

(51) International Patent Classification:

<i>C12P 7/04</i> (2006.01)	<i>C12P 11/00</i> (2006.01)
<i>C12P 7/26</i> (2006.01)	<i>C12P 13/00</i> (2006.01)
<i>C12P 7/42</i> (2006.01)	<i>C12P 19/32</i> (2006.01)
<i>C12P 7/44</i> (2006.01)	<i>C12N 1/21</i> (2006.01)

(21) International Application Number:

PCT/US2016/061287

(22) International Filing Date:

10 November 2016 (10.11.2016)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/255,303 13 November 2015 (13.11.2015) US

(71) **Applicant:** INVISTA NORTH AMERICA S.A.R.L.
[LU/US]; Three Little Falls Centre, 2801 Centerville Road,
Wilmington, DE 19808 (US).

(72) **Inventors: BOTES, Adriana, Leonora**; Red House Farm, Daleside Road, Rosedale East YO18 8RL (GB). **CONRADIE, Alex, Van Eck**; 4 Aberdovey Drive, Eaglescliffe, Cleveland TS16 9EZ (GB).

(74) **Agents:** CHAPMAN, Ernest, F. et al.; Finnegan, Henderson, Farabow, Garrett & Dunner, LLP, 901 New York Avenue, NW, Washington, DC 20001-4413 (US).

(81) Designated States (*unless otherwise indicated, for every*

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

(84) Designated States (*unless otherwise indicated, for every*

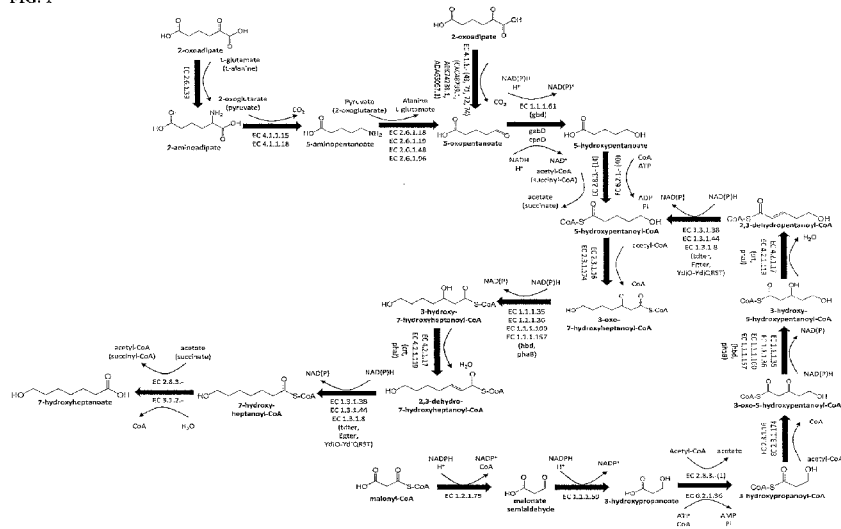
kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS,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

Published:

— with international search report (Art. 21(3))

(54) Title: METHODS AND MATERIALS FOR PRODUCING 5 AND 7-CARBON MONOMERS

FIG. 1



(S7) Abstract: This document describes biochemical pathways for biosynthesizing a 3-oxo-7-hydroxyheptanoyl-CoA intermediate using a *β -ketothiolase*, and enzymatically converting 3-oxo-7-hydroxyheptanoyl-CoA to 7-hydroxyheptanoic acid. 7-hydroxyheptanoic acid can be further enzymatically converted to pimelic acid, 7-aminoheptanoic acid, heptamethylenediamine or 1,7-heptanediol. This document also describes recombinant hosts producing 7-hydroxyheptanoic acid as well as pimelic acid, 7-aminoheptanoic acid, heptamethylenediamine and 1,7-heptanediol.

Methods and Materials for Producing 5 and 7-Carbon Monomers

CROSS-REFERENCE

This application claims the benefit of U.S. Provisional Application No. 62/255,303, filed November 13, 2015, the entire contents of which are incorporated herein by reference.

TECHNICAL FIELD

This invention provides materials and methods for biosynthesizing 5 carbon and 7 carbon monomers. For example, the invention provides 3-oxo-7-hydroxyheptanoyl-CoA using a β -ketothiolase, and enzymatically converting 3-oxo-7-hydroxyheptanoyl-CoA to 7-hydroxyheptanoic acid using one or more of an isolated *3-hydroxyacyl-CoA dehydrogenase*, a *3-oxoacyl-CoA reductase*, an *enoyl-CoA hydratase*, a *trans-2-enoyl-CoA reductase*, and a *thioesterase*, or using host having at least one exogenous nucleic acid capable of making one or more of such enzymes. This invention also provides methods for converting 7-hydroxyheptanoic acid to one or more of pimelic acid, 7-aminoheptanoic acid, heptamethylenediamine, and 1,7-heptanediol using one or more isolated enzymes such as *dehydrogenases*, *reductases*, *hydratases*, *thioesterases*, *monooxygenases*, and *transaminases* or using host cells expressing one or more such enzymes.

BACKGROUND

Nylons are polyamides which are generally synthesized by the condensation polymerization of a diamine with a dicarboxylic acid. Similarly, Nylons also may be produced by the condensation polymerization of lactams. Nylon 7 is produced by polymerisation of 7-aminoheptanoic acid, whereas Nylon 7,7 is produced by condensation polymerisation of pimelic acid and heptamethylenediamine. No economically viable petrochemical routes exist for producing the monomers for Nylon 7 and Nylon 7,7.

Given that there are no economically viable petrochemical monomer feedstocks, biotechnology offers an alternative approach via biocatalysis. Biocatalysis is the use of

biological catalysts, such as enzymes, to perform biochemical transformations of chemical compounds.

Both bioderived feedstocks and petrochemical feedstocks can be viable starting materials for the biocatalysis processes.

SUMMARY

Accordingly, against this background, it is clear that there is a need for sustainable methods for producing one or more of pimelic acid, 7-hydroxyheptanoic acid, 7-aminoheptanoic acid, heptamethylenediamine, and 1,7-heptanediol or derivatives thereof, wherein the methods are biocatalyst based. This document is based at least in part on the discovery that it is possible to construct biochemical pathways for using, *inter alia*, a β -ketothiolase to produce, for example, 7-hydroxyheptanoate, which can be converted in one or more enzymatic steps to pimelic acid, 7-aminoheptanoic acid, heptamethylenediamine, or 1,7-heptanediol. "Pimelic acid" and "pimelate," "7-hydroxyheptanoic acid" and "7-hydroxyheptanoate," and "7-aminoheptanoic acid" and "7-aminoheptanoate" are used interchangeably herein to refer to the compound in any of its neutral or ionized forms, including any salt forms thereof. It is understood by those skilled in the art that the specific form will depend on *inter alia* pH.

For compounds containing carboxylic acid groups such as organic monoacids, hydroxyacids, aminoacids and dicarboxylic acids, these compounds may be formed or converted to their ionic salt form when an acidic proton present in the parent compound either is replaced by a metal ion, e.g., an alkali metal ion, an alkaline earth ion, or an aluminum ion; or coordinates with an organic base. Acceptable organic bases include ethanolamine, diethanolamine, triethanolamine, tromethamine, N-methylglucamine, and the like. Acceptable inorganic bases include aluminum hydroxide, calcium hydroxide, potassium hydroxide, sodium carbonate, sodium hydroxide, and the like. The salt can be isolated as is from the system as the salt or converted to the free acid by reducing the pH to below the pKa through addition of acid or treatment with an acidic ion exchange resin.

For compounds containing amine groups such as but not limited to organic amines, aminoacids and diamine, these compounds may be formed or converted to their

ionic salt form by addition of an acidic proton to the amine to form the ammonium salt, formed with inorganic acids such as hydrochloric acid, hydrobromic acid, sulfuric acid, nitric acid, phosphoric acid, and the like; or formed with organic acids such as acetic acid, propionic acid, hexanoic acid, cyclopentanepropionic acid, glycolic acid, pyruvic acid, lactic acid, malonic acid, succinic acid, malic acid, maleic acid, fumaric acid, tartaric acid, citric acid, benzoic acid, 3-(4-hydroxybenzoyl)benzoic acid, cinnamic acid, mandelic acid, methanesulfonic acid, ethanesulfonic acid, 1,2-ethanedisulfonic acid, 2-hydroxyethanesulfonic acid, benzenesulfonic acid, 2-naphthalenesulfonic acid, 4-methylbicyclo-[2.2.2]oct-2-ene-1-carboxylic acid, glucoheptonic acid, 4,4'-methylenebis-(3-hydroxy-2-ene-1-carboxylic acid), 3-phenylpropionic acid, trimethylacetic acid, tertiary butylacetic acid, lauryl sulfuric acid, gluconic acid, glutamic acid, hydroxynaphthoic acid, salicylic acid, stearic acid or muconic acid. Acceptable inorganic bases are known in the art and include aluminum hydroxide, calcium hydroxide, potassium hydroxide, sodium carbonate, sodium hydroxide, and the like. The salt can be isolated as is from the system as a salt or converted to the free amine by raising the pH to above the pK_b through addition of base or treatment with a basic ion exchange resin.

For compounds containing both amine groups and carboxylic acid groups such as but not limited to aminoacids, these compounds may be formed or converted to their ionic salt form by either 1) acid addition salts, formed with inorganic acids such as hydrochloric acid, hydrobromic acid, sulfuric acid, nitric acid, phosphoric acid, and the like; or formed with organic acids such as acetic acid, propionic acid, hexanoic acid, cyclopentanepropionic acid, glycolic acid, pyruvic acid, lactic acid, malonic acid, succinic acid, malic acid, maleic acid, fumaric acid, tartaric acid, citric acid, benzoic acid, 3-(4-hydroxybenzoyl)benzoic acid, cinnamic acid, mandelic acid, methanesulfonic acid, ethanesulfonic acid, 1,2-ethanedisulfonic acid, 2-hydroxyethanesulfonic acid, benzenesulfonic acid, 2-naphthalenesulfonic acid, 4-methylbicyclo-[2.2.2]oct-2-ene-1-carboxylic acid, glucoheptonic acid, 4,4'-methylenebis-(3-hydroxy-2-ene-1-carboxylic acid), 3-phenylpropionic acid, trimethylacetic acid, tertiary butylacetic acid, lauryl sulfuric acid, gluconic acid, glutamic acid, hydroxynaphthoic acid, salicylic acid, stearic acid, muconic acid. Acceptable inorganic bases include aluminum hydroxide, calcium

hydroxide, potassium hydroxide, sodium carbonate, sodium hydroxide, and the like or 2) when an acidic proton present in the parent compound either is replaced by a metal ion, e.g., an alkali metal ion, an alkaline earth ion, or an aluminum ion; or coordinates with an organic base. Acceptable organic bases are known in the art and include ethanolamine, diethanolamine, triethanolamine, tromethamine, N-methylglucamine, and the like. Acceptable inorganic bases are known in the art and include aluminum hydroxide, calcium hydroxide, potassium hydroxide, sodium carbonate, sodium hydroxide, and the like. The salt can be isolated as is from the system or converted to the free acid by reducing the pH to below the pKa through addition of acid or treatment with an acidic ion exchange resin.

In the face of the optimality principle, it surprisingly has been discovered that appropriate non-natural metabolic pathways, feedstocks, cells (hosts or microorganisms), attenuation(s) to a cell's biochemical network, and/or cultivation strategies may be combined to efficiently produce 7-hydroxyheptanoate as a C7 (7-carbon) building block, or convert 7-hydroxyheptanoate to other C7 building blocks such as pimelic acid, 7-aminoheptanoic acid, heptamethylenediamine, or 1,7-heptanediol.

In some embodiments, a terminal carboxyl group can be enzymatically formed using a *thioesterase*, an *aldehyde dehydrogenase*, a *6-oxohexanoate dehydrogenase*, a *7-oxoheptanoate dehydrogenase*, or a *monooxygenase* (e.g., in combination with an *oxidoreductase* and ferredoxin). See FIG. 1 and FIG. 2.

In some embodiments, a terminal amine group can be enzymatically formed using a *ω -transaminase* or a *deacylase*. See FIG. 4. The *ω -transaminase* can have at least 70% sequence identity to any one of the amino acid sequences set forth in SEQ ID NOs.: 7 – 12.

In some embodiments, a terminal hydroxyl group can be enzymatically formed using an *alcohol dehydrogenase*. See FIG. 1 and FIG. 5.

In one aspect, this document features a method of producing 3-oxo-7-hydroxyheptanoyl-CoA. The method includes enzymatically converting 5-hydroxypentanoyl-CoA to 3-oxo-7-hydroxyheptanoyl-CoA using a *β -ketothiolase* classified under EC. 2.3.1.- (e.g., EC 2.3.1.16 or EC 2.3.1.174). The *β -ketothiolase* can

have at least 70% sequence identity to the amino acid sequence set forth in SEQ ID NO:1 or SEQ ID NO:13.

In one aspect, this document features a method of producing 3-oxo-7-hydroxyheptanoyl-CoA. The method includes enzymatically converting 5-hydroxypentanoyl-CoA to 3-oxo-7-hydroxyheptanoyl-CoA using a *thiolase*. The *thiolase* can have at least 70% sequence identity to the amino acid sequence set forth in any one of SEQ ID NOs: 15-36.

In one aspect, this document features a method of producing 5-hydroxypentanoyl-CoA. The method includes enzymatically converting 5-hydroxypentanoate to 5-hydroxypentanoyl-CoA using a *5-hydroxyvalerate CoA transferase*. The *5-hydroxyvalerate CoA transferase* can have at least 70% sequence identity to the amino acid sequence set forth in SEQ ID NO: 37.

In one aspect, this document features a method of producing 3-oxo-7-hydroxyheptanoyl-CoA from 5-hydroxypentanoate in a two-step enzymatic reaction catalyzed by a *5-hydroxyvalerate CoA transferase* and a *thiolase*. The *5-hydroxyvalerate CoA transferase* can have at least 70% sequence identity to the amino acid sequence set forth in SEQ ID NO: 37. The *thiolase* can have at least 70% sequence identity to the amino acid sequence set forth in any one of SEQ ID NOs: 15-36.

In one aspect, this document features a method for enzymatically converting 3-oxo-7-hydroxyheptanoyl-CoA to 7-hydroxyheptanoate using a *3-hydroxyacyl-CoA dehydrogenase* or a *3-oxoacyl-CoA reductase*, an *enoyl-CoA hydratase*, a *trans-2-enoyl-CoA reductase*, and a *thioesterase* or a *CoA transferase*. The *3-hydroxyacyl-CoA dehydrogenase* or *3-oxoacyl-CoA reductase* can be classified under EC 1.1.1.35, EC 1.1.1.36, EC 1.1.1.100, or EC 1.1.1.157. The *enoyl-CoA hydratase* can be classified under EC 4.2.1.17 or EC 4.2.1.119. The *trans-2-enoyl-CoA reductase* can be classified under EC 1.3.1.38, EC 1.3.1.44, or EC 1.3.1.8.

In one aspect, this document features a method for biosynthesizing 7-hydroxyheptanoate. The method includes enzymatically synthesizing 3-oxo-7-hydroxyheptanoyl-CoA from 5-hydroxypentanoyl-CoA using a β -ketothiolase classified under EC. 2.3.1.- (e.g., EC 2.3.1.16 or EC 2.3.1.174) and enzymatically converting 3-oxo-7-hydroxyheptanoyl-CoA to 7-hydroxyheptanoate. The β -ketothiolase can have at least 70% sequence identity to the amino acid sequence set forth in SEQ ID NO:1 or SEQ

ID NO:13. 3-oxo-7-hydroxyheptanoyl-CoA can be converted to 3-hydroxy-7-hydroxyheptanoyl-CoA using a *3-hydroxyacyl-CoA dehydrogenase* or a *3-oxoacyl-CoA reductase*, 3-hydroxy-7-hydroxyheptanoyl-CoA can be converted to 2,3-dehydro-7-hydroxyheptanoyl-CoA using an *enoyl-CoA hydratase*, 2,3-dehydro-7-hydroxyheptanoyl-CoA can be converted to 7-hydroxyheptanoyl-CoA using a *trans-2-enoyl-CoA reductase*, and 7-hydroxyheptanoyl-CoA can be converted to 7-hydroxyheptanoate using a *thioesterase* or a *CoA transferase*.

Any of the methods further can include enzymatically converting 7-hydroxyheptanoate to pimelic acid, 7-aminoheptanoate, heptamethylenediamine, or 1,7-heptanediol in one or more steps.

For example, 7-hydroxyheptanoate can be enzymatically converted to pimelic acid using one or more of a *monooxygenase*, an *alcohol dehydrogenase*, a *4-hydroxybutyrate dehydrogenase*, a *5-hydroxyvalerate dehydrogenase*, a *6-hydroxyhexanoate dehydrogenase*, a *7-oxoheptanoate dehydrogenase*, a *6-oxohexanoate dehydrogenase*, a *5-oxovalerate dehydrogenase*, or an *aldehyde dehydrogenase*.

For example, 7-hydroxyheptanoate can be converted to 7-aminoheptanoate using one or more of an *alcohol dehydrogenase*, a *6-hydroxyhexanoate dehydrogenase*, a *5-hydroxypentanoate dehydrogenase*, a *4-hydroxybutyrate dehydrogenase*, and a *ω-transaminase*. The *ω-transaminase* can have at least 70% sequence identity to any one of the amino acid sequences set forth in SEQ ID NOs.: 7 – 12.

For example, 7-hydroxyheptanoate can be converted to heptamethylenediamine using one or more of a *carboxylate reductase*, a *ω-transaminase*, an *alcohol dehydrogenase*, an *N-acetyltransferase*, and an *acetylputrescine deacylase*. The *ω-transaminase* can have at least 70% sequence identity to any one of the amino acid sequences set forth in SEQ ID NOs.: 7 – 12.

For example, 7-hydroxyheptanoate can be converted to 1,7-heptanediol using a *carboxylate reductase* and an *alcohol dehydrogenase*. The *carboxylate reductase* can have at least 70% sequence identity to any one of the amino acid sequences set forth in SEQ ID NOs.: 2 – 6.

In any of the methods, 5-hydroxypentanoyl-CoA can be enzymatically produced from 2-oxoadipate. For example, 5-hydroxypentanoyl-CoA can be enzymatically produced from 2-oxoadipate using one or more of a α -aminotransaminase; a 2-oxoadipate decarboxylase; a branch chain decarboxylase; a glutamate decarboxylase; a ω -transaminase; a CoA transferase, a CoA ligase, and an alcohol dehydrogenase.

In any of the methods, 5-hydroxypentanoyl-CoA can be enzymatically produced from malonyl-CoA. For example, 5-hydroxypentanoyl-CoA can be enzymatically produced from malonyl-CoA using one or more of a malonyl-CoA reductase; a 3-hydroxypropionate dehydrogenase; a 3-hydroxypropionyl-CoA synthase; a CoA-transferase; a β -ketothiolase; a 3-hydroxyacyl-CoA dehydrogenase; a 3-oxoacyl-CoA reductase, an enoyl-CoA hydratase, and a trans-2-enoyl-CoA reductase.

In any of the methods described herein, pimelic acid can be produced by forming the second terminal functional group in pimelate semialdehyde (also known as 7-oxoheptanoate) using (i) an aldehyde dehydrogenase classified under EC 1.2.1.3, (ii) a 7-oxohexanoate dehydrogenase classified under EC 1.2.1.63 such as that encoded by *ChnE* or a 7-oxoheptanoate dehydrogenase classified under EC 1.2.1.- (e.g., the gene product of *ThnG*), or iii) a monooxygenase in the cytochrome P450 family.

In any of the methods described herein, 7-aminoheptanoic acid can be produced by forming the second terminal functional group in pimelate semialdehyde using a ω -transaminase classified under EC 2.6.1.18, EC 2.6.1.19, EC 2.6.1.29, EC 2.6.1.48, or EC 2.6.1.82.

In any of the methods described herein, heptamethylenediamine can be produced by forming a second terminal functional group in (i) 7-aminoheptanal using a ω -transaminase classified under EC 2.6.1.18, EC 2.6.1.19, EC 2.6.1.29, EC 2.6.1.48 or EC 2.6.1.82, or in (ii) N7-acetyl-1,7-diaminoheptane using a deacylase classified, for example, under EC 3.5.1.17.

In any of the methods described herein, 1,7-heptanediol can be produced by forming the second terminal functional group in 7-hydroxyheptanal using an alcohol dehydrogenase classified under EC 1.1.1.- (e.g., EC 1.1.1.1, 1.1.1.2, 1.1.1.21, or 1.1.1.184) such as that encoded by *YMR318C*, *YqhD*, or *CAA81612.1*.

In some embodiments, the biological feedstock can be or can derive from monosaccharides, disaccharides, lignocellulose, hemicellulose, cellulose, lignin, levulinic acid and formic acid, triglycerides, glycerol, fatty acids, agricultural waste, condensed distillers' solubles, or municipal waste.

In some embodiments, the non-biological feedstock can be or can derive from natural gas, syngas, CO₂/H₂, methanol, ethanol, benzoate, non-volatile residue (NVR) or a caustic wash waste stream from cycloheptane oxidation processes, or terephthalic acid / isophthalic acid mixture waste streams.

In some embodiments, the host microorganism's tolerance to high concentrations of one or more C7 building blocks is improved through continuous cultivation in a selective environment.

In some embodiments, the host microorganism's biochemical network is attenuated or augmented to (1) ensure the intracellular availability of acetyl-CoA and 5-hydroxypentanoyl-CoA, (2) create an NADH or NADPH imbalance that may only be balanced via the formation of one or more C7 building blocks, (3) prevent degradation of central metabolites, central precursors leading to and including C7 building blocks, and/or (4) ensure efficient efflux from the cell.

In some embodiments, a cultivation strategy is used to achieve anaerobic, micro-aerobic, or aerobic cultivation conditions.

In some embodiments, the cultivation strategy includes limiting nutrients, such as limiting nitrogen, phosphate or oxygen.

In some embodiments, one or more C7 building blocks are produced by a single type of microorganism, e.g., a recombinant host containing one or more exogenous nucleic acids, using, for example, a fermentation strategy.

In some embodiments, one or more C7 building blocks are produced by a single type of microorganism having one or more exogenous nucleic acids which encode a polypeptide having a *β*-ketothiolase activity, (ii) a *thioesterase* activity or a *CoA transferase* activity, and one or more of (iii) a *3-hydroxyacyl-CoA dehydrogenase* activity or a *3-oxoacyl-CoA reductase* activity, (iv) an *enoyl-CoA hydratase* activity, and (v) a *trans-2-enoyl-CoA reductase* activity, using, for example, a fermentation strategy.

In another aspect, this document features a recombinant host that includes at least one exogenous nucleic acid encoding (i) a β -ketothiolase, (ii) a thioesterase or a CoA transferase, and one or more of (iii) a 3-hydroxyacyl-CoA dehydrogenase or a 3-oxoacyl-CoA reductase, (iv) an enoyl-CoA hydratase, and (v) a trans-2-enoyl-CoA reductase, the host producing 7-hydroxyheptanoate.

A host producing 7-hydroxyheptanoate further can include one or more of the following exogenous enzymes: a monooxygenase, an alcohol dehydrogenase, a 4-hydroxybutyrate dehydrogenase, a 5-hydroxyvalerate dehydrogenase, a 6-hydroxyhexanoate dehydrogenase, a 7-oxoheptanoate dehydrogenase, a 6-oxohexanoate dehydrogenase, a 5-oxovalerate dehydrogenase, or an aldehyde dehydrogenase, the host further producing pimelic acid.

A host producing 7-hydroxyheptanoate further can include one or more of the following exogenous enzymes: a monooxygenase, a transaminase, a 6-hydroxyhexanoate dehydrogenase, a 4-hydroxybutyrate dehydrogenase, a 5-hydroxypentanoate dehydrogenase, and an alcohol dehydrogenase, the host further producing 7-aminoheptanoate. -

A host producing 7-hydroxyheptanoate further can include one or more of the following exogenous enzymes: a carboxylate reductase, a ω -transaminase, a deacylase, a N-acetyl transferase, or an alcohol dehydrogenase, said host further producing heptamethylenediamine.

A host producing 7-hydroxyheptanoate further can include an exogenous carboxylate reductase and an exogenous alcohol dehydrogenase, the host further producing 1,7-heptanediol.

Any of the recombinant hosts described herein further can include one or more of the following exogenous enzymes: an α -aminotransaminase; a 2-oxoadipate decarboxylase; a branch-chain decarboxylase; a glutamate decarboxylase; a ω -transaminase; a CoA-ligase; a CoA-transferase; and an alcohol dehydrogenase.

Any of the recombinant hosts described herein further can include one or more of the following exogenous enzymes: a malonyl-CoA reductase; a 3-hydroxypropionate dehydrogenase; a 3-hydroxypropionyl-CoA synthase; a CoA-transferase; a β -

ketothiolase; a *3-hydroxyacyl-CoA dehydrogenase*; a *3-oxoacyl-CoA reductase*; an *enoyl-CoA hydratase*; and a *trans-2-enoyl-CoA reductase*.

Any of the recombinant hosts can be a prokaryote such as a prokaryote from a genus selected from the group consisting of *Escherichia*; *Clostridia*; *Corynebacteria*; *Cupriavidus*; *Pseudomonas*; *Delftia*; *Bacillus*; *Lactobacillus*; *Lactococcus*; and *Rhodococcus*. For example, the prokaryote can be selected from the group consisting of *Escherichia coli*, *Clostridium ljungdahlii*, *Clostridium autoethanogenum*, *Clostridium kluyveri*, *Corynebacterium glutamicum*, *Cupriavidus necator*, *Cupriavidus metallidurans*, *Pseudomonas fluorescens*, *Pseudomonas putida*, *Pseudomonas oleovorans*, *Delftia acidovorans*, *Bacillus subtilis*, *Lactobacillus delbrueckii*, *Lactococcus lactis*, and *Rhodococcus equi*. Such prokaryotes also can be sources of genes for constructing recombinant host cells described herein that are capable of producing C7 building blocks.

Any of the recombinant hosts can be a eukaryote such as a eukaryote from a genus selected from the group consisting of *Aspergillus*, *Saccharomyces*, *Pichia*, *Yarrowia*, *Issatchenkia*, *Debaryomyces*, *Arxula*, and *Kluyveromyces*. For example, the eukaryote can be selected from the group consisting of *Aspergillus niger*, *Saccharomyces cerevisiae*, *Pichia pastoris*, *Yarrowia lipolytica*, *Issatchenkia orientalis*, *Debaryomyces hansenii*, *Arxula adeninivorans*, and *Kluyveromyces lactis*. Such eukaryotes also can be sources of genes for constructing recombinant host cells described herein that are capable of producing C7 building blocks.

Any of the recombinant hosts described herein further can include attenuation of one or more of the following enzymes: a *polyhydroxyalkanoate synthase*, an *acetyl-CoA thioesterase*, a *phosphotransacetylase forming acetate*, an *acetate kinase*, a *lactate dehydrogenase*, a *menaquinol-fumarate oxidoreductase*, an *alcohol dehydrogenase forming ethanol*, a *triose phosphate isomerase*, a *pyruvate decarboxylase*, a *glucose-6-phosphate isomerase*, *NADH-consuming transhydrogenase*, an *NADH-specific glutamate dehydrogenase*, a *NADH/NADPH-utilizing glutamate dehydrogenase*, a *pimeloyl-CoA dehydrogenase*, an *acyl-CoA dehydrogenase* accepting C7 building blocks and central precursors as substrates, a *butyryl-CoA dehydrogenase*, or an *adipyl-CoA synthetase*.

Any of the recombinant hosts described herein further can overexpress one or more genes encoding: an *acetyl-CoA synthetase*; a *6-phosphogluconate dehydrogenase*; a *transketolase*; a *puridine nucleotide transhydrogenase*; a *glyceraldehyde-3P-dehydrogenase*; a *malic enzyme*; a *glucose-6-phosphate dehydrogenase*; a *glucose dehydrogenase*; a *fructose 1,6 diphosphatase*; a *L-alanine dehydrogenase*; a *L-glutamate dehydrogenase*; a *formate dehydrogenase*; a *L-glutamine synthetase*; a *diamine transporter*; a *dicarboxylate transporter*; and/or a *multidrug transporter*.

This document also features a biobased polymer comprising the bioderived pimelic acid, 7-aminoheptanoic acid, heptamethylenediamine, or 1,7-heptanediol.

This document also features a biobased resin comprising the bioderived pimelic acid, 7-aminoheptanoic acid, heptamethylenediamine, or 1,7-heptanediol, as well as a molded product obtained by molding a biobased resin.

In another aspect, this document features a process for producing a biobased polymer that includes chemically reacting the bioderived pimelic acid, 7-aminoheptanoic acid, heptamethylenediamine, or 1,7-heptanediol, with itself or another compound in a polymer producing reaction.

In another aspect, this document features a process for producing a biobased resin that includes chemically reacting the bioderived pimelic acid, 7-aminoheptanoic acid, heptamethylenediamine, or 1,7-heptanediol, with itself or another compound in a resin producing reaction.

Also, described herein is a biochemical network comprising a polypeptide having β -ketothiolase activity, wherein the polypeptide having β -ketothiolase activity enzymatically converts 5-hydroxypentanoyl-CoA to 3-oxo-7-hydroxyheptanoyl-CoA.

The biochemical network can further include a polypeptide having *3-hydroxyacyl-CoA dehydrogenase* or a polypeptide having *3-oxoacyl-CoA reductase* activity, a polypeptide having *enoyl-CoA hydratase* activity, a polypeptide having *thioesterase* activity, a polypeptide having *CoA transferase*, and a polypeptide having *trans-2-enoyl-CoA reductase* activity. In one aspect, the biochemical network is a non-naturally occurring biochemical network comprising at least one substrate of Fig. 1, at least one exogenous nucleic acid encoding a polypeptide having the activity of at least one enzyme

of Fig. 1 and at least one product of Fig. 1. In another aspect of the invention, the biochemical network is a non-naturally occurring biochemical network comprising a 3-hydroxypropionyl-CoA, an exogenous nucleic acid encoding a polypeptide having the activity of a β -ketothiolase classified under EC. 2.3.1 and a 3-hydroxypropionyl-CoA.

In one aspect, this document features a method for producing a bioderived five and seven carbon compounds. The method for producing a bioderived five and seven carbon compounds can include culturing or growing a recombinant host as described herein under conditions and for a sufficient period of time to produce the bioderived five and seven carbon compounds, wherein, optionally, the bioderived carbon compound is pimelic acid, 7-aminoheptanoic acid, heptamethylenediamine, or 1,7-heptanediol.

In one aspect, this document features composition comprising a bioderived five and seven carbon compounds as described herein and a compound other than the bioderived five and seven carbon compound, wherein the bioderived carbon compound is pimelic acid, 7-aminoheptanoic acid, heptamethylenediamine, and 1,7-heptanediol. For example, the bioderived four carbon compound is a cellular portion of a host cell or an organism.

In one aspect, this document features nucleic acid constructs or expression vectors comprising a polynucleotide encoding a polypeptide having β -ketothiolase activity, wherein the polynucleotide is operably linked to one or more heterologous control sequences that direct production of the polypeptide and wherein the polypeptide having β -ketothiolase activity is selected from the group consisting of: (a) a polypeptide having at least 70% sequence identity to the polypeptide of SEQ ID NOs: 1 or 13; a polynucleotide encoding a polypeptide having ω -transaminase activity, wherein the polynucleotide is operably linked to one or more heterologous control sequences that direct production of the polypeptide and wherein the polypeptide having ω -transaminase activity is selected from the group consisting of: (a) a polypeptide having at least 70% sequence identity to the polypeptide of SEQ ID NOs: 7-12. Further, this document features compositions comprising the nucleic acid construct or expression vector as described above.

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 to practice the 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.

The details of one or more embodiments of the invention are set forth in the accompanying drawings and the description below. Other features, objects, and advantages of the invention will be apparent from the application, including the written description and drawings and the claims. The word “comprising” in the claims may be replaced by “consisting essentially of” or with “consisting of,” according to standard practice in patent law.

DESCRIPTION OF DRAWINGS

FIG. 1 is a schematic of exemplary biochemical pathways leading to 7-hydroxyheptanoate using 2-oxo-adipate and malonyl-CoA as central metabolites.

FIG. 2 is a schematic of exemplary biochemical pathways leading to pimelic acid using 7-hydroxyheptanoate as a central precursor.

FIG. 3 is a schematic of an exemplary biochemical pathway leading to 7-aminoheptanoate using 7-hydroxyheptanoate as a central precursor.

FIG. 4 is a schematic of exemplary biochemical pathways leading to heptamethylenediamine using 7-aminoheptanoate, 7-hydroxyheptanoate, pimelate semialdehyde, or 1,7-heptanediol as a central precursor.

FIG. 5 is a schematic of an exemplary biochemical pathway leading to 1,7-heptanediol using 7-hydroxyheptanoate as a central precursor.

FIG. 6 contains the amino acid sequences of a *Cupriavidus necator* β -ketothiolase (see GenBank Accession No. AAC38322.1, SEQ ID NO: 1), a *Mycobacterium marinum* carboxylate reductase (see Genbank Accession No. ACC40567.1, SEQ ID NO: 2), a

Mycobacterium smegmatis carboxylate reductase (see Genbank Accession No. ABK71854.1, SEQ ID NO: 3), a *Segniliparus rugosus* carboxylate reductase (see Genbank Accession No. EFV11917.1, SEQ ID NO: 4), a *Mycobacterium massiliense* carboxylate reductase (see Genbank Accession No. EIV11143.1, SEQ ID NO: 5), a *Segniliparus rotundus* carboxylate reductase (see Genbank Accession No. ADG98140.1, SEQ ID NO: 6), a *Chromobacterium violaceum* ω -transaminase (see Genbank Accession No. AAQ59697.1, SEQ ID NO: 7), a *Pseudomonas aeruginosa* ω -transaminase (see Genbank Accession No. AAG08191.1, SEQ ID NO: 8), a *Pseudomonas syringae* ω -transaminase (see Genbank Accession No. AAY39893.1, SEQ ID NO: 9), a *Rhodobacter sphaeroides* ω -transaminase (see Genbank Accession No. ABA81135.1, SEQ ID NO: 10), an *Escherichia coli* ω -transaminase (see Genbank Accession No. AAA57874.1, SEQ ID NO: 11), a *Vibrio fluvialis* ω -transaminase (See Genbank Accession No. AEA39183.1, SEQ ID NO: 12), an *Escherichia coli* β -ketothiolase (see GenBank Accession No. AAC74479.1, SEQ ID NO: 13), a *Mycobacterium smegmatis* carboxylate reductase (see Genbank Accession No. ABK75684.1, SEQ ID NO: 14), a *Pseudomonas putida* thiolase (see Genbank Accession No. AAN70209.2, SEQ ID NO: 15), a *Sphingomonas wittichii* thiolase (see Genbank Accession No. ABQ69245.1, SEQ ID NO: 16), a *Pseudomonas reinekei* thiolase (see Genbank Accession No. ACZ63623.1, SEQ ID NO: 17), a *Pseudomonas putida* thiolase (see Genbank Accession No. AAA85138.1, SEQ ID NO: 18), a *Burkholderia xenovorans* thiolase (see Genbank Accession No. ABE28745.1, SEQ ID NO: 19), a *Burkholderia xenovorans* thiolase (see Genbank Accession No. ABE33819.1, SEQ ID NO: 20), a *Rhodococcus jostii* thiolase (see Genbank Accession No. ABG94668.1, SEQ ID NO: 21), a *Bdellovibrio bacteriovorus* thiolase (see Genbank Accession No. CAE79693.1, SEQ ID NO: 22), a *Cronobacter turicensis* thiolase (see Genbank Accession No. CBA32535.1, SEQ ID NO: 23), an *Arthrobacter* sp. thiolase (see Genbank Accession No. ABK03524.1, SEQ ID NO: 24), a *Caulobacter segnis* thiolase (see Genbank Accession No. ADG08907.1, SEQ ID NO: 25), a *Dinoroseobacter shibae* thiolase (see Genbank Accession No. ABV92581.1, SEQ ID NO: 26), a *Burkholderia xenovorans* thiolase (see Genbank Accession No. ABE36495.1, SEQ ID NO: 27), a *Geobacillus kaustophilus* thiolase (see Genbank

Accession No. BAD75605.1, SEQ ID NO: 28), a *Beijerinckia indica* thiolase (see Genbank Accession No. ACB95386.1, SEQ ID NO: 29), a *Citrobacter freundii* thiolase (see Genbank Accession No. EKS55037.1, SEQ ID NO: 30), a *Cupriavidus necator* thiolase (see Genbank Accession No. AEI75849.1, SEQ ID NO: 31), a *Gordonia bronchialis* thiolase (see Genbank Accession No. ACY20886.1, SEQ ID NO: 32), a *Burkholderia sp.* thiolase (see Genbank Accession No. ADG18081.1, SEQ ID NO: 33), a *Glutamicibacter arilaitensis* thiolase (see Genbank Accession No. CBT74677.1, SEQ ID NO: 34), an *Escherichia coli* thiolase (see Genbank Accession No. AAC74479.1, SEQ ID NO: 35), a *Cupriavidus necator* thiolase (see Genbank Accession No. AAC38322.1, SEQ ID NO: 36), and a *Clostridium viride* 5-hydroxyvalerate CoA transferase (see NCBI Reference Sequence: NZ_KK211198.1, SEQ ID NO: 37).

FIG. 7 is a bar graph summarizing the change in absorbance at 340 nm after 20 minutes, which is a measure of the consumption of NADPH and activity of six *carboxylate reductase* preparations in enzyme only controls (no substrate).

FIG. 8 is a bar graph of the change in absorbance at 340 nm after 20 minutes, which is a measure of the consumption of NADPH and the activity of two *carboxylate reductase* preparations for converting pimelate to pimelate semialdehyde relative to the empty vector control.

FIG. 9 is a bar graph of the change in absorbance at 340 nm after 20 minutes, which is a measure of the consumption of NADPH and the activity of six *carboxylate reductase* preparations for converting 7-hydroxyheptanoate to 7-hydroxyheptanal relative to the empty vector control.

FIG. 10 is a bar graph of the change in absorbance at 340 nm after 20 minutes, which is a measure of the consumption of NADPH and the activity of three *carboxylate reductase* preparations for converting N7-acetyl-7-aminoheptanoate to N7-acetyl-7-aminoheptanal relative to the empty vector control.

FIG. 11 is a bar graph of the change in absorbance at 340 nm after 20 minutes, which is a measure of the consumption of NADPH and activity of a *carboxylate reductase* preparation for converting pimelate semialdehyde to heptanedial relative to the empty vector control.

FIG. 12 is a bar graph summarizing the percent conversion of pyruvate to L-alanine (mol/mol) as a measure of the ω -transaminase activity of the enzyme only controls (no substrate).

FIG. 13 is a bar graph of the percent conversion after 4 hours of pyruvate to L-alanine (mol/mol) as a measure of the ω -transaminase activity of four ω -transaminase preparations for converting 7-aminoheptanoate to pimelate semialdehyde relative to the empty vector control.

FIG. 14 is a bar graph of the percent conversion after 4 hours of L-alanine to pyruvate (mol/mol) as a measure of the ω -transaminase activity of three ω -transaminase preparations for converting pimelate semialdehyde to 7-aminoheptanoate relative to the empty vector control.

FIG. 15 is a bar graph of the percent conversion after 4 hours of pyruvate to L-alanine (mol/mol) as a measure of the ω -transaminase activity of six ω -transaminase preparations for converting heptamethylenediamine to 7-aminoheptanal relative to the empty vector control.

FIG. 16 is a bar graph of the percent conversion after 4 hours of pyruvate to L-alanine (mol/mol) as a measure of the ω -transaminase activity of six ω -transaminase preparations for converting N7-acetyl-1,7-diaminoheptane to N7-acetyl-7-aminoheptanal relative to the empty vector control.

FIG. 17 is a bar graph of the percent conversion after 4 hours of pyruvate to L-alanine (mol/mol) as a measure of the ω -transaminase activity of three ω -transaminase preparations for converting 7-aminoheptanol to 7-oxoheptanol relative to the empty vector control.

FIG. 18 is a bar graph of the production of 3-oxo-7-hydroxyheptanoyl-CoA after 3 hours and 9 hours in a two-step enzymatic reaction catalyzed by a 5-hydroxyvalerate CoA transferase and an enzyme from the thiolase family, relative to the empty vector control.

FIG. 19 contains a nucleic acid sequence (SEQ ID NO: 38) encoding a *Pseudomonas putida* thiolase, a nucleic acid sequence (SEQ ID NO: 39) encoding a *Sphingomonas wittichii* thiolase, a nucleic acid sequence (SEQ ID NO: 40) encoding a

Pseudomonas reinekei thiolase, a nucleic acid sequence (SEQ ID NO: 41) encoding a
Pseudomonas putida thiolase, a nucleic acid sequence (SEQ ID NO: 42) encoding a
Burkholderia xenovorans thiolase, a nucleic acid sequence (SEQ ID NO: 43) encoding a
Burkholderia xenovorans thiolase, a nucleic acid sequence (SEQ ID NO: 44) encoding a
Rhodococcus jostii thiolase, a nucleic acid sequence (SEQ ID NO: 45) encoding a
Bdellovibrio bacteriovorus thiolase, a nucleic acid sequence (SEQ ID NO: 46) encoding
a *Cronobacter turicensis* thiolase, a nucleic acid sequence (SEQ ID NO: 47) encoding an
Arthrobacter sp. thiolase, a nucleic acid sequence (SEQ ID NO: 48) encoding a
Caulobacter segnis thiolase, a nucleic acid sequence (SEQ ID NO: 49) encoding a
Dinoroseobacter shibae thiolase, a nucleic acid sequence (SEQ ID NO: 50) encoding a
Burkholderia xenovorans thiolase, a nucleic acid sequence (SEQ ID NO: 51) encoding a
Geobacillus kaustophilus thiolase, a nucleic acid sequence (SEQ ID NO: 52) encoding a
Beijerinckia indica thiolase, a nucleic acid sequence (SEQ ID NO: 53) encoding a
Citrobacter freundii thiolase, a nucleic acid sequence (SEQ ID NO: 54) encoding a
Cupriavidus necator thiolase, a nucleic acid sequence (SEQ ID NO: 55) encoding a
Gordonia bronchialis thiolase, a nucleic acid sequence (SEQ ID NO: 56) encoding a
Burkholderia sp. thiolase, a nucleic acid sequence (SEQ ID NO: 57) encoding a
Glutamicibacter arilaitensis thiolase, a nucleic acid sequence (SEQ ID NO: 58) encoding
an *Escherichia coli* thiolase, a nucleic acid sequence (SEQ ID NO: 59) encoding a
Cupriavidus necator thiolase, and a nucleic acid sequence (SEQ ID NO: 60) encoding a
Clostridium viride 5-hydroxyvalerate CoA transferase.

DETAILED DESCRIPTION

In general, this document provides enzymes, non-natural pathways, cultivation strategies, feedstocks, host microorganisms and attenuations to the host's biochemical network, for producing 7-hydroxyheptanoate or one or more of pimelic acid, 7-aminoheptanoic acid, heptamethylenediamine, or 1,7-heptanediol, all of which are referred to as C7 building blocks herein. As used herein, the term "central precursor" is used to denote any metabolite in any metabolic pathway shown herein leading to the

synthesis of a C7 building block. The term “central metabolite” is used herein to denote a metabolite that is produced in all microorganisms to support growth.

Host microorganisms described herein can include endogenous pathways that can be manipulated such that 7-hydroxyheptanoate or one or more other C7 building blocks can be produced. In an endogenous pathway, the host microorganism naturally expresses all of the enzymes catalyzing the reactions within the pathway. In one aspect of the invention, a host microorganism containing an engineered pathway does not naturally express all of the enzymes catalyzing the reactions within the pathway but has been engineered such that all of the enzymes within the pathway are expressed in the host.

The term “exogenous” as used herein with reference to a nucleic acid (or a protein) and a host refers to a nucleic acid that does not occur in (and cannot be obtained from) a cell of that particular type as it is found in nature or a protein encoded by such a nucleic acid. Thus, a non-naturally-occurring nucleic acid is considered to be exogenous to a host once in the host. It is important to note that non-naturally-occurring nucleic acids can contain nucleic acid subsequences 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 a non-naturally-occurring nucleic acid, and thus is exogenous to a host cell once introduced into the host, 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. 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. 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 host microorganism. For example, an entire chromosome isolated from a cell of yeast x is an exogenous nucleic

acid with respect to a cell of yeast *y* once that chromosome is introduced into a cell of yeast *y*.

In contrast, the term “endogenous” as used herein with reference to a nucleic acid (e.g., a gene) (or a protein) and a host refers to a nucleic acid (or protein) that does occur in (and can be obtained from) that particular host as it is found in nature. Moreover, a cell “endogenously expressing” a nucleic acid (or protein) expresses that nucleic acid (or protein) as does a host of the same particular type as it is found in nature. Moreover, a host “endogenously producing” or that “endogenously produces” a nucleic acid, protein, or other compound produces that nucleic acid, protein, or compound as does a host of the same particular type as it is found in nature.

For example, depending on the host and the compounds produced by the host, one or more of the following enzymes may be expressed in the host in addition to a β -ketothiolase: a 3-hydroxyacyl-CoA dehydrogenase, a 3-oxoacyl-CoA reductase, an enoyl-CoA hydratase, a trans-2-enoyl-CoA reductase, a thioesterase, a CoA transferase, an aldehyde dehydrogenase, a monooxygenase, an alcohol dehydrogenase, a 7-oxoheptanoate dehydrogenase, a 7-oxoheptanoate dehydrogenase, a ω transaminase, a 6-hydroxyhexanoate dehydrogenase, a 4-hydroxybutyrate dehydrogenase, a 5-hydroxypentanoate dehydrogenase, a carboxylate reductase, a deacylase, an N-acetyl transferase, a ω -transaminase, or an amidohydrolase. In recombinant hosts expressing a carboxylate reductase, a phosphopantetheinyl transferase also can be expressed as it enhances activity of the carboxylate reductase. In recombinant hosts expressing a monooxygenase, an electron transfer chain protein such as an oxidoreductase or ferredoxin polypeptide also can be expressed.

For example, a recombinant host can include an exogenous β -ketothiolase and produce 3-oxo-7-hydroxyheptanoyl-CoA, which can be converted to 7-hydroxyheptanoate.

For example, a recombinant host can include an exogenous enzyme from the *thiolase* family and produce 3-oxo-7-hydroxyheptanoyl-CoA, which can be converted to 7-hydroxyheptanoate.

For example, a recombinant host can include an exogenous β -ketothiolase and an exogenous thioesterase or CoA-transferase, and one or more of the following exogenous enzymes: 3-hydroxyacyl-CoA dehydrogenase or a 3-oxoacyl-CoA reductase, an enoyl-CoA hydratase, and a trans-2-enoyl-CoA reductase, and produce 7-hydroxyheptanoate. For example, a recombinant host can include an exogenous β -ketothiolase, an exogenous thioesterase or CoA-transferase, an enoyl-CoA hydratase, an exogenous trans-2-enoyl-CoA reductase, and an exogenous 3-hydroxyacyl-CoA dehydrogenase or an exogenous 3-oxoacyl-CoA reductase, and produce 7-hydroxyheptanoate.

For example, a recombinant host producing 7-hydroxyheptanoate can include one or more of the following exogenous enzymes: a monooxygenase, an alcohol dehydrogenase, a 4-hydroxybutyrate dehydrogenase, a 5-hydroxypentanoate dehydrogenase, a 6-hydroxyhexanoate dehydrogenase, a 7-oxoheptanoate dehydrogenase, a 5-oxovalerate dehydrogenase, a 6-oxohexanoate dehydrogenase, or an aldehyde dehydrogenase, and further produce pimelic acid. For example, a recombinant host producing 7-hydroxyheptanoate can include an exogenous monooxygenase and produce pimelic acid. For example, a recombinant host producing 7-hydroxyheptanoate can include an exogenous 6-hydroxyhexanoate dehydrogenase and an aldehyde dehydrogenase and produce pimelic acid. For example, a recombinant host producing 7-hydroxyheptanoate can include an exogenous alcohol dehydrogenase and one of the following exogenous enzymes: a 5-oxovalerate dehydrogenase, a 6-oxohexanoate dehydrogenase, or a 7-oxoheptanoate dehydrogenase, and produce pimelic acid.

For example, a recombinant host producing 7-hydroxyheptanoate can include one or more of the following exogenous enzymes: an alcohol dehydrogenase, a 6-hydroxyhexanoate dehydrogenase, or a transaminase, and further produce 7-aminoheptanoate. For example, a recombinant host producing 7-hydroxyheptanoate can include an exogenous alcohol dehydrogenase and an exogenous transaminase and produce 7-aminoheptanoate. For example, a recombinant host producing 7-hydroxyheptanoate can include an exogenous 6-hydroxyhexanoate dehydrogenase and an exogenous transaminase and produce 7-aminoheptanoate.

For example, a recombinant host producing 7-hydroxyheptanoate can include one or more of the following exogenous enzymes: a *carboxylate reductase*, a ω -*transaminase*, a *deacylase*, an *N-acetyl transferase*, or an *alcohol dehydrogenase*, and produce heptamethylenediamine. For example, a recombinant host producing 7-hydroxyheptanoate can include an exogenous *carboxylate reductase*, an exogenous *alcohol dehydrogenase*, and one or more exogenous *transaminases* (e.g., one *transaminase* or two different *transaminases*), and produce heptamethylenediamine. For example, a recombinant host producing 7-hydroxyheptanoate can include an exogenous *carboxylate reductase* and one or more exogenous *transaminases* (e.g., one *transaminase* or two different *transaminases*) and produce heptamethylenediamine. For example, a recombinant host producing 7-hydroxyheptanoate can include an exogenous *alcohol dehydrogenase*, an exogenous *carboxylate reductase*, and one or more exogenous *transaminases* (e.g., one *transaminase*, or two or three different *transaminases*) and produce heptamethylenediamine. For example, a recombinant host producing 7-hydroxyheptanoate can include an exogenous *alcohol dehydrogenase*, an exogenous *N-acetyl transferase*, a *carboxylate reductase*, a *deacylase*, and one or more exogenous *transaminases* (e.g., one *transaminase* or two different *transaminases*) and produce heptamethylenediamine.

For example, a recombinant host producing 7-hydroxyheptanoate can include one or more of the following exogenous enzymes: a *carboxylate reductase* and an exogenous *alcohol dehydrogenase*, and further produce 1,7-heptanediol.

In any of the recombinant hosts, the recombinant host also can include one or more (e.g., one, two, three, or four) of the following exogenous enzymes used to convert 2-oxoadipate - to 5-hydroxypentanoyl-CoA: an *alpha-aminotransaminase*; a 2-*oxoadipate decarboxylase*; a *branch-chain decarboxylase*; a *glutamate decarboxylase*; a *CoA-ligase*; a *CoA-transferase*; a ω -*transaminase*; and an *alcohol dehydrogenase*. For example, a recombinant host can include an exogenous *alpha-aminotransaminase*; a *glutamate decarboxylase*; a *CoA-ligase* or a *CoA-transferase*; a ω -*transaminase*; and an *alcohol dehydrogenase*. For example, a recombinant host can include an exogenous 2-

oxoadipate decarboxylase or a *branch-chain decarboxylase*; a *CoA-ligase*; a *CoA-transferase*; and an *alcohol dehydrogenase*.

In any of the recombinant hosts, the recombinant host also can include one or more (e.g., one, two, three, or four) of the following exogenous enzymes used to convert manonyl-CoA to 5-hydroxypentanoyl-CoA: a *malonyl-CoA reductase*; a *3-hydroxypropionate dehydrogenase*; a *3-hydroxypropionyl-CoA synthase*; a *CoA-transferase*; a β -ketothiolase; a *3-hydroxyacyl-CoA dehydrogenase*; a *3-oxoacyl-CoA reductase*; an *enoyl-CoA hydratase*; and a *trans-2-enoyl-CoA reductase*. For example, a recombinant host can include an exogenous *malonyl-CoA reductase*; a *3-hydroxypropionate dehydrogenase*; a *3-hydroxypropionyl-CoA synthase*; a *CoA-transferase*; a β -ketothiolase; a *3-hydroxyacyl-CoA dehydrogenase*; a *3-oxoacyl-CoA reductase*; an *enoyl-CoA hydratase*; and a *trans-2-enoyl-CoA reductase*.

Within an engineered pathway, 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.

Any of the enzymes described herein that can be used for production of one or more C7 building blocks can have at least 70% sequence identity (homology) (e.g., at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 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) or on the basis of the immature enzyme (e.g., with any signal sequence included). It also will be appreciated that the initial methionine residue may or may not be present on any of the enzyme sequences described herein.

For example, a β -ketothiolase described herein can have at least 70% sequence identity (homology) (e.g., at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100%) to the amino acid sequence of a *Cupriavidus necator*

(see GenBank Accession No. AAC38322.1, SEQ ID NO: 1) or an *Escherichia coli* (see GenBank Accession No. AAC74479.1, SEQ ID NO: 13) β -ketothiolase. See FIG. 6.

For example, a *carboxylate reductase* described herein can have at least 70% sequence identity (homology) (e.g., at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100%) to the amino acid sequence of a *Mycobacterium marinum* (see Genbank Accession No. ACC40567.1, SEQ ID NO: 2), a *Mycobacterium smegmatis* (see Genbank Accession No. ABK71854.1, SEQ ID NO: 3), a *Segniliparus rugosus* (see Genbank Accession No. EFV11917.1, SEQ ID NO: 4), a *Mycobacterium massiliense* (see Genbank Accession No. EIV11143.1, SEQ ID NO: 5), or a *Segniliparus rotundus* (see Genbank Accession No. ADG98140.1, SEQ ID NO: 6) *carboxylate reductase*. See, FIG. 6.

For example, a ω -transaminase described herein can have at least 70% sequence identity (homology) (e.g., at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100%) to the amino acid sequence of a *Chromobacterium violaceum* (see Genbank Accession No. AAQ59697.1, SEQ ID NO: 7), a *Pseudomonas aeruginosa* (see Genbank Accession No. AAG08191.1, SEQ ID NO: 8), a *Pseudomonas syringae* (see Genbank Accession No. AAY39893.1, SEQ ID NO: 9), a *Rhodobacter sphaeroides* (see Genbank Accession No. ABA81135.1, SEQ ID NO: 10), an *Escherichia coli* (see Genbank Accession No. AAA57874.1, SEQ ID NO: 11), or a *Vibrio fluvialis* (see Genbank Accession No. AEA39183.1, SEQ ID NO: 12) ω -transaminase. Some of these ω -transaminases are *diamine ω -transaminases*. See, FIG. 6.

For example, a *thiolase* described herein can have at least 70% sequence identity (homology) (e.g., at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100%) to the amino acid sequence of a *Pseudomonas putida thiolase* (see Genbank Accession No. AAN70209.2, SEQ ID NO: 15), a *Sphingomonas wittichii thiolase* (see Genbank Accession No. ABQ69245.1, SEQ ID NO: 16), a *Pseudomonas reinekei thiolase* (see Genbank Accession No. ACZ63623.1, SEQ ID NO: 17), a *Pseudomonas putida thiolase* (see Genbank Accession No. AAA85138.1, SEQ ID NO: 18), a *Burkholderia xenovorans thiolase* (see Genbank Accession No. ABE28745.1, SEQ ID NO: 19), a *Burkholderia xenovorans thiolase* (see Genbank Accession No.

ABE33819.1, SEQ ID NO: 20), a *Rhodococcus jostii* thiolase (see Genbank Accession No. ABG94668.1, SEQ ID NO: 21), a *Bdellovibrio bacteriovorus* thiolase (see Genbank Accession No. CAE79693.1, SEQ ID NO: 22), a *Cronobacter turicensis* thiolase (see Genbank Accession No. CBA32535.1, SEQ ID NO: 23), an *Arthrobacter sp.* thiolase (see Genbank Accession No. ABK03524.1, SEQ ID NO: 24), a *Caulobacter segnis* thiolase (see Genbank Accession No. ADG08907.1, SEQ ID NO: 25), a *Dinoroseobacter shibae* thiolase (see Genbank Accession No. ABV92581.1, SEQ ID NO: 26), a *Burkholderia xenovorans* thiolase (see Genbank Accession No. ABE36495.1, SEQ ID NO: 27), a *Geobacillus kaustophilus* thiolase (see Genbank Accession No. BAD75605.1, SEQ ID NO: 28), a *Beijerinckia indica* thiolase (see Genbank Accession No. ACB95386.1, SEQ ID NO: 29), a *Citrobacter freundii* thiolase (see Genbank Accession No. EKS55037.1, SEQ ID NO: 30), a *Cupriavidus necator* thiolase (see Genbank Accession No. AEI75849.1, SEQ ID NO: 31), a *Gordonia bronchialis* thiolase (see Genbank Accession No. ACY20886.1, SEQ ID NO: 32), a *Burkholderia sp.* thiolase (see Genbank Accession No. ADG18081.1, SEQ ID NO: 33), a *Glutamicibacter arilaitensis* thiolase (see Genbank Accession No. CBT74677.1, SEQ ID NO: 34), an *Escherichia coli* thiolase (see Genbank Accession No. AAC74479.1, SEQ ID NO: 35), or a *Cupriavidus necator* thiolase (see Genbank Accession No. AAC38322.1, SEQ ID NO: 36). See FIG. 6.

For example, a 5-hydroxyvalerate CoA transferase described herein can have at least 70% sequence identity (homology) (e.g., at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100%) to the amino acid sequence of a *Clostridium viride* 5-hydroxyvalerate CoA transferase (see NCBI Reference Sequence: NZ_KK211198.1, SEQ ID NO: 37). See, FIG. 6.

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 (Bl2seq) program from the stand-alone version of BLASTZ containing BLASTP version 2.0.14. This stand-alone version of BLASTZ can be obtained from the U.S. government's National Center for Biotechnology Information web site (www.ncbi.nlm.nih.gov). Instructions explaining how to use the Bl2seq program can be found in the readme file accompanying BLASTZ. Bl2seq performs a comparison

between two amino acid sequences using the BLASTP algorithm. To compare two amino acid sequences, the options of Bl2seq are set as follows: -i is set to a file containing the first amino acid sequence to be compared (e.g., C:\seq1.txt); -j is set to a file containing the second amino acid sequence to be compared (e.g., C:\seq2.txt); -p is set to blastp; -o is set to any desired file name (e.g., C:\output.txt); and all other options are left at their default setting. For example, the following command can be used to generate an output file containing a comparison between two amino acid sequences: C:\Bl2seq -i c:\seq1.txt -j c:\seq2.txt -p blastp -o c:\output.txt. If the two compared sequences share homology (identity), then the designated output file will present those regions of homology as aligned sequences. If the two compared sequences do not share homology (identity), then the designated output file will not present aligned sequences. Similar procedures can be following for nucleic acid sequences except that blastn is used.

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) is obtained, using appropriate codon bias tables for that species.

Functional fragments of any of the enzymes described herein can also be used in the methods of the document. The term "functional fragment" as used herein refers to a peptide fragment of a protein that has at least 25% (e.g., at least: 30%; 40%; 50%; 60%; 70%; 75%; 80%; 85%; 90%; 95%; 98%; 99%; 100%; or even greater than 100%) of the

activity of the corresponding mature, full-length, wild-type protein. The functional fragment can generally, but not always, be comprised of a continuous region of the protein, wherein the region has functional activity.

This document also provides (i) functional variants of the enzymes used in the methods of the document and (ii) functional variants of the functional fragments described above. Functional variants of the enzymes and functional fragments can contain additions, deletions, or substitutions relative to the corresponding wild-type sequences. In one aspect of the invention, enzymes with substitutions will generally have not more than 50 (e.g., not more than one, two, three, four, five, six, seven, eight, nine, ten, 12, 15, 20, 25, 30, 35, 40, or 50) amino acid substitutions (e.g., conservative substitutions). This applies to any of the enzymes described herein and functional fragments. A conservative substitution is a substitution of one amino acid for another with similar characteristics. Conservative substitutions include substitutions within the following groups: valine, alanine, and glycine; leucine, valine, and isoleucine; aspartic acid and glutamic acid; asparagine and glutamine; serine, cysteine, and threonine; lysine and arginine; and phenylalanine and tyrosine. The nonpolar hydrophobic amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine. The polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine. The positively charged (basic) amino acids include arginine, lysine, and histidine. The negatively charged (acidic) amino acids include aspartic acid and glutamic acid. Any substitution of one member of the above-mentioned polar, basic, or acidic groups by another member of the same group can be deemed a conservative substitution. By contrast, a nonconservative substitution is a substitution of one amino acid for another with dissimilar characteristics.

Deletion variants can lack one, two, three, four, five, six, seven, eight, nine, ten, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 amino acid segments (of two or more amino acids) or non-contiguous single amino acids. Additions (addition variants) include fusion proteins containing: (a) any of the enzymes described herein or a fragment thereof; and (b) internal or terminal (C or N) irrelevant or heterologous amino acid sequences. In the context of such fusion proteins, the term "heterologous amino acid sequences" refers to

an amino acid sequence other than (a). A heterologous sequence can be, for example a sequence used for purification of the recombinant protein (e.g., FLAG, polyhistidine (e.g., heptahistidine), hemagglutinin (HA), glutathione-S-transferase (GST), or maltosebinding protein (MBP)). Heterologous sequences also can be proteins useful as detectable markers, for example, luciferase, green fluorescent protein (GFP), or chloramphenicol acetyl transferase (CAT). In some embodiments, the fusion protein contains a signal sequence from another protein. In certain host cells (e.g., yeast host cells), expression and/or secretion of the target protein can be increased through use of a heterologous signal sequence. In some embodiments, the fusion protein can contain a carrier (e.g., KLH) useful, e.g., in eliciting an immune response for antibody generation or ER or Golgi apparatus retention signals. Heterologous sequences can be of varying length and in some cases can be longer sequences than the full-length target proteins to which the heterologous sequences are attached.

Engineered hosts can naturally express none or some (e.g., one or more, two or more, three or more, four or more, five or more, or six or more) of the enzymes of the pathways described herein. Thus, a pathway within an engineered host can include all exogenous enzymes, or can include both endogenous and exogenous enzymes. Endogenous genes of the engineered hosts also can be disrupted to prevent the formation of undesirable metabolites or prevent the loss of intermediates in the pathway through other enzymes acting on such intermediates. Engineered hosts can be referred to as recombinant hosts or recombinant host cells. As described herein recombinant hosts can include nucleic acids encoding one or more of a β -ketothiolase, a dehydrogenase, a synthase, a decarboxylase, a reductase, a hydratase, a thioesterase, a monooxygenase, a thioesterase, amidohydrolase, and transaminase as described herein.

In addition, the production of C7 building blocks can be performed *in vitro* using the isolated enzymes described herein, using a lysate (e.g., a cell lysate) from a host microorganism as a source of the enzymes, or using a plurality of lysates from different host microorganisms as the sources of the enzymes.

The reactions of the pathways described herein can be performed in one or more host strains (a) naturally expressing one or more relevant enzymes, (b) genetically

engineered to express one or more relevant enzymes, or (c) naturally expressing one or more relevant enzymes and genetically engineered to express one or more relevant enzymes. Alternatively, relevant enzymes can be isolated from of the above types of host cells and used in a purified or semi-purified form. Moreover, such isolates or extracts include lysates (e.g. cell lysates) that can be used as sources of relevant enzymes. In the methods provided by the document, all the steps can be performed in host cells, all the steps can be performed using isolated or extracted enzymes, or some of the steps can be performed in cells and others can be performed using isolated or extracted enzymes.

Enzymes generating 7-hydroxyheptanoate

As depicted in FIG. 1, 7-hydroxyheptanoate can be biosynthesized from 2-oxoadipate or malonyl-CoA through the intermediate 3-oxo-7-hydroxyheptanoyl-CoA, which can be produced from 5-hydroxypentanoyl-CoA using a β -ketothiolase. 3-oxo-7-hydroxyheptanoyl-CoA can be converted to 7-hydroxyheptanoate using a 3-hydroxyacyl-CoA dehydrogenase or 3-oxoacyl-CoA dehydrogenase, an enoyl-CoA hydratase, a trans-2-enoyl-CoA reductase, and a thioesterase or a CoA transferase.

In some embodiments, a β -ketothiolase may be classified under EC 2.3.1.16, such as the gene product of *bktB*, or may be classified under EC 2.3.1.174 such as the gene product of *paaJ*. The β -ketothiolase encoded by *bktB* from *Cupriavidus necator* accepts acetyl-CoA and butanoyl-CoA as substrates, forming a CoA-activated C7 aliphatic backbone (see, e.g., Haywood *et al.*, *FEMS Microbiology Letters*, 1988, 52:91-96; Slater *et al.*, *J. Bacteriol.*, 1998, 180(8):1979 - 1987). The β -ketothiolase encoded by *paaJ* from *Escherichia coli* accepts succinyl-CoA and acetyl-CoA as substrates, forming a CoA-activated backbone (Nogales *et al.*, *Microbiology*, 2007, 153, 357-365). See, for example, SEQ ID NO:1 and SEQ ID NO:13 in FIG. 6.

In some embodiments, a 3-hydroxyacyl-CoA dehydrogenase or 3-oxoacyl-CoA dehydrogenase can be classified under EC 1.1.1.-. For example, the 3-hydroxyacyl-CoA dehydrogenase can be classified under EC 1.1.1.35, such as the gene product of *fadB*; classified under EC 1.1.1.157, such as the gene product of *hbd* (also referred to as a 3-hydroxybutyryl-CoA dehydrogenase); or classified under EC 1.1.1.36, such as the

acetoacetyl-CoA reductase gene product of *phaB* (Liu & Chen, *Appl. Microbiol. Biotechnol.*, 2007, 76(5):1153 – 1159; Shen *et al.*, *Appl. Environ. Microbiol.*, 2011, 77(9):2905 – 2915; Budde *et al.*, *J. Bacteriol.*, 2010, 192(20):5319 – 5328).

In some embodiments, a *3-oxoacyl-CoA reductase* can be classified under EC 1.1.1.100, such as the gene product of *fabG* (Budde *et al.*, *J. Bacteriol.*, 2010, 192(20):5319 – 5328; Nomura *et al.*, *Appl. Environ. Microbiol.*, 2005, 71(8):4297 – 4306).

In some embodiments, an *enoyl-CoA hydratase* can be classified under EC 4.2.1.17, such as the gene product of *crt*, or classified under EC 4.2.1.119, such as the gene product of *phaJ* (Shen *et al.*, 2011, *supra*; Fukui *et al.*, *J. Bacteriol.*, 1998, 180(3):667 – 673).

In some embodiments, a *trans-2-enoyl-CoA reductase* can be classified under EC 1.3.1.38 or EC 1.3.1.44, such as the gene product of *ter* (Nishimaki *et al.*, *J. Biochem.*, 1984, 95:1315 – 1321; Shen *et al.*, 2011, *supra*) or *tdter* (Bond-Watts *et al.*, *Biochemistry*, 2012, 51:6827 – 6837), or EC 1.3.1.8 (Inui *et al.*, *Eur. J. Biochem.*, 1984, 142, 121 – 126).

In some embodiments, the terminal carboxyl group leading to the synthesis of 7-hydroxyheptanoate is enzymatically formed in 7-hydroxyheptanoyl-CoA by a *thioesterase* classified under EC 3.1.2.-, resulting in the production of 7-hydroxyheptanoate. The *thioesterase* can be the gene product of *YciA* or *Acot13* (Cantu *et al.*, *Protein Science*, 2010, 19, 1281 – 1295; Zhuang *et al.*, *Biochemistry*, 2008, 47(9):2789 – 2796; Naggert *et al.*, *J. Biol. Chem.*, 1991, 266(17):11044 – 11050).

In some embodiments, the terminal carboxyl group leading to the synthesis of 7-hydroxyheptanoate is enzymatically formed in 7-hydroxyheptanoyl-CoA by a *CoA-transferase* classified under, for example, EC 2.8.3- such as the gene product of *cat2* from *Clostridium kluyveri*, *abfT* from *Clostridium aminobutyricum*, or the *4-hydroxybutyrate CoA-transferase* from *Clostridium viride*.

Enzymes generating the terminal carboxyl groups in the biosynthesis of pimelic acid

As depicted in FIG. 2, the terminal carboxyl group leading to the production of pimelic acid can be enzymatically formed using an *aldehyde dehydrogenase*, a *5-oxovalerate dehydrogenase*, a *6-oxohexanoate dehydrogenase*, a *7-oxoheptanoate dehydrogenase*, or a *monooxygenase*.

In some embodiments, the second terminal carboxyl group leading to the synthesis of pimelic acid can be enzymatically formed in pimelate semialdehyde by an *aldehyde dehydrogenase* classified under EC 1.2.1.3 (Guerrillot & Vandecasteele, *Eur. J. Biochem.*, 1977, 81, 185 – 192). See, FIG. 2.

In some embodiments, the second terminal carboxyl group leading to the synthesis of pimelic acid is enzymatically formed in pimelate semialdehyde by EC 1.2.1.- such as a *5-oxovalerate dehydrogenase* classified, for example, under EC 1.2.1.20, such as the gene product of *CpnE*; a *6-oxohexanoate dehydrogenase* classified, for example, EC 1.2.1.63 such as the gene product of *ChnE* from *Acinetobacter sp.*; or a *7-oxoheptanoate dehydrogenase* such as the gene product of *ThnG* from *Sphingomonas macrogolita* (Iwaki *et al.*, *Appl. Environ. Microbiol.*, 1999, 65(11), 5158 – 5162; López-Sánchez *et al.*, *Appl. Environ. Microbiol.*, 2010, 76(1), 110 - 118)). See, FIG. 2.

In some embodiments, the second terminal carboxyl group leading to the synthesis of pimelic acid is enzymatically formed in pimelate semialdehyde by a *monooxygenase* in the *cytochrome P450* family such as CYP4F3B (see, e.g., Sanders *et al.*, *J. Lipid Research*, 2005, 46(5):1001-1008; Sanders *et al.*, *The FASEB Journal*, 2008, 22(6):2064 – 2071). See, FIG. 2.

Enzymes generating the terminal amine groups in the biosynthesis of Heptamethylenediamine or 7-aminoheptanoate

As depicted in FIG. 3 and FIG. 4, terminal amine groups can be enzymatically formed using a *ω -transaminase* or a *deacylase*.

In some embodiments, a terminal amine group leading to the synthesis of 7-aminoheptanoic acid is enzymatically formed in pimelate semialdehyde by a *ω -transaminase* classified, for example, under EC 2.6.1.- (e.g., EC 2.6.1.18, EC 2.6.1.19,

EC 2.6.1.29, EC 2.6.1.48, or EC 2.6.1.82) such as that obtained from *Chromobacterium violaceum* (Genbank Accession No. AAQ59697.1, SEQ ID NO: 7), *Pseudomonas aeruginosa* (Genbank Accession No. AAG08191.1, SEQ ID NO: 8), *Pseudomonas syringae* (Genbank Accession No. AAY39893.1, SEQ ID NO: 9), *Rhodobacter sphaeroides* (Genbank Accession No. ABA81135.1, SEQ ID NO: 10), *Vibrio fluvialis* (Genbank Accession No. AEA39183.1, SEQ ID NO: 12), *Streptomyces griseus*, or *Clostridium viride*. See, FIG. 3.

An additional ω -transaminase that can be used in the methods and hosts described herein is from *Escherichia coli* (Genbank Accession No. AAA57874.1, SEQ ID NO: 11). Some of the ω -transaminases classified, for example, under EC 2.6.1.29 or EC 2.6.1.82 are diamine ω -transaminases (e.g., SEQ ID NO: 11).

The reversible ω -transaminase from *Chromobacterium violaceum* (Genbank Accession No. AAQ59697.1, SEQ ID NO: 7) has demonstrated analogous activity accepting 7-aminoheptanoic acid as an amino donor, thus forming the first terminal amine group in pimelate semialdehyde (Kaulmann *et al.*, *Enzyme and Microbial Technology*, 2007, 41, 628 – 637).

The reversible 4-aminobutyrate:2-oxoadipate transaminase from *Streptomyces griseus* has demonstrated activity for the conversion of 7-aminoheptanoate to pimelate semialdehyde (Yonaha *et al.*, *Eur. J. Biochem.*, 1985, 146, 101 - 106).

The reversible 5-aminovalerate transaminase from *Clostridium viride* has demonstrated activity for the conversion of 7-aminoheptanoate to pimelate semialdehyde (Barker *et al.*, *J. Biol. Chem.*, 1987, 262(19), 8994 – 9003).

In some embodiments, the second terminal amine group leading to the synthesis of heptamethylenediamine is enzymatically formed in 7-aminoheptanal by a diamine transaminase classified, for example, under EC 2.6.1.29; or classified, for example, under EC 2.6.1.82, such as the gene product of *YgjG* from *E. coli* (Genbank Accession No. AAA57874.1, SEQ ID NO: 12). The transaminases set forth in SEQ ID NOs:7-10 and 11 also can be used to produce heptamethylenediamine. See, FIG. 4.

The gene product of *ygjG* accepts a broad range of diamine carbon chain length substrates, such as putrescine, cadaverine, and spermidine (Samsonova *et al.*, *BMC Microbiology*, 2003, 3:2).

The *diamine transaminase* from *E.coli* strain B has demonstrated activity for 1,7-diaminoheptane (Kim, *The Journal of Chemistry*, 1964, 239(3), 783 – 786).

In some embodiments, the second terminal amine group leading to the synthesis of heptamethylenediamine is enzymatically formed in N7-acetyl-1,7-diaminoheptane by a *deacylase* classified, for example, under EC 3.5.1.17 such as an *acyl lysine deacylase*.

Enzymes generating the terminal hydroxyl groups in the biosynthesis of 1,7 heptanediol

As depicted in FIG. 5, the terminal hydroxyl group can be enzymatically formed using an *alcohol dehydrogenase*. For example, the second terminal hydroxyl group leading to the synthesis of 1,7 heptanediol can be enzymatically formed in 7-hydroxyheptanal by an *alcohol dehydrogenase* classified under EC 1.1.1.- (e.g., EC 1.1.1.1, 1.1.1.2, 1.1.1.21, or 1.1.1.184) such as the gene product of *YMR318C* or *YqhD* (Liu *et al.*, *Microbiology*, 2009, 155, 2078 – 2085; Larroy *et al.*, 2002, *Biochem J.*, 361(Pt 1), 163 – 172; Jarboe, 2011, *Appl. Microbiol. Biotechnol.*, 89(2), 249 - 257) or the protein having GenBank Accession No. CAA81612.1.

Biochemical pathways

Pathways to 7-hydroxyheptanoate

In some embodiments, 5-hydroxypentanoyl-CoA is synthesized from the central metabolite, 2-oxoadipate, by conversion of 2-oxoadipate to 2-aminoadipate by a α -*aminotransaminase* classified, for example, under EC 2.6.1.39; followed by conversion of 2-aminoadipate to 5-aminopentanoate by a *glutamate decarboxylase* classified, for example, under EC 4.1.1.15; followed by conversion of 5-aminopentanoate to 5-oxopentanoate - by a ω -*transaminase* classified, for example, under EC 2.6.1.18, EC 2.6.1.19, EC 2.6.1.48, or EC 2.6.1.96 such as the gene product of *gabT* from *Escherichia coli* (Bartsch *et al.*, *J. Bacteriol.*, 1990, 172(12), 7035); followed by conversion of 5-

oxopentanoate to 5-hydroxypentanoate by an *alcohol dehydrogenase* classified, for example, under EC 1.1.1.61 such as the gene product of *gbd* (e.g., from *Sorangium cellulosum*), *gabD* (Bartsch *et al.*, *J. Bacteriol.*, 1990, 172(12), 7035), or *YihU* (Saito *et al.*, *J. Biol. Chem.*, 2009, 284(24), 16442 – 16452), or a *5-hydroxyvalerate dehydrogenase* such as the gene product of *cpnD* (see, for example, Iwaki *et al.*, 2002, *Appl. Environ. Microbiol.*, 68(11):5671 – 5684); followed by conversion of 5-hydroxypentanoate to 5-hydroxypentanoyl-CoA using a *CoA-ligase* classified under, for example, EC 6.2.1- (e.g., EC 6.2.1.40), or a *CoA-transferase* classified under, for example, EC 2.8.3.- such as the gene product of *cat2* from *Clostridium kluyveri*, *abhT* from *Clostridium aminobutyricum*, or the *4-hydroxybutyrate CoA-transferase* from *Clostridium viride*. See FIG. 1.

In some embodiments, 2-oxoadipate is converted to 5-oxopentanoate -using a *2-oxoadipate decarboxylase* classified, for example, under EC 4.1.1.71 or a *branch-chain decarboxylase* classified, for example, under EC 4.1.1.72 such as the gene product of *kdcA* or *kivD*. 5-oxopentanoate - produced in this fashion can be converted to 5-hydroxypentanoyl-CoA as described above. See, FIG. 1.

In some embodiments, 5-hydroxypentanoyl-CoA is synthesized from the central metabolite, malonyl-CoA, by conversion of malonyl-CoA to malonate semialdehyde by a *malonyl-CoA reductase* classified, for example, under EC 1.2.1.75; followed by conversion of malonate semialdehyde to 3-hydroxypropanoate by a *3-hydroxypropionate dehydrogenase* classified, for example, under EC 1.1.1.59; followed by conversion of 3-hydroxypropanoate to 3-hydroxypropanoyl-CoA by a *3-hydroxypropionyl-CoA synthase* classified, for example, under EC 6.2.1.36, or a *CoA-transferase* classified, for example, under EC 2.8.3.1; followed by conversion of 3-hydroxypropanoyl-CoA to 3-oxo-5-hydroxypentanoyl-CoA using a β -ketothiolase classified, for example, under EC 2.3.1.16 or EC 2.3.1.174 such as the gene product of *bktB* or *paaJ* (e.g., SEQ ID NO: 1 or 13); followed by conversion to 3-hydroxy-5-hydroxypentanoyl-CoA using a *3-hydroxyacyl-CoA dehydrogenase* classified, for example, under EC 1.1.1.- such as EC 1.1.1.35 (e.g., the gene product of *fadB*), EC 1.1.1.36 (e.g., the gene product of *phaB*), or EC 1.1.1.157 (e.g., the gene product of *hbd*), or a *3-oxoacyl-CoA reductase* classified, for example,

under EC 1.1.1.100, such as the gene product of *fabG*; followed by conversion of 3-hydroxy-5-hydroxypentanoyl-CoA to 2,3-dehydro-5-hydroxypentanoyl-CoA using an *enoyl-CoA hydratase* classified, for example, under EC 4.2.1.17 such as the gene product of *crt* or classified under EC 4.2.1.119 such as the gene product of *phaJ*; followed by conversion of 2,3-dehydro-5-hydroxypentanoyl-CoA to 5-hydroxypentanoyl-CoA by a *trans-2-enoyl-CoA reductase* classified, for example, under EC 1.3.1.38, EC 1.3.1.44, or EC 1.3.1.8 such as the gene product of *ter* or *tdter*. See FIG. 1.

In some embodiments, 7-hydroxyheptanoate is synthesized from the central precursor, 5-hydroxypentanoyl-CoA, by conversion of 5-hydroxypentanoyl-CoA to 3-oxo-7-hydroxyheptanoyl-CoA using a β -ketothiolase classified, for example, under EC 2.3.1.16 or EC 2.3.1.174 such as the gene product of *bktB* or *paaJ* (e.g., SEQ ID NO: 1 or 13); followed by conversion to 3-hydroxy-7-hydroxyheptanoyl-CoA using a 3-hydroxyacyl-CoA dehydrogenase classified, for example, under EC 1.1.1.- such as EC 1.1.1.35 (e.g., the gene product of *fadB*), EC 1.1.1.36 (e.g., the gene product of *phaB*), or EC 1.1.1.157 (e.g., the gene product of *hbd*), or a 3-oxoacyl-CoA reductase classified, for example, under EC 1.1.1.100 such as the gene product of *fabG*; followed by conversion of 3-hydroxy-7-hydroxyheptanoyl-CoA to 2,3-dehydro-7-hydroxyheptanoyl-CoA using an *enoyl-CoA hydratase* classified, for example, under EC 4.2.1.17 such as the gene product of *crt*, or classified under EC 4.2.1.119 such as the gene product of *phaJ*; followed by conversion of 2,3-dehydro-7-hydroxyheptanoyl-CoA to 7-hydroxyheptanoyl-CoA by a *trans-2-enoyl-CoA reductase* classified, for example, under EC 1.3.1.38, EC 1.3.1.44, or EC 1.3.1.8 such as the gene product of *ter* or *tdter*; followed by conversion of 7-hydroxyheptanoyl-CoA to 7-hydroxyheptanoate by a *thioesterase* classified, for example, under EC 3.1.2.- such as the gene product of *YciA* or *Acot13*, or a *CoA-transferase* classified, for example, under EC 2.8.3.-. See FIG. 1.

Pathways using 7-hydroxyheptanoate as central precursor to pimelic acid

In some embodiments, pimelic acid is synthesized from 7-hydroxyheptanoate, by conversion of 7-hydroxyheptanoate to pimelate semialdehyde by an *alcohol dehydrogenase* classified under EC 1.1.1.- such as the gene product of *YMR318C*

(classified, for example, under EC 1.1.1.2, see Genbank Accession No. CAA90836.1) (Larroy *et al.*, 2002, *Biochem J.*, 361(Pt 1), 163 – 172), *cpnD* (Iwaki *et al.*, 2002, *Appl. Environ. Microbiol.*, 68(11):5671 – 5684), or *gabD* (Lütke-Eversloh & Steinbüchel, 1999, *FEMS Microbiology Letters*, 181(1):63 – 71), or a *6-hydroxyheptanoate dehydrogenase* classified, for example, under EC 1.1.1.258 such as the gene product of *ChnD* (Iwaki *et al.*, *Appl. Environ. Microbiol.*, 1999, 65(11):5158 - 5162); followed by conversion of pimelate semialdehyde to pimelic acid by a *dehydrogenase* classified, for example, under EC 1.2.1.- such as a *7-oxoheptanoate dehydrogenase* (e.g., the gene product of *ThnG*), a *7-oxoheptanoate dehydrogenase* (e.g., the gene product of *ChnE*), a *glutarate semialdehyde dehydrogenase* classified, for example, under EC 1.2.1.20, a *5-oxovalerate dehydrogenase* such as the gene product of *CpnE*, or an *aldehyde dehydrogenase* classified under EC 1.2.1.3. See FIG. 2. The *alcohol dehydrogenase* encoded by YMR318C has broad substrate specificity, including the oxidation of C7 alcohols.

In some embodiments, pimelic acid is synthesized from the central precursor, 7-hydroxyheptanoate, by conversion of 7-hydroxyheptanoate to pimelate semialdehyde by a *cytochrome P450* (Sanders *et al.*, *J. Lipid Research*, 2005, 46(5), 1001-1008; Sanders *et al.*, *The FASEB Journal*, 2008, 22(6), 2064 – 2071); followed by conversion of pimelate semialdehyde to pimelic acid by a *monooxygenase* in the *cytochrome P450* family such as CYP4F3B. See FIG. 2.

Pathway using 7-hydroxyheptanoate as central precursor to 7-aminoheptanoate -

In some embodiments, 7-aminoheptanoate is synthesized from the central precursor, 7-hydroxyheptanoate, by conversion of 7-hydroxyheptanoate to pimelate semialdehyde by an *alcohol dehydrogenase* classified, for example, under EC 1.1.1.2 such as the gene product of YMR318C, a *6-hydroxyhexanoate dehydrogenase* classified, for example, under EC 1.1.1.258, a *5-hydroxypentanoate dehydrogenase* classified, for example, under EC 1.1.1.- such as the gene product of *cpnD*, or a *4-hydroxybutyrate dehydrogenase* classified, for example, under EC 1.1.1.- such as the gene product of *gabD*; followed by conversion of pimelate semialdehyde to 7-aminoheptanoate by a ω -

transaminase (EC 2.6.1.18, EC 2.6.1.19, EC 2.6.1.29, EC 2.6.1.48, or EC 2.6.1.82 such as one of SEQ ID NOs:7-10 or 12, see above). See FIG. 3.

Pathway using 7-aminoheptanoate, 7-hydroxyheptanoate, pimelate semialdehyde, or 1,7-heptanediol as a central precursor to heptamethylenediamine

In some embodiments, heptamethylenediamine is synthesized from the central precursor, 7-aminoheptanoate, by conversion of 7-aminoheptanoate to 7-aminoheptanal by a *carboxylate reductase* classified, for example, under EC 1.2.99.6 such as the gene product of *car* in combination with a *phosphopantetheine transferase* enhancer (e.g., encoded by a *sfp* gene from *Bacillus subtilis* or *npt* gene from *Nocardia*), or the gene products of *GriC* and *GriD* from *Streptomyces griseus* (Suzuki *et al.*, *J. Antibiot.*, 2007, 60(6), 380 – 387); followed by conversion of 7-aminoheptanal to heptamethylenediamine by a ω -*transaminase* such as a ω -*transaminase* in EC 2.6.1.-, (e.g., EC 2.6.1.18, EC 2.6.1.19, EC 2.6.1.48, EC 2.6.1.82 such as SEQ ID NOs:7-12). The *carboxylate reductase* can be obtained, for example, from *Mycobacterium marinum* (Genbank Accession No. ACC40567.1, SEQ ID NO: 2), *Mycobacterium smegmatis* (Genbank Accession No. ABK71854.1, SEQ ID NO: 3), *Segniliparus rugosus* (Genbank Accession No. EFV11917.1, SEQ ID NO: 4), *Mycobacterium massiliense* (Genbank Accession No. EIV11143.1, SEQ ID NO: 5), or *Segniliparus rotundus* (Genbank Accession No. ADG98140.1, SEQ ID NO: 6). See FIG. 4.

The *carboxylate reductase* encoded by the gene product of *car* and enhancer *npt* or *sfp* has broad substrate specificity, including terminal difunctional C4 and C5 carboxylic acids (Venkitasubramanian *et al.*, *Enzyme and Microbial Technology*, 2008, 42, 130 – 137).

In some embodiments, heptamethylenediamine is synthesized from the central precursor, 7-hydroxyheptanoate (which can be produced as described in FIG. 1), by conversion of 7-hydroxyheptanoate to 7-hydroxyheptanal by a *carboxylate reductase* classified, for example, under EC 1.2.99.6 such as the gene product of *car* (see above) in combination with a *phosphopantetheine transferase* enhancer (e.g., encoded by a *sfp* gene from *Bacillus subtilis* or *npt* gene from *Nocardia*), or the gene product of *GriC* & *GriD*

(Suzuki *et al.*, 2007, *supra*); followed by conversion of 7-hydroxyheptanal to 7-aminoheptanol by a *ω*-transaminase classified, for example, under EC 2.6.1.18, EC 2.6.1.19, EC 2.6.1.29, EC 2.6.1.48, or EC 2.6.1.82 such as SEQ ID NOs:7-12, see above; followed by conversion to 7-aminoheptanal by an *alcohol dehydrogenase* classified, for example, under EC 1.1.1.- (e.g., EC 1.1.1.1, EC 1.1.1.2, EC 1.1.1.21, or EC 1.1.1.184) such as the gene product of *YMR318C* or *YqhD* (Liu *et al.*, *Microbiology*, 2009, 155, 2078 – 2085; Larroy *et al.*, 2002, *Biochem J.*, 361(Pt 1), 163 – 172; Jarboe, 2011, *Appl. Microbiol. Biotechnol.*, 89(2), 249 - 257) or the protein having GenBank Accession No. CAA81612.1; followed by conversion to heptamethylenediamine by a *ω*-transaminase classified, for example, under EC 2.6.1.18, EC 2.6.1.19, EC 2.6.1.29, EC 2.6.1.48, or EC 2.6.1.82 such as SEQ ID NOs:7-12, see above. See FIG. 4.

In some embodiments, heptamethylenediamine is synthesized from the central precursor, 7-aminoheptanoate, by conversion of 7-aminoheptanoate to N7-acetyl-7-aminoheptanoate by an *N*-acetyltransferase such as a *lysine N*-acetyltransferase classified, for example, under EC 2.3.1.32; followed by conversion to N7-acetyl-7-aminoheptanal by a *carboxylate reductase* classified, for example, under EC 1.2.99.6 such as the gene product of *car* (see above, e.g., SEQ ID NO: 4, 5, or 6) in combination with a *phosphopantetheine transferase* enhancer (e.g., encoded by a *sfp* gene from *Bacillus subtilis* or *npt* gene from *Nocardia*), or the gene product of *GriC* & *GriD*; followed by conversion to N7-acetyl-1,7-diaminoheptane by a *ω*-transaminase classified, for example, under EC 2.6.1.18, EC 2.6.1.19, EC 2.6.1.29, EC 2.6.1.48, or EC 2.6.1.82 such as SEQ ID NOs:7-12, see above; followed by conversion to heptamethylenediamine by an *acyl lysine deacylase* classified, for example, under EC 3.5.1.17. See, FIG. 4.

In some embodiments, heptamethylenediamine is synthesized from the central precursor, pimelate semialdehyde, by conversion of pimelate semialdehyde to heptanedial by a *carboxylate reductase* classified, for example, under EC 1.2.99.6 such as the gene product of *car* (see above, e.g., SEQ ID NO:6) in combination with a *phosphopantetheine transferase* enhancer (e.g., encoded by a *sfp* gene from *Bacillus subtilis* or *npt* gene from *Nocardia*) or the gene product of *GriC* & *GriD*; followed by conversion to 7-aminoheptanal by a *ω*-transaminase classified, for example, under EC

2.6.1.18, EC 2.6.1.19, EC 2.6.1.29, EC 2.6.1.48, or EC 2.6.1.82; followed by conversion to heptamethylenediamine by a *ω-transaminase* classified, for example, under EC 2.6.1.18, EC 2.6.1.19, EC 2.6.1.29, EC 2.6.1.48, or EC 2.6.1.82 such as SEQ ID NOs:7-12. See FIG. 4.

In some embodiments, heptamethylenediamine is synthesized from 1,7-heptanediol by conversion of 1,7-heptanedion to 7-hydroxyheptanal using an *alcohol dehydrogenase* classified, for example, under EC 1.1.1.- (e.g., EC 1.1.1.1, EC 1.1.1.2, EC 1.1.1.21, or EC 1.1.1.184) such as the gene product of *YMR318C* or *YqhD* or the protein having GenBank Accession No. CAA81612.1; followed by conversion to 7-aminoheptanol by a *ω-transaminase* classified, for example, under EC 2.6.1.18, EC 2.6.1.19, EC 2.6.1.29, EC 2.6.1.48, or EC 2.6.1.82 such as SEQ ID NOs:7-12, followed by conversion to 7-aminoheptanal by an *alcohol dehydrogenase* classified, for example, under EC 1.1.1.- (e.g., EC 1.1.1.1, EC 1.1.1.2, EC 1.1.1.21, or EC 1.1.1.184) such as the gene product of *YMR318C* or *YqhD* or the protein having GenBank Accession No. CAA81612.1, followed by conversion to heptamethylenediamine by a *ω-transaminase* classified, for example, under EC 2.6.1.18, EC 2.6.1.19, EC 2.6.1.29, EC 2.6.1.48, or EC 2.6.1.82 such as SEQ ID NOs:7-12. See FIG. 4.

Pathways using 7-hydroxyheptanoate as central precursor to 1,7-heptanediol

In some embodiments, 1,7 heptanediol is synthesized from the central precursor, 7-hydroxyheptanoate, by conversion of 7-hydroxyheptanoate to 7-hydroxyheptanal by a *carboxylate reductase* classified, for example, under EC 1.2.99.6 such as the gene product of *car* (see above, e.g., SEQ ID NO: 2, 3, 4, 5, or 6) in combination with a *phosphopantetheine transferase* enhancer (e.g., encoded by a *sfp* gene from *Bacillus subtilis* or *npt* gene from *Nocardia*) or the gene products of *GriC* and *GriD* from *Streptomyces griseus* (Suzuki *et al.*, *J. Antibiot.*, 2007, 60(6), 380 – 387); followed by conversion of 7-hydroxyheptanal to 1,7 heptanediol by an *alcohol dehydrogenase* (classified, for example, under EC 1.1.1.- such as EC 1.1.1.1, EC 1.1.1.2, EC 1.1.1.21, or EC 1.1.1.184) such as the gene product of *YMR318C* or *YqhD* (from *E. coli*, GenBank Accession No. AAA69178.1) (see, e.g., Liu *et al.*, *Microbiology*, 2009, 155, 2078 – 2085;

Larroy *et al.*, 2002, *Biochem J.*, 361(Pt 1), 163 – 172; or Jarboe, 2011, *Appl. Microbiol. Biotechnol.*, 89(2), 249 - 257), or the protein having GenBank Accession No. CAA81612.1 (from *Geobacillus stearothermophilus*). See, FIG. 5.

Cultivation strategy

In some embodiments, one or more C7 building blocks are biosynthesized in a recombinant host using anaerobic, aerobic, or micro-aerobic cultivation conditions. In some embodiments, the cultivation strategy entails nutrient limitation such as nitrogen, phosphate, or oxygen limitation.

In some embodiments, a cell retention strategy using, for example, ceramic hollow fiber membranes can be employed to achieve and maintain a high cell density during either fed-batch or continuous fermentation.

In some embodiments, the principal carbon source fed to the fermentation in the synthesis of one or more C7 building blocks can derive from biological or non-biological feedstocks.

In some embodiments, the biological feedstock can be or can derive from, monosaccharides, disaccharides, lignocellulose, hemicellulose, cellulose, lignin, levulinic acid and formic acid, triglycerides, glycerol, fatty acids, agricultural waste, condensed distillers' solubles, or municipal waste.

The efficient catabolism of crude glycerol stemming from the production of biodiesel has been demonstrated in several microorganisms such as *Escherichia coli*, *Cupriavidus necator*, *Pseudomonas oleovorans*, *Pseudomonas putida*, and *Yarrowia lipolytica* (Lee *et al.*, *Appl. Biochem. Biotechnol.*, 2012, 166:1801 – 1813; Yang *et al.*, *Biotechnology for Biofuels*, 2012, 5:13; Meijnen *et al.*, *Appl. Microbiol. Biotechnol.*, 2011, 90:885 - 893).

The efficient catabolism of lignocellulosic-derived levulinic acid has been demonstrated in several organisms such as *Cupriavidus necator* and *Pseudomonas putida* in the synthesis of 3-hydroxyvalerate via the precursor propanoyl-CoA (Jaremko and Yu, 2011, *supra*; Martin and Prather, *J. Biotechnol.*, 2009, 139:61 – 67).

The efficient catabolism of lignin-derived aromatic compounds such as benzoate analogues has been demonstrated in several microorganisms such as *Pseudomonas putida* and *Cupriavidus necator* (Bugg *et al.*, *Current Opinion in Biotechnology*, 2011, 22, 394 – 400; Pérez-Pantoja *et al.*, *FEMS Microbiol. Rev.*, 2008, 32, 736 – 794).

The efficient utilization of agricultural waste, such as olive mill waste water has been demonstrated in several microorganisms, including *Yarrowia lipolytica* (Papanikolaou *et al.*, *Bioresour. Technol.*, 2008, 99(7):2419 - 2428).

The efficient utilization of fermentable sugars such as monosaccharides and disaccharides derived from cellulosic, hemicellulosic, cane and beet molasses, cassava, corn, and other agricultural sources has been demonstrated for several microorganisms such as *Escherichia coli*, *Corynebacterium glutamicum*, *Lactobacillus delbrueckii*, and *Lactococcus lactis* (see, e.g., Hermann *et al.*, *J. Biotechnol.*, 2003, 104:155 – 172; Wee *et al.*, *Food Technol. Biotechnol.*, 2006, 44(2):163 – 172; Ohashi *et al.*, *J. Bioscience and Bioengineering*, 1999, 87(5):647 - 654).

The efficient utilization of furfural, derived from a variety of agricultural lignocellulosic sources, has been demonstrated for *Cupriavidus necator* (Li *et al.*, *Biodegradation*, 2011, 22:1215 – 1225).

In some embodiments, the non-biological feedstock can be or can derive from natural gas, syngas, CO₂/H₂, methanol, ethanol, benzoate, non-volatile residue (NVR) or a caustic wash waste stream from cycloheptane oxidation processes, or terephthalic acid / isophthalic acid mixture waste streams.

The efficient catabolism of methanol has been demonstrated for the methylotrophic yeast *Pichia pastoris*.

The efficient catabolism of ethanol has been demonstrated for *Clostridium kluyveri* (Seedorf *et al.*, *Proc. Natl. Acad. Sci. USA*, 2008, 105(6) 2128 - 2133).

The efficient catabolism of CO₂ and H₂, which may be derived from natural gas and other chemical and petrochemical sources, has been demonstrated for *Cupriavidus necator* (Prybylski *et al.*, *Energy, Sustainability and Society*, 2012, 2:11).

The efficient catabolism of syngas has been demonstrated for numerous microorganisms, such as *Clostridium ljungdahlii* and *Clostridium autoethanogenum* (Köpke *et al.*, *Applied and Environmental Microbiology*, 2011, 77(15):5467 – 5475).

The efficient catabolism of the non-volatile residue waste stream from cycloheptane processes has been demonstrated for numerous microorganisms, such as *Delftia acidovorans* and *Cupriavidus necator* (Ramsay *et al.*, *Applied and Environmental Microbiology*, 1986, 52(1):152 – 156).

In some embodiments, the host microorganism is a prokaryote. For example, the prokaryote can be a bacterium from the genus *Escherichia* such as *Escherichia coli*; from the genus *Clostridia* such as *Clostridium ljungdahlii*, *Clostridium autoethanogenum* or *Clostridium kluyveri*; from the genus *Corynebacteria* such as *Corynebacterium glutamicum*; from the genus *Cupriavidus* such as *Cupriavidus necator*, or *Cupriavidus metallidurans*; from the genus *Pseudomonas* such as *Pseudomonas fluorescens*, *Pseudomonas putida*, or *Pseudomonas oleovorans*; from the genus *Delftia* such as *Delftia acidovorans*; from the genus *Bacillus* such as *Bacillus subtilis*; from the genus *Lactobacillus* such as *Lactobacillus delbrueckii*; or from the genus *Lactococcus* such as *Lactococcus lactis*. Such prokaryotes also can be a source of genes to construct recombinant host cells described herein that are capable of producing one or more C7 building blocks.

In some embodiments, the host microorganism is a eukaryote. For example, the eukaryote can be a filamentous fungus, e.g., one from the genus *Aspergillus* such as *Aspergillus niger*. Alternatively, the eukaryote can be a yeast, e.g., one from the genus *Saccharomyces* such as *Saccharomyces cerevisiae*; from the genus *Pichia* such as *Pichia pastoris*; or from the genus *Yarrowia* such as *Yarrowia lipolytica*; from the genus *Issatchenkia* such as *Issatchenkia orientalis*; from the genus *Debaryomyces* such as *Debaryomyces hansenii*; from the genus *Arxula* such as *Arxula adenoinivorans*; or from the genus *Kluyveromyces* such as *Kluyveromyces lactis*. Such eukaryotes also can be a source of genes to construct recombinant host cells described herein that are capable of producing one or more C7 building blocks.

Metabolic engineering

The present document provides methods involving less than all the steps described for all the above pathways. Such methods can involve, for example, one, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve or more of such steps. Where less than all the steps are included in such a method, the first, and in some embodiments the only, step can be any one of the steps listed.

Furthermore, recombinant hosts described herein can include any combination of the above enzymes such that one or more of the steps, e.g., one, two, three, four, five, six, seven, eight, nine, ten, or more of such steps, can be performed within a recombinant host. This document provides host cells of any of the genera and species listed and genetically engineered to express one or more (e.g., two, three, four, five, six, seven, eight, nine, 10, 11, 12 or more) recombinant forms of any of the enzymes recited in the document. Thus, for example, the host cells can contain exogenous nucleic acids encoding enzymes catalyzing one or more of the steps of any of the pathways described herein.

In addition, this document recognizes that where enzymes have been described as accepting CoA-activated substrates, analogous enzyme activities associated with [acp]-bound substrates exist that are not necessarily in the same enzyme class.

Also, this document recognizes that where enzymes have been described accepting (R)-enantiomers of substrate, analogous enzyme activities associated with (S)-enantiomer substrates exist that are not necessarily in the same enzyme class.

This document also recognizes that where an enzyme is shown to accept a particular co-factor, such as NADPH, or co-substrate, such as acetyl-CoA, many enzymes are promiscuous in terms of accepting a number of different co-factors or co-substrates in catalyzing a particular enzyme activity. Also, this document recognizes that where enzymes have high specificity for e.g., a particular co-factor such as NADH, an enzyme with similar or identical activity that has high specificity for the co-factor NADPH may be in a different enzyme class.

In some embodiments, the enzymes in the pathways outlined herein are the result of enzyme engineering via non-direct or rational enzyme design approaches with aims of

improving activity, improving specificity, reducing feedback inhibition, reducing repression, improving enzyme solubility, changing stereo-specificity, or changing co-factor specificity.

In some embodiments, the enzymes in the pathways outlined here can be gene dosed, i.e., overexpressed, into the resulting genetically modified organism via episomal or chromosomal integration approaches.

In some embodiments, genome-scale system biology techniques such as Flux Balance Analysis can be utilized to devise genome scale attenuation or knockout strategies for directing carbon flux to a C7 building block.

Attenuation strategies include, but are not limited to the use of transposons, homologous recombination (double cross-over approach), mutagenesis, enzyme inhibitors, and RNAi interference.

In some embodiments, fluxomic, metabolomic, and transcriptomal data can be utilized to inform or support genome-scale system biology techniques, thereby devising genome scale attenuation or knockout strategies in directing carbon flux to a C7 building block.

In some embodiments, the host microorganism's tolerance to high concentrations of a C7 building block can be improved through continuous cultivation in a selective environment.

In some embodiments, the host microorganism's endogenous biochemical network can be attenuated or augmented to (1) ensure the intracellular availability of acetyl-CoA and 5-hydroxypentanoyl-CoA, (2) create an NADH or NADPH imbalance that may only be balanced via the formation of one or more C7 building blocks, (3) prevent degradation of central metabolites/central precursors leading to and including one or more C7 building blocks, and/or (4) ensure efficient efflux from the cell.

In some embodiments requiring intracellular availability of acetyl-CoA for C7 building block synthesis, endogenous enzymes catalyzing the hydrolysis of acetyl-CoA such as short-chain length *thioesterases* can be attenuated in the host organism.

In some embodiments requiring the intracellular availability of acetyl-CoA for C7 building block synthesis, an endogenous *phosphotransacetylase* generating acetate such as *pta* can be attenuated (Shen *et al.*, *Appl. Environ. Microbiol.*, 2011, 77(9):2905–2915).

In some embodiments requiring the intracellular availability of acetyl-CoA for C7 building block synthesis, an endogenous gene in an acetate synthesis pathway encoding an *acetate kinase*, such as *ack*, can be attenuated.

In some embodiments requiring the intracellular availability of acetyl-CoA and NADH for C7 building block synthesis, an endogenous gene encoding an enzyme that catalyzes the degradation of pyruvate to lactate such as *lactate dehydrogenase* encoded by *ldhA* can be attenuated (Shen *et al.*, 2011, *supra*).

In some embodiments, enzymes that catalyze anapleurotic reactions such as *PEP carboxylase* and/or *pyruvate carboxylase* can be overexpressed in the host organism.

In some embodiments requiring the intracellular availability of acetyl-CoA and NADH for C7 building block synthesis, endogenous genes encoding enzymes, such as *menaquinol-fumarate oxidoreductase*, that catalyze the degradation of phosphoenolpyruvate to succinate such as *frdBC* can be attenuated (see, e.g., Shen *et al.*, 2011, *supra*).

In some embodiments requiring the intracellular availability of acetyl-CoA and NADH for C7 building block synthesis, an endogenous gene encoding an enzyme that catalyzes the degradation of acetyl-CoA to ethanol such as the *alcohol dehydrogenase* encoded by *adhE* can be attenuated (Shen *et al.*, 2011, *supra*).

In some embodiments, where pathways require excess NADH co-factor for C7 building block synthesis, a recombinant *formate dehydrogenase* gene can be overexpressed in the host organism (Shen *et al.*, 2011, *supra*).

In some embodiments, where pathways require excess NADH co-factor for C7 building block synthesis, a recombinant NADH-consuming *transhydrogenase* can be attenuated.

In some embodiments, an endogenous gene encoding an enzyme that catalyzes the degradation of pyruvate to ethanol such as *pyruvate decarboxylase* can be attenuated.

In some embodiments requiring the intracellular availability of acetyl-CoA for C7 building block synthesis, a recombinant *acetyl-CoA synthetase* such as the gene product of *acs* can be overexpressed in the microorganism (Sato et al., *J. Bioscience and Bioengineering*, 2003, 95(4):335 – 341).

In some embodiments, carbon flux can be directed into the pentose phosphate cycle to increase the supply of NADPH by attenuating an endogenous *glucose-6-phosphate isomerase* (EC 5.3.1.9).

In some embodiments, carbon flux can be redirected into the pentose phosphate cycle to increase the supply of NADPH by overexpression of a *6-phosphogluconate dehydrogenase* and/or a *transketolase* (Lee et al., 2003, *Biotechnology Progress*, 19(5), 1444 – 1449).

In some embodiments, where pathways require excess NADPH co-factor in the synthesis of a C7 building block, a gene such as *UdhA* encoding a *puridine nucleotide transhydrogenase* can be overexpressed in the host organisms (Brigham et al., *Advanced Biofuels and Bioproducts*, 2012, Chapter 39, 1065 - 1090).

In some embodiments, where pathways require excess NADPH co-factor in the synthesis of a C7 Building Block, a recombinant *glyceraldehyde-3-phosphate-dehydrogenase* gene such as *GapN* can be overexpressed in the host organisms (Brigham et al., 2012, *supra*).

In some embodiments, where pathways require excess NADPH co-factor in the synthesis of a C7 building block, a recombinant *malic enzyme* gene such as *maeA* or *maeB* can be overexpressed in the host organisms (Brigham et al., 2012, *supra*).

In some embodiments, where pathways require excess NADPH co-factor in the synthesis of a C7 building block, a recombinant *glucose-6-phosphate dehydrogenase* gene such as *zwf* can be overexpressed in the host organisms (Lim et al., *J. Bioscience and Bioengineering*, 2002, 93(6), 543 - 549).

In some embodiments, where pathways require excess NADPH co-factor in the synthesis of a C7 building block, a recombinant *fructose 1,6 diphosphatase* gene such as *fbp* can be overexpressed in the host organisms (Becker et al., *J. Biotechnol.*, 2007, 132:99 - 109).

In some embodiments, where pathways require excess NADPH co-factor in the synthesis of a C7 building block, endogenous *triose phosphate isomerase* (EC 5.3.1.1) can be attenuated.

In some embodiments, where pathways require excess NADPH co-factor in the synthesis of a C7 building block, a recombinant *glucose dehydrogenase* such as the gene product of *gdh* can be overexpressed in the host organism (Sato *et al.*, *J. Bioscience and Bioengineering*, 2003, 95(4):335 – 341).

In some embodiments, endogenous enzymes facilitating the conversion of NADPH to NADH can be attenuated, such as the NADH generation cycle that may be generated via inter-conversion of *glutamate dehydrogenases* classified under EC 1.4.1.2 (NADH-specific) and EC 1.4.1.4 (NADPH-specific).

In some embodiments, an endogenous *glutamate dehydrogenase* (EC 1.4.1.3) that utilizes both NADH and NADPH as co-factors can be attenuated.

In some embodiments, a membrane-bound *cytochrome P450* such as *CYP4F3B* can be solubilized by only expressing the cytosolic domain and not the N-terminal region that anchors the P450 to the endoplasmic reticulum (Scheller *et al.*, *J. Biol. Chem.*, 1994, 269(17):12779-12783).

In some embodiments, an *enoyl-CoA reductase* can be solubilized via expression as a fusion protein with a small soluble protein, for example, the maltose binding protein (Gloerich *et al.*, *FEBS Letters*, 2006, 580, 2092 – 2096).

In some embodiments using hosts that naturally accumulate polyhydroxyalkanoates, the endogenous *polymer synthase* enzymes can be attenuated in the host strain.

In some embodiments, a *L-alanine dehydrogenase* can be overexpressed in the host to regenerate L-alanine from pyruvate as an amino donor for ω -transaminase reactions.

In some embodiments, a *L-glutamate dehydrogenase*, a *L-glutamine synthetase*, or a *alpha-aminotransaminase* can be overexpressed in the host to regenerate L-glutamate from 2-oxoadipate as an amino donor for ω -transaminase reactions.

In some embodiments, enzymes such as a *pimeloyl-CoA dehydrogenase* classified under, EC 1.3.1.62; an *acyl-CoA dehydrogenase* classified, for example, under EC 1.3.8.7, EC 1.3.8.1, or EC 1.3.99.-; and/or a *butyryl-CoA dehydrogenase* classified, for example, under EC 1.3.8.6 that degrade central metabolites and central precursors leading to and including C7 building blocks can be attenuated.

In some embodiments, endogenous enzymes activating C7 building blocks via Coenzyme A esterification such as *CoA-ligases* (e.g., an *adipyl-CoA synthetase*) classified under, for example, EC 6.2.1.- can be attenuated.

In some embodiments, the efflux of a C7 building block across the cell membrane to the extracellular media can be enhanced or amplified by genetically engineering structural modifications to the cell membrane or increasing any associated transporter activity for a C7 building block.

The efflux of heptamethylenediamine can be enhanced or amplified by overexpressing broad substrate range multidrug transporters such as *Blt* from *Bacillus subtilis* (Woolridge *et al.*, 1997, *J. Biol. Chem.*, 272(14):8864 – 8866), *AcrB* and *AcrD* from *Escherichia coli* (Elkins & Nikaido, 2002, *J. Bacteriol.*, 184(23), 6490 – 6499), *NorA* from *Staphylococcus aureus* (Ng *et al.*, 1994, *Antimicrob Agents Chemother*, 38(6), 1345 – 1355), or *Bmr* from *Bacillus subtilis* (Neyfakh, 1992, *Antimicrob Agents Chemother*, 36(2), 484 – 485).

The efflux of 7-aminoheptanoate and heptamethylenediamine can be enhanced or amplified by overexpressing the solute transporters such as the *lysE* transporter from *Corynebacterium glutamicum* (Bellmann *et al.*, 2001, *Microbiology*, 147, 1765 – 1774).

The efflux of pimelic acid can be enhanced or amplified by overexpressing a dicarboxylate transporter such as the *SucE* transporter from *Corynebacterium glutamicum* (Huhn *et al.*, *Appl. Microbiol. & Biotech.*, 89(2), 327 – 335).

Producing C7 Building Blocks Using a Recombinant Host

Typically, one or more C7 building blocks can be produced by providing a host microorganism and culturing the provided microorganism with a culture medium containing a suitable carbon source as described above. In general, the culture media

and/or culture conditions can be such that the microorganisms grow to an adequate density and produce a C7 building block efficiently. 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 tank (e.g., a 100 gallon, 200 gallon, 500 gallon, or more tank) 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 is reached, the broth containing the microorganisms can be transferred to a second tank. This second tank can be any size. For example, the second tank can be larger, smaller, or the same size as the first tank. Typically, the second tank is larger than the first such that additional culture medium can be added to the broth from the first tank. In addition, the culture medium within this second tank can be the same as, or different from, that used in the first tank.

Once transferred, the microorganisms can be incubated to allow for the production of a C7 building block. Once produced, any method can be used to isolate C7 building blocks. For example, C7 building blocks can be recovered selectively from the fermentation broth via adsorption processes. In the case of pimelic acid and 7-aminoheptanoic acid, the resulting eluate can be further concentrated via evaporation, crystallized via evaporative and/or cooling crystallization, and the crystals recovered via centrifugation. In the case of heptamethylenediamine and 1,7-heptanediol, distillation may be employed to achieve the desired product purity.

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

Enzyme activity of ω -transaminase using pimelate semialdehyde as substrate and forming 7-aminoheptanoate

A nucleotide sequence encoding an N-terminal His-tag was added to the nucleic acid sequences from *Chromobacterium violaceum*, *Pseudomonas syringae*, *Rhodobacter sphaeroides*, and *Vibrio fluvialis* encoding the ω -transaminases of SEQ ID NOs: 7, 9, 10, and 12, respectively (see FIG. 6) such that N-terminal HIS tagged ω -transaminases could be produced. Each of the resulting modified genes was cloned into a pET21a expression vector under control of the T7 promoter and each expression vector was transformed into a BL21[DE3] *E. coli* host. The resulting recombinant *E. coli* strains were cultivated at 37 °C in a 250 mL shake flask culture containing 50 mL LB media and antibiotic selection pressure, with shaking at 230 rpm. Each culture was induced overnight at 16 °C using 1 mM IPTG.

The pellet from each induced shake flask culture was harvested via centrifugation. Each pellet was resuspended and lysed via sonication. The cell debris was separated from the supernatant via centrifugation and the cell free extract was used immediately in enzyme activity assays.

Enzyme activity assays in the reverse direction (i.e., 7-aminoheptanoate to pimelate semialdehyde) were performed in a buffer composed of a final concentration of 50 mM HEPES buffer (pH = 7.5), 10 mM 7-aminoheptanoate, 10 mM pyruvate, and 100 μ M pyridoxyl 5' phosphate. Each enzyme activity assay reaction was initiated by adding cell free extract of the ω -transaminase gene product or the empty vector control to the assay buffer containing the 7-aminoheptanoate and incubated at 25 °C for 4 h, with shaking at 250 rpm. The formation of L-alanine from pyruvate was quantified via RP-HPLC.

Each enzyme only control without 7-aminoheptanoate demonstrated low base line conversion of pyruvate to L-alanine. See FIG. 12. The gene products of SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 10, and SEQ ID NO: 12 accepted 7-aminoheptanoate as substrate as confirmed against the empty vector control. See FIG. 13.

Enzyme activity in the forward direction (i.e., pimelate semialdehyde to 7-aminoheptanoate) was confirmed for the *transaminases* of SEQ ID NO: 9, SEQ ID NO: 10, and SEQ ID NO: 12. Enzyme activity assays were performed in a buffer composed of a final concentration of 50 mM HEPES buffer (pH = 7.5), 10 mM pimelate

semialdehyde, 10 mM L-alanine, and 100 μ M pyridoxyl 5' phosphate. Each enzyme activity assay reaction was initiated by adding a cell free extract of the ω -transaminase gene product or the empty vector control to the assay buffer containing the pimelate semialdehyde and incubated at 25 °C for 4 h, with shaking at 250 rpm. The formation of pyruvate was quantified via RP-HPLC.

The gene products of SEQ ID NO: 9, SEQ ID NO: 10, and SEQ ID NO: 12 accepted pimelate semialdehyde as substrate as confirmed against the empty vector control. See FIG. 14. The reversibility of the ω -transaminase activity was confirmed, demonstrating that the ω -transaminases of SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 10, and SEQ ID NO: 12 accepted pimelate semialdehyde as substrate and synthesized 7-aminoheptanoate as a reaction product.

EXAMPLE 2

Enzyme activity of *carboxylate reductase* using pimelate as substrate and forming pimelate semialdehyde

A nucleotide sequence encoding a HIS-tag was added to the nucleic acid sequences from *Segniliparus rugosus* and *Segniliparus rotundus* that encode the *carboxylate reductases* of SEQ ID NOs: 4 (EFV11917.1) and 6 (ADG98140.1), respectively (see FIG. 6), such that N-terminal HIS tagged *carboxylate reductases* could be produced. Each of the modified genes was cloned into a pET Duet expression vector along with a *sfp* gene encoding a HIS-tagged *phosphopantetheine transferase* from *Bacillus subtilis*, both under the T7 promoter. Each expression vector was transformed into a BL21[DE3] *E. coli* host and the resulting recombinant *E. coli* strains were cultivated at 37 °C in a 250 mL shake flask culture containing 50 mL LB media and antibiotic selection pressure, with shaking at 230 rpm. Each culture was induced overnight at 37 °C using an auto-induction media.

The pellet from each induced shake flask culture was harvested via centrifugation. Each pellet was resuspended and lysed via sonication, and the cell debris was separated from the supernatant via centrifugation. The *carboxylate reductases* and *phosphopantetheine transferases* were purified from the supernatant using Ni-affinity

chromatography, diluted 10-fold into 50 mM HEPES buffer (pH = 7.5), and concentrated via ultrafiltration.

Enzyme activity assays (i.e., from pimelate to pimelate semialdehyde) were performed in triplicate in a buffer composed of a final concentration of 50 mM HEPES buffer (pH = 7.5), 2 mM pimelate, 10 mM MgCl₂, 1 mM ATP, and 1 mM NADPH. Each enzyme activity assay reaction was initiated by adding purified *carboxylate reductase* and *phosphopantetheine transferase* gene products or the empty vector control to the assay buffer containing the pimelate and then incubated at room temperature for 20 min. The consumption of NADPH was monitored by absorbance at 340 nm. Each enzyme only control without pimelate demonstrated low base line consumption of NADPH. See bars for EFV11917.1 and ADG98140.1 in FIG. 7.

The gene products of SEQ ID NO: 4 (EFV11917.1) and SEQ ID NO: 6 (ADG98140.1), enhanced by the gene product of *sfp*, accepted pimelate as substrate, as confirmed against the empty vector control (see FIG. 8), and synthesized pimelate semialdehyde.

EXAMPLE 3

Enzyme activity of *carboxylate reductase* using 7-hydroxyheptanoate as substrate and forming 7-hydroxyheptanal

A nucleotide sequence encoding a His-tag was added to the nucleic acids from *Mycobacterium marinum*, *Mycobacterium smegmatis*, *Segniliparus rugosus*, *Mycobacterium abscessus subsp. bolletii*, *Segniliparus rotundus*, and *Mycobacterium smegmatis* that encode the *carboxylate reductases* of SEQ ID NOs: 2-6 and 14, respectively (GenBank Accession Nos. ACC40567.1, ABK71854.1, EFV11917.1, EIV11143.1, ADG98140.1, and ABK75684.1, respectively) (see FIG. 6) such that N-terminal HIS tagged *carboxylate reductases* could be produced. Each of the modified genes was cloned into a pET Duet expression vector alongside a *sfp* gene encoding a His-tagged *phosphopantetheine transferase* from *Bacillus subtilis*, both under control of the T7 promoter. Each expression vector was transformed into a BL21[DE3] *E. coli* host along with the expression vectors from Example 3. Each resulting recombinant *E. coli*

strain was cultivated at 37 °C in a 250 mL shake flask culture containing 50 mL LB media and antibiotic selection pressure, with shaking at 230 rpm. Each culture was induced overnight at 37 °C using an auto-induction media.

The pellet from each induced shake flask culture was harvested via centrifugation. Each pellet was resuspended and lysed via sonication. The cell debris was separated from the supernatant via centrifugation. The *carboxylate reductases* and *phosphopantetheine transferase* were purified from the supernatant using Ni-affinity chromatography, diluted 10-fold into 50 mM HEPES buffer (pH = 7.5), and concentrated via ultrafiltration.

Enzyme activity (i.e., 7-hydroxyheptanoate to 7-hydroxyheptanal) assays were performed in triplicate in a buffer composed of a final concentration of 50 mM HEPES buffer (pH = 7.5), 2 mM 7-hydroxyheptanal, 10 mM MgCl₂, 1 mM ATP, and 1 mM NADPH. Each enzyme activity assay reaction was initiated by adding purified *carboxylate reductase* and *phosphopantetheine transferase* or the empty vector control to the assay buffer containing the 7-hydroxyheptanoate and then incubated at room temperature for 20 min. The consumption of NADPH was monitored by absorbance at 340 nm. Each enzyme only control without 7-hydroxyheptanoate demonstrated low base line consumption of NADPH. See FIG. 7.

The gene products of SEQ ID NOs: 2 -- 6 and 14, enhanced by the gene product of *sfp*, accepted 7-hydroxyheptanoate as substrate as confirmed against the empty vector control (see FIG. 9), and synthesized 7-hydroxyheptanal.

EXAMPLE 4

Enzyme activity of ω -transaminase for 7-aminoheptanol, forming 7-oxoheptanol

A nucleotide sequence encoding an N-terminal His-tag was added to the *Chromobacterium violaceum*, *Pseudomonas syringae*, and *Rhodobacter sphaeroides* nucleic acids encoding the ω -transaminases of SEQ ID NOs: 7, 9 and 10, respectively (see FIG. 6), such that N-terminal HIS tagged ω -transaminases could be produced. The modified genes were cloned into a pET21a expression vector under the T7 promoter. Each expression vector was transformed into a BL21[DE3] *E. coli* host. Each resulting recombinant *E. coli* strain was cultivated at 37 °C in a 250 mL shake flask culture

containing 50 mL LB media and antibiotic selection pressure, with shaking at 230 rpm. Each culture was induced overnight at 16 °C using 1 mM IPTG.

The pellet from each induced shake flask culture was harvested via centrifugation. Each pellet was resuspended and lysed via sonication. The cell debris was separated from the supernatant via centrifugation and the cell free extract was used immediately in enzyme activity assays.

Enzyme activity assays in the reverse direction (i.e., 7-aminoheptanol to 7-oxoheptanol) were performed in a buffer composed of a final concentration of 50 mM HEPES buffer (pH = 7.5), 10 mM 7-aminoheptanol, 10 mM pyruvate, and 100 μ M pyridoxyl 5' phosphate. Each enzyme activity assay reaction was initiated by adding cell free extract of the ω -transaminase gene product or the empty vector control to the assay buffer containing the 7-aminoheptanol and then incubated at 25 °C for 4 h, with shaking at 250 rpm. The formation of L-alanine was quantified via RP-HPLC.

Each enzyme only control without 7-aminoheptanol had low base line conversion of pyruvate to L-alanine. See FIG. 12.

The gene products of SEQ ID NOs: 7, 9 & 10 accepted 7-aminoheptanol as substrate as confirmed against the empty vector control (see FIG. 17) and synthesized 7-oxoheptanol as reaction product. Given the reversibility of the ω -transaminase activity (see Example 1), it can be concluded that the gene products of SEQ ID NOs: 7, 9 & 10 accept 7-oxoheptanol as substrate and form 7-aminoheptanol.

EXAMPLE 5

Enzyme activity of ω -transaminase using heptamethylenediamine as substrate and forming 7-aminoheptanal

A nucleotide sequence encoding an N-terminal His-tag was added to the *Chromobacterium violaceum*, *Pseudomonas aeruginosa*, *Pseudomonas syringae*, *Rhodobacter sphaeroides*, *Escherichia coli*, and *Vibrio fluvialis* nucleic acids encoding the ω -transaminases of SEQ ID NOs: 7 - 12, respectively (see FIG. 6) such that N-terminal HIS tagged ω -transaminases could be produced. The modified genes were cloned into a pET21a expression vector under the T7 promoter. Each expression vector

was transformed into a BL21[DE3] *E. coli* host. Each resulting recombinant *E. coli* strain was cultivated at 37 °C in a 250 mL shake flask culture containing 50 mL LB media and antibiotic selection pressure, with shaking at 230 rpm. Each culture was induced overnight at 16 °C using 1 mM IPTG.

The pellet from each induced shake flask culture was harvested via centrifugation. Each pellet was resuspended and lysed via sonication. The cell debris was separated from the supernatant via centrifugation and the cell free extract was used immediately in enzyme activity assays.

Enzyme activity assays in the reverse direction (i.e., heptamethylenediamine to 7-aminoheptanal) were performed in a buffer composed of a final concentration of 50 mM HEPES buffer (pH = 7.5), 10 mM heptamethylenediamine, 10 mM pyruvate, and 100 μ M pyridoxyl 5' phosphate. Each enzyme activity assay reaction was initiated by adding cell free extract of the ω -transaminase gene product or the empty vector control to the assay buffer containing the heptamethylenediamine and then incubated at 25 °C for 4 h, with shaking at 250 rpm. The formation of L-alanine was quantified via RP-HPLC.

Each enzyme only control without heptamethylenediamine had low base line conversion of pyruvate to L-alanine. See FIG. 12.

The gene products of SEQ ID NOs: 7 - 12 accepted heptamethylenediamine as substrate as confirmed against the empty vector control (see FIG. 15) and synthesized 7-aminoheptanal as reaction product. Given the reversibility of the ω -transaminase activity (see Example 1), it can be concluded that the gene products of SEQ ID NOs: 7 – 12 accept 7-aminoheptanal as substrate and form heptamethylenediamine.

EXAMPLE 6

Enzyme activity of *carboxylate reductase* for N7-acetyl-7-aminoheptanoate, forming N7-acetyl-7-aminoheptanal

The activity of each of the N-terminal His-tagged *carboxylate reductases* of SEQ ID NOs: 3, 5, and 6 (see Examples 2 and 3, and FIG. 6) for converting N7-acetyl-7-aminoheptanoate to N7-acetyl-7-aminoheptanal was assayed in triplicate in a buffer composed of a final concentration of 50 mM HEPES buffer (pH = 7.5), 2 mM N7-acetyl-

7-aminoheptanoate, 10 mM MgCl₂, 1 mM ATP, and 1 mM NADPH. The assays were initiated by adding purified *carboxylate reductase* and *phosphopantetheine transferase* or the empty vector control to the assay buffer containing the N7-acetyl-7-aminoheptanoate then incubated at room temperature for 20 min. The consumption of NADPH was monitored by absorbance at 340 nm. Each enzyme only control without N7-acetyl-7-aminoheptanoate demonstrated low base line consumption of NADPH. See FIG. 7.

The gene products of SEQ ID NOs: 3, 5, and 6, enhanced by the gene product of *sfp*, accepted N7-acetyl-7-aminoheptanoate as substrate as confirmed against the empty vector control (see FIG. 10), and synthesized N7-acetyl-7-aminoheptanal.

EXAMPLE 7

Enzyme activity of ω -transaminase using N7-acetyl-1,7-diaminoheptane, and forming N7-acetyl-7-aminoheptanal

The activity of the N-terminal His-tagged ω -transaminases of SEQ ID NOs: 7 – 12 (see Example 5, and FIG. 6) for converting N7-acetyl-1,7-diaminoheptane to N7-acetyl-7-aminoheptanal was assayed using a buffer composed of a final concentration of 50 mM HEPES buffer (pH = 7.5), 10 mM N7-acetyl-1,7-diaminoheptane, 10 mM pyruvate, and 100 μ M pyridoxyl 5' phosphate. Each enzyme activity assay reaction was initiated by adding a cell free extract of the ω -transaminase or the empty vector control to the assay buffer containing the N7-acetyl-1,7-diaminoheptane then incubated at 25 °C for 4 h, with shaking at 250 rpm. The formation of L-alanine was quantified via RP-HPLC.

Each enzyme only control without N7-acetyl-1,7-diaminoheptane demonstrated low base line conversion of pyruvate to L-alanine. See FIG. 12.

The gene products of SEQ ID NOs: 7 - 12 accepted N7-acetyl-1,7-diaminoheptane as substrate as confirmed against the empty vector control (see FIG. 16) and synthesized N7-acetyl-7-aminoheptanal as reaction product.

Given the reversibility of the ω -transaminase activity (see Example 1), the gene products of SEQ ID NOs: 7 – 12 accept N7-acetyl-7-aminoheptanal as substrate forming N7-acetyl-1,7-diaminoheptane.

EXAMPLE 8

Enzyme activity of *carboxylate reductase* using pimelate semialdehyde as substrate and forming heptanedial

The N-terminal His-tagged *carboxylate reductase* of SEQ ID NO: 6 (see Example 3 and FIG. 6) was assayed using pimelate semialdehyde as substrate. The enzyme activity assay was performed in triplicate in a buffer composed of a final concentration of 50 mM HEPES buffer (pH = 7.5), 2 mM pimelate semialdehyde, 10 mM MgCl₂, 1 mM ATP, and 1 mM NADPH. The enzyme activity assay reaction was initiated by adding purified *carboxylate reductase* and *phosphopantetheine transferase* or the empty vector control to the assay buffer containing the pimelate semialdehyde and then incubated at room temperature for 20 min. The consumption of NADPH was monitored by absorbance at 340 nm. The enzyme only control without pimelate semialdehyde demonstrated low baseline consumption of NADPH. See FIG. 7.

The gene product of SEQ ID NO: 6, enhanced by the gene product of *sfp*, accepted pimelate semialdehyde as substrate as confirmed against the empty vector control (see FIG. 11) and synthesized heptanedial.

EXAMPLE 9

Biotransformation of 5-hydroxypentanoate to 3-oxo-7-hydroxyheptanoyl-CoA in a two-step enzymatic reaction catalyzed by a 5-hydroxyvalerate CoA transferase and an enzyme from the thiolase family

An enzyme with a *5-hydroxyvalerate CoA transferase* activity was identified following the sequencing of *Clostridium viride* genome and subsequent analysis for putative CoA transferases. Five enzyme sequences were identified and two of these transferases acted upon 5-hydroxyvalerate. The CoA transferase selected for the reaction described in this Example is further called BDIGENE# 246. The sequence is currently available publicly as NCBI Reference Sequence WP_027096059.1 (SEQ ID NO: 37). The gene (SEQ ID NO: 60) encoding the *5-hydroxyvalerate CoA transferase* was codon-optimized for expression in *E. coli*.

Selected genes (SEQ ID NOs: 38-57) encoding various enzymes from thiolase family (see FIG. 18 for the list of the thiolases tested) were codon-optimized for expression in *E. coli*, synthesized and cloned into pET15b vector with the *NdeI* and *BamHI* restriction sites. Gene sequences were checked to ensure they did not contain the recognition sequences of these two restriction enzymes prior to cloning. The selection was supplemented by *paaI* from *E. coli* which was cloned out of the respective genomic DNA. For P0C7L2 and Q0KBP1, the genes (SEQ ID NO: 58, and SEQ ID NO: 59, respectively) were cloned from naturally occurring sources. Specifically, the gene encoding P0C7L2 was from *Escherichia coli*, and the gene encoding Q0KBP1 was from *Cupriavidus necator*. The following chart shows the list of the genes expressed in this Example, as well as their corresponding thiolase designations as shown FIG. 18. See also FIG. 19.

SEQ ID NO	Designation in FIG. 18
SEQ ID NO: 38	Q88E32 from <i>Pseudomonas putida</i>
SEQ ID NO: 39	A5VAC9 from <i>Sphingomonas wittichii</i>
SEQ ID NO: 40	D2I940 from <i>Pseudomonas reinekei</i>
SEQ ID NO: 41	Q51956 from <i>Pseudomonas putida</i>
SEQ ID NO: 42	Q146J4 from <i>Burkholderia xenovorans</i>
SEQ ID NO: 43	Q13Q20 from <i>Burkholderia xenovorans</i>
SEQ ID NO: 44	Q0SCR8 from <i>Rhodococcus jostii</i>
SEQ ID NO: 45	Q6MM13 from <i>Bdellovibrio bacteriovorus</i>
SEQ ID NO: 46	C9XWW6 from <i>Cronobacter turicensis</i>
SEQ ID NO: 47	A0JWV4 from <i>Arthrobacter sp.</i>
SEQ ID NO: 48	D5VE84 from <i>Caulobacter segnis</i>
SEQ ID NO: 49	A8LRD1 from <i>Dinoroseobacter shibae</i>
SEQ ID NO: 50	Q13HE4 from <i>Burkholderia xenovorans</i>
SEQ ID NO: 51	GK1320 from <i>Geobacillus kaustophilus</i>
SEQ ID NO: 52	B2ICW2 from <i>Beijerinckia indica</i>
SEQ ID NO: 53	K8QJZ5 from <i>Citrobacter freundii</i>
SEQ ID NO: 54	G0EVE5 from <i>Cupriavidus necator</i>
SEQ ID NO: 55	D0L7M3 from <i>Gordonia bronchialis</i>
SEQ ID NO: 56	D5WHY6 from <i>Burkholderia sp.</i>
SEQ ID NO: 57	E1VSS4 from <i>Arthrobacter arilaitensis</i>
SEQ ID NO: 58	P0C7L2 from <i>Escherichia coli</i>
SEQ ID NO: 59	Q0KBP1 from <i>Cupriavidus necator</i>

For each culture, expression plasmids were freshly transformed into BL21(DE3) (Agilent). Colonies from fresh agar plates were used to inoculate 20ml LB in 250ml flask overnight pre-culture. For all constructs except 237, ampicillin was used, for 237 kanamycin was required. After incubation overnight at 37°C and 200rpm, pre-cultures were used to inoculate (1:100) larger cultures, 350ml LB, with respective antibiotic in 1l flasks. After 2.5h of shaking with 200rpm at 37°C, cultures reached OD600 in the range 0.5-0.7 at which point they were induced with 1mM IPTG and the temperature of the culture was changed to 25°C. Incubation continued overnight. Cells were harvested the next day and stored at -20°C until protein purification was performed.

To purify an overexpressed protein, a bacterial pellet from a 350ml culture was resuspended in up to total 20ml of Binding Buffer (20mM Sodium phosphate pH 7.4, 500mM Sodium chloride, 20mM imidazole, 5% glycerol) and sonicated for 2min with a microtip (amplitude 50%, 1s pulse ON, 2s OFF). Cell suspension was centrifuged with 20,000G for 30min at 4°C. As all the tested proteins were His-tagged, the purification with Immobilized Metal Affinity Chromatography was sufficient to ensure both necessary purity and amount for the enzymatic assays. HisTrap (GE Healthcare) purification has been performed exactly according to manufacturer's protocol. Proteins were eluted with 2x 60µl of Elution Buffer (20mM Sodium phosphate pH 7.4, 500mM Sodium chloride, 500mM imidazole, 5% glycerol) and stored frozen at -20°C.

Prior to tests with non-native substrates, enzymes were assayed for their native biological activity. Many of the selected enzymes are putative proteins, whose existence was inferred from homology to known thiolases and thus had not yet been biochemically investigated. In most cases there was no information available on what could be a possible native substrate. Therefore to assess their activity acetoacetyl-CoA was successfully used for all of them as a test substrate. The method from Slater (Slater *et al.*, *J. Bacteriol.*, 1998, 180(8):1979 - 1987) was used with minor modifications for assay of native enzyme activity. Assays were performed in disposable Corning UV-transparent 96-well plates without cover. Typically 200µl reaction mix contained 150mM Tris-HCl pH 8.0, 50mM MgCl₂, 100µM CoA, with or without 40µM acetoacetyl-CoA (Sigma). Reaction was started by the addition of 10µl enzyme prepared as described above.

Absorbance at 304nm, corresponding to the disappearing acetoacetyl-CoA, was followed every minute for 2 hours. Negative controls with/without main substrate or with/without enzyme, as well as with empty vector were always included. The method showed that there was a good correlation between the solubility of the enzyme as judged from the SDS-PAGE gels and the apparent thiolase activity tested with this assay.

The set of thiolases was tested to investigate whether they would be able to accept 5-hydroxyvaleryl-CoA as a substrate to produce 3-oxo-7-hydroxyheptanoyl-CoA. Selected thiolases were tested in a coupled assay with BDIGENE# 246 CoA transferase using 5-hydroxyvalerate and acetyl-CoA as substrates.

LC-MS based assay performed in 96-well format with 300µl total reaction volume in each well. Reaction mix contained 50mM potassium phosphate pH 6.8, 1mM acetyl-CoA (Applichem) and 1mM of 5-hydroxyvalerate. Control samples without the substrate and/or without enzyme were always included. Samples with the empty vector control were included in each run as a reference. The reaction was initiated by the addition of 10µl of enzyme aliquote to 300µl reaction mix. The plate was covered with adhesive tape to minimize the evaporation and incubated at 30°C with gentle shaking (600rpm in the Eppendorf thermomixer) for 3 hours. After the incubation samples were transferred with multichannel pipette to the 96-format filter (AcroPrep Advanced Filter Plate 0.2µm Supor) and filtered by centrifugation at 1500G for 4min to a clean 96-well plate before the LC-MS analysis. There were no analytical standards available for the reaction products hence the analysis relied exclusively on the presence of ions with expected molecular weight.

Analysis of acyl-CoAs was performed using an Agilent Technologies 1290 Series Infinity HPLC system coupled to an Agilent 6530 Series Q-ToF mass spectrometer equipped with an ESI interface operating in positive polarity with centroid data storage. QToF parameters were as follows: source temperature 350 °C, drying gas flow rate 13 l/min, nebulizer pressure 60 psig, sheath gas temperature 400 °C, sheath gas flow rate 12 l/min, VCap 3500 V, nozzle voltage 1000 V, fragmentor voltage 100 V, skimmer voltage 60 V, Octopole 1 RF voltage 750 V; scan range 50 – 1100 m/z, scan rate 2 spectra/sec; reference masses of 121.0509 m/z and 922.0098 m/z were infused directly into the source

of the MS to ensure accurate mass correction of the instrument. Mobile phase A was 10 mM ammonium acetate and mobile phase B was acetonitrile.

For acyl-CoA screening a C18 column measuring 2.1 mm x 50 mm with a 1.8 μ m particle size, 95 Å pore size and 0.5 μ m frit was used (Agilent part no: 959757-902). Autosampler was kept at 4 °C and 5 μ l of analyte was injected from each sample. Between injections the needle was washed with mobile phase in the flush port for 3 s to decrease the possibility of carryover contamination. Acyl-CoAs were eluted according to varying the percentage of mobile phase B as follows: 0 – 1 min, 3 % B; 4 – 5.2 min, 100 % B; 5.5 – 6.8 min, 3 % B. Total run time was 6.8 minutes per sample. The intermediate product of the coupled reactions, 5-hydroxyvaleryl-CoA, was detected in all samples containing both substrates and BDIGENE# 246. The final product, 3-oxo-7-hydroxyheptanoyl-CoA, was detected in several samples as illustrated in FIG. 18.

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.

WHAT IS CLAIMED IS:

1. A method of producing 3-oxo-7-hydroxyheptanoyl-CoA or a salt thereof, said method comprising enzymatically converting 5-hydroxypentanoyl-CoA to 3-oxo-7-hydroxyheptanoyl-CoA using a polypeptide having the activity of a *β -ketothiolase* classified under EC. 2.3.1.-.
2. The method of claim 1, wherein said *β -ketothiolase* is classified under EC 2.3.1.16.
3. The method of claim 1, wherein said *β -ketothiolase* is classified under EC 2.3.1.174.
4. The method of any one of claims 1-3, wherein said *β -ketothiolase* (a) has at least 70% sequence identity to the amino acid sequence set forth in SEQ ID NO:1 or SEQ ID NO: 13 or (b) has at least 70% sequence identity to the amino acid sequence set forth in SEQ ID NO:1 or SEQ ID NO: 13 and is capable of converting 5-hydroxypentanoyl-CoA to 3-oxo-7-hydroxyheptanoyl-CoA.
5. The method of any one of claims 1-4, further comprising enzymatically converting 3-oxo-7-hydroxyheptanoyl-CoA to 7-hydroxyheptanoate using a polypeptide having the activity of a *3-hydroxyacyl-CoA dehydrogenase* or a *3-oxoacyl-CoA reductase*, an *enoyl-CoA hydratase*, a *trans-2-enoyl-CoA reductase*, and a *thioesterase* or a *CoA transferase*.
6. The method of claim 5, wherein said *3-hydroxyacyl-CoA dehydrogenase* or said *3-oxoacyl-CoA reductase* is classified under EC 1.1.1.35, EC 1.1.1.36, EC 1.1.1.100, or EC 1.1.1.157.
7. The method of claim 5 or claim 6, wherein said *enoyl-CoA hydratase* is classified under EC 4.2.1.17 or EC 4.2.1.119.
8. The method of any one of claims 5-7, wherein said *trans-2-enoyl-CoA reductase* is classified under EC 1.3.1.38, EC 1.3.1.44, or EC 1.3.1.8.
9. A method for biosynthesizing 7-hydroxyheptanoate, said method comprising enzymatically synthesizing 3-oxo-7-hydroxyheptanoyl-CoA from 5-hydroxypentanoyl-CoA using a polypeptide having the activity of a *β -ketothiolase* classified under EC. 2.3.1.- and enzymatically converting 3-oxo-7-hydroxyheptanoyl-CoA to 7-hydroxyheptanoate.

10. The method of claim 9, wherein 3-oxo-7-hydroxyheptanoyl-CoA is converted to 3-hydroxy-7-hydroxyheptanoyl-CoA using a polypeptide having the activity of a 3-hydroxyacyl-CoA dehydrogenase or a 3-oxoacyl-CoA reductase, 3-hydroxy-7-hydroxyheptanoyl-CoA is converted to 2,3-dehydro-7-hydroxyheptanoyl-CoA using a polypeptide having the activity of an enoyl-CoA hydratase, 2,3-dehydro-7-hydroxyheptanoyl-CoA is converted to 7-hydroxyheptanoyl-CoA using a polypeptide having the activity of a trans-2-enoyl-CoA reductase, and 7-hydroxyheptanoyl-CoA is converted to 7-hydroxyheptanoate using a polypeptide having the activity of a thioesterase or a CoA transferase.
11. The method of any one of claims 5-10, said method further comprising enzymatically converting 7-hydroxyheptanoate to pimelic acid, 7-aminoheptanoate, heptamethylenediamine, or 1,7-heptanediol in one or more steps.
12. The method of claim 11, wherein 7-hydroxyheptanoate is converted to pimelic acid using one or more of a monooxygenase, an alcohol dehydrogenase, a 4-hydroxybutyrate dehydrogenase, a 5-hydroxyvalerate dehydrogenase, a 6-hydroxyhexanoate dehydrogenase, a 7-oxoheptanoate dehydrogenase, a 6-oxohexanoate dehydrogenase, a 5-oxovalerate dehydrogenase, or an aldehyde dehydrogenase.
13. The method of claim 11, wherein 7-hydroxyheptanoate is converted to 7-aminoheptanoate using one or more of a polypeptide having the activity of an alcohol dehydrogenase, a 6-hydroxyhexanoate dehydrogenase, a 5-hydroxypentanoate dehydrogenase, a 4-hydroxybutyrate dehydrogenase, and a ω -transaminase.
14. The method of claim 11, wherein 7-hydroxyheptanoate is converted to heptamethylenediamine using one or more of a carboxylate reductase, a ω -transaminase, an alcohol dehydrogenase, an N-acetyltransferase, and an acetylputrescine deacylase.
15. The method of claim 13 or claim 14, wherein said ω -transaminase has at least 70% sequence identity to any one of the amino acid sequences set forth in SEQ ID NOs.: 7 – 12.

16. The method of claim 11, wherein 7-hydroxyheptanoate is converted to 1,7-heptanediol using a *carboxylate reductase* and an *alcohol dehydrogenase*.
17. The method of claim 15 or claim 16, wherein said *carboxylate reductase* has at least 70% sequence identity to any one of the amino acid sequences set forth in SEQ ID NOs.: 2 – 6.
18. The method of any one of claims 1-17, wherein said 5-hydroxypentanoyl-CoA is enzymatically produced from 2-oxoadipate or malonyl-CoA.
19. The method of claim 18, wherein 5-hydroxypentanoyl-CoA is enzymatically produced from 2-oxoadipate -using one or more of an *alpha-aminotransaminase*; a 2-oxoadipate decarboxylase; a branch chain decarboxylase; a glutamate decarboxylase; a ω -transaminase; a CoA transferase, a CoA ligase, and an alcohol dehydrogenase.
20. The method of claim 18, wherein 5-hydroxypentanoyl-CoA is enzymatically produced from malonyl-CoA using one or more of a *malonyl-CoA reductase*; a 3-hydroxypropionate dehydrogenase; a 3-hydroxypropionyl-CoA synthase; a CoA-transferase; a β -ketothiolase; a 3-hydroxyacyl-CoA dehydrogenase; a 3-oxoacyl-CoA reductase, an enoyl-CoA hydratase, and a trans-2-enoyl-CoA reductase.
21. The method of any of the preceding claims, wherein said method is performed in a recombinant organism.
22. The method of claim 21, wherein said organism is subjected to a cultivation strategy under aerobic, anaerobic or, micro-aerobic cultivation conditions.
23. The method of claim 21 or claim 22, wherein said organism is cultured under conditions of nutrient limitation.
24. The method according to any one of claims 21-23, wherein said organism is retained using a ceramic hollow fiber membrane to maintain a high cell density during fermentation.
25. The method of any one of claims 21-24, wherein the principal carbon source fed to the fermentation derives from a biological feedstock.
26. The method of claim 25, wherein the biological feedstock is, or derives from, monosaccharides, disaccharides, lignocellulose, hemicellulose, cellulose, lignin,

- 88 levulinic acid, formic acid, triglycerides, glycerol, fatty acids, agricultural waste,
89 condensed distillers' solubles, or municipal waste.
- 90 27. The method of any one of claims 21-24, wherein the principal carbon source fed to
91 the fermentation derives from a non-biological feedstock.
- 92 28. The method of claim 27, wherein the non-biological feedstock is, or derives from,
93 natural gas, syngas, CO₂/H₂, methanol, ethanol, benzoate, non-volatile residue (NVR)
94 caustic wash waste stream from cycloheptane oxidation processes, or terephthalic
95 acid / isophthalic acid mixture waste streams.
- 96 29. The method of any one of claims 21-28, wherein the organism is a prokaryote.
- 97 30. The method of claim 29, wherein said prokaryote is from a genus selected from the
98 group consisting of *Escherichia*; *Clostridia*; *Corynebacteria*; *Cupriavidus*;
99 *Pseudomonas*; *Delftia*; *Bacillus*; *Lactobacillus*; *Lactococcus*; and *Rhodococcus*.
- 100 31. The method of claim 30, wherein said prokaryote is selected from the group
101 consisting of *Escherichia coli*, *Clostridium ljungdahlii*, *Clostridium*
102 *autoethanogenum*, *Clostridium kluyveri*, *Corynebacterium glutamicum*, *Cupriavidus*
103 *necator*, *Cupriavidus metallidurans*, *Pseudomonas fluorescens*, *Pseudomonas putida*,
104 *Pseudomonas oleovorans*, *Delftia acidovorans*, *Bacillus subtilis*, *Lactobacillus*
105 *delbrueckii*, *Lactococcus lactis*, and *Rhodococcus equi*.
- 106 32. The method of any one of claims 21-28, wherein the organism is a eukaryote.
- 107 33. The method of claim 32, wherein said eukaryote is from a genus selected from the
108 group consisting of *Aspergillus*, *Saccharomyces*, *Pichia*, *Yarrowia*, *Issatchenkia*,
109 *Debaryomyces*, *Arxula*, and *Kluyveromyces*.
- 110 34. The method of claim 33, wherein said eukaryote is selected from the group consisting
111 of *Aspergillus niger*, *Saccharomyces cerevisiae*, *Pichia pastoris*, *Yarrowia lipolytica*,
112 *Issathenkia orientalis*, *Debaryomyces hansenii*, *Arxula adeninivorans*, and
113 *Kluyveromyces lactis*.
- 114 35. The method of claim 21, wherein the organism's tolerance to high concentrations of a
115 C7 building block is improved through continuous cultivation in a selective
116 environment.

- 117 36. The method of any one of claims 21-35, wherein said host comprises an attenuation
118 to one or more of the following enzymes: *polyhydroxyalkanoate synthase*, an *acetyl-*
119 *CoA thioesterase*, a *phosphotransacetylase* forming acetate, an *acetate kinase*, a
120 *lactate dehydrogenase*, a *menaquinol-fumarate oxidoreductase*, an *alcohol*
121 *dehydrogenase* forming ethanol, a *triose phosphate isomerase*, a *pyruvate*
122 *decarboxylase*, a *glucose-6-phosphate isomerase*, NADH-consuming
123 *transhydrogenase*, an NADH-specific *glutamate dehydrogenase*, a NADH/NADPH-
124 utilizing *glutamate dehydrogenase*, a *pimeloyl-CoA dehydrogenase*, an *acyl-CoA*
125 *dehydrogenase* accepting C7 building blocks and central precursors as substrates, a
126 *butaryl-CoA dehydrogenase*, or an *adipyl-CoA synthetase*.
- 127 37. The method of any one of claims 21-36, wherein said organism overexpresses one or
128 more genes encoding: an *acetyl-CoA synthetase*; a *6-phosphogluconate*
129 *dehydrogenase*; a *transketolase*; a *puridine nucleotide transhydrogenase*; a
130 *glyceraldehyde-3P-dehydrogenase*; a *malic enzyme*; a *glucose-6-phosphate*
131 *dehydrogenase*; a *glucose dehydrogenase*; a *fructose 1,6 diphosphatase*; a *L-alanine*
132 *dehydrogenase*; a *L-glutamate dehydrogenase*; a *formate dehydrogenase*; a *L-*
133 *glutamine synthetase*; a *diamine transporter*; a *dicarboxylate transporter*; and/or a
134 *multidrug transporter*.
- 135 38. A recombinant organism comprising at least one exogenous nucleic acid encoding (i)
136 a β -ketothiolase, (ii) a *thioesterase* or a *CoA transferase*, and one or more of (iii) a 3-
137 *hydroxyacyl-CoA dehydrogenase* or a *3-oxoacyl-CoA reductase*, (iv) an *enoyl-CoA*
138 *hydratase*, and (v) a *trans-2-enoyl-CoA reductase*, said organism producing 7-
139 *hydroxyheptanoate*.
- 140 39. The recombinant organism of claim 38, said host further comprising one or more of
141 the following exogenous enzymes: a *monooxygenase*, an *alcohol dehydrogenase*, a 4-
142 *hydroxybutyrate dehydrogenase*, a *5-hydroxyvalerate dehydrogenase*, a 6-
143 *hydroxyheptanoate dehydrogenase*, a *7-oxoheptanoate dehydrogenase*, a 7-
144 *oxoheptanoate dehydrogenase*, a *5-oxovalerate dehydrogenase*, or an *aldehyde*
145 *dehydrogenase*, said host further producing pimelic acid.

- 146 40. The recombinant organism of claim 38, said host further comprising one or more of
147 the following exogenous enzymes: a *monooxygenase*, a *transaminase*, a 6-
148 *hydroxyhexanoate dehydrogenase*, a 5-*hydroxypentanoate dehydrogenase*, a 4-
149 *hydroxybutyrate dehydrogenase*, and an *alcohol dehydrogenase*, said host further
150 producing 7-aminoheptanoate.
- 151 41. The recombinant organism of claim 38, said host further comprising one or more of
152 the following exogenous enzymes: a *carboxylate reductase*, a ω -*transaminase*, a
153 *deacylase*, an *N-acetyl transferase*, or an *alcohol dehydrogenase*, said host further
154 producing heptamethylenediamine.
- 155 42. The recombinant organism of claim 38, said host further comprising an exogenous
156 *carboxylate reductase* and an exogenous *alcohol dehydrogenase*, said host further
157 producing 1,7-heptanediol.
- 158 43. The recombinant organism of any one of claims 38-42, said host further comprising
159 one or more of the following exogenous enzymes: an *alpha-aminotransaminase*; a 2-
160 *oxoadipate decarboxylase*; a *branch-chain decarboxylase*; a *glutamate*
161 *decarboxylase*; a ω -*transaminase*; a *CoA-ligase*; a *CoA-transferase*; and an *alcohol*
162 *dehydrogenase*.
- 163 44. A non-naturally occurring organism comprising at least one exogenous nucleic acid
164 encoding at least one polypeptide having the activity of at least one enzyme depicted
165 in any one of Figs. 1 to 5.
- 166 45. A nucleic acid construct or expression vector comprising
167 (a) a polynucleotide encoding a polypeptide having β -*ketothiolase* activity, wherein
168 the polynucleotide is operably linked to one or more heterologous control sequences
169 that direct production of the polypeptide and wherein the polypeptide having β -
170 *ketothiolase* activity is selected from the group consisting of: (a) a polypeptide having
171 at least 70% sequence identity to the polypeptide of SEQ ID NOs: 1 or 13;
172 (b) a polynucleotide encoding a polypeptide having ω -*transaminase* activity, wherein
173 the polynucleotide is operably linked to one or more heterologous control sequences
174 that direct production of the polypeptide and wherein the polypeptide having ω -

transaminase activity is selected from the group consisting of: (a) a polypeptide having at least 70% sequence identity to the polypeptide of SEQ ID NOs: 7-12.

46. A composition comprising the nucleic acid construct or expression vector of claim 45.

47. A non-naturally occurring biochemical network comprising at least one substrate of Fig. 1, at least one exogenous nucleic acid encoding a polypeptide having the activity of at least one enzyme of Fig. 1 and at least one product of Fig. 1.

48. A bio-derived product, bio-based product or fermentation-derived product, wherein said product comprises:

i. a composition comprising at least one bio-derived, bio-based or fermentation-derived compound according to any one of claims 1-37, or of Figure 1 or any combination thereof,

ii. a bio-derived, bio-based or fermentation-derived polymer comprising the bio-derived, bio-based or fermentation-derived composition or compound of i., or any combination thereof,

iii. a bio-derived, bio-based or fermentation-derived resin comprising the bio-derived, bio-based or fermentation-derived compound or bio-derived, bio-based or fermentation-derived composition of i. or any combination thereof or the bio-derived, bio-based or fermentation-derived polymer of ii. or any combination thereof,

iv. a molded substance obtained by molding the bio-derived, bio-based or fermentation-derived polymer of ii. or the bio-derived, bio-based or fermentation-derived resin of iii., or any combination thereof,

v. a bio-derived, bio-based or fermentation-derived formulation comprising the bio-derived, bio-based or fermentation-derived composition of i., bio-derived, bio-based or fermentation-derived compound of i., bio-derived, bio-based or fermentation-derived polymer of ii., bio-derived, bio-based or fermentation-derived resin of iii., or bio-derived, bio-based or fermentation-derived molded substance of iv, or any combination thereof, or

vi. a bio-derived, bio-based or fermentation-derived semi-solid or a non-semi-solid stream, comprising the bio-derived, bio-based or fermentation-derived composition of i., bio-derived, bio-based or fermentation-derived compound of i., bio-derived, bio-

205 based or fermentation-derived polymer of ii., bio-derived, bio-based or fermentation-
206 derived resin of iii., bio-derived, bio-based or fermentation-derived formulation of v.,
207 or bio-derived, bio-based or fermentation-derived molded substance of iv., or any
208 combination thereof.
209
210

SUBSTITUTE SHEET (RULE 26)



FIG. 2

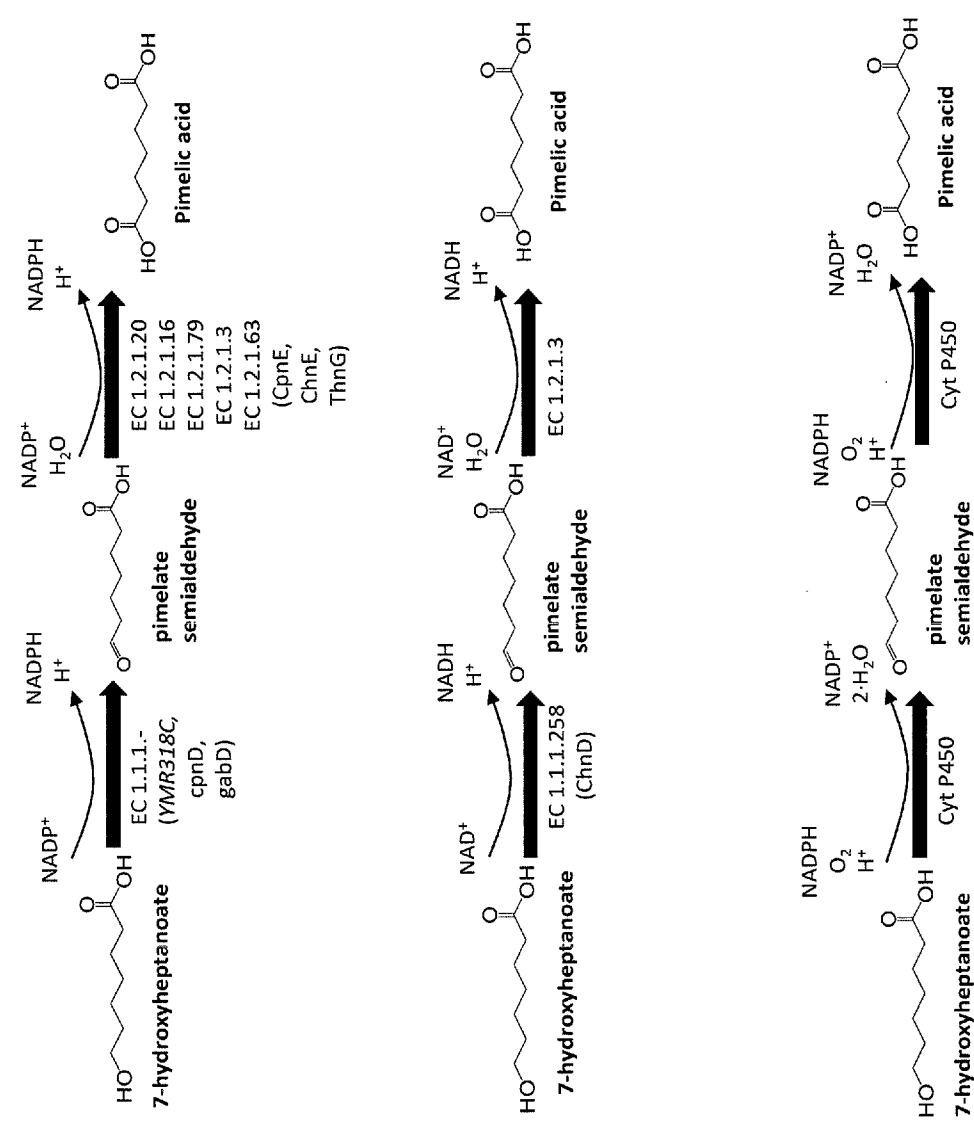
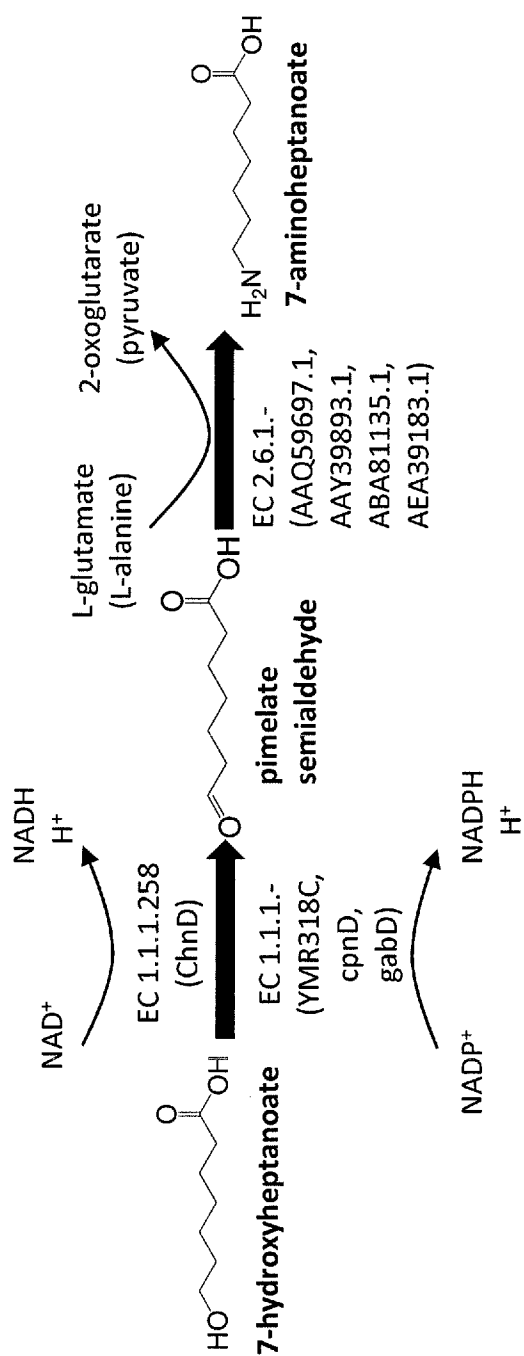


FIG. 3



[illegible]

FIG. 5

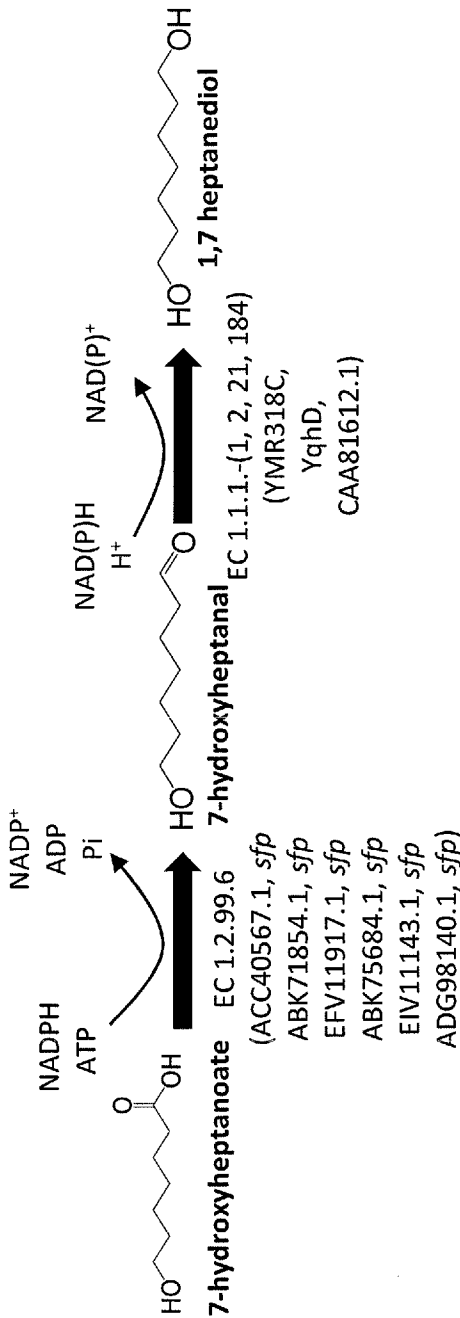


FIG. 6

SEQ ID NO	Organism	GENBANK reference	Amino acid sequence
1	<i>Cupriavidus necator</i>	AAC38322.1	MTREVVVSGVRTAIGTFGGSLKDVAPAELGALVVREALARAQVSGDDVGHVVFGNVIQT EPRDMLGRVAAVNGGVNTINAPALTVNRLCGSLQAIVSAAQTLLGDTDV AIGGGAESM SRAPYLAPAAARWGARMGDAGLVDMMLGALHDPFHRHMGVTAENVAKKEYDISRAQQDEAA LESHRRASAAIKAGYFKDQIVPVVSKGRKGDVTFDTHDVHRDA TIDDMTKLRPVPVKEN GTVTAGNASGLNDA AAA VMMERAEAEERRGLKPLARLVSYGHAGVDPKAMGIGPVPATKI ALERAGLQVSDLDVIEANEFAAQA CAVTKALGLDPAKVNPNGSGISLGHPIGATGALIT VKALHELNRVQGRYALVTMCIGGGQGIAAFERI MSPITREERLRIQDLYANDPQFAAAKPATAITAAIERPGLPLPQIETVM TGYAD RPA LAQRSVEFVTDAGTGHTTLRLPHFETISY GELWDRISALADVLSTEQT V KPGDRVCLLG FNSVDYATIDMTLARLGAVAVPLQTSAAITQLQPIV AETQPTMIAASVDALADATELALS GQTATRVLVFDHHRQVDAHRAA VESARERLAGSAVVE TLAEAIARGDVPRGASAGSAPGT DVSDDSLALLIYTSGSTGAPKGAMYPRRNVAITFWRKRTWFEGGYEPSITLNFMPMSHVMG RQILYGTLNCGGTAYFVAKSDLSTLFEDLALVRPTLTFVPRVWDMVDFEQSEVDRRLV DGADRVALEAQVKAIEIRNDVVLGGRYTSALTGSAPISDEMKA WVEELDMHLVEGYGSTE A GMILIDGAIRRP AVL DYKLVDVVDLGYFLTDRPHPRGELLVKTDSLFPGY YQRAEVTADV FDADGFYRTGDMAEVGP EQFVYLD RNNV LKLSQQGEFVTVSKLEAVFGDSPLVRQIYIY GNSARAYLLAVIVPTQEALDAVPVEELKARLGDSLQEVAKAAGLQSYEIPRDFIETTPW TLENGLLTGIRKLARPQLKKHYGELLEQIYTDLAHQQADELSLRQSGADAPVLVTVCR A AAALLGGSASDVQPD AHFTDLGGDSL SALSFTNLLHEIFDIEVPVGVIVSPANDLQALAD YVEAARKPGSSRPTFASVHGASNGQVTEVHAGDLSLDKFDIDAATLAEAPRLPAANTQVRT VLLTGATGFLGRYLALEWLERMDLVDGKLICLVRAKSDTEARARLDKTFDSGDPELLAHY RALAGDHLEVL AGDKGEADLGLDRQTWQRLADTVDLIVDPAALVNHVLPYSQLFGPNALG TAE LLRLALTSKIPYSYTSITGVADQIPPSAFTEDADIRVISATRAVDDSYANGYSNSK WAGEVLLREAHDL CGLPVA VFRCMDILADTTWAGQLNVPDMFTRMILSLAATGIAPGSFY ELAADGARQRAHYDGLPVEFIAEAI STLGAQSQDGFHTYHVMNPNYDDGIGLDEFVDWLN E SGCPIQRIADYGDWLQRFETALRALPDRQRHSSLLPLLHNYRQPERPVRGSIAPTDRFRA AVQEA KIGPKDIPVHGAPIVKYVSDRLLLGLL
2	<i>Mycobacterium marinum</i>	ACC40567.1	

SEQ ID NO	Organism	GENBANK reference	Amino acid sequence
3	<i>Mycobacterium smegmatis</i>	ABK71854.1	MTSDVHDATDGVTTETALDDEQSTRRIAELYATDPEFAA AAPLPAVVDAAHKPGRLAEIL QTLFTGYGDRPALGYRARELATDEGGRTVTRLLPRFDLTLYAQVWSRVQAVAAALRHNF QPIYPGDVAATIGFASPDYLTLDLVCAYLGLVSVPLQHNAPVSRILAEVEPRILTVS AEYLDLA VESVRDVNSVSQLVVFDDHPEVDDHRRDALARAREQLAGKGIATTTLDIADEG AGLPAEPIYTADHDQRLAMILYTSGSTGAPK GAMYTEAMV ARLWTMSFITGDPPTVINVN FMPLNHLGGRIPISTAVQNGGTSYFVPESDMSTLFEDLALVRPTELGLVPRVADMLYQHH LATVDRLVTQGADELTAEKQAGAELEQVLGGRVITGFVSTAPLAAEMRAFLDITLGAHI VDGYGLTETGAVTRDGVIVRPPVIDYKLIDVPELGYFSTDKPYPRGELLVRSQTLTPGY KRPEVTASVFDRLDGYHTGDVMAETAPDHLVYVDRRNNVLKLAQGEFVAVANLEAVFSGA ALVRQIFVYGNRSERFLLAVVPTPEALEQYDPAALKAAALADSLQRTARD AELQSYEVPA DFIVETEFSAAANGLLSGVGKLLRPNLKD RYGRLEQMYADIAATQANQLRELRAAATQ PVIDTLTQAAA TILGTGSEVASDAHFTDLGGDSL SAL.TLSNLLSDFGFEVPVGTIVNPA TNLAQLAQHIEAQRTAGDRRPSFTTVHGADATEIRASELTLDKFIDAE/TLRAAPGLPKVT TEPRTVLLSGANGWLGRFLTQWLERLAPVGGTITIVRGRDDAAAARALTQAYDIDPEL SRRFAELADRLRVVAGDIGDPNLGLTPEIWHRLAAEVDLVVHPAALVNHVLPYRQLFGP NVVGTAEVVKLALTERIKPVTYLSTVSVAMGIPDFEEDGDIRTVSPVRPLDGGYANGYGN SKWAGEVLLREAHDL CGLPVA TFRSDMILAHPRYRGQVNVPMFTRLLLSLLITGVAPRS FYIGDGERPRAHY PGLTVDFV AEA VTTLGAQQREGYVSYDVMNPHDDGISLDVFDWLIR AGHPIDRVDDYDDWVRRFETALTALPEKRR AQTVPLLLHAFRA PQAPLRGAPEPTEVFHA AVRTAKVGP GDIPHLDEALIDKYIRD LREFGLI

SEQ ID NO	Organism	GENBANK reference	Amino acid sequence
4	<i>Segniliparus rugosus</i>	EFV11917.1	MGDGEERAKRFFQRIGELSA TD PQFAAAA PDPAVVEAVSDPSLSFTRYLD TLMRGYAERP ALAH RVGAGYETISY GELWARVGAIAAAWQADGLVATPGDFVATVGFTSPDYVAVDLAAARS GLVSVPLQAGASLAQLVGILEETEPKVLAA SASSLEGA VACALAAPSVQRLVVF DLRGPD ASESAADERRGALADAEEQ LARAGRAV VVETLADLAARGEALPEAPLFEPAEGEDPLALL IYTS GSTGAPKGAMYSQRLVSQLWGRTPVVP GMPNLSHYMPLSHSYGRAVLAGALSAGG TAHFTANS DLS TLFE DIALARPTFLALVPRVCEMLFQESQRGQDV AELRERV LGGRLLVA VCGSAPLSPEMRAFMEEVLGFPLLDGYGSTEALGV MRNGIIQRPVIDYKLVDPVPELGYR TTDKPYPRGELCIRSTLSISGYYKRPEITAEVFD AQGYKTDGVMAEIAPDHLVYVDRSK NVLKL SQGEFVAVAKLEAA YGTSPYVKQIFVYGN SERSFLLAVVVPNAEVLGARDQEEAK PLIAASLQKIAKEAGLSYEVPDRFLIETEPFTTQNGLLSEVGKLLRPKLKARYGEALEA RYDEIAHGQADELRALRDGAGQRPV VETVVRAAVAISGSEGA EVGPEANFADLGGDSL SA LSLANLLHDVFEVEVPVRIIGPTASLAGIAKHIEAERAGASAPTAA SVHGAGATRIRAS ELTLEKFLPEDLLAAAKGLPAADQVRTVLLTGANGWLGRFLALEQLERLARSQDGGKLI CLVRGKDA AARRRIEETLGTDPALAA RFAELAEGRLEVVP GDVGEPKFGLDDAAWDRLA EEVDVIVHPAALVNHVLPYHQLFGPNVVGTA EIIRLAI TAKRKPVTYLSTVA AAGVEPS SFEEDGDIRAVVPERPLGDGYANGYGN SKWAGEVLLREAH ELVGLPVAVFRSDMILAHTR YTGQLNVPDQFTRLVLSLLATGIAPKSFYQQGAAGERQRAHYDGPVDFTA EAITLGAE PSWFDGGAGFRSFDVFNPHHDGVGLDEFVDW LIEAGHPISRIDDHK EWFA RFETA VRGLP EAQRQHSLLPLLRAYSFPHPPVDG SVYPTGKFQGA VKAAQVGSDDHDVPHLGKALIVKYAD DLKALGLL

SEQ ID NO	Organism	GENBANK reference	Amino acid sequence
5	<i>Mycobacterium abscessus subsp. bolletii</i>	EIV11143.1	MTNETNPQQEQLSRRIESLRESDPQFRAAQDPVAVAEQVLRPGLHLSEALAMTGYAER PALGERARELVIDQDGRITLLPRFDTTTTYGELWSRTTSVAAA WHHDATHPVKAGDLVA TLGFTSIDYTVLDLAIMILGGVAVPLQTSAPASQWTTILAEAPNTLA VSIELIGAAAMES VRATPSIKQVVVEDYTPEVDQREAFEAASQTLAGTGIALETDLDAVIARGAALPAAPLYA PSAGDDPLALLIYTSGSTGAPKGAMHSENIIVRRWWIREDVMAGTENLPMGLNFMFPMMSHI MGRGTLTSTLSTGGTGYFAASSDMSTLFEDMELIRPTALALVPRVCDMVFRFQTEVDRR LASGDTASAEAAEVKADIRDNLFGGRVSAVMVGSAPLSEELGEFIESCFELNLTGYYG STEAGMVFDRDGVQRPPIYKLVDPVPELGYFSTDKPHPRGELLKKTGDMFLGYYKRPV TASVFDADGFYMTGDIVAELAHDNIEIIDRRNNVVKLSQGEFVAVATLEAEYANSPVVHQ IYVYGSSERSYLLAVVPTPEAAAKGDAAALKTTIADSLQDIAKEIQLQSYEVPRDFI IEPQFTQGNGLLTGIAKLARPNLKAHYGPRLEQMYAEIAEQQAELRALHGVDPDKPAL ETVLKAAQALLGVSSAEAAADAHFTDLGGDSLSALSFSDLLRDIFAVEVPVGVIVSAAND LGGVAKFVDEQRHSGGTRPTAETVHGAGHTEIRAADLTLDKFIDEATLHAAPSLPKAAGI PHTVLLTGSNGYLGHYLALEWLERLDKTDGKLIIVIRGKNAEAAAYGRLEEAFDTGDTTELL AHFRSLADKHLEVLAGDIGDPNLGLDADTWQRLADTVDVIVHPAALVNHVLPYNQLFGPN VVGTAIEIKLAITTKIKPVTYLSTVAAAYVDPPTTFDEESDIRLISAVRPIDDGANGYG NAKWAGEVLLREAHDLCLGPVAVFRSDMILAHSRYTGQLNVPDQFTRJLSLATGIAPG SFYQAQTTGERPLAHYDGLPGDFTAEAITTLGTQVPEGSEGFVTYDCVNPHADGISLDNF VDWLEAGYPIARIDNYTEWFTREDTAIRGLSEKQKQHSLLPLHAFEQPSAAENHGVVVP AKRFQHAVQAAAGIGPVGQDGTDDIPHLRRLIVKYAKDLEQLGLL

SEQ ID NO	Organism	GENBANK reference	Amino acid sequence
6	<i>Segniliparus rotundus</i>	ADG98140.1	MTQSHQTGPQASAAHSRLARRAAELLATDPQAAATLPDPEVVROATRPGLRLAERVDAIL SGYADRPALGQRSFQTVKDPITGRSSVELLPFTDTITYRELREERATAIASDLAHPQAP/A KPGDFLASIGFISVDYVAIDYAGVFAGLTAVPLQTGATLATLTAITAETAPTFLFAASIEH LPTAVDAVLATPSVRLLVFDYRAGSDEDEAREAVEAKRKIADAGSSVLVDVLDEVIARGK SAPKAPLPPATDAGDDSLSLIYTSGSTGTPKGAMYPERNVAHFVGWVAAAFEDEDAAPP VPAINITFLPLSHVASRLSLMPTLARGGLMHFVAKSDSLTFEDLKLARPTNLFVPRVV EMLYQHYQSELDRRGVQDGTREAEAVKDDRLTGLLGRILTAGFGSAPLSAELAGFIESL LQIHLVDGYGSTEAGPVWRDGYLVKPPVTDYKLIDVPELGYFSTDSHPHPRGELAIAKTQTI LPGYKRPETTAEVFDEDEGFYLTGDVVAQIGPEQFAYVDRRKNVLKLSQGEFVTLAKLEA AYSSSPLVRQLFVYGSSERSYLLAVIVPTDALKKFGVGEAAKALGESLQKIARDEGLQ SYEVRDFIETDPTFVENGLLSDARKSLRPLKEHYGERLEAMYKELADGQANELRDIR RGVQQRPTLETVRRAAAAMLGASAAEIKPDHFTDLGGDSLSALTFSNFLHDLFEVDVVPV GVVSAANTLGSVAEHIDAQLAGGRARPFTFATVHGKSGSTTIKASDLTLDKFIDEQTEAA KHLKPADPPRTVLLTGANGWLGRFLALEWLERLAPAGGKLITIVRGKDAQAQAKARLDA YESGDPKLAGHYQDLAATTLEVLADGFSEPLGLDEATWNLRADEVDVFIHPGALVNHVL PYNQLFGPNVAGVAEIIKLAITTRIKPVTYLSTVAVAGVEPSALDEDDGDIRTVSAERSV DEGYANGYGNSKWGGEVLLREAHDRITGLPVRFRSDMILAHQKYTGQVNAIDQFTRLVQS LLATGLAPKSFYELDAQGNRQRAHYDGPVDTAESITTLGGDLEGYRSYNVFNPHRDG VGLDEFVDWLEAGHPITRIDDDYDQWLSRFETSLRGLPESKRQASVPLLLHAFARPGPAV DGSFPRNTVFRDQVQAKIGAEHDIPHLGKALVLKYADDIKQLGLL
7	<i>Chromobacterium violaceum</i>	AAQ59697.1	MQQRITTSQWRELDAAHHLHPFTDTASLNQAGARVMTRGEGVYLWDESGNKIIDGMAGLW CVNVGYGRKDFAEAAARRQMEELPFYNTFFKTTTHPAVVELSSLLAEVTPAGFDRVFYTNNG SESVDTMIRMVRRYWDVQKGPEKKTIGRWNGYHGSTIGGASLGGMKYMHQGGDLPIPGM AHIEQPWVYKHGKDMTPDEFVVAARWLEEKILEIGADKVAAFVGEPIQGAGGVVPPAT YWPEIERCKYDVLVADDEVICGFGRTGEWFGHQHFGFQPDFTAAKGLSSGYLPIGAV FVGKRVAEGLIAGGDFNHGFTYSGHPVCAVAHANVAALRDEGIVQRVKDDIGPYMQKRW RETFSRFEHVDDVRGVGMVQAFTLVKNKAKRELFPDFGEIGTLCRDIFFRNNLMRACGD HIVSAPPLVMTRAEVDEMLAVAERCLEEFEQTLKARGLA
8	<i>Pseudomonas aeruginosa</i>	AAG08191.1	MNARLHATSPLGDADLVRADQAHYMHGYHVFDDHRRVNGSLNIAAGDGAYIYDTAGNRYLD AVGGMWCTNIGLGREEMARTVAEQTRLLAYSNPFCDMANPRAIELCRKLAELAPGDLDHV FLTTGGSTAVDTAIRLMHYQNCRGKRAKKHVITRINAYHGSTFLGMSLGGKSADRPAAEF DFLDERIHHLACPVYVYRAPEGLGEAEFLDGLVDEFERKILELGADRVGAFISEPVFGSGG VIVPPAGYHRRMWELCQRYDVLVYISDEVVTSFGRGLGHFFASQAVFGVQPDILTAKGLTS GYQPLGACIFSRRIWIEVIAEPDKGRCSHGFTYSGHPVACAAALKNIEIEREGLLAHAD EVGRYFEERLQSLRDLPIVGDVVRGMRFMACVEFVADKASKALFPESLNIGEWVHLRAQKR GLLVRPVHNLNVMSPPLILTREQVDTVVRVLRVRESIEETVEDLVRAGHR

SEQ ID NO	Organism	GENBANK reference	Amino acid sequence
9	<i>Pseudomonas syringae</i>	AA Y39893.1	MSANNPQTLEWQALSSSEHHLAPFSDYKQKKEKGPRITRAEGVYLWDSEGNKILDGMSGL WCVAIGYGREELADAAASKQRELPPYNNLFFQTAHPVLELAKAISDIAPEGMNHVFTGS GSEGNDTMLRMVRHYWALKGQPNKKTIIISVNGYHGSTVAGASLGGMTYMHQGDLP VVHIPQYWFEGEGDMTPDEGIWAAEQLEKKILEKLVENVGAFIAEQAGGAGVIVPPD SYWPKIKEILSRDYDLFAADEVICGFGRTSEWFGSDFYGLRPMMTIAKGLTSGYVPMGG LIVRDEIVAVLNEGDFNHGFTYSGHPVAAAVALENIRILREEKIVERVSETAPYLQKR LRELSDHPLVGEVRGVLLGAIELVKDKTTRERYTDKGAGMICRTFCFDNGLIMRAVGDT MIIAPPLVISFAQIDELVEKARTCLDLTLAVLQG
10	<i>Rhodobacter sphaeroides</i>	ABA81135.1	MTRNDATNAAGAVGAAMRDHILLPAQEMAKLGKSAQPVLTHAEGYVHTEDGRRLIDGPA GMWCAQVGYGRREIVDAMAHQAMVLPYASPWYMATSPAARLAEKIALTPGDNLRIFFTT GGSTAVDSALRFSEFYNNVLGRPQKKRIIVRYDGYHGSTALTAACRTGTGNWPNFDIAQD RISFLSSPNRHAGNRSQEAFLDDLVEFEDRIESLGPDTIAAFLAEPIASGGVIIPPA GYHARFKAICEKHDILYSISDEVVTGFGRCGEWFASEKVFVGVVDPDIITFAKGVTSYVPLG GLAISEAVLARISGENAKGSWFTNGYTYSNQPVACAAALANIELMEREGIVDQAREMADY FAAALASLRDLPVGAETRSVGLVGCVCQLLDPTRADGTAEDKAFITLKIDERCFCFELGLIVR PLGDLCSIPPLIISRAQIDEMVAIMRQAITEVSAAHGLTAKEPAAV
11	<i>Escherichia coli</i>	AAA57874.1	MNRLPSSASALACSAHALNLEKRTLDHEEMKALNREVIEYFKEHVNPFGFLEYRKSVTAG GDYGAWEWQAGSLNTLVDTQGEFIDCLGFGIFNVGHRNPVVSAVQNQLAKQPLHSQE LLDPLRAMLAKTLAALTPGKLKYSFFCNSGTESVEAALKLAKAYQSPRGKFTFIATSGAF HGKSLGALSATAKSTFRKPFMPLLPGRFHVFPFGNIEAMRTALNECKKTGDDVAAVILEPI QGEVVILPPPGLYLTAVRKLCEFGALMILDEVQTMGRGTGKMFACENHVQPDILCLAK ALGGVMPIGATIAEEVSVLFDNPFLLHTTTFGGNPLACAAALATINVLLEQNLPAAAE QKGDMLLDGFRQLAREYDDLVEARGKGMMLMAIEFVDNEIGYNFASEMFRQVRVLVAGTLN NAKTIRIEPLTLTIEQCELVIKAARKALAAAMRVSVEEA
12	<i>Vibrio fluvialis</i>	AEA39183.1	MNKPQSWEARAETYSLYGFTDMPSLHQRGTVVVTHGEGPYIVDVNGRRYLDANSGLWNMV AGFDHKGLIDAAKAQYERFPGYHAFFGRMSDQTVMLSEKLVEVSPFDSGRVFTYNSGSEA NDTMVKMLWFLHAAEGKPQKRKILTRWNAHYGVTAVSASMTGKPNYSVFGPLPGFVHLT CPHYWRYGEEGETEEQFVARLARELEETIQREGADTIAGFFAEPVMGAGGVIPPAKGYFQ AILPILRKYDIPVISDEVICGFGRTGNTWGCVTYDFTPDIAISSKNLTAGFFPMGAVILG PELSKRLETAIEAIEEFPHGFTASGHPVGCALKAIDVVMNEGLAENVRRLAPRFEERL KHIAERPNI GEYRGIFGMWALEAVKDKASKTTPFDGNLSVSEKANTCTDLGLICRPLGQS VVLCPFFILTEAQMDMEFMDKLEKALDKVFAEVA

SEQ ID NO	Organism	GENBANK reference	Amino acid sequence
13	<i>Escherichia coli</i>	AAC74479.1	MREAFICDGIRTPIGRYGGALSSVRADDLAAIPLRELLVRNPRDLAECIDDDVILGCANQA GEDNRNVARMA TL LAGLPQSVSGTTINRLCGSLDALGFAARAIAKAGDGDLLIAGGVESM SRAPFVMGKAASAFSRQAEMEDTTIGWRVFNPLMAQQFGTDSMPETAENVAELLKISRED QDSFALRSQORTAKAQSSGILAEIIVPVVLKNKGVVTEIQHDEHLRPETTTLEQLRGLKA PFRANGVITAGNASGVNDGAAALIIASEQMAAAQGLTPRARIVAMATAGVEPRLMGLGPV PATRRVLERAGLSIHDMDVIELNEAFAAQALGVLRGLPDDAPHVNPNGGAIALGHPLG MSGARLALAASHELHRRNGRYALCTMCIGVGQGIAMILERV
14	<i>Mycobacterium smegmatis</i>	ABK75684.1	MTIETREDRFRNRIDHLEFETDPQFAAARPDEAISAASAAADPELRLPAAVKQILAGYADRPALGKRA VEFVTDDEGRITAKLLPRFDITIRYQLAGRIQAVTNAWHNHPVNAAGDRVAAILGFTSVDTTIDIA LLELGAVSVPLQTSAPVAQLQPIVAETEPKVIASSVDFLADAVALVESGPAPSRLLVVFDSHEVD DQREAFEAAKGLAGTGVVVETITDALDRGRSLADAPLYVPDEADPLTLIIYTSGSTGTPKGAM YPESKATAMWQAGSKARWDETLGVMPISITLNFMPMSHVMMGRGILCSTLASGGTAYFAARSCLS TFLEDLALVRPTQLNFVRIWDMLFQEYQSRLDNRRAESESDRAEAAVLEEVRTQLLGGRFVSA LTGSAPISAEMKSWVEDLLDMHLEGGYSTEAGAVFIDGQIQRPVIDYKLVDVPDLGYFATDRP YPRGELLVKSEQMFPGYKRPETAEEMFEDGYRTGDIVAEELGPDHLEYLDRNNNVKLSSQGE FVTVSKLEAVFGDSPLVRQIYVYGNARSYLLAVVVPTEALSRWDGDELKSRISDSLQDAARA AGLQSYEIPRDFLVEPTPTFLENGLLTGIRKLARPKLKAHYGERLEQLYTDLAEGQANELRELRR NGADRPVETVSRAAVALLGASVTDLRSDAHFTDLGGDSLSALSFSNLLHEIFDVPVGVIVSP ATDLAGVAAVIEGELRGSKRPTYASVHGRDATEVRARDLALGKFIDAKTLA AAPGLPRSGTEIRT VLLTGATGFLGRYLALWLERMDLVGKVICLVRAKSDDEARARLDATFDTGDATLLEHYRAL AADHLEVIAGDKGEADLGLDHDWTQRLADTVLIVDPAALVNHVLPYSQMFQPNALGTAEELRI ALTTTIKPYVYVSTIGVGQGISPEAFVEDADIREISATRRVDDSYANGYGNSKWAGEVLLREAH WCGLPVSVFRCDMILADTTYSGQLNPLDMFTRLMLSLVATGIAPGFSFYELDADGNRQRAHYDGL PVEFIAEAISTIGSQVTDGFETFHVMNPYDDGIGLDEYVDWLIEAGYPVHRVDDYATWLSRFETA LRALPERQRQASLLPLLNHYQQPSPVCGAMAPTDRFRAAVQDAKIGPKDIPHVTDVIVKYIS NLQMLGLL
15	<i>Pseudomonas putida</i>	AAN70209.2	MNDVVIVAATRITAIGSFQGALATVPAVDLGA AAVIKQLLKQTGLDPAQVDEVILGQVLTAG AGQNPARQAIAKAGLPFSVPALTLNKVCGSLKALHLAAQAIRCGDAEFVVIAGGQENMSL APYVMPARSATGQRMGHGQLIDSMITDGLWD AFNDYHMGITAENLVDKYGLSREQQDAFAA ESQRKAVAAIEAGRFDAEITPVLVLPQKKGPEKVFARDEQPRPDTTAE SLAKLRPAFKKDG SVTAGNASSLNDGAAAVLLMSAAKAEALGLPVLA KIAIYASAGVDPAIMGIGPVSATQRC LDKAGWQLAELDLIEANEAFAAQALAVGNALAWDAARVNVNGGAIALGHPIGASGCRVLV TLIHEMIKRDVKKGLATLCIGGGOGVALAIER

SEQ ID NO	Organism	GENBANK reference	Amino acid sequence
16	<i>Sphingomonas wittichii</i>	ABQ69245.1	MEDIYVGAARTAIADFGGALKDVPADLGVIVARAALERAGLEPGDVQNVVMQVMPTE PRDAYLARVMVGVTAGVPIETPALTNLRLCGSGVEAIVTGAKAMVLGESDIVLAGGAEVMS RVPHVVKGARWGTGMGNVEMTDGLIEALSDPFDK VHMGITAEVNAERYQITREAAQDALAL QGHQRAARAIAEGRFKAQIVPVEVKTRKGVVAFDTHVHVGDVSAEELAKLRPVFKKDDGT VTAANASGINDGAAMVVLATKKA VDAKGLKPLARILSWGHAGVEPLYMGIGPVKAVPIAL ERAGLTADIDVIEANEAFAAQACAVAQELGFDPKVNPNGSGVALGHPVGATGAILTVK TVYELERIGGRYGLITMCIGGGQGIAMVVERCA
17	<i>Pseudomonas reinekei</i>	ACZ63623.1	MKNALIVSPLRTPIGKFGGALAPLTAEHLASFMSQVMARTGVPGHSLDEVIVAQSYASS EAPCIGRYAALSAGLPVEVPGYTLDRRCGSLQAVIDASMMVKGTGNAEAVLVVGVESMSN IEYYSTDMRWGARAGSVRFHDLRLGRGRSPSERFGHISGMPETADNLALDYGISREEA DSFSVRSHQNAAAA WREGRFADDEVVAVDVPGRGAVTRVTIDEGREDALESMMKALRLI RPEGVCTAGNSSQQNDAAAGCLVVSPEYAAHGLTPMARLV DWAAA GCEPSRMGIGPVPA TQKLLMRTGLSLAELDLIELNEAFAAQALAVLKTWGLDDL SRVNVNGSGISLGHPIGATG VRIMTTLHEMRREARYGLETCIGGGQGLAALFERV
18	<i>Pseudomonas putida</i>	AAA85138.1	MRDVFICDAIRTPIGRFGGALAGVRADDLAAVPLKALIEPNPAVQWDQVDEVFFGCANQA GEDNRNVARMALLAGLPESIPGVTLNRLC ASGMDAIGTAFRAIASGEMELAIAGGVESM SRAPFVMGKAESGYSRNMKLEDTTIGWRFINPLMKSYGVDSMPETADNVADDYQVSRAD QDAFALRSQQKAAAAQAAAGFFAEIIPVRIAHHKKGGETIVERDEHLRPETTLKLPV NGPDKTVTAGNASGVNDGAAALILASAEAVKHKHGLTPRARVLGMASGGVAPRVMGIGPVV AVRKLTERLGVAVSDFDVIELNEAFASQGLAVLRELGVADDAQVNPNGGAIALGHPGLGM SGARLVLTALHOLEKSGGRKGLATMCVGVGQGLALAIERV
19	<i>Burkholderia xenovorans</i>	ABE28745.1	MSETHMSGTKADPIVIVGVARTPMAAFQGDFAITLSAPQLGSVAIQAAVQRAGLKPEQIDE VVMGCVLPAGLGQAPARQAALGAGLPLATGSTTVNKMCGSGMRMAAMFAHDMMLAAGSVDVI VAGGMESMTNAPYLLPKARAGMRMGHGQVIDHMFYDGLDA YEKGRLMGSFAEECAASF FTREAQDAFAVESLARAKRANEDGSFAWEIAPVKVESRKGEVTIDRDEQPFKANIEKIPT LKPAFSKTGTVTAAANSSISDGAALVMMRESTAKRLGVQPIARVVGHSTLAQEPAKFTT APVGAIRKLFEKNGWRADDEVLDLFEVNEAFVVTMAAMKEHHLPHKEKVN NVNGGACALGHP GASGARILVTLIGALKRGGKRGVATLCIGGGEATAMGIELV
20	<i>Burkholderia xenovorans</i>	ABE33819.1	MTEAFLCDAIRTPIGRYAGALSSVRADDLGA VPLKALMERKEVDWNAIDDDVIYGCANQA GEDNRNVARMSLLLAGLPQGVPGTTVNRLCGSGMDAVGIAARAKSGEAAALMVAGGVESM SRAPFVTGKATSAFSRQAEIYDTTIGWRFNPLMKLYGVDSMPETGENVATDYNISRAD QDAFALRSQQKAAARAQRDGTLAQEI VGTIAQKKGDPVTVSQDEHPRETSLDALAKLKG VRPDGTVTAGNASGVNDGAAALLANEETARRFGLTPRARVLGIATAGVAPRVMGIGPAP ATQKLLARLNMSLDQFDVIELNEAFASQGLAVLRALGVADDDTRVNPNGGAIALGHPGLGM SGARLVTTAMYQLHRTQGRFALCTMCIGVGQGLAIAERV

SEQ ID NO	Organism	GENBANK reference	Amino acid sequence
21	<i>Rhodococcus jostii</i>	ABG94668.1	MAEVELVDGARTPQGRYGGALAGVRPDDLAGLVAAEAARRAGIPGDVDEVILGAANQAG EDNRDVARMALLAGLPDSVPGYTVNRKLCASGLTAVASAAHTIRSGEADIVAGGVESMT RAPWVMAKPGTWPARGVEADTSLGWRFNPRFTAADRDVPAAGAPDVVKVTLISMGETAE EVAALGVTAEADAFALRSQERAIAAVDAGRFERIVPVVRDGGELAADETPRRGTTLE KLGSLKPVFRTGGIVTAGSSSSLDGAAALVASEAAVEKYGLTVRGRIVTSASAGIAPN VMGLGPVPA TRKALARANWSISDLGAAELNEAFAAQSLGVIRQLKLDSEIVNADGGAIAL GHPLGCSGARILLTLGRMEREGARRGLATLCVGVGGQVAMLEAP
22	<i>Bdellovibrio bacteriovorus</i>	CAE79693.1	MKSPRDVVLVEGVRTPFKAKGTLKKVHPAELKVALKQVIAQTNDVNLVDEVIIGNTG NPPDSVNISRVVNALNAGIPKTSAYTVHRNCASALEISNGYKIKSGTMDVILAGGTEN MSQMPTLPPKKFQEYKELFAAKGPKQALPLLWSLFKADVKQKALLSGNMRDEYFPVIS VMMGLTDPFVGINMGQTAEILAKEWGLSRETQDKFALRSHQLASKAMKEGRMREEIAPVY LAPEYKEVISEDIGPRDTQTMEALAKLKPFFDKATGSITAGNSCPIIDGAAVLMMSREK AEALGYKPLATIRSYGFAGLEPERMGLGPVYSTPVALKRAGLSMKDIGLVELNEAFAAQV LSCQKAFDSDKFGQEKLGSSKIGEIRDDIILNVNGGAIALGHPVGATGTRIVLTLAKEMK RRNTQFGLATLCIGGGGGSMILENEG
23	<i>Cronobacter turicensis</i>	CBA32535.1	MFSLLQGNVMSQALPLVTRQGDRIAVSGLRTPFARQATAYHGVPAVDLGMVVGELLAR SEIPPDVIEQLVFGQVVMPEAPNIAREIVLGTGMSVHTDAYSVSRACATSFQAVANVAE SLMAGTIRAGIAGGADSSSVLPIGVSKKLARTLVDANKARTAGQRLKLFSLRLRDLLPV PPAVAEYSTGLRMGDTAEQMAKTHGITREQQDALAHRSHQLAAQAWAEGKLREEVMTAYT PPYREPLSEDNNIRKNSSLADYTKLRPAFDRKHGTVTAANSTPLTDGAAAAILMTESRAR ELGLTPLGLYRSYAFTAVDVVWQDMLGPASTPLALERAGLTMADLTIDMHEAFASQTL ANLKLASDRFAREVLRGSRQATGEVDESKFNVLGGSIAYGHPFAATGARMITQTLNELRR RGGGGLVTACAAGGLGAAMVLEAE
24	<i>Arthrobacter sp.</i>	ABK03524.1	MSFNGQSATGPDESAAAAPAAATPGAGLLRKA VVGGNRIPFARTGGAYTKSSNQDMLTAAL DGLIARFGLADERIGEVAAGA VLKHSRDFNL TREAVLGSALS AETPAYDLQQACATGLET VLGLANKIKLGOIDS AIAAGGVDSASDAPIAVSEGLREVLLDLNRAKTL PQRLLKVLGRLRP KDLAPDAPNTGEPTGLSMGEHQALTTAQWKITREAOQDELA YNSHRNLAAAYDAGFFDDL LTPYRGLNRDSNLRADTTREKLS TLKPVFGKNLGAETMTAGNSTPLTDGASTVLLASEE WADAHELKPLATVVDGEAAA VDFVHGKDGLLMAPAFAPVRLLRNGLTLDIDDDFFEIHEA FAGTVLSTLAAWEDEEFGRTRLGLDGPLGSDIRAKLNVNGSSLAAGHPFAATGGRIVATL AKMLHDKGQVDGRPARGLISICAAAGGQGVVAILEAS

SEQ ID NO	Organism	GENBANK reference	Amino acid sequence
25	<i>Caulobacter segnis</i>	ADG08907.1	MATASSAASSGVWLAAGVRSPFAKVDGALAGHDAIGLSVPVVKAMLARAKPDFAVWGT VIPNLTSNLAREVLDDAGDPTIPAFSTVMACSTSMIGAIEAAGMVDGRGRDLALVGGV ESMSRVQLGLSVALSDFWIRKFNQAKTGQQLRALGALNKLDFIPKVNNRVVTGLSMGE HTEITAKEWNLSRADQDAIALASHQGA VKGWESGFFDDLVPVGEVKRDGIPRKDTSLEK LAKLGPADFDTSGKGTLTAGNSSPLTDGAAVWVGSAAAGMARLPGETPKVRLVDYEVTSI DLRHEGLLMAPAYGVPRMLARNGLTYADVGLWEIHEFAAQVLSHIAWESAKFLSEKAG VTTPMGAFPRERMNPNNGSLALGHPFGATGARIISQTVKELAAPRKGERAIVSICADGGG GTMMLLESA
26	<i>Dinoroseobacter shibae</i>	ABV92581.1	MTEAYIYDAIRSPRGKGRKDGSLHEVTA VLSAQTLNAIKDRNGLTGHAVEDVIWGNVTQ VMEQGGCLARTAVLASDLDESIPGLAINRFCASGLEAVNLAA NQVRGGGQAYIAGGVEM MGRVPMGSDGAAIAADPSVAMKTYFPQGISAADIIATEYGISRDDADALAVASQRRAKAA WDENRFNGSVFTVRDQNLNLDHDEYMRPETDMQSLGALKPAFKDMGEQMPGFDKIALM KYPHLEKIEHIIHAGNSSGIVDGSAAALLIGNKAFGEAHGLKPRAVIKATAKIGTDTIML TGPVPA TEKILADSGMSISDIDLFEVNEAFSSVLRFMQAFDVHDHDKVNVNGGAIAMGHP LGATGAMILGTLLELERTGKGTGLATLCVASGMGAATIERV
27	<i>Burkholderia xenovorans</i>	ABE36495.1	MTRDTRDVVIVDAVRTPIGKFRGALAGVRADHLGALVIDELIRRAGVKPQAVNDVVFGCV TQIGEQSANIARTSVLGAGWPETIPGLTIDRCKGSGEEAVHIAAGLIAFGAADVIVAGGA ESMSRVPMGNSRDLHGEAFGWMA SERFELTSQGEAAERLDCDCWALTRAQLDAYSVESHRR AAAAAAEGWFAREIVPVVGVQVREKSLEGEAALFAADETIRPGTNADKLATLKSSFRSDG RLTAGNSSQISDGA AALLMSSDKARELGVKARARVRAVTTVGS DPTLMLTGPILATCQV LEKAGLGLSDIDLFEINEAFAPVPLVWMKEFGVPHAKLVNNGGAIALGHPLGASGARIMT SMLHELERRGARYGLQAICCAGGMGTATLIERLD
28	<i>Geobacillus kaustophilus</i>	BAD75605.1	MREAVIVEAVRTPVGKRNQVFRDVHPVHLAAVVLDEVVRRAGMDKGAVEDIVMGCVTPVA EQGYNIGRLAALEAGFPIEVPAVQINRMCGSGQQAIIHFAAQEIRSGMDMDVTIAAGVESMT KVPILSDGNERTIPPSLHEKYEYFIHQGVSAERIAKKYGLTREELDAYAYESHQALAAALR EGKFRAEIVPVKGLDRDGREILVTDDEGRADTSPEAL AALKPVFQEDGLITAGNASQMS DGAAAVLLMEREAAARRFGLKPKARIVAQTVVGS DPTYMLDGVIPATRQVLKAGLSIDDI DLIEINEAFAPVVLAWQKEIGAPLEK VNVNNGGAIALGHPLGATGAKLMTSLVHELERGG RYGLLTICIGHGMATATIERE
29	<i>Beijerinckia indica</i>	ACB95386.1	MTKVVIAGYIRSPFTLAKKGELATVRPDDLA AQVVKGLIKKTGIPAEIEDLLLGCAFPE GEQGFNVARLVSLAGLPLSVGASTVNRFCGSSMTTVHMAAGAIQMNAAGNAFIAAGVESM SRVPMMGFNPLPNPELAATMPGAYMGMGDTAENVA AKWTISRKEQEEFALRSHQRATAAQ KEGRLTGEIPIITGRKGTITDGCIRPDTTLEGLAELKPAFSANGVVV TAGTSSPLTDGAA AVLVCSEDYAKHHHLVDVLSVKAIAVSGCSP EIMGIPVAA SRKALARAGLEAGQIDIVE LNEAFASQSIACMRELNLS PDRVNIIDGGAIALGHPLGATGARIVGKAASLLKREKGYAL ATQCIGGGGGIATVLEAF

SEQ ID NO	Organism	GENBANK reference	Amino acid sequence
30	<i>Citrobacter freundii</i>	EKS55037.1	MEQVIVDAIRTPMGRSKGGAFNRNVRAEDLSAHLMRSLARNPALDPTALDDIYWGCVQQ TLEQGFNIARNAALLAEIPHSVPAVTNRLCGSSMQALHDAARMIMTGDAQAACLIGGVEH MGHVPMSHGVDFHPGMSRNVAKAAGMMGLTAEMLSRMHGMQDFAAARSHARAWAAT QSGAFKNEIPTGGHDADGVLKQFNYDEVIRPETTVEALSTLRPAFDPVSGTVTAGTSSA LSDGAAAMLVMSESARELGLTPRARIRSMVAVVCGDPSIMGYGPVASKLALKAGLSTS DIGLFEMNEAFAAQILPCIKDLGLMEQIDEKINLNGAIALGHPLGCSGARISTTLLNLM ERKDVQFGLATMCIGLGGQIATVFERV
31	<i>Cupriavidus necator</i>	AEI75849.1	MKQLQDAYIVAATRSPIGKAPKGAFKNTRPDDLLATILKAAVAQVNPNDPKLIEDAIVGC AIPEAQQLNVARIGALLSGLPNTVGGITVNRFCASGVSAMAAADRIRVGESDVMIAAG VESMSMVPMMGNSSPMSPEIFTRDENVGIAYGMGLTAEKVAQQWQVSRDQDAFSLASHQ KAIAAQAGGEFKDEITPIEIVERFPDLASGQVNVKTRTISLDEGPRPETSLEGLKLRPV FANKGSVTAGNSSQTS DGAGALILVSEKILKQFNLVPLARFVSFAVRGVPPEIMGIPKE AIPAALKAAGLTQDQLDWIELNEAFAAQSLAVMRDLQDPAKVNRMGGAIALGHPLGATG AIRSATVVHALLRRHNLKYGMVMTMCVGTGMGAAGIFERV
32	<i>Gordonia bronchialis</i>	ACY20886.1	MAPCSVKAMPEAVIVAHARSPIGRAGKGS�KDVPRDEL SRQMVAALAKVPELAPSDIED IHWGIGQPGGGGYNRIARVIAVELGYDHPGVTVNRYCSSSLQTRMALHAIKAGEADVL ISGGVESVSSFGGADGAPDSKNPVFDDAQARTAKAEGGAPAWTDPREQGLIPDVYI AMGQTAENVASFSGISREDQDRWSVLSONRAEEAINAGFFEREIDPVTLPDGSVTNTDDG PRAGTTYEKVSQLKPVFRPDGTVTAGNACPLNDGAAALVIMSDSKAKQLGLTPLARVVAT AATGLSPEIMGLGPIEAIRKVLISGMSLSIDLVEINEAFVQVLGSANELGIDHDKLN VSGGAIALGHPFGMTGARITTTLLNNLQTRDKTFGIESMCVGGGQGMAMVLERLS
33	<i>Burkholderia sp.</i>	ADG18081.1	MREAIVIVSTARTPLTKAHRGEFNITPGPTLASFAVRAAVERSGVDPDIIEDAILGCGYPE GTTGRNVARQSVIRAGLPLSIAGTTVNRFCASGLQALMAAAGRIVVDGAPAMAGGVESI SNIQTREDGVSGLDPWIVEHKPSLYTAMIDTADIVARRYGISREAOQDFSVESQRRTAEA QQAGRYADEIIPVTTTMAITDKETRAVSYSREVTVSADNCNRP GTTYEALAKLAPVKGPDQ FITAGNASQNADGASACVLEAKAERANFAPLGAFRGLAGCEPDEMIGIPVLAVPKL LARHGLTVDDIGLWELNEAFASQAVYCCQKRLEIPSERLNVNNGGAISIGHPFGMTGSRLVG HVLIEGRRRGVKYAVVTMCMAGMGMAAGLFEIY
34	<i>Glutamicibacter arilaitensis</i>	CBT74677.1	MQQAYLYDAIRTPFGKIGGALSSHRPDDLAHVVRRELVARSPKLDVADIDESIFGNANGA GEENRNVARMATLLAGLPTSLPGTTMNRLCGSSLDASIASRQIATGDADLVLVGGVESM SRAPWVLPKTERPFPMSNLEANTLWRLVNPAMPGEWTVSLGEATEQLREKHGHSRED QDEFSAASHQRAAAAWQAGKYDNLVVPVPANKRGTEVTRDETIRADSTAQTL SKLRTVF RTGENATVTAGNASPMSDGAFAFISERGEGELLGAAPARIASNGAAALDPQFFGFAPV EAANKALAKAGLKWSDIAA VELNEAFAAQSLACIRAWDIDPAIVNAWGGGAISIGHPLGAS GLRILGTVARRLAESGERYGLAICIGVGQGLAVVVENINATK

SEQ ID NO	Organism	GENBANK reference	Amino acid sequence
35	<i>Escherichia coli</i>	AAC74479.1	MREAFCIDGIRTPIGRYGGALSSVRADDLAAIPRELLVRNPRLDAECIDDDVILGCANQA GEDNRNVARMATLLA GLPQSVSGTTNRLCGSLDALGFAARAIAKAGDGDLLIAGGVESM SRAPFVMGKAASAFSRQAEMFDTTIGWRFNPLMAQQFTDSMPETAENVAELLKISRED QDSFALRSQRTAKAQSSGILAEIIVPVVLKNNKGVVTEIQHDEHLRPETTTLEQLRGLKA PFRANGVITAGNASGVNDGAAALIIASEQMAAAQGLTPRARIVAMATAGVEPRLMGLGPV PATRRVLERAGLSIHDMDVIELNEAFAAQALGVLRELGLPDDAPHVNPNGGAIALGHPLG MSGARLALAAASHELHRRNRYALCTMCIGVGQGIAMILERV
36	<i>Cupriavidus necator</i>	AAC38322.1	MTREVVVVSGVRTAIGTFGGSLKDVAPAEALGVVREALARAQVSGDDVGHVVFNGVVIQT EPRDMYLGRLVAAVNGGVTTINAPALTVNRLCGSLQAIVSAAQTILLGDTDVAIAGGGAESM SRAPYLAPAAARWGARMGDAGLVDMMMLGALHDPFHRIHMGVTAENVAKKEYDISRAQQDEAA LESHRRASAAIKAGYFKDQIVPVVSKGRKGDVTFDTDEHVRHDAITIDDMTKLRPVFVKEN GTVTAGNASGLNDAAA VVMMERAEAEERRGLKPLARLVSYGHAGVDPKAMGIGPVPA TKI ALERAGLQVSDLDVIEANEAFAAQACAVTKALGLDPAKVNPNGSGISLGHPIGATGALIT VKALHELNRVQGRYALVTMCIGGGQGI AIFERI
37	<i>Clostridium viride</i>	NCBI Reference Sequence: NZ_KK211198.1	MAQFVTAQEA VKHIPNGSRVVLAHSTGEPRTLVKAMVENEYEQYKDVEVCHMLGLGPYEYTNPE MKGHLWHNSLFMGPGGRKAFNENRLDFTPGYFCDTSIKFFREGYLPVDVLMMTVSPDCKHGYVT CGITCDFTMPAFECAKMMVIVQVNKNMPRTFGQTAIHLDDIDFAVEADDDPLYGSPGELTDIDRKI GEHCASLINDGACLQIGIGIPNAVLTYLTKNDMGIHSEMLSDGILQLIKAGNINNSKKQIHVGK SAVTFLNGSQELYDYVDDNPSVEFYVDYINDPYVIGKNDNMVSVNSALSVDLMGQIVADNLSA TRQISGAGGFVDFVRGATISKGGISIVAMPSTAAGGKASRIEMMFDA GRPITLTRFESFYVVEYGI AKMRGNSLRTRARQLIEIAHPDFRDEMKEFYE KRFGEKY

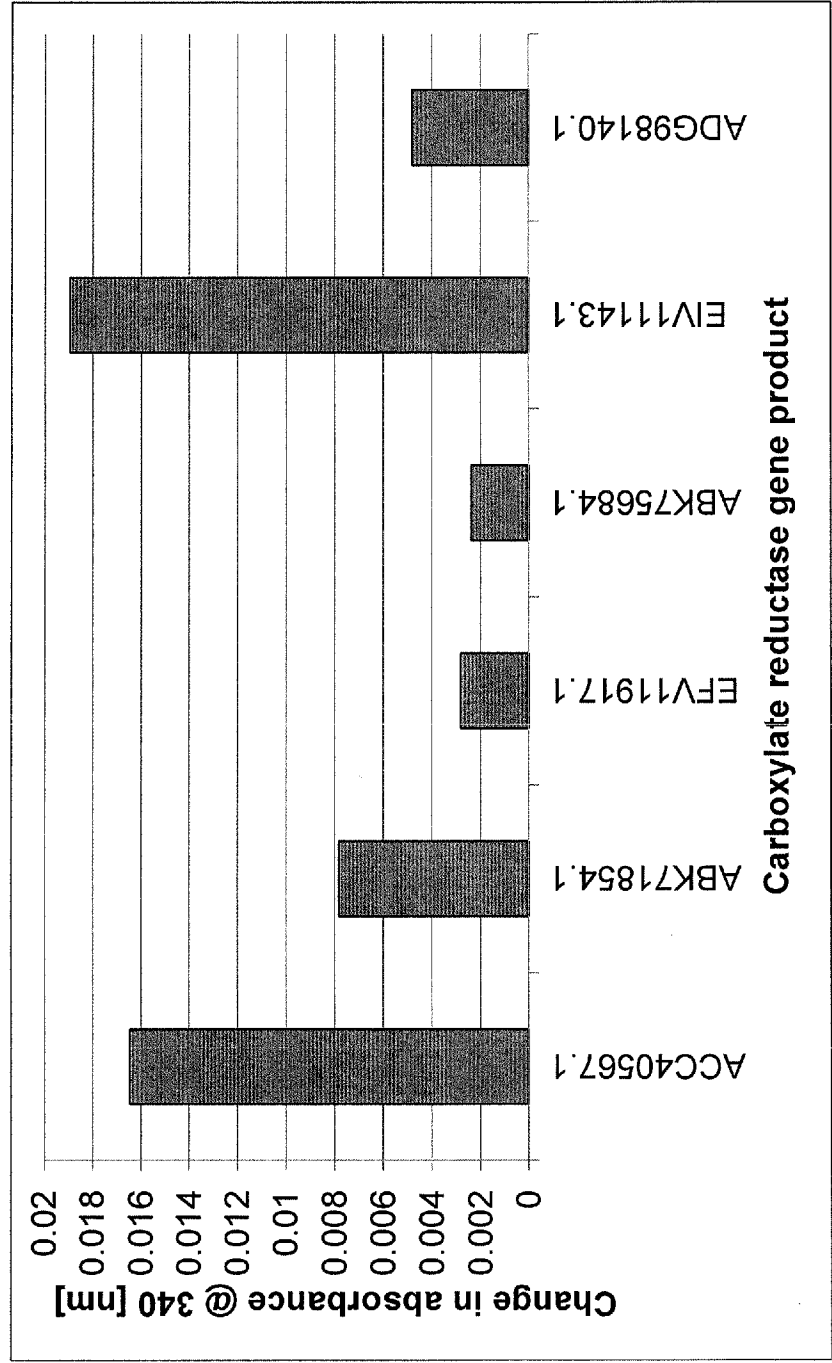


FIG. 7

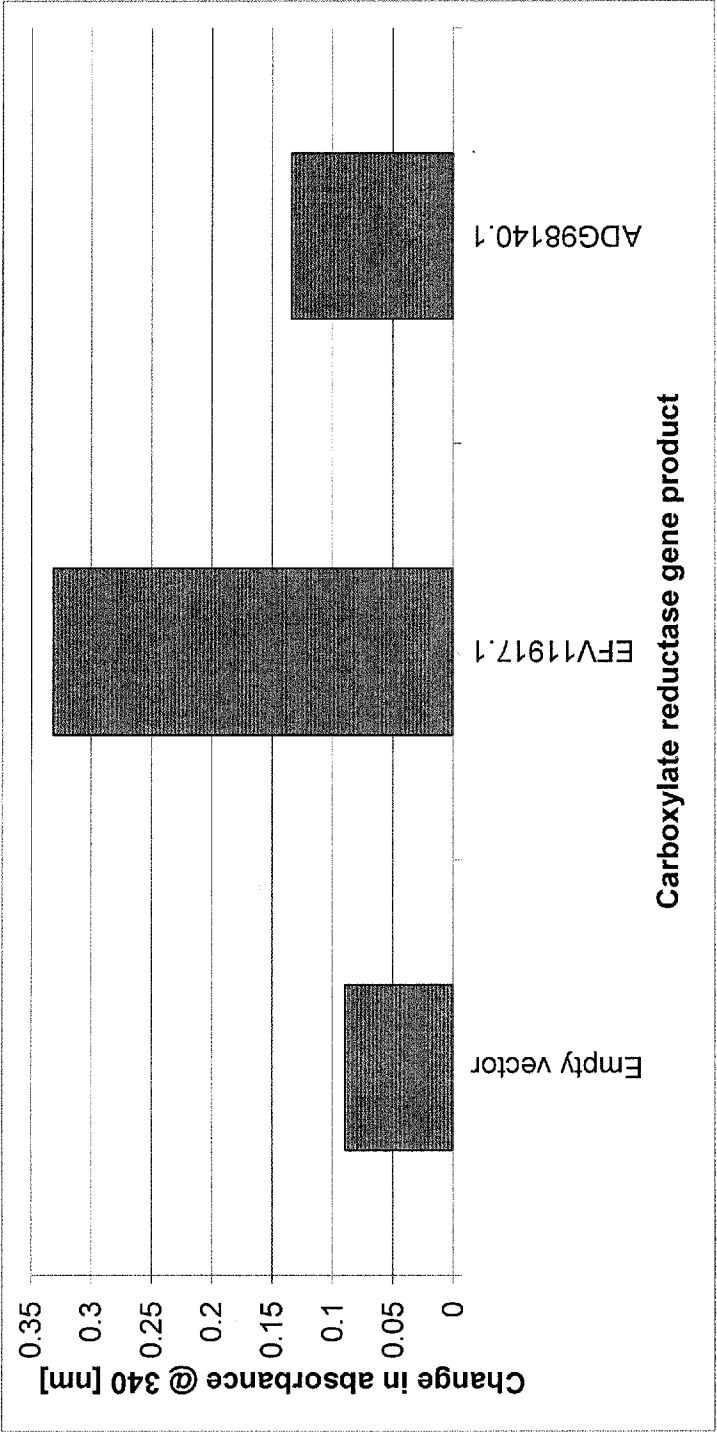


FIG. 8

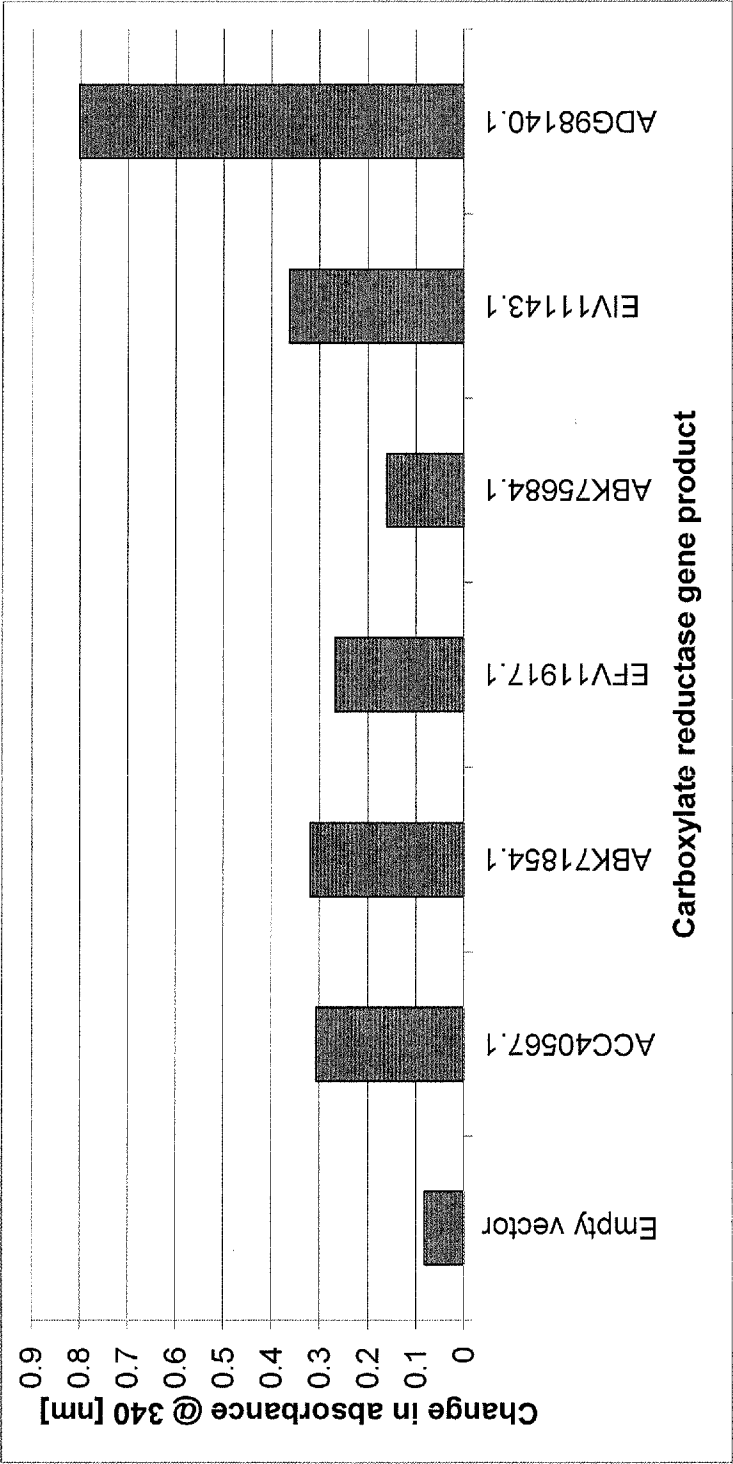


FIG. 9

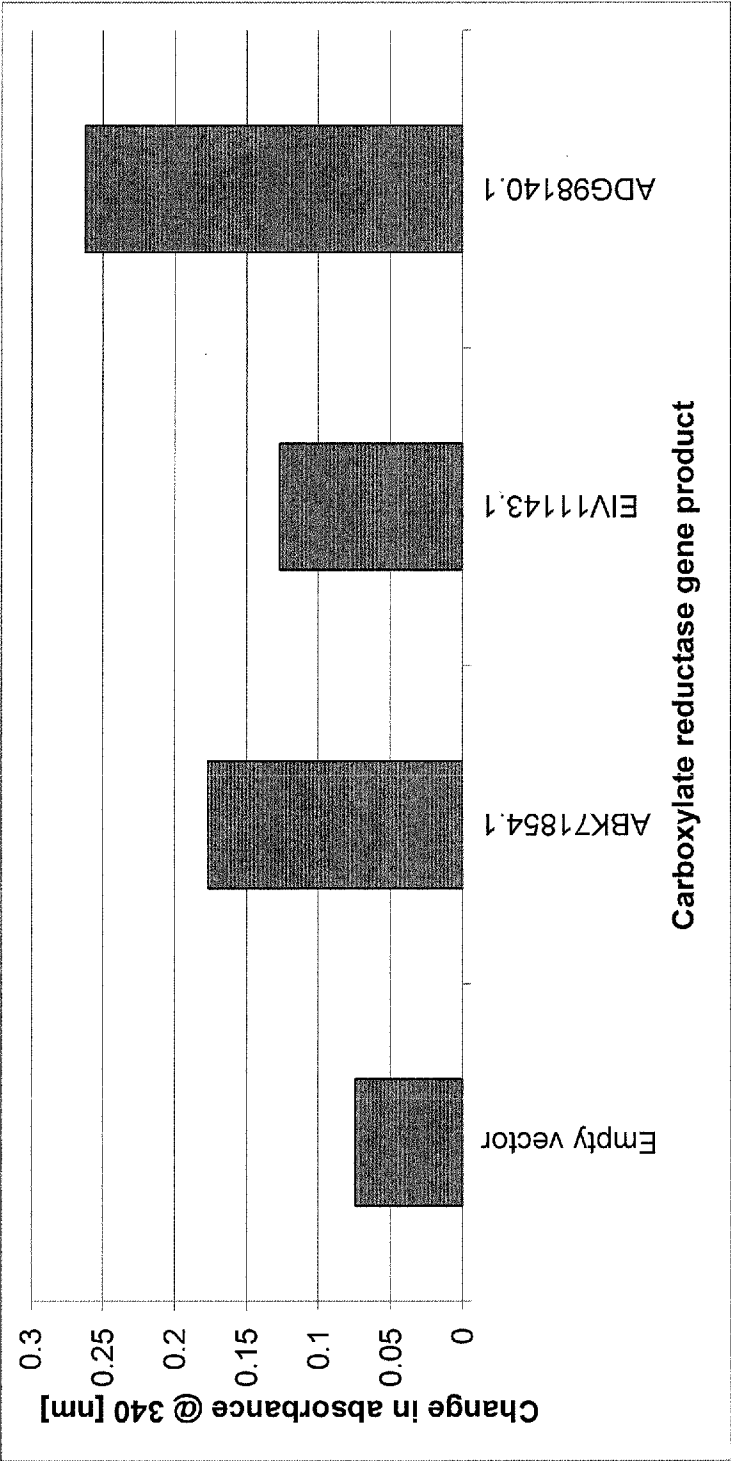


FIG. 10

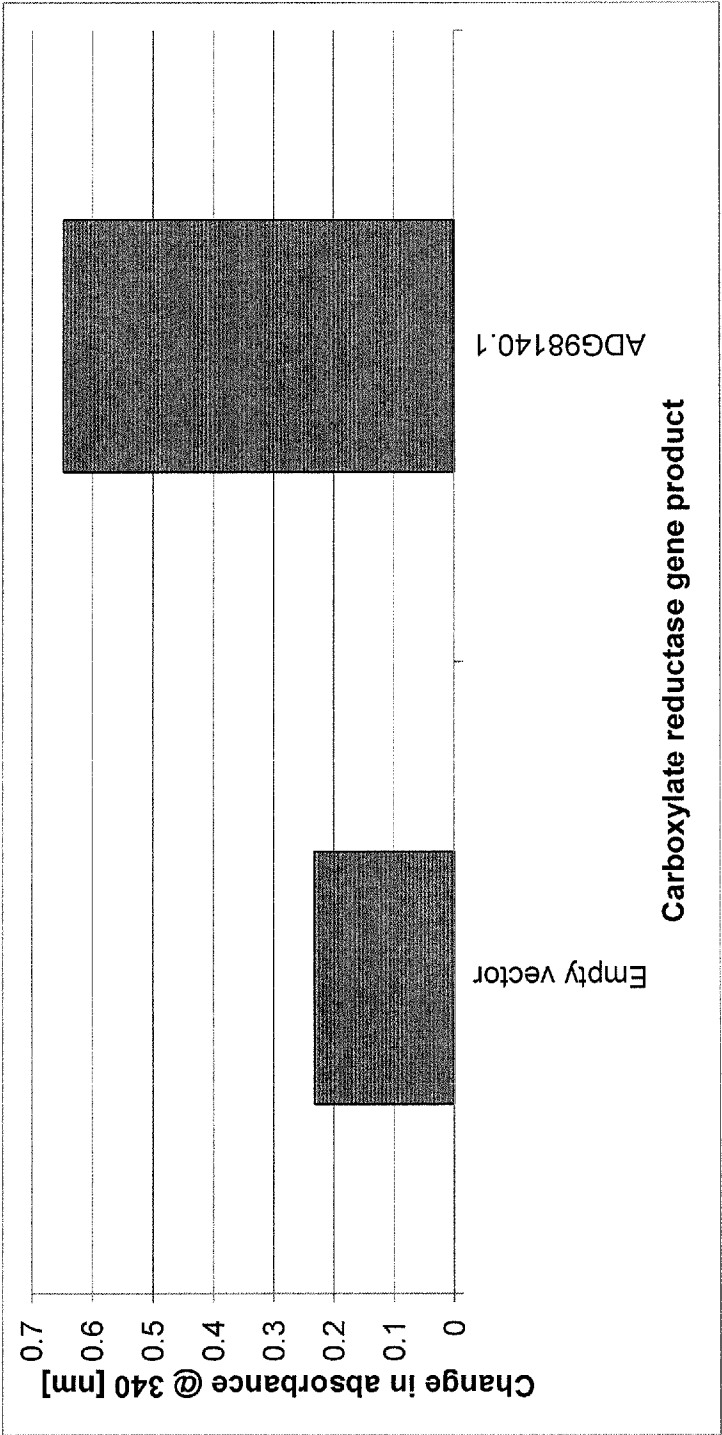


FIG. 11

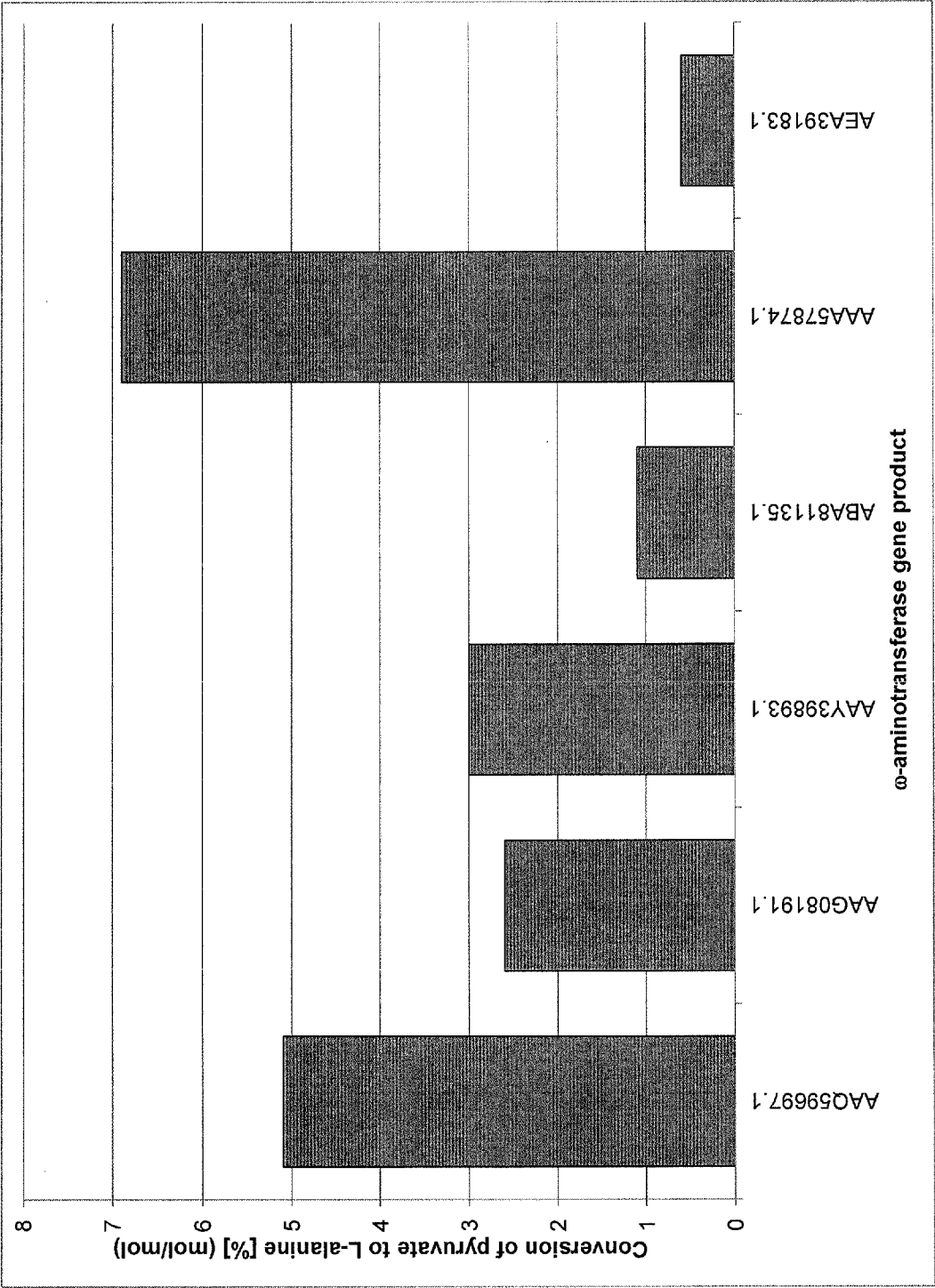


FIG. 12

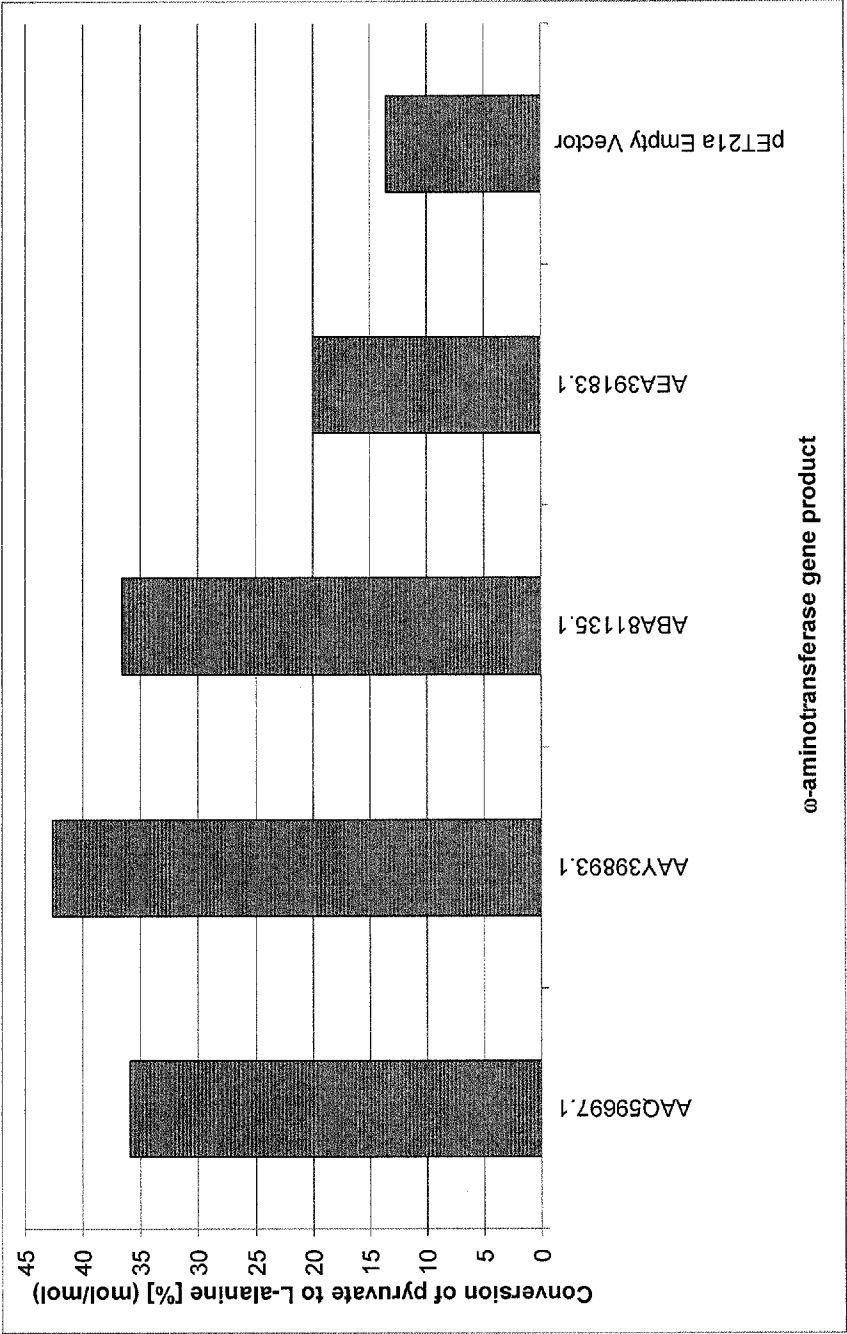


FIG. 13

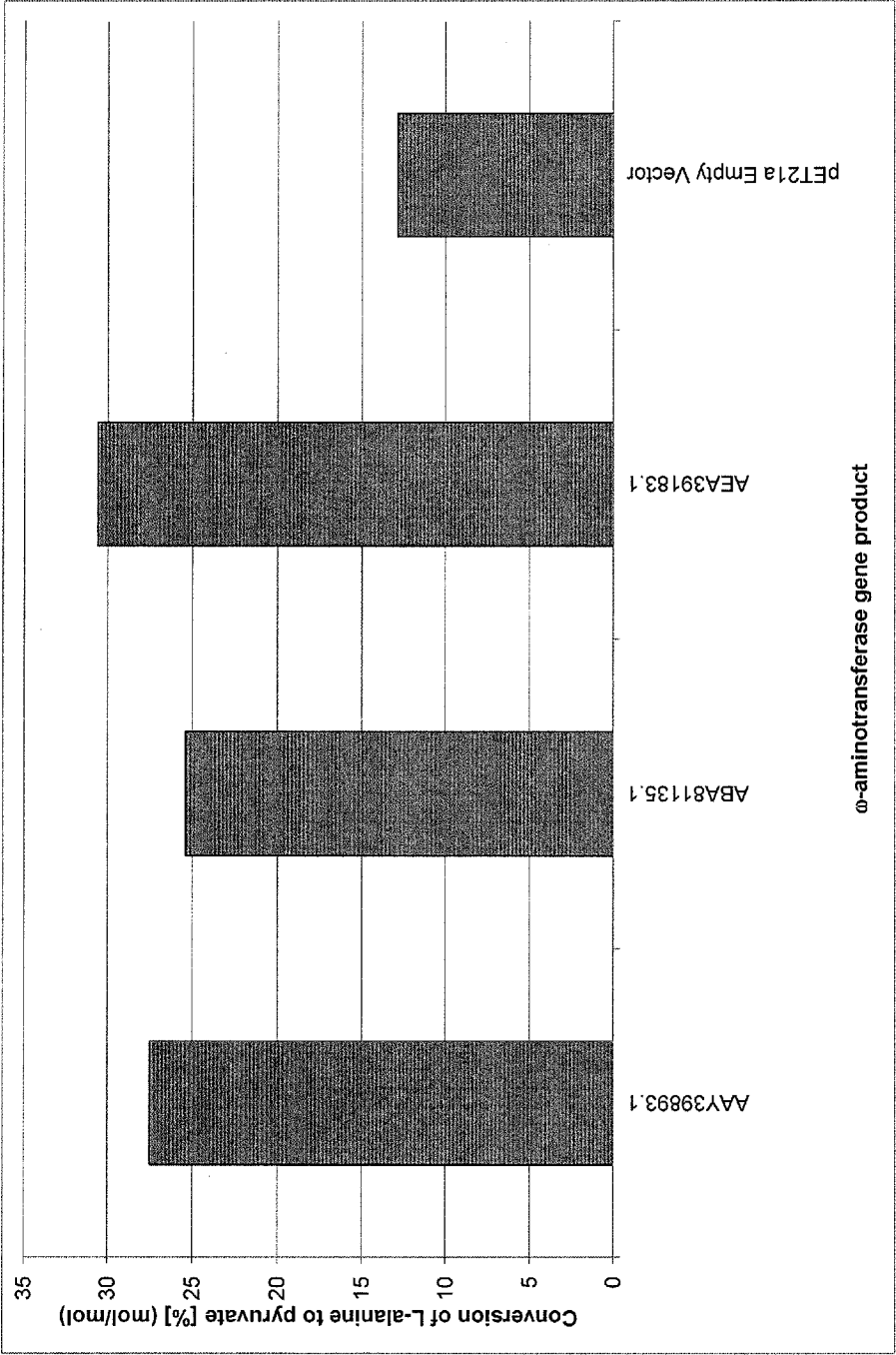


FIG. 14

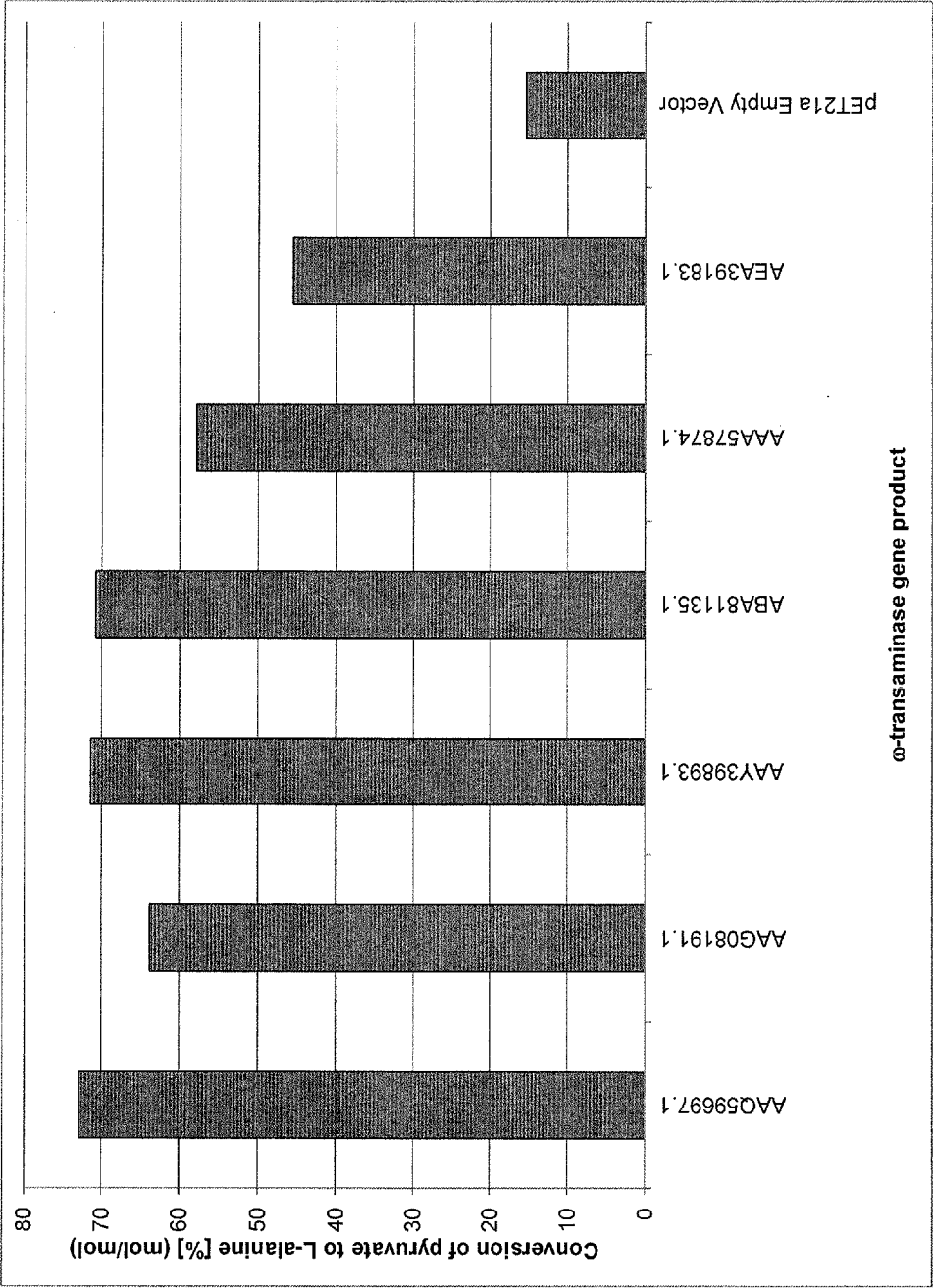


FIG. 15

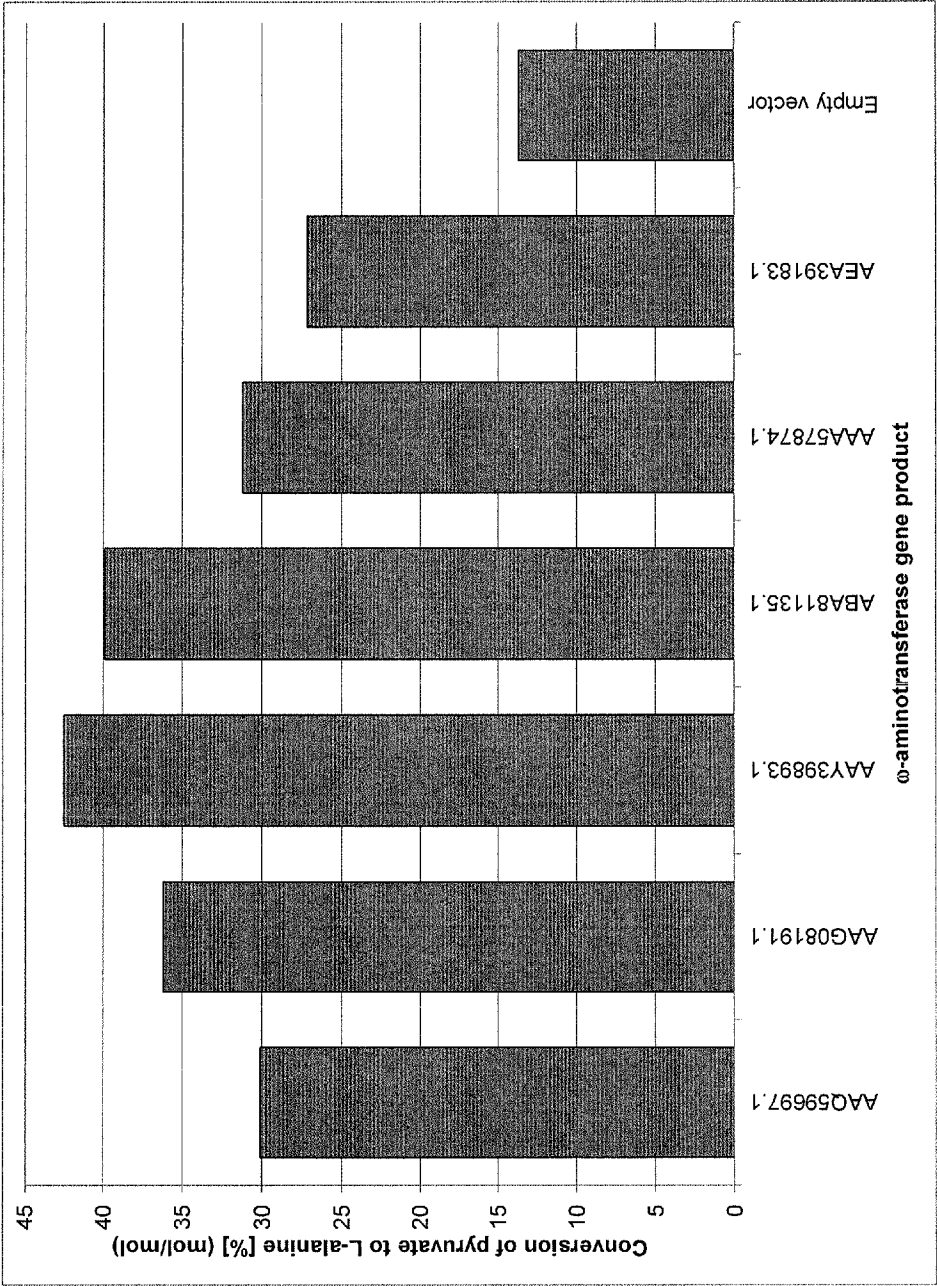


FIG. 16

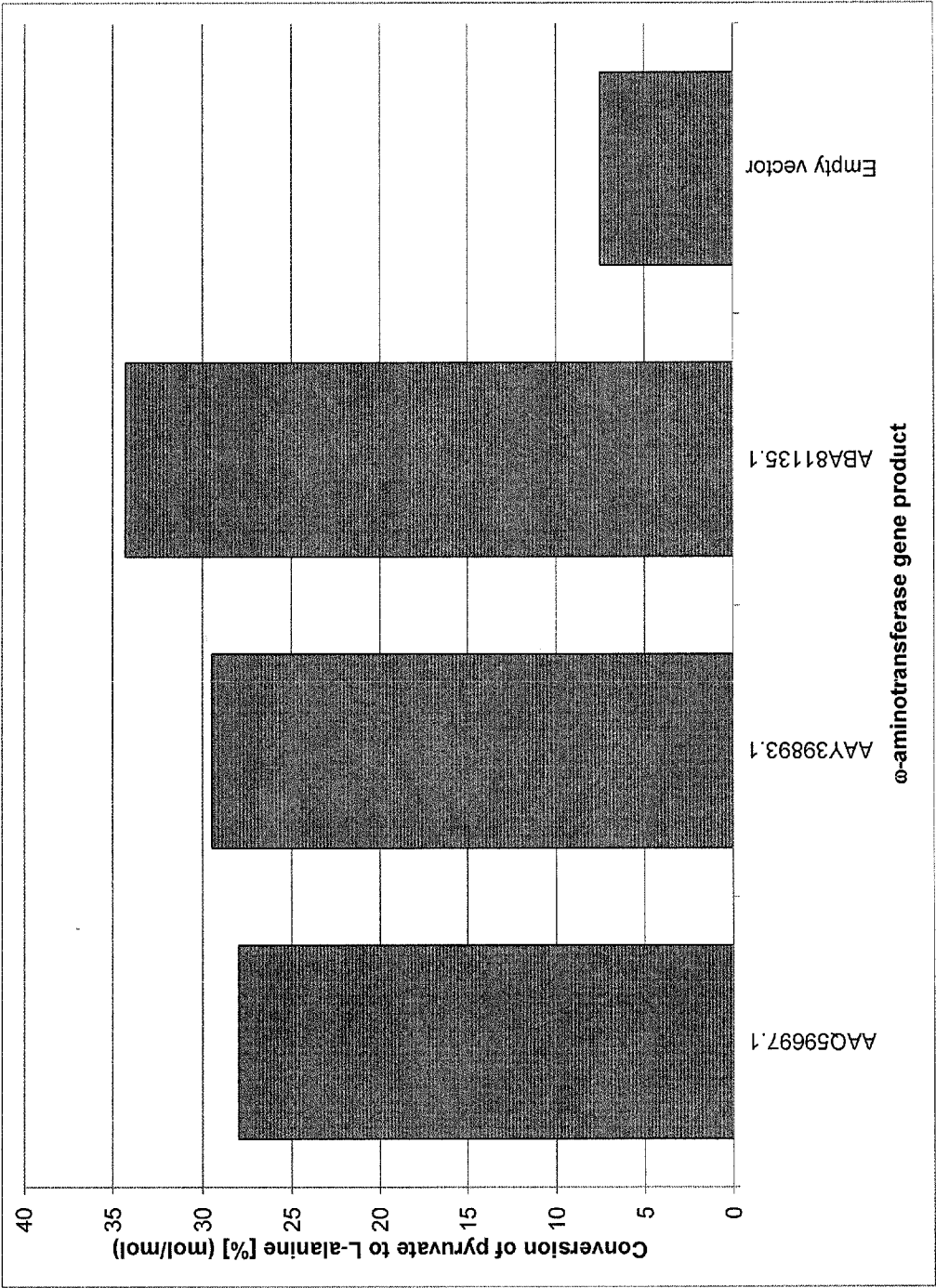


FIG. 17

FIG. 18

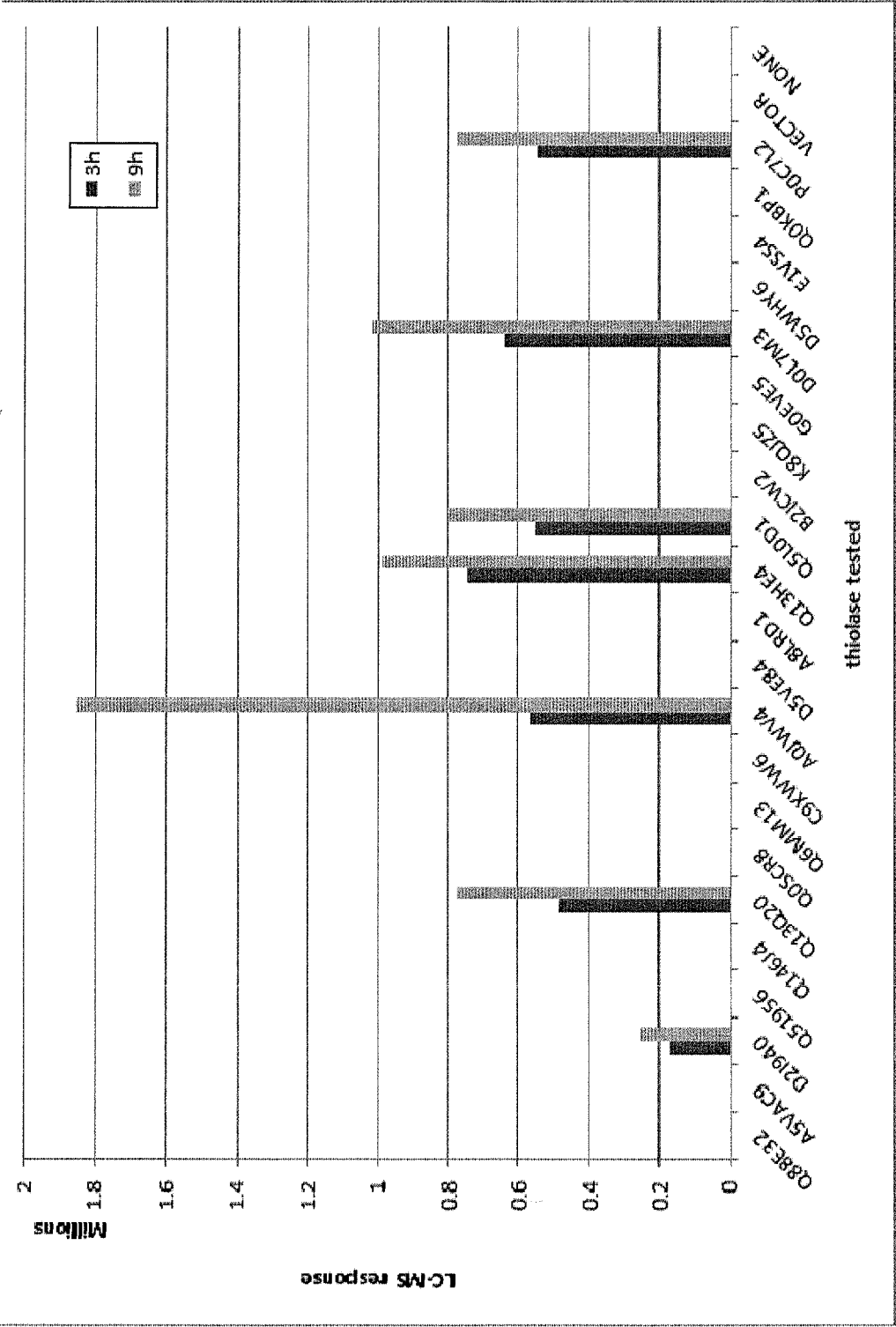


FIG. 19

SEQ ID NO	Nucleic acid sequence
38	ATG AAC GAT GTG GTC ATT GTG GCT GCA ACG CGC ACT GCA ATT GGC AGC TTT CAG GGA GCG TTG GCT ACG GTG CCA GCC GTA GAT CTG GGA GCT GCC GTC ATC AAA CAG CTC GGT GGC AAA CAG ACC GGC TTA GAC CCA GCA CAG GTA GAT GAG GTA ATC CTG CCG TTT TCG GTT CCG GCA TTA ACC CTG AAC AAA GTG TGT GGC TCT GGT CTG GCA ATC AAA GCT GGA CTG CCG TTT TCG GTT CCG GCA TTA ACC CTG AAC AAA GTG TGT GGC TCT GGT GGC CAG AAA GCA CTG CAT CTG GCA GCA CAA GCG ATT CCG CCT AGC GCG GGT ACC GGC CAA CCG CAT GGC CAG CAG GAG AAC ATG TCC TTG GCC CCT TAT GTG ATG CCG GGT TTA TGG GAT GCG TTC AAT GAC TAC CAC ATG GGC ATT ACG GCC CTC ATT GAT AGC ATG ATT ACC GAC GGT TTA TGG GAT GCG TTC AAT GAC TAC CAC ATG GGC ATT ACG GCC GAG AAC CTG GTG GAC AAG TAC GGC CTG TCT CCG GAA CAG CAG GAT GCT TTT GCG GCT GAA TCG CAG CGC AAA GCG GTC GCG GCT ATC GAA GCG GGT TTT GAC CCG GAG ATC ACG CCC ATT GTG TTT GCG CAG AAG AAA GGG GAA CCG AAA GTG TTC GCA CCG GAT GAA CAA CCG GGT CCG GAT ACC ACA GCC GAA TCG CTT GCC AAA TTA CGT CCT GCG TTC AAG AAA GAT GGC AGT GTA ACA GCC GAG AAC TCA AGC CTG AAT GAT GGC GCT GCC GCC GTT CTG CTG ATG AGC GCC GCT AAA GCG GAA CCG GTG GGC TTA CCC GTT TTG GCG AAA ATC GCC GCT TAT GCG TCA GCC GGT GTC GAT CCG GCG ATT ATG GGT ATC GGT CCA GTG TCC GCC ACT CAG CGT TGC TTG GAC AAA GCC GGT TGG CAG CTG GCG GAA CTT GAC CTG ATT GAA CCG AAC GAA GCC TTT GCG GCG CAA GCA CTG GCC GTT GGC AAT GCA CTT GCG TGG GAT GCG GCA CCG GTG AAC GTT AAT GGC GGT GCG ATT GCC CTT GGA CAT CCG ATT GCG GCT AGT GGT TGC CCG GTC CTG GTC ACC CTC CAG GAA ATG ATC AAA CGG GAC GTC AAG AAA GGC CTG GCG ACC CTG TGC ATT GGT GGT CAA GGC GGT GCG CTG GCC ATT GAA CGC taa
39	ATG GAG GAC ATT TAC ATT GTT GGC GCT GCC CGT ACG GCA ATT GCG GAC TTT GGA GGC GCG TTA AAG GAC GTT CCA CCA GCT GAT CTT GGC GTC ATT GTG GCA CGT GCT GCT GAA CCG GGT GGC GCG CTC GAA CCG GGT GAT GTT CAG AAC GTA ATG GGC CAG GTG ATG CCT ACC GAA CCG CGT GAT GCC TAC TTA GCT CGC ATG GTG GGT GTG ACT GCT GGT GTC CCG ATC GAA ACC CCA GCC CTC ACA CTG AAT CGC CTG TGT GGA AGC GGA GTT GAG GCA ATC GTT ACC GGC GCA AAA GCC ATG GTT CTG GGA GAA TCG GAT ATT GTC CTT GCG GGT GGC GCG GAA GTC ATG AGC CGT GTT CCT CAC GTC GTA AAA GGT GCG CGT TGG GGT ACC AAA ATG GGC AAT GTC GAG ATG ACC GAT GGT CTG ATC GAG GCG TTG TCC GAT CCG TTC GAC AAA GTG CAC ATG GGC ATT ACC GCG GAA AAC GTC GCC GAA CCG TAC CAG ATC ACT CGC GAA GCA CAG GAT GCT CTT GTC CAG GGT CAT CAA CGT GCG GCA CCG GCG ATC GCC GAG GGT CCG TTC AAA GCC CAG ATT GTC CCT GTG GAA GTG AAA ACG CGC AAA GGC GTT GTG GCG TTC GAT ACC GAC GAG CAC GTT CCG GGT GAT GTG TCT GCG GAA GAA CTG GCG AAA CTG CGT CCG GTC TTT AAG AAG GAT GGC ACC GTA ACA GCC GCG AAT GCG TCA GGC GAA GGT GGC GCG GCA GCA ATG GTG GTC TTG GCA ACG AAG AAA GCC GTC GAT GTC GAC GCG AAA GGG TTG AAA CCC TTA GCC CGC ATC TTG TCG TGG GGT CAT GCA GGG TTA ACT CTG GCG GAT ATC GAC GGT ATT GAG GCC AAT GAA GCT GTT CCG ATT GCG CTG GAA CGC GCA GGC TTA ACT CTG GCG GAT ATC GAC CCG GAT AAA GTG AAC CCG AAC GGC AGT GGC CCG CAA GCA TGC GCA GTT GCG GAG GAA CTC GGG TTT GAC CCG GAT AAA GTG AAC CCG AAC GGC AGT GGC CCG CAA CTG GGC CAT CCG GTA GGT GCG ACA GGT GCG ATC CTG ACC GTG AAA ACG GTG TAT GAG CTG GAA CCG ATT GGA GGT CCG TAT GGT CTG ATC ACG ATG TGC ATT GGC GGT CAA GGC ATT GCC ATG GTG GTG GAA CGT TGT GCG taa

SEQ ID NO	Nucleic acid sequence
40	ATG AAG AAC GCG TTG ATT GTC TCT CCT CTG CGT ACT CCG ATC GGG AAG TTT GGC GGT GCA CTG GCT CCG CTT ACC GCG GAA CAT CTC GCC AGT TTC ATG ATC TCG CAA GTG ATG GCG CGT ACC GGC GTG CCA GGC CAT TCG CTG GAT GAG GTG ATT GTG GCC CAG TCT TAT GCG AGC TCC GAA GCC CCG TGC ATT GGT CGC TAT GCG GCG TTG AGT GCC GGC TTA CCG GTG GAA GTA CCC GGA TAT ACC CTG GAT CGC CGC TGT TCA GGG CTG CAG GCT GTC ATT GAT GCC AGC ATG ATG GTC AAA ACC GGT AAT GCC GAA GCC GTT CTG GTG GTA GGG GTT GAA AGC ATG TCG AAT ATC GAG TAC TCA ACC GAT ATG CGC TGG GGT GCT CGC GCC GGT AGT GTC CGG TTT CAT GAT CCG TTG GAG CGT GGT GCG GAG CGC AGT CAA CCG AGC GAA CGC TTT GGC CAC ATT TCG GGG ATG CCA GAA ACG GCG GAC AAT CTG GCC CTC GAC TAT GGC ATC TCA CCG GAA GAG GCC GAT AGC TTC AGC GTT CGT AGC CAC CAG AAT GCA GCT GCG GCG TGG CGT GAG GGT GAT GAG GGT ATT CGC GAA GAT GCA GTG GAC GTA CCT GGT AAA CGT GGC GCT GTG ACA CGC GTC ACG ATT GAT GAG GGT TGC ACT GCG AAC AGT TCG CAG CTG GAG TCC ATG AAA GCT TTA CGC TTG ATC CGT CCG GAA GGC GTT TGC ACT GCG CAT GGC CTG ACT CCG CAG AAC GAT GCG GCG GCA GGC TGT GTT GTA TCC CCG GAA TAC GCA GCT CGC CAT GGC CTG ACT CCG ATG GCT CGT CTG GTC GAC TGG GCG GCA GCA GGC TGT GAA CCA TCC CGC ATG GGG ATT GGC CCC GTT CCT GCG ACC CAG AAA CTG CTG ATG CGT ACA GGG TTA TCT CTG GCA GAA CTG GGT CTG GAT GAC CTG TCC CGC GTT AAC GAA GCG TTT GCA GCC CAA CCG CTC GCG GTA CTG AAA ACG TGG GGT CTG GAT GAG CTG TCC CGC GTT AAC GTC AAC GGA TCA GGC ATT AGC TTA GGC CAT CCG ATC GGT GCA ACG GGT GTT CGC ATC ATG ACC ACC CTT CTG CAC GAA ATG CGT CGT GGT GAA GCC CGC TAT GGC CTG GAA ACG ATG TGC ATT GGA GGT GGC CAG GGA TTA GCG GCA CTT TTC GAA CGC GTG taa
41	ATG CGC GAT GTC TTT ATC TGC GAC GCC ATC CGT ACG CCG ATT GGT CGG TTT GGC GGT GCA CTG GCG GGT GTT CGC GCA GAT GAT CTG GCG GCC GTT CCG TTG AAA GCC CTT ATC GAA CCG AAT CCA GCC GTA CAG TGG GAT CAG GTT GAC GAG GTG TTC TGC TGT GCG AAC CAA GCG GGA GAG GAC AAT CGC AAC GTC GCG CGC ATG GCC CTG CCG ATC GCG ATC GGT ACT GCG TTT CCG TCT GCG GAA ATG GAA CTG GCG ATT GCA GGT GGC ATG GAT GCG ATG TCG CGT GCA CCC TTT GTG ATG GGC AAA CCG GAA TCC GGT TAT AGC CGC AAC ATG AAG CTT GAG GAT ACC ACT ATT GGT TGG CGC TTC ATC AAC CCG CTG ATG AAA TCC CAG TAT GGA GTA GAC TCC ATG CCC GAA ACT GCG GAC AAC GTT GCG GAT GAC TAC CAG GTT AGC CTT GCC GAA AAT CAA GAT GCG TTC GCA CTG CCG TCA CAA CAG AAA GCC GCA GCT GCA CAA GCG GGG TTC TTT GCC GAA ATC GTG CCG GTC CCG ATT GCT CAC AAG AAA GGC GAA ACC ATT GTG GAA CCG GAC GAA CAT CTG CCG CCG GAA ACC ACG TTA GAG CCG TTG ACC AAG CTC AAA CCG GTG AAT GGC CCG GAT AAA ACC GTT ACG GGT GGG AAT GCA AGC GGC GTG AAC GAT GGT GCT GCA GCC TTG ATC CTG GCG TCT GAT GAG GCC GTG AAG AAA CAC GGT CTG ACA CCA CGT GCG GTT GTA CTT GGC ATG GCC AGT GGC GGT GTG CCT CGC GTC ATG GGA ATT GGC CCG GTT CCT GCG GTG CCG AAA CTG ACA GAA CGC TTA GGC GTG GCA GTG AGC GAC TTT GAT GTG AAT GAA CTG AAC GAA GCG TTT GCG AGT CAG GGG TTG GCA GTA CTG CCG GAA CTG GGT GAT GAT GCC CCA CAA GTC AAC CCT AAT GGC GGT GCA ATT GCC CTG GGT CAT CCG CTC GGC ATG TCA GGG GCC CGT CTG TTA ACC GCG CTG CAT CAG CTG GAG AAA TCG GCG GGC CTG GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGA CTC GCG CTT GCT ATC GAA CGT GTC taa

SEQ ID NO	Nucleic acid sequence
42	ATG AGC GAA ACG CAC ATG AGC GGC ACC AAA GCC GAT CCC ATC GTG ATT GTC GGA GTG GCT CGC ACC CCT ATG GCA GCC TTT CAG GGC GAC TTT GCG ACC TTG TCG GCA CCG CAG TTA GGG TCA GTG GCC ATT CAA GCC GCT GTT CAG CGT GCC GGA CTG AAA CCG GAA CAG ATC GAT GAA GTC GTA ATG GGG TGT GTA CTG CCA GCT GGT CTG GGT CAG GCT CCT GCA CGG CAA GCT GCG CTT GGC ATC GAT GAA GTC GGA TTA CCG CTG GCC ACT GGC TCA ACG ACG GTC AAC AAA ATG TGC GGC TCT GGC ATG CCG ATG CCG ATG GCG ATG TTT GCG CAC GAC ATG CTC GCT GCC GGC TCA GTT GAC GAA ATA ATT GTA GCG GGC CAT GGG CAG GTG GAA TCC ATG ACA AAT GCG CCG TAT CTG CTT CCC AAG GCG CGT GCG GGA ATG CCG ATG CCG CTT ATG GGC GGC CAT GGC AGC TTT GCC GAG GAA TGT GCA GCG AGT TTC GAC TTC ACC CGT GAA TAC GAG AAA GGT GCT TTT GCG GTC GAA TCG CTG GCC GGT GCG AAA CGT GCA AAT GAA GAT GGC TCG TTT GCC GCG CAA GAT GCT TTT GCG CCA GTT AAA GTC GAA TCT CGC AAA GGT GAA GTG ACG ATC GAT CCG GAT GAG CAG CCG TTT AAA GCG AAC ATC GAG AAA ATC CCG ACC TTA AAG CCG GCG TTC AGC AAA ACG GGG ACT GTT ACG GCG GCC AAC AGC AGT TCC ATT TCC GAT GGT GCA GCT GCG CTG GTT ATG ATG CGC GAA AGC ACT GCC AAA CGC CTC GGC GTT CAG CCG ATT GCG GCG GTT GTC GGC CAT TCT ACC CTG GCA CAA GAA CCG GCG AAA TTT ACC ACC GCC CCA GTT GGT GCT ATC CGC AAG CTG TTC GAG AAG AAT GGC TGG CGT GCG GAT GAA GTC GAC CTG TTC GAA GTG AAC GAA CCG TTT CCG GTG ACA ATG GCG GCA ATG AAA GAA CAC CAT CTG CCT CAT GAG AAA GTG AAC GTC AAT GGT GGT GCC TGC GCA TTA GGC CAT CCG ATT GCG GCA AGT GGA GCA CGC ATC TTG GTA ACC TTG ATT GGT GCC CTG AAG AAA CGC GGT GGT AAA CGT GGT GTG GCT ACT CTG TGC ATT GGC GGT GGC GAG GCA ACA GCG ATG GGG ATC GAG CTC GTG taa
43	ATG ACG GAA GCG TTC CTG CTG TGT GAT GCG ATC CGC ACG CCG ATT GGC CCG TAT GCG GGC GCT AGC TTC GTA CGC GCA GAT GAT CTG CTG GGT GCG GTA CCG TTG AAA GCC CTT ATG GAG CCG TAT GCG GGC GAT GTC GAC TGG AAT GCC ATC GAC GAC GTT ATC TAC GGG TGC GCA AAC CAA GCT GGA ACG ACA GTG AAT CCG AAC CGC AAC GAT GCC CGG ATG TCG CTC CTC TTA GCT TTA GCT GGC ATC GCA CGT CCG CAA GGT GTT CCG TCA AAG TCG AAA GCG GCT TTT TCC CGG CAG AGT GGC ATG GAC GCG GTT GGC ATT GGC CCG CCG TTT GTG ACG GGA AAA GCG ACC AGC GCG TTT TCC CGG CAG GCA GGG GTG GAA AGC ATG TCA CGT GCG CCG TTT GTG ACG GGA AAA GCG ACC AGC GCG TTT TCC CGG CAG GCA GAA ATC TAC GAT ACC ACC ATT GGT TGG CCG TTT GTC AAC CCG CTG ATG AAG AAA CTC TAT GGC GTC GAT AGC ATG CCG GAA ACC GGT GAG AAT GTT GCC ACC GAC TAC AAC CCG CTG GCG GCA GAT CAA GAC GCT TTC GCC TTA CGC TCT CAG CAG AAA CCG GCA CGT GCG CAG GAT GGC ACT CTG GCA CAG GAA ATC GTG GGT GTT ACC ATT GCC CAG AAG AAA GGC GAT CCA GTC ACC GTG TCG CAG GAC GAA CAT CCG CGC GAA ACC AGT CTT GAC GCC CTG GCT AAA CTG AAA GGT GTG CTG CCC GAT GCG ACC GTA ACT GCT GGC AAC GCT AGT GGG GTC AAC GAT GGG GCA CCG GCG TTG CTG CTG GCG AAT GAG GAA ACA GCC GTG ATG GCG ATT GGA CCA GCC CCC CGT GCA CGC GTG CTT GGT ATT GCC ACC GCA GGT GTA GCT CCG GTG CAG TTC GAT GTC ATC GAA CTG AAC CCT GCG ACG CAG AAA CTG TTA GCG CCG ATT GCG GTC CTG CGC GCG TTA GGA GTC GCT GAT GAC GAT ACC CGT GTG GAA GCG TTT GCG TCG CAA GGC ATT GCG GTC CTG CGC GCG TTA GGA GTC GCT GGT TCT TCG CGC CTG AAC AAT CCG AAT GGC GGT GCC ATT GCC CTC GGT CAT CCA CTG CGC GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GCG ATG TAT CAA CTG CAC CGC ACA CAA GGC CGC TTT GCT CTG TGC ACG ATG TGC ATC GGC GTC GGT GGT GGT GGC ATT GCG ATT GCC ATT GAA CGC GTA taa

SEQ ID NO	Nucleic acid sequence
44	ATG GCC GAA GTG TTC CTC GTG GAT GGT GCG CGT ACC CCA CAA GGT GGA GGT GCC CTG GCG GGT GTT CGC CCT GAC GAT CTG GCA GGC TTG GTG GTT GCG GAA GCC GCG CGT CGC GCA GGA ATT CCG GGA GAT GCG GTG GAC GAA GTG ATC CTT GGG GCA GCG AAT CAA GCA GGC GAG GAT AAC CGC GAT GTG GCT CGT ATG GCT GTA CTC CTG GCG GGC CTG CCC GAT AGC GTC CCA GGC TAT ACC GTC AAT CGC CTG TGT GCG TCG GCG CTT ACT GCT GTG GCG AGT GCC GCG CAC ACC AAT CGC AGC GGC GAA GCT GAC ATC GTC ATT GCG GGA GGG GTG GAA TCA ATG ACG CGT GCT TGG GTA ATG GGC AAA CCG GGT ACC CCA TGG GCA CGT CCG GGT GAA GTT GCG GAC ACG AGC TTG GGC TGG TGC CCG AAT CCC CGT TTT ACC GCT GAT CGC GAG GAA GTT GCC CGC GCC GGT GCT GGG CTT GAT GTG CGC AAA GTA ACC CTG TCG ATG GGC GAA ACT CGC GAG GAA GTT GCC GCG CTG GAA GGC GTA ACC CGT GCG GAA TCC GAC GCA TTT GCT CTC CGT TCT CAG GAA CGT GCG ATC GCA GCG GTT GAT GCC GGT CGG TTT GAA CCG GAG AAT GTT CCG GTG CCG GTT CGC GAT GGC GAA CTG GCA GCG GAC GAG ACA CCG CGT CGT GGT ACG ACA CTG GAG AAA TTA GGG AGT CTG AAA CCC GTG TTT CGT ACG GCG GGT ATT GTC ACT GCA GGG TCC AGC AGT TCT CTG TCT GAT GGC GCG GCA GGC TTA GTT GTC GCC AGC GAA GCA GCG GTG GAA AAG TAC GGC CTG ACC GTG CCA GCG ACT CGC AAA GCG ATT GTC ACG TCG GCG GGT GCA CCG AAC GTG ATG GGT CTT GGT GCG GAG TTG AAC GCG TTC GCA CAG AGT CTG GCG GTT ATC CGC CAG TTA AAG CTG GAC GCT GCG ATC GTA AAT GCT GAC GGT GGG GCC AAT GCT CTC GGA CAT CCG TTG GCG TGC TCA GGT GCC CGC ATC TTA CTG ACC CTG TTA GGG CCG ATG GAA CCG GAA GGC GCC CGT CCG GGT CTT GCC ACA CTG TGC GTG GGT GTT GGC CAG GCG GTG GCC ATG CTG ATT GAG GCC CCG taa
45	ATG AAG TCA CCA CCG GAC GTT GTG CTC GTG GAA GGT GTT CGG ACA CCG TTC GCA AAA GCG GGT ACC AAA CTG AAA AAA GTC CAC CCA GCT GAA CTG GGC AAA GTG GCG CTT AAA CAG GTC ATT GCG CAA ACG AAC CTG GAT GTC AAC CTT GTG GAT GAA GTG ATC ATC GGG AAT ACT GGA AAC CCA CCA GAT TCG GTG AAC ATC TCT CGT GTT GTT GCC CTG AAT GCC GGT AAT CCG CTG CAC ACC TCA GCG TAC ACG GTA CAC CGC AAT TGC GCA TCT GCC CTT GAA AGC ATT AGT AAC GGC TAT GAG AAA ATC AAA TCC GGC ACG ATG GAT GTT ATT CTG GCG GGT GGT ACC GAG AAT ATG TCA CAA ATG CCG ACT TTA CCG CCG AAG AAA TTC CAG GAA ATC TAT GAG AAA CTG TTT GCT GCA AAA GGC CCG AAA CAA GCT CTG CCC CTG CTG TGG AGT CTG TTC AAA CCG GAT GTG AAG CAG ATC AAA GCC CTG TTA AGT GGG AAT ATG CGT GAC GAG TAC TTC CCG GTG ATT TCG GTG ATG ATG GCG TTA ACC GAT CCC TTT GTG GGC ATT AAC CTG GGC CAG ACT GCC GAA ATC TTG GCG AAA GAA TGG GCG CTG TCT CGC GAA ACC CAG GAT AAG TTC GCC CTG CGT TCC CAT CAG CTG GCG AGC AAA GCC ATG AAA GAG GGT CGT ATG CGG GAA GAA AAT GCA CCG GTG TAT CTC GCC GAA TAC AAG GAA GTT ATT AGC GAG GAT ATC GGT CCT CGT GAT ACC CAG ACG ATG GAA GCG CTG GCC AAA CCG TTT GAT GAC AAA GCG ACA GGC TCT ATT ACC GCA GGC AAC AGC TGT CCC ATT ACC GAT GGA GCG ACT ATG TTT CTG ATG TCG CGC GAG AAA GCA AAG GCG CTG GGC TAT AAG CCT TTG GCA ACC ATT CCG TCC TAT GGG TTT GCG GGT GAA CCG GAA CGC ATG GGT TTA GGT CCT GTA TAC TCC ACA CCT GTA CTT CTT CAA GTC TTA TCG TGC CAG AAA GCG TTT GAT GAC ATT GGG TTG GTC GAA TTG AAT GAA GCG TTT GCT GCT CAA GTC TTA TCG TGC CAG AAA GCG TTT GAT AGC GAC AAG TTT GGC CAG GAA AAG CTT GGT CTG AGT AGC AAA ATC GGC GAC GAC ATT CTG AAC GTG AAT GGC GGT GCT ATT GCA CTG GGT CAT CCG GTA GGT GCC ACC GGT ACG CGC GAT GGT CTG ACG CTG GCG AAA GAG ATG AAA CGT GGT AAC ACC CAG TTT GGG TTG GCC ACT CTG TGT ATT GCG GGA GCG CAA GGA GGT AGC ATG ATT CTC GAA AAT GAG GGG taa

SEQ ID NO	Nucleic acid sequence
46	ATG TTT TCG CTT CTG CAG GGG AAC GTC ATG ATG TCT CAA GCT CTG CCG CTT GTC ACG CGT CAG GGC GAT CGC ATT GCG ATC GTT TCC GGC AAA ATG GTA GTT GGG GAG CTG TTT GCC CGC CAA GCT ACC GCG TAT CAC GGG GTT CCT GCC GTG GAT CTG GGC AAA ATG GTA GTT GGG GAG CTG TTT GCC CGC CAA GCT ACC GCG TAT CAC GGG GTT CCT GCC CAG CTG GTA TTC GGT CAG GTA ACA GAT GCG TAT GGC AAG AAA CTG CGC TTA CGC GAT ACG TTA CGC TTA CGC TTA CGC ACG GGT ATG TCC GTT CAC ACA GAT GCG TAT GGC AAG AAA CTG CGC TTA CGC TTA CGC TTA CGC TTA CGC TTA CGC GCC AAT GTG GCG GAA AGT CTG ATG AGC AAG AAA CTG CGC TTA CGC TTA CGC TTA CGC TTA CGC TTA CGC TTA CGC AGT GTG CTG CCG ATT GGT GTC AGC AAG AAA CTG CGC TTA CGC TTA CGC TTA CGC TTA CGC TTA CGC TTA CGC GCC GGC CAG CCG CTG AAA GGT TTA CGC ATG GGA GAT ACG GGC CAA CTC GAA CAG ATG GCG AAA ACG CAT GGC ATC ACG GCT GAG TAC TCA ACG GAT GCA CTT GCG CAT CGT TCT CAC CCG TAT CCG TAT CCG TAT CCG TAT CCG TAT CCG TAT CTC GAA CAG GAA GTT ATG ACC GCT TAC ACC CCA CCG TAT CCG TAT CCG TAT CCG TAT CCG TAT CCG TAT CCG TAT CGG AAG AAT TCG AGC TTG GCG GAC TAC ACC AAA CTG CGT CCG GCA GTG ATC CTG ATG ACA GAA TCA CGC GCC ACA GCT GCA AAC TCC ACC CCT CTT ACG GAT GGC GCG TCG TCG TCG TCG TCG TCG TCG TCG TCG TCG TCG TCG TCG CGC GAA CTG GGG TTG ACC CCG TTG GGT TAT CTG CGC TCG TCG TCG TCG TCG TCG TCG TCG TCG TCG TCG TCG TCG GAC ATG TTG TTA GGT CCC GCG TGG AGT ACT CCC CTC GCG CTG GAA CCG GCG GGT TTA ACC ATG GCC GAT CTG ACG CTG ATC GAC ATG CAT GAA GCG TTT GCC TCT CAG ACC CTT GCG AAC CTC AAA CTG CTC GCG TCA GAT CGC TTT GCA CGC GAA GTG TTA GGT CGT TCC CAA GCT ACC GGC GAA GTA GAC GAG AGC AAG TTC AAT GTG TTG GGT GGC AGC ATT GCG TAT GGC CAT CCG TTT GCG GCG ACT GGT GCG GCG ATG ATT ACC CAG ACT CTG AAC GAA CTG CGT CGT CGC GGT GGA GGC TTT GGC TTA GGT GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCA GCC ATG GTG CTG GAA GCC GAA taa

SEQ ID NO	Nucleic acid sequence
47	ATG TCC TTT AAC GGG CAA TCC GCA ACC GGT CCG GAT GAA AGT GCA GCG GCG CCT GCT GCT GCC ACA CCT GGT GCG GGT CTG TTG CGC AAA GCG GTA GTT GTG GGC GCG GAT GAA AAC CGC ATC CCA TTT GCC CGT ACG GGT GGT GCT TAT ACG AAG AGC AGC AAT CAG GAC ATG CTG ACA GCG GGC CTG ATT GCT CGC TTT GGC CTG GCG GAT GAA GCG ATT GGG GAA GTG GCG GCT TCT GCG GGT GCG GAT TCA CGC GAC TTC AAC TTA ACG CGG GAA GCC GTG TTA GGC AGT GCC CTC CTG GCC AAC AAA ATC AAA TTA GGC CAG GAT TCC GCC ATT GCC ACA GGA CTG GAA ACG GTA CTC GGT GAC GCG CAA CCC CGT ACC GGA GAG TAC AAC AGC CAT CAG GCC CTG ACA ACT GGT GGA GTT GAT TCT GCG AAA ACC CTG CCG GAA CCC CGT ACC GGA GAG TTG GCA TAC AAC AGC CAT CGC AAC CTC GCG GCT CTG AAT CGC GCG AAA AAT ACG GGA GAA CCC CGT ACC GCG CAG GAC TTG ACG CCG TAT CGT GGT TTT GGC AAT CTG GCG GAA CCA GAT GCG CCG AAT ACG GGA GAA CCC CGT ACC GCG CAG GAC TTG ACG CCG TAT CGT GGT TTT GGC AAT CTG GCG AGC GAG GCA CAG TGG AAG ATT ACC CGC GAA GAA TTT GAC GAC CTC TTG ACG CCG TAT CGT GGT TTT GGC AAT CTG GCG AGC GAG GCG GCG GAT ACC ACG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG ACT ATG ACC GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GAA TGG GCT GAT GCG CAC GAA TTA CCG AAA CTG GCC ACT GTG GCG GAT GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG TTT GTC CAT GCG AAA GAC GGT CTG TTA ATG GCG CCA GCG TTC GCG GTA CCT CGG CTG TTG GCA CGC AAT GGC CTT ACT CTC GAC GAT ATC GAT TTC TTA GAG ATT CAT GAG GCC TTT GCT GGC ACC GTG TTG AGT ACC CTT GCA GCT TGG GAA GAT GAA GAG TTC GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GAT CGT GCA AAA CTG AAT GTG AAT GGG TCG TCA TTG GCA GCT GGT GCG CCG TTT GCG GCG GCG GCG GCG GCG GCG GCG CGT ATC GTA GCC ACC CTG GCC AAA ATG CTG CAC GAT AAA GGG CAA GTT GAT GGT CGG CCG CCG GCG GCG GCG GCG GCG CTG ATC AGC ATC TGC GCA GCC GGC GGT CAA GGC GTC GTT GCC ATT CTG GAA GCT AGC

SEQ ID NO	Nucleic acid sequence
48	ATG GCC ACA GCG AGT TCC TCG GCG GGA TCC TCT TCC GGT GTC TGA GCT GCG GGT GTT CGC AGC CCA TTT GCC AAA GTA GAT GGT GCG TTA GCC GGT CAT GAC GCC ATT TGG GGT ACT GTG ATC CCG AAC CTC ACG TGG AGT AAC ATG TTA GCA CGT GCC AAA CCG GAC TTT GCG GTT TGG GGT GAT CCG ACG ATT CCG GCG TTT TCC ACG GTT ATG GCT CTG GCC CGT GAG GTC CTG TTG GAT GCC GGT GAT CCG ACG ATT CCG GCG TTT TCC ACG GTT ATG GCT TGC AGC ACC AGC ATG ATT GGC GCG ATT GAA GCA GCC GGC ATG GCG GAT GTG GAT GGC CCG GGT GAT CCG GCG CTG GTA GGC GGC GTG GAA AGC ATG TCA CGT GTG CAG TTG GCG GCT CTG GGA GGC CTT AAC CTG AAA GAT CGC AAG TTC CAG AAC GCG AAA ACC GGC CAA CAG CGG CTC GCG GCT TCA ATG GGA GAA CAT ACC GAG ATT GTA CGC TTA TTC ATC CCG AAG TTC TCT CGT GCA GAC CAA GAT GCC ATT GCT CTG GCG GAA CAT ACC GAG ATT ACC GCG AAA GAG TGG AAT TTG TCT CGT GCA GAC CAA GAT GCC ATT CCA ATT CCA GCG GAA CAT ACC GAG ATT GTG AAA GGC TGG GAA AGC GGA TTC TTT GAC GAC CTC GTC GCG CCA GCG TTT GAC AAA ACC GAG GCG ATC CCG CGC AAA GAC ACA TCT CTG GAG AAG CTC GCA AAA CTG GCG CCA GCG TTT GAC AAA ACC GAG GCG AAA GGC ACC CTG ACC GCG GGT AAT TCC TCA CCC CTG ACC GAT GGT GCG GCT GGT TGG GTA GGT TCT GCA GCT GGC ATG GCG CGT TTA CCG GGT GAG ACA CCG AAG GTT CGT CTG GTC GAC TAC GAA GTG ACC AGC ATC GAT CTG CGT CAT GAA GGC CTG TTG ATG GCG CCG GCT TAT GGG GTA CCG CGC ATG CTC GCC CGC AAT GGG CTG ACT TAT GCC GAT GTT GGG CTG TGG GAA ATC CAC GAA GCC GCG GTG ACT ACG CCG GTG TGG AGC CAC ATT GCG GCA TGG GAA TCG GCC AAG TTC CTG AGC GAG AAA GCC GCG GTG ACT ACG CCG GTG TGG AGC CAC CCT CGC GAA CGC ATG AAC CCG AAT GGC GGC TCT CTG GCT CTT GGT CAC CCA TTT GGT GCA ACT GCG GCG CGC ATC ATT TCG CAG ACC GTG AAA GAA CTG GCA GCG CGC CCT AAA GGG GAA CCG GCG ATC GTG AGT ATC TGT GCG GAT GGT GGG CAA GGC ACG ATG ATG CTT CTG GAA TCA GCC taa
49	ATG ACG GAA GCC TAC ATC TAT GAC GCG ATT CGC TCA CCA CGT GGT AAA GGG CGC AAA GAT GGC TCC CTG CAT GAA GTC ACT GCG GTC AGC CTG AGT GCC CAA ACC CTG AAT GCG ATT AAA GAC CGC AAT GGC CTG ACC GGG CAT GCG GTT GAA GAT GTG ATT TGG GGT AAC GTC ACC CAG GTA ATG GAA CAG GGT GGA TGT CTG GCA CGT ACG GCG GTA CTT GCG TCT GAT CTG GAT GAG AGC AAT CAA GTG CGT GGA GGC TTA GCG ATC AAC CGC AGC GGT TTA GAA GCC GTC AAT TTG GCG GCC AAT CAA GTG CGT GGT AGC GAT GGC GCG GCA ATC GCA CGC GGC GGT GAT GAG ATG GGT GTC GGT CCC ATG ATG GGT TCC GCG GAT ATC ATC GCA ACG GAA TAT GGC ATT GTG GCG ATG AAA ACC TAC TTT GTC CCG CAA GGT GTC GCG TCG CAA CGT CGG GCT AAA GCG TGG GAT GAG AGC CGC GAT GAT GCG GAT GCC CTG GCT GTA GCG CTG AAG CAT CAG AAC GGT CTG AAC ATT CTG GAC GAG CGC TTC AAT GGG TCC GTG TTT ACC GTT CGC GAT CAG AAC CCG GCT TTC AAA GAT ATG GGT GAA CAG ATG CGT CCT GAA ACT GAC ATG CAG AGC CTG GGT GCA CTG AAG TAT CCG CAT TTA GAG AAA ATC GAA CAC ATT ATG CCC GGT TTC GAC AAG ATT GCG CTC ATG AAG TAT CCG CAT TTA GAG AAA ATC GAA CAC ATT CAT CAC GCC GGT AAC TCG TCA GGC ATT GTT GAT GGG AGT GCA GCG CTC CTT ATC GGC AAT AAA GCC TTC GCG GAA GCT CAT GGC CTT AAA CCA CGT GCA GTG ATC AAA GCC ACA GCG AAG ATC GGA ACC GAT CCG ACG ATT ATG CTG ACT GGC CCG GTT CCA GCG ACC GAA AAG ATT TTG GCA GAT TCT GGC ATG AGT ATC TCA GAC ATT GAT CTG TTC GAA GTG AAC GAG GCC TTT TCG TCG GTG TTG CCG TTT ATG CAG GCG TTT GAT GTC GAC CAC GAC AAA GTG AAC GTG AAT GGC GGT GCG ATT GCA ATG GCG CAT CCG TTG GGT GCA ACA GCG ATG ATT CTG GGC ACC CTG CTC GAT GAG CTG GAA CGT ACG GCG AAA GGG ACC TTA GCT ACA CTG TGC GTT GCT AGT GGA ATG GGC GCG GCT ACG ATC GAA CGC GTA taa

SEQ ID NO	Nucleic acid sequence
50	ATG ACT CGC GAT ACG CGC GAC GTG GTG ATC GTA GAT GCG GTT CGC ACA CCA ATT GGC AAA TTC CGT GGT GCC TTG GCT GGT GTA CCG GTG AAC GAT CAT GTG GTC TTT GGT GCA CTC GTC ATT GAG GAA CTG ATC CGC CGT GCC GGT GTC AAA CCG CAG GCG TCT GTG TTA GGA GTG CAC ATT GCC GGT GCG CCG GAA ACC ATC CCA GGC CTG ACC ATT GAC CGC AAG ATT GCC CGC ACC TCT GTG TTA GGA GTG CAC ATT GCC GGT GCG CCG GAA ACC ATC CCA GGC CTG ACC ATT GAC CGC AAG TGT GGT AGT GGT GGC GAA GAA GCA GCA TCT ATG TCG CGT GTC CCG ACC TCG CAA GGC GAG GCG GAT CTG GGC GAA GCA GTT GCC GGT GGT GGC GAA GAA GCA TCT ATG TCG CGT GTC CCG ACC TCG CAA GGC GAG GCG GAT CTG GGC GAA GCA TTC GGC TGG ATG GCT TCA GAG CGC TTA GAG CGC CAA CTT GAC GCG TAC TCG GTA GAG TCT CAT CGC GCG GCG GCT TGT TGG GCG CTG ACA CCG TGG TTT GCG GCG CCG GAT GGT GCG CCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCT GAA GGC TGG TTT GCG GCG TTA TTT GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG ACC TTG AAA TCC TCG TTT CGC AGT GAT GGT GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCA GCA CTT CTG CTG ATG AGC AGT GAT CCG ACC CTA GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG CGT GCG GTT ACC ACG ACG GTG GGT GAT CCG ACC CTA GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GTA CTG GAG AAA GCG GCG TTA GGT TTG AGC GAT ATC GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GTA CCG CTG GTC TGG ATG AAG GAA TTC GGT GGT CCC CAC GCA AAA CTG AAC GTG AAT GGC GGC GCG ATT GCA CTG GCG CAT CCG TTA GGC GCA TCC GCG GCA CGT ATC ATG ACG AGC ATG TTG CAC GAA CTG GAA CGC CGG GGA GCA CGT TAT GGC CTG CAA GCC ATT TGC TGC GCT GGA GGT ATG GCG ACT GCC ACG CTG ATT GAA CGC TTA GAT taa
51	ATG CGT GAA GCT GTG ATT GTA GAA GCC GTA CGC ACA CCT GTC GCG AAA CGG AAT GGC GTA TTC CGC GAT GTG CAT CCC GTA CAT CTG GCA GCG GTG TGT GTG ACC CCT GTG GCA GAG CAA GGT TAC AAC ATT GGC CGT TTA GCG GTT GAG GAC ATC GTG ATT GCG TTT GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCA GCT CTG GAG GCC GCG TTT GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG CAG CAG GCC ATC CAT TTT GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GTG GAA TCG ATG ACC AAA GTC CAC CAA GGT GTA TCT GCG GAA CGC ATT GCG AAG AAA TAC GGA CTG ACA CGC GAA AAG TAC GAG TTC ATC CAC CAA GGT GTA TCT GCG GAA CGC ATT GCG AAG AAA TAC GGA CTG ACA CGC GAA GAG TTG GAC GCG TAT GCG TAT GAA AGC CAC CAG CGT GCG CTT GCT GCA CTG CGT GAA GGC AAA TTT CGG GCG GAA ATC GTT CCG GTA AAA GCG CTT GAT GGT GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GGT CCG CGT GAT ACT AGC CCG GAA GCC TTA GCA GCG CTG AAA CCG GTG TTT CAG GAA GAT GGC CTG ATT ACC GCG GGT AAT GCT TCC CAG ATG TCT GAT GGC GCA GCC GCG GCG GCG GCG GCG GCG GCG GCG GCA CGT CGC TTT GCG CTG AAA CCT AAA GCT CGC ATT GTC GCG CAA ACC GTT GTG GGA TCA GAC CCG ACC TAT ATG CTG GAT GGT GTT ATT CCG GCA ACT CGT CAG GTC CTG AAG AAG GCC GCG GCG GCG GCG GCG GCG ATT GAC CTG ATC GAA ATC AAC GAG GCG TTT GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG CCG CTG GAG AAA GTG AAC GTG AAT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT AAA CTG ATG ACG TCC TTG GTT CAC GAG TTA GAA CCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG ATT GCG CAT GGT ATG GCG ACG GCT ACG ATC ATT GAA CGC GCG GCG GCG GCG GCG GCG GCG GCG GCG

SEQ ID NO	Nucleic acid sequence
52	ATG ACC AAA GTG GTC ATT GCA GGC TAC ATT CGC TCA CCG TTT ACC CTG GCG AAG AAA GGG GAA CTG GCC ACG GTG CGT CCT GAC GAT TTG GCA GCT CAA GGT TGT GCG TTT CCA GAA GGC TTT AAC GGC ATT CCG GCC GAG GAC ATC GAA GAT CTG CTT CTG GGT CTG CCG CTG CCG GCA GGA GGC ATT CAG ATG AAG AAC GGC TTT TGT GGC TCT CTG GTT TCG TTT CTG GCG GGT CTG CCG GGT CTG CCG GGT CTG CCG GGT CTG CCG GGT CTG CCG GGT CTG CCG TCG ATG ACT ACC GTT CAC ATG GCA GGA GGC ATT CAG ATG AAG AAC GGC TTT TGT GGC GGT GGT GTC GAA AGC ATG AGC GCG GTA CCG ATG GGT GAT ACG GCT GGT GCT GCT GCT GCT GCT GCT GCT GCT ACA ATG CCT GGC GGC TAT ATG GGC ATG GGT GAT ACG GCT GGT GCT GCT GCT GCT GCT GCT GCT GCT CGG AAA GAA CAG GAA GAA TTC GGC GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT TTG ACC GGT GAA ATT ATC CCG ATT ACC GGC GGC AAA CCG GCG TTT CCA GCT AAT GGA GTC GTA ACC GCG GGT GAT ACC ACA CTT GAG GGT CTG GCA GAA CTC AAA CCG GCG TTT CCA GCT AAT GGA GTC GTA ACC GCG GGT ACT TCC TCT CCG CTG GCG GGT TTA GCG AGC GAT GGC GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT CAC TTA GAC GGT TTA GCG AGC GAT GGC GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT CCA GTT GCG GCG TCA CCG AAG GCG TTA GCG GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GAA TTG AAC GAA GCT TTT GCG AGT CAG TCC ATT GCG TGT ATG CGC GAA CTG AAC CTG AGT CCG GAC CGC GTT AAC ATC GAT GGT GGC GCG ATT GCA CTG GGC CAT CCA CTC GCG ACA GCG GCG GCG GCG GCG GCG GCG AAA GCA GCA TCG CTG CTC AAA CCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG CAA GGT ATC GCG ACC GTT CTT GAG GCA TTT taa
53	ATG GAA CAG GTC GTG ATC GTC GAT GCC ATT CGC ACG CCA ATG GGC CGT TCA AAA GGT GGA GCC TTT CGC AAT GTG CGC GCG GAA GAT CTG TCT GCA CAT CTG ATG CGC TCT CTG CTG GCG TTT CCA GGT GGA TTA GAT CCG ACA GCG CTG GAC GAC ATC TAT TGG GGC TGT GTG CAG CAG ACA CTG GAA CAA GGG TTC AAC ATT CCG CGT AAT GCT GCT CTG CTG CCG GAA ATC CCT CAC AGT GTC CCA GCA GTA ACG GTT AAT CGC TTG TGT GGC TCC AGC ATG CAA GCT CTC CAT GAT GCT GCA CGC ATG ATG ACC GCG GAT GCG CAA GCG TGC CTC ATT GGT GGT GTA GAA CAC CAC ATG GGC CAT GTC CCC ATG TCA CAC GCG GTA GAC TTT CAC CCG GGT ATG AGC CGC AAT GTG GCG AAA GCA GCG GGT TTT GCA GCG AGT CAT GCG GAG ATG TCG GCG GCG GCG GCG GCG GCG GCG CGT GAA ATG CAG GAT GCC TTT GCA GCG AGT CAT GCG GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT TTC AAG AAC GAG ATT ATT CCG ACT GGC GGC CAT GAT GCG GGT GGT GGT GGT GGT GGT GGT GGT GGT GAG GTG ATT CCG CCG GAA ACC ACC GGT GAA GCC TTT TCG ACC TTA CCG GCG GCG GCG GCG GCG GCG GCG GGT ACC GTT ACC GCG GCG AGC TCG TCC GCG TTT AGC GAC GGT GCG GCG GCG GCG GCG GCG GCG GCG GCG AGC CGT GCT GGT GAA CTT GGC TTA ACA CCG GGT GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG CCG AGC ATT ATG GCG TAT GGA CCA CTC CCG GCT TCG AAA CTC GCG GGT GCG GCG GCG GCG GCG GCG GCG TCC GAC ATT GGG CTG TTT GAA ATG AAC GAA GCT TTT GCA GCG CAG ATT CTG CCG TGC ATC AAA GAC CTG GGT CTG ATG GAG CAG ATC GAC GAG AAA ATC AAC CTG AAT GCG GGT GCA ATC GCG CTT GCG CAT CCT CTG GGT TGC AGT GGT GCG CGC ATT TCG ACC ACG CTG TTA AAC CTG ATG GAA CCG AAA GAT GTC CAG TTC GGT CTT GCG ACC ATG TGC ATT GGA CTG GGT CAA GGC ATC GCA ACC GTG TTC GAG CCG GTT taa

39/42

Nucleic acid sequence

SEQ ID NO	Nucleic acid sequence
58	ATGCGTGAAGCCCTTATTTGTGACGGAATTCGTACGCCAAATGGTCGTACGGCGGGGCATATCAAGTG TTCGGGCTGATGATCTGGCTGCTATCCCTTTGCGGAACTGCTGGTGGGAAACCCGCGTCTCGATGCGGA GTGATCGATGATGATCTCGGCTGTCTAATCAGCGGGGAGAGATAACCGTAACGTAACCGCGGATG GCGACTTTACTGGCGGGGCTGCCGAGAGTGTTCGGGCACAACCATTAACCGCTTGTGTGGTCCGGGC TGGACGCACTGGGTTTGGCGACGGCGGATTAAGCGGGCGATGGCGATTTGCTGATCGCCGGTGGCGT GGAGTCAATGTACGGGCAACCGTTTGTATGGCAAGCGCAGTGCATTTCTCGTCAGGCTGAGATG TTCGATACCACTATTGGCTGGCGATTTGTGAACCCGCTCATGGCTCAGCAATTTGAACTGACAGCATGC CGGAAACGGCAGAGAAATGTAGCTGAACCTTAAATAATCTCAGAGAAAGATCAAGATAGTTTTCGGCTACG CAGTCAGCAACGTACGGCAAAAGCGCAATCCTCAGGCATTTCTGGCTGAGGAGATTTTCGGTTGTGTG AAAAACAAGAAAGGTGTGTAAACAGAAATACAACATGATGAGCATCTGCCCGGAAACGACGCTGGAAC AGTTACGTGGGTAAAGCACCATTTCTGCCAATGGGTGATTAACCGAGGCAATGCTTCGGGGTGAA TGACGGAGCCGCTGCGTTGATTAATGCCAGTGAACAGATGGCAGCAGCAAGGACTGACACCGCGGGCG CGTATCGTAGCCATGGCAACCGCGGGGTGAACCGCCTGATGGGGCTTGGTCCGGTGCCTGCAACTC GCCGGTGTGGAACCGCAGGGCTGAGTATTCACGATATGACGTGATGAACGAACGAGCGTTTCGC GGCCAGGCGTTGGGTGTAACCGCAATTTGGGCTGCCTGATGATGCCCCACATGTTAACCCCAACGGA GGCGCTATCGCCTTAGCCCATCCGTTGGGAATGAGTGTGCCCGCTGGCACTGGCTGCCAGCCATGAGC TGATCGGCGTAACGCTGTTACGCATTTGTCAACCATGTGCATCGGTGTCGGTCAGGCGATCGCCATGAT TCTGAGCGTGTGTA
59	ATGACGCGTGAAGTGTAGTGTAAAGCGGTGTCGTACCGCATCGGACCTTTGGCGGC AGCCTGAAGGATGTGGACCGGGGAGCTGGCGCACTGGTGGTGGCGAGGCGCTGGCG CGCGCGAGGTGTGGCGGACGATGTGGCCACGTGGTATCGGCAACGTGATCCAGACC GAGCCGCGGACATGATGTGGCGCGGTGGCGCGCTCAACGGCGGTGACGATCAAC GCCCCCGCGCTGACCGTGAACCGCCTGTGGGCTCGGGCTGAGGCGCATTTGTACGCGCC GCGCAGACCATCCTGTGGCGGATACCGACGTCCCATCGGCGCGCGGAAAGCATG AGCCGCGCACCGTACCTGGCGCGGACGCGCTGGGCGCACGATGGCGGACGCGCGC CTGGTCGACATGATGTGGTGGCTGCACGATCCCTTCATCGCATCCACATGGGCGTG ACCGCGAGAAATGTCCCAAGGAATACGACATCTCGCGCGGCGAGCAGGACGAGGCCGCG CTGGAATCGCACCGCGCGCTTCGGCAGCGATCAAGCGCGCTACTTCAAGGACAGATC GTCCCGTGGTGAGCAAGGCGCAAGGCGACGTGACCTTCGACACCGCAGACGACGTCG CGCCATGACGCCACCATCGACGACATGACCAAGCTCAGGCCGTCTTCGTCAGGAAAC GGCAGGTACGGCCGCAATGCTCGGGCTGAACGACCGCGCGCGGTGGTGATG ATGGAGCGCGCGAAGCGGCGCGCGCTGAAGCGCTGGCCCGCTGGTGTCTGTAC GGCCATGCCGCGTGAACCGAAGGCCATGGGCATCGGCCCGGTGCCGCGCACGAAGATC GCGCTGAGCGCGCGCTGCAGGTGTCGACCTGGACGTGATCGAAGCCCAACGAAAGCC TTTGCCGCGACAGGCGTGGCGCGTGACCAAGGCGCTCGGTCTGGACCCGGCCAAAGTTAAC CCGAACGGCTCGGGCATCTCGTGGGCCACCCGATCGCGCCACCGGTGCCCTGATCACG GTGAAGGCGCTGCATGAGCTGAACCGGTGAGGGCGCTACGCGCTGGTGACGATGTGC ATCGCGGGCGGCGAGGCGATTCGCCCATCTTCGAGCGTATCTGA

SEQ ID NO	Nucleic acid sequence
60	ATG GCG CAG TTT GTC ACT GCT CAG GAA GCC GTT AAA CAC ATC CCG AAT GGC AGC CGT GTG GTG CTT GCG CAT TCT ACA GGA GAA CCG CGT ACT CTG GTG AAG GCA ATG GTT GAA AAC TAC GAG CAG TAT AAA GAC GTC GAG GTT TGC CAC ATG CTG GGC CCA GGT GGA CCT TAC GAA TAC ACC AAT CCG GAG ATG AAA GGG CAT CTG TGG CAC AAC TCG CTC TTT ATG GGC CCA GGT GGA CGG AAA GCC TTT AAC GAA AAT CGC CTT GAC TTT ACG CCA GGG TAC TTC TGC GAT AGC ATC AAA TTC TTT CGT GAG GGC TAT CTG CCC GTT GAT GTG CAT ATG ATG ACC GTA TCG CCT CCA GAT AAA CAT GGG TAT GTG ACG TGT GGG ATT ACG TGC GAT TTC ACT ATG CCA GCA TTT GAA TGC GCC AAA ATG GTC ATC GTC CAG GTG AAC AAG AAT ATG CCG CGC ACG TTC GGT CAA ACC GCA ATC CAC CTG GAC GAC ATC GAT TTC GCG GTA GAA GCA GAT GAT CCG CTG TAT GGC AGT GTA CCG GGT GAA TTG ACA GAC ATT GAT CGC AAA ATT GGT GAA CAT TGT GCC TCG TTA ATC AAC GAT GGC GCT TGT CTG CAA TTA GGG ATT GGC GGC ATT CCG AAT GCC GTC TTG ACC TAT CTC ACC GAA AAA AAC AAT ATG GGC ATT CAG TCC GAG ATG CTC TCT GAT GGC ATT CTG CAG CTG ATT AAA GCC GGC AAC ATC AAC AAT AGC AAG AAA CAG ATT CAC GTG GGT AAA TCA GCG GTT ACC TTC TTG AAC GGT AGT CAG GAA CTG TAC GAT TAT GTG GAC GAT AAT CCG AGC GTA GAA TTT TAT CCG GTG GAT TAC ATC AAC GAC CCC TAC GTT ATT GGC AAG AAC GAC AAT ATG GTG TCC GTT AAT TCA GCG TTA TCG GTG GAT CTG ATG GGG CAA ATT GTT GCA GAT AAC CTG AGT GCG ACG CGC CAG ATC TCT GGT GCT GGT TTC GTA GAC TTT GTC CGT GGA GCC ACC ATC TCA AAA GGC GGC ATC AGC ATT GTG GCT ATG CCT AGC ACT GCG GCT GGT GGT AAA GCG AGT CCG ATT GAA ATG ATG TTT GAT GCC GGT CGC CCG ATT ACC CTG ACA CGC TTT GAG AGC TTC TAT GTT GTC ACG GAA TAC GGC ATT GCG AAA ATG CGC GGT AAT TCC TTA CGT ACC CGT GCA CGC CAA CTT ATC GAA ATT GCG CAT CCG GAT TTT CGT GAC GAA ATG AAA GAG TTC TAT GAA AAG CGC TTT GGC GAG AAA TAT taa

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2016/061287

A. CLASSIFICATION OF SUBJECT MATTER

INV. C12P7/04 C12P7/26 C12P7/42 C12P7/44 C12P11/00
C12P13/00 C12P19/32 C12N1/21

ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12P C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, COMPENDEX, FSTA, BIOSIS, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2014/248673 A1 (BOTES ET AL) 4 September 2014 (2014-09-04) * See Figure 1 (from pentanoyl-CoA, EC. 2.3.1.16) *	1-48
Y	----- KR 10-20 15-01 04450 A (KYUNGPOOK NATIONAL UNIVERSITY ET AL) 15 September 2015 (2015-09-15) * See Figure 1 (model reactions) and SEQ 1 (Ralstonia eutropha = Cupriavidus necator / SEQ 1 of the Application) *	1-48
Y	----- US 2014/193863 A1 (BOTES ET AL) 10 July 2014 (2014-07-10) * See Figures 4 and 6 (= Figure 5 of the Application) and SEQs 1-4 = 1-4 of the Application) *	11-20
	----- -/-	



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

20 February 2017

Date of mailing of the international search report

02/03/2017

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Korsner, Sven-Erik

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2016/061287

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y,P	WO 2016/077745 A1 (INVISTA TECHNOLOGIES S.A.R.L) 19 May 2016 (2016-05-19) * See Figure 1 (from 3-hydroxypropionyl-CoA) and SEQ ID NO 1 (Cupriavidus necator); earlier priority right to the extent that it is valid) * -----	1-48

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2016/061287

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2014248673 A1	04-09-2014	CN 105073996 A	18-11-2015
		EP 2931907 A2	21-10-2015
		US 2014248673 A1	04-09-2014
		WO 2014093847 A2	19-06-2014

KR 10-2015-0104450 A	15-09-2015	-----	
US 2014193863 A1	10-07-2014	CN 105189764 A	23-12-2015
		EP 2938736 A2	04-11-2015
		US 2014193863 A1	10-07-2014
		WO 2014105790 A2	03-07-2014

WO 2016077745 A1	19-05-2016	US 2016138052 A1	19-05-2016
		WO 2016077745 A1	19-05-2016
