatgacttggaaat aa 6702

```

SEQ ID NO: 138 moltype = DNA length = 1632
FEATURE Location/Qualifiers
source 1..1632
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 138
atggaaatgtt ctacttttc ttttgggtt gtttgtaaaa ttattttttt tttttttct 60
ttaatatttc aaacttctat tgctaattcc agagaaaaatt tttgaaatg ttttctcaa 120
tatattccaa ataatgtcac taatttggaaa ttgggttata ctcaaaataa tcattgtat 180
atgtctgttt tgaattctac tattcataat ttgagattta ctctgtatac tactccaaa 240
ccatgtgtta ttgttactcc atctcatgtt tctcatattc aaggactatc ttgtgtct 300
aaaaaaagtgtt gtttgcataat tagaactaga ttgggtgtc atgatgtcg aggtatgtc 360
tatatttctc aagtccattttt tttttttttt gattttggaaat atagatgatc tattttttttt 420
gatgttcattt ctc当地actgc ttgggttgaat gctgggtcta ctttgggtga agttttttt 480
ttgggttataat aaaaatgtt aattttgtct ttgggtgtcg ttatgttgc aacttgtttt 540
gtctgggtgtt attttgggtt tgggtgttgc ttgggttgc tttttttttt tttttttttt 600
gtctgatataat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 660
tctatgggtt aagatgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 720
atttgttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 780
aaaattttatgg aattttcatgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 840
aatatgtataat aagatgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 900
caaggtaaaa ataaaactgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 960
gattttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1020
tgttagacaat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1080
actgtatattttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1140
aaaattttatgg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1200
gaaaaattgtt atgaagaaga tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1260
attatgtatgg aatattttctat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1320
gaatttgtgtt atattttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1380
atttagaaata tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1440
ttgaatttataat gaggatgtgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1500
caaggatgtatgg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1560
aaaatttttttgg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1620
agacatagacat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1680

```

-continued

ttagggcggac	acttttctgg	cgggggttat	ggtgcttga	tgaggaaacta	tggtttggcc	600
gcccataata	taattgacgc	ccatcttgc	aacgtcgacg	ggaagggtct	ggaccgtaaa	660
tctatgggtt	aagatttatt	ctggggcgata	agagggtggcg	ggggagagaga	cttgggtat	720
atcgacgtt	gtaagattaa	ftgttagttgt	gtccccctcaa	aagcaacaat	tttttcagtgt	780
aaagaacaaac	tggaaatccca	cgggtttgtt	aaactgttta	taataatggca	gaaatattggc	840
tacaaatacag	ataaggattt	gatgttgaca	acacatttca	gaactagaaaa	tattactgac	900
aaccacggaa	agaacaagac	aaccgtccat	ggatattttt	gttctatTTT	cttagggcgaa	960
gttggattccac	tagtagactt	aatgaacaaag	tctttccccc	aatttgggaaat	aaaaaaaaacc	1020
gattggcaagg	aattattcttc	gatagataca	aaatatctt	actctggagt	cgttaatatt	1080
aatacggccca	actttaaagaa	ggaaatattat	ttagatcggt	ccgcaggtaa	aaagacagot	1140
ttttccataaa	aattggacta	cgtcaaaaaa	ttaatttctg	agacagccat	ggtaaaaata	1200
tttggaaaaat	tgtacgaaga	ggagggttagc	gtgggtatgt	atgttgttata	cccatacggt	1260
ggtttattatgg	atggaaatttc	tgagagcgct	attcccttc	cccatctgtgc	aggatataat	1320
tatgatattat	ggtacacagc	aacatgggaa	aaacaagagg	ataacgaaaa	gcataattat	1380
tgggtacgtt	gtgtgtacaa	ctttaacgaca	ccttaacgtgt	ccccaaatcc	aaagattagcgt	1440
tatttgaact	atagagactt	agatttagt	aaaacaaacc	ctgagcttcc	aaataattac	1500
acccaaggcca	ggattttgggg	tggaaaatac	ttccggaaaaaa	atttcaatag	attgtttaag	1560
gtttaaaaatca	aggccggatcc	aaacaatttt	tttagaaatg	agcagagatgt	tccggccccgt	1620
ctcccaadac	accat					1635

SEQ ID NO: 140 moltype = DNA length = 2160
FEATURE Location/Qualifiers
source 1..2160
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 140

atgggtaaaa	attataaaatc	tttggatct	gttgttgctt	ctgatTTtat	tgctttgggt	60
attacttcgt	aagtgtcgta	aaccttgcat	ggtagatTTgg	ctgaaaatTTg	ttgtttaattt	12
ggtgtcgta	ctcccaaaac	tttggatata	attgttcaatc	atattttgtc	tccagatttg	18
ccatTTtctt	tgcatcaat	gttgtttat	ggttgttata	aagatTTgg	tccagctcca	24
ccagcttgg	ttccagatcc	agaaaaaagt	aaatctacta	atttgggtgc	tttggggaa	30
aaaaggagg	aagaatTTtt	gggtgtttaa	tataaagato	caatttcttc	tttttctcat	36
tttcaagaat	tttctgttg	aaatccaggaa	gtttatTTgg	gaactgttt	tgatggatga	42
atggaaattt	cttttttcaa	agatccaggaa	tgttatttga	gaagagatga	tattaataat	48
ccagggttgtt	ctgaatgggt	gccagggtgt	tatTTgatt	ctgtttaaaa	ttgttgaat	54
gttataatcc	ataaaaaattt	gaatgtatct	atgtatTTtt	ggagagatga	aggtaatgt	60
gatTTggcat	tgaatTTttt	gactttggat	caatttggaaa	aaagatTTt	tttgggttgtt	66
tatgttttgg	aagaatTTgg	tttggaaaaa	ggttgtgtca	ttgttattga	tatggcaatg	72
catgtttagt	ctgttggat	tttatttggct	atgttttgg	ctgggttatgt	tgttggTTCT	78
gtctatgttt	cttttttgc	ttccaggaaa	tctactatgt	tgatggatgc	taaagcttaa	84
gctatTTttt	ctcaagatca	tatttattaa	ggttttaaaa	gaattttccat	ttatTTctaga	90
gttggtaaag	ctaaatccc	aatggotatt	gttattttccat	gttctgggtc	taatatttgg	96
gtctgaatgt	gagatggta	tatttttctgg	gattttttt	tggaaagagc	taaagaatTT	102
aaaatttgg	aattttactgc	tagagaacaa	ccaggTTgtat	cttataactaa	tattttttgtt	108
tttcttggta	ctactggta	accaaaaatg	atccatggaa	ctcaagatca	tccattttggaa	114
gctgtgtgtt	atgggtggc	tcatttggat	attagaaaa	gtgatgttt	tgttggcoca	120
actaatttgg	gttggatgtat	gggtccatgg	ttggTTtATG	cttctttgtt	gaatgggtgt	126
tctatTTgtt	tgttataatgg	tttccatTT	gttTGTtgggt	ttgtttaattt	tgttcaatgt	132
gtctaaatgtt	ctatgttgg	ttgttggca	tctatTTgtt	gatTTggaa	atcttcaat	138
tgtgtttctt	gtttagatgt	gtctactatt	agatTTTTT	cttcttctgg	tgaagttct	144
aatgtttagt	aatTTttgt	gtttaggtt	agagctaa	ataaaccaat	tattgaaatg	150
tgtgggtgg	ctgaaatTTgg	ttgtgttt	tctgtgttt	cttttttggaa	agotcaatct	156
tgttctttt	tttttttcca	atgtatTTgtt	tgtacttTgt	atatttttgg	taaaaaatgtt	162
tatccaatgc	caaaaaatTT	accaggatTT	ggtgaatTTg	cttTgggtoc	atgtatTTtt	168
ggtgcttcta	aaactttttt	gaatggtaat	catcatgt	tttattttaa	aggtatgtcca	174
actttgtat	gtgaagtTT	gagaagacat	gggtatTTt	ttgttggatcc	tttcaatgtt	180
tattatcatg	ctcatggat	agctgtatgt	actatggata	ttttttgggtat	taaaaaatttct	186
tctatttggaa	ttgaaaagat	ttgttaatggaa	gttggatata	gagTTTTGA	aactactgt	192
atgggtgtt	caccatttgg	ttgttggtcc	gaacaatttgg	tttatttttt	ttgtttggaa	198
gattcttaat	atactactat	ttgttggat	caatttggat	tgttctttaa	tttgggtttt	204
caaaaaatTT	tgaatTTccat	ttttttggaa	actagatTTt	ttccatTTgtc	ttttttggcc	210
agaactgtca	ctaataaaat	ttgttggat	ttttttggat	ttttttggat	ttttttggat	216

-continued

```

gaaattaatc cattgggttgc tactaaagat ggtgtttatg ttttggattt ggctgctaaa 660
gttgatgcta ctgctgatata tatttgcataa gttaaatggg gtgatattga atttccacca 720
ccatttggta gagaagctta tccagaagaa gcttatattt ctgatttggta tgctaaatct 780
ggtgttcttt tgaaattgac ttgttgcataa cccaaaggta gaatttggac tatggttct 840
ggtgtgtgtt ctctgttgc ttatctgtat actatatttgc atttgggtgg tgtaatgaa 900
ttggctaaattt atggtaataa ttctgttgc ccatctgaaac aacaaactta tgattatgt 960
aaaacttattt tgcattttgc gacttagagaa aaacatccatg atggtaaaaat ttgttattt 1020
ggtggttctta ttgtctaaat tacaatgtt gctgtactt ttaaaggat ttttagagct 1080
attagagattt atcaaggatc ttgttgcataa catgaaggta ctatttttgt tagaagagg 1140
ggtcacaaattt atcaaggatc ttgttgcataa catgaaggta ctatttttgt tagaagagg 1200
ccaaatctatg ttgttgcataa tgaaatctatgactgtta ttgttgcata tttgttgcata 1260
catagaccaa ttccaaatca accaccaact gctgtcataa ctgctaaat tttgttgcata 1320
gttctgttgc ctacttctca tccagctca tctagaactt ctcttttgc tgaatctaga 1380
gtctgtatggat ttgttgcataa taaaaggat aacccagta tgccacaaga ttgttgcata 1440
tttccaaatg ttgttgcataa taaaaggat aacccagta tgccacaaga ttgttgcata 1500
gtttgggtta tgcaaaacttag agctgttca ggtatgttgc attttgcata tttgttgcata 1560
agagatgaaatc catctgttgc tgctatgttgc tttccatata ctgggtatca taaacaaaa 1620
ttttatttggg gtcataaaga aattttgttgc ccaatgttgc taaatatggc tgatgtatg 1680
agaaaaacatc cagaatgttgc tggttgcataa aattttgttgc ccaatgttgc ttttgcata 1740
tctactatgg aaactatgaa ttatgtctaa attagaacta ttgttgcata ttttgcata 1800
attccagaag ctgttgcataa aaaaattgttgc aaaaaggatc atcaaaaaagg ttgttgcata 1860
attgttgcataa ctacttgcataa ttgttgcataa ccagggttgc ttttgcata ttttgcata 1920
ggtattgttgc ataatatgttgc ggttgcataa ttgttgcata cagggttgc ttttgcata 1980
tctagatctg tggttgcataa taatgttgcata aataatattt ttttgcata ttttgcata 2040
gtttatgaaatc ttgttgcataa ttgttgcata agatatccat gtttgcata ttttgcata 2100
gttttgcataa ttgttgcataa ttgttgcata aaaaattgttgc ttgttgcata ttttgcata 2160
ggtaactgttgc aataaaaaat ttgttgcata aaaaaggatc ttttgcata ttttgcata 2220
gttttgcataa ttgttgcata aaaaaggatc ttttgcata ttttgcata ttttgcata 2280
gtctgtgttgc ttgttgcata aaaaaggatc ttttgcata ttttgcata ttttgcata 2340
gaaggctgttgc ttgttgcata aaaaaggatc ttttgcata ttttgcata ttttgcata 2400
ttatgaaatc ttgttgcata ttgttgcata gtttgcata ttttgcata ttttgcata 2460
gttccaaatggg atatatttgc ttgttgcata aaaaaggatc ttttgcata ttttgcata 2520
atgacttctta ttgttgcata aaaaaggatc ttttgcata ttttgcata ttttgcata 2580
gaaggatggaaatc aaaaaggatc ttgttgcata aaaaaggatc ttttgcata ttttgcata 2640
agatttgcata aatatttgcata ttttgcata ttttgcata ttttgcata ttttgcata 2700
ggtccagctg ttgttgcata ttttgcata ttttgcata ttttgcata ttttgcata 2760
gttttgcataa ttgttgcata ttgttgcata ttttgcata ttttgcata ttttgcata 2820
gtctgtgttgc aataatgttgc ttttgcata ttttgcata ttttgcata ttttgcata 2880
aataaaaaatc aaaaaggatc ttttgcata ttttgcata ttttgcata ttttgcata 2940
aataatccatc atatgttgcata ttttgcata ttttgcata ttttgcata ttttgcata 3000
actccattgttgc ttgttgcata ttttgcata ttttgcata ttttgcata ttttgcata 3060
ttgttgcata aatatttgcata ttgttgcata ttttgcata ttttgcata ttttgcata 3120
ggtttgcataa ttgttgcata ttttgcata ttttgcata ttttgcata ttttgcata 3180
ttttttttttt gtagatcttgc ttttgcata ttttgcata ttttgcata ttttgcata 3240
caagggttgc atagacatcc atgggtatgc ttttgcata ttttgcata ttttgcata 3300
atgaaatgttgc ttgttgcata ttttgcata ttttgcata ttttgcata ttttgcata 3360
gaaggctgttgc atatatttgcata ttgttgcata ttttgcata ttttgcata ttttgcata 3420
gaatattatc aaaaaggatc ttttgcata ttttgcata ttttgcata ttttgcata 3480
ggtaatgttgc attataaaatc aaaaaggatc ttttgcata ttttgcata ttttgcata 3540
tttgcata aatatttgcata ttttgcata ttttgcata ttttgcata ttttgcata 3600
caagggttgc ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 3660
aatttgcata ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 3720
caaaaaggatc ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 3780
ttgttgcata ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 3840
tatcaaaaatc ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 3900
gcttatttttgcata aaaaaggatc ttttgcata ttttgcata ttttgcata ttttgcata 3960
tttgcata ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 4020
ggtgtgtgttgc ttgttgcata ttttgcata ttttgcata ttttgcata ttttgcata 4080
ttgttgcata aaaaaggatc ttttgcata ttttgcata ttttgcata ttttgcata 4140
tttttttttgcata aatatttgcata ttttgcata ttttgcata ttttgcata ttttgcata 4200
gcttatttttgcata aaaaaggatc ttttgcata ttttgcata ttttgcata ttttgcata 4260
tctgtgtgttgc ttgttgcata ttttgcata ttttgcata ttttgcata ttttgcata 4320
aaaggctgttgc ttgttgcata ttttgcata ttttgcata ttttgcata ttttgcata 4380
catcatcatc atatttgcata ttttgcata ttttgcata ttttgcata ttttgcata 4440
agaaaatgttgc attttgcata ttttgcata ttttgcata ttttgcata ttttgcata 4500
agagatgaaatc aaaaaggatc ttttgcata ttttgcata ttttgcata ttttgcata 4560
gaaaatgttgc ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 4620
gaaaatgttgc ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 4680
gaaaatgttgc ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 4740
gaaaatgttgc ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 4800
gaaaatgttgc ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 4860
gaaaatgttgc ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 4920
gaaaatgttgc ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 4980
gaaaatgttgc ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 5040
gaaaatgttgc ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 5100
gaaaatgttgc ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 5160
gaaaatgttgc ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 5220
gaaaatgttgc ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 5280
gaaaatgttgc ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 5340
gaaaatgttgc ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 5400
gaaaatgttgc ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 5460
gaaaatgttgc ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 5520
gaaaatgttgc ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 5580
gaaaatgttgc ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 5640
gaaaatgttgc ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 5700
gaaaatgttgc ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 5760
gaaaatgttgc ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 5820
gaaaatgttgc ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 5880
gaaaatgttgc ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 5940
gaaaatgttgc ttttgcata ttttgcata ttttgcata ttttgcata ttttgcata 6000

```

SEQ ID NO: 142 moltype = DNA length = 2463
 FEATURE Location/Qualifiers
 source 1..2463
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 142

```

atgaaaaattt ttgttgcata aaaaaggatc ttgggttctt taatgggtct 60
ttggcttctta ttgttgcata tgatttgggt gctactgtta taaaaggatc tatttgcata 120
gctaaaaattt atcttcataa ttgttgcata gttattatgg gtaatgtttt gcaaggctgtt 180
ttgggtcaaa atccagcttgc acaaggcttgc ttgttgcata aactgtttt 240
ggtttactgc ttttgcata ttttgcata ttttgcata ttttgcata 300

```

-continued

```

gctattcaag ctggtaaagg tcaatctatt gttgctgggt gtagggaaa tatgtcttg 360
gctccatatt ttttggatgc taaagctaga tctgggtata gatgggtga tggtaaagg 420
tatgtatgtt ttttggagaga tgggttggatc tggtaacttc atggatatac tatgggtatt 480
actgctgaaa atgttgcataa agaatatggt attactagag aaatgcaaga tgaattggct 540
ttgcatttc aaagaaaagc tgctgctgtt atggatctg tgctgtttac tgctgaaatt 600
gttccagtttta atgttgcataa tagaaaaaaa acgtttgtt tttctcaaga tgaatttcca 660
aaagctaattt ctactgctga agctttgggt gctttagac cagctttga taaagctgt 720
actgttactg ctggtaatgc ttctggatt aatgtatggt ctgtctgtt gggttattatg 780
gaagaatctg ctgggttgc tgctgggtt actccattgg ctagaattaa atctttagtgc 840
tctgggtgtt tttccaccagg tttggatggt atgggtccaa tggttgcata tc当地aaagct 900
ttgcaattttt ctgggttgc atggctgtt atggattgtt tttggatggt tgaagcttt 960
gctgctcaat ttttggctgt tggtaaaaat ttgggtttt attctgaaa agttaatgtt 1020
atgttgcataa atggatcttgc ggttgcataa atggatctg ctgggtctg aattttgtt 1080
actttgttgc atgtatgtca agttagatgaaa acgtttggt gtttgcata ttttgttatt 1140
gggtgggttgc aaggtattgc tttggattt gaaaaggatgtt ataaatttgc tgggtgggt 1200
gggtctgggt gttgggttgc tgggtgggtt ggttctgtt aagctggta taatttgggt 1260
atgttgcataa ataaacaagg ttttggatggt aatattatca aatattatca aatattatca 1320
gaatttggatc caaataatgc tttggatggt ttttggatggt ttttggatggt 1380
gggtgatttgc aaaaatgttca ttttggatggt ttttggatggt ttttggatggt 1440
gctgaagctt ggtataattt gggtaatgtt ttttggatggt ttttggatggt ttttggatggt 1500
atgttgcataa atggatcttgc ttttggatggt gatccaaatc ataaatttgc ttttggatggt 1560
aaaatatttgc ttttggatggt ttttggatggt ggttgcata aatattatca ttttggatggt 1620
caaaaatgttca ttttggatggt ttttggatggt ttttggatggt ttttggatggt 1680
tatttataac aaggttgcataa ttttggatggt ttttggatggt ttttggatggt 1740
gatccaaatc atgttgcataa ttttggatggt ttttggatggt ttttggatggt 1800
tatcaaaaatc attttgcataa ttttggatggt ttttggatggt ttttggatggt 1860
agatctgttgc ttttggatggt ttttggatggt ttttggatggt ttttggatggt 1920
tatttatacatc attttgcataa ttttggatggt ttttggatggt ttttggatggt 1980
gggttctggatc aatgttgcataa ttttggatggt ttttggatggt ttttggatggt 2040
tatttggatc aaaaatgttca ttttggatggt ttttggatggt ttttggatggt 2100
gcttcataaaatc aaggttctgc ttttggatggt ttttggatggt ttttggatggt 2160
gctgctgtca aaaaatgttca ttttggatggt ttttggatggt ttttggatggt 2220
gaatttgggtt ttttggatggt ttttggatggt ttttggatggt ttttggatggt 2280
aaaaaaatgttca ttttggatggt ttttggatggt ttttggatggt ttttggatggt 2340
tttggatggatc aaaaatgttca ttttggatggt ttttggatggt ttttggatggt 2400
tttggatggatc aaaaatgttca ttttggatggt ttttggatggt ttttggatggt 2460
gaaatgttca ttttggatggt ttttggatggt ttttggatggt ttttggatggt 2463

```

SEQ ID NO: 143 moltype = DNA length = 2328

FEATURE Location/Qualifiers

source 1..2328

mol_type = other DNA

organism = synthetic construct

SEQUENCE: 143

```

atggaaaaaaat ttttggatggt ttttggatggt ttttggatggt ttttggatggt 60
gtgtgttgc aatgttgcataa ttttggatggt ttttggatggt ttttggatggt 120
tttggatggatc ttttggatggt ttttggatggt ttttggatggt ttttggatggt 180
actaaatgttgc ttttggatggt ttttggatggt ttttggatggt ttttggatggt 240
tggatggatc ttttggatggt ttttggatggt ttttggatggt ttttggatggt 300
gatggatggatc ttttggatggt ttttggatggt ttttggatggt ttttggatggt 360
attactgttgc ttttggatggt ttttggatggt ttttggatggt ttttggatggt 420
atccatccaaatc ttttggatggt ttttggatggt ttttggatggt ttttggatggt 480
acttttgcataa ttttggatggt ttttggatggt ttttggatggt ttttggatggt 540
gaagcttgcataa ttttggatggt ttttggatggt ttttggatggt ttttggatggt 600
atttttgcataa ttttggatggt ttttggatggt ttttggatggt ttttggatggt 660
atccatccaaatc ttttggatggt ttttggatggt ttttggatggt ttttggatggt 720
attatggatc ttttggatggt ttttggatggt ttttggatggt ttttggatggt 780
aaaaaaatgttca ttttggatggt ttttggatggt ttttggatggt ttttggatggt 840
tcttggatggatc ttttggatggt ttttggatggt ttttggatggt ttttggatggt 900
gctgttgcataa ttttggatggt ttttggatggt ttttggatggt ttttggatggt 960
attgttgcataa ttttggatggt ttttggatggt ttttggatggt ttttggatggt 1020
tttggatggatc ttttggatggt ttttggatggt ttttggatggt ttttggatggt 1080
gcttggatggatc ttttggatggt ttttggatggt ttttggatggt ttttggatggt 1140
aaacaatgttca ttttggatggt ttttggatggt ttttggatggt ttttggatggt 1200
aataatttgc ttttggatggt ttttggatggt ttttggatggt ttttggatggt 1260
tatcaaaaatc ttttggatggt ttttggatggt ttttggatggt ttttggatggt 1320
gcttggatggatc ttttggatggt ttttggatggt ttttggatggt ttttggatggt 1380
tatttatacatc ttttggatggt ttttggatggt ttttggatggt ttttggatggt 1440
atgttgcataa ttttggatggt ttttggatggt ttttggatggt ttttggatggt 1500
gaatttggatc ttttggatggt ttttggatggt ttttggatggt ttttggatggt 1560
tcttggatggatc ttttggatggt ttttggatggt ttttggatggt ttttggatggt 1620
gggttctggatc ttttggatggt ttttggatggt ttttggatggt ttttggatggt 1680
tggatggatc ttttggatggt ttttggatggt ttttggatggt ttttggatggt 1740
catatttgcataa ttttggatggt ttttggatggt ttttggatggt ttttggatggt 1800
gggttctggatc ttttggatggt ttttggatggt ttttggatggt ttttggatggt 1860
gggttctggatc ttttggatggt ttttggatggt ttttggatggt ttttggatggt 1920
aaaatgttca ttttggatggt ttttggatggt ttttggatggt ttttggatggt 1980
acttttgcataa ttttggatggt ttttggatggt ttttggatggt ttttggatggt 2040
agagaatgttca ttttggatggt ttttggatggt ttttggatggt ttttggatggt 2100

```

-continued

```
attccaat ttgttggtt gttggatact tatccatata atccacaat ttgttttgtt 2160
aaaccaact cttctatgac tattaaaact ggtaaacatg ttgatgctaa tggtaaaatt 2220
tatttgcct atttgcatttga atggaaacat ccacaatctg atttgggtt tttgattcaa 2280
gttatgatg ttgttttgtt tgatgaacca ccagttttt cttagacca 2328
```

SEQ ID NO: 144 moltype = DNA length = 2223

FEATURE Location/Qualifiers

source 1..2223

mol_type = other DNA

organism = synthetic construct

SEQUENCE: 144

```
atggaattga ataatgttat ttggaaaaaa gaaggtaaag ttgtgttgt tactattaat 60
agaccaaaaag ctttgaatgc ttgttattt gatacttga aagaatgaa ttatgtttt 120
gttggaaattt tttatgttgc ttgttgggtt gttgttattt tgactggc tggtgaaaaaa 180
tttttgttgc ttgggtctga tatttctgaa atgaaagaaa tgaatactat tgaaggtaga 240
aaatttggta ttttgggtaa taaagttttt agaagattgg aattgttgg aaaaccatgtt 300
attgtgtgtt ttaatgtttt tgcttgggtt gttgggttggt aaattgttgc ttcttggat 360
attagaatttgc ttcttcctaa tgcttagattt ggtcaaccagg aagttgggtt ggggttact 420
ccagggtttt gtggacttca aagatgttgc agatgggtt gtatgggtt ggctaaacaa 480
tttggattttca ttgtcaaaa tattaaagct gatgaagctt tgagaattgg tttggtaat 540
aaagtgttgc aaccatctg attgtatgtt actgtctaaag aaattgttca taaaattgtt 600
tctaattgtc cagttgtgtt taaatgttgc taaacaaatgtt taaatagagg tattgtcaatgt 660
gatattgtata ctgtttggc tttttgtatc gaagctttt tgtaatgtt ttctactgaa 720
gatcaaaaag atgtatgtac tgcttttattt gaaaaaaagaa aaattgttgc tttttttttttt 780
agaaaaattgtt ctgggtgttgg tgggttctgtt gttgggttggt ctgggtgttgg tgggttctgt 840
gaagctgtgtt ataaattttttt taaatgttgc tttttttttttt tttttttttttt tttttttttttt 900
gaatattatc aaaaatgttggttaa gaaatgttgc tttttttttttt tttttttttttt tttttttttttt 960
ggtatgtttt attataaaca aggttattttt caaaaatgttgc tttttttttttt tttttttttttt 1020
tttggatttttt accaaaaataa tgctgtttttt tttttttttttt tttttttttttt tttttttttttt 1080
caagggttatttcaatcaaaaatgttgc tttttttttttt tttttttttttt tttttttttttt tttttttttttt 1140
aattttgttgc ctgtttttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt 1200
caaaaatgttgc tttttttttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt 1260
ttgttataattt tttttttttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt 1320
tatcaaaaatgttgc tttttttttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt 1380
gctttttttttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt 1440
tttggatccaaataatgttgc tttttttttttt tttttttttttt tttttttttttt tttttttttttt 1500
gggtgggtgttgg tttttttttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt 1560
gctccaccaccaaaaatgttgc tttttttttttt tttttttttttt tttttttttttt tttttttttttt 1620
atgcattttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt 1680
tttggatccaaataatgttgc tttttttttttt tttttttttttt tttttttttttt tttttttttttt 1740
gttttgttgc tttttttttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt 1800
gctgggttgc tttttttttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt 1860
gggttctgtgttgc tttttttttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt 1920
ccaaactgttgc tttttttttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt 1980
agaccaaaaatgttgc tttttttttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt 2040
actatgttgc tttttttttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt 2100
gaaaaaaaacacatattttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt 2160
ccatctttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt 2220
gtt 2223
```

SEQ ID NO: 145 moltype = DNA length = 2460

FEATURE Location/Qualifiers

source 1..2460

mol_type = other DNA

organism = synthetic construct

SEQUENCE: 145

```
atgattgtta aaccaatgttgc tagaaataat atttgggttgc atgctcatcc acaagggttgc 60
aaaaaaagggttgc tttttttttttt tttttttttttt tttttttttttt tttttttttttt 120
gctgggtgttgc tttttttttttt tttttttttttt tttttttttttt tttttttttttt 180
gctttctttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt 240
aaagctgttttgc tttttttttttt tttttttttttt tttttttttttt tttttttttttt 300
gaagctgttttgc tttttttttttt tttttttttttt tttttttttttt tttttttttttt 360
gaaatattttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt 420
gtttttttttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt 480
gtttttttttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt 540
tttggatccaaataatgttgc tttttttttttt tttttttttttt tttttttttttt tttttttttttt 600
atgggtttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt 660
gaaggtttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt 720
agaaaaagggttgc tttttttttttt tttttttttttt tttttttttttt tttttttttttt 780
aaagaaaaatc catcttattttt tttttttttttt tttttttttttt tttttttttttt 840
tctgtttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt 900
aaaggtaatc atgaaatgttgc tttttttttttt tttttttttttt tttttttttttt 960
agaaaaagggttgc tttttttttttt tttttttttttt tttttttttttt tttttttttttt 1020
tttggatccaaataatgttgc tttttttttttt tttttttttttt tttttttttttt tttttttttttt 1080
gctgtttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt 1140
gatgtttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt 1200
ggtgggtttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt 1260
aattttttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt 1320
aaagctttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt 1380
```

-continued

tataaaacaag gtgattatca aaaagctatt gaatattatc aaaaagctt ggaattggat 1440
 ccaaataatg ctgaagctt gtataatttg ggtatgcattt attaaacaca aggtgattat 1500
 caaaaagcta ttgaagatta tcaaaaagct ttgaaattgg atccaaatcaa tttcaagct 1560
 gaagcttggaa aaaatttggg taatgcattt tataaacaag gtgattatca aaaagctatt 1620
 gaatattatc aaaagctttt ggaattggat ccaaataatg cttctgctt gtataatttg 1680
 ggtatgcattt attataaaca aggtgattt caaaagcta ttgaatattt tcaaaaagct 1740
 ttgaaattgg atccaaatcaa tgctaaagct tggatagaa gaggtaatgc ttattataaa 1800
 caagggtattt atcaaaaagc tattgaatg tatcaaaaag ctttggattt ggatccaaat 1860
 aatagatcta gatctgttgg tggatgttgg tctggatgttgg gtggatctgg tggatgttgg 1920
 gettcttattt attatcatca tcatcatc tatttggat ctacttctt gtataaaaaaa 1980
 gctggatcttgg ttttggaaaa agaaaaagaa aaattgagaa aaaaaaaa acaattttt 2040
 gaaaatcatg ttttggaaaa agaaaaagaa aaattgagaa aaaaaaaa acaattttt 2100
 aataaaatggg aagcttataa aggttctgtt ggttctgtt ctggatctgg tgaattttt 2160
 ttctgttgg aattttgttc ttcttattt catcatcatc atcatattt ggaatctact 2220
 ggttctgttgg aattttgttc ttcttattt catcatcatc atcatattt ggaatctact 2280
 tctttgtata aaaaagctgg ttctgttctt aaaaagattt ctatttttag aaaaaaaattt 2340
 gctgcttgg aaaaagataa tgctaaattt gaaaaagata ttgctaaattt ggaaaaatgaa 2400
 attgaaagat tgattaaaga aattaaaact ttgaaaatg aagtgttcc tcatgaacaa 2460

SEQ ID NO: 146 moltype = DNA length = 2478
 FEATURE Location/Qualifiers
 source 1..2478
 mol_type = other DNA
 organism = synthetic construct
 SEQUENCE: 146
 atgactagag aagttgttgg tttttcttgg gttagaactg ctattgttac ttttggatgtt 60
 tctttggaaa atgttgcctt agctgaaattt ggttctttgg ttgttagaga agctttggct 120
 agagctcaag tttctgttgg tgatgttgg catgttggat ttgttagatgt tattcaaaact 180
 gaaccaaaag atatgttattt gggtagatgtt gctgttggat atggatgttgg tactatttaat 240
 gctccagttt tgactgtttaa tagattgtgtt ggttctgttgg tgcaagatgt tggttctgtt 300
 gctccaaactt ttttggatgttgg tgatgttggat gtttctgttgg tggttctgttgg tgaatctatg 360
 tcttagagctc catatgttggc tccagctgtt agatgggggtt cttagatggg tgatgttggat 420
 ttgttggatgttgg tgatgttggat gatccatttc atagaatcca tatggatgtt 480
 actgtcttgg aatgttgcataa agaaatgttcatg atttcttagat ctcaacaaga tgaagctgtt 540
 ttggatgttgc atagaagac ttctgttgcg attaaagctt gttatttttaa agatcaaaattt 600
 gttccagttt tttctaaagg tagaaaaggtt gatgttactt ttgatactga tgaacatgtt 660
 agacatgtatg ctactattgtt tgatgttactt aaatttggatc agttttttgt taaagaaaaat 720
 ggtactgttgc tgctgttggat tgcttctgtt ttgtatgttgc tgctgttgc tggttggatgtt 780
 atggaaaagatg ctgaagctgtt aagaagatgtt tgaaaccat tggatgttggat ggtttctt 840
 ggttcatgttgc tggttgcattt aaaaagctttt ggttattgttgc cagttccatc tactaaaattt 900
 gctttggaaa gagctgttggat gcaagtttgc gatgttggatg ttatttggatc taatgttggat 960
 ttgttgcgtt aagcttgcgtt tggttacttgc gctttggatgtt tggttgcgtt taaatgttgc 1020
 ccaaatgttgc tggttacttgc gttttttttt ccaattttggatc tttttttttt tttttttttt 1080
 gttttttttt tgcatgttgc ttgttacttgc gttttttttt tttttttttt tttttttttt 1140
 attggatgttgc tgcaaggttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc 1200
 ggttctgttgc tggttgcgtt tggttacttgc gttttttttt tttttttttt tttttttttt 1260
 aatgttgcattt ataaacaaaagg tgatttttttgc aaaaatgttgc ttgttacttgc ttgttacttgc 1320
 gaatttggatc caaaatgttgc tggttacttgc ttgttacttgc ttgttacttgc ttgttacttgc 1380
 ggttacttgc aaaaagcttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc 1440
 gctgttgcattt ggttacttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc 1500
 attggatgttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc 1560
 aaaaatgttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc 1620
 caaaaagcttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc 1680
 tttttttttt ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc 1740
 gatccaaatgttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc 1800
 tttttttttt ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc 1860
 agatctgttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc 1920
 gtttatgttgg aatataatgttgc gggaaaatgttgc ttgttacttgc ttgttacttgc ttgttacttgc 1980
 tctgttgcattt tgcaaaatgttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc 2040
 aatatttttttgc caccagggttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc 2100
 actggatgttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc 2160
 aaaaagatgttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc 2220
 tctgttgcattt attttttttttgc gtttttttttgc ttgttacttgc ttgttacttgc ttgttacttgc 2280
 agattttttttgc atcatccatc aatatttttttgc ttgttacttgc ttgttacttgc ttgttacttgc 2340
 aaagaacaatgttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc 2400
 actgttgcattt ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc 2460
 ttggatgttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc 2478

SEQ ID NO: 147 moltype = DNA length = 2550
 FEATURE Location/Qualifiers
 source 1..2550
 mol_type = other DNA
 organism = synthetic construct
 SEQUENCE: 147
 atgaaactctt caactaaatgttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc 60
 caacaacaat tacacaatac aaacttgcattt ggttacttgc ttgttacttgc ttgttacttgc ttgttacttgc 120
 gaacaaaaaaa ccagacatc aatgttgcgtt attaaaggatgttgc ttgttacttgc ttgttacttgc ttgttacttgc 180
 caatgttgcattt tgcaaaatgttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc 240
 attggatgttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc ttgttacttgc 300

-continued

tccctaactg ttttgtctaa gttgatcaag agttacaaca tcgacaccaa caaaatttgt 360
 agattagaag tcggtaactga aactctgatt gacaagtcca agtctgtcaa gtctgttgc 420
 atgcaattgt ttggtaaaaa cactgacgtc gaaggatting acacgttcaa tgccgttac 480
 ggtggtaacc aacgctgtt caactcttg aactggatting aatctaaccgc atggatgg 540
 agagacgcoa ttgttagttt cggtgatatt gccatctacg ataagggtgc cgcaagacca 600
 accgggtgtc cgggtactgt tgctatgtgg atcggctctg atgtccaat tgtagtttgac 660
 tctgttaagag ctcttacat ggaacacgcoa taegattttt acaaggcaga tttcaccacg 720
 gaataatcctt acgtcgatgg tcattttca ttaacttggt acgtcaaggc tcttgatcaa 780
 gtttacaaga gttatccaa gaaggcttca tcttaaagggt tggttagcga tcccgcttgg 840
 tcggatgtt tgaacgtttt gaaatatttc gactacaacg tttccatgt tccaaacctgt 900
 aaatttggtca caaaatcata cggtagatta ctatataacg atttccagacg caatcccaa 960
 ttgttcccg aagttgacgc cgaatttagct actcgcgatt atgacgaatc tttaccggat 1020
 aagaacatgg aaaaaactt tttttatgtt gtaaggccat tccacaaga gagagtggcc 1080
 caatctttaa ttgttccaaac aacacagggt aacatgtaca cccgatctgt ttatggcc 1140
 ttgcatttc tattaaacta tggttgatct gacgacttac aaggcaaggc tggtgttta 1200
 ttttcttacg gtccgggtt agtgcattct ctatattctt gcaaaaattgt tggtgacgtc 1260
 caaacatatta tcaaggaaat agatattact aacaaatttag ccaagagaat cacggaaact 1320
 ccaaggattt acgaagctgc catcgattt agagaaaatgg cccatttggaa gaagaacttc 1380
 aaaaacctaaatggttccatgca gcatggca aatcttggat ttttttttttccatcgat 1440
 gacaaaatttaa gaagatctt cgtatggtaaa aaataaaaat ttttttttttccatcgat 1500
 ggtgggtgtt gtttgggtgg tggtgttgc gtttgggttgc ggttataattt ggtaatgtgt 1560
 tattataaac aagggtgttca tcaaggaaat atttggatattt atcaaaaaggc tttggaaattt 1620
 gatccaaatata atgctgaaat ttgtttaat ttttttttttccatcgat 1680
 tatcaaaaatgg ctattgtata ttatcaaaaaa gtttggat ttttttttttccatcgat 1740
 gtttggatata ttgttggtaa ttgtttaat ttttttttttccatcgat 1800
 gattatcaaa aaggctttggaa atttggatcca aataatttgc aagctgaaatc ttggaaaattt 1860
 ttggtaatgt ttatttttttccatcgat 1920
 gtttggat ttttttttttccatcgat 1980
 aacaaagggtt atttcaaaa aagcttggaa ttttttttttccatcgat 2040
 aataatgtca aaggctttggaa ttttttttttccatcgat 2100
 aaaggcttggtgg aaggatattca aaaaaggcttggaa gtttggatcca aataatatttgc atcttagatct 2160
 gtttggatgtt gtttggatgtt ttttttttttccatcgat 2220
 ttggccaccatg gtttggatgtt ttttttttttccatcgat 2280
 aataatagaaatctacttcaatt tactgtatcca agatgtcatg gtttggatgtt gtttggatgtt 2340
 gtttggatgtt aatttggatgtt ttttttttttccatcgat 2400
 ttttttttttccatcgat 2460
 ccagggttggg aaaaaaagaac tgattctaat gtttggatgtt ttttttttttccatcgat 2520
 aagaattacttcaatggttccatcgat 2550

SEQ ID NO: 148 moltype = DNA length = 2478
 FEATURE Location/Qualifiers
 source 1..2478
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 148
 atgggtggcc tacgttaggaa ggcttttca attttggcaag aagctccgtt atttgcattt 60
 gatcgtttac catataaaaaa ttatgactac gaccgcgtat ttggcgttgc ttgtaaaaat 120
 gttttaggtt acatgcctt gcccgggtt gttttaggcctt ctttgggttat cgtatgttaca 180
 tctttatcata taccatggc aactacagag gtttgggtttgg tagtctctgc catcggtggc 240
 tgaaggccaa tcaatgttgc cgggtgttgc acaactgttt taactaaaggc ttgtatgaca 300
 agaggccatg tagtccgtt cccaaactttt aaaaaggatctg ttgtatgttgc gatatggta 360
 gactcagaag aggacaaaaa cgcattttttaa aaaaaggcttta actcttacatc aagatttgc 420
 cgtctgcaac atattcaaac ttgtttagca ggagattttac ttttcatgag atttagaaaca 480
 actactgttgc acgcataatgg ttttttttttccatcgat 540
 caaatgtttagt aagagtatgg ctggaaatgg atggagggttgc ttttccatcgat 600
 ttttccatcgat 660
 ttttccatcgat 720
 ttggatgtt gtttggatgtt ttttccatcgat 780
 ggatggatgtt gtttggatgtt ttttccatcgat 840
 ctttccatcgat 900
 ttggatgtt gtttggatgtt ttttccatcgat 960
 cttagaaccac aaggcttgcattt gtttggatgtt ttttccatcgat 1020
 ctttccatcgat 1080
 ttatcccttgc ttttccatcgat 1140
 agggaaattgtt gtttggatgtt ttttccatcgat 1200
 gtttggatgtt gtttggatgtt ttttccatcgat 1260
 gtttggatgtt gtttggatgtt ttttccatcgat 1320
 gtttggatgtt gtttggatgtt ttttccatcgat 1380
 ttggatgtt gtttggatgtt ttttccatcgat 1440
 caaggttgcattt gtttggatgtt ttttccatcgat 1500
 aatttgcattt gtttggatgtt ttttccatcgat 1560
 caaaaaggctt gtttggatgtt ttttccatcgat 1620
 ttttccatcgat 1680
 ttttccatcgat 1740
 gtttggatgtt ttttccatcgat 1800
 ttggatgtt gtttggatgtt ttttccatcgat 1860
 gtttggatgtt gtttggatgtt ttttccatcgat 1920
 ttggatgtt gtttggatgtt ttttccatcgat 1980
 aaaaaggctt gtttggatgtt ttttccatcgat 2040

-continued

SEQ ID NO: 149 moltype = DNA length = 3141
FEATURE Location/Qualifiers
source 1..3141
mol_type = other DNA
organism = synthetic constru

SEQUENCE : 149

```
SEQ ID NO: 150          moltype = DNA  length = 2604
FEATURE                  Location/Qualifiers
source                   1..2604
                         mol_type = other DNA
                         organism = synthetic constru
```

-continued

tcgataggcg gatctaagaa ccctttcatt	aaaaaagtta tcgctaactgt atttagtac	300
ttaaaccta acatggacga ctactgcaat	agaaaacttgt tcgttattga tatttcct	360
gatgatgc accattctca ggaggatgc	gttaccgaaac atcgccaa cagaaggat	420
agtttcatt cgcacagaat tgaagaagtt	ccaaaacacg ggctggctc ctgcggcagg	480
ttagtacag tttaactac agcttggcc	tcctttttg tatcgacact ggaaaataat	540
gtagacaat atagagaagt tattcataat	ttagcacaag ttgetcatg tcaagctcag	600
ggtaaaatg gaagcgggtt tgatgttagcg	gcgcgcagcat atggatctat cagatataga	660
agattcccc accgcattaat ctctaat	ccagatattg gaagtgcac ttacggcagt	720
aaactggcgc atttgggtga tgaagaagac	tggaatattt cgattaaaag taaccat	780
ccttcggat taactttatg gatggcgat	attaagaatg ttcccgaaac agtaaaactg	840
gttcagaaagg taaaaattt gtatgatcg	catatgcgcg aaagcttgg aataatataca	900
gaactcgatc atgcaattc tagatttat	gatggactat ctaaactaga tcgcttacac	960
gagactcatg accgttacag cgatcagat	tttgagtttc ttgagggaa tgactgtac	1020
tgtaaaatg atctcgatc cacagaatg	agagatgcac ttgcacaaat tagacgttcc	1080
tttagaaaaa taactaaaga atctggc	gatatcgac ctcccgatca aactagctt	1140
ttggatgatt gccagaccc aaaaggagtt	cttacttgcg taatacctgg tgctgggt	1200
atgacgcoca ttgcgtat tactaaggaa	gatgtgtac tttagggcata aaccgctaat	1260
gacaaaatgg ttctaaagggt tcaatggctg	gatgtactc aggctgactg ggggtttag	1320
aaagaaaaaa atccggaaac ttatctgtat	aaataaaaat tgcgttggg tggtgttct	1380
ggtgtgtgtt gttctgggt	ttgtgttctc gctgaagctt ggtataattt	1440
tattataaac aagggtgatc tcaaaaaggt	ttgtgttctc gatcaatattt atcaaaaagc	1500
gatccaaata atgctgaaac ttgggtataat	ttgggtatg ctttattataa acaagggtgat	1560
tatcaaaaag ctatgtatc ttatcaaaaa	gcttggaaat tggatccaaat taatgtcgaa	1620
gcttggataa atttggtaa tgcttattat	aaacaagggtt attatcaaaa agtatttgc	1680
gattatacaa aagcttggaa atggatcca	aataatttgc aagctgacg ttggaaaaat	1740
tttgggtatc ttatcaaaa acaagggtat	ttatcaaaaatg ctatgtatc ttatcaaaa	1800
gcttggaaat tggatccaaat taatgtcg	ttttgggtaa tgcttattat	1860
aaacaagggtt attatcaaaa agtatttgc	tattatcaaa aagcttggaa attggatcca	1920
aataatgtca aagcttggta tagaaagggt	aatgttattt ataaacaagg tgattatcaa	1980
aaaggatgtt aagattatca aaaaggcttgc	aaaagggttgc gaaataatag atcttagatc	2040
gcttgggtgtt gttttctgg	ttgtgttctg tctgggttgc ttcttattat	2100
catcatcatc atcatcattt ggaatctact	tctttgtata aaaaagctgg ttctgggtct	2160
caaaaagggtt aagaattgaa aaataaaaat	gctgaatttgcg aaaaatagaaa tgctgttaaa	2220
aaaaatagatg ttgtctgtc gaaaaaggaa	attgttattt tgaaaggatg atttggctgt	2280
catgatattt aagggtctgc ttgttctgt	ttgtgttctg gtgaatttgg ttctgtcgaa	2340
gctgctgtca aagaagctgc tgctaaagct	ttttgggtctg gttctgtc tggttctgtt	2400
gaatttgggtt ttcttttata tcatcatcat	catcatcattt tggaatctac ttctttgtat	2460
aaaaaaagggtt gttctgttca ttttggaaat	ttttgggtttt gttactcatg aatttatttt	2520
gaaaatggaaa atgctaaattt gagaaggat	ttttgggtttt ggtacttgc aatggatcttgc	2580
ttgagaaatg aagttgttgc ttgttctgtt	ttttgggtttt ggtacttgc aatggatcttgc	2640

SEQ ID NO: 151 moltype = DNA length = 3027
 FEATURE Location/Qualifiers
 source 1..3027
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 151

atgaccgtt acacagcatc	cgttaccgca cccgtcaaca tcgcaaccct	taagtattgg	60			
ggggaaaaggc acacgaagtt	gaatctgccc accaattcg	ccatatcg	120			
caagatgacc tccaaacgtt	gaccctctgg	gctactgtc	180			
ttgtgtgtttaa atggagaacc	acacagcatc	gacaatggaa	gaaactcaaa	ttgtgtcg	240	
gacactccggc aattaagaaa	ggaaatggaa	tcgaaggacg	cctcattggc	cacattatct	300	
caatggaaac tccacattgt	ctccggaaat	aactttctca	cagcagctgg	tttagcttcc	360	
tccgctgtcg gttttgtctc	atttgc	ttttgtctc	gcaatttgc	agttatcacca	attaccacag	420
tcaacttcag aaatattctag	aatagcaaga	aagggttgc	tttgcgttgc	tagatcg	480	
tttggcggat acgttggctgc	ggaaatggaa	aaaggcttgc	tttgcgttgc	tttgcgttgc	540	
gtacaatcg cagacagctc	tgacttgc	ttttgtgtct	ttttgtgtct	ttttgtgtct	600	
gtatattaaa aggatgttgc	tttgcgttgc	ttttgtgtct	ttttgtgtct	ttttgtgtct	660	
ctatattaaatg aagaaggatgttgc	ttttgtgtct	ttttgtgtct	ttttgtgtct	ttttgtgtct	720	
attttgttgc aagatgttgc	ttttgtgtct	ttttgtgtct	ttttgtgtct	ttttgtgtct	780	
catgccacat gtttggactc	ttttgtgtct	ttttgtgtct	ttttgtgtct	ttttgtgtct	840	
atcatcgatgttgc	ttttgtgtct	ttttgtgtct	ttttgtgtct	ttttgtgtct	900	
ttttgtgtct gttttgtgtct	ttttgtgtct	ttttgtgtct	ttttgtgtct	ttttgtgtct	960	
gtcatttatctt ataaattttgttgc	ttttgtgtct	ttttgtgtct	ttttgtgtct	ttttgtgtct	1020	
cagcttgcggg ctttcaacca	ttttgtgtct	ttttgtgtct	ttttgtgtct	ttttgtgtct	1080	
tttgagtttgc aaaaaggatgttgc	ttttgtgtct	ttttgtgtct	ttttgtgtct	ttttgtgtct	1140	
aaaaacaaacg aatcttgcgttgc	ttttgtgtct	ttttgtgtct	ttttgtgtct	ttttgtgtct	1200	
gggtgtgtgtt gttttgtgtct	ttttgtgtct	ttttgtgtct	ttttgtgtct	ttttgtgtct	1260	
attttgtgtgtt atgcttgcgttgc	ttttgtgtct	ttttgtgtct	ttttgtgtct	ttttgtgtct	1320	
aaagcttggg aatttggatcc	ttttgtgtct	ttttgtgtct	ttttgtgtct	ttttgtgtct	1380	
tataaacaacg gtgattatca	aaaaggcttgc	ttttgtgtct	ttttgtgtct	ttttgtgtct	1440	
ccaaataatcg ctggaaatgttgc	ttttgtgtct	ttttgtgtct	ttttgtgtct	ttttgtgtct	1500	
caaaaaggatgttgc ttggaaatgttgc	ttttgtgtct	ttttgtgtct	ttttgtgtct	ttttgtgtct	1560	
gaagcttggaa aaaaattttgttgc	ttttgtgtct	ttttgtgtct	ttttgtgtct	ttttgtgtct	1620	
gaatattatc aaaaaggatgttgc	ttttgtgtct	ttttgtgtct	ttttgtgtct	ttttgtgtct	1680	
ggtaatgttgc atttataacatcg	ttttgtgtct	ttttgtgtct	ttttgtgtct	ttttgtgtct	1740	
tttggaaatgg atccaaatcg tggtaatgttgc	ttttgtgtct	ttttgtgtct	ttttgtgtct	ttttgtgtct	1800	
caagggttgc atccaaatcg ttttggaaatgttgc	ttttgtgtct	ttttgtgtct	ttttgtgtct	ttttgtgtct	1860	
aatagatgttgc gatctgttgc ttgtgtgttgc	ttttgtgtct	ttttgtgtct	ttttgtgtct	ttttgtgtct	1920	

-continued

gcttcgtcata	tggctgatTTT	ggaaacaaaa	gtttggaaa	tggaaGCTTC	tacttatgat	1980
gggttttttta	tttggaaaat	ttctgatTTT	ccaagaaaaa	gacaagaAGC	tgttgctgg	2040
agaatcccaag	ctatTTTTTC	tccagCTTT	tataCTCTA	gatatGGTTA	taaaatGTG	2100
tttggaaattt	atTTGAATGG	tgtatGGTACT	ggtagAGGTA	ctcatTTGTC	tttttttttt	2160
tttggatTTG	aaggTCCTAA	tgtatGCTT	tttagatGGC	cattaATCA	aaaatGTTACT	2220
tttggatTTG	ttggatCAAAA	taatAGAGAA	catgttATTG	atgttttag	accagatGTT	2280
acttCTTCTT	cttttcaag	accagttaat	gatatGAATA	ttgcTTCTGG	ttgtCCATTG	2340
ttttgtcccaag	ttttcaaat	ggaaGCTAA	aattCTTATG	tttagatGAG	tgtatTTTTT	2400
attaaAGCTT	tttgtGATT	gactGGTTG	ggttcTGCTG	ttttGTGTC	ttggTTGTTG	2460
gaatttGGT	ctgtcGAAGC	tgctGCTAA	gaagctGCTG	ctaaAGCTGG	ttctGCTGGT	2520
tctgtGCTG	gttctGGTGA	atttggTTCT	gtttcttATA	aattGCAATC	ttctGATGGT	2580
gaaaatTTTTT	aaggTGTGAT	tgaaaATGCT	aaacaATCTG	ttaacttAA	aactatGTTG	2640
gaaggatTTGG	gtatGGATGA	tgaaaggTGT	gtatGATCCAG	ttccatTTGC	aaatGTTTA	2700
gtgtGCTATT	tgaaaAAAGT	tattcaATGAT	tgtactCTCAT	taaaatGATGA	ttccaccACCA	2760
ccagaAGATG	atggaaaATAA	agaaaaAAAGA	actgtatGATA	ttccAGTTTG	ggatCAAGAA	2820
tttttGAAAG	ttgtatCAAGG	tactTTGTTT	gaattGATT	tggctGCTAA	ttatttGGAT	2880
attaaAGGTT	tgttgatGATG	tactGTAA	actgttGCTA	atatGATTAA	aggtaaaaACT	2940
ccagaAGAGG	tttagaaaaAC	ttttaatATT	aaaaatGATT	ttactGAAGA	agaAGAAAGCT	3000
caagtatGAA	aaagaaaaATCA	atqgtGT				3027

SEQ ID NO: 152 moltype = DNA length = 2154
FEATURE Location/Qualifiers
source 1..2154
mol_type = other DNA
organism = synthetic construct

```
SEQ ID NO: 153          moltype = DNA    length = 2448
FEATURE
source                 Location/Qualifiers
                      1..2448
mol_type = other DNA
organism = synthetic construct
```

```

SEQUENCE: 153
atggcttcag aaaaggaaat aagaagagaaa agattcttgc acgttattccc aaagttagtt 60
gaagaatttg a cgcgtatgtt gtttagctt ggtatgccta aagaagcgtc cgatggtat 120
gctcaactt taaactacaa tactccagggt ggtaaatttg a atagagggtt gagttcgatgtt 180
gatactttagt ctatcttgc taacaaaaaa gtggaaacat taggtcaaga aagatacgaa 240
aaaggctcgatc tcttgggtt gtgttattgaa ttgttgcagg catactttt gggtggcgat 300
gacatgtatgg ataagtctat aacaagaaga ggtcaaccat gctggtacaa agttccagaa 360
gttggtgaaat tagccataaa tgatgtttt atgttggaa ccgcatacttataaaatgttgg 420
aagtccatcat tcgaaacacg aaaaactac atcgatattt ccgaaattttt ccacgaaatgtt 480
atcttccaaat cagaatgttgg tcaattgtatg gattgttataa ctgcaccttgc aatgttgcatt 540
gacttgttcaa atgtttccctt gaagaaacat tcaattcatgc tcaccccttgc aactgttgcatt 600

```

-continued

tactcccttc	atttgcgg	cgccttggt	atgtacgt	ctggattac	tgatgaaaa	660
gacttgaagc	aagcaaga	tgtttgata	ccttgggt	aatactcca	aatccaag	720
gactactag	actgttcc	tatccaa	caaatacgta	aaatccgt	agatattcc	780
gacaataagt	gcagttgg	tattaaca	gcttggat	tagcatgc	cgacaaga	840
aagacttgg	atgaaaact	cggtaaaa	gactcagg	ctgaagca	gtgttaaaaa	900
attttaatg	atttgaag	tgaacaatt	taccatga	acgaagaat	catcgctaa	960
gacttaaagg	caaaggat	tcaagg	gaatcaag	gtttttaa	cgacgtt	1020
acagcttc	tgaataag	tcaagag	tcaaaatg	aatttgcgg	ttgggtgtt	1080
tctgggtgt	ttgggtctt	ttgttgggt	tctgtcaag	cttggata	ttttggtaat	1140
gcttattata	aacaagg	ttatcaaaa	gttattga	attatcaaa	agcttggaa	1200
ttggatccaa	ataatgc	agcttgg	atttggta	atgtctt	taaacaagg	1260
gattatcaa	aagcttatt	attatcca	aaagottt	aatttggat	aaataatcg	1320
gaagcttgg	attatgg	taatgtt	tataaca	tgatattca	aaaatgtt	1380
gaagattatc	aaaaagg	ggaattgg	ccaaataatt	tgcaagct	agcttggaa	1440
atttggta	atgttatt	taaacaagg	gattatca	aagcttatt	gaatattca	1500
aaagcttgg	aatttggat	aaataatcg	tctgtgtt	ataatttgg	taatgtt	1560
tataacaag	gttgatt	aaagat	gaatattt	aaaagctt	ggaaatgg	1620
ccaaataatg	cttaaagct	gtataga	ggtaaatgtt	attatcaa	agggttatt	1680
aaaaaagcta	ttgaagat	tcaaaagct	ttggatatt	atccaataa	tagatctaga	1740
tctgtgtgt	ttgggtgtt	ttgttgggt	ggtttgggt	ttgggtgtt	ttctttgtt	1800
actatggaaa	aaggccat	ttgttatgg	tttatttg	attctgtt	atctttaacc	1860
ggtaattttt	tttagatct	tgtatcg	tcttcagg	aaatcttgg	tttgagagot	1920
caagataga	tttgttga	taatgg	tgtatgg	gttacaaaca	ttggatgtt	1980
gtttctgtca	tttagagctt	ttgtatgt	actaaatgtt	ttgttggta	tagagaagg	2040
tctgtgtgt	ctgtctgtt	tttgggtgaa	tttggatct	ctgaagct	tgcttaaaa	2100
gtgtctgtca	aaatgtgtt	tgtgtgtt	gtctgtgtt	cttggtaatt	ttgttctt	2160
tctgggtgtca	ttatatttac	ttgttgaat	aaaagat	ttgtgtccatt	gggttattact	2220
atttctgtt	ctgaagaacc	atttgc	atttattt	cttcttgc	taaagggtgt	2280
tctgggtgtt	gaactctgt	tatcatatt	ggtgtat	ttttggctat	taatttctt	2340
tctttgtt	gtaaaccatt	gtctgtt	atttattt	tgcaatggc	ttggtaaaact	2400
tttacttgc	aaataaaaaa	acaaactgt	gtcaaccag	cttcttct		2448

-continued

tgtctgtta	aatttggtaa	tgatgttcaa	cattttaaag	ttttgagaga	tgggtcttgt	2520
aaatattttt	tgtgggttgt	taaatttaat	tctttgaatg	aattggttga	ttatccataga	2580
tctatcttc	tttctagaaa	tcaacaatt	tttttggagag	atattgaaca	agttccacaa	2640
caaccaact						2649

SEQ ID NO: 155 moltype = DNA length = 1995
FEATURE Location/Qualifiers
source 1..1995
mol_type = other DNA
organism = synthetic construc

```
SEQ ID NO: 156          moltype = DNA    length = 2391
FEATURE
source           Location/Qualifiers
                 1..2391
mol_type = other DNA
organism = synthetic construct
```

-continued

tatcaaaaag	ctttggaaatt	ggatccaaat	aatgcttctg	cttggataaa	tttgggtaat	1680
gcttattata	aacaagggtga	ttatcaaaaa	gctattgaat	attatcaaa	agcttggaa	1740
tttggatccaa	ataatgctaa	agcttggat	agaagaggta	atgcttata	taaacaagggt	1800
gattatcaa	aagctattga	agattatcaa	aaagcttgg	aattggatcc	aaataataga	1860
tctagatctg	ctgggtgggg	tgggtctgtt	gggtgggtgg	ctgggtgggg	tggtgtctt	1920
gctgaatatg	ttagagcttt	gtttgatgtt	aatggtaat	atgagaaga	tttgcattt	1980
aaaaaaagggt	atatttttag	aattagagat	aaaccagaag	aacaatggtg	gaatgctgaa	2040
gattctgaag	gtaaaaggagg	tatgattcca	gttccatatg	ttgaaaata	tggttctgtct	2100
gggttctgtg	ctgggtctgg	tgaatttgg	tctgctgaag	ctgctgtaaa	agaagctgtct	2160
gctaaagggt	gttctgtgg	ttctgtctgt	gggttctgg	aatttgggtc	tttgattaaa	2220
catatgagag	ctgaaggctt	gtttgatttt	actggtaat	ctaaatttgg	attgtaattt	2280
aaagctgggt	atgttatttt	tttggtgtct	agaattaata	aagattgggt	ggaagggtact	2340
tttagaggtg	ctactggtat	tttttcattt	tcttttgtta	aaattttgaa	a	2391

SEQ ID NO: 157 moltype = DNA length = 8226
 FEATURE Location/Qualifiers
 source 1..8226
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 157

atgagcgaag	aaagcttatt	cgagtcttct	ccacagaaga	tggagtagca	aattacaaac	60
tactcagaaa	gacatacaga	acttccagg	catttcattt	gcctcaatac	agttagataaa	120
ctagaggagt	ccccgttaag	ggactttgtt	aagagtca	gtggcacac	ggtcataatcc	180
aagatcctga	tagcaaataa	ttgttatttcc	gccgtgaaag	aaatttagatc	cgtcagaaaa	240
tgggcatacg	aacgttcgg	cgtatcaca	accgttccaa	tctgtccat	ggccacccca	300
gaagatctgg	aggccaaacg	agaatataat	cgtatggcc	atcaatacat	tgaatgtcc	360
gggtgtacta	ataataaaca	ctacgctaa	gttagactga	tctgttagacat	cgccgaaaga	420
gcagacgtag	acgcgtatg	ggctggctgg	ggtcacgc	ccgagaatcc	actattgcct	480
aaaaaaattgt	cccagtctaa	gaggaaagtc	atctttat	ggcctccagg	taacggcat	540
aggttcttag	gtgtaaaaat	ctcttctact	atgttgcotc	aaatgtgtaa	agtccatgt	600
attccatgg	ctggtacccg	tgttgcaccc	gttcacgtt	acggaaaaac	cggtctgttc	660
tctgtcgacg	atgacatcta	tcaaaagggt	tgttgcactt	ctccgtaa	ttgtttacaa	720
aggccaaacg	gtattggttt	tcctgtcatg	attaaggcat	ccgaagggtgg	ttgtgttaaa	780
gtatcagac	aaatggaaacg	tgaagaaat	ttcactgcet	tataccacca	ggcggccaaac	840
aaaaattccgg	gtccccccat	tttcatcatg	aatgtggccg	gttagggcgc	tcacttggaa	900
gttcaactgc	tagcaatca	gtacggtaca	aatatttcc	tgttgcgtt	agactgttc	960
gttcagagac	gtcatcaaaa	aattatcgaa	gaagcaccag	ttacaattgc	caaggctgaa	1020
acatttcacg	agatggaaaa	ggctggccgtc	agactgggg	aactagtccg	ttatgtctct	1080
ggccgttaccg	ttggagatct	atattctcat	gtatgtggaa	aatttctactt	ttttagattt	1140
aacccaagat	tacaatgcg	gcacccaaca	acggaaatgg	tctccgggtt	taacttacat	1200
gcagctcaat	tacaatatcgc	tatgggtatc	cctatgcata	gaataagtga	cattagaact	1260
ttatatgtta	tgaaatcctca	ttctgcctca	gaatatcgat	tgcatttca	aactcaagat	1320
gccaccaaga	aacaaacaag	acatttcca	aaagggtatt	gtatgttgc	tcgtatcaca	1380
tcaagaatgc	aaacacatgg	attcaagcc	tcgggtgtt	ctttgcattt	actaaacttc	1440
cgttcttctt	ctaatagttt	gggttacttc	tccgtgggtt	acaatggtaa	tattcactcc	1500
tttccggact	ctcagttccgg	ccatattttt	gtttttgggt	aaaatagaca	agcttccagg	1560
aaacacatgg	tttgcgttcc	gaaggaaattt	tccatttaggg	gtgatttcag	aactactgt	1620
gaataacttgc	tcaaaactttt	ggaaactgaa	gatttgcagg	ataacactat	taccacccgt	1680
ttgttggacg	atttgattac	tcataaaaatg	accgttgc	agcctgtatcc	aactcttgc	1740
gttatttgcg	gtgcggctac	aaaggcttcc	tttagcatct	aagaaggccg	ccacaagat	1800
atcgaatctt	tacaaaagggg	acaatgttca	tctaaagacc	tacttgc	tatgttccct	1860
gttagattttt	tccatgaggg	taaaagatac	aaatgttcc	tagtcaat	cggtatgtac	1920
cgttacacat	tattttatca	tggttctaaa	tgtgatata	tactgtgtca	actatctgt	1980
gtgtgttccat	tgtatttgc	aggcggtaaa	tgcgcatacc	tctatttggaa	agaagaagtt	2040
gtctgtacaa	tgcgttccgt	tgcgttccgt	tactctat	tggaaatgtt	aaacgtatcc	2100
acccaggatgc	gtactccatc	ccctgtgaaa	ttgtttaat	tcttgcgtt	aaatgtgtaa	2160
cacattatca	agggccaaacc	atatgcagaa	attgaagtta	tggaaaatgca	aatgccttgc	2220
gtttctcaag	aaaatgttat	cggttccat	ttaaagcaac	ctgggttccat	cattgttgc	2280
gggtgtatata	tggcttccat	gacttgcatt	gatccat	aaatgttgc	cgcttacca	2340
tttgaaggta	ttgttgc	ttttgttct	ccatgttgc	aaggaaacca	acctgcctat	2400
aaattcaagt	cattatgtt	tactttggaa	aaatgttgc	agggttat	caaccaagtt	2460
attatgaacg	cttccttgca	acaatgttgc	gagggttttt	gaaatccaaa	actgccttac	2520
tcagaatgg	aactacacat	ctctgttca	cattcaat	tgcgttgc	tttgcgttgc	2580
caaatggaa	agttgttgc	acgttccat	acgtgttgc	ctgtttccc	agcttgcata	2640
ttaagttaat	tgattgtat	ggccgttgc	aatcctgtat	acaacccca	caaattgttgc	2700
ggccgttgc	ttggaaaccatt	ggccgttgc	gtctatgttgc	actcttacgg	tttgcgttgc	2760
catgaacatt	ctatatttgc	ccatattttt	gaagaat	acgttgcgttgc	tttgcgttgc	2820
aatgttccaa	atgttgc	ggaaaatatac	atttgcataat	tgcgttgc	aaacccat	2880
gatctgtat	aatgttgcgtt	aactgttttgc	gttgcatttgc	aaatgttgc	tttgcgttgc	2940
ctgtatcttgc	ctatcttgc	acattatca	ccattgttgc	atgttgc	ttaatgttgc	3000
gccatatttgc	ctactccat	acaacatatt	gttgcatttgc	aatcttgc	taccgttgc	3060
gtcgcttac	aagcaagaga	aatttgtt	caaggcgctt	tacccgttgc	caaggaaaga	3120
actgaacaaa	ttgaacat	ctttaatcc	tctgttgc	agggttgc	tggcttccat	3180
aatccaaacg	gtctgtat	agttgttgc	atcttgc	aaatgttgc	tttgcgttgc	3240
tttgcgttgc	atgtttact	tcaatttgc	accatgttgc	acccgttgc	tttgcgttgc	3300
gtctgtat	tttgcgttgc	gtctatatttgc	cggttgc	tttgcgttgc	tttgcgttgc	3360
cacgaagggt	tcacagttcc	atttgc	tggaaatttgc	aactacccat	agcttgcgttgc	3420
tccacccat	caactgttta	atctaaat	ttgtatgttgc	gggttgc	tttgcgttgc	3480
ttgttgc	ttgttgc	tttgcgttgc	ccgttgc	aaatgttgc	tttgcgttgc	3540

-continued

-continued

SEQ ID NO: 158 moltype = DNA length = 10860
 FEATURE Location/Qualifiers
 source ..1..10860 mol_type = other DNA
 organism = synthetic construct
 SEQUENCE: 158
 atgggttctg ctgggttctgc tgctggttc ggtgaatttg gttctgtgg ttctgtgtc 60
 gggtctgtt aattttgttc tgctggttc gctgtgttt ctggtaattt ttcttattat 120
 catcatcatc atcatcattt ggaatctact tctttgtata aaaaagctgg ttctgtgtc 180
 gctagaatatg cttatgttag aaaaaaaatt gctagattga aaaaagataa ttgcaattt 240
 gaaagatgg aacaaaatg ggaaaaaaatt atgttattt tgagatgtga aattgtctg 300
 ttggaaaatg aagtgtctc tcataacaat ggttctgtg gtgtctgtc tggttctgtt 360
 gaatttgcg aagctgtcg taaagaagct gctgttaaaag ctgggtctgc tggttctgt 420
 gctgggtctg gtgaattttc ttattatcat catcatcatc atcatttggaa atctacttct 480
 ttgtataaaa aagctgggtc ttgggtctc aatttgaaaat tgaatggaaa agatgttgt 540
 tctttggaaa atgaaaatg aactttggaa aaaaaaaatt tgatcataaaa agatgttatt 600
 gcttatttgg aaaaagaaat tgctaattt agaaaaaaaaa ttgagaagg ttctgtgtt 660
 tctgtgtctg gttctgtgtg atttggttt gctgaagctg ctgtctaaaga agtctgtgt 720
 aaagaagctg ctgtctaaaga agctgtgtc aaagctgtt ctgtgtgtt tggtctgt 780
 tctgtgtat ttttttttc ttattatcat catcatcatc atcatttggaa atctacttct 840
 ttgtataaaa aagctgggtc ttgggtctca aagttgtctg aattggaaaat tagatgttgt 900
 gttaaattga atagaaatga acaattggaa aataaaatgtt aagaatttggaa aaatagaaaat 960
 gcttatttgg aaaaatgttgg ggttacttgc gaaaatggaaat ttgttgcattt ggaaatgtt 1020
 gtttgcgaaat gttctgtgtt ttctgtgtc ggttctgtg aattttgtca agtctgtgt 1080
 aaagaagctg ctgtctaaagg ttgggtctgtt ggttctgtg tggttctgtg tgatattttt 1140
 tattatcatc atcatcatca tcatttggaa tctacttctt tttttttttt agctgggtt 1200
 ggttctaaatg aagttactac ttggaaaat gatgtgttt ttattttggaa tggaaatgtt 1260
 tattttggaaa aaaaatttgc tagatttggaa aaaaagaaaaat ctgttttgatg aataataggat 1320
 gtcataaaa aaggttctgc ttgggtctgt gtcgttctgtg tgatatttttgg ttctgtgtt 1380
 gtcgtctgtt aagaagctgc tgctaaagaa gtcgtctgtt aagaagctgc tgctaaagct 1440
 gtttgcgtgtc gtttgcgtgtc ttgggtctgtt gatattttgtt ctggaccacc aacttatttt 1500
 aatccaccaat catttggattt ttgtgttcaat catcatcatc ttgggtctgtc ttgttctgtt 1560
 gtcgttctgtg tgtaatttgc tgaatgttgc gtcataaaaatg ctgtctgttgc agtctgtgtt 1620
 getgggttctg ctgtgtgtt ttgggtatattt aattttttgtc aatctagacc agaaccact 1680
 gtcctccaaatg aagaatcttgc tagatttggtt ggttctgtgtc ttctgtgtc tggttctgtt 1740
 gaattttgtt ctgtgtgtt tggtgttcaat gaaatgtgtc ctggaccacc aacttattttt 1800
 aagatgtgtt ctgggttctgtt ttgtgtgtt tggtgtgtt tggtgtgtt tggtgtgtt 1860
 tctaaaggta ctgggttggaa tccaaatgtt aagtttggc aaaaatttgc tccaggtaat 1920
 ggttctgtgtc gtttgcgtgtc ttgggtctgtt gatatttgc aagtttgcgtc taaaaggatgtt 1980
 gtcgttcaatg ctgggttctgtc ttgggttgcgtt gtttgcgtt gtcgttgcgtt aatgttgcgtt 2040
 actacttttgc aacattttgtt gtcgttgcgtt gtcgttgcgtt aatgttgcgtt 2100
 ttcgtgtgtc gtttgcgtgtc ttgggttgcgtt gtcgttgcgtt ctggaccacc aacttattttt 2160
 aaagaagctg ctgtctaaaga agctgtgtt aagatgtgtt ctgtgtgttgc tggttctgtgtt 2220
 tctgtgtat ttgggttctt ttattatcatc catcatcatc atcatttggaa atctacttctt 2280
 ttgtataaaa aagctgggtc ttgggttcaat gaaatgtgtt ttgttgcgtt aaaaatgtt 2340
 gcttggaaaat aagataatgc taaaatttggaa aaaaatgttgc taaaatttggaa aaaaatgtt 2400
 gaaagattga taaaaggaaat taaaatttggt gaaaatgttgc ttgttctca tgaacaaggat 2460
 ttcgtgtgtt ctgtgtgttgc ttgtgtgtt gtttgcgtt aatgttgcgtt aagatgtgtt 2520
 gtcataaggat gtttgcgtgtc ttgtgtgtt gtcgttgcgtt aatgttgcgtt 2580
 catcatcatc atttggatc ttcattttttt gtttgcgtt gtttgcgtt aatgttgcgtt 2640
 ttgggtctact tgagatctac tggtgtgtt ttggaaaatg aaaaatcatgt ttggaaaatgg 2700
 gaaaatggaaa aatttggaaaat gaaaatggaaaat gtttgcgtt gtttgcgtt aatgttgcgtt 2760
 gtttgcgtgtc gtttgcgtgtc ttgtgtgtt gatattttgtt ctgtgtgttgc tggttctgtt 2820
 gaaatgtgtt ctggaccacc aatgttgcgtt gtttgcgtt gtttgcgtt aatgttgcgtt 2880
 ttcgtgtgtc gtttgcgtgtc ttgggttgcgtt ccagatcttgc tcaacatcc accaccacca 2940
 ccagggttata gatctcaatc tccatctcat gtttgcgtt gtttgcgtt gtttgcgtt 3000
 gaattttgtt ctgtgtgtt aatgttgcgtt gtttgcgtt aatgttgcgtt 3060
 gtttgcgtgtc gtttgcgtgtc ttgtgtgtt gtttgcgtt aatgttgcgtt 3120
 ttggaaaatgg gatgtatggg ttgtgtgtt gtttgcgtt aatgttgcgtt 3180
 gtcgttcaatg ctgtgtgttgc ttgtgtgtt gtttgcgtt aatgttgcgtt 3240
 aaaaatgttgcgtt ctgtgtgttgc ttgtgtgtt gtttgcgtt aatgttgcgtt 3300
 atttttttttgc caccaccattt gtttgcgtt gtttgcgtt aatgttgcgtt 3360
 ttcgtgtgtc gtttgcgtgtc ttgtgtgtt gtttgcgtt aatgttgcgtt 3420
 gtttgcgtgtc gtttgcgtgtc ttgtgtgtt gtttgcgtt aatgttgcgtt 3480
 ccaatgtgtt caccaccaatg atctttttttt gtttgcgtt gtttgcgtt aatgttgcgtt 3540
 ttgggtgtt gtttgcgtgtc ttgtgtgtt gtttgcgtt aatgttgcgtt 3600
 gtcataaggat ctgtgtgttgc ttgtgtgtt gtttgcgtt aatgttgcgtt 3660
 gtttgcgtgtc gtttgcgtgtc ttgtgtgtt gtttgcgtt aatgttgcgtt 3720
 gtttgcgtgtc gtttgcgtgtc ttgtgtgtt gtttgcgtt aatgttgcgtt 3780
 gtttgcgtgtc gtttgcgtgtc ttgtgtgtt gtttgcgtt aatgttgcgtt 3840
 ttgggtgtt gtttgcgtgtc ttgtgtgtt gtttgcgtt aatgttgcgtt 3900
 gtttgcgtgtc gtttgcgtgtc ttgtgtgtt gtttgcgtt aatgttgcgtt 3960
 gtttgcgtgtc gtttgcgtgtc ttgtgtgtt gtttgcgtt aatgttgcgtt 4020
 gtttgcgtgtc gtttgcgtgtc ttgtgtgtt gtttgcgtt aatgttgcgtt 4080
 atttttttttgc ataaaaatgg gtttgcgtgtc ttgtgtgtt gtttgcgtt aatgttgcgtt 4140
 atttttttttgc ttttttttttgc ttgtgtgtt gtttgcgtt aatgttgcgtt 4200
 gaaatggaaaat gtttgcgtgtc ttgtgtgtt gtttgcgtt aatgttgcgtt 4260
 caagggttgcgtt ctgggttgcgtt gtttgcgtt aatgttgcgtt 4320

-continued

gctgtctgtca aagctgggtc tgctgggttct gctgtctgtt ctgggtgaatt ttcttattat 4380
catcatcatc atcatcatcatt ggaattttact tctttgtata aaaaagctgg ttctgggttct 4440
aattttgttg ctactttgg atctactgtct gctgtttttgg aaaaatgaaaa tcatgttttg 4500
gaaaagaaa aagaaaaaaa gagaagaaaaaa aagaacaat tttgttaataa attggaaact 4560
tataaagggt ctgtctggtc tgctgtttgt tctgggtaaat ttgggtctgc tgaagctgt 4620
gctaaagaag ctgtctgtaa agaagctgtct gctaaagaag ctgtctgtaa agctgggttct 4680
gctgtttgtt ctgtctggtc tggtgtattt ggttctgtt tttgtgtatg tgctgtctgt 4740
tatgaaccac caccatcttaa taatggaaaaa gctttgggtt ctgtctggtc tgctgtctgt 4800
tctgggtat ttgtctgaagc tgctgtttaa gaagctgtct ttttgtttgg ttctgtctgtt 4860
tctgtctgtt gttctgggtg attttagagaa ttgtttgtatg atccatctta tttgttaatgtt 4920
caaattttgtt ataaaggatcg acaagggtct gctgtttgtt ctgtctgggtt tttgtgtatg 4980
ggttctgtgt aagtctgttc taagaagact gctgttaaag aagtctgttc taagaagact 5040
gtctctaaag ctgtctggtc tggttctgtt gctgtttgtt ctgtttttgg ttctttttttttt 5100
actaaatcta tgaattttga taatccagtt tatagaaaaaa ctactgaaga agaaagggtct 5160
gctgtttgtt ctgtctggtc tggtgtattt gctgtttaaag aagtctgttc tgctgtctgt 5220
aaaggctgtt ctgtctgtt ctgtgtttgtt tctgggtat tttatgtttt gccatctact 5280
ttggattgaaa ataaatttga tttgtatgtt gatccaaattt ggggttctgc ttgttctgtt 5340
gtctgtttgtt gtgtttttttt ttctgtgttt gctgtctgtt aagaagctgc tgctttttttt 5400
gtctctgtgtt aagaagctgc tgctgtttaaag gtttctgtt gtttctgttc tggttctgtt 5460
gaattttgtt ctgtgtttgtt taatttcttca ccacacgtt tttccacaaaaaaaagacaaa 5520
tctgtccat ctgtctggtc tggttctgtt gctgtttgtt ctgtttttttt tggttctgtt 5580
gctaaagaag ctgtctgtttt aagttttttt gctgtttgtt ctgtctgggtt tttgtgtatg 5640
acttcaagat ctaaacccaca accagctgtt ccaccaagac catctgtgtt tttgtttttt 5700
ggttctgtgtt gtttctgttc tggttctgtt gtttgggtt ctgtgtttttt tggttctgtt 5760
gaaagctgtgtt ctgtgtttgtt tggtgtttttt tttttttttttt tggttctgtt 5820
tctgtctgtt gtttctgtt atttttttt actgtatgtt aagagagaaga aactgtttttt 5880
gaagttttttt tttttttttttt tactactttt ggttctgtt gtttctgtt tggttctgtt 5940
gaattttgtt aagctgtgtt aagaagactt gctgtttaaag ctgtctgtt gtttctgtt 6000
gctgtttgtt tgtaattttt tggtgtttttt tttttttttttt tggttctgtt 6060
gttagaactt attctttgtt ttctgtttttt tttttttttttt tggttctgtt 6120
gtctgtgtttt ctgtctgtttt aatggatgtt aagagatgtt tggttctgtt 6180
aaaggctgtt ctgtctggtc tgctgtttgtt tttttttttt tggttctgtt 6240
catcatcatc atcatcttgg atctactttt tttttttttttt tggttctgtt 6300
aaaggctgtt aattttttttt tagactgttt tttttttttttt tggttctgtt 6360
aatattttgtt cttagattttttaa aatgtataat gcttaatttttgg aaaaagatat tgctttttt 6420
gaaaagatgtt ttgtctttttt tggttctgtt aatggatgtt tggttctgtt 6480
ggttctgtgtt aattttgtt aagctgtgtt aagagatgtt tggttctgtt 6540
ggttctgtgtt ctgtctgggtt tttttttttttt tggttctgtt 6600
tctactttttt tttttttttttt tggttctgtt tttttttttttt tggttctgtt 6660
tatattttttt aattttttttt aagctgtgtt tggttctgtt tttttttttttt tggttctgtt 6720
ttttgtctttt ctgtctgtt aattttttttt aatggatgtt tggttctgtt 6780
tctgtctgtt gtttctgtt aattttttttt tggttctgtt tttttttttttt tggttctgtt 6840
aaaggctgtt ctgtctgtttt aatggatgtt tggttctgtt tttttttttttt tggttctgtt 6900
tctgtgtttt ttgttctttt tggttctgtt tttttttttttt tggttctgtt 6960
tctgtctgtt aatctgtatgg ttctgtttttt tggttctgtt gtttctgtt tggttctgtt 7020
gtctctgtgtt aagaagctgtt tgctttttttt tggttctgtt tttttttttttt tggttctgtt 7080
gaattttttt ctccaaatgtt tggttctgtt gaaacaaactt caaaaaaaacc aggttttggaa 7140
agaaggggtt ctgtctggtc tgctgtttttt tggttctgtt tttttttttttt tggttctgtt 7200
gtctaaaggaa ctgtctgtttt aagagatgtt gctaaaggaa ctgtctgtttt tggttctgtt 7260
gtctgtttttt ctgtctggtc tggtgtttttt tggttctgtt tttttttttttt tggttctgtt 7320
cattttttt ctactttttt tgtaaaaaaa gctgtttttt tggttctgtt tttttttttttt tggttctgtt 7380
catgtttttt tttttttttttt tggttctgtt aatggatgtt aatggagaag attggaaact 7440
aaatggaaa gagaatttttgc tagattttttt gatgtttttt tggttctgtt tttttttttttt tggttctgtt 7500
tctgtctgtt gtttctgtt aattttttttt tggttctgtt gctgtttttt tggttctgtt 7560
ggttctgtgtt gtttctgtt tggttctgtt gtttctgtt tttttttttttt tggttctgtt 7620
cattttttt ctactttttt tgtaaaaaaa gctgtttttt tggttctgtt tttttttttttt tggttctgtt 7680
ttttttttt aattttttttt aatggatgtt aagaaatgtt tttttttttttt tggttctgtt 7740
cattttttt aagaaatgtt tgtaaaaaaa gatgtttttt tggttctgtt tttttttttttt tggttctgtt 7800
tctgtctgtt ctgtctgtttt tggttctgtt tttttttttttt tggttctgtt 7860
getgtctgtt aagaagctgtt tgctttttttt tggttctgtt aagctgtttt tggttctgtt 7920
getgtctgtt ctgtctgtttt tggttctgtt tttttttttttt tggttctgtt 7980
gattttttt tgatgtttttt tggttctgtt tttttttttttt tggttctgtt 8040
ttttttttt ctgtctgtttt aagagatgtt tgctttttttt tggttctgtt tttttttttttt tggttctgtt 8100
ggttctgtgtt aattttttttt tgctacttttca attttttttttt tggttctgtt tttttttttttt tggttctgtt 8160
gattttttt tggttctgtt tgctttttttt tggttctgtt tttttttttttt tggttctgtt 8220
gaagctgtgtt cttaaaaggaa ctgtctgtttt tggttctgtt tttttttttttt tggttctgtt 8280
gtctgtttttt ctgtctggtc tggttctgtt gtttctgtt tttttttttttt tggttctgtt 8340
catcatcatc attttttttt tgctttttttt tggttctgtt tttttttttttt tggttctgtt 8400
gtttttttt tgtaaaaaaa aattttttttt tggttctgtt tttttttttttt tggttctgtt 8460
gatattttttt tgtaaaaaaa aattttttttt tggttctgtt tttttttttttt tggttctgtt 8520
getgtctgtt tgtaaaaaaa aattttttttt tggttctgtt tttttttttttt tggttctgtt 8580
gtctgtttttt ctgtctggtc tggttctgtt tttttttttttt tggttctgtt 8640
gaatctttttt ctgtttttttt tggttctgtt tttttttttttt tggttctgtt 8700
atggctgtgtt cttaatgtt aatggatgtt agagaattttt tggttctgtt tttttttttttt tggttctgtt 8760
actgtatcat tttttttttt tggttctgtt tttttttttttt tggttctgtt 8820
tggatgtttt tggttctgtt tttttttttttt tggttctgtt 8880
gattttttt tggttctgtt tttttttttttt tggttctgtt 8940
ggttctgtgtt tggttctgtt tttttttttttt tggttctgtt 9000
aaaggctgtt ctgtctgtttt tggttctgtt tttttttttttt tggttctgtt 9060

-continued

tcttttcaaa	tgccagcgtga	tactccaccca	ccagcatttt	tgccaccaga	agatccaatg	9120
actgggttctg	ctgggttctgc	tgctgggttct	ggtaatttg	ctgaagctgc	tgctaaagaaa	9180
gctgtcgota	aaggctgggtc	tgctgggttct	gctgtcggtt	ctgggtgaatt	tgaaaagaaaa	9240
tctaatgaag	aaccacccac	accatatgaa	ggccatattt	gggttaatgg	tggttctgt	9300
gggttcgtgt	ctgggttctgg	tgatattttg	tctgtctga	ctgtctgttt	aaagactgtct	9360
gctaaagaag	ctgctgtcaa	agaagctgt	gctaaagctg	gttctgttgg	ttctgtctgt	9420
gggtctgggt	aatttggttc	ttcttattat	catcatatc	atcatcattt	ggaattactt	9480
tctttgtata	aaaagctgg	tgctgggttct	caaaaagttt	ctgaattgaa	aaatagagtt	9540
gctgttaat	tgaatagaaa	tgaacattt	aaaaataaaatg	ttgaagaattt	aaaaatagata	9600
aatgcttatt	tgaaaaatga	atggctact	ttggaaaatg	agttgttag	attggaaaat	9660
gtgttgtgt	aagggttctgc	tggttctgt	gtctgggttctg	gtgaatttgc	tgaatgtct	9720
gctaaagaag	ctgctgtcaa	agctgggttct	gtctgggttctg	ctgtgttgc	tggtgaaattt	9780
tcttattatc	atcatcatca	tcattttttt	gtactttttt	ctttgtataa	aaaagctgtgt	9840
tctgggttata	atgaagttac	tactttggaa	aatgtatgt	cttttattttt	aaatgaaaat	9900
gttttattttt	aaaagaaaat	tgcttagattt	agaaaagaaa	aaactgtcttt	gagaatata	9960
ttggctcatca	aaaatcttata	ttatcatcat	catcatatc	atttggaaat	tactttttttt	10020
tataaaaaaa	ctgggttctgg	ttctgtctaga	aatgtttttt	tgaaaaaaa	aatttgtatg	10080
ttgaaaaaaa	ataatttgc	atttggaaaga	gttggaaaaaa	atttggaaaa	aatttattgt	10140
aatttggag	atggaaatttgc	tagattggaa	aatggatgtt	cttctcatga	acaaggttct	10200
gtgttgttct	ctgtctgggtc	tggttgaattt	gtctggatgt	ctgtctaaaga	agctgtctgt	10260
aaagctgtgt	ctgtctgggtc	tgctgtttttt	tctggatgt	tttcttattt	tcatcatcat	10320
catcatattt	tggaaatctac	ttctttgtat	aaaaaaatgt	gttctgttcc	taatttggtt	10380
gctcaattgg	aaaatgttgt	tgcttttttgc	ggaaatgaaa	atgaaacttt	ggaaaaaaa	10440
aatttgcata	aaaagatgtt	gttgcattttt	ttggaaaaaa	aaatgtctaa	tttggaaaaaa	10500
aaaatgttgt	aaagggttctgc	tggttctgt	gtctgggttctg	tggttattttt	ttctgtctgt	10560
gtctgtgtca	aagaagctgc	tgctaaagaa	gtctgtgtca	aaagactgtgc	tgctaaagct	10620
ggttctgtgt	gttctgtctgc	tggttctgtt	gaatttggtt	ctgaacaaaa	attgtattttt	10680
gaagaagatt	tggaaacaaaa	attgtttttt	gaagaagattt	ttggaaacaaaa	attgtttttt	10740
gaagaagatt	tggttctgtc	tggttctgt	gtctgggttctg	tggttattttt	ttctgtctgt	10800
ttctgtctgt	tttctgtgt	atttttttttt	gtctgggttctg	ctgtctgttgc	ttggtaattttt	10860

```

SEQUENCE: 159
atgggttctg ctggttctgc tgctggttct ggtgaatttg gttctgctgg ttctgctgt 60
ggtgttctggg aattttgggtc tgctgggtt gctgtgggtt ctgggtgatt ttcttattat 120
catcatcatc atcatcattt ggaatctact tctttgtata aaaaagctgg ttctggttct 180
gcttagaaatg cttatgttag aaaaaaaattt gctagattga aaaaagataa ttgcattt 240
gaaagagatg aacaaaattt ggaaaaaattt attgtcaatt tgagatgatg aattgtctaga 300
ttggaaatgt aagtgttc tcatgaaacaa ggttctgctg gttctgctgc tgggtctgg 360
gaatttgcgtg aagctgtcgtc taaagaaggtt gctgtttttt ctgggttctgc tgggtctgt 420
gctggttctg gtgaattttt ttattatcat catcatcatc atcatttggg atctacttct 480
ttgtataaaa aagctgtgtc tggttctaat ttgggttctgc aatttgaaaaa tgaatgtgt 540
tctttggaaa atgaaaaatgt aacttggaaa aaaaaaaattt tgcataaaaa agattgtatt 600
gctttatggg aaaaagaaat tgcttaattt tgaaaaaaaat ttggaaagg ttctgtctgg 660
tctgtctgtc gttctggta atttgggtt gctgaagctg ctgtctaaaga agctgtctgt 720
aaagaagctc ctgctaaaga agctgtctgt aaagctgtttt ctgggttctgc tgggtctgtt 780
tctgtgtat ttgggttcttc tgctactaga gaatttgatgg aatttgatggc ttcttggat 840
gatttaaaaa ttcaagggtt ttctgtctgt tctgtctgtc ttgggttctgc ttcttggat 900
getgtctgtca aagaagctgc tgctaaagct gggttctgtc gttctgctgc tgggtctgg 960
gaatttggat tggctttgtc tgaaaattgg gctcaagaat ttgggtctgc tgggtatgtct 1020
gttggatgtt ctgctgtgtc tgctgtctgt tctgggtat ttgggttctgc tgaactgtct 1080
gctaaagaag ctgtgtctaa agaaagctgtc gctaaaagag ctgtgtctaa agctgtgttct 1140
gttgggttctg ctgtgtgtt ctgggtattt gggttctgtt ataaagatgt tgatgataaa 1200
gattataaag atgatgtat taaagatgtt aaagatgtatg atgataaagg ttctgtctgtt 1260
ttctgtctgtc gttctgtgtc atttgggtt gctgggttctgc ctgtgtgttctc tgggtatattt 1320
gggttctgtca gttctgtgtc tggttctgtt ataaatgtt 1356

```

SEQ ID NO: 160 moltype = AA length = 1532
FEATURE Location/Qualifiers
source 1..1532
mol_type = protein
organism = synthetic construct

SEQUENCE: 160		Organism = Synthetic construct	
MSAKAISEQT	GKELLYKFC	TTSQAIQNRFK YARVTPPTDW ARLLQDHPWL LSQNLVVVKPD	60
QLIKRRGKLG	LGVVNLTLDG	VKSWLKPRLQ QEATVGKATG FLKNFLIEPF VPNSQAEFFY	120
VCIYATREGD	YVLFLHHEGGV	DVGDVDAAQ KLLVGVDEKL NPEDIKKHNL VHAPEDKKEI	180
LASFISGLFN	FYEDLFYFTYL	EINPLVVTKD GVYVLDLAAK VDATADYICK VKWGDIEFPP	240
PFGREAYPEE	AYIAIDLAKS	GASLKLTLNN PKGRGIWTMVA GGGASVVSYD TICDGGVNE	300
LANYEGSYGA	PSEQQTYDYA	KTILSMLTRTE KHPDGKLLII GGSIANFTNV AATFKGIVRA	360
IRDYQGPPLKE	HEVTIFVRG	GPNYQEGLRV MGEVGKTTGI PIHVFGTETH MTAIVGMALG	420
HRPIPQNQPPT	AAHTANFLLN	ASGSTSTPAP SRTASFESR ADEVAPAKKA KPAMPQDSVP	480
SPRSLQCKST	TLFSRSRTHKA	VWGMQTRAVQ GMGLDFDVCS RDEPSVAAMV YPFTGDKHQK	540
FYWHGHKEILY	PVFKNMADAM	RKHPEVDVLN NFASLRSAJD STMETMNYAQ IRTIAIIAEG	600
IPEALTRKLII	KKADQKGVTI	IGPATVGGIK PGFKCIGNTG GMLDNILASK LYRPGSVAYV	660
SRSGGMSNEL	NNIISERTTDG	RYPGSTMDH VLRYODTPGV KMIVLLEGEIG	720

-continued

GTEEYKICRG IKEGRLTkpI	VCWCIGTCAT	MFSSEVQFgh	AGACANQASE	TAVAKNQALK	780
EAGVFVPRSF DELGEIIQSV	YEDLVANGVI	VPAQEVPPT	VPMDSWARE	LGLIRKPASF	840
MTSICDERGQ ELIYAGMPIT	EVFKEEGMG	GVLGLLWFQk	RLPKYSCQFI	EMCLMVTAHD	900
GPAVSGAHNT II CARAGKDL	VSSLTSGLLT	IGDRFGGALD	AAAKMFSKAF	DSGIIPMEFV	960
NKMKKEGKL MGIGHRVKS	NPNDPMDVQJL	KDYVRQHFP	TPLLDYALEV	EKITTSKPN	1020
LILNVVDGLIG VAFVDMRLNC	GSFTREADE	YIDIGALNGI	FVLGRSMGFI	GHYLDQKRLK	1080
QGLYRHPWDD ISYVLPFEHMS	MKLSGGGGSG	GGGGGGGSA	EAWYNLGNAY	YKQGDYQKAI	1140
EYYQKALELD PNNAEAWYNL	GNAYYKQGDY	QKAIEYYQKA	LELDPNNAEA	WYNLGNAYYK	1200
QGDYQKAIER YQKALELDPN	NLQAEAWKNL	GNAYYKQGDY	QKAIEYYQKA	LELDPNNASA	1260
WYNLGNAYYK QGDYQKAIY	YQKALELDPN	NAKAWYRRGN	AYYKQGDYQK	AIEDYQKALE	1320
LDPNNRSRSA CGGGGGGGGS	GGGGASSYYH	HHHHHLESTS	LYKKAGSGSN	LVAQLENNEVA	1380
SLENENETLK KKNLHKKDLI	AYLEKEIANL	RKKIEEGSAG	SAAGSGEFGS	AEAAAEEAAA	1440
KAGSAGSAAAG SGEFGSSYYH	HHHHHLESTS	LYKKAGSGSA	RNAYLRKKIA	RLKKDNLQLE	1500
RDEQNLEKII ANLRDEIARL	ENEVASHEQG	SG			1532

SEQ ID NO: 161 moltype = AA length = 824
 FEATURE Location/Qualifiers
 source 1..824
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 161
 MKNCVIVSVA RTAIGSFNGS LASTSAIDLG ATVKAAIER AKIDSQHVDE VIMGNVLQAG 60
 LGONPARQAL LKSGLAETVC GFTVNKVCGS GLKSVLAQAO AIQAGQAQSI VAGGMENMSL 120
 APYLLDAKAR SGYRLGDGVQ YDVILRDGLM CATHGYHMG1 TAENVAKEYG ITREMQDELA 180
 LHSQRKAAAA IESGAFTAEL VPVNVTTRKK TFVFSQDFEP KANSTAAEALG ALRPAFDKAG 240
 TVTAGNSAGI NDGAALAVIM EESAALAAAGL TPLARIKSYA SGGVPPALMG MGPVPATQKA 300
 LQLAGLQLAD IDLIEANEAF AAQFLAVGKLN LGPDSEKVN VNGGAIALGHP IGASGARILV 360
 TLLHAMQARD KTLGLATLCI GGGQGIAMVI ERLNKLSSGG GSGGGGSGGG GSAAEAWYNLG 420
 NAYYKQGDYQ KAIYEYQKAL ELDPNNAEAW YNLGNAYYKQ GDYQKAIIEYY QKALELDPNN 480
 AEAWYNLGNA YYKQGDYQKA IEDYQKALEL DPNNLQAEAW KNLGNAYYKQ GDYQKAIIEYY 540
 QKALELDPNN ASAWYNLGNA YYKQGDYQKA IEYQKALEL DPNNAKAWYR RGNAYYKQGD 600
 YQKAIEDYQK ALELDPNNRS RSAGGGGSGG GGSAGGGGASS YYHHHHHLE STSLYKKAGS 660
 GSNEVTTLEN DAAFIENENA YLEKEIARLR KEKAALRNRL AHKKGSAGSA AGSGEFGSAB 720
 AAKKEAAAKA GSAGSAAGSG EFGSSYYHHH HHHLLESTSLLY KKAGSGSQKV AELKNRVAVK 780
 LRNNEQLKNK VEELKNRNYA LKNELATLEN EVARLENDVA EGSS 824

SEQ ID NO: 162 moltype = AA length = 779
 FEATURE Location/Qualifiers
 source 1..779
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 162
 MKKCVIGAG TMGSGIAQAF AAKGFVEVRLR DIKDEFVDRG LDFINKNLSK LVKKGKIEEA 60
 TKVEILTRIS GTVDLNMAAD CDLVIEAAVE RMDIKKQIFA DLDNICKPET ILASNISTSLS 120
 ITEVASATKR PDKVIGHMHFF NPAPVMKLV IRVGIATSQE TFDAVKETSI AIGKDPVEVA 180
 EAPGFVVNR LIPMINEAVG ILAEGIASVE DIDKAMKLGA NHPMGPLELG DFIGLDICLA 240
 IMDVLYSETG DSKYRPHTLL YYKVVRAGWLQ RKSKGKGFYDNG LKGNAKKQGD YQKAIIEYYQK 300
 AEAWYNLGNA YYKQGDYQKA IEYQKALEL DPNNNAEAWYLN LGNAKKQGD YQKAIIEYYQK 360
 ALELDPNNAE AWYNLGNAYY KQGDYQKAIE DYQKALELDP NNQAEAWKN LGNAKKQGD 420
 YQKAIIEYYQK ALELDPNNAS AWYNLGNAYY KQGDYQKAIE YYQKALELDP NNQAEAWKN LGNAKKQGD 480
 NAYYKQGDYQ KAIEDYQKAL ELDPNNRSSRS AGGGGSGGGG SGGGASENL YFQGENLYFQ 540
 GDSSESCWNCG RKASETCSG CNTARYCGSF CQHKDWEXHH HICGQTLQAO QGSAGSAAGS 600
 GEFGSAAEAAA KEAAAKAGSA GSAAGSGEFG SMAVSESOLK KMVSKYKYRD LTVRETNVI 660
 TLYKDLKPVL DSYVFDNGSS RELMNLGTI PVPYRGNTYI IPICLWLLDT YPYNPICFV 720
 KPTSSMTIKT GKHDVANGKI YLPYLHEWKH PQSDLGLLIQ VMIVVFGDEP PVFSRPGSG 779

SEQ ID NO: 163 moltype = AA length = 744
 FEATURE Location/Qualifiers
 source 1..744
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 163
 MELNNVILEK EGKVAVVTIN RPKALNALNS DTLKEMDYVI GEIENDSEVL AVILTGAGEK 60
 SFVAGADISE MKEMNTIEGR KFGILGNKVF RRLELLEPKV IAAVNGFALG GGCEIAMSCD 120
 IRIASSNARF GQPEVGLGIT PGFGGTQRLS RLVGMGMAQK LIFTAQNIKA DEALRIGLVN 180
 KVVEPSELNMN TAKETANKIV SNAPAVAKLS KQAINRGMQC DIDTALAFES EAFGECFSTE 240
 DQKDAMTAFI EKRKIEGFKN RKLSGGGGSG GGGGGGGSSA EAWYNLGNAY YKQGDYQKAI 300
 EYYQKALELD PNNAEAWYNL GNAYYKQGDY QKAIEYYQKA LELEDPNNAEA WYNLGNAYYK 360
 QGDYQKAIER YQKALELDPN NLQAEAWKNL GNAYYKQGDY QKAIEYYQKA LELEDPNNASA 420
 WYNLGNAYYK QGDYQKAIY YQKALELDPN NAKAWYRRGN AYYKQGDYQK AIEDYQKALE 480
 LDPNNRSRSA CGGGGGGGGS GGGGASGPLG SPLTASMLAS APPQEOKQML GERLPLIQA 540
 MHPTLAGKIT GMLLEIDNSE LLHMLESPES LRSKVDEAVA VLQAHQAKEA AQKAGSAGSA 600
 AGSGEFGSAAE AAAKEAAAKA GSAGSAAGSG EFGSNTNMVS PTDGAVTTSQ IPASEQETLV 660
 RPKPLLLKLL KSVGQAKDTY TMKEVLFYLG QYIMTKRLLY EKQQHIVYCS NDLLGDLFGV 720
 PSFSVKEHRK IYTMIYRNVL VGSG 744

SEQ ID NO: 164 moltype = AA length = 823
 FEATURE Location/Qualifiers
 source 1..823

-continued

mol_type = protein
 organism = synthetic construct

SEQUENCE: 164
 MIVKPMVRNN ICLNAHPQGC KKGVEDQIEY TKKRITADEVK AGAKAPKNVL VLGSNSGYGL 60
 ASRITAAGFY GAATIGVSFS KAGSETKYGT PGWYNNLAFD EAAKREGLYS VTIDGDAFSD 120
 EIKAQVIEEA KKKGKIFDLI VYSLASPVRT DPDTGIMHKS VLKPFGKTFT GKTVDPTGE 180
 LKEISAEPEAN DEEAAATVKV MGGEDWERWI KQLSKEGLLE EGCITLAYSY IGPEATQALY 240
 RKGTIGKAKE HLEATAHRLN KENPSIRAFV SVNKGVLTRA SAVIPVPLY LASLFKVMKE 300
 KGNHEGCTEQ ITRLYAERLY RKDGTTIPVDE ENRIRIDDE LEEDVQKAVS ALMEKVTCGEN 360
 AESLTDLAGY RHDFLASNGF DVEGINYEAE VERFDRIKLS GGGGSGGGGS GGGGSAAEWY 420
 NLGNAYYKQDQY DYQKAIETYQKALEDPNNA EAWYNLGNAY YKQGDYQKAI EYYQKALED 480
 PNNAEAWYNL GNAYYKQGDY QKAIEDYQKA LEELDPNNLQA EAWKNLGNAY YKQGDYQKAI 540
 EYYQKALELD PNNAEAWYNL GNAYYKQGDY QKAIETYQKALEDPNNAKA WYRRGNAYYK 600
 QGDYQKAIED YQKALELPN NRSRSAAGGGG SGCCGCGGGC ASSYYHHHH HLESTSLYKK 660
 AGSGSNLLAT LRSTAALLEN ENHVLEKEKE KLRKEKEQOLL NKLEAYKGSA GSAAGSGEFG 720
 SAEAAAKEAA AKAGSAGSAA GSGEFGSSYY HHHHHHLEST SLYKKAGSGS KRIAYLRKKI 780
 AALKKDANAL EKDIANLENE IERLIKEIKT LENEVASHEQ GSG 823

SEQ ID NO: 165 moltype = AA length = 829
FEATURE Location/Qualifiers
source 1..829
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 165
 MTRVVVVVSG VRTAIGTFGG SLKDVAPPAEL GALVVREALA RAQVSGDDVG HVVFGNVIQT 60
 EPRDMYLGRV AAVNGGVTL APALTVNRLC GSGLQAIWSA AQTILLGDTD VAIGGGAESM 120
 SRAPYLTLPAA RWGARMGDAV LVDMLMLGALH DPFHRIHMVG TAENVAKEYD ISRAQDDEAA 180
 LESHRRASAA IKAGYFKDQI VPVVSKGRKG DVTFTDTEHV RHADATIDDMT KLRPVFKEN 240
 GTVTAGNASG LNDVAAAVNE MERAEAERRG LKPLARLVSY GHAGVDPKAM GIGPVPATKI 300
 ALERAGLQVS LDLDVIEANEFA FAAQACAVTK ALGLDPAKVN PNGSGISLGH PIGATGALIT 360
 VKALHELNRV QGRYALVTCMC IGGGGIAAI FERIKLSCGGG GSAGGGSAGGG GSAAEWYNLG 420
 NAYYKQGDYQ KAIEYYQKAL ELDPNNAEAW YNLGNAYYKQ GDYQKAIETYQKALEDPNNA 480
 AEAWYNLGNNA YYKQGDYQKA IEDYQKALEL DPNNLQAEAW KNLGNAAYYKQ GDYQKAIETYQ 540
 QKALELDPNN ASAOWNLGNNA YYKQGDYQKA IEYYQKALEL DPNNNAKAWYR RGNAYYKQGD 600
 YQKAIEDYQK ALELDPNRS RSAGGGGSGG GGSGGGGASD VMWEYKWEWT GDAELYGPFT 660
 SAQMOTWVSE GFYPDGVYCR KLDPPGGQFY NSKRIDFDLY TGSAGSAAGS GEFGSAEEAA 720
 KEAAAKGSA GSAAGSGEFG SESDSVEFNN AISYVNKIKT RFLDHPEIYR SFLEILHTYQ 780
 KBQLHTKGRP PRGMSEEEVF TEVANLFRQG EDLLSEFGQF LPEAKRGSG 829

SEQ ID NO: 166 moltype = AA length = 852
FEATURE Location/Qualifiers
source 1..852
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 166
 MLLSTKLCWC GIKGRLRPQK QQQLHNTNLQ MTELKKQKTA EQKTRPQNVG IKGIQIYIPT 60
 QCVNQSELEK PDGVSQGKYT IGLGQTNMSF VNDREDIYSM SLTVLSKLK SYNIDTNKIG 120
 RLEVGTETLI DKSCKVKSVL MQLFGENTDSV EGIDTLNACY GGTNALFNSL NWIESNAWDG 180
 RDAIVVCGDI AIYDKGAARP TGGAGTVAMW IGPDAPIVFD SVRASYMEHA YDFYKPDFTS 240
 EYPYVGDGHFS LTCYVVKALDY VYKSYSKKAI SKGLVSDPAG SDALNVLKTYF DYNVFHVPCT 300
 KLVTKSYGRL LYNDPRANPQ LFPEVDAELA TRDYDESITL KNIETKTFVNV AKPFHKERVA 360
 QSLIVPNTNG MMYTASVYAA FASLNNYVG5 DDLQGKRVGL FSYGSGLAAS LYSCKIVGVD 420
 QHIICKELDT NKLAKRITET PKDYEAAIEL RENAHLKNF KPQGSIEHLQ SGVYLYTNID 480
 DKPFRSYDVK KKLSSGGGGS GGGSGGGSA EAWYNLGNAY YKQGDYQKAI EYYQKALEL 540
 PNNAEAWYNL GNAYYKQGDY QKAIETYQKALEL DPNNNAEAW YNLGNAYYK QGDYQKAIED 600
 YQKALELDPN NLQAEAWKLN GNAYYKQGDY QKAIETYQKALEL DPNNNAEAW YNLGNAYYK 660
 QGDYQKAIETY YQKALELDPN NAKAWYRRGN AYYKQGDYQK AIEDYQKALEL DPNNRSRSA 720
 GGGGSGGGGS GGGGASLGPL PPGWEVRSTV SGRYFVDHN NRTTQFTDPR LHGSAGSAAG 780
 SGEFGSAEEAA AKEAAAKAGS AGSAGSGEFG GSGAMGPLPP GWEKRTDSNG RVYFVNHNTR 840
 ITQWEDPRSG SG 852

SEQ ID NO: 167 moltype = AA length = 829
FEATURE Location/Qualifiers
source 1..829
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 167
 MVAVRRKALS ILAEAPVLAS DRPLPKNYDY DRVFGACCEN VIGYMPPLPVG VIGPLVIDGT 60
 SYHIPMATTG CCLVVASAMRG CKAINAGGGT TTBLTKDGMT RGPVVRPTL KRSGACKIWL 120
 DSEEGQNAIK KAFNSTSRFA RLQHJQTCLA GDLFLMFRFT TTGDAMGMNM ISKGVEYSLK 180
 QMVEEYGWED MEVVSVSGNY CTDKKPAAIN WIEGRGKSVV AEATIPGDDV RKVLKSDVSA 240
 LVELNITAKNL VGSAMAGSVG GFNAHAANLV TAVFLALGQD PAQNVESSNC ITLMKEVDGD 300
 LRISVSMPSI EVGTIGGGTV LEPQGAML DL LGVRGPHATA PGTNARQLAR IVACAVLAGE 360
 LSLCAALAAAG HLVQSHMTHN RKLSCGGGGSG GGGGGGGSA EAWYNLGNAY YKQGDYQKAI 420
 EYYQKALELD PNNAEAWYNL GNAYYKQGDY QKAIETYQKALEL DPNNNAEAW YNLGNAYYK 480
 QGDYQKAIED YQKALELDPN NLQAEAWKLN GNAYYKQGDY QKAIETYQKALEL DPNNNAEAW 540
 WYNLGNAYYK QGDYQKAIETY YQKALELDPN NAKAWYRRGN AYYKQGDYQK AIEDYQKALEL 600
 LPNNRSRSA GGGGSGGGGS GGGGASSYYH HHHHHLESTS LYKKAGSEFF RRERNKMAAA 660
 KCRNRNRELT DTLQAETDQL EDEKSALQTE IANLLKEKEK LEFILAAHRP ACKIPDDLGF 720

-continued

PEEMSLEGSAGA GSAAGSGEFG SAEAAAKEAA AKAGSAGSAA GSGEFGSSYY HHHHHHLEST	780
SLYKKAGSGS QKVESLKQKI EELKQRKAQL KNDIANLEKE IAYAETGSG	829

SEQ ID NO: 168 moltype = AA length = 1049
 FEATURE Location/Qualifiers
 REGION 1..1049
 note = chimeric enzyme
 source 1..1049
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 168

MSLPFLTSAP GKVIIFGEHS AVYNKPAVAA SVSALRTYLL ISESSAPDTI ELDFPDISFN	60
HKWISINDFNA ITEQDVSQNSLQ LAKAQCATDQ LSQELVSLSD PLLAQLSESF HYHAAFCFLY	120
MFVCLCPHAK NIKFSLKSTL PIGAGLSSA SISVSLALAM AYLGLIGSN DLEKLSENDK	180
HIVNQWAFIG EKCIHGTPSC IDNAVATYGN ALLFEKDSHN GTINTNNFKL LDDFFPAIPMI	240
LTYTRIPRST KDLVARVRVL VTEKFPEVMK PILDAMGECA LQGLEIMTKL SKCKGTDDEA	300
VETNNELYEQ LLELIRINHG LLVSGIVSHP GLELIKNLSD DLRIGSTKLT GAGGGGCSLT	360
LLRRDITQEQ 1DSFKKLQD DFSYETPFD LGGTGCCLLS AKNLNDLKI KSLVFQLFEN	420
KTTTKQQQIDD LLLPGNTNLP WTSKLSSGGG SGGGGSSGGG SAEAWYNLGN AYYQGDYQK	480
AIEYYQKALE LDPNNAEAWY NLGNAYYKQG DYQKAIEYYQ KALELDPNNA EAWYNLGNAY	540
YKQGDYQKAI EDYQKALEPD PNNLQAEAWK NLGNAYYKQG DYQKAIEYYQ KALELDPNNA	600
SAWYNLGNAY EYYQKALEPD PNNAKQGDYR GNAYYKQGDY QKAIEDYQKA	660
LELDPNNRSR SAGGGGSSGG GSGGGGASME PAMEPETLEA RINRATNPLN KELDWASING	720
FCEQLNEDFE GPPLATRLLA HKIQSPQEW E AIQALTIVLET CMKSCGKRHF DEVGKFRFLN	780
ELIKVVSPKY LGSRITSEKVK NKILELLYSW TVGLPPEEVKI AEAYQMLKKQ GIVKSGSAGS	840
AAGSGEFGSA EAAAKEAAAK AGSAGSAAGS GEFGSGAMGS MAAEAGESLE SWLNKATNPS	900
NRQEDWEYII GFCDQINKEL EGPOIAVRLL AHKIQSPOEW EALQALTGLE ACMNCGRRF	960
HNEVGKFRFL NELIKVVSPK YLGDRVSEKV KTKVIELLYS WTMALPEEAK IKDAYHMLKR	1020
QGIVQSDPPI PVDRTLIPSP PPRPKNGSG	1049

SEQ ID NO: 169 moltype = AA length = 871
 FEATURE Location/Qualifiers
 source 1..871
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 169

MSELRAFSAP GKALLAGGYL VLDTKYEAJV VGLSARMHAV AHPYGSLQGS DKFEVRVSKS	60
QFKDGEWLH SIGPKSGFIPV SIGGSKNPFI EKVIANVFSY FKPNMDDYCN RNLFVIDIFS	120
DDAYHSQEDS VIEHGRNRL SFHSHRIEEV PKTGLGSSAG GLVTVLTTAL ASFFFVDSLLEN	180
NVDKYREVIH NLAQVAHCQA QGKIGSGFDV AAAAYGSIRY RRFPPALISN LPDIGSATYG	240
SKLAHLVDEE DWNITIKSNH LPSGLTLWNG DIKNGSETVK LVQVKVNWYD SHMPESLKIY	300
TELDHANSRF MDGLSKLDRD HETHDDYSDQ IFESLERNDTC TCQKYPEITE VRDAVATIRR	360
SFRKITKESG ADIEPPVQTS LDDDCQTLKG VLTCLIPGAG GDYDAIAVITK QDVDLRAQTA	420
NDKRFSKVQW LDVTQADWGV RKEKDPTELY DKKLSSGGGGS GGGGSSGGGS AEAWYNLGN	480
YYQGDYQKAI EYYQKALEPD DPNNNAEAWY LGNAYYKQGD YQKAIEYYQK ALELDPNNAE	540
AWYNLGNAYY KQGDYQKAI E DYQKAELDP NNLQAEAWK LGNAYYKQGD YQKAIEYYQK	600
ALELDPNNAS AWYNLGNAYY KQGDYQKAI E YYQKALEPD NNAKAWYRNG NAYYQGDYQ	660
KAIEDYQKAL ELDPPNRSRS AGGGGSSGGG SGGGGASSYY HHHHHHLEST SLYKKAGSGS	720
QKVEELNKI AELENRNAVK KNRVAHLKQE IAYLKDLEAA HEFEGSAGSA AGSGEFGSAAE	780
AAAKEAAAKA GSAGSAAGS EFGSSYYHHH HHHLESTSLY KKAGSGSFEN VTHEFILATL	840
ENENAKLRLR EAKLERELAR LRNEVAWLGS G	871

SEQ ID NO: 170 moltype = AA length = 1011
 FEATURE Location/Qualifiers
 source 1..1011
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 170

MTVYTASVTA PVNIATLKYW GKRDTKLNLNQ TNSSISVTLN QDDLRLTSA ATAPEFERDT	60
LWLNGEHPHI DNERTONCLR DLRLQLRKEME SKDASLPTLS QWKLHIVSEN NFPTAAGLAS	120
SAAGFAALVS AIAKLYQLPQ STSEISRIAR KGSGSACRSL FGGYVAWEMG KAEGDHDSMA	180
VQJADSDWP QMKACVLVVS DIKKDVSSTQ GMQLTVATSE LFKERIEHVV PKRFEVMRKA	240
IVEKDFATFA KETMMDNSF HATCLDSFPP IFYMNDTSKR IIISWCHTINQ FYGETIVAYT	300
FDAGPNAVYL YLAENESKLF AFIYKLFGSV PGWDKKFTTE QLEAFNHQFE SSNFTARELD	360
LELQKDVARL ILTQVGSGPQ ETNESLIDAK TGLPKEKLSS GGGGGGGGG GGGSAEAWY	420
LGNAYYKQGD YQKAIEYYQK ALELDPNNAE AWYNLGNAYY KQGDYQKAI E YYQKALELP	480
NNAAEAWYLNQ NAYYQGDYQ KAIEDYQKAL ELDPPNLLQAE AWKNLGNAYY KQGDYQKAI	540
YYQKALEIDY QKAAEAWYLNQ NAYYQGDYQ KAIEDYQKAL ELDPPNNAKAW YRRGNAYYQK	600
GDYQKAIEDY QKAAEAWYLNQ NAYYQGDYQ KAIEDYQKAL ELDPPNNAKAW YRRGNAYYQK	660
VFIWKISDFP RKRQEAVAGR IPAIFSPAFY TSRYGYKMCL RIYLNGDGTG RGTHLSLFFV	720
VMKGPNDALL RWPFNQKVTL MLLDQNNREH VIDAFRPDVT SSSFQRPVND MNIAASGCPFL	780
CPVSKMEEAKN SYVRDAAIFI KAIVDLTGLG SAGSAAGSGE FGSAAEAAKE AAAKAGSAGS	840
AAGSGEFGSA SIKLQSSDGE IFEVDVIEAK QSVTIKTMLE DLGMDDEGDD DPVPLPVNA	900
AIIKKVIQWC THHKDPPP EDDENKEKRT DDIPVWDQEF LKVDQGTLFE LILAANYLDI	960
KGLLDVTCKT VANMIKGKTP EEIRKTFNIK NDFTEEEEAO VRKENQWCWS G	1011

SEQ ID NO: 171 moltype = AA length = 720
 FEATURE Location/Qualifiers
 source 1..720

-continued

```

mol_type = protein
organism = synthetic construct

SEQUENCE: 171
MTADNNNSMPH GAVSSYAKLV QNQTPEDILE EFPEIIPLQQ RPNTTRSSETS NDESGETCFS 60
GHDEEQIKLM NENCIVLDWD DNAIAGATGK VCHLMENIEK GLLHRAFSVF IFNEQGELL 120
QQRATEKITF PDLWTNTCCS HPLCILDELG LKGKLEDDKIK GAITAAVRKL DHELGIPED 180
TKTRGKFHFH MRIHYMAPSN EPWGEHEIDY ILFYKINAKE NLTVNPVNVE VRDFKWSPN 240
DLKTMFADPS YKFTPWFKII CENYLFNWWQE QLDDLSEVEN DRQIHRMLKL SGGGGSGGG 300
SGGGGSAEAW YNLGNAYYKQ GDYQKAIEYY QKALELDPNN AEAWYNLGNNA YYKQGDYQKA 360
IEYYQKALEL DPNNAEAAWY LGNAYYQKQD YQKAIEDYQK ALEELDPNNLQ AEAWKNLGN 420
YYKQGDYQKA IEYYQKALEL DPNNASAWYN LGNAYYQKQD YQKAIEYYQK ALEELDPNNAK 480
AWYRRGNAYY KQGDYQKAIE DYQKALELDP NNRSRSAGGG GSAGGGGSGGG GASSYYHHHH 540
HHLESTSLYK KAGSGSNTV ELKNYIQELE ERNAEKLNLK EHLKFAKAEF EFEELAAHKFE 600
GSAGSAAGSG EFGSABAAAK EAAAKAGSAG SAAGSGEFGS SYYHHHHHHHL ESTSLYKAG 660
SGSQKVAQLK NRVAYKLKEN AKLENANLEKD IANLEKDIAN LERDVARGSG 720

SEQ ID NO: 172      moltype = AA length = 801
FEATURE          Location/Qualifiers
source           1..801
mol_type = protein
organism = synthetic construct

SEQUENCE: 172
MEAKIDELIN NDPVWSSQNE SLISKPYNHLL KPGKKNFRL NLIVQINRVM NLPKDQLAIV 60
SQIVELLHNS SLLIDDEDN APLRRGQTTS HLIWGPSTI NTANYMYFRA MQLVSQQLT 120
EPLYHWLITI FNEELINLH GQGLDIYWRD FLEPIIPTQE MYLNMMVNKT GGLFRLTLLR 180
MEALHSPSHH GHSLVPFINL LGIYQIRDD YLNLKDFQMS SEKGFAEDIT EGKLSFPIVH 240
ALNFTKTKQQ TEQHNEBILRI LLLRTSDKDI KLKLQILEF DTNSLAYTKN FINQLVNM 300
NDNENKYLDP LASHSDTATN LHDELLYIID HLSELKLSSG GGSGGGGSGGG GGSAEAWYN 360
GNAYYQKQDY QKAIYEYQKA LEELDPNNAEA WYNLGNAYYK QGDYQKAIEY YQKALELDPN 420
NAEAWYNLGN AYYKQGDYQK AIEDYQKALE LDPNNLQAEWA WKNLGNAYYK QGDYQKAIEY 480
YQKALELDPN NASAWYNLGN AYYKQGDYQK AIEYYQKALE LDPNNAKAWY RRGNAYYKQG 540
DYQKAIEDYQ KALELDPNNR SRSAGGGGSG GGGSGGGGAS LCTMKKGPSG YGFNLHSOKS 600
KPGQFIRSVD PDSPAFAASL RAQDRIVEVN GVCMEKGKOH DVVSAIRAGG DETKLLVVDR 660
EGSAGSAAGS GEFGSAAAAA KEAAAKAGSA GSAAGSGEFQ SSSGAIYTVELKRYGGPLG 720
ITISGTEEPF DPIISSLTK GGLAERTGAI HIDGRILAIN SSSLKGKPLS EAIHLLQMAG 780
ETVTLKIKQ TDAQPASSGS G 801

SEQ ID NO: 173      moltype = AA length = 547
FEATURE          Location/Qualifiers
source           1..547
mol_type = protein
organism = synthetic construct

SEQUENCE: 173
MKCSTFSFWF VCKIIFFFFS FNIQTSIANP RENFLKCFSQ YIPNNATNLK LVYTQNNPLY 60
MSVLNSTIHN LRFTSDTTPK PLVIVTPSHV SHIQGTILCS KKVGQJIRTR SGGHDSEGMS 120
YISQVPFIV DLRNMRSYGLA DVHSQTAWE AGATLGEVYY WWNEKNENLS LAAGYCPTVC 180
AGGHFGGGY GALPMRNYGLA ADNIIDAHLV NVHGKVLDRK SMGEDLFWAL RGGGAESFGI 240
IAAWKIRLVA VPKSTMFSVK KIMEIHELVK LVNKWQNIAY KYDKDLLMFT HFITRNITDN 300
QGKNTKTAHT YFSSVFLGGV DSLVDLMNKS FPELGIKKTD CRQLSWIDTI IFYSGVVNYD 360
TDFNFNKBILL DRSAQGNGAF KIKLDYVKKP IPESVVFVQIL EKLYEEDIIGA GMYALYPYGG 420
IMDEISESAI PPPFHRRAGILY ELWYICSWEK QEDNEKHLMW IRNIYNFMTP YVSKNPRLAY 480
LNYRDLDDIGI NDPKNPNNYT QARIWGEKY GKNFDRLVK KTLVDPNNFF RNEQSIPPLP 540
RHRHGSG 547

SEQ ID NO: 174      moltype = AA length = 548
FEATURE          Location/Qualifiers
source           1..548
mol_type = protein
organism = synthetic construct

SEQUENCE: 174
MNCSTFSFWF VCKIIFFFLS FNIQTSIANP QENFLKCFSE YIPNNPANPK FIYTQHDQLY 60
MSVLNSTIQN LRFTSDTTPK PLVIVTPSNV SHIQASILCS KKVGQJIRTR SGGHDAEGLS 120
YISQVPAIV DLRNMHTVKV DIHSQTAWE AGATLGEVYY WWNEMNENFS FPGGYCPTVG 180
VGGHFGGGY GALPMRNYGLA ADNIIDAHLV NVHGKVLDRK SMGEDLFWAI RGGGENFGI 240
IAACKIKLKV VPSKATIFSV KKNMEIHGLV KLFNWKQNIY KYDKDMLT THFRTRNITD 300
NHGKNTKTVH GFYFSSIFLGG VDSLVDLMNK SFPELGIKKT DCKELSWIDT TIFYSGVVNY 360
NTANFKKEIL LDRSAGKTKA FSIKLDYVKK LIPETAMVKI LEKLYEEEVG VGMYVLYPYG 420
GIMDEISESAI IPPFPHRAGIM YELWYTATWE KQEDNEKHIN WVRSVYNFTT PYVSQNPRLA 480
YLNRYRDLDDLG KTNPESPNYY TQARIWGEKY FGKKNFNRLLVK VKTKADPNNF FRNEQSIPPLP 540
PPRHHGSG 548

SEQ ID NO: 175      moltype = AA length = 886
FEATURE          Location/Qualifiers
source           1..886
mol_type = protein
organism = synthetic construct

SEQUENCE: 175
MNHLRAEGPA SVLAIGTANP ENILLQDEFY DYYFRVTKSE HMTQLKEKFR KICDKSMIRK 60
RNCFLNNEEHL KQNPRLVEHE MQTLDARQDM LVVEVPKLGK DACAKAIKEW GQPKSKITHL 120

```

-continued

IFTSASTTDM PGADYHCAKL LGLSPSVKRV MMYQLGCYGG GTVLRIAKDI AENNKGARVL 180
 AVCCDIMACL FRGPSESDLLE LLVGQAIFGD GAAAIVVGA PDESVGERPI FELVSTGQTI 240
 LPNSEGTIGG HIREAGLIFD LHKDVPMLIS NNIEKCLIEA FTPIGISDWN SIFWITHPGG 300
 KAILDKVEEK LHLKSDKFVD SRHVLSEHGN MSSSTVLFVM DELRKRSLEE GKSTTGDGFE 360
 WGVLFGFPGP LTVERVVRS VPIKPGTSLGG GGSGGGGSGG GSAAEAWYNL GNAYYKQGDY 420
 QKAIEYYQKA LELEDPNNAEA WYNLGNAYYK QGDYQKAIEY YQKALELDPN NAEAWYNLGN 480
 AYYKQGDYQK AIEDYQKALE LDPPNLQABA WKNLGNAYYK QGDYQKAIEY YQKALELDPN 540
 NASAWYNLGN AYYKQGDYQK AIEYYQKALE LDPPNNAKAWY RRGNAYYKQG DYQKAIEDYQ 600
 KALELDPNR SRSAGGGSG GGGSGGGAS GNNLETYEWY NKSISRDKAEE KLLLDTGKEG 660
 AFMVRDSRTP TGYTVSFTK AIISENPCIK HYHIKETNDs PKRYVAEKY VFDSIPLLIQ 720
 YHQYNGGLV TRLRYPVCGG SAGSAAGSGE FGSAEAAAKEA AAKAGSAGS AAGSGEFGSG 780
 SHPWFFGKIP RAKAEEMLSK QRHDGAFLIR ESESAPGDFS LSVKFGNDVQ HFKVLRDGAG 840
 KYFLWWVKFN SLNELVVDYHR STSVSRNQQI FLRDIEQVPO QPTGSG 886

SEQ ID NO: 176 moltype = AA length = 668
 FEATURE Location/Qualifiers
 source 1..668
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 176
 MAVKHLIVLK FKDEITEAQK EEEFKTYVNL VNIIIPAMKDV YWGDVDTQKN KEEGYTHIVE 60
 VTFESVETIQ DYIIPAHVG FGDVYRSFWE KLLIFDYTPR KKLISGGGGSG GGGSGGGGSA 120
 EAWEYNLGNAY YKQGDYQKAIEY EYYQKALELD PNNAEAWYNL GNAYYKQGDY QKAIEYYQKA 180
 LELEDPNNAEA WYNLGNAYYK QGDYQKAIED YQKALELDPN NLQAEAWKLN GNAYYKQGDY 240
 QKAIEYYQKA LELEDPNNASA WYNLGNAYYK QGDYQKAIEY YQKALELDPN NAKAWYRGN 300
 AYYKQGDYQK AIEDYQKALE LDPPNNSRSA GGGGSGGGGS GGGGASQDR SEATLIKRFK 360
 GEGVRYKAKL IGIDEVAAR GDKLICQDSMM KLKGVVAGAR SKGEHKQKIF LTISPGGIKI 420
 FDEKTGALQH HHAVHEISYI AKDITDHRAF GYVCVGKEGNH RFVAIKTAQA AEPVILDLRD 480
 LFQLIYELKQ REELEKKAGS AGSAAGSGEF GSAAEAAAKEA AAKAGSAGSA AGSGEFGSGS 540
 HMGSQFWTS QTKEASERCG LQGSYIILRVE AEKLTLLTLG AQSQILEPLL FWPYTLLRRY 600
 GRDKVMMSFE AGRRCPSPGPG TFTFQTQSQN DIFQAVEAAI QQQKAQGKVG QAQDILRLHEH 660
 HHHHHGSG 668

SEQ ID NO: 177 moltype = AA length = 800
 FEATURE Location/Qualifiers
 source 1..800
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 177
 MGSSSVCTFS FQTNYHTLLN PHNNNPKTSL LCYRHPKTPY KYSYNNFPSK HCSTKSFHQ 60
 NKCSESLSIA KNSIRAATTN QTEPPESDNH SVATKILNFG KACWKLQRPY TIIAFTSCAC 120
 GLFGKELLHN TNLISWLSMF KAFFFNLVAIL CIASFPTTIN QIYDLHIDRI NKPDLPLASG 180
 EISVNTAWIM SIIVALFGLI ITIKMKGGPL YIFGYCFCGIF GGIVYSPPF RWKQNPSTAF 240
 LLNFLAHIIT NFTFYIASRA ALGLPFELRP SFTPLLAFAFM SMGSALALIK DASDVEGDTK 300
 FGISTLASKY CSRNLTLFCS GIVLSSYVAA ILAGIIWPQA FNSNVMLSH AILAPWLILQ 360
 TRDFALTNYD PEAGGRFYEE MWKLIIYAEYL VVVFILKLSGG GGSGGGGSGG GGSABEWYNL 420
 GNAYYKQGDY QKAIEYYQKA LELEDPNNAEA WYNLGNAYYK QGDYQKAIEY YQKALELDPN 480
 NAEAWYNLGN AYYKQGDYQK AIEDYQKALE LDPPNQLQAEA WKNLGNAYYK QGDYQKAIEY 540
 YQKALELDPN NASAWYNLGN AYYKQGDYQK AIEYYQKALE LDPPNNAKAWY RRGNAYYKQG 600
 DYQKAIEDYQ KALELEDPNN SRSAGGGSG GGGGGGGAS AEYVRALFD NGNDEEDLPF 660
 KKGDIILRQD PKEEQWNNAE DSEGKRGMI P VPYVEKYGSA GSAAAGSGEF GSAAEAAKEA 720
 AKAGSAGSAA GSGEFGSLIK HMRAEALDFD TGNSKLELNF KAGDVIFLLS RINKDWLEGT 780
 VRGATGIFPL SFVKILKGSG 800

SEQ ID NO: 178 moltype = AA length = 2744
 FEATURE Location/Qualifiers
 source 1..2744
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 178
 MSEEESLFESS PQKMEYEITN YSERHTELPG HFIGLNTVDK LEESPLRDFV KSHGGHTVIS 60
 KILIANNGIA AVKEIRSVR K WAYETFGDDR TVQFVAMATP EDLEANAEYI RMADQYIEVP 120
 GGTNNNNNYAN VDLIVDIAER ADVDAVWAGW GHASENPLLP EKLSQSQRKRV IFIGPPGNAM 180
 RSLGDKISST IVAQSAKVPC IPWSGTCVDT VHDEKTLGLV SVDDDIYQKG CCTSPEDGLQ 240
 KAKRIGFPVM IKASEGGGGK GIRQVEREED FIALYHQAAN EIPGSPIFIM KLAGRARHLE 300
 VQLLADQYGT NISLFGRDCS VQRRLHQKIE EAPVTIAKAE TFHEMKAACV RLGKLVGYVS 360
 AGTVEYLYSH DDGKFYFLEL NPRLQVEHPT TEMVSGVNLP AAQLQIAMGI PMHRISDIRT 420
 LYGMNPHSAS EIDFEFKTQD ATKKQPRIP KGHCTACRIT GDIMAIMTLD DPSKVHALP 480
 RSSSNVWGYF SVGNNGNIHS FSDSQFGHIF AFGENRQASR KHMVVALKEI SIRGDFRTTV 540
 EYLIKLLTE DFEDNTITTG WLDDLITHKM TAEKPDPTLA VICGAATKAF LASEEARHKY 600
 IESLQKGQVL SKDLLQTMFP VDFIHEGKRY KFTVAKSGND RYTLFINGSK CDIILRQLSD 660
 GGLLIAIGGK SHTIYKKEEV AATRLSVDMS TTLLEVENDP TQLRTPSPKL LVKFLVENG 720
 HIIKGQPYAE IEVMMQMPL VSQENGIVQL LKQPGSTIVA GDIMAIMTLD DPSKVHALP 780
 FEGLMLPDFGS PVIETGKPAV KFKSLVSTLE NILKGYDNQV IMNASLQLQI EVLRNPKLPY 840
 SEWKLHISAL HSRLPAKLDE QMEELVARSL RRGAVFPARQ LSKLIDMAVK NPEYNPDKLL 900
 GAVVEPLADI AHKYSNGLEA HEHSIFVHFL EEEYVEKLF NGPNVREENI ILKLRDENPK 960
 DLDKVALTVL SHSKVSAKNN LILAILKHQ PLCKLSSKVS AIFSTPLQHI VELESKATAK 1020
 VALQAREILI QGALPSVKER TEQIEHILKS SVVKVAYGSS NPKRSEPDLN ILKDLIDSNY 1080
 VVFDVLLQFL THQDPVVTAA AAQVYIRRAY RAYTIGDIRV HEGVTVPIVE WKFQPLSAAF 1140

-continued

STFPPTVSKM	GMMRAVSVSD	LSYVANSQSS	PLREGILMAV	DHLEDDVDEIL	SQSLEVIPIRH	1200
QSSSNGPAPD	RSGSSASLSN	VANVCVASTE	GFESEEEILV	RLREILDLNK	QELINASIRR	1260
IITFMGFKDG	SYPKYTFNG	PNYNENETIR	HIEPALAFQL	ELGRLSNFNI	KPIFTDNRNI	1320
HVYEAVSKTS	PLDKRFFTRG	IIRTGHIRDD	ISIQEYLTS	ANRLMSDILD	NLEVTDTNS	1380
DLNHIFINF1	AVFDISPEDL	EAAFGFGLER	FGKRLRLRLRV	SSAEIRRIIK	DPQTGAPVPL	1440
RALINNVSGY	VIKTEMYTEV	NAKAGEWVK	SLGKPGSMHL	RPIATPYPVK	EWLQPKRYKA	1500
HLMGTTVYD	FPELFQASS	SQWNKFSADV	KLTDDFFISN	ELIEDENGEL	TEVEREPGAN	1560
AIGMVAFKIT	VKTPEYPRGR	QFVVVANDIT	FKIGSFGPQE	DEFFNKVTEY	ARKRGIPRIY	1620
LAANSGARIG	MAEEIVPLFQ	VAWNDAAANDP	KGPQYLYLTS	EGMETLKKFD	KENSVLTERT	1680
VINGEERFV1	KTIIGSEDGL	GVECLRGSGL	IAAGTSRAYH	DIFTITLVTC	RSVGIGAYLV	1740
RLGQRAIQC1	COPIIILTGP	AINKMLGREV	YTSNLQLGGT	QIMYNNGVSH	LTAVIDLAGV	1800
EKIVEWMSYV	PAKRNMVPPI	LETKDTWDRP	VDFTPNTDET	YDVRWMIEGR	ETESGFYGL	1860
FDKGSFFETL	SGWAKGVVVG	RARLGGIPLG	VIGVETRTVE	NLIPADPAND	NSAFTLIQEP	1920
GQWHPNSAF	KTAQAINDFN	NGEQLPMLN	ANWRGFSGQQ	RDMFNEVLKY	GSPFIVDALVD	1980
YKQPIIIYIP	PTGELRGGSW	VVVDPTINAD	QMEMYADVNA	RAGVLEPQGM	VGIKPRREKL	2040
LDTMNRLDDK	YRELRSQLSN	KSLAPEVHQ	ISKQLADRER	ELLPPIYGQIS	LQFADLHDRS	2100
SRMVAKGVIS	KELEWEARLK	FFFWRLLRRL	NEEYLIKRLS	HQVGEASRLE	KIARIRSWYP	2160
ASVDHEDDRO	VATWIEENYK	TLDKLLGKLX	LESYFAQDLAK	KIRSDHDNAI	DGLSEVIKML	2220
STDDEKEKLX	TLKKLSSGGGG	SGGGGSSGGG	SABAOWNLGN	AYYKQGDYQK	AIEYYQKALE	2280
LDPNNAEAWY	NLGNAYYKQG	DYQKAIEYYQ	KALELDPNNA	EAWYNLGNAY	YKQGDYQKAI	2340
EDYQKALELD	PNNLQAEAWI	NLGNAYYKQG	DYQKAIEYYQ	KALELDPNNA	SAWYNLGNAY	2400
YKQGDYQKAI	EYYQKALELD	PNNAKAQDNL	HNMLGAQSIO	PTANLDRDTD	LVYLNVMELV	2460
SAGGGGSGGG	GGGGGGASGS	SGEFQSSYH	HHHHHLESTS	LYKKAGSSQ	KVAELKNRVA	2520
AQLPPEGYVV	VVKNVGLTLR	KLIGSVDDLL	PSLPSSSRTE	IEGTQKLLNK	DLAELINKMR	2580
LAQONAVTSL	SEECKRQMLT	ASHTLAVIDAK	NLLDAVDQAK	VLANLAHPA	EGSAGSAAGS	2640
GEFGSAEAAA	KEAAAKAGSA	GSAAGSGEFG	SGAMATPGSE	NVLPREPLIA	TAVKFLQNSR	2700
VRQSPATR	AFLKKKG LTD	EEIDMAFQQS	GTAADEPSSL	WGSG		2744

SEQ ID NO: 179 moltype = AA length = 3623

FEATURE Location/Qualifiers

source 1..3623

mol_type = protein

organism = synthetic construct

SEQUENCE: 179

MGSAGSAAGS	GEFGSGSAA	GSGEFGSAGS	AAGSGEFSYY	HHHHHHLEST	SLYKKAGSGS	60
ARNAYLRKKI	ARLKKDNLQL	ERDQNLEKI	IANLRDEIAR	LENEVASHEQ	GSAGSAAGSG	120
EFAAAAEEA	AAKAGSAGSA	AGSGEFSYYH	HHHHHLESTS	LYKKAGSSN	LVAQLENNEVA	180
SLENENETL	KKLNHKKDLDI	AYLEKEI	RKKIEEGSAG	SAAGSGEFGS	AEAAAEEAAA	240
KEAAAKEAAA	KAGSAGSAAG	SGEFQSSYH	HHHHHLESTS	LYKKAGSSQ	KVAELKNRVA	300
VKLNRNEOLQK	NKVEELKNRN	AYLKNELATL	ENEVARLEND	VAEGSAGSAA	GSGEFAEEAAA	360
KEAAAKAGSA	CSAAGSGEFS	YHHHHHHHLE	STSLYKKAGS	GSNEVTTLEN	DAAFIENENA	420
YLEKEIARLR	KEKAALRNRL	AHKMGAGSA	AGSGEFGSAA	AAKEAAAEE	AAKEAAAEE	480
GSAGSAAGSG	EGFSRPPTIS	NPPPLISSAK	HPSVGSAGSA	AGSGEFAEEAA	AAKEAAAKAGS	540
AGSAAGSGEF	NFLQSRPEPT	APPEESFRSG	GSAGSAAGSG	EFGSAAEAAAK	EAAKEAAAEE	600
EAAAAGSAG	SAAGSGEFGS	SKGTGLNPNA	KVWQEIAPGN	GSAGSAAGSG	EFAEEAAAEE	660
AKAGSAGSA	AGSGEFPDGG	TTFEHLWSSL	EPDSTYGSAG	SAAGSGEFGS	EEAAAKEAAA	720
KEAAAKEAAA	KAGSAGSAAG	SGEFQSSYH	HHHHHLESTS	LYKKAGSGSK	RIAYLRKIA	780
ALKKDNANL	KDIANLENET	ERLKEIKTL	ENEVASHBQG	SAGSAAGSGE	FAEEAAKEAA	840
AKAGSAGSA	SGGEFSYYHH	HHHHLESTS	YKKAGSGSNL	LATLRSTAAB	LENENHVLEK	900
EKEKLKEKE	QLLNKLEAYK	GSAGSAAGSG	EFGSAAEAAAK	EEAAKEAAAK	EEAAKAGSAG	960
SAAGSGEFGS	PATSQHPPP	PGHRSQAPSH	GSAGSAAGSG	EFAEEAAKEA	AAKAGSAGSA	1020
AGSGEFLNS	LLILLEAAEY	LERRDRGSAG	SAAGSGEFGS	EEAAAKEAAA	KEAAKEAAA	1080
KAGSAGSAAG	SGEFGSRPPT	ISNPPPLISS	AKHPVGSAG	SAAGSGEFAE	AAKEAAAKAA	1140
GSAGSAAGSG	EFPNFLQSRP	PTTAPPEESFR	SGGSAGSAAG	SGEFGSAEEAA	AKEEAAKEAA	1200
AKEEAAAKAGS	AGSAAGSGEF	GSSKGTLNP	NAKVWQEIA	GNGSAGSAAG	SGEFAEEAAAK	1260
EAAAAGSAG	SAAGSGEFPD	GGTTFEHLWS	SLEPDSTYGS	AGSAAGSGE	GSAAEAAKEA	1320
AAKEAAAKEA	AAKAGSAGSA	AGSGEFGSSY	YHHHHHLES	TSLYKKAGSG	SKRIAYLRKK	1380
IAALKKDNAN	LEKDANLEN	EIERLKEIK	TLENEVASHE	QGSAGSAAGS	GEFAEAAAKE	1440
AAAKAGSAGS	AAGSGEFSYY	HHHHHLESTS	SLYKKAGSGS	NLLATLRSTA	AVLENENHV	1500
EKEKEKLKE	KEQLLNKLEA	YKGAGSAGA	SGEFGSABAA	EEAAKEAAKEA	AAKEAAAKAGS	1560
AGSAAGSGEF	GSALVDDAAD	YEPPPSNNEE	ALGSAGSAAG	SGEFAEAAAK	EEAAKAGSAG	1620
SAAGSGEFLRE	LFDDDPYVN	QNLDKARQGS	AGSAAGSGEF	GSAAEAAKEA	AAKEAAAKEA	1680
AAKAGSAGSA	AGSGEFGSKN	TKSMNFDNPV	YRKTTTEEGS	AGSAAGSGEF	EEAAAKEAAA	1740
KAGSAGSAAG	SGEFRSLPST	WIENKLGYMS	DPNWGSAGSA	AGSGEFGSAAE	AAKEAAAKEA	1800
AAAKEAAAKA	GSAGSAAGSG	EFGSVVDNSP	PPALPPKKRQ	SAPSGSAGSA	AGSGEFAEAA	1860
AKEAAAKAGS	AGSAAGSGEF	TQRSKPQPAV	PPRPSADLIL	GSAGSAAGSG	EFGSAAEAAAK	1920
EAAAKEAAAK	EEAAKAGSAG	SAAGSGEFGS	TDEEREETEE	EVYLLNSTTL	GSAGSAAGSG	1980
EFAEEAAAKA	AAKAGSAGSA	AGSGEFDGVS	SGTQRLLSAT	VRTYSCGSAG	SAAGSGEFGS	2040
EEAAAKEAAA	KEAAKEAAA	KAGSAGSAAG	SGEFQSSYH	HHHHHLESTS	LYKKAGSGSQ	2100
KVAQLKNRVA	YKLKENAKLE	NIVARLEND	ANLEKDIANL	EKDIANLERD	VARGSAGSAA	2160
GSGEFAEEAA	KEAAAKAGSA	GSAGSGEFS	YHHHHHHHLE	STSLYKKAGS	GSNTVKELKN	2220
Y1QELEERNA	ELKLNKLEHLK	FAKAELEFEL	AAHKFEQSAG	SAAGSGEFGS	EEAAAKEAAA	2280
KEAAAKEAAA	KAGSAGSAAG	SGEFGSHDDS	LPHPQQATDD	SGHESDGSAG	SAAGSGEFAE	2340
AAKEAAAKA	GSAGSAAGSG	EFGSPNAGSV	EQTPKKPGLR	RRGSAGSAAG	SGEFGSAAEAA	2400
AKEAAAKEAA	AAEAAAKAGS	AGSAAGSGEF	GSSYYHHHH	HLESTSLYKK	AGSGSFENVT	2460
HEFILATLEN	ENAKLRLREA	KLERELARLR	NEAWLGSAG	SAAGSGEFAE	AAKEAAAKA	2520
GSAGSAAGSG	EFSYHHHHHH	HLESTSLYKK	AGSGSQKVEE	LKNKIAELEN	RNAVKKNRVA	2580
HLKQEIAYLK	DELAHEFEG	SAGSAAGSGE	FGSAAEAAAKE	AAKEAAAKA	AAKAGSAGS	2640
AAGSGEFGSV	SSTKLVSFH	DSDEDLLHIG	SAGSAAGSGE	FAEAAAKEAA	AAKAGSAGSAA	2700

-continued

GSGEFAAATP ISTFHDDSDE DLLHVGSGS AAGSGEFGSA EAAAKEAAAK EAAAKEAAAK 2760
 AGSAGSAAGS GFGFSSYYHH HHHHLESTSL YKKAGSGSQK VESLKQKIEE LKQRKAQLKN 2820
 DIANLEKEIA YAETGSAGSA AGSGEFAEEA AKEAAAKAGS AGSAAGSGEF SYYHHHHHHL 2880
 ESTSLYKKAG SEFFRRERNK MAAAKCRNRR RELTDTLQAE TDQLEDEKSA LQTEIANLLK 2940
 EKEKLEPILA AHRPACKIPD DLGPFEEEMSL EGSGAGSAAGS GEFGSAAEEE KEAAAKEAAA 3000
 KEAAAKAGSA GSAAGSGEFG SFQMPADTPP PAYLPPPEDPM TGSAGSAAGS GEFAEAAAKE 3060
 AAKAGSAGS AAGSGEFERE SNEEPPPPVE DPYWGNNGSA GSAAGSGEFG SAEAAAKEEA 3120
 AKEAAAKEAA AKAGSAGSAA GSGEFGSSYY HHHHHLEST SLYKKAGSGS QKVAELKNRV 3180
 AVKLRNRNEQL KNKVEELKNS NAYKLKNELT LENEVARLEN DVAEGSGAGSAA AGSGEFAEEA 3240
 AKEAAAKAGS AGSAAGSGEF SYYHHHHHL ESTSLYKKAG SNSNEVTTLE NDAAFIENEN 3300
 AYLEKEATRL RKEKAALRNK LAHKKSYYHH HHHHLESTSL YKKAGSGSAR NAYLRKKIAR 3360
 LKKDNQLER DEQNLEKIIA NLRDEIARLE NEVASHEQGS AGSAAGSGEF AEAAAKEAAA 3420
 KAGSGASAG SGEOFSSYYHH HHHHLESTSL KKAGSGSNLV AQLENEVASL ENENETLKKK 3480
 NIHKDKLIAY LEKEIANLRK KIEEGSAGSA AGSGEFGSAA AAAKEAAAKE AAAKEAAAKA 3540
 GSAGSAAGSG EFGSEQKLIS EEDLEQKLIS EEDLGSAGSA AGSGEFGSAG 3600
 SAAGSGEFGS AGSAAGSGEF GSG 3623

SEQ ID NO: 180 moltype = AA length = 455
 FEATURE Location/Qualifiers
 source 1..455
 mol_type = protein
 organism = synthetic construct
 SEQUENCE: 180
 MGSGAGSAAGS GEFGSAGSAA GSGEFGSAGS AAGSGEFSYY HHHHHLEST SLYKKAGSGS 60
 ARNAYLRKKI ARLKKDNQL ERDQNLEK IANLRDEIAR LENEVASHEQ GSAGSAAGSG 120
 EFAEAAAKEA AAKAGSAGSA AGSGEFSYYHH HHHHLESTS LYKKAGSGSN LVAQLENNEVA 180
 SLENENETLK KKNLHKDKDL AYLEKEIANL RKKIEEGSAG SAAGSGEFGS AEAAAKEAAA 240
 KEAAAKEAAA KAGSGAGSAAG SGEFGSSATR ELDELMASLS DFKIQGGSAG SAAGSGEFAE 300
 AAKEAAAKA GSAGSAAGSG EFDLALSENW AQEFLAAGDA VDGSAGSAAG SGEFGSAAEA 360
 AKEAAAKEAA AKEAAAKAGS AGSAAGSGEF GSDYKDDDDK DYKDDDDKDY KDDDKGSAG 420
 SAAGSGEFGS AGSAAGSGEF GSAGSAAGSG EFGSG 455

SEQ ID NO: 181 moltype = DNA length = 4596
 FEATURE Location/Qualifiers
 source 1 .. 4596
 mol_type = other DNA
 organism = synthetic construct
 SEQUENCE: 181
 atgagtgtca aggcaatttc tgaacaaact ggtaaagaat tggtttgcataa gtttatttgt 60
 actacatcatc ccatccaaaa tagattcaaa tacgcttagt ttaccccgaga tactgactgg 120
 gctagattgt tacaagatca tccatgggtt ttatctcaaa acttgggttgc caaacctgac 180
 caattaatca agagaaggagg taaatgggtt ttatgggttgc atttggatgg 240
 gtaaaggctt ggttgaacc accatgttgc caagaaggcca cagttgttgc agctaccgg 300
 ttcttgaaa atttcttgcat cgaaccattt gtccttcattt cacaagccgaa agaattctat 360
 gtatgtatct acgttacttagt agagggttgc tatgttttat ttcatcacgc aggtgggttc 420
 gagttgggtt acgttgcacgc caaggctcaa aagttgttgc ttgggttgc tgaaaagg 480
 aacccagaag aacccagaataaa gcattttgttgc ttccacgcac ccataagataaa aaaggaaata 540
 ttggccctct ttataatgttgc ttgtttaat ttctacgcac atttgcattt caccctacttgc 600
 gaaattaacc cattatgttgc tacaaggat ggtgtatatg ttttggactt agctgcacaaa 660
 gtttgcacca cagccgacta cattttgttgc gtcacatggg ttgcacatcga atttccacct 720
 catttcggta gagaaggatca tcccaagaaa gcttacatttgc ttgttgc tgcaatgttgc 780
 ggttgcacatc tgaagggttgc atttgcacgc cttaaaggta gaattttggc catgggttgc 840
 ggttgggtgtt ctatgttgcgtt atattcttgcat actatatgttgc acttgggttgc ttgttacgc 900
 ttggccaaact acgggttgcata ctcaggttgc ccatccgaaac aacaaacatca cgattacgc 960
 aagaccatct tgcctttat gactatgggaa aacccatcgc ttgttacgcattt cttgttacgc 1020
 ggttggtagt tcccaatattt tacaatgggat gcccgcacatc tcaagggttgc cgttgcacatc 1080
 ataagagattt accaagggtcc atttgcacccatc accaaggatca cacgttgcatc aatatttgcgtt tagaaggat 1140
 ggttgcacatc accaaggatcc ttgttgcacccatc atgggttgcacccatc taggttacgc 1200
 ccaatccatc tctttgttgc acgggttgcata ctcaggttgc ccatccgaaac aacaaacatca cgattacgc 1260
 catagacccaa tccctaatca acctccaaacc gcaaggccacca ctgcacatc tttgttacgc 1320
 gcttctgggtt caacttccac accagtcctt tcttgcacccatc tcaagggttgc cgttgcacatc 1380
 gcttgcacccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 1440
 agtccctatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 1500
 gtatgggttgc tccatccatc acgggttgcata ctcaggttgc ccatccgaaac aacaaacatca cgattacgc 1560
 agatgttgcacccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 1620
 ttcttacttggg gtcacaaaggatcc aatatttgcacccatc tccatccatc tccatccatc tccatccatc tccatccatc 1680
 agaaaacatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 1740
 tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 1800
 attccatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 1860
 atccatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 1920
 ggtatgttgc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 1980
 tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 2040
 gtatgttgc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 2100
 gtatgttgc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 2160
 ggtactgttgc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 2220
 gtttgttgc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 2280
 gcaaggccatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 2340
 gaaggccatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 2400
 tttgttgc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc tccatccatc 2460

-continued

gtccctatgg attactcttggcagaagagaa ttgggtttaa ttagaaaggcc agctagttt 2520
 atgacctcta tatgtatga aagaggtaa gaattgtatc atgctggat gcctattact 2580
 gaagtattca aagaagaaat gggtatcggtt ggtgttttag gttgttgg tttccaaag 2640
 agattgc当地 aactacttgc tcaattcatt gaaatgtgt taatggttac agctgatcat 2700
 ggtcctgtt ttcaggtgc acacaatcc ataattctgc ctagagttt gaaagattt 2760
 gtttcttctt tgaccctcagg ttgttactt atgggtgaca gattgggtt tgcattagac 2820
 gcccgtc当地 agatgtttc aaaagcttc gattccggta taatcccaat ggaattcggt 2880
 aataagatga aaaaggaggg taaattgata atgggtatcg gtcatcggt taagtctatc 2940
 aataaccctg atagactagatc acaaattcttggcagaagactt ttagacaaca ctttcagcc 3000
 acaccccttgg tagattatgc tttggaaatggc aaaaaggatataa ccacttctaa aaaggccaaat 3060
 ttgatcttgc acgttgatgg tttatgttgg gttgttgg tgcacatgtt gagaactgt 3120
 ggttccttca cttagagaaga agctgatgaa tatatcgaca ttgggtgattt gaaatggatc 3180
 ttgtttagt tagatctatc ggggttattt ggttactt ggttactt gggatcaaaa gagattaaag 3240
 caagggttgc acagacatcc atgggtatcg atttcttgc ttttacatcg acacatgtca 3300
 atgaaatttgc ttgggttggg ttgttcttgg ggtgttggt ctgggttggg tggtagtgcc 3360
 gaagcttggg acaatttggg taacgc当地 tacaaggccgg gttactacca aaaggcaatt 3420
 gaatattacc aaaaggccctt ggaatttagac ccaataaaatc cagaaggccgtt gtataatttt 3480
 ggtatgttcc attataaaca ggggtactt cttttttttt gttttttttt gttttttttt 3540
 ttgaaatgg acccttataa cgttgc当地 ttgttataat tgggttggcc ttattataag 3600
 cagggtgact atccaaaagc catgc当地 taccaaaagg ctttggattt agatccaaat 3660
 aacttcaacg cccaaatcttgg gaaatattt ggttacttcc acttataaaca ggggtactac 3720
 caaaaaggccaa ttgaatacta tcaaaatgtt gttttttttt gttttttttt gttttttttt 3780
 ttgttacattt ttgggttgc ttacttataatc cagggtgactt atcagaaggcc ctttggat 3840
 tatccaaaagg cttttagattt ggttccaaat aacgcttataat aacgcttataat ctttggat 3900
 gtttattttt aacagggttgc ctaccatggaa gccatttgc当地 atttataaaa ggctttggaa 3960
 ttggatcttgc acaacatgc tagatcttgc tttgggttgg gttctgggtt tgggttgggtt 4020
 ggttgggttgc ttgttcttgc ttattttttt caccatccatc atcattttttt gttttttttt 4080
 ttatccaaaagg aggcttggcc ttgttccaaat tttttttttt tttttttttt tttttttttt 4140
 tttttttttt atgaaaacgc aaccccttgggaaa aagaaaaactt tacataaaaagggg agattttttt 4200
 gtttatttttgg aaaaaggaaat cggccatggggaaa ttggaaagggg tagttttttt 4260
 ttccggcttgc ttgttgggttgc ttgttccaaat gttttttttt gttttttttt gttttttttt 4320
 aaaggccgtt cagcttggcc ttgttgggttgc ttgttccaaat tttttttttt tttttttttt 4380
 catcaccatc atcacttggcc ttccactatgtt ttatataaaa aaggccgttcc ttgttccaaat 4440
 agaaatgcctt acttggccaaat gaaaatgttcc agatataactt gttttttttt gttttttttt 4500
 agagatgaaatc aaccccttgggaaaat gttttttttt gttttttttt gttttttttt gttttttttt 4560
 gaaaatggatc ttgttccaaat tttttttttt tttttttttt tttttttttt tttttttttt 4596

SEQ ID NO: 182 moltype = DNA length = 2472
 FEATURE Location/Qualifiers
 source 1..2472
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 182
 atgaaaaacttgc当地 ttctgttgc当地 agaactgc当地 ttgggttccctt taatggtagt 60
 ttggcctcta catcagctat tgattttgg gctaccgtca tcaaagctgc aatttgc当地 120
 gccaaggatc attcttcaaca ttgttgc当地 gtaataatgc当地 gtaacgtttt gcaaggcttgc当地 180
 tttaggttccaaatccatc当地 acaaggcccttgg gttttttttt gttttttttt gttttttttt 240
 ggtttccatc当地 tttaggttccaaatccatc当地 ttgttgc当地 gttttttttt gttttttttt 300
 gcttataatgc当地 cttttttttt gttttttttt gttttttttt gttttttttt gttttttttt 360
 gaccccttgc当地 ttgttgc当地 gttttttttt gttttttttt gttttttttt gttttttttt 420
 taatggatc ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 480
 acaggccatc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 540
 ttatccatccc aaaaaggccatc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 600
 gttccatc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 660
 aaaggccatc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 720
 accggtaatgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 780
 gaagaatccatc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 840
 ttatccatccc aaaaaggccatc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 900
 ttatccatccc aaaaaggccatc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 960
 gttccatc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 1020
 aacgggttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 1080
 acatccatc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 1140
 ggttgggttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 1200
 ggttctgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 1260
 aacggccatc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 1320
 gaatttagatc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 1380
 ggttggatc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 1440
 gccc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 1500
 atcggccatc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 1560
 aagaatccatc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 1620
 caaaaaggccatc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 1680
 tacttataatc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 1740
 gatccatccc aaaaaggccatc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 1800
 taccatccatc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 1860
 agaaggccatc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 1920
 tattaccatc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 1980
 ggttggatc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 2040
 tacttggaaaatccatc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 2100
 gccc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 ttgttgc当地 2160

-continued

```
ggcgctgca aagaagccgc tgcaaaggca ggttctgccc gttcagccgc tggtgtctgg 2220  
gaatttcgggtt ccagttaacta tcaccatcac catcatcaact tggaaatctac ttcatatat 2280  
aaaaaggccgc gttccggtag tcaaaaatgc gctgaattaa agaacagact agctgttaag 2340  
tttgcacacaga acgaacaataag gaaaataag gtagaagaat tgaaaataat aaacgcctca 2400  
tttaaagaatg aatttgcac ac attggaaaac gaagtgccta gattggaaaa tgatgttagca 2460  
gaagggttctg gt 2472
```

SEQ ID NO: 183 moltype = DNA length = 2337
 FEATURE Location/Qualifiers
 source 1..2337
 mol_type = other DNA
 organism = synthetic construct
 SEQUENCE: 183
 atggaaaaagg tttgtgtcat tggtgcgtgtt accatgggtt ctggatatgc acaaggcctt 60
 gtcgaaagg gttcgaaagt tgtcttgaga gatatacagg acgaaattcg tggataggtt 120
 ttggacttca tcaataagaa cttgtcttaag ttggtaaaaat agggtaaaat cgaagaagct 180
 acaaaggtag aaatcttgac cagaatttca ggtactgtt atttgataat ggcgcgtat 240
 tggacttgc taatcgaaac agccgtgtaa agaaatggata ttaaaggaaaca aatcttcgc 300
 gatggccaa acatctggaa acatcgaaata atcttagctt caaacaccc tttatgtcc 360
 attactgaag tcgctatgtc aacaaaaaaa ccagataggaa taatggcat gcatttctt 420
 aatccagctc ctgttatgaa gttggtagaa gttttagag gtatgcacac attcctaaagaa 480
 accttttgac ctgttatgaa aaccttcaata gcaatcggtt aagatccgtt cgaatgtac 540
 gaagcttgc gtttctgtt taacagataat ttgatatacta tgatcaacga agtctgtgtt 600
 atcttggctt aaggatattgc atctgtcgaa gatattggaca aagccatggaa gtttaggtgt 660
 aatcaccctt tgggtccccc ggaattgggtt gactttttt gtttggacat atgttttagct 720
 atatcgaccc ttttgtatc cgaaacagggtt gacagttttt acagaccata taccttgggtt 780
 aagaaatatg ttagagcagg tttgtttagt agaaatgtctg tttttttt ctatggattac 840
 tctttttttt tttttttttt tttttttttt ggtttttttt gttttttttt tttttttttt 900
 gcagaaggctt ggttacattt gggtaacgct tactacaaagc agggttacta cccaaaggcc 960
 atagaatactt accaaaaaggc tttggatggat gatccatata acgtctggaa atgtttataat 1020
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1080
 gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1140
 gctttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1200
 aacgggggtt actatcaaaa agctatcgaa gattacaaaaggccatggatctt 1260
 aacaactttt aacggcggaaatctt tttttttttt tttttttttt tttttttttt tttttttttt 1320
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1380
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1440
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1500
 gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1560
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1620
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1680
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1740
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1800
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1860
 gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1920
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1980
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2040
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2100
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2160
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2220
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2280
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2337

-continued

SEQ ID NO: 185 moltype = DNA length = 2469
 FEATURE Location/Qualifiers
 source 1..2469
 mol_type = other DNA
 organism = synthetic construct
 SEQUENCE: 185
 atgatcgtaaa agccaatggta tagaaacaacatctgtttga acgctcatcc tcaaggttgc 60
 aaaaagggttgc tagaaagatca aatcgataac accaaaaaaga gaatcactgc agaattaa 120
 cggcggtctaa aaggccatcaa gaatgttttg gtcttaggtt gttccaaacgg ttatggtttgc 180
 gtcgtatgaa taacagctgc atttggttac ggtgcgcgta ccatcggtgt ttccctcgaa 240
 aaggctggta gtgaaaccaa atatggtaact ccagggttgg acaataactt ggcatattgtat 300
 gaaggcagccaa agagagaagg ttatatactt gtcactatag atgtggacgc ttcttcgat 360
 gaaatcaagg cacaagtat tgaagaagac aaaaaggaaag tgataaaattt cgatggatc 420
 gtttactctt tagcaagttc agtcagaacaa gatcttgaca ccggatataat gcataaggttc 480
 gtttgaagc cattcggtaa aactttcaca ggtaaaacag tcgatccccc caccggtaa 540
 ttgaaagaaa tatctgtgtc accagcaat gatgaagaag ctgcggccaa actgtaaaatgt 600
 atgggttgtt aagactggaa aagatggatc aagcaatgtt cccaaagaagg ttttgtggaa 660
 gaagggtgttacccatggc ttatccatgtt atgggttgtt aagccactca agttttgtat 720
 agaaaaggtt caatcggtaa agcttaaagaa catttggaaag ccaccgctca cagattaaat 780
 aaggaaaacc catctatcatc agcattttgtt tctgtataataa agggttttagt tactagcata 840
 tccgcgttacccatgtt cttttgttgc ttggctgtt tggttaaagg tataaggaa 900
 aagggttacccatgttcatggaaatccatgtt gatcaacaa atcactatgt tgatcgcgaa aagattatata 960
 agaaaaggatgttacccatgttcatggaaatccatgtt gatcaacaa atcactatgt tgatcgcgaa 1020
 ttggaagaagatgtccaaaaggcgtatcttgccttaatgg aaaaaggatccatgttcatggaaatccatgtt 1080
 gtcgtatcgttgcgtatgttgcgtatgttgcgtatgttgcgtatgttgcgtatgttgcgtatgtt 1140
 gatgtcgaaatgttacccatgttcatggaaatccatgtt gatcaacaa atcactatgt tgatcgcgaa 1200
 ggtgttgtgttgcgtatgttcatggaaatccatgtt gatcaacaa atcactatgt tgatcgcgaa 1260
 aatttgggttacccatgttcatggaaatccatgtt gatcaacaa atcactatgt tgatcgcgaa 1320
 aagggttgttgcgtatgttgcgtatgttgcgtatgttgcgtatgttgcgtatgttgcgtatgtt 1380
 tataaaacagggttacccatgttcatggaaatccatgtt gatcaacaa atcactatgt tgatcgcgaa 1440
 cccaaataacccatgttcatggaaatccatgtt gatcaacaa atcactatgt tgatcgcgaa 1500
 cccaaataacccatgttcatggaaatccatgtt gatcaacaa atcactatgt tgatcgcgaa 1560
 cccaaataacccatgttcatggaaatccatgtt gatcaacaa atcactatgt tgatcgcgaa 1620
 gaaatcgatggaaatccatgttcatggaaatccatgtt gatcaacaa atcactatgt tgatcgcgaa 1680
 ggttactatccatgttcatggaaatccatgtt gatcaacaa atcactatgt tgatcgcgaa 1740
 tttagaattggatccatgttcatggaaatccatgtt gatcaacaa atcactatgt tgatcgcgaa 1800
 cagggtgtactatccatgttcatggaaatccatgtt gatcaacaa atcactatgt tgatcgcgaa 1860
 aatccatgttcatggaaatccatgtt gatcaacaa atcactatgt tgatcgcgaa 1920
 gtttccatgttcatggaaatccatgtt gatcaacaa atcactatgt tgatcgcgaa 1980
 ggggggttgttgcgtatgttgcgtatgttgcgtatgttgcgtatgttgcgtatgttgcgtatgtt 2040
 gaaatccatgttcatggaaatccatgtt gatcaacaa atcactatgt tgatcgcgaa 2100
 aaaaaccatgttcatggaaatccatgtt gatcaacaa atcactatgt tgatcgcgaa 2160
 aatccatgttcatggaaatccatgtt gatcaacaa atcactatgt tgatcgcgaa 2220
 gtttccatgttcatggaaatccatgtt gatcaacaa atcactatgt tgatcgcgaa 2280
 agtttatataatccatgttcatggaaatccatgtt gatcaacaa atcactatgt tgatcgcgaa 2340
 gtcgtatcgttcatggaaatccatgtt gatcaacaa atcactatgt tgatcgcgaa 2400
 atccatgttcatggaaatccatgtt gatcaacaa atcactatgt tgatcgcgaa 2460
 gtttccatgttcatggaaatccatgtt gatcaacaa atcactatgt tgatcgcgaa 2520

```
SEQ ID NO: 186 moltype = DNA length = 2487
FEATURE Location/Qualifiers
source 1..2487
        mol_type = other DNA
        organism = synthetic construct
SEQUENCE: 186
atgactagag aagtgtcg agttagtggt gttagaacag ctattggta ctttgggtg 60
tctttaaaag atgttgcacc agccgaattt ggtgcattag tcgtaagaga agctttggca 120
agagccccaa ttccaggtga cgatgtccgt catgttgtct tcgttcacat tatccaaaca 180
gaaccaaagat atatgttattt gggtagagata gtcgtcaat tttttttttt tttttttttt 240
qctccctgcataaactgtcaac qcaattttttt qqttaatgtttt tttttttttt tttttttttt 300
```

-continued

gctcaaaca	tattgtttag	tgacaccgc	gttgtatcg	gtgggtgtc	tgaatctatg	360
tcaagagcc	catacttag	tcctgcgac	agatgggg	ccagaatgg	tgacgctgg	420
tttgttgc	tgatgttgg	tgcttgc	gatccatcc	atagaatcc	cattgggtc	480
actgcagaa	acgttgc	ggaatagat	atctcaag	cacaacta	cgaactgc	540
ttagatcac	acagaagac	atccgc	attaagcc	gttactta	ggatcaata	600
gttccagt	tttctaaag	tagaaagg	gacgttac	tcgatactg	cgaacatgtt	660
agacacgc	ctactattg	tgacatgac	aagttaa	ctgttccgt	caaggaaa	720
ggatctgtt	cagctgg	tgcatctgg	ttgaacgat	cagccgtc	agtctaatg	780
atggaaag	cgaaagctg	aagaagagg	ttgaacat	tgactgat	gttttttat	840
ggtcatctg	gtgtcgat	taaagca	ggtataggt	cagttctgc	tactaagatc	900
gcatggaa	gaggccgtt	acaagtct	gattggacy	taatggaa	caatggact	960
tttgcgc	aageatgt	cggttac	gccttgggt	tagatcc	taaggctcaat	1020
cctaaccgt	gtggtatct	tttaggtc	ccaatttgg	caacccgtc	cttgataact	1080
gttaaggct	tgcacaa	gaacagat	caaggtag	atgcattag	tacaatgtc	1140
atcggtgg	gtcaaggat	tgacccat	ttcgaaa	ttaaatgtc	ttgggtgtt	1200
ggttctgtt	gtgggtgt	tgggtgtt	ggtagtgc	aagcatgtt	caatttgggt	1260
aacgttact	acaacgg	tgactacc	aaggcaatc	aatattacc	aaaaggctt	1320
gaatttagac	caaataac	cgaagctt	tataatttgg	gtaatgc	ttataaaac	1380
gggtgactat	aaaacgtat	agaactact	aaaaaggcat	ttggaaat	cccttataaac	1440
cgacaaggc	ggtataatt	gggtaaac	tattataa	agggtgact	tcaaaaggcc	1500
atagaagac	accaaaagg	tttggaa	gatccaa	acttacaa	tgaaatcgat	1560
agaatttgg	gtaacgctt	ttacaa	ggtgactacc	aaaaagctat	tgaatactat	1620
caaaggctt	tagaattg	cccta	ggcttgc	ttggat	gggtatgtc	1680
tactataaa	agggtgat	ccagaagg	atcgatatt	atcaaaa	tttagat	1740
gacc	acgc	ctggat	cgtgtt	catttacaa	gcagggtgc	1800
taccagaa	ctattgaa	ttatcaa	gcattgg	tggatctt	caacagatcc	1860
agaagtgc	gttgttgg	ttctgttgg	gggtgttct	gttgttgg	tgcttctgt	1920
gttgttgg	aatataatg	ggaaatata	gggtgc	attatcc	tcctttact	1980
ttagcaca	tgcaaaatc	ggatccg	ggttat	ctgtatgtt	ttactcg	2040
aaatttagac	cacccgtt	tcaatttac	aactcaaa	gaatagatt	cgactgtac	2100
accgggttc	ctgggtcc	tgcaagg	ggtaattt	gttccgc	agccgtgc	2160
aaaaggcc	ctgcaaaagg	tggtagtgc	gggttcc	ctgtgttgg	tgaatttgg	2220
ttagcata	atcccg	atccaat	gcttat	acgttaaa	gattaaacc	2280
agattttg	atcatcc	aatctatag	tcatttt	aaatcttgc	tacataccaa	2340
aaagaacaat	tgccacaa	gggttagac	ttcaagg	tgtccgaa	agaagtctt	2400
actgaatgt	ctatattt	tgaggat	gaagat	tgtcgaat	cggtcata	2460
ttcccaaaq	caaaaaqqq	ttcccg				2487

-continued

agaactgctt caaacaaat tatgagaaga gtcttgagac aacaattcag tcattttgaa 2160
ggttctgggt 2169

-continued

aaaggcttgg	ataattttggg	taacgcata	tacaaggcagg	gtgactacca	aaaggccatc	1260
gaatactacc	aaaaggctt	ggaaattggac	ccaaataaacg	cagaagctg	gtacaatttg	1320
ggtaatgtct	actataaaca	gggtgtac	caaaaaggcaa	ttgaatatta	ccaaaaggcc	1380
tttggatatt	accctaataa	cgtgtaaagca	tggtaatcaa	tggtaaacgc	ctactataa	1440
cagggtgact	atccaaaaggc	tattttaaagac	tacccaaagg	cattgttgcatt	agatccaaat	1500
aacttgcac	ccgaaagctt	gaaaaatttg	ggtaacgcctt	actacaaaca	gggtgactac	1560
caaaaaggct	tttgcataact	tccaaaaggct	tttggatattgg	accctaataa	cgcattctgc	1620
tggtataatt	tgggtatgc	tttattataaa	cagggtgact	accgaaaggc	aatagaataac	1680
atccaaaagg	ccttggattt	agacccaaat	aacgcotaaag	catggtagat	acgttgtat	1740
gtttactata	agcagggtg	ctaccagaaa	gctatagaag	attatcffff	ggcattggaa	1800
tttggatccca	acaacagatc	tagatcgat	gggtgggttg	gttttttttttt	tttttttttttt	1860
gggtgggttg	gtgtctccag	tttattaccat	caccatcacc	atcactttgg	atccactagt	1920
tttataaaaa	aggcaggtc	agaatttttc	agaagagaaa	gaaaatataat	ggccgcgtca	1980
aaatgttagaa	acagaagaag	agaatttgaca	gataccctac	aagctgaaac	cgatcaatt	2040
gaagacgaaa	agtcgtccat	gcaaaatgtaa	atagocat	tgttggaaag	aaaggaaag	2100
tttggatcca	tttttagcccc	ttatagaccc	gcttgcaaaa	ttcttgcatg	cttttttttttt	2160
cccaaaaaaa	tgtttttttag	aggttccgc	ggtagtgcag	ccggttccgg	ttgttttttttt	2220
agtgctgtgg	ctgcagccaa	ggaaagctgc	gccaatgt	gttctgcagg	ttcagctgc	2280
ggttccgggt	atttcggt	ttcataactat	caccatcacc	atcatcactt	gaaatctacc	2340
tcattatata	aaaaggcttgg	tttccggtagt	caaaaagg	tttgcataactt	gcaaaatgtt	2400
aaagaatata	agcggaaagaa	agcccaattt	aagaatgtata	tcgttgcatt	agaaaaggaa	2460
atcgccatcg	ctgaaaacttgg	tttgcgtt				2487

-continued

tggactatgg	ccttgectga	agaagctaag	atcaaaggatg	cataccatat	gttgaagaga	3060
caaggtagat	tccaaatcaga	tccacctatc	ccagtagaca	gaactttat	tccatctcca	3120
ccaccaagac	ctaaaaatgg	ttcccggt				3147

SEQ ID NO: 191 moltype = DNA length = 2610
 FEATURE Location/Qualifiers
 source 1..2610
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 191
 atgtccgaat taagagcttt tagtgcacct ggtaaagect tggtagctgg tggttatttg 60
 gttttggata caaagtacga agcatcggtt gtcgggttgc cagccagaat gcatgcgtc 120
 gcccacccctt acgggtttttt acaaggatctt gataaggatc aagtaaaggat caagtctaag 180
 caattcaagg acgggtatgg gttataccat atatctcaaa agtcaggtt tattctgtt 240
 tccataggtt gtatgtaaaaa tccatccatc gaaaagggtt ttgcaaaatgt cttttcttac 300
 ttcaageccta acatggatga ctactgtaa agaaaacttgt tcgtcatcga tatattctct 360
 gatgacgccc atcatctca agaaagactca gtaactgaaac acagaggtaa tagaagattg 420
 tcctttcata gtcacagaat tgaaggatg ccaaaaaccgg gtttaggttcc ttcagctgg 480
 gggttagtca ctgtattgtac tacagcttta gcatcctttt cgatgttttgc tttggaaaac 540
 aacgttagaca agtacagaga agtttattcat aatttggcac aagttagccca ctgccaagca 600
 caaaggtaaaa tccggtccgg ttttgcgtt gtcgcagccg ctatgttca aatttagatac 660
 agaagattcc caccgttccat gatatgttttca ttgcccagata tgggttgc tacatatgg 720
 tcaaaaggatgg cacatgttgc tgatgaagaa gactggaaac tcacaattaa atccaacat 780
 ttgcctatgtt gtttgcacctt atggatgggt gacattaaga atggttctga aactgttaag 840
 ttggtccaaa aagtaaagaa ctggtagcat ttcatcatatgc cagaatcattt gaagatctac 900
 acagaatttag accatgttca ttccaggatc atggatgggt tgatgtttt aagacagattg 960
 catacccacg atgactactc tgatcaatc ttccggatcat tggaaagaaa cgactgtact 1020
 tgccaaaaat acccagaaat cacagaagta agagatgccc ttgttccat aagaagatct 1080
 tttagaaaaa tcaactaaaggatc acaggatgtca gatatcgaaac caccgttca aacatcttg 1140
 tttagatgtact gtcacaaaccctt gaagggttgc ttaacttgc tgatgttttgc tttttttttt 1200
 tatgtatgc tggccgtcat cactaaacaaat gatgttagact tgatgttca aacagcaaac 1260
 gataagagat ttcaaaaggat ccaatgggtt gatgttcaacc aagctgactg ggggtttttaga 1320
 aaagaaaaagg atccgtaaac tttacttggatc aaaaaggat ttttttttttgc tttttttttt 1380
 gttgggttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 1440
 tacaaggcgg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 1500
 ccaaataacg ccggaaatctt gtttttttttgc ttttttttttgc ttttttttttgc tttttttttt 1560
 caaaaggcga ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 1620
 tgggttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 1680
 taccaaaaagg ctttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 1740
 ggttacggcgtt acttataaaa ggggttgc ttttttttttgc ttttttttttgc tttttttttt 1800
 ttggaaatttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 1860
 ctttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 1920
 aacgggttgcact atcagaaggatc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 1980
 gtttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 2040
 gtttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 2100
 caccatcacc atcacttggatc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 2160
 aagggttgcact aatttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 2220
 aacagatgtcc ctttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 2280
 gaatttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 2340
 gtttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 2400
 ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 2460
 aaggccgggtt ctttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 2520
 aacggaaaacg caaaggatggatc aaggatggatc gccaaggatggatc aaggatggatc aaggatggatc 2580
 agaaatgttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 2610

SEQ ID NO: 192 moltype = DNA length = 3033
 FEATURE Location/Qualifiers
 source 1..3033
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 192
 atgacagttt ataccgttcc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 60
 ggtttaaaagg atactaaaggat gtttttttttgc ttttttttttgc ttttttttttgc tttttttttt 120
 caaaggatgttca ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 180
 ttgttgggttca ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 240
 gattttggatc aattttggatc gtttttttttgc ttttttttttgc ttttttttttgc tttttttttt 300
 caatgttggatc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 360
 agtgcggccg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 420
 agtacatcttgc aattttggatc gtttttttttgc ttttttttttgc ttttttttttgc tttttttttt 480
 ttgttgggttca ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 540
 gtacaaaatgttca ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 600
 gacatcaaaaaggatgttca ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 660
 ttgtttaaaaggatgttca ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 720
 atcgttagaaaaggatgttca ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 780
 catgcaactt gtttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 840
 ataattcttca ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 900
 ttccgtatgttca ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 960
 gtttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 1020
 caatgttggatc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc tttttttttt 1080

-continued

SEQ ID NO: 193 moltype = DNA length = 2160
FEATURE Location/Qualifiers
source 1..2160
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 193

	organism = synthetic construct					
atgactgtctg	ataataaactc	tatgccacat	ggtgccgtat	cttcatacg	taagtgggtt	60
caaaaaccaa	caccgtgaaga	tatcttggaa	gaattcccg	aaatcatccc	tttgcacaac	120
agaccaaaaa	ctagatcccg	tggaaacatcc	aacgatgaaa	gtgggtaaac	ctgtttttc	180
ggtcatgacg	agaacaaat	taattatgtat	aacgaaaaat	gcatcgatt	qgatggat	240
gacaatgc	taggtgcgg	tactaagaaa	gtttgcatt	tgtatggaaaa	catagaaaag	300
ggtttgttgc	acagagett	ctccgtttt	atattaatcc	aacagggttgc	attgttatttgc	360
caacaaagag	caacagaaaa	gatcaccctt	ccagatgttgc	ggactaaat	atgtgtctc	420
catcccttgt	gcattgtat	cgaattagtgc	tttgcggat	aatttgatgc	caaaaatgg	480
ggtgctataa	ctgctgcgt	cagaaaaattt	gatcatgat	tgggtatacc	agaagacgaa	540
accaagactc	gtggtaaattt	ccatcttca	aacagaatcc	actatatggc	tccatctaa	600
gaaccccttgg	gttgcacat	aatcgat	atctttttt	acaagattaa	tgcacaaagg	660
aacttgcacg	ttaaacccaa	cgttataat	gtcagatgttgc	tccatgttgc	ttctcttcaat	720
gatttgc	ccatgtttgc	tgaccatca	tataatgttgc	ctccttgc	caagatcatc	780
tgtgaaaact	acttgtttaa	ctgggtggaa	caattatgt	acttgtctgc	agttgaaaac	840
gatagacaaa	tccatataat	gttgcatttgc	tctgggtgttgc	gttgcatttgc	tgggtgtgttgc	900
tctgggttgc	gttgcatttgc	cgaagcttgc	tacaattttgc	gttgcatttgc	cttgcacacg	960
ggtgactacc	aaaaggcaat	acttactact	caaaaggcc	tggatatttgc	cccaaaataac	1020
gcagaaggct	ggtataattt	gggtatgc	tattataaac	agggtgacta	tcaaaaggct	1080
atgtatattt	accaaaaaggc	atggatgg	gaccctata	acgtcgacg	atgtatatttgc	1140
tgggttgc	cttattat	gcgggttgc	tatcaaaaat	ccatcgaa	cttaccaaaa	1200
gcttggaaat	tggatccaa	caacttacaa	ggccaaat	ggaaatattt	gggttacgc	1260
tattacaaac	agggtgacta	ccaaaaaagct	attgtatact	atcaaaaagc	cttgcataat	1320
gaccctataa	acgcacat	ctggatcaat	ttgggtatgc	cttaccaaaa	gcagggttgc	1380
tatcagaagg	cttgcataat	cttaccaaaa	gcattgttgc	tggatccaa	taatgcataat	1440
gcatgttgc	gacgttgc	tgcattatcc	aacgagggttgc	atcatcaaaa	ggcgatgttgc	1500
gattatcaa	aagcttgg	attggatcc	aacaacat	cttgcacat	tgggtgtgttgc	1560
ggttctgttgc	gttgcatttgc	tgggtgtgttgc	gggtgcatttgc	catattacca	tcaccatcac	1620
catcacttgc	aatccacaa	tgttgcataat	aagggttgc	ctggatccaa	caccgttcaat	1680
gaatttgc	actacatcc	agaatttgc	gaaagaaacg	cagaatgttgc	aaatgttgc	1740
gaacatgttgc	atgttgc	ggcttgcatttgc	gaatttgc	tggccgtca	caaaaatgttgc	1800
ggttccgttgc	gtatgtc	cggttccgttgc	gaatttgc	gttgcacat	tgcacggccaa	1860
gaagctgtcg	ccaaggcttgc	ttctgcatttgc	tcagctgc	gttgcatttgc	attttgttcc	1920
atgttactatc	accatcacca	tcatcacttgc	gaatccacta	gttgcatttgc	aaaaggccgttgc	1980
tctgggttgc	aaaaggcttgc	acaatcaat	aatagatgttgc	caccgttcaat	gaaagaaacg	2040
gctaagggttgc	aaaaggcttgc	cgcaatgttgc	gaaacatgttgc	atgcaccaat	ggaaaaaaacg	2100
atcgatgttgc	tggaaaaaqqa	tatttgc	caacatgttgc	tttgc	tttgc	2160

SEQ ID NO: 194 moltype = DNA length = 2403
FEATURE Location/Qualifiers

-continued

-continued

aacgcttctgt	cttggtacaa	tttgggcac	gcttactaca	aacagggtga	ctaccaaaa	1680
gctatcgaa	attatcaaaa	ggctctggaa	ctagatccaa	ataacgccaa	ggcttggtat	1740
agaaggggaa	atgtctttaa	taaacgggt	gactaccaga	aagcaattga	agactaccaa	1800
aaaggccctt	aactggatcc	taataacaga	tctagaacgc	ctgggggtgg	ttgttctgggt	1860
ggtgttgggtt	cttgggtgggg	ttggtctttt	ggttaacacta	taaaacata	cgagtggtag	1920
aataagtcta	tttctagaga	taaggccaa	aagttaactac	ttgacacccgg	taaagaagggt	1980
gcttttatgg	tttagatgtt	tagactcca	ggtaacttata	cagtctctgt	attcacaaaag	2040
gctatcatct	cagaaaaacc	atgtatcaag	cattaccaca	tcaggaaac	caagactct	2100
cctaaaagat	attacgtggc	agaaaaatgc	gtttttgtt	caatcccaact	gttggatcaa	2160
tatcatcagt	acaatggtgg	ttggtttgggt	actagattga	ggtacccctgt	ttggcgggtgt	2220
agcgcagggtt	cggtctgcagg	atcaggcgaa	tttggttccg	ccgaggccgc	tgcaaaaagaa	2280
gcccgtgc	aggctggatc	ttcaggctca	ccgcgtgtt	ttggggaaat	ttgttctgggt	2340
tctcatccct	ggtttttccg	taaaaatccca	agggcaaaagg	ccggaaataat	ttgttctaaat	2400
caaagacacg	acgggtgcatt	tttgataagg	gaaagtggaga	gcgcacccctgg	tgacttttcg	2460
tttgttctgtt	aattcggtaat	tgatgtccaa	catttcacgg	tatttgagaga	ttggtctgtgtt	2520
aaatactttt	tgtgggtctt	aaagttcataat	tccttgaac	aattagtggaa	ttacccataga	2580
ttcaacttcgg	tttagtaggaa	ccaaacagat	ttcttggagag	atatacgaaaca	agttccacaa	2640
cagcctacag	gttctggaa					2658

```

SEQ ID NO: 197          moltype = DNA    length = 2400
FEATURE                Location/Qualifiers
source                 1..2400
                      mol_type = other DNA
                      organism = synthetic construct

SEQUENCE: 197
atgggttttgcctcacaacacaatccaaacttcttgcatacaccatcttgcgtgaac 60
atgggttttgcctcacaacacaatccaaacttcttgcatacaccatcttgcgtgaac 120
atggatagctacaacaatttccatcgaaaattttgttagtcaagagttccattgcaa 180
aataagtgtccgaatctttgtctatcgctaagaactcaaatttgcgtgc aactacaat 240
cagacggaaacaccgtggatccatcgccaaatatttgcgttggt 300
aaagcttggatggaaatgtcaaggaccatcaataataggcttcacccctgtgtgttgc 360
ggtttgtttgttgcataaacacaatatttgcgttggttgcgttgc 420
aaggcatttttcttgcgttgcatacccttgcgttgcaccgactattaat 480
caaatactatgatattgcacatcgacagaattaaataaggcccgatttgcactagttcggt 540
gaaatcccgatataacttgcatggattatgtcaatcatgttcgccttgcgtttaatc 600
atccacaatataatgaaatggatccatgttgccttgcatacttgcgtactatgttgcgtt 660
gggtgttagtatatccatgttgccttgcataatgttgcgttgcgttgcgttgc 720
ttactaaattttcttgcacatcatcatcaca aacttgcacttcttactacgc ttctagatgt 780

```

-continued

gctttgggtt	tgccattcga	attaagacca	tctttatcat	tttgcgtggc	ttttatgaaa	840
tcgtatgggt	ctgcattggc	cttgattaaa	gatgcacatcg	acgttgaagg	tgacacaaaa	900
ttcggcatca	gtaccttggc	tagcaactac	ggttcttagaa	atttgactct	gttttgttgc	960
ggtatcgat	tgttatacct	cgtggcagcc	attttacccg	gtatcatttg	gccacaactg	1020
tttaacatgt	atgtcatgt	acttacccggc	gcaatatgg	ccttcgttgc	gatcttgcag	1080
acgagagatt	ttgtttaac	taattatgac	cctgaggcag	gtagaagatt	ctacgaattc	1140
atgttggaaacg	tgtactacgc	tgaatatttg	gtttacgtct	ttttaatgg	gtcttgttgc	1200
ggttgttgcgt	ttgttgggtt	ttcttgggtt	ggttgttagtgc	ctgaacatgt	gtacaaacttca	1260
ggcaacgcat	actacaacga	gggttactact	cagaaggcaa	tttggatattt	ccaaaaggccc	1320
tttagaactag	acccaaacca	tgccgaggct	ttgttataact	ttggcaatgc	tttattacaaa	1380
cagggtgact	atccaaaagc	tatagaatat	tacccaaagg	cacttgagct	ggaccctaa	1440
aatgtcggaa	cctgttataa	ctttagggat	gctttatata	agccgggtgc	ctatcggaa	1500
gccatcgagg	actaccaaaa	ggctttggaa	ctggatccaa	aaatcttgc	ggctgtaa	1560
tggaaagaat	ttgggttacgc	ttactatataa	cagggtgact	atcagaaaago	aatatataatc	1620
tacccaaaag	cccttagaact	tgaccctaa	aatgcctctg	cttggtacaa	cttgggttaat	1680
gtttactata	agcagggtg	ctacccaaa	gctatcgat	attacccaaa	agcactggaa	1740
ttggacccaa	acaatgc	ggccctggat	agaagaggt	acgcctacta	caaacagggt	1800
gactacccaa	aggttattga	agattacca	aaggctctgg	aatagatcc	ttaaacacaga	1860
tcttagatccg	ctgggttgggg	ttgttctgtt	ggtgggtgtt	ctgggttgg	ttgtgttct	1920
ggagaatact	tttagactct	gttcgtatc	aacggtaac	atgaagagg	cttgccttt	1980
aaggaaagggt	acattttgag	attcaggac	aaacccaaag	agcaatgttg	gaatgtgtaa	2040
gatttcgtgg	gtttaaaagg	aatgttccct	gttccctat	tcgaaaggta	ccgtccgcaga	2100
ggttccgcgt	caggatctgg	cgaatttccgt	tcagccgggg	ccgctgtcaaa	agaagccgc	2160
gcaaaaggctg	gaagtgcagg	cagcgcgcgt	ggttccggag	aatttggtag	tttggattaaa	2220
catatggatg	cggaaatgtt	atttcgtat	actgttaact	ccaaatcttgc	actgaatttc	2280
aaaggcagggt	cggttattttt	tttgcgtat	agaattata	aggactgtt	ggaaaggtaat	2340
tttagagggtt	ctactgttgc	attcccaactt	tcttttgtga	aaatcttgc	ggggctcagggt	2400

SEQ ID NO: 198 moltype = DNA length = 1641
FEATURE Location/Qualifiers
source 1..1641
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 198

	organism	synthetic construct	60			
atgaaatgt	gcacttttc	tttctgggtc	gtttgaaga	tcattttctt	tttcttttct	120
ttaataatcc	aaacttcgat	cgccaaatcca	agagaaaaac	tcattaaatgc	ttttccacca	180
tatcatctca	ataaaggccac	ggatggaa	ctggtatata	ctcgaacaaa	ccctatgtac	240
atgagcgtgc	taaactcgac	aatccataat	ttgagattca	cttcggatata	tacacccaaa	300
ccatttagta	tcgtgacacc	ttctcatgtt	tcacacattc	aaggaacca	actatgtctc	360
aagaaggctc	gttgcagat	tagaaacagg	ttctgggtgc	atgatagtga	aggatgtcgc	420
tatcatcgatc	aaggttccat	cgttatgc	ttttggagaa	acatgggtc	tatcaaataa	480
gacgttcaact	cacagacggc	ttgggtcgag	gcagggtcca	ctttggggaa	agtttactac	540
tgggtcaacg	aaaagaatga	aaatttgtct	cttgctgcag	gttactgtcc	aactgtctgc	600
gctgggtggc	atttttgggg	ttttgtgttt	ggacccctta	tgaaacta	cgggttggc	660
gctgtataata	tcatttgcgc	acatttgtta	aatgtgcacg	gttaaagtct	agatagaaag	720
tcataatgggt	aaagattgtt	tttgggcattt	agaggggtgt	gtgtgttata	ctttttgtata	780
atcgtagatc	ggaaaatattag	attttgttgc	gtcccaaaat	ctacaatgtt	ctcagttaa	840
aaaattatgtt	aaatccatga	gtctggtaa	ttttgtgtata	agtggcaaaa	cattcgctta	900
aaagtatgtca	aggactgtgt	gtatgttgc	cattttatca	cgaaacacat	cattgttataa	960
cagggtaaaa	ataagacgc	aatacacacc	tactttctt	cagttttctt	gggtgggttt	1020
gatcccttag	ttggatttgat	gaataagagt	ttccctgttgc	ttgggtattta	aaaaactgtat	1080
tgttagacata	tgagctggat	cgacacaaatc	atattttata	gttgggtttgt	caactatgtat	1140
actggaaact	tcaaaacaaa	aatcccttgc	gatagaatgt	ccggacaaaa	tggcgttttcc	1200
aaaattaaatgt	ttggactactgt	tttttttttttt	ataccggat	cattttttttt	gttggatccctt	1260
aaaaaaactgt	atgaagagga	tattttttttt	ggaatgttgc	cattttatcc	atccgggttt	1320
ataatgttgc	aaatctccga	ggatgttgcata	ccattttttcc	atagatgttgc	tatctttgtat	1380
gaacttgtgtt	acatatgttc	ttttttttttt	ttttttttttt	ttttttttttt	ttttttttttt	1440
atccggaaac	tctataaattt	catgacttct	ttttttttttt	ttttttttttt	ttttttttttt	1500
ttttaatttaca	gagattttgc	cataggttata	ttttttttttt	ttttttttttt	ttttttttttt	1560
caacggaaac	ttttttttttt	aaatgttactt	ttttttttttt	ttttttttttt	ttttttttttt	1620
aaacgttttag	ttttttttttt	ttttttttttt	ttttttttttt	ttttttttttt	ttttttttttt	1680
agacataggc	acgggtttttt	t				

SEQ ID NO: 199 moltype = DNA length = 1644
FEATURE Location/Qualifiers
source 1..1644
mol_type = other DNA
organism = synthetic construct

```

SEQUENCE: 199          organism = Synchetaeon celatae
atgaactgt a gcaacttttc ttttgggtc gtttgcaga taatattttt ctttttgccc 60
ttaaatatcc a aatcgatcg ccggaaaccc a caggaaact tttaaagtgc ttctctgag 120
tacatccccca a caaaaaacccg a taaccctaa g ttatataata cacaacatcg tcagctgtc 180
atggcgatgt t gaaactcgac cattcaaaa ttgaggatca ctctcgacac tacacctaag 240
cccttggtc a tagtaatccc ttctaaatgc tcacatatac aagttcttat ctgtgtctt 300
aagaaaaggttt gtttgcagat tagaacaagg tctgggtgtc acgatgcaga aggtttatcc 360
tatatttagtc a aatgcataatc ttggcatatgc gatggagaa atatgcataatc tgtaaaatgtt 420
gacatacatac c acagactgc ttgggtggaa gcagggtgcac cattggggaga gggttactac 480
tggatcaacg a agatgcaga aaatcttgc ttgggtgtgc tgacttgc cacaatcggt 540
ttgggtgttc attttctgg tgggtgttat ggagcttaa tgagaaacta cggttgggt 600

```

-continued

```

gcagataata tcattgacgc acatttgggt aacggttgatg gtaaaggttct tgacagaaaa 660
tcatgggtg aagatttgtt ttggctatc agaggtggg tggtgaaaa tttcggtata 720
atccgcgtt gcaaaaattaa gttgggtgtc gtaccttagca aagctactat tttctctgtc 780
aaaaagaaca tggaaatcca tggtttagta aagttgttta ataagtggca aaacatcgca 840
tacaagtacg ataaggattt gatgttacc acgcatttca gaactaggaa catcacagat 900
aaccatggta aaaataagac tacagttca ggtacttctt ctcaatttt ctgggtgg 960
gttggatttc ttgttggattt gatgaataag tcattcccaag aactgggtat taaaagaca 1020
gattgttaagg aactgagctg gatcgacacc acgattttctt atagtgggtg ggttaattac 1080
aacaccgcoca acttcaaaa gggaaatcttg ctggatagat cgcgtggtaa aaagaccgct 1140
tttcttattt aacttgcata cgtttagaaaa ctgtatccctg aactgtcaat ggttaagata 1200
ttggagaaagc tggatcgaga ggaatcgccg gtggcatgt acgttttgc tccatagcg 1260
ggtataatgg atgagatctc cgaaatgtccg ataccatttc ctcatagagc tggtatcatg 1320
tatgaattttt ggttacaccgc tacgtgggg aqgcaagaqg ataacgagaa acacataaac 1380
tgggtcatgt atgtatcaa cttaatcata cttcaatcggtt ctcaagacc aagattggca 1440
tatttgcatac acagagattt ggacttggg aaaaccaacc cggaaatctcc aaataactat 1500
acgcaagcaa gaatttgggg tggaaatgtc ttccgtttaaa atttcaacag attgttgaag 1560
gtttaagacaa aagccgatcc aaacaacttc tttagaaacg aacaatctat tccaccattt 1620
ccaccaagac atcatgggtc cgcc 1644

```

```

SEQ ID NO: 200 moltype = DNA length = 8232
FEATURE Location/Qualifiers
source 1..8232
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 200
atgtcagaag agtccttatt tgaatcttc ccacaaaaga tggagttacga aatcactaac 60
tactctgaga gacatcaca attgccttgc cacttcatcg gtttgaacac agttgacaaag 120
ctggaaagatg ctccatttgc agatttcggtc aagtccatg gtggtcacac cgttaattgt 180
aaatgttgc tagctaaca cggtatcggtc gcagtcaagg aaatttagtgc tggttagaaag 240
tgggcattatg aaaccccttgg tgacatgata acgggtccatc tggtagatgat ggcacactt 300
gaagacttgg aggccaaatgg tggatataatc agaatggccc atcaatataat tgaagtttca 360
gttggtacaa ataaatataatc ctacgctaat gtcgacttac tagtagatat cgctgaaaga 420
gcagacgtgg atggccgttgc ggctgggtgg ggacatgttgc cggaaaaccc tttgttaccc 480
gaaaatgttgc ttcagatgttgc gggaaatgttgc atttttttttt gtcggccatggg aaatgtcaatg 540
agatcatttttgc tgacatgttgc atccatgttgc atcggtccatc aatcggccaa agttccatgt 600
attcccttgc cccggccccgg tggatgttgc gtttgcattt atgaaaagac tggtttgg 660
tctgttagatg acgatataatc tcagaaggga tggatgttgc caccatgttgc tggtttgc 720
aaaggcttgc gaaatgttgc cccatgttgc atcaaggatc cagaagggttgc tgggtgttgc 780
ggatgttgc gggatgttgc aagagatgttgc ttcatcgctc tggatgttgc agccgtcaat 840
gaaaatccgc gtttccatgc ttttataatc aaactgttgc gaaaggccaa acatgttgc 900
gttcagatgc ttgttgcacca atacggcactt aatattttctt tggatgttgc agattgttgc 960
gttccaaatgg gacatcataatc gattatgttgc gggccaccc tcactatagc aaaagccgaa 1020
acatccatgc agatggaaa ggcacatgttgc agatgttgc aatttgggttgc atatgttgc 1080
gtctggaaatgc tccatgttgc gtcacatgttgc aatcgttgc ttttgc 1140
aaccatgttgc tacaatgttgc gcaatgttgc aatcgttgc ttttgc 1200
gtctggaaatgc tccatgttgc gtcacatgttgc aatcgttgc ttttgc 1260
ctgttgcacca tggatgttgc cactgttgc gagatgttgc tggatgttgc aatcgttgc 1320
gcacatgttgc aacaaatggg accaatccca aagggttgc ttttgc 1380
tccatgttgc ttttgc 1440
agaatgttgc tcaatgttgc gggatgttgc ttcgttgc acaacgttgc aatcgttgc 1500
tttttgc 1560
aaccatgttgc tggatgttgc ttttgc 1620
aaatgttgc ttttgc 1680
tggatgttgc ttttgc 1740
gttcatgttgc ttttgc 1800
atagaaatgc tggatgttgc ttttgc 1860
gttggatttgc ttttgc 1920
aggatgttgc ttttgc 1980
gttgggttgc ttttgc 2040
gcacatgttgc ttttgc 2100
actcaatgttgc ttttgc 2160
catataatgc ttttgc 2220
gtttcttgc ttttgc 2280
gttggatgttgc ttttgc 2340
tttggatgttgc ttttgc 2400
aaggatgttgc ttttgc 2460
attatgttgc ttttgc 2520
tcacatgttgc ttttgc 2580
caaatgttgc ttttgc 2640
tttggatgttgc ttttgc 2700
gttggccatgttgc ttttgc 2760
catgttgc ttttgc 2820
aacgggttgc ttttgc 2880
gacttggatgttgc ttttgc 2940
ttaatccatgttgc ttttgc 3000
gcaatatgttgc ttttgc 3060
gttggatgttgc ttttgc 3120
acagagatgttgc ttttgc 3180
aaccatgttgc ttttgc 3240

```

-continued

gtatgttttgc acgttttact acagtttcta actcatcaag atcctgttgt caca	3300
ccgcctcaag tctataataag aaggccat aagacatcata cttcggtga cattaggta	3360
cacgaaggcg tgacagtcc aatcggtaa tgaaatcc aattgcctc cgccggctt	3420
agtacaccttcc caacggtaaa gtcggaaatg ggtatgaaca gagctgttgc tggttctgtat	3480
tttagctatg tggcttaatcc gcaatcatcc ctcttaaagaa aaggatctt aatggctgtg	3540
gaccatttg acaatgttga tggaaatcc tctcaatctt aatggatgtt tccaagacac	3600
caaaggact cgaatggtcc cgccggat aggtctggat ttccagcaag tttaaagcaac	3660
gtagccaatg tgggtgttgc ttccactgtggat gggtttgaaa gtgaagagga aatcttgg	3720
agattgagag aaattttggaa ttcaacaaa caagaatgttga ttaatgttca catcagaagg	3780
atcacattca tggctgggtt taaatgttgc agttaccatc agtactacac cttaatgtt	3840
cccaactaca acggaaacga aactatcata catatcgaa ctgccttagc ttccaaattt	3900
gaactggta gattgtcaaa ctcaacatc aagccaattt tcactgttataa cagaacatc	3960
catgtgtacg aacgtgttcc aaagacatcc ccattgttataa agagatttt caccagaggc	4020
atcatttagga tagatgttgc acatgttataa aagatgttgc aagatgttgc 4080	
gctaacatg taatgttca catcttagat aattttggaa ttaccgcac gtcgaactct	4140
gatttgaacc atatctttat taacttcatc gcagtgttgc acatatctcc tgaggatgtt	4200
gaagctgtcat ttgggtgtt cttggaaaga ttccggtaaaa gattgttgcg attggatgtc	4260
tccagtgttca aataatgttca catatgttgc gatccacaaa ctgggtcccc tggatgttgc	4320
agagctgttca tcaataatgttctgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	4380
aagaatgttca aagggtgttgc ggtattcaag agttgggttgc aaccggctc gatgcactt	4440
agaccaatttgc caacaccata tcctgttcaaa gaatgggttgc aacctaagaa atacaagcc	4500
cacttaatgg tgactacata ctgggttgc acatgttgc ttcccgaaat tggatgttgc	4560
tctcaatgttca aataatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	4620
gaactaatgttca aagggtgttgc aacccatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	4680
gcccattggca tgggtgttgc ttcaatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	4740
caatttgcgttgggttcaaa cggatcatcaccatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	4800
gatggatgttcaaa tcaataatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	4860
ttggccgttca aatttgcgttgggttcaaa tcaataatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	4920
gttgcgttca aacggccgtc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	4980
gagggtatgg aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	5040
gttattatgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	5100
ggcggtggatgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	5160
gatatttttcaatgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	5220
agatttgggttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	5280
gttattatgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	5340
caaatcatgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	5400
gaaaagatgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	5460
tttggaaacttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	5520
tatgtatgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	5580
ttcgataaagggttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	5640
agagctatgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	5700
aatttgcgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	5760
ggtcaatgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	5820
aacgggttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	5880
agggatgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	5940
tacaaggacatgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	6000
gttgcgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	6060
agagcgtgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	6120
ctggatacttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	6180
aagtcttttgc aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	6240
gaatttgggttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	6300
tccagaatgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	6360
tttttttttgc aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	6420
catcaatgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	6480
gcatcgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	6540
acactggatgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	6600
aagatttgggttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	6660
agcaactgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	6720
tctgggttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	6780
gttattatgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	6840
ctggatcttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	6900
gacttacaaa aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	6960
gaggcatgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	7020
gaagatttgcgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	7080
aacttagtta aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	7140
agggttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	7200
tataaaggatgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	7260
cctaataatgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	7320
aaaaaaatgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	7380
tccgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	7440
catatgtatgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	7500
tttgcgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	7560
gcacaatttgcgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	7620
aatgttgcgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	7680
atagaggatgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	7740
tttgcgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	7800
gttcttcata aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	7860
gttttagtca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	7920
ggcgttca aacatgttgc acatgttgc gatccacaaa ctgggtcccc tggatgttgc	7980

-continued

ggtagtgcgt	ctggtagcg	agaatttgg	tctgggtc	tggctactcc	tggttcagaa	8040
aacgttcac	caagagaacc	attgattgc	acagccgt	agttctgc	gaactctaga	8100
gttagacaat	ctccattggc	aactagaaga	gcattttga	aaaagaagg	tttgcaccac	8160
gaggaaattg	atatggctt	ccaacagtcc	ggtaactgc	ccgatgaacc	atcttcattg	8220
tggggaaatg	gc					8232

SEQ ID NO: 201 moltype = DNA length = 10869
 FEATURE Location/Qualifiers
 source 1..10869
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 201

atgggttc	ctggttc	c	tgcagg	ttt	ggtaattc	gttccgc	ttgtccgt	60
gttctgg	aatttgg	t	tgctgg	ta	gcagccgt	ctgg	gaaatt	120
catcaccatc	accatcactt	g	aatctact	t	attataca	aaaagg	ttccgt	180
gccagaaacg	cttaactgag	aa	aaaaatt	ct	gattga	agaaagataa	tttgcatt	240
gaaaagatg	aacaaaat	gg	aaaaggat	atc	gcta	tgagatga	aatagcaaga	300
ttggaaaatg	aagtgc	t	catgaa	aca	gg	gtatgc	ccgg	360
gaatttgc	aagccgtc	aa	ggd	aa	gcaaa	cagg	tttc	420
gtctgg	gt	ta	ttactat	ca	catcaccatc	atcactt	gat	480
ttatataaa	aggccgtt	c	ggtagt	aa	ttgg	ttgtc	att	540
tcattggaa	acgaaacg	a	actt	gaa	aa	gaaaact	tataa	600
gttacttag	aaaaggaa	a	gcaaa	attt	gg	aa	taga	660
agtgac	gtatgg	at	tcggt	ttt	gtcaag	ct	gg	720
aaagaagccg	ctgtaa	ag	ctgc	gg	aa	actg	ctgc	780
tcgggt	at	gg	ccgt	ttt	catc	accac	accact	840
ttatacaaga	aagctgtt	c	ggtagt	ca	aa	gg	tcgt	900
gttaagt	tgca	ac	aa	ttt	ttt	ttt	ttt	960
gttactt	aaaacgatt	ggca	actt	ttt	gaa	aa	acg	1020
gttgc	taag	ttt	ctc	gtt	ttt	ttt	ttt	1080
aaggaa	ccgt	cc	gg	ttt	ttt	ttt	ttt	1140
tactat	cacc	at	cc	ttt	ttt	ttt	ttt	1200
ggtagtaatg	aagttaact	at	ttt	ttt	ttt	ttt	ttt	1260
tacttagaa	aggaaatc	ttt	ttt	ttt	ttt	ttt	ttt	1320
gtctataa	aaagt	ttt	ttt	ttt	ttt	ttt	ttt	1380
gcccgt	ttt	ttt	ttt	ttt	ttt	ttt	ttt	1440
ggttccgc	gtt	ttt	ttt	ttt	ttt	ttt	ttt	1500
aatccac	catt	ttt	ttt	ttt	ttt	ttt	ttt	1560
gcccgt	tg	ttt	ttt	ttt	ttt	ttt	ttt	1620
gcagg	ttt	ttt	ttt	ttt	ttt	ttt	ttt	1680
gctcctt	cc	ttt	ttt	ttt	ttt	ttt	ttt	1740
gaatttgg	ccgc	ttt	ttt	ttt	ttt	ttt	ttt	1800
gaagctgt	caa	ttt	ttt	ttt	ttt	ttt	ttt	1860
tcaaa	ggta	ttt	ttt	ttt	ttt	ttt	ttt	1920
ggttctgc	gtt	ttt	ttt	ttt	ttt	ttt	ttt	1980
gcagccaa	agg	ttt	ttt	ttt	ttt	ttt	ttt	2040
accactt	tc	ttt	ttt	ttt	ttt	ttt	ttt	2100
tctgc	cag	ttt	ttt	ttt	ttt	ttt	ttt	2160
aagga	act	ttt	ttt	ttt	ttt	ttt	ttt	2220
tcaat	ttt	ttt	ttt	ttt	ttt	ttt	ttt	2280
ttataca	aa	ttt	ttt	ttt	ttt	ttt	ttt	2340
gcatt	taa	ttt	ttt	ttt	ttt	ttt	ttt	2400
gaaagat	ttt	ttt	ttt	ttt	ttt	ttt	ttt	2460
tcagcc	ttt	ttt	ttt	ttt	ttt	ttt	ttt	2520
gctaagg	ttt	ttt	ttt	ttt	ttt	ttt	ttt	2580
caccatc	actt	ttt	ttt	ttt	ttt	ttt	ttt	2640
ttagca	at	ttt	ttt	ttt	ttt	ttt	ttt	2700
aaaaagg	aa	ttt	ttt	ttt	ttt	ttt	ttt	2760
ggttc	ag	ttt	ttt	ttt	ttt	ttt	ttt	2820
gaagctg	ctaa	ttt	ttt	ttt	ttt	ttt	ttt	2880
tccgc	gc	ttt	ttt	ttt	ttt	ttt	ttt	2940
cctgtt	cata	ttt	ttt	ttt	ttt	ttt	ttt	3000
gaattc	aa	ttt	ttt	ttt	ttt	ttt	ttt	3060
gcagg	ttt	ttt	ttt	ttt	ttt	ttt	ttt	3120
ttggaa	aa	ttt	ttt	ttt	ttt	ttt	ttt	3180
gcaga	agg	ttt	ttt	ttt	ttt	ttt	ttt	3240
aaaggcc	ttt	ttt	ttt	ttt	ttt	ttt	ttt	3300
atttca	atc	ttt	ttt	ttt	ttt	ttt	ttt	3360
tcag	ctt	ttt	ttt	ttt	ttt	ttt	ttt	3420
ggtt	cac	ttt	ttt	ttt	ttt	ttt	ttt	3480
cattacc	tc	ttt	ttt	ttt	ttt	ttt	ttt	3540
tctgg	ata	ttt	ttt	ttt	ttt	ttt	ttt	3600
cgca	agg	ttt	ttt	ttt	ttt	ttt	ttt	3660
ggttt	ttt	ttt	ttt	ttt	ttt	ttt	ttt	3720
ggta	ac	ttt	ttt	ttt	ttt	ttt	ttt	3780
gaggcc	cc	ttt	ttt	ttt	ttt	ttt	ttt	3840
gttgt	aca	ttt	ttt	ttt	ttt	ttt	ttt	3900
gttgtt	cc	ttt	ttt	ttt	ttt	ttt	ttt	3960
ggcc	ccaa	ttt	ttt	ttt	ttt	ttt	ttt	4020

-continued

gtggcagtg gcgaatttgg ttcttcatat tataccatc atcatcatca cttggaaatct 4080
acttcattat acaagaaga aggttcccg tctaaaaggaa ttgtctactt aagaaagaaa 4140
atcgggctt tgaagaaga taatgcctaaat ttagaaaaag atattgccaa cttggaaat 4200
gaaatcgaaa gattaattaa ggaatattaa acatggaaa acaaagtgc acatcatgaa 4260
caagggttcg ctgggtccgc tgccgggtcc ggcaattt cagaagccgc cgccaaaggaa 4320
gccgcaccca aagctggtag tgccgggtt gcccgtggc ctggcgaattt ttcttactat 4380
catcatccac atcacccactt ggaatctact tcattatata aagaaagcccc ttccgggtt 4440
aacttggtag caactttaa atctacagcc gctgttttag aaaaatggaaat ccatgttca 4500
gaaaaaaaggaa aggaaaaggaa gagaaggaa aaggaaacaat tattaaataa gtttagaaagcc 4560
tacaagggtt cagcagggtc cgccggcaggc tcaggcgaat ttgggtctgc agaaggcgct 4620
gctaaggaaag ctggcgcaaa ggaagcgct gctaaggagg ccgtcgcaaa ggctgggtt 4680
gttgggtccg ccgggggcgc tggagaattt ggttccgtt tgggttgcgat cgccgtgt 4740
tatgaacctc cacccatcaa taatcgaaagaa gcttttaggt ccgtcggttc cgctcgagg 4800
tccggcgagt tgcgagaagc cgccggcaaa gaagccgcg ctaaggcagg tagtgcgg 4860
tccggcgctt cgcgtggaga attcgagaa ttgttgcgtt accatctta cgctcaactg 4920
caaaatttgg ataaagctg acaagggttcc gcccgggttgc caggggatc tggggaaattt 4980
ggttctgcag aagctggccgc caaaagaatg cgacgttcaaaag aagccgcgc caaagaagct 5040
gctgctaagg ccgggtctgc tgggttctgc gcaggatctg gggaaatttgg ttccaaagaaat 5100
actaaagata tgaacttgcg taaccctgt tacagaaaaa atcacaagaaga aagaatgtca 5160
gcccgttccg ccggcggtt cgggttgcgat cgacggatctg ccgtctaaaga ggctggcg 5220
aaggccggta gtgtcggttgc tgccgggttgc tccggaaat tcaatctt cgcctccaca 5280
tggattgaaa acaaattata cggcatgtca gaccctaaat ggggttctgc aggttctgc 5340
gggggatctg gtgaatttgc ttccggcaga gcccgcgc gggaaaggccg tgcaaaaggag 5400
geccgtgcgc aagaagcgcg tgcgttgcgtt cctccaaagac catctgtca ctgttgcattt 5460
gaattttgtt gtgttgcgat taatttccca cctccaaatg tggctccaaa aaaaagacaaat 5520
tctgtccat ctgggttcagc aggttctgcgat gctgggttgc gtaatttgc cgaagcgct 5580
gccaaggaaag ctggcgccaa ggggggcgtt gcaagggttgcg ctgggggttc tgggtgaaatc 5640
actcaaaatg cttaaacccaa accttgcgtt cctccaaagac catctgtca ctgttgcattt 5700
ggttccggccg gtccgcgcg tgggttctgcgat tggccggatcc tccgtcgagg ccggccgtt 5760
aagcgccgcg cttaaaggaggc agccgttcaaa gaggccgcgc ctaaagcagg ttctgcgg 5820
tcagcagcag gtgttgcgttgc attttgcgtt acatgtaaag aaaaagaaaga aacccaaagaa 5880
gaagttttt gtgttgcgttgc taccacttgc ggttgcgttgc ggttgcgtc ggggttctgcg 5940
gaattttgcag aagcagctgc taaggaaaggc gcccgcgc gggccaaagg ctgggttctgc 6000
cgagggtctg gtgaatttgcg tggtaatgtc tctgtgttactt aaaaatggaaat ctcaatctt 6060
gttagaactt attcatgcggc ttctggccgtt agtgcagccg gctctgggaa attcgggttcc 6120
gcaagacgcg ctggcaaaatg aaggccgtcga aaaaagccgcg ctggccaaagg ggctggcg 6180
aaggcagggtt ccggcggttgc tgccgggttgc tccggaaat tgggttgcgtt tctatccatc 6240
catcatcacc accatctggaa atccaaatgtt tttatataaaga aactctggttc tgggttccaaat 6300
aaggtagctc aattttggaaa tagagggtca tacaaggtaa atggaaacgc taattttggaa 6360
aacatagtagt caagatggaa aacatgttcaat gcttaattttgg aaaaagacat ccacaaatgg 6420
gaaaaggatgt tagctaaatg gggaaagatgt gttgtctggat gtttgcgttgc tagtgcgc 6480
ggctctggccg aattctggatc agcttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 6540
gggttctgcag cggggcttgcg tgaatttttgc tattatcacc atcatcatca tcacttggaa 6600
tccaccatgttatacaagaagaa agccggctctt ggttcaaaatc tctgttgcgttgc ttttgcgttgc 6660
tacattcaag aattttggaaa aaaaacgtt gattttggaaa atttgcgttgc ttttgcgttgc 6720
tttgcgcgaaatggatc atccgttgcgat ttttgcgttgc ttttgcgttgc ttttgcgttgc 6780
tcttgcgttgc gatctgggaaat ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 6840
aaagaggctg cagctaaaggaa agcttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 6900
tccggcgatgtt ttgggttccatc tgatgttcc ttttgcgttgc ttttgcgttgc ttttgcgttgc 6960
tcttgcgttgc atccgttgcgat ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 7020
geccgtgtca aagaggctgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 7080
gaattttgtt ccccaatgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 7140
agaagaggatgt tggttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 7200
gccaaggaggatgt tggttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 7260
gttgggtccg cggcaggctc cgggggtttttttt ggttcttccatc ttttgcgttgc ttttgcgttgc 7320
caacttggatc ccacaatgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 7380
catgttccatc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 7440
aattttggaaa gagaatgttgc tagatgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 7500
tcggccgttgc gcttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 7560
gggttctgcgttgc gtttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 7620
caacttggatc ccacaatgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 7680
tttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 7740
catgttccatc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 7800
aattttggaaa gagaatgttgc tagatgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 7860
gttggcgggttccatc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 7920
gettggcgatgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 7980
caacttggatc ccacaatgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 8040
tttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 8100
ggttagtgggttgc aatttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 8160
gacttgggttgc atgttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 8220
gaagcttgcgc ccaaggaggatgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 8280
gcccgttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 8340
catcatcacc acttggatc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 8400
gttggaaatcat taaaacaaaatgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 8460
gatatttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 8520
gcccgttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 8580
gcaacttggatc ccacaatgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 8640
gatatttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 8700
atggcaggccgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc 8760

-continued

acagatcaat	tagaagacga	aaaatcgct	ttgcaaaccg	aaatcgcaaa	tttgggtgaaa	8820
aaaaaaagaaa	aattggaaat	cattttagca	gcccatagac	cagcttgcaa	aatacctgtat	8880
gacttgggtt	ttccagaaga	aatgtctta	gaaggatgt	ccggtagtcg	cgctggctca	8940
ggtaattttgc	gtatgcaga	agtcggcg	aaaagagccg	cagctaaga	agctggccgc	9000
aaagggcgcg	ccgcaaggcg	aggctcaga	ggttcaggt	ccgggtccgg	ggaaatttggt	9060
tcattccaaa	tgccagctga	cactcctca	cctgcatt	tgccacactga	agatctat	9120
acaggtagtg	ccgggtctgc	tgccgggtt	ggcgaattcg	ctgaagccgc	tgctaaaggag	9180
gctgcagctt	aggccggctc	tgccaggttc	gctgcagggt	cagggttaat	tgaagagaaa	9240
tctaacaag	aaccacttcc	acccatgtaa	gatccatata	ggggtatgg	tggtttgc	9300
ggtagtgccg	ccggcgtcagg	cgaatttgg	tctgcggagg	ctgctcggaaa	ggaaagtgcgc	9360
gccaaggaa	ctggcgcaaa	agaggctgt	gccaaggccg	gttcagcagg	ttcagcagct	9420
gggtccgggt	aatttggttc	cagtattat	caccacatc	atcaccaactt	ggaatttacc	9480
tcattatata	agaaagcggg	ttccggtagt	aaaaaaatgg	cagaatttga	aaacagaggat	9540
gctgtcaat	taataatggaa	tgccaggtt	aaaaataatgg	tcgaggagtt	gaaaatagaa	9600
aacgcatact	tgaaaaatga	attggctact	ttggaaaacg	aagtgcgaag	attagaaaat	9660
gatgttagct	aaggctctgc	tggttccgc	gggggtctcg	gtgaatttgc	cgaagcagcc	9720
gceaaggaa	ctggcgctaa	ggccggctca	gcagggtctg	ccggccggaa	cggtaattt	9780
tcttattacc	accacccaca	tcacccatct	gatctacta	cattatacaa	gaaaggcccc	9840
tcgggtatg	acgaagtcc	aaccttagaa	aatgtatcg	cctttataga	aaacgaaaat	9900
gcctacttag	aaaaagaaaat	tgccaggatt	agaaaggaaa	aagctgcatt	gagaacacaga	9960
ttggcccaaa	agaaatctt	ctatcacac	catcatatc	acttggaaat	tacatcatt	10020
tacaagaa	cgggtccgg	tagtgcata	aatgcctact	taagaaagaa	aatagccaga	10080
ttgaagaa	acaatttgc	attagagaga	gatggacaga	acttagaaaa	gattatagcc	10140
aatttggag	atggaaatttgc	tagatttaga	aatgaagtag	cttctcatga	acaaggtat	10200
gctggctcc	ctggcgctc	cggagaattt	gcccgaatgt	ccggcaacgg	agccggccgc	10260
aagggttgtt	ccgtgttgtt	tctgcgggg	tctggagaat	tttcttattt	ccatcatcatt	10320
catcatcatt	tggaaatctac	atccattata	aaagaacggc	gttctgttcc	taacttggtc	10380
gccccattgg	agaacgaagt	cgccatcttgc	gagaacgaaa	acgaaaccc	gaagaaaaag	10440
aacttacaca	aaaaggattt	gatacttgc	tttagaaaaa	aaatcgctaa	tttggaaaag	10500
aaaatgttgg	aaggtagtgc	aggttcagcc	gctggctccg	gttgaatttgg	ttcagcgggg	10560
gctggctca	aggggcgcg	cgctaaagaa	gccccggctc	aggagggtgc	agccaaagca	10620
ggttccggcg	gttctgcagc	gggttccgg	gaatttgg	ctgacaaaa	gttgcattct	10680
gaagaagatt	tggaaacaaa	gttgcata	gaagaagact	tggaaacaaa	attaatatca	10740
gaagaagatt	tgggttagtgc	aggttcagca	gctggtttgc	gagaatttgg	ttcagcagggt	10800
tctgcgtgt	gaagtggcga	attcggtat	ggccggctccg	ctgtcgctc	tggcgaattt	10860
qgttctgt						10869

SEQUENCE:	202	Organism	- Synthetic construct			
atgggttctg	ctgggttcagc	tgcagggtct	ggtaaatttg	gttccgcagg	tagtgcgcgt	60
ggttctggtg	aattcggttc	tgctgggtca	gcagccccgt	ctgggttaat	ttcatattac	120
catcacacc	accatcacat	ggaaatccacc	agtttataca	aaaaggctgt	ttctgggtca	180
gcttagaaacg	cataatggag	aaagaaaaat	gctgatggta	aaaagatata	cttgcatttg	240
gaaagagatg	aacaaaattt	ggaaaagatt	atcgccaaact	taagagatga	aatagcaaga	300
ttggaaaacg	aagttagctc	tcatgaccaa	ggttccgcag	gtatgcgcg	ttgttctgtt	360
gaatttgcgt	aaggccgtgc	aaaggaagcc	gctgaaaaaa	ctgttccgc	ttgttgcggc	420
gctggttccg	gtgaattcag	ttatcatccat	catcaccatc	atcacttgg	atccactagt	480
ttatataaaa	aggccgttc	ttggttcaat	ttggttgc	atttagaaaa	cgaatgcgc	540
tcttttagaaa	acgaaaacga	aacattggaa	aagaaaaatt	tgcataagaa	agatttgatc	600
gcttttttgg	aaaaggaaat	cgccaaatcg	agaaaaaaaat	tagagaagg	ttccgcgtgt	660
tctgcgtgc	gttccgggt	atttggtca	gtctgatgc	cagecaagga	agetcgcagc	720
aaagaaggccg	ctgtcaaaa	agctgcgcag	aaagcagggt	tcgcgggtc	ttgcggcagggt	780
tccgggtaat	tccgggtctc	agctactaga	gaattggatg	aattgtatgc	atccctaagt	840
gacttcaaga	tacaagggtt	ttccgcgtgt	tctgcagccg	gtcttgcgcg	atttcgcagaa	900
gcggccgcgt	agggaaacgc	cgctaaacgt	ggtttcgcag	gttctgcgtc	cggttctgtt	960
gaatttcgatt	ttgggttgc	tgaaaaactgg	gcacaaagaat	tttgcgtgc	agggtgcgt	1020
gttgcgtgtt	ctgcgttgt	tgctgcgggt	tcaggtaat	ttggtagtgc	ttgaactgt	1080
gccaagaaga	cagccgctaa	agaagctgt	gccaaggaaag	ctgcgcgtaa	agcagggttcc	1140
gccgggttctg	ccggccggc	cggttcagatt	ggttcagatt	ataaggatga	cgatgacaag	1200
gattcaaaacg	acgtatgtat	caaggattat	aaagatgcac	atgacaaagg	ttccgcgtgt	1260
atggccgcgc	gtctggaga	atccgttgc	gcgggttcag	tcgcggcgtc	cggagaattt	1320
ggttccgcgt	gtatgtcagc	cggttcagggt	gaatttcgtt	ctgggt		1365

SEQ ID NO: 203 moltype = DNA length = 19104
FEATURE Location/Qualifiers
source 1..19104
mol_type = other DNA
organism = synthetic construct

```

SEQUENCE: 203
          organism = synthetic construct

atgagtgcta aggcaattt c tgaacaaact ggttaaaaat ttgttgtacaa gtttattttgt 60
actacatcg ccatccaaaa tagattcaaa tacgtctagag ttacccccaaga tactgactgg 120
gtctatgtt tacaatgcata tccatggttt ttatctcaaa acttgggtt ccaaacctgac 180
caattaatta agagaagagg ttaattgggt tttagtaggt ttaatttgtac attggatcggt 240
gtaaagtctt gggtggaaacc aagatgttagt caagaaggcca cagtggtaa agcttccgggt 300
ttcttggaaaa atttttgtat cgaaaccattt gtcccttcatc cacaaggccga aagttttat 360

```

-continued

gtatgtatct acgtctactag agagggtgc tatgttttat ttcatcacga aggtgggtgc 420
gacgttagtg acgttgacgc caaggotcaa aagtgttgg ttgggtcga taaaagttg 480
aacccagaac acattaaaaa gcattttgg gtcacgcac ctgaataaa aaaaagaata 540
ttggccctct ttataagtgg ttgtttaat ttctacaagat attttactt caccacttg 600
gaaattaacc catttagt tactaaggat ggtgtatatg tttggactt agetgcacaa 660
gttgatgca cagccgacta catttgaag gtcaaatggg gtgacatcg atttccacct 720
ccatccgtgaa gagaaggtt ccacaaagaa gcctacattt ctgtatggc cgtaatgt 780
ggtgtccatcat tgaatgttgc acatgttgcac ctaaaaggta gaatttggc catgttgct 840
gggtgggtgt ctatgtcgat atatctgt actataatcg acttgggtgg tgtaaacgaa 900
ttggcaact acggtaata ctcagggtcc ccatccgaac aacaaacata cgattacgct 960
aagaccatct tgccttaat gactgaga aagcatccct atggtaaaaat ctgtatcatc 1020
gggtggtagt tcgcaatatt tactaacgtt cgccgetatc tcaagggtt cgtcagact 1080
ataagagatt accaagggttcc attgaaaggaa cacaagtaa caatattcgta taagaagggt 1140
ggtcttaact accaagaagg tttgagagtc atgggtgaag tagttaaaaac cactgttata 1200
ccatccatcg tctttgtgtac agaaaaaccac atgactgcac tagtggatgtt ggcttaggt 1260
catagaccaa tcccttaatac acettcaacaa gcggccaca ctgcacattt ctgtttaac 1320
gcctctggta caactccaccc accgttccct tctgaaacag caaattttcc tgaatcaaga 1380
gtgtgatgaag tcgctccagc taagaaagca aaaccagcca tgectcaaga ctccgttcca 1440
agtccatgtat ctttgagggg taaattttact acctttttt ctgacatcata taaggctata 1500
gtatggggta tgcaacaaac agcagttca ggcatgttgg atttgcacta tggtttagt 1560
agatgacac catctgttgc tgcacatgttgc tatectttt ctgggtgacca taacaaaaaa 1620
ttctactggg gtcacaagga aatattgtatcc ccagtttttta agaacatggc cgatgtat 1680
agaaaaacatc ctgaagtgcg cgtattgtt aacttccgtt cattaagatc cgcttacgt 1740
tcttataatcg aaaccatggaa ctacgttca ataagaacca tgcgtatcat tgcagaaggt 1800
atttccaaagag ctttgcacta aatgttgc aaaaagctg atcaaaaaagg tgtcacataa 1860
atcggtccatc ctaccgttgc tggttattaa cctgggttgg tcaagatcg taaactgtt 1920
ggtagtgggata aatacatatt ggcacatcaag ttgtatagac caggttcaatg cgcttacgt 1980
tcccaagatc gtggatgtc caacgatgg aacaaatca tcaatggaa tacagatgtt 2040
gtatcgaag gtgttgcgtc ttgggtgc acataccac gtttctactt tattgtatcat 2100
gtttagatgtt atcaacacac acctgggtttt aaaaatgtt tggttgggg tgaatagtgtt 2160
ggtagtggaa aataacaatgtt atgcacaggtt atcaaaaggaa gtagattgtac aaaggccaaatc 2220
gtttgtgttgc gcatgttgcat ttgtgcacaa atgtttttt cagaatgttca atccgttgc 2280
geagggccatc ggcataatca acgttcaaa acagcgttgc ccaaaacca agcattaaaa 2340
aaaggccgtt tttttgtccc tagatcttcc gatgatatttgc tgcatttttgc tcaatcgatc 2400
tatgaagact tggtagctt tgggttatttgc gttccacgc aagaatgttcc tccacactt 2460
gtccctatggg attactcttgc ggcacaaagaa ttgggttttta ttagaaaggcc agcttgggtt 2520
atgacatcttca tattgtatca aagagggttca gattgtatc atgctgtat gcttattact 2580
gaatgttca aagaatgttca aagggatccgat gttgttttagt gttgttgg tttccaaagg 2640
agattgtccaa agtactcttgc tcaatttatttgc gaaatgttgc tattgttgc acgttgcatt 2700
ggtcctgtct tctcagggttgc acacaaatacc ataatcttgc cttagactgg taaaatgtt 2760
gttttttttgc tgcacccatc ttttttttttgc gatttttttgc gtttttttgc tgcattttttgc 2820
ggccgtccaa agatgttttca aaaatgttca gattgttgc tatttttttttgc gtttttttgc 2880
aataagatgtt aaaaaggaggaa taaattgtata atgggtatcg gtcacatgttgc taatgttgc 2940
ataacccttgc atatgtatgtt acaaaatcttgc aaggactatg ttagacaaca cttttccagoc 3000
acaccccttgc tagatttgc ttttttttttgc gaaatgttgc cacttcttca aaaaaggccaaatc 3060
ttgtatgttgc acgttgcatttgc ttttttttttgc gtttttttgc tgcacatgttgc gaaatgttgc 3120
ggttccctca cttagaaaggaa agtgcattttgc ttttttttttgc gtttttttgc tgcacatgttgc 3180
tttttttttgc gtagatcttgc gggtttccatttgc gtttgcattact ttttttttttgc gtttttttgc 3240
caagggttgc acagacatcc atgggtatcg ttttttttttgc acacatgttgc ttttttttgc 3300
atggatgttgc ttttttttttgc gtttttttgc gtttttttgc ttttttttgc ttttttttgc 3360
atggatgttgc ttttttttttgc gtttttttgc ttttttttgc ttttttttgc ttttttttgc 3420
gaaggttgc acatatttttttgc ttttttttttgc ttttttttgc ttttttttgc ttttttttgc 3480
aatatttacc aaaaaggccatc ggaatttagac ccaaaatatac cagaaggccgtt gtataattttgc 3540
ggtagtgcatttgc attataaaaca ggggttgcatttgc ttttttttttgc ttttttttgc ttttttttgc 3600
tttggatatttgc acccttaatca ctttgcatttgc ttttttttttgc ttttttttgc ttttttttgc 3660
cagggttgc acatcaaaaggaa ctttgcatttgc ttttttttttgc ttttttttgc ttttttttgc 3720
aacttataacatc ctttgcatttgc ttttttttttgc ttttttttgc ttttttttgc ttttttttgc 3780
caaaaacatc ttttttttttgc ttttttttttgc ttttttttgc ttttttttgc ttttttttgc 3840
tggtagtgcatttgc ttttttttttgc ttttttttttgc ttttttttgc ttttttttgc ttttttttgc 3900
tatttttttgc ttttttttttgc ttttttttttgc ttttttttgc ttttttttgc ttttttttgc 3960
tatttttttgc ttttttttttgc ttttttttttgc ttttttttgc ttttttttgc ttttttttgc 4020
gggtgggtgttgc ttttttttttgc ttttttttttgc ttttttttgc ttttttttgc ttttttttgc 4080
tttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttgc ttttttttgc 4140
tttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttgc 4200
tttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttgc 4260
gttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttgc 4320
aaaggccgtt ctttgcatttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttgc 4380
catccatccatc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttgc 4440
tttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttgc 4500
agaaaatgttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttgc 4560
agatgttgc acatccatc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 4620
gaaaatgttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttgc 4680
caagggttgc acgttgcatttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttgc 4740
gtttagtgcatttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttgc 4800
gggtgttgcatttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttgc 4860
gaaggttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 4920
tttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 4980
atcggttgcatttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttgc 5040
agatgttgcatttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttgc 5100

-continued

-continued

tttttcagaaa gattggaaat gttggaaaaa cctgttatacg ctgcagtcgaa tggtttcgct 9900
ttgggtgtt gttgtgaat cgcaatgtcc tgcgatatta gaatagtc ttcaaacgc 9960
agatgttgtt aaccqaaga ctggtttagt attacacctg gtttcgttg tacccaaaga 10020
ttatcttagat tgggtgtt aaggatggcc aagcaatgtt aaaaaaaaaacttc 10080
aaggctgatg aagcatgtt aatcggtt gttataataa tagttaaaaa atctgtatgg 10140
atgaacaccg ccaaggaaat cgctaataa agtggtttca atgtccatgt tgctgtcaag 10200
ttgagtaaag aagctataaa tcgtgtgtt caatgtgtt tgccagactg attggccctt 10260
gaatctgaag cattttgtt atgttttca aacaagatgg aaaaagacgc aatggccgc 10320
tttatcgaaa agagaagatg aagaaggttt aaaaacaaaa agttatctgg tgggtgtt 10380
tctgggtgtt gtggttctgg tgggtgtt agtgcgtt aatggatcca ttgggttaac 10440
gcttactaca aecagggtt cttacaaaaa gcaatcgaa actacaaaaa ggccgttggaa 10500
ttggacccaa ataacccggc agctgggtt aatgggtt atgcctttaa aacagggtt 10560
gactatcaa aagctataga atactaccaa aaggatgggg aatggacccca taatacgca 10620
gaagcctgtt ataatttggg taacgcctt tataaggcgg gtgactatca aaaggccata 10680
gaagactacc aaaaggctt ggaattggat ccaaaacaact tacaagtcg agatggaaag 10740
aattttgggtt acgcttataa caacagggtt gactacaaaaa aagctatgtt atatttttt 10800
aaaggcttag aatttagaccc taataacgc ttcgttgggtt acaatgggg taatgcctac 10860
tataaaacagg gtgactacca gaaggctttaa atattttttt aaaaagctt agaattggat 10920
ccaaatcaacg cttacggccgt gtacagactg gtaatggctt attacaagca gggtgactac 10980
aaacaaacggc ttgaaatggat tcaaaaaatggt ttggtatgg atccataacaa cagatccaga 11040
agtgcgttggt tgggtgtt tgggtgtt gttctgttgg tgggtgttgc ttctgttcca 11100
ttgggttccc ctttgactgc atcaatgtt gttccggc caccccaaga acaaaaaagca 11160
atgttgggtt aaagattttt cccattgata caagctatgc atccatctt agcagggttttt 11220
atcacaggcc tgggtgttgg aatcgatcaat tcgttatttt tacatgttggt aaaaaccccc 11280
gaaagggttgg gatctaaatgg tgcacggcc gtagctgttgg tgcacgttca tcaacaaaaa 11340
aaaggccgttcc aaaaggccggg ttcaatgttgg tcccgacggc gtagtgggttgg atttgttctt 11400
gctgaagctg ctttgcggcc agctgcggcc aaggccgggtt gttccgggttgc tgctgtcagg 11460
ctctgggtat tgggttccaa tttttttttt aatggatggt gttctgttgg tgggtgttgc ttctgttcca 11520
tctcaatcc tcgtcatcga acaagaaatgg tttttttttt tttttttttt tttttttttt 11580
ttggttggat cttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 11640
ttgggttcaat atcatcatgac aaagagatgg tttttttttt tttttttttt tttttttttt 11700
tggtaaaccg atttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 11760
agaaaggatct atacaatgtt atacagaaaaat tttttttttt tttttttttt tttttttttt 11820
tctttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 11880
atgggttggat acaacatctt tttttttttt tttttttttt tttttttttt tttttttttt 11940
gtacaaatcc aatcacccaa aaaaaggatcc tttttttttt tttttttttt tttttttttt 12000
cttaaaatgg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 12060
gtgcatttttgg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 12120
acccaaatatg gtactccagg tttttttttt tttttttttt tttttttttt tttttttttt 12180
gaagggtttat actctgttcc tttttttttt tttttttttt tttttttttt tttttttttt 12240
atgggttggat aaggccaaaaa tttttttttt tttttttttt tttttttttt tttttttttt 12300
agtccatgttca aacacatggc tttttttttt tttttttttt tttttttttt tttttttttt 12360
ggttaaaactt tcacaggtaa aacaggctt tttttttttt tttttttttt tttttttttt 12420
gctgaaccatgg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 12480
caaatatgtt gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 12540
tgggttggat ggttcaatgg tttttttttt tttttttttt tttttttttt tttttttttt 12600
tttagctttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 12660
ggttaaaatgg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 12720
atccatgttca aacacatggc tttttttttt tttttttttt tttttttttt tttttttttt 12780
ggttccatgtt aacaaatccatgg tttttttttt tttttttttt tttttttttt tttttttttt 12840
atccatgttca aacacatggc tttttttttt tttttttttt tttttttttt tttttttttt 12900
caaaaggccatgg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 12960
ggttccatgttca aacacatggc tttttttttt tttttttttt tttttttttt tttttttttt 13020
aactatgttca aacacatggc tttttttttt tttttttttt tttttttttt tttttttttt 13080
cggccatgttca aacacatggc tttttttttt tttttttttt tttttttttt tttttttttt 13140
tattttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 13200
gacccttata aacccggccatgg tttttttttt tttttttttt tttttttttt tttttttttt 13260
tattttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 13320
ggttccatgttca aacacatggc tttttttttt tttttttttt tttttttttt tttttttttt 13380
aactatgttca aacacatggc tttttttttt tttttttttt tttttttttt tttttttttt 13440
tattttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 13500
ggttccatgttca aacacatggc tttttttttt tttttttttt tttttttttt tttttttttt 13560
aactatgttca aacacatggc tttttttttt tttttttttt tttttttttt tttttttttt 13620
aataacccatgg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 13680
aactatgttca aacacatggc tttttttttt tttttttttt tttttttttt tttttttttt 13740
ggttccatgttca aacacatggc tttttttttt tttttttttt tttttttttt tttttttttt 13800
catccatccatgg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 13860
aattttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 13920
gaaaaggatgg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 13980
tacaaaggatgg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 14040
gctaaaggatgg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 14100
ggttccatgttca aacacatggc tttttttttt tttttttttt tttttttttt tttttttttt 14160
ggttccatgttca aacacatggc tttttttttt tttttttttt tttttttttt tttttttttt 14220
gataacccatgg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 14280
aactatgttca aacacatggc tttttttttt tttttttttt tttttttttt tttttttttt 14340
aactatgttca aacacatggc tttttttttt tttttttttt tttttttttt tttttttttt 14400
gaaaggatgg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 14460
gatgttccatgttca aacacatggc tttttttttt tttttttttt tttttttttt tttttttttt 14520
gttccatgttca aacacatggc tttttttttt tttttttttt tttttttttt tttttttttt 14580

-continued

gatatgtatt tggtagagt agctgcagtt aatgggttgt ttaccataaa cgctcctgca	14640
ttaactgtca acagattgtg tggtagtgt ttacaagcta ttgttctgc cgctcaaaa	14700
atattgttag gtgacaccga cgttgctata ggtgggttgt ctgaatctat gtcaagagcc	14760
ccatacttag ctctgcagc cagatgggt gcccagaatgg gtgacgctgg tttgggtgac	14820
atgatgtgg gtgccttgc tgatccatc catagaatcatc acatgggtgt aactgcagaa	14880
aacgttgcga aggaatacga tatctcaaga gcacaacaag acgaagctgc attagaatca	14940
cacagaagag catccgcgc tattaaaggc ggttacttta aggataaat agtcccgat	15000
gtttctaaag gtagaaaggg tgacgttacc ttgcatactg acgaacatgt tagacacgac	15060
gtactattt atgacatgc aaagttaaga ctgttttcg tcaaggaaaa tggtaactgtt	15120
acagctgta atgcattcg tttgaacat gcaagccgtg cagtcgtat gatggaaaga	15180
gcggaaatgtt aagaagaggg tttagatcat ttagatcat tttgttctt tggatcatgt	15240
ggtgtcgatc taaagcaat gggtaggtt ccgttctgtt ctactaagat cgcattggaa	15300
agagccgtt tacaatgtc tgatgggtt gtaattgttgc ccaatgtcgc ttttgcgcgt	15360
caagcatgtt cgtttttttt aacgttgggtt ttagatcaat ctaaggatcaat tccatcacgg	15420
agtggatctt cttaggtca tccaattgtt gcaacccgtt cttgtataac tggtaaggct	15480
ttgcacgaat tgaacagagt acaaggatgtatcattatgtt tttttttttt tttttttttt	15540
ggtcaaggta ttgcagccat attcggaaattaatgtt tttttttttt tttttttttt	15600
ggtgtgtgtt ctgggtgtt gttttttttt tttttttttt tttttttttt tttttttttt	15660
tacaaggcagg gtgacttacca aaaggcaatc gaatattacc aaaaaggctt ggaatttagac	15720
ccaaataacg cggaaatgtt gtataattt ggttaatgcattttaaaa aaaaaggctt ggaatttagac	15780
aaaaaaagctt tagaaatataa cccaaaggcc ttggaaattt acccttaataa cgcagaagcc	15840
ttgtataattt tggtaacgc ctatccatc cttttttttt tttttttttt tttttttttt	15900
tacccaaagg cttttttttt ggtttttttt tttttttttt tttttttttt tttttttttt	15960
ggtaacgtttt attacaaaca gggtagttttttt tttttttttt tttttttttt tttttttttt	16020
ttagaattttt acccttataatc ccccttctgtt tttttttttt tttttttttt tttttttttt	16080
cagggtgactt accaggaaatc tttttttttt tttttttttt tttttttttt tttttttttt	16140
aacggcaagg cttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	16200
gttattttttt attatcaaaa ggcattttttt tttttttttt tttttttttt tttttttttt	16260
ggtgtgtgtt gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	16320
gaatataatgtt gggaaatataa aacgttggatc gttttttttt tttttttttt tttttttttt	16380
atgcacccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	16440
ccacccgttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	16500
gtctgggttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	16560
gttgcacccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	16620
gattccgtttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	16680
gatcatccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	16740
ttgcacccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	16800
gtctttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	16860
gaaaaatccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	16920
gttgcacccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	16980
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	17040
attttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	17100
tttgcacccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	17160
tttgcacccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	17220
ggttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	17280
tccatccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	17340
gtttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	17400
gttgcacccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	17460
aaaaaaatgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	17520
gttgcacccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	17580
gtttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	17640
atcgatgttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	17700
tacgtttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	17760
ttatccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	17820
cctttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	17880
ggttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	17940
agagccaaat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	18000
actaaatccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	18060
ggccacccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	18120
gttgcacccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	18180
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	18240
aaatccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	18300
tggaaatccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	18360
tccatccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	18420
ccatccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	18480
ttacaagttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	18540
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	18600
ggttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	18660
tccatccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	18720
ttaacccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	18780
ggttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	18840
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	18900
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	18960
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	19020
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	19080
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	19104

-continued

source 1..18738
 mol_type = other DNA
 organism = synthetic construct
 SEQUENCE: 204
 atgaaggtaat ctactaaatt gtgttggc ggttataagg gtatagttaa accacaaaag 60
 caacaacaat tgccataaac acacttgca atgaccgat tgaagaaaca aaagactgtc 120
 gaacaagaag ctagaccaca aaacgttgg attaaatgtc tcacaatctc tatccctaca 180
 caatgtgtc atcaatctgaa atggaaaaag ttgtatgtc tataccaggaa taatcacat 240
 atcggttag gtcaaaacaa catgttttc gttaacgcata gaaagacatc tattatgt 300
 tcattgtactg tttgtccaa gttgataaaa agttacaaca tcgatataaaa caagattgt 360
 agatgttggaa ttgttaccga aactttgtc gataatgtc agagtgtca gtctgttatt 420
 atgtcaatgtc tggcgatggaa tccatgtt ggaggatcactc acactttaaa tgctgttatt 480
 ttgttgcatac acgcattaa caattatgtc aactgtgtc acatccatgc ctggatgtt 540
 agagatgtca ttgttgcata cggtgacatc gctatctatg acaaagggtc tgcaagacc 600
 accgggtggc caggtaactgt tgccatgtt atagggtccatc atgcacccatc eggtttgtc 660
 ttgttgcatac catcatatc ggaatgtcc tacgatattc acaaaggccatc cttcacccct 720
 gaaatccatc acgttgcattc tcaatgttgc ttgtacttgc acgtcaaggc ttggccaa 780
 gtatacaatgtc ttactctaa gaaagcaata tctaagggtt tggttccaga tccagctgtt 840
 tccgacgcata taaaatgttgc gaaatgttgc gattacaacg tttccatgtt ccctatgtc 900
 aagtgttggaa ccaatgttgc cggtagatc ttgtatgtc aatccatgc taacccacaa 960
 ttgttgcatac aagtgcacgc tttgtatgtc actagatgtc acgcacaaatc ttgttgcata 1020
 aagaacatcg aaaagactt cgtaaagctt gcaaaaggccatc tccacaaaga aagagtggcc 1080
 caatcattaa ttgtccatc aaataccgtt aacatgttgc cagccatgtt ttacggccgt 1140
 ttgttgcatac ttgttataatgtc tggtaggtt gatgttgc aagggtttaaa agttgttgg 1200
 ttctccatgtt gtatgttgc acgcacccatc ttgtacttgc ttgtatgtt aggtgacgtt 1260
 caacacattaa ttaaggatgtt ggcacatc aataatgttgc ctaagagaaatc cactgaaaca 1320
 ccaaaggatt atgaagctgc aatcgatgtt agagaaaacg cacatttgc gaaaaatttc 1380
 aaacccatcg gtatgtatgc acacttgcata tctgggttgc actacttacaa acacatcgat 1440
 gacaaatccatc gaagatcata cgtatgttaaa aagaatgttgc ctgggttgc tggttctgt 1500
 ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 1560
 tacaaggcagg gtgactacca aaaagcaatc gaatattacc aaaaggccatc ggaatttagac 1620
 ccaaataatcg ccgaaggatgtc gtacaatgtt ggtatgtc actataaaaca gggtgactat 1680
 ccaaaggatc tggatataactt cccaaaggca ttggatatt accctataaaatc cgcgaaagcc 1740
 ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 1800
 taccatgttgc ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 1860
 ggtatgtcactt acataacaa ggtgttgcact cccaaaggccatc ttgtatgttgc ttgttgcatac 1920
 ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 1980
 cagggttgcact acccagaacgc ctttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 2040
 aacgcacaaatcg cctggatgttgc acgttgcatac gcctactaca agcagggttgc ctatcagaag 2100
 gctatgttgcact attatgttgc acgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 2160
 ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 2220
 ccacatgttgc ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 2280
 aacagaactt cacaatttccatc ogacccaaaga ttgcacgggtt ctgttgcatac agccgttgcgtt 2340
 ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 2400
 gctatgttgcact cccgggttgc tggatgttgc ttgttgcatac ttgttgcatac ttgttgcatac 2460
 ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 2520
 attactcaat ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 2580
 ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 2640
 caacatgttgc ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 2700
 ttcatttttttttccatc ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 2760
 ggtgtatgtc ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 2820
 ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 2880
 gctatgttgc ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 2940
 ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 3000
 aaaaagggttcaatccatc ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 3060
 gccgggttgcact ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 3120
 atgtatgttgc acgggttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 3180
 gatgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 3240
 aactgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 3300
 gtttagaaagg ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 3360
 ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 3420
 ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 3480
 ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 3540
 atagaatgtc ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 3600
 ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 3660
 agaattgtc ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 3720
 ggtgtatgtc ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 3780
 ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 3840
 ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 3900
 gacccaaatcg acgcacaaatcg ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 3960
 taccatgttgc acgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 4020
 gcatgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 4080
 gacttacaaatcg acgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 4140
 ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 4200
 gcttggatgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 4260
 aaaaagggttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 4320
 aataacgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 4380
 aaaaagggttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 4440
 gcttggatgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac ttgttgcatac 4500

-continued

cataccattt accatcaactt ggaatccaca agtttataca aaaaggcagg ttcaaaattt	4560
ttcagaagag aaagaataaa gatggccgt gcaaatgtt gaaacagaag aagagaattt	4620
acagataacct tacaagctt aaccgtttttt ttggaaagacg aaaagtctgc attgcaactt	4680
gaaaatagcca atttggtaa ggaaaaggaa aagttggat tcatttttc cgctcataga	4740
ccagcttgc aaatttctgt tgacttgggt ttcccagaagaa atatgtctt agaagggtcc	4800
cgaggtagtg cagccgttc cggtaattt ggttagtgcg aagtcgcgc caaggaaagtt	4860
gcagccaaag ctggttctgc aggttcgtt gcaggttccg gtgaatttccg ttcttcatac	4920
tatcaccatc accatcatca ctggaaatctt accttattt aaaaaaggc tggttccgg	4980
agtccaaagg ttgaatctttt gaagaaaaatg attgaagaat tgaagcaaaag aaaaggccaa	5040
tttgaagaatg atatcgatca ctttagatcc gaaatcgcc acgtgaaatc tggttcttgt	5100
gtctactact tttttttttt ggaaagca ggtgacgtt aagaaaaatcc aggttccatg	5160
agtttaccat ttttgacatc tgcttcttgtt aaagtttata tattcggtt acaatgttgc	5220
gtctataata agccatgtt cgctgtatc gtatcgatc tgaaacata ttgttgcate	5280
tctgaatattt cagcacatca taccatcgaa ttggattttt cagacatctc attcaatcac	5340
aagtggtc ttaatgattt caacgcttcc accggaaatc aagttttttt acaaaaagg	5400
gccaaagctc aacaagcaac tgatggttt tcacaagaat tagtttccctt gtttagacca	5460
ttgttggctc aatttgcgtt aagtttccat taccacgccc ctttctgtt ttgttacatg	5520
ttcgtttttt tttttttttt tgtaaaatc atcaatattt ctttggaaatc tacttttgc	5580
atttggtcgat gtttagttt cagttccctt atatcgatc ttttttttt ggccatgg	5640
tatttgggtt gtttgcattt tagtaacgtt ttggaaaatg tttttttttt cgacaaggcat	5700
atcgtaacc aatggccattt categgttcaaaatgttcc acggtaacttcc tagttgttata	5760
gataatgttgcgat tttttttttt tggttacatc ttgttgcattt aaaaaggactc tcataaacgg	5820
accatcaaca ctaacaactt caagtttgcgtt gatgttttcc ttgttgcattt aatgttgc	5880
acttacacaa gaatttcaag atctactaaa gatttttttgcgtt gttttttttt agtattttttt	5940
acagaaaaatg tccctgttgcgtt tgtaaaatc atctttttttt gttttttttt atgttgc	6000
caagggttttggaaatcatgttcaaaatgttgcgtt aagtttttttgcgtt gttttttttt	6060
gaaaacaaataa acaaaatgttgcgtt acaaaatgttgcgtt ttggaaatgttgcgtt aatgttgc	6120
tttagtttcaat ttgggttgcgtt ccacccatgttgcgtt ttggaaatgttgcgtt aatgttgc	6180
tttaaagatcg gtttttttttatttgcgtt gtttttttttgcgtt gtttttttttgcgtt	6240
ttaaagatcg atatcaatcacttca aaaaaaaaaatgcgttgcgtt aaaaaaaaaatgcgtt	6300
tttcttttttgcgtt acaacatgttgcgtt acaacatgttgcgtt ttggaaatgttgcgtt	6360
aagaatttttgcgtt acaaaatgttgcgtt acaacatgttgcgtt ttggaaatgttgcgtt	6420
actacaacca acaaaatgttgcgtt acaacatgttgcgtt ttggaaatgttgcgtt	6480
accttacaaat ttttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt	6540
gttgcgttgcgtt gtttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt	6600
atcgatactt accaaaaggc attttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt	6660
tttgggttgcgtt ttttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt	6720
gttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt	6780
aaggcagggttgcgtt acaaaatgttgcgtt acaacatgttgcgtt ttggaaatgttgcgtt	6840
aataacttgc aacgttgcgtt atggaaaaat ttggaaatgttgcgtt ttggaaatgttgcgtt	6900
tacccaaaatgcgtt acaacatgttgcgtt acaacatgttgcgtt ttggaaatgttgcgtt	6960
gttttttttgcgtt acaacatgttgcgtt acaacatgttgcgtt ttggaaatgttgcgtt	7020
tactatcaat aacgttgcgtt acaacatgttgcgtt acaacatgttgcgtt ttggaaatgttgcgtt	7080
aatgttactt acaaaatgttgcgtt acaacatgttgcgtt acaacatgttgcgtt ttggaaatgttgcgtt	7140
gaatttagacc caaataacag aagtttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt	7200
tctgggttgcgtt ttttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt	7260
atcaatagatcc ttttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt	7320
tgttgcgtt acaacatgttgcgtt acaacatgttgcgtt ttggaaatgttgcgtt	7380
aaaatttcaat caccacaaatgcgtt acaacatgttgcgtt acaacatgttgcgtt ttggaaatgttgcgtt	7440
atgttgcgtt gtttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt	7500
tttgcgtt acaacatgttgcgtt acaacatgttgcgtt ttggaaatgttgcgtt	7560
aaagatcttttgcgtt acaacatgttgcgtt acaacatgttgcgtt ttggaaatgttgcgtt	7620
gaagcgttgcgtt acaacatgttgcgtt acaacatgttgcgtt ttggaaatgttgcgtt	7680
gttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt	7740
gttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt	7800
gttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt	7860
agacaagaatgcgtt acaacatgttgcgtt acaacatgttgcgtt ttggaaatgttgcgtt	7920
gttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt	7980
gttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt	8040
aacgttgcgtt acaacatgttgcgtt acaacatgttgcgtt ttggaaatgttgcgtt	8100
tttaggttgcgtt acaacatgttgcgtt acaacatgttgcgtt ttggaaatgttgcgtt	8160
actatgttgcgtt acaacatgttgcgtt acaacatgttgcgtt ttggaaatgttgcgtt	8220
gttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt	8280
ccaaacatgttgcgtt acaacatgttgcgtt acaacatgttgcgtt ttggaaatgttgcgtt	8340
gttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt	8400
tttaggttgcgtt acaacatgttgcgtt acaacatgttgcgtt ttggaaatgttgcgtt	8460
tcacccatgttgcgtt acaacatgttgcgtt acaacatgttgcgtt ttggaaatgttgcgtt	8520
gaagtttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt	8580
aagtttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt	8640
attgttgcgtt acaacatgttgcgtt acaacatgttgcgtt ttggaaatgttgcgtt	8700
tttgcgtt acaacatgttgcgtt acaacatgttgcgtt ttggaaatgttgcgtt	8760
cacaggttgcgtt acaacatgttgcgtt acaacatgttgcgtt ttggaaatgttgcgtt	8820
gttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt	8880
tttgcgtt acaacatgttgcgtt acaacatgttgcgtt ttggaaatgttgcgtt	8940
caagtttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt	9000
gttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt gtttttttttgcgtt	9060
atcggttgcgtt acaacatgttgcgtt acaacatgttgcgtt ttggaaatgttgcgtt	9120
atcacaatcaat acaacatgttgcgtt acaacatgttgcgtt ttggaaatgttgcgtt	9180
aatgttgcgtt acaacatgttgcgtt acaacatgttgcgtt ttggaaatgttgcgtt	9240

-continued

ccagaatcat tgaagatcta cacagaatta gaccatgcta attccagatt catggatgtt	9300
tttagataat tagacaggat gcataccac gatgactact ctgatcaat cttcaatca	9360
ttggaaagaa acgactgtac ttgcacaaaa tacccagaaa tcacagaagt aagagatgcc	9420
gttgcatacca taagaagatc tttagaaag atcaactaagg aatcaggatgc agatatcgaa	9480
ccacctgttc aaacatcttt gttagatgc tgtaaacact tgaagggtgt cttaacttgc	9540
ttgattccag gtgctgggtt ttatgatca atagccgtca tcataaaca agatgttagac	9600
tttagagotc aaacagcaaa cgataagaga tttcacaaagg tccatgtt agatgttaacc	9660
caagctgact ggggtgttag aaaagaaaaag gatccctgaa cttacttggaa caaaaagttt	9720
tctgggtgtt gggtgtctgg ttgggtgtgt tctgggtgtt gtggtagtgc tgaagcatgg	9780
tacaatttgg gtaacgcata ctacaaaggac ggtgacttaccc aaaaaggccat agaataactac	9840
caaagggtt tgaaatggta cccaaaggccat gcccggactt ggttaatattt gggtaatgt	9900
tattataaac agggtgacta tcaaaaggca atcgaataact accaaaaaggc cttggaaat	9960
gacccttaata accgcagaacg ctgttataat ttgggttaacg ttattataaa gcagggttgac	10020
tatcaaaaag otatcgaaat ctacaaaaggac gtatggaaat tagatccaa taacttgcac	10080
gtctgaacat ggaagaaattt ggtaacgcgt tactataac accgggtacta cccaaaagcc	10140
attgaatatt atcaaaaagc ttggaaattt gatccataat acgcctctgc ttggtaacat	10200
ttgggtaaatg ctatataatc gcagggttac tttcagaagg ctatgtaaat ttatcaaaag	10260
gtcttagata tggaccctaa taacgcaaaac ggctgggttac gacgttggtaa cgcttattac	10320
aaacagggtt actaccagaa agtattgtt gattatcaa accgcatttga attggatcc	10380
aacaacagat ccagaagtgcc ttgggtgtgtt gggtctgggtt gtgggtgttc ttgggtgtgtt	10440
gggtgcctca gttattacca tcaccatcatc catcaacttgc aatctacatc attatacaaa	10500
aggctgttgtt ccggtagtca aaagggttac gaaatggaaata agatgtac cgaattggaa	10560
aacagaaacg ctgttataaaa gaacagatgc gcacatttga aacaagaaat accctacttgc	10620
aggatgtat tagcagcccc tgaatttggaa ggttctggcc gttcagatgc aggttctgg	10680
gaattcggtt cagctgaaacg cgctgcacaa gaaggccgtt cttccgcgtt ttccgcgtt	10740
agtgcctgtt gttctgggtt atttgggtt tcataacttgc accatcacca tcatcacttgc	10800
gaatctactt cattatataaa aaaggccgtt tccggtagtgc tggaaaactgt tacatcgaa	10860
ttcattttgg ctaccttggaa aaacggaaac gcaaaatgttac gaaatgttgc agccaatgtt	10920
gaaaagagaat tagcttagat gagaatgttac gttcgtatgtt taggttctgg tgctactaac	10980
ttctctttgt tgaagaaacg aggttgcgtt gaaagaaaatc caggatccaaat gacagtttt	11040
accgccttcg taccgcacc tggtaatattt gctacttgc aatactgggg taaaagagat	11100
actaagtgtt atttgcacac aaactcttca atctcgttgc cattgtccca agatgtacttgc	11160
agaaaccttgc ctctgttgc aactgttgc aactgttgc aatgttgc aagatgttgc	11220
ggtaaccacat atctatgtca caacggaaacg actcaaaaactt gtttggagaga ttggagacaa	11280
ttgagaaagg aatggagat tggatgttgc ttttgcttca cattgttgc aatggaaatgtt	11340
cacatagttt ctgaaaacaa cttcccaacc gccgctgtt tggcatccag tgccgcgtt	11400
ttcgctgtat tagtctctgc aatcgccaaat ttgttaccaat tgccacaatg tacatcttgc	11460
atcagtagaa tccgttgc ggggttgcgtt tccgcgttgc gatctttatt ttgggtgttac	11520
gtcgcgttgc gaaatgggttac agccggacat ggttgcgttgc catatggccgtt acaaataatgt	11580
gactcttcgtt attggcttca aatggaaatgttgc tggcttgc tttgtcttgc catcaaaaag	11640
gatgtatccca gtacacaagg catcaatttgc actgttgc aatccgttgc gttttaagaa	11700
agaatcgac acgttgcgttcc aaaaaggatc gaaatgttgc gaaagggttgc tggaaatggaa	11760
gatttcgcgtt ctttcgttgc gggaaatgttgc tggatgttgc actcttcgc tgcaacttgc	11820
ttggatccat ttccacatgc ttcttatgtt aatggatccat ctttgcgttgc aatctcttgc	11880
tgccacacta tcaaccatc ctacgggttgc acaatgttgc ttttgcgttgc cttacaccatc	11940
cctaattgcgtt ctttgcgttgc aacgttgc aatgttgcgttgc ttttgcgttgc ttttgcgttgc	12000
aagttgtttt gttccgttgc ggggttgcgttgc aatggatccat ctttgcgttgc ttttgcgttgc	12060
ttaatcatc aatcgcaatc ttcaacttgc acttgcgttgc aatggatccat ctttgcgttgc	12120
aggatgttag cttaggttgc ttttgcgttgc aatggatccat ctttgcgttgc ttttgcgttgc	12180
ttcttgcgttgc atgttgcgttgc aatggatccat ctttgcgttgc ttttgcgttgc ttttgcgttgc	12240
gttgggtgtt gttctgggttgc ttttgcgttgc aatggatccat ctttgcgttgc ttttgcgttgc	12300
tattacaaacgggttgc aatggatccat ctttgcgttgc ttttgcgttgc ttttgcgttgc	12360
gacccttataa acggccgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc	12420
tatcaaaaacgggttgc aatggatccat ctttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc	12480
gttgggtgttgc aatggatccat ctttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc	12540
gatttacaaacgggttgc aatggatccat ctttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc	12600
ttgggttgcgttgc aatggatccat ctttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc	12660
gttttgcgttgc aatggatccat ctttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc	12720
aaacagggttgc aatggatccat ctttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc	12780
aataacccaa aacgttgcgttgc aatggatccat ctttgcgttgc ttttgcgttgc ttttgcgttgc	12840
aggcgatgttgc aatggatccat ctttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc	12900
gttgggtgttgc aatggatccat ctttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc	12960
gatttggaaatgggttgc aatggatccat ctttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc	13020
aatgttgcgttgc aatggatccat ctttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc	13080
tttagtccatc ctttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc	13140
aatggatccat ctttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc	13200
cctaattgcgttgc aatggatccat ctttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc	13260
caaacaacgggttgc aatggatccat ctttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc	13320
caaaggccatc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc	13380
aatgttgcgttgc aatggatccat ctttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc	13440
gatttgcgttgc aatggatccat ctttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc	13500
gaacgcgttgc aatggatccat ctttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc	13560
ggcgaattttg gtagtgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc	13620
gttgggtgttgc aatggatccat ctttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc	13680
gttgggtgttgc aatggatccat ctttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc	13740
aaagtttgcgttgc aatggatccat ctttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc	13800
aaagtttgcgttgc aatggatccat ctttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc	13860
caaggttgcgttgc aatggatccat ctttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc	13920
gttgggtgttgc aatggatccat ctttgcgttgc ttttgcgttgc ttttgcgttgc ttttgcgttgc	13980

-continued

aagaccttta atattaagaa tgatttact gaagaagaag aagcacaagt tagaaaggaa	14040
aaccaatggt gcggttctgg tgctactaac ttctcttgc tgaagcaagc aggtgacgtt	14100
gaagaaaatc caggccaat gactgctgat aataactcta tgccacatgg tgccgtatct	14160
tcatacgcta agttggttca aaaccacca cctgaagata tcttggaga attcccagaa	14220
atcatccctt tgcaacaaag accaaacact agatccagtg aaacatccaa cgatgaaagt	14280
ggtgaaaccc tttttcagg tcatgacgaa gaacaaatta aattgatgaa cgaaaactgc	14340
atcgatattttt atgggatga caatgcaata ggtgccggta ctaagaagt ttgtcattt	14400
atggaaaaca tagaaaaggg ttgttgac agagctttt attcaatgaa	14460
cagggtgaaat tgtttgcg acaaaggagca acagaaaaaaga tcacccccc agattttgtt	14520
actaatacatc ttgtctca ttctttgtc attgtatgacg aatttagttt gaagggttaaa	14580
ttggatgaca aatattaaggg tgctataact gctgcgtca gaaaatttga tcatgattt	14640
ggtataccag aagcgaac caagactcg ggttaattcc atttcttaa cagaatccac	14700
tatatggc catctacgca accttgggtt gaacatgaaat cttgattcat cttatatttac	14760
aagattaatc aaaaaggaaa ctgtacatgaa aacccaaacg ttaatgaaat cagagattt	14820
aaatgggtt ctcctaatgaa ttgttgcg atggttgcg accccatcata taagtttact	14880
ccttgggtca agatcatctg taaaactac ttgtttaact ggtgggaaaca attagatgac	14940
ttgtctcaag ttggaaaacg tagacaaatc catagaatgt tggaaatttgc ttgtgggtt	15000
gggtctgggtc ttgttgcg tggtagtgcg aaggttggta caatttgggt	15060
aacgcttact acaaggcaggg tttgttgcg aatactacca aaggccatcg aatactacca	15120
gaatttggacc caaataacgc agaaggctgg tataatttgc gtaatgcata ttataaacag	15180
ggtgactatc aaaaggctatc tgaatatttcaaaaaggcat tggaaatttgc ccctaataac	15240
gctgaagcat ggttataatttcaaaaaggcat tggtaacgcg tattataatgcg aagggtgacta	15300
atcgaagact accaaaaggc tttgttgcgatttgcgatttgcgatttgcgatcg cgaagcttgc	15360
aagaattttt gtaacgcttta ttacaaacag ggtgacttcc aaaaagctat tgaatactat	15420
caaaaaggct tagaatttgc cccataataatc gcatctgcctt ggtacaattt ggttaatgc	15480
tattacaatc aggggtgacta tcagaaggctt atggatataactt accaaaaaagc atttggat	15540
gatccaaata acgctaaaggc atggatcaga cgtggatcata ctttgcataa gcagggtgtac	15600
tatcaaaagg cgatttgcgaa ttatcaaaaa gctttggat tggatcataa caacagatct	15660
agatcgttgc ttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgc	15720
tattaccatc accatccacca tcacttagaa tccacaatgt tttgttgcg tttgttgcg	15780
gggtcaacaa ccgttacggaa atttggaaatc tttgttgcg tttgttgcg tttgttgcg	15840
gaatttgcgaa atttggaaatc acatttgcg tttgttgcg tttgttgcg tttgttgcg	15900
gcccgtccaca aatgggttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	15960
cgagaagctc cggccaaatggc agtgcgcggc aaggcttgcg tttgttgcg tttgttgcg	16020
tcttggtaat ttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	16080
ttgttataaaa aggccgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	16140
tacaagtgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	16200
gccaacttgcg aaaaaggatc tttgttgcg tttgttgcg tttgttgcg tttgttgcg	16260
tttgcgatgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	16320
gaaaatccatc tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	16380
tcttgcataaa acgaaatccatc gatcgatcgc tttgttgcg tttgttgcg tttgttgcg	16440
aaaaattttca gatcgatcgc tttgttgcg tttgttgcg tttgttgcg tttgttgcg	16500
caatatttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	16560
gacatcgatcgc tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	16620
gtcccttagtca tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	16680
caatttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	16740
atcaatttttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	16800
atcaccttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	16860
tttgcgatgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	16920
ccttgcataaa acgaaatccatc gatcgatcgc tttgttgcg tttgttgcg tttgttgcg	16980
gatttccaaatc tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	17040
tttgcgatgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	17100
gaaatcttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	17160
caaaatcttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	17220
gttacatgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	17280
gacactcgatcgc tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	17340
aaattatcttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	17400
gccttgcgatcgc tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	17460
tactaccatc tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	17520
aacgcattatc tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	17580
gaatttggacc ctaataacgc cgaagcttgcg tttgttgcg tttgttgcg tttgttgcg	17640
ggtgactatc aaaaaggatc tttgttgcg tttgttgcg tttgttgcg tttgttgcg	17700
tttgcgatgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	17760
aaaggccatttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	17820
tataacttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	17880
caaaaaggcttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	17940
tactaccatc tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	18000
gatcccttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	18060
gggtgggttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	18120
cattctgata agtcaaaatgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	18180
gaaggccatgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	18240
ggtaaacaatc tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	18300
tttgcgatgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	18360
gctgaagccgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	18420
tcttggtaatc tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	18480
gggtgggttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	18540
agttcttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	18600
atatttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	18660
tttgcgatgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg tttgttgcg	18720

-continued

-continued

gcagccaaag aagctgcagc caaggccggtc tctgctgggt cagctgcagg ctctgggtgaa	4320
tttggttctgttctat gggttctcaa ttttgggttaa cttctcaaaa gactgaagct	4380
tccgagagat gtgggttgca aggctccat attttaagggtt cggaaagccgaa gaagcttacc	4440
ctacttacgc tgggtgcaca gagtcataata ttggaaacccc tttgttctg gccatataact	4500
ttatttagaa gatacggtat agatataagggtt atgttcaagt tcgaaggtt tagaagatgc	4560
ccaagcggtc ctggaaactttt tacatccatg acatcacaagaa gcaatgtat cttttagggca	4620
gttgaagcgtt ctatcaaca gcaaaaaggcc cagggttaaagggc tggacaggcc tcaagacat	4680
ctaagattgg aacaccatca ccatacatcat ggttctgggtt ctactaactt ctctttttt	4740
aagcaagcgtt gtggacgttga agaaaaatccaa ggttccatgtt gtttgcgtt actttgttaca	4800
ttcttcttccaa aacacgactt ccataatccatc acaacaacaa tcccaaaactt	4860
tctttgttgc ttgttgcatacatacataatgtt cttttttttt caatccca	4920
tcgaaacattt gtagtactaa gagttccatgtt cttttttttt agtgccttgc atctttgttct	4980
atcgctaaaga actcaatggc agtgcgtactt acaaatcaga cggaaacccccc tgagtcggat	5040
aatcaacttgc tagccaccaaa aattttgcactt ttttttttgcactt gtttgcgtt gtttgcgtt	5100
ccatacacaatataatgcctt caccctctgtt gtttgcgtt gtttgcgtt gtttgcgtt	5160
cataacacaa atttgcatttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt	5220
atcccttgc tccgccttctt accatccatg attaataatccaa ttttttttgcgtt gtttgcgtt	5280
agaattttttt acggccatgtt gtttgcgtt gtttgcgtt gtttgcgtt	5340
attatgttcaatccatgtt gtttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt	5400
ccatttgtaca ttttttttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt	5460
ccttttttttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt	5520
atcacaaacttccatgtt gtttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt	5580
agaccatcttccatgtt gtttgcgtt gtttgcgtt gtttgcgtt	5640
attaaagatgtt catctgttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt	5700
aagttacgtt gtttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt	5760
gcagccatgtt gtttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt	5820
agccacgcataatttttttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt	5880
tatgaccctg agggcgtt gtttgcgtt gtttgcgtt gtttgcgtt	5940
tatttttttttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt	6000
gttgggtgtt gtttgcgtt gtttgcgtt gtttgcgtt	6060
gacttacccaaatccatgtt gtttgcgtt gtttgcgtt gtttgcgtt	6120
gaggcttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt	6180
gaatatttttttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt	6240
ggcaatgtt gtttgcgtt gtttgcgtt gtttgcgtt	6300
tttggacttgc gtttgcgtt gtttgcgtt gtttgcgtt	6360
tataaaacagg gtttgcgtt gtttgcgtt gtttgcgtt	6420
cctaacaatgtt gtttgcgtt gtttgcgtt gtttgcgtt	6480
caaaaaatccatgtt gtttgcgtt gtttgcgtt gtttgcgtt	6540
tggtatagaa gggcgtt gtttgcgtt gtttgcgtt	6600
tacccaaatggcgtt gtttgcgtt gtttgcgtt gtttgcgtt	6660
tctgggtgtt gtttgcgtt gtttgcgtt gtttgcgtt	6720
gatttttttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt	6780
agggacaaatccatgtt gtttgcgtt gtttgcgtt gtttgcgtt	6840
attcccttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt	6900
ttcgggttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt	6960
ggccgttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt	7020
gatttttttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt	7080
tttggatagaaatccatgtt gtttgcgtt gtttgcgtt	7140
ccacttttttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt	7200
caaggcgtt gtttgcgtt gtttgcgtt gtttgcgtt	7260
tttggatagaaatccatgtt gtttgcgtt gtttgcgtt	7320
ccaaagaaaaatccatgtt gtttgcgtt gtttgcgtt	7380
aagctgtt gtttgcgtt gtttgcgtt gtttgcgtt	7440
aatttttttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt	7500
gtttccatgtt gtttgcgtt gtttgcgtt gtttgcgtt	7560
aggctgtt gtttgcgtt gtttgcgtt gtttgcgtt	7620
gtcgattttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt	7680
gaggcgtt gtttgcgtt gtttgcgtt gtttgcgtt	7740
tcttttttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt	7800
tatggacccctt gtttgcgtt gtttgcgtt gtttgcgtt	7860
gttaaatgtt gtttgcgtt gtttgcgtt gtttgcgtt	7920
tttggatagaaatccatgtt gtttgcgtt gtttgcgtt	7980
gcagtccatgtt gtttgcgtt gtttgcgtt gtttgcgtt	8040
aagtttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt	8100
acccttttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt	8160
acctacttttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt	8220
agtttccctt gtttgcgtt gtttgcgtt gtttgcgtt	8280
atccatatttttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt	8340
cttggatagaaatccatgtt gtttgcgtt gtttgcgtt	8400
ccttacccctt gtttgcgtt gtttgcgtt gtttgcgtt	8460
gctggatgtt gtttgcgtt gtttgcgtt gtttgcgtt	8520
ataccatccatgtt gtttgcgtt gtttgcgtt gtttgcgtt	8580
aaacaaagagg gtttgcgtt gtttgcgtt gtttgcgtt	8640
ccttacccctt gtttgcgtt gtttgcgtt gtttgcgtt	8700
atcaacacatgtt gtttgcgtt gtttgcgtt gtttgcgtt	8760
ttcgggttttttgcgtt gtttgcgtt gtttgcgtt	8820
ttcgggttttttgcgtt gtttgcgtt gtttgcgtt	8880
aacttctt gtttgcgtt gtttgcgtt gtttgcgtt	8940
agcacttttttgcgtt gtttgcgtt gtttgcgtt	9000

-continued

caaatcgtatcgccaaaccc acaggaaaaac tttttaaagt gtttctctga gtacatcccc 9060
aacaaccccg ctaaccctaa gtttatatac acacaacatg atcagctgtg catgaggcgt 9120
ttgaactcg ccattcaaaa ttggatgtt actcttcgaca ctacacccaa gccccttggt 9180
atagtaactc cttctatgt ctcacatata caagcttata tcttgcgtc taagaagg 9240
ggttgtcaga ttagaaacaag gtcgtgggtt cagcatgtt aaggtttatc ctatattatg 9300
caagtcggat ttgcccata gatgttggaa aatatgcata ctgtgaaagt tgacatcac 9360
tcacagactc ttgggtgtt gggatgttcc acatggggg aggtttacta ctgggttcaac 9420
gagatgttcc aaaaatggg ttccatggt ggttactgtt ccacagtcg tggttgtt 9480
cattttctg gtgggtgtt tggagctta atgagaatc acggtttgc tgacatataa 9540
atattgtacg cacatttgtt gacgttgc gttaaagtgc ttgacagaaa atcaatgggt 9600
gaaggtttt tttgggttcat caggtgtt ggtgggtt gttttttttt aatccggcc 9660
tgcggatata agttgttgc cgtatctggc aagactgtt ttttctgtt caaaaggaa 9720
atggaaatcc atgggttgc aaggttttgc aataatgttgc aacatcgca atacaatgc 9780
gataaggatt tgatgttac cagcatttca agaacttagga acatcacaga taaccatgtt 9840
aaaataaaga ctacatgttca cggatcttcc ttttcaattt ttttgggtt tggtgtt 9900
ttttgtgt tgatgtatata cttatccca gaactgggtt tttttttttt aatccggcc 9960
gaactgtactt gggatgttcc cttatccca ttttgggtt tttttttttt aacccggcc 10020
aacttcaaaa agggatctt gctggataga tccgtgttca aaaaagccgc tttttctatt 10080
aaaatgttactt acgtttaaaggaa actgtatccctt gaaactgttca ttttgggtt 10140
ctgtacgaaagg ggaaatgttcc cttatccca ttttgggtt tttttttttt aacccggcc 10200
gtatgttgc cttatccca ttttgggtt tttttttttt aacccggcc 10260
tggtacaccg ctacgtgggaa gaagcaagaa gataacggaa aacacataaa ctgggtcaga 10320
tctgtatatac acttcaatc accttacgtt ttttgggtt tttttttttt aacccggcc 10380
tacagatgtt tggactgtt ttttgggtt tttttttttt aacccggcc 10440
agaatttggg gtggaaatgtt ttttgggtt tttttttttt aacccggcc 10500
aaagccgatcc caaacaactt ttttgggtt tttttttttt aacccggcc 10560
catcatgtt cccggcttac taatcttctt ttttgggtt tttttttttt aacccggcc 10620
aatcccgatcc caatgttca gaaatgttca ttttgggtt tttttttttt aacccggcc 10680
gaaatcaactt actactgttca gggatgttca gaaatgttca ttttgggtt tttttttttt aacccggcc 10740
acagttgaca agctgttca gtttccatggt ttttgggtt tttttttttt aacccggcc 10800
accgttattttt gtaatgttca gtttccatggt ttttgggtt tttttttttt aacccggcc 10860
tctgttagaa agttggccat ttttgggtt tttttttttt aacccggcc 10920
atggcaactt ctggatgttcc gggatgttca ttttgggtt tttttttttt aacccggcc 10980
atggatgttcc cttatccca ttttgggtt tttttttttt aacccggcc 11040
atcgctgaaatgttcc gggatgttca ttttgggtt tttttttttt aacccggcc 11100
cctttgttac cttatccca ttttgggtt tttttttttt aacccggcc 11160
gaaatgttca ttttgggtt tttttttttt aacccggcc 11220
atggatgttcc ttttgggtt tttttttttt aacccggcc 11280
actgggttttcc ttttgggtt tttttttttt aacccggcc 11340
gatgggttttcc ttttgggtt tttttttttt aacccggcc 11400
gggttgggtt ttttgggtt tttttttttt aacccggcc 11460
gggttgggtt ttttgggtt tttttttttt aacccggcc 11520
gggttgggtt ttttgggtt tttttttttt aacccggcc 11580
gggttgggtt ttttgggtt tttttttttt aacccggcc 11640
gcaaaacccggcc aacatccatcc cttatccca ttttgggtt tttttttttt aacccggcc 11700
ggatgttcaatgtt ttttgggtt tttttttttt aacccggcc 11760
tttttgggtt ttttgggtt tttttttttt aacccggcc 11820
gttaattttcc ttttgggtt tttttttttt aacccggcc 11880
gatgttccatcc ttttgggtt tttttttttt aacccggcc 11940
aaatccatcc ttttgggtt tttttttttt aacccggcc 12000
tgcggatataatgtt ttttgggtt tttttttttt aacccggcc 12060
gaaactaaactt ttttgggtt tttttttttt aacccggcc 12120
aacatccatcc ttttgggtt tttttttttt aacccggcc 12180
cttccatcc ttttgggtt tttttttttt aacccggcc 12240
aaacccatcc ttttgggtt tttttttttt aacccggcc 12300
attactacatgtt ttttgggtt tttttttttt aacccggcc 12360
cccacccatcc ttttgggtt tttttttttt aacccggcc 12420
gggttgggtt ttttgggtt tttttttttt aacccggcc 12480
gacatgttcc ttttgggtt tttttttttt aacccggcc 12540
gggttgggtt ttttgggtt tttttttttt aacccggcc 12600
caacttccatcc ttttgggtt tttttttttt aacccggcc 12660
gggttgggtt ttttgggtt tttttttttt aacccggcc 12720
gggttgggtt ttttgggtt tttttttttt aacccggcc 12780
gggttgggtt ttttgggtt tttttttttt aacccggcc 12840
gggttgggtt ttttgggtt tttttttttt aacccggcc 12900
accatgttcc ttttgggtt tttttttttt aacccggcc 12960
catgttccatcc ttttgggtt tttttttttt aacccggcc 13020
aaacccatcc ttttgggtt tttttttttt aacccggcc 13080
gataaccatcc ttttgggtt tttttttttt aacccggcc 13140
aagtttccatcc ttttgggtt tttttttttt aacccggcc 13200
aaatccatcc ttttgggtt tttttttttt aacccggcc 13260
ccacccatcc ttttgggtt tttttttttt aacccggcc 13320
gataaccatcc ttttgggtt tttttttttt aacccggcc 13380
gggttgggtt ttttgggtt tttttttttt aacccggcc 13440
gggttgggtt ttttgggtt tttttttttt aacccggcc 13500
gggttgggtt ttttgggtt tttttttttt aacccggcc 13560
gccaatccatcc ttttgggtt tttttttttt aacccggcc 13620
tccaaatgtt ttttgggtt tttttttttt aacccggcc 13680
gttaccgttcc ttttgggtt tttttttttt aacccggcc 13740

-continued

-continued

gatcaagcaa	aaggtttagc	caatttggtc	caccacactg	cgaaagggttc	tgcgtggatca	18540
gctggcaggat	ccggcgaat	tggttctgt	gaaggccctg	caaaaaggacg	tgcgtccaaaa	18600
gttggatctg	caggtagtgc	tgcgttgtc	ggagaatttg	gttctggtgc	catggctact	18660
cctggttcag	aaaacgttct	accaaaaggaa	ccatgttgc	caacagccgt	gaaggttcttg	18720
cagaactcta	gagtttagaca	atctccattt	gcaacttagaa	gagcattttt	aaaaaaagaaa	18780
ggtttgcacc	acggaggaaat	tgtatgtgtt	tttccaaacagt	ccggtaactgc	agccgatgaa	18840
ccatcttcat	tgtggggaaat	ttggc				18864

SEQ ID NO: 206 moltype = DNA length = 19473
 FEATURE Location/Qualifiers
 source 1..19473
 mol_type = other DNA
 organism = synthetic construct
 SEQUENCE: 206
 atgggttcgt ctggttcage tgcagggtct ggtgaattcg gttccgctgg tagtgcgcgt 60
 ggttcgtgtt aatgggttc tgctggttca gcagccgggt ctggtaatt ctccttattac 120
 catccccatc accatcactt ggaatctact tcattataca aaaaggctgg ttccggtagt 180
 gccagaaacg cttacttgag aaagaaaatt gctagattga agaaagataa ttgcatttg 240
 gaaagagatg aacaaaactt ggaaaggatc atgcattt tgagagatg aatgcaga 300
 ttggaaaatg aagtgcgtc tcattgaaaca ggttccgag gtagtgcgc cggttctgt 360
 gaatttgcgt aaggccgtc aaaggaagcc gtcgcaaaag cagggttgc cggttgcgg 420
 gctggtagt gtgaattttc ttatcatc catcaccatc atcacttgg atctacactca 480
 ttatataaa aggccgtgtt cggtagtaac ttgggtcgtc aattggaaa tgaatgcgc 540
 tcattggaaa acgaaaaacg aactttgaaa aagaaaaattt tacataaaga agattgtatc 600
 gtttacttag aaaaaggat agcattttt gaaaggaaaa tagaagaagg ttccgttgcgtt 660
 agtgcagccg gtagtgggtg attcggtct gctgaagctg cagccaagg agctgcagcc 720
 aaagaagccg ctgtttaaaa agtcgcagcc aaagctgtt ctgcagggtc tgcccgagg 780
 tccggtagt ttgggtcttc atatcatc catcaccacc accacttgg atctacactca 840
 ttatataaaa aagtcgttgc cggtagtca aaggctgtc aattggaaa cagagtatgt 900
 gtttacttagt acgaaaaacg acaattggaa aataaggtag aagaattgaa aaatagaaaac 960
 gtttacttagt aaaaaggat tggcaactttt gaaaaggccg tagttagattt agaaaaacgt 1020
 gtttctgtgg gttctgtgtt ctgtgtgtt aatttcgttgc agcagccgtc 1080
 aaggaagccg cccgtttaaaa cgggttccggc ggttctgtc cgggctctgg tgaattttcc 1140
 tactatcacc atcatcatca ccacttggaa tctacatcat tatataaaga agccggttcc 1200
 ggttagtaatg aagttactac attggaaaac gatgtctt ttattggaaa cggaaaacgc 1260
 tacttagaaa aggaaatccg tagattgaga aagggaaaagg ccgttcttagg aaatagatta 1320
 gtttacttagt aaggctgtc tggttagcgtc gtttgcgtc tgtaatttttgg ttccgccc 1380
 gccgtgtcta aggaagccg tggccaaaagg gccgtgtccca aggaagccgc tgtaatggct 1440
 ggttccggcc gttcagctgc aggctctgtt gaattcggtt cttagaccacc taccatctt 1500
 aatccacccatc catgttattt cagtgtaaaa catccatccg tccgtatgtc aggttccgt 1560
 gccggcttgtc gcgaaatttgc cgaagctgtt gcccggaaag cagccgttca agctggttca 1620
 cgagggtccg ctggccgttcc tggttccatc aatttttttc aatctagacc aaactact 1680
 gtcctccatc aagaaatgtt cagatctgtt ggttccgtc gtccgcgcg aggatctggc 1740
 gaattttgtt ccgcggaaatc tgccgtttaaaa gaagctgtc cggaaaaggc agccgcacca 1800
 gaagctgtc cggaaaaggccg tagtgcgtt tgcgtccg gtttccgtga attcgttct 1860
 tccaaaggta ccgggttaaa tccaaaggatc aaagggtttccg aagaaatgtc ttccgtttaac 1920
 ggttctgtcg gttccgcage tggttccgtt gaatttcgcg aggccgtgc taaggaaagca 1980
 gcaaggccaaatc caggtagtgc tggttccgcga gtttccgttgc tgtaattttcc agacgggttgc 2040
 accactttccg aacattttgtt tgccgtttaa gacacttgc ttccgttgcggttccgttgc 2100
 ttcgtccgcg cgcagggttgc attcggttgc gccggatgtc ttccgttgcggttccgttgc 2160
 aaggaagctg ctgttacttagt aaggccgttca gtttccgttgc tgccgggttgc tgctgcgg 2220
 tcagggtgtt attttttttttccatc caccaccacc atcacttggaa atctacatcat 2280
 ttatataaaa aaggccgttgc tggttagtaag agaaatccgt acttaagaaa gaaaatccgt 2340
 gcatgttacttagt aaggatccgc aacattttggaa aaggacatccg ttccgttgcggttccgttgc 2400
 gaaagatgtt aaaaaggatc caaaccttgc gaaaatggaa ttgcatttca tgaacaaagg 2460
 tcagccgtt ctgcaggccgg ctccggtaa ttggccaaag ctgcaggaaa agaagctgc 2520
 gctaaggccgtt gtagtgcgtt ttccgttgc ggcaggccgtt aatttttttca caccaccatc 2580
 caccaccatc acttggatac ttccgttcaata tataaggaaaagg cagggttctgg tagtaactt 2640
 ttagcaatcatc taagatctac cgttgcgttgc ttggaaaaggc aaaaaccatgtt atttggaaa 2700
 gaaaaggaaa aattggaaaaggaa gggaaaaggaa caattttgttgc ataaggatgttgc agcttacaaa 2760
 ggttccgttgc gtttccgttgc ggggttccgttgc gaatttgcgtt ccggccaaaggc tgccggaaatc 2820
 gaagctgtcga cttaaaggatc cgttgcgttgc aagatgttgc cggaaaaggagg tagtgcgttgc 2880
 tccggccgcg gtcggccgttccatc aatttttttca ccccaatcc tccaccatcc 2940
 cttccgttcaata gatcttcaatc ttccgttcaata ggttccgttgc gtagtgcgtc tggtatccgttgc 3000
 gaatttccgttgc aagatgttgcgta taaggaaaagg cttccgttgc ggttccgttgc tggttccgttgc 3060
 gcaaggccgttgc tgtaatggaa attttttttccatc caccaccacc atcacttggaa atctacatcat 3120
 ttggaaaaggaa gggatggaaaaggatggaaaaggaa gggatggaaaaggaa gggatggaaaaggaa 3180
 gcaaggccgttgc ggggttccgttgc aatttttttca ccccaatcc tccaccatcc 3240
 cttccgttcaata gatcttcaatc ttccgttcaata ggttccgttgc gtagtgcgtc tggtatccgttgc 3300
 atttccatcaccaccatc gatcttccatc gccaaggatc catccgttgc tggtatccgttgc 3360
 tcagggtccgttgc tgtagtgcgttgc ttccgttgc ggggttccgttgc aatttttttca caccaccatc 3420
 ggttccgttgc gtttccgttgc aatttttttca gatcttcaatc ttccgttgc tggtatccgttgc 3480
 cttccgttcaata gatcttcaatc ttccgttcaata ggttccgttgc gtagtgcgtc tggtatccgttgc 3540
 tccgttcaata gatcttcaatc ttccgttcaata ggttccgttgc gtagtgcgtc tggtatccgttgc 3600
 gcaaggccgttgc tgtagtgcgttgc ttccgttgc ggggttccgttgc aatttttttca caccaccatc 3660
 gcaaggccgttgc tgtagtgcgttgc ttccgttgc ggggttccgttgc aatttttttca caccaccatc 3720
 ggttccgttgc gtttccgttgc aatttttttca gatcttcaatc ttccgttgc tggtatccgttgc 3780
 cttccgttcaata gatcttcaatc ttccgttcaata ggttccgttgc gtagtgcgtc tggtatccgttgc 3840
 ggttccgttgc tgtagtgcgttgc ttccgttgc ggggttccgttgc aatttttttca caccaccatc 3900

-continued

gctgggtccg	ctgcaggatc	tggcgaattc	ggttccgggg	aagccgcgc	aaaagaagcc	3960
gcccacaag	aagccgcgc	aaaggaaagcc	gcagcaagg	caggtatgc	cggctccgc	4020
gctggcagt	gcaatttgg	ttcttcata	tatcaccatc	atcatcatca	cttggaaatct	4080
acttcattat	acaagaagc	aggttccgg	tctaaaagaa	ttgttactt	aagaaagaaa	4140
atcgccgott	tgaagaaaga	caatgtcaac	tttagaaaaag	atattgccaa	cttggaaaaat	4200
gaaatcgaa	gattaattaa	ggaaattaa	acattggaa	acgaagttgc	atcacatgaa	4260
caagggttag	ctgggtccg	tgccgggtcc	ggcgaattt	cagaagcgc	cggcaaggaa	4320
gccgcagcca	aagctggtag	tgccaggttct	ggccgtgg	ctggcgaatt	ttcttactat	4380
catcatcacc	atcaccactt	ggaatctact	tcattataca	agaaagogg	ttcaggttct	4440
aacttgttag	caacttttaag	atctacagcc	gctgttttag	aaaatgaaaa	ccatgttcta	4500
gaaaaagaaa	aggaaagttt	gagaaggaa	aaggaaacaat	tataaaat	gtttagaaagcc	4560
tacaagggtt	cagcaggttc	cgcagcagg	tcaggcgaat	ttggttctgc	agaagcgcgt	4620
gctaaggaaag	ctgcggcaaa	ggaaagcgt	gctaaggagg	ccgctgcaaa	ggctggttct	4680
getgggtccg	tgccgggg	tggagaattt	ggttccgg	tgggtatga	cgccgtgtat	4740
atgaaacctc	cacccctaaa	taaccaaaa	ggtttaggtt	ccgctgttc	cgctgcagg	4800
tccggcaggt	tcgcagaagc	cgcagcaaa	gaagccgcag	ctaaggcagg	tagtgcgg	4860
tccggccgtc	gcagttggaa	attcaggaa	ttgttctgt	accatctt	cgtcaacgt	4920
caaaatttgg	ataaagcttag	acaaggttc	ggccgttctg	cagegggatc	tggggaaattt	4980
ggttctgcag	aagctgcgc	caaaaagct	gcagctaaag	aaggcccaagc	caaqaagct	5040
gctgctaagg	ccgggtctgc	tggttctgc	gcaggatctg	gggaatttgg	ttccaagaat	5100
actaagatgt	tgaatttgc	taacccagg	tacagaaaag	ctacagaaga	agaagggttca	5160
gcccgggtcg	ccgcgggttc	cggttgcattt	gcagagggt	cgcgttcaaa	ggctgcccgt	5220
aaggccggta	tgctgttgc	tgccggccgc	tccggaaat	ttagatctt	gccatccaca	5280
tggattgaaa	acaaattata	cggeatgtca	gaccctattt	ggggttctgc	aggttcagct	5340
gcgggatctg	gtgaatttgc	ttcagcagaa	ggccgcagcc	aggaaggccg	tgccaaaggag	5400
gecgctccca	aagaagcgc	tgcttgcagg	ggttcagccg	tttccgcagc	cgccgtgtt	5460
gaatttggta	gttgttgc	taatttgc	cctccagett	tgcccttcaa	gaaaagacaa	5520
tctgcttcat	ctgggttgc	aggttcaged	gttgggttgc	gtttagaaat	cgaagcagct	5580
gccaaggaaag	ctgcggccaa	ggccgggca	ggccgttgc	ctgcgggg	tggtaattt	5640
acttcaaaat	ctaaaccaca	acctgcagg	cctccaaag	cattctgtca	tttgatttt	5700
ggttccgcgc	tttccgcgc	tggttctgg	gaatttgc	ccgctgagg	tgccgttcaa	5760
gaagcggccgc	ctaaqaaggc	agccgttca	gaggccggcc	ctaaagcagg	tttgcaggt	5820
tcagcagcag	gttgttgc	attttggtt	acagatga	aaagagaaga	aaccgaagaa	5880
gaattttttt	tgttgttgc	tacccatgtt	tttttttttt	ggtttgcgtt	gggttcttgc	5940
gaatttgcag	aagcagctgc	ttaggaagcc	ggcccaaaa	tttttttttt	gggttccgc	6000
gcagggttgc	gtgaatttgc	tggtaatgt	tttttttttt	tttttttttt	tttttttttt	6060
gttagaaactt	attcatgcgg	tttgcgggtt	agtgcagccg	gtcttgcgg	attcggttcc	6120
gcagaagcgc	ctgcggcaaa	agccgttgc	aaagaaggcc	ctgcggaa	ggctgtgtt	6180
aaggccgggtt	ccgcgggtt	tgctgggtt	ttccggcaat	tttttttttt	tttttttttt	6240
catcatcacc	accacttgc	atccacaat	tttataaaat	aaagcttgc	tggtttccaa	6300
aaggtagtct	aattgaaaaa	taggttgc	tacaaggttt	aggaaaacgc	taatgttgc	6360
aacatagtag	caagatgtt	aaacgataat	gtaatttttgc	aaaaggacat	cgcaatatttgc	6420
gaaaaggata	tagtcaattt	ggaaaggat	tttttttttt	tttttttttt	tttttttttt	6480
ggctctggcc	aattcgttgc	agcttgcgtt	aaagaggcttgc	ccggctaaagc	tttttttttt	6540
ggttctgcag	ccccgggttgc	tgaatttttgc	tattatcacc	atcatcacca	tcacttgc	6600
tccaccagg	tatacaagaa	agccgttgc	tttttttttt	tttttttttt	tttttttttt	6660
catatcaag	aatttgc	aaagaaacgc	tttttttttt	tttttttttt	tttttttttt	6720
tttgcggaaa	ccggatgttgc	tttgcggat	tttttttttt	tttttttttt	tttttttttt	6780
tctggccgcg	gatcttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	6840
aaagagggtt	tcggcgttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	6900
tccggcgttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	6960
tctgggtat	aatccgcgg	tttgcggat	tttttttttt	tttttttttt	tttttttttt	7020
gccgctctca	aagggttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	7080
gaattttgtt	ccccaaatgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	7140
agaagaggata	tgctgttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	7200
gccaaggagg	ctggcgttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	7260
gttgggttgc	ccggcaggatc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	7320
caacttggat	ccacaaatgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	7380
catgaatttca	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	7440
aagtggaaa	gagaatttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	7500
tccggcgttgc	gtcttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	7560
ggttctggcc	gttccgcgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	7620
cacttggat	ccacaaatgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	7680
tttggaaaata	agatgttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	7740
catttggaa	aaagaaatgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	7800
agtggccgtt	ccgttgcgttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	7860
gttggccgcoca	aagaagcgc	ccggccaa	tttttttttt	tttttttttt	tttttttttt	7920
gttggccgtt	ccggccgttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	7980
gacagtgttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	8040
tttgcagagg	cagctgttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	8100
ggtagtgggt	aatttgcgttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	8160
gacttgggtc	atgttgcgttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	8220
gaagctgttgc	ccaaaggatc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	8280
gcccggcttgc	tttgggttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	8340
catcatcacc	acttggatc	cacaaggat	tataaggaa	ccgggttgc	tttccaaa	8400
gttggatcat	aaaaacaaa	gatttgc	tttgc	tttgc	tttgc	8460
gatatttgc	atttggaaa	ggaaatgc	tacgc	caggatgc	tttgc	8520
gccccggcttgc	tttgggttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	8580
gcagggttgc	tttgggttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	8640

-continued

gaatccacaa	gtttatacaa	gaaagcgggt	agtgaatttt	tcagaagaga	aagaacaag	8700
atggcagcgc	ctaagtgtag	aaacagaaga	agagaatttg	ctgatacatt	acaagctgaa	8760
acagatcaat	tagaagacga	aaaatcagct	ttgcaaaccg	aaatcgaaa	tttgttggaaa	8820
aaaaaaagaaa	aatttggatt	catttttagca	gcccatacag	cagcttgc	aatacctgtat	8880
gacttgggtt	ttccagaaga	aatgtctta	gaaggtagtgc	cggtagtgc	cgctggctca	8940
ggtagatttgc	gtatgtcaga	agctgcccgg	aaagaagccg	cagctaaaga	agctgcccgc	9000
aaagaggcgc	ccgcaaggc	aggttcaga	ggttcagctg	ccgggtccgg	ggaatttgg	9060
tcattccaa	tgccagctga	cactcctcca	cctgcattt	tgccacactg	agatcctatg	9120
acaggtagtgc	ccgggtctgc	tgcgggtt	ggcgaatttc	ctgaagcgc	tgctaaggag	9180
gctgcagctc	aggccatctc	tgcagggttc	gctgcagggt	cagggtgaa	tgaaagagaa	9240
tctaacaag	aaccacccctc	accttatgaa	gatccatact	ggggtaatgg	ttgttctgc	9300
ggtagtgcgc	ccggttcagg	cgaatttgg	tctgcggagg	ctgtgc	aaaagctgcg	9360
gccaaggaa	ctgcggcaaa	agaggtctgc	gccaaggccg	tttcagcagg	ttcagcagct	9420
gggtccgg	aatttgggt	cagtttatt	caccacate	atcaccactt	ggaatctacc	9480
tcattatata	agaaagcggg	ttccggtagt	caaaaatgtt	cagaatttga	aaacagagat	9540
gctgtcaat	taaatagaaa	tgagcagtt	aaaataagg	tcgaggagtt	aaaaaataga	9600
aacgcatact	tgaaaaat	atggctact	ttggaaaacg	agtcgc	aaatggaaat	9660
gatgttagtgc	aaggctctgc	tggttccgca	gccccgtca	gtgaaattc	cgaagcagcc	9720
gcaaggaa	ctgcggctaa	ggccggctca	qcaaggcttc	ccggccggaa	cggtgaaatt	9780
tcttattacc	accaccacca	tcaccactt	gaatctactt	cattatacaa	gaaagcgggg	9840
tccggtagta	acgaagtac	aaccttagaa	aatgtatgc	cctttataga	aaacgaaaat	9900
gcctacttag	aaaagaaat	tgcaagattt	agaaaggaaa	aagtcgcatt	gagaacacga	9960
ttagccccaca	aaacatctta	ctatcacca	catcatcatc	atcttgcatt	tacatcattt	10020
tacaaggaa	cgggctccgg	tagtgcata	aatgcctact	taagaaagaa	aatagccaga	10080
ttgaagaa	acaatgtca	attagagaga	gatgaacaga	acttagaaa	gattatagcc	10140
aatttggag	atgaaattgc	tagatttt	aatgaagtagt	cttctcatga	acaaggtagt	10200
gctggctcgc	ctgcggctc	cgggagattt	gcccgaatgc	ccggccaa	agccgcggcc	10260
aaggctgg	ccgtctgg	tgcggccga	tctggagaat	tttcttattt	ccatcatcat	10320
catcatcatt	ttgaatctac	atcattatac	aagaaacgg	gatctgg	ttacttgc	10380
gcccacatttgc	agaacaaat	cgcatcttac	gagaacaccc	acgaaaccc	gagaaaaaaa	10440
aacttacaca	aaaaggatttgc	gatagcttac	tttagaaaaaa	aaatcgctt	tttgagaaaa	10500
aaaatttgaag	aaggtagtgc	aggttcagcc	gctggctccg	gtgaaatttgg	ttcagcggag	10560
gctgcccgtt	aggaggcagc	cgctaaagaa	gcaaggcgtt	aggaggctgc	agcaaaagca	10620
ggttccgcag	gttctgcgc	gggttccgg	gatattttgg	ctgaaacaaa	gttgatctt	10680
gaagaagat	ttggaa	ttgtatattc	gagaagact	ttgaaacaaa	attaatatca	10740
gaagaagatt	ttggtagtgc	aggttcagca	gctggttctg	gagaatttgg	ttcagcagg	10800
tctgcccgt	gaagtggcga	attcggtat	gcccggctcc	ctgctgg	tggcgaattt	10860
ggttctgg	ctactaactt	cttcttgg	aagcaagcag	gtgacgttgc	agaaaatcc	10920
ggttcaatgg	gttctgttgc	ttcagctgc	ggttctgg	aatttgg	tcgcaggat	10980
gcccgtgg	ttgggttgcatt	cggttctgc	ggttgcag	ccgggttcttgc	tgaatttca	11040
tattaccatc	accatcacca	tcacttggaa	tccaccagt	tataaaaaaa	ggctgg	11100
ggtttagtgc	gaaacgcata	tttggagaaa	aaaatttgc	gattgaagaa	agataacttgc	11160
caatttggaa	gagatggaa	aaatttggaa	agattatgc	ccaacttgc	agatgaata	11220
gcaaggatgg	aaaacaaat	agcttctcat	gaacaagggt	ccgcaggtag	tgca	11280
tctggtaat	ttgtcgttgc	cgctgc	tttttttttt	tttttttttt	tttttttttt	11340
tcagccgttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	11400
acaaggtttat	ataaaaaaat	cggttcttc	tcaatttttt	ttgtcattt	agaaaaacgaa	11460
gtgcacatctt	tagaaaaat	aaacggca	ttggaaaaaa	aaaatttgc	taagaaagat	11520
tttgcattt	atttggaaa	ggaaatcgc	aacttggaa	agaaaataga	agaagg	11580
gtctgg	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	11640
gcagccaa	aaacccgtc	taaaaggat	gcagccaa	cagggttctgc	cggttctgc	11700
gcaggttcc	gttgcatttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	11760
ttaagtgcact	tcaagatata	agggttttcc	gctggttctg	cagccgg	tggcattt	11820
gcagaaggcag	ccgtcttgc	agcagccgt	aaagctgttgc	ctgcagg	ttgttccgg	11880
tctggtaat	tcgattttgc	tttgcatttgc	tttttttttt	tttttttttt	tttttttttt	11940
gacgctgttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	12000
gctgttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	12060
ggttccgcgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	12120
gacaaggat	aaaaaaat	tttttttttt	tttttttttt	tttttttttt	tttttttttt	12180
gctgttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	12240
gaattttgg	ccgtcttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	12300
tttctttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	12360
gcaattttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	12420
atccaaaata	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	12480
caagatcata	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	12540
agaagagat	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	12600
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	12660
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	12720
gtctacttag	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	12780
gtttgacccca	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	12840
attaaaaatgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	12900
ataatgtttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	12960
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	13020
gcccactaca	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	13080
gaagcttata	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	13140
aatgttgcata	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	13200
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	13260
ggttaataact	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	13320
tccttaatgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	13380

-continued

gcaaattta ctaacgttgc cgctcatttc aagggtatcg tcagagctat aagagattac 13440
 caaggccat tgaaggaaaca cgaagaataca atattcgta gaagagggtgg tccttaactac 13500
 caagaagggt tgagactgtc gggtaaagta ggtaaaaacca ctgttatacc aatccatgtc 13560
 ttggtagac aaaccacat gactgcataa gtggatgttgc ctttaggtca tagaccaatc 13620
 cctaatacac ctccaaacgc agccccacact gcaaattttc tgtaaacgcg ctctgggtca 13680
 acttccacac cagctccctt tagaacagca agtttcttg aatcaagac tgatgaagtc 13740
 gtcggcagta agaaagcaaa accggccatg cctcaagact cccgttccaa tcctatgatc 13800
 ttgcagggtt aatctactac ttgtttttt agacataatac aggtatagt atgggtatg 13860
 caaaacaaagag caggtaaagg catgttgatg ttgcactatg ttgttagatg agatgaacca 13920
 tctgttgcgtc caatggtcta tcctttact ggtgaccata agccaaaattt ctactgggt 13980
 cacaaggaaa tattgtatccc agtttttaag aacatggccg atgtatgag aaaacatcc 14040
 gaagtcgacg tattgtatcaa ctccgcctca ttaagatccg cttacgattc tacaatggaa 14100
 accatgaaact acgctcaat aagaaccatc gctatcatc cagaagggtt tccaaagggc 14160
 ttgactagaa agttgtatcaa gaaagctgtat caaaaagggtg tcacaataat cgggtccagct 14220
 accgttaggtg gtattaaggc ttgggttttca aagatcgtaa acactgttgg tatgttggat 14280
 aacatattgg catcaatggt gtatagacca ggttcaatcg cttacgatc cagaatgggt 14340
 ggtatgtcca acgaattgaa caacatcatc agtagaaacta cagatgggtt atacaagggt 14400
 ttgtgtatgg ttgggtacag atacccagggt tctacttttta tggtatcatgt attgagat 14460
 caagacacac ctgggtttaa aatgttggt ttgttgggtt aaatagggtt tactgttggaa 14520
 caaaatggat gcaagggat caaaagggtt agatgtacaa agccatcatgt ttgttgggtc 14580
 atttgttactt gtgcacaatc ttgttcttaa gaaatgttca tcggatcatgc aggttgcgtc 14640
 gtaatacaag ctccagaaac agcagttgcc aagaaccaag cattaaaaga agccgggttt 14700
 ttgttcccta gatcttgcg tgaattatggt gaaatcatc aatcgttctca tgaagacttg 14760
 gtagatcaatg gtgtatatttgc tccagcaca aaggttccctc caccatctgt ccctatgggt 14820
 tactcttggg caaagaaatggggttttaa aagaaaggccatgattttgcgacatccatgttata 14880
 tttgtgtatggaa gaggtcaaga attgatctat gctggatgtc ctattactga agtattcaaa 14940
 gaagaaatgg gtatcggtt tggttttaggt ttgttgggtt tccaaaaggag attgccaagg 15000
 tactcttgcg aattcttgcg aatgttgcg atggttacag ctgatcatgg tcctctgttc 15060
 tcagggtgcac aacatccatc aatctgcgtc agagcttgcgta aagatgggtt ttcttttttgc 15120
 acctcgatgtt tggttactat ttgtgtacaga ttgttgggtt cattagacgc cgctgcggaa 15180
 atgttttcaaa aagcttgcg ttccggatata atccaaatgg aattctgttta taatgtgaaa 15240
 aaggagggtt aattgtatcaat ggttgcgtt catcggttgc agtctatcaa taaccctgtat 15300
 attagaggtac aatcttgcg ggactatgtt agacaacact tttccagccac acctttgtta 15360
 gattacgtttc tggaaatgttca aaaaatggatcc acatcttcaaa agccaaatttt gatcttgcac 15420
 ttgtgtgtt taattgggttgc ttgttttgc gacatgttgc gaaactgttgg ttcttcaact 15480
 agagaagaaatgc ctgtatgatca ttgcacatt ggttgcgttgc atggatcttgc ttgtttaggt 15540
 agatctatgg ttttcatgg tccatcttcgatc gatcaaaaaggaa gatcaaaagca aggttgcgt 15600
 agacatccatc gggatgcatt ttcttcgttccatc ttacatggac acatgttcaat gaaaatgttgc 15660
 ggttgggttgc ttcttgcgttgg tgggtgttct ggttgggttgc gtagtgcgcg agcttgggtac 15720
 aattttgggttgc acgcataacta caagcagggtt gactaccaaa aggcaatgg atattaccaa 15780
 aaggcccttgg aatttagaccaaa ataaatcgcgaa gaaatgggttgcgatgggatcttgc 15840
 tataaaacagggtgtatcataacta aaaaatggccatc gatcaatggc aaaaaggccattt ggaatggatqac 15900
 cctaataacgc ctggaaatggatccatc ttatggatgttgc ttatggatgttgc ttatggatgttgc 15960
 cccaaagggttcc tggaaatggacttccaaagggttgc ttatggatgttgc ttatggatgttgc 16020
 gaagcttggaa agaaatgggttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 16080
 gaataatctatc aaaaatggatccatc ttatggatgttgc ttatggatgttgc ttatggatgttgc 16140
 ggttgcgtt actatataatggatccatc ttatggatgttgc ttatggatgttgc ttatggatgttgc 16200
 tttagaaatgg atccaaatataatggatccatc ttatggatgttgc ttatggatgttgc ttatggatgttgc 16260
 cagggtgtatccatc ttatggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 16320
 aacatgtatca gatcgatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 16380
 gtttcttcat attccatcaatc ccataccatc cacttggatccatc ttatggatgttgc ttatggatgttgc 16440
 gtttgcgttgcgtt gttcaatttgcgttgc gttcaatggatccatc ttatggatgttgc ttatggatgttgc 16500
 gaaacaggaaatccatc ttatggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 16560
 aaggaaatctcaatc ttatggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 16620
 ttctggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 16680
 gtttgcgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 16740
 cacttggatccatc ttatggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 16800
 tttagaaatggatccatc ttatggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 16860
 aatggaaatctcaatc ttatggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 16920
 gcatccatc ttatggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 16980
 gtttgcgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 17040
 atttggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 17100
 atcaaaatctcaatc ttatggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 17160
 ggttgcgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 17220
 gtttgcgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 17280
 tcgtatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 17340
 ggttgcgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 17400
 agatgtatggatccatc ttatggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 17460
 catggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 17520
 gaaatctcaatc ttatggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 17580
 ggttgcgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 17640
 ttcttccaaatctcaatc ttatggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 17700
 cctgtatc ttatggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 17760
 gccgtgtatc ttatggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 17820
 gtttgcgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 17880
 gtttgcgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 17940
 gtttgcgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 18000
 gtttgcgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 18060
 tcaatgtatcaatc ttatggatgttgc ttatggatgttgc ttatggatgttgc ttatggatgttgc 18120

-continued

gtttagtgc ctttgttat cggtgggtt caaggatcg caatggttat cgaaagattg 18180
 aataagtgt ctgggtggg tggctctgt ggtgggttt ctgggtgtt ggtagtgca 18240
 gaaggctgtt acaattttgg taacgttac tacaaggcgg gtactacca aaaggcaatc 18300
 gaatactacc aaaaggccctt ggaatttagat ccaaataacg ctgaagcatg gtatataattg 18360
 ggtatgcctt attataaaca gggtgactat caaaagota ttgaatatta caaaaggca 18420
 ttggaatttag atcctaataa cgccgaagct tggtataatt tgggttaacgc ctattataag 18480
 cagggtgact atcaaaaaggc catcgaagat tacaaaaaagg ctttggattt ggatccaac 18540
 aacttgcgaag cagaaggctt gagaatttg ggttaacgctt attacaaca agggtgactatc 18600
 caaaaggactt ttgaactata tcaaaaaggcc ttgaatttg ttgcctaataa cgcttcgtca 18660
 tggtaactattt tgggtatgtc ctactataaa cagggtgact accagaaggc tattgaataac 18720
 taccaaaaag cattagaattt ggtccaaat aacgccaagg cttggtagc acgtggtaat 18780
 gccttattaca agcagggtgtc ctaccggaaa gccatagaa actatcaaaa agcccttggat 18840
 ttggactcta acaacagatc cagaagtgtt ggtgggtgtt gttctgggtt tgggtgttgc 18900
 ggtgggtgtt gtgtcttcattt accatcatca caccatcacc atcacttggat atcatacatca 18960
 ttatacaaaa aggctggttc cggttagtaat gaagttacta cattggaaaa cgatggccgt 19020
 ttatcgaaa acgaaaacgc atacttggaa aaggaaatcg ccagtagtggaa aaggaaaaag 19080
 gcaaggcttgc gaaatagat agccatcataa aagggttccg ctggtagtgc tgccgttgc 19140
 ggttagattt gttcagctga agccctgtca aaaaaggccc ctgcggatcggc agtttgcgc 19200
 ggttcagccg ctgggtctgg tgaattcggt tccagttact atcaccatca ccatcatcac 19260
 ttggaaatcta cttcattata taaaaggcc ggttccggta gtcaaaaaggc cgctgttgc 19320
 aagaacagac tagctgttaa gttgaacaga aacgaacat tggaaaaataa ggttagaagaa 19380
 ttggaaaataa gaaaggccctt tttttttttt gaaatggccaa cattggaaaaa cgaaggctgc 19440
 agatggaaaaa atgtatgttgc agaagggtttt ggt 19473

SEQ ID NO: 207 moltype = DNA length = 7125
FEATURE Location/Qualifiers
source 1..7125
mol_type = other DNA
organism = synthetic construc

SEQUENCE: 207

atgagtgcta	aggcaattt	tgaacaact	ggtaaagaat	tgttgtaca	gttttattgt	60
actacatcg	ccatccaaa	tagatccaa	tacgctag	ttaccggcaga	tactgactgt	120
gcttagattt	tacaagata	tccatgggt	ttatctcaa	atctgggt	caaacctgc	180
caattatta	agagaagagg	taaattgggt	tttagtaggt	ttaatttgc	attggatgg	240
gtaaagtctt	ggttgaaacc	aagatttagt	caagaagcca	cagttgtaa	agctaccgt	300
ttctgtaaaa	attctctgt	cgaaccatt	gtcccttatt	cacaaggcga	agaatctat	360
gtatgtatct	acgctactgt	aggggtgac	tagttttt	tcatcacga	aggttgtgc	420
gacgtatgtt	acgttgacgc	caaggctca	aagtggttg	tttgtgtca	tgaaaagg	480
aaccaggaaag	acattaaaaaa	gcattttgtt	gttcacgcac	ctgaagataa	aaaggaaata	540
ttggccctcc	ttataaagg	tttggtaat	ttctacgaa	atttggactt	cacctactt	600
gaaataacc	catttagt	tactaaggat	ggtgtat	ttttggactt	agctcaaaa	660
gttgatgc	cagccgacta	catttggta	ccatcaatgg	ttgcacatcg	atttccac	720
ccatcggt	gagaagctt	tccagaagaa	gcctacattt	ctgatttgg	cgctaagtct	780
ggtgcatcat	tgaagg	atttgttgc	cctaaaggta	gaatttggc	catgttgtct	840
ggtgggtgtt	ctatgtgt	atatttgc	actatatgc	acttgggtt	ttgttacaa	900
ttggcaaac	acggtyataa	ctcgggtcc	ccatccgg	aaacaacata	cgattacgt	960
aagaccatct	tgtccttaat	gactagagaa	aagcatctt	atggtaaaat	cttgatcatc	1020
ggtggtagta	tcgcaattt	taactaacgtt	gccgtacat	tcaagggtat	cgtcagact	1080
ataaagagat	accagggtt	attgaaggaa	cacgaaat	caatatttgc	tagaagaggt	1140
ggtccaaact	accaaggaa	tttggagatc	atgggttgc	tagtttttt	cactgttata	1200
ccaatccatg	tctttggta	agaaaaccac	atgactgc	tagttgtat	ggccttaggt	1260
catagaccaa	tccctaata	accttccaa	gcagccca	ctgcaattt	cttggtaa	1320
gcctctgggt	caacttccac	accaggctt	tcttagaa	caatgttctc	tgaatca	1380
gtgtatgaag	tgcgtcc	taaaggaa	aaacaggca	tgcccaaga	ctccgttca	1440
agtccatagat	ctttggagg	taaatctact	actttttt	ctagacata	taaggctata	1500
gtatgggtt	tgcaaaaca	agcagttca	ggcatgtt	attcgacta	ttgttgtat	1560
agagatgtac	cattcttgtc	tgcaatgtt	tatctttt	ctgggtgacca	taaaggaaaa	1620
ttctactgg	gtcacaagg	aaatattgtc	ccagttttt	agaacatggc	cgatgtat	1680
agaaaacatc	ctgaagtca	cgtattgtt	atcttgcct	cattaaatc	cgcttgcgt	1740
tctacaatgg	aaaccatgaa	ctacgctaa	ataagaacca	tcgctatcat	tgcaaggtt	1800
attccagaa	ccttgactat	aaatgtt	aagaaagtc	atcaaaaagg	tgtcacaata	1860
ggctggcc	ctaccgtt	ttgtttaa	cctgggtt	tcaagatcg	taacactgtt	1920
ggtagtgtt	ataacatatt	ggcatctaa	ttgtat	tcgttgc	cgcttgcgt	1980
tccagaagtg	gtggatgtc	caacgattt	aacaacata	tcgttgcgt	tacagatgtt	2040
gtatagca	gtgttgttat	ttgttgt	agataccca	tttctactt	tatgtatcat	2100
gtttagat	atcaagacac	accttgg	aaaatgtt	ttgttgttgg	tgaataggt	2160
ggtagtca	aatacaat	atcgaggt	atcaaagg	tgatgtac	aaaggcaatc	2220
gttttgtt	gcatttgtac	tttgtca	atgttttt	cagaatgtca	attccgtcat	2280
geagggtgc	gcgctaaat	agcttca	acagcagtt	ccaagaacca	agcattaaaa	2340
gaagccgtt	tttttgtcc	tagatctt	gatgaattt	tgaaatcat	tcaatcgatc	2400
tatgaatgt	tggtagtca	tggtagtatt	ttttccac	aaagaaatcc	tccacccat	2460
gtccctatgg	attactctt	ggcaagagaa	ttgggtt	ttagaaagg	agctgtttt	2520
atgcattctt	tatgtatgt	aaaggttca	aaatgtt	atctgttgt	gccttactt	2580
gaagtattca	aagaagaaa	gggtatcgtt	ggtgtttt	tttttgtt	ttgttgcgt	2640
agattgcca	agtactctt	tcaattt	gaaatgtt	ttatgttac	agctgtatcat	2700
ggtcctctgt	tctcagg	acacaaat	ataatctt	cttgcgtt	taaagattt	2760
gtttttctt	tgacactt	tttggtaat	tttgttgc	gattttgtt	tgccatagac	2820
ggccgtc	caagat	ttttttt	aaaagg	tttttttt	tttttttt	2880
aataagat	aaaaggagg	taaattt	tttttttt	tttttttt	tttttttt	2940

-continued

-continued

SEQ ID NO: 209	moltype = AA length = 723
FEATURE	Location/Qualifiers
source	1..723
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 209	
MGKNYKSLDS VVASDFIALG ITSEVAETLH GRLAEIVCNY GAATPQTWIN IANHILSPDL	60
PFLSLHQMLFY GCYKDFGPAP PAWIPDPEKV KSTNLGALLE KRGKEFLGVK YKDPISSFSH	120
FQEFQFVNPYE VYWRTVLMD MKISFSKDPPE CILRRDDINN PGGSEWLPGG YLNSAKNCLN	180
VNSNNKKLNDT MIVWRDEGND DLPLNKLTLD QLRKRVWLVG YALEEMGLEK GCAIAIDMPM	240
HVDAAVVIYLA IVALAGYVVVS IADFSFAPEI STRLRLSKAK AIFTQDHIIIR GKKRIPLYSR	300
VVEAKSPMAI VIPCGSGSNIG AELRDGDISW DYFLERAKEF KNCEFTAREQ PVDAYTNILF	360
SSCTTGEPKA IPWTQATPLA AAADGWSHLD IRKGDVIVWP TNLGWMMPGW LVYASLLNGA	420
SIALYNGSPLKA AKVTMLGVVP SIVRSWKSTN CVSGYDWSTI RCFSSSGEAS	480
NVDEYLWLMG RANYKPVIEM CGGTETIGGAF SAGSFLOQAOQS LSSFSSOCMG CTLYILDKNG	540
YPMPKNKPGI GELALGPVMF GASKTLLNGN HHDDVYFKGMP TLNGEVLRHH GDIFELTSNG	600
YYHAHGRADD TMNIIGGIKIS SIEIERVCNE VDDRVFETTA IGVVPLGGP EQLVIFVFV рк	660
DSNDTTIDLN QRLLSFNGLQ QKQLNPLFKV TRVVPLSSLP RTATNKIMRR VLRQQFSHFE	720
GSG	723
SEQ ID NO: 210	moltype = AA length = 127
FEATURE	Location/Qualifiers
source	1..127
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 210	
GSHMGSQFWV TSQKTEASER CGLQGSYIILR VEAELKLLT LGAQSQILEP LLFWPYTLLR	60
RYGRDKVMFS PEAGRRCPSG PGTFTFQTSQ GNDIFQAVEA AIQQQKAQGK VGQAQDILRL	120
EHHHHHH	127
SEQ ID NO: 211	moltype = AA length = 2741
FEATURE	Location/Qualifiers
source	1..2741
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 211	
MSEESLFESS POKMEYEITN YSERHTELPQ HFIGLNTVDK LEESPLRDFV KSHGGHTVIS	60
KILIANNGIA AVKEIRSVRK WAYETFGDRR TVQFVAMATP EDLEANAEEYI RMADQYIEVP	120
GGTNNNNNYAN VDLIVIDAER ADVDAWAGW GHASENPLLP EKLSQSQRKV IFIGPPGNAM	180
RSLGDKISST IVAQSAKVP CIPWSGTGVDT VHDEKTGLV SVDDDIYQKG CCTSPEDGLQ	240
KAKRIGFPVM IKASEGGGGK GIRQVEREED FIALYHQAN EIPGSPFIM KLAGRARHLE	300
VQLLADQYGT NISLPRGDCS VQRQHQKII EAPVTIAKAE TFHEMEKAAB RLGKLVGYVS	360
AGTVEYLISH DDGKFYFLFEL NPRQLQVEHPT TEMVSGVNLP AAQLQIAMGI PMHRISDIRT	420
LYGMNPHSAS RIDFEFKTQD ATKQORRPIP KGHTACRIT SEDPNQDFKP SGGLTLHELNF	480
RSSSNVWGYF SVGNNNGNIHS FSDSQFGHIF AFGENRQASR KHMVVALKE L SIRGDFRTTV	540
EYLIKLLETE DFEDNTITTC WLDDLITHKM TAEKPDPTLA VICGAATKAF LASEEARHKY	600
IESTLQKGVL QSDLLQTMFP WDFIHEKGRL KFTVAKSGND RTYLFINGSK CDIILRQLSD	660
GGGLIAIGGK SHTIWKEEV AATRLSVDSM TTLLEVENDP TQLRTPSPKL LVKFLVENE	720
HIIGKQPYAE IEVMKMQMPL VSQENGIVQL LKQPGSTIVA GDIMAINTLD DPSVKHALP	780
FEGMLPDFGS PVIETKTPAY KFKSLVSTLE NILKQYDMQV IMNASLQQLI EVLNRPKLY	840
SEWKLHISAL HSRLPAKLDE QMEELVARSL RRGAVFPARQ LSKLIDMAVK NPEYNPDKLL	900
GAVVEPLADI AHKYSNGLEA HEHSIFVHPL EEEYVEVLF NGPNVREENI ILKLRDENPK	960
DLDKVALTVL SHSKVSAKNN LILAILKHYQ PLCKLSSKVS AIFSTPLQHI VELESKATAK	1020
VALQARILLI OGALPVSKEF TEQIEHILKS SVVKVAYGSS NPKRSEPDLN ILKDLIDSNA	1080
VVFDVLLQFL THQDPVTTAA AAQVYIRRAY RAYTIGDIRV HEGTVPIVE WKFQLPSAAF	1140
STFPTVSKSM CMNRAVSVD LSYVANSQSS PLREGILMAV DHLDDVDEIL SQSLEVIPRH	1200
QSSSNPGPAPD RSGSSASLSN VANCVASTE GFESEEEILV RLREILDLNK QELINASIRR	1260
ITPMFGFKDG SYPKYTFNG PPNYNETETR HIEPALAQFL ELGRLSNFNI KPIFTDNRNI	1320
HVYEAVSKRD IIRTGHIRDT ISIQEYLTSE ANRLMSDILD NLEVTDTSNS	1380
DLNHIFINFI AVFDISPEDV EAAFCGFLER FGKRLRLRV SSAEIRRIIK DPQTGAPVPL	1440
RALINNVSgy VIKTEMYTEV KNAKGEWVK SLGKPGSMHL RIATPYPVK EWLQPKRYKA	1500
HIMGTTYYD FPFLPFRQASS SQWNKNSADV KLTDFFISN ELIEDENGEL TEVEREPGAN	1560
AIGMVAFKIT VKTPPEYPRGR QFVVFVANDIT FKIGSFGPQK DFFNKVTEY ARKRGIPRIY	1620
LAANSGARIG MAAEIVPLFQ VAWNDAANPD KGQYLYLTS EGMETLKFKD KENSVLTERT	1680
VINGEERFVI KTIIGSEDGL GVECLRGSGL IAGATSRAYH DIFTITLVT C RSVGIGAYLV	1740
RLQORAIOVE QOPIIILTGP AINKMLGREV YTSLNQLCGT QIMYNNGVSH LTAVDDLAGV	1800
EKIVEWMSYV PAKRNMPVP1 LETKDTWDTRP VDFTPTNDT YDVRWMIEGH ETESGFEYGL	1860
FDKGSFFETL SWGRNKGWVVG RARLGGIPLG VIGVETRTVE NLIPADPANP NSAETLQEP	1920
GQVWHPNPSAF KTAQAIINDFN NGEQLPMMIL ANWRGFSGQQ RDMFNEVLYK GSFIVDALVD	1980
YKQPIIIYIP PTGELRGGSW VVVDPTINAD QMEMYADVN A RAVGLEPQGM VGIKFRREKL	2040
LDTMNRLDDK YRELRSQSLN KSLAPEVHQO ISKQLADRER ELLPIYGQIS LQFADLHDRS	2100
SRMVAKGVIS KELEWEARKE FFFWRLRRRL NEEYLIKRLS HQVGEASRLE KIARIRSWYP	2160
ASVDHEDDRQ VATWIEENYK TLDKLKGLK LESFAQDLAK KIRSDHDNAI DGLSEVIKML	2220
STDDEKELLK TLKKLSSGGG SGGGGGGGG SABAWYNLGN AYYKQGDYQK AIEYYQKALE	2280
LDPNNAEAWY NLGNAYYKQG DYQKAIEYYQ KALELDPNNA EAWYNLGNAY YKQGDYQKAI	2340
EDYQKALELD PNBLQAEAWK NLGNAYYKQG DYQKAIEYYQ KALELDPNNA SAWYNLGNAY	2400
YKQGDYQKAI EYYQKALELD PNNAKAWYRR GNAYYKQGDY QKAIEDYQKA LEIDPNNRSR	2460
SAGGGGGGGG GSAGGGASGS HMRLGAQSIQ PTANLDRDTD LVYLNVMELV RAVLELKNE	2520
AQLPPEGYVV VVKNVGLTLR KLIGSVDDLL PSLPSSSRTE IEGTQKLLNK DLAELINKMR	2580

-continued

LAQQNAVTLS SSEECKRQMLT ASHTLAVDAK NLLDAVDQAK VLANLAHPPA EGSAGSAAGS	2640
GEFGSAAAAA KEAAAKAGSA GSAAGSGEFG SGAMATPGSE NVLPREPLIA TAVKFLQNSR	2700
VRQSPLATRR AFLKKKGLTD EEIDMAFQQS GTAADEPSSL W	2741

SEQ ID NO: 212 moltype = DNA length = 57
 FEATURE Location/Qualifiers
 source 1..57
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 212
 gctactaact tcttttggtaa agagcaagca ggtgacgttg aagaaaatcc aggtccaa 57

What is claimed is:

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.

217

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-motif1-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.

218

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.

* * * * *