

(43) International Publication Date
27 December 2012 (27.12.2012)(51) International Patent Classification:
C12P 7/18 (2006.01)(21) International Application Number:
PCT/US2012/043714(22) International Filing Date:
22 June 2012 (22.06.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/500,106 22 June 2011 (22.06.2011) US
61/620,377 4 April 2012 (04.04.2012) US(71) Applicant (for all designated States except US): **GENOMATICA, INC.** [US/US]; 10520 Wateridge Circle, San Diego, CA 92121 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **BURGARD, Anthony, P.** [US/US]; 10520 Wateridge Circle, San Diego, CA 92121 (US). **OSTERHOUT, Robin, E.** [US/US]; 10520 Wateridge Circle, San Diego, CA 92121 (US). **SUN, Jun** [CN/US]; 10520 Wateridge Circle, San Diego, CA 92121 (US). **PHARKYA, Priti** [IN/US]; 10520 Wateridge Circle, San Diego, CA 92121 (US).(74) Agents: **GAY, David, A.** et al.; Jones Day, 222 East 41st Street, New York, NY 10017-6702 (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, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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, 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, ML, MR, NE, SN, TD, TG).

Published:

- without international search report and to be republished upon receipt of that report (Rule 48.2(g))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: MICROORGANISMS FOR PRODUCING ETHYLENE GLYCOL AND METHODS RELATED THERETO

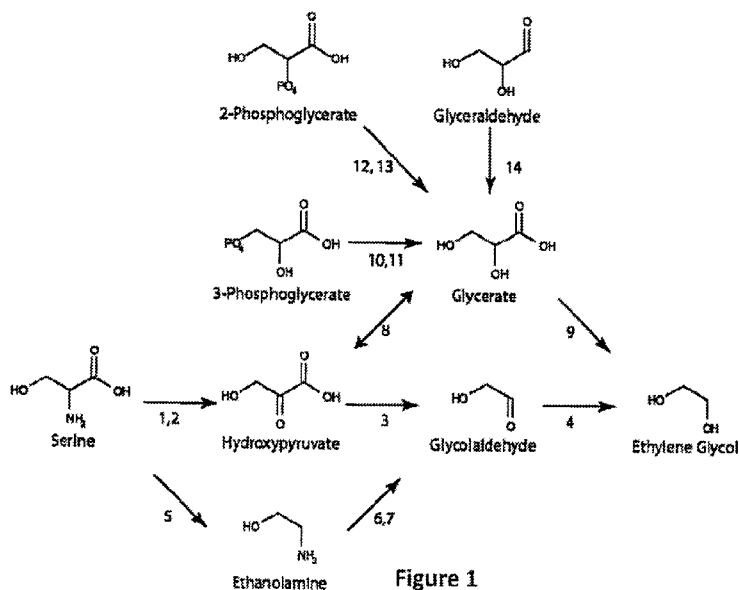


Figure 1

(57) Abstract: The invention provides non-naturally occurring microbial organisms having an ethylene glycol pathway. The invention additionally provides methods of using such organisms to produce ethylene glycol.

MICROORGANISMS FOR PRODUCING ETHYLENE GLYCOL AND METHODS RELATED THERETO

This application claims the benefit of priority of United States Provisional application serial
5 No. 61/500,106, filed June 22, 2011, United States Provisional application serial No. 61/620,377, filed April 4, 2012, the entire contents of which are incorporated herein by reference.

The instant application contains a Sequence Listing which has been submitted via EFS-Web and is hereby incorporated by reference in its entirety. Said ASCII copy, created on June 20,
10 2012, is named 871943-228181_PCT_Sequence_Listing.txt and is 77,803 bytes in size.

BACKGROUND OF THE INVENTION

The present invention relates generally to biosynthetic processes, and more specifically to organisms having ethylene glycol biosynthetic capability.

Ethylene glycol is a chemical commonly used in many commercial and industrial
15 applications including production of antifreezes and coolants. Ethylene glycol is also used as a raw material in the production of a wide range of products including polyester fibers for clothes, upholstery, carpet and pillows; fiberglass used in products such as jet skis, bathtubs, and bowling balls; and polyethylene terephthalate resin used in packaging film and bottles. Around 82% of ethylene glycol consumed worldwide is used in the production of polyester
20 fibers, resins and films. Strong growth in polyester demand has led to global growth rates of 5-6%/year for ethylene glycol. The second largest market for ethylene is antifreeze formulations.

Typically, in the manufacture of ethylene glycol, ethylene oxide is first produced by the oxidation of ethylene in the presence of oxygen or air and a silver oxide catalyst. A crude
25 ethylene glycol mixture is then produced by the hydrolysis of ethylene oxide with water under pressure. Fractional distillation under vacuum is used to separate the ethylene glycol from the higher glycols. Ethylene glycol was previously manufactured by the hydrolysis of ethylene oxide, which was produced via ethylene chlorohydrin but this method has been superseded by the direct oxidation route. Ethylene glycol is a colorless, odorless, viscous,
30 hygroscopic sweet-tasting liquid and is classified as harmful by the EC Dangerous Substances Directive.

Microbial organisms and methods for effectively producing commercial quantities of ethylene glycol are described herein and include related advantages.

SUMMARY OF INVENTION

5 The invention provides non-naturally occurring microbial organisms containing ethylene glycol pathways comprising at least one exogenous nucleic acid encoding an ethylene glycol pathway enzyme expressed in a sufficient amount to produce ethylene glycol. The invention additionally provides methods of using such microbial organisms to produce ethylene glycol, by culturing a non-naturally occurring microbial organism containing ethylene glycol pathways as described herein under conditions and for a sufficient period of time to produce
10 ethylene glycol.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows exemplary pathways for production of ethylene glycol. Enzymes for transformation of identified substrates to products include: 1) Serine aminotransferase, 2) Serine oxidoreductase (deaminating), 3) Hydroxypyruvate decarboxylase, 4) Glycolaldehyde
15 reductase, 5) Serine decarboxylase, 6) Ethanolamine aminotransferase, 7) Ethanolamine oxidoreductase (deaminating), 8) Hydroxypyruvate reductase, 9) Glycerate decarboxylase, 10) 3-Phosphoglycerate phosphatase, 11) Glycerate kinase, 12) 2-Phosphoglycerate phosphatase, 13) Glycerate-2-kinase and 14) Glyceraldehyde dehydrogenase.

Figure 2 shows an exemplary pathway for production of ethylene glycol. Enzymes for
20 transformation of identified substrates to products include: 1) Glyoxylate carboligase, 2) Hydroxypyruvate isomerase, 3) Hydroxypyruvate decarboxylase, 4) Glycolaldehyde reductase and 5) Glycerate dehydrogenase.

Figure 3 shows exemplary pathways for production of ethylene glycol. Enzymes for transformation of identified substrates to products include: 1) Glyoxylate reductase, 2)
25 Glycolyl-CoA transferase, 3) Glycolyl-CoA synthetase, 4) Glycolyl-CoA reductase (aldehyde forming), 5) Glycolaldehyde reductase, 6) Glycolate reductase, 7) Glycolate kinase, 8) Phosphotransglycolylase, 9) Glycolylphosphate reductase and 10) Glycolyl-CoA reductase (alcohol forming).

Figure 4 shows the reverse TCA cycle for fixation of CO₂ on carbohydrates as substrates. The enzymatic transformations are carried out by the enzymes as shown.

Figure 5 shows the pathway for the reverse TCA cycle coupled with carbon monoxide dehydrogenase and hydrogenase for the conversion of syngas to acetyl-CoA.

- 5 Figure 6 shows Western blots of 10 micrograms ACS90 (lane 1), ACS91 (lane2), Mta98/99 (lanes 3 and 4) cell extracts with size standards (lane 5) and controls of *M. thermoacetica* CODH (Moth_1202/1203) or Mtr (Moth_1197) proteins (50, 150, 250, 350, 450, 500, 750, 900, and 1000 ng).

- 10 Figure 7 shows CO oxidation assay results. Cells (*M. thermoacetica* or *E. coli* with the CODH/ACS operon; ACS90 or ACS91 or empty vector: pZA33S) were grown and extracts prepared. Assays were performed at 55°C at various times on the day the extracts were prepared. Reduction of methylviologen was followed at 578 nm over a 120 sec time course.

- 15 Figure 8A shows the nucleotide sequence (SEQ ID NO: 1) of carboxylic acid reductase from *Nocardia iowensis* (GNM_720), and Figure 8B shows the encoded amino acid sequence (SEQ ID NO: 2).

Figure 9A shows the nucleotide sequence (SEQ ID NO: 3) of phosphopantetheine transferase, which was codon optimized, and Figure 9B shows the encoded amino acid sequence (SEQ ID NO: 4).

- 20 Figure 10A shows the nucleotide sequence (SEQ ID NO: 5) of carboxylic acid reductase from *Mycobacterium smegmatis* mc(2)155 (designated 890), and Figure 10B shows the encoded amino acid sequence (SEQ ID NO: 6).

Figure 11A shows the nucleotide sequence (SEQ ID NO: 7) of carboxylic acid reductase from *Mycobacterium avium* subspecies paratuberculosis K-10 (designated 891), and Figure 11B shows the encoded amino acid sequence (SEQ ID NO: 8).

- 25 Figure 12A shows the nucleotide sequence (SEQ ID NO: 9) of carboxylic acid reductase from *Mycobacterium marinum* M (designated 892), and Figure 12B shows the encoded amino acid sequence (SEQ ID NO: 10).

Figure 13A shows the nucleotide sequence (SEQ ID NO: 11) of carboxylic acid reductase designated 891GA, and Figure 13B shows the encoded amino acid sequence (SEQ ID NO: 12).

Figure 14 shows exemplary pathways for production of ethylene glycol from 3-

- 5 phosphohydroxypyruvate. Enzymes for transformation of the identified substrants to products include: A. 3-phosphohydroxypyruvate decarboxylase; B. 2-oxoethyl-phosphate reductase; C. 2-hydroxyethyl-phosphate phosphatase or ethylene glycol kinase; D. 2-oxoethyl-phosphate phosphatase or glycolaldehyde kinase; E. glycolaldehyde reductase.

- 10 Figure 15 shows an exemplary flux distribution for achieving the maximum theoretical ethylene glycol yield from glucose and CO₂. 3PG is 3-phosphoglycerate, 3PHP is 3-phosphohydroxypyruvate, OEP is 2-oxoethyl-phosphate, HEP is 2-hydroxyethyl-phosphate, 2PG is 2-phosphoglycerate, PEP is phosphoenolpyruvate, CIT is citrate, AKG is alpha-ketoglutarate, SUCCOA is succinyl-CoA, SUCC is succinate, FUM is fumarate, MAL is malate, OAA is oxaloacetate, [H] indicates a reducing equivalent.

15 DETAILED DESCRIPTION OF THE INVENTION

The present invention is directed to the design and production of cells and organisms having biosynthetic production capabilities for ethylene glycol. The invention, in particular, relates to the design of microbial organism capable of producing ethylene glycol by introducing one or more nucleic acids encoding an ethylene glycol pathway enzyme.

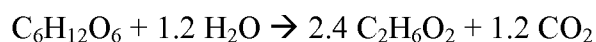
- 20 In one embodiment, the invention utilizes *in silico* stoichiometric models of *Escherichia coli* metabolism that identify metabolic designs for biosynthetic production of ethylene glycol. The results described herein indicate that metabolic pathways can be designed and recombinantly engineered to achieve the biosynthesis of ethylene glycol in *Escherichia coli* and other cells or organisms. Biosynthetic production of ethylene glycol, for example, for the
- 25 *in silico* designs can be confirmed by construction of strains having the designed metabolic genotype. These metabolically engineered cells or organisms also can be subjected to adaptive evolution to further augment ethylene glycol biosynthesis, including under conditions approaching theoretical maximum growth.

- 30 In certain embodiments, the ethylene glycol biosynthesis characteristics of the designed strains make them genetically stable and particularly useful in continuous bioprocesses.

Separate strain design strategies were identified with incorporation of different non-native or heterologous reaction capabilities into *E. coli* or other host organisms leading to ethylene glycol producing metabolic pathways from either serine, 3-phosphohydroxypyruvate, 3-phosphoglycerate or glyoxylate. *In silico* metabolic designs were identified that resulted in the biosynthesis of ethylene glycol in microorganisms from each of these substrates or metabolic intermediates.

Strains identified via the computational component of the platform can be put into actual production by genetically engineering any of the predicted metabolic alterations, which lead to the biosynthetic production of ethylene glycol or other intermediate and/or downstream products. In yet a further embodiment, strains exhibiting biosynthetic production of these compounds can be further subjected to adaptive evolution to further augment product biosynthesis. The levels of product biosynthesis yield following adaptive evolution also can be predicted by the computational component of the system.

The maximum theoretical ethylene glycol yield from glucose is 2.4 mol/mol (0.834 g/g), according to the equation:



The pathways presented in Figures 1-3 and Figure 14 achieve a yield of 2 moles ethylene glycol per mole of glucose utilized. Increasing product yields to 2.4 mol/mol is possible if cells are capable of fixing CO₂ through pathways such as the reductive TCA cycle or the Wood-Ljungdahl pathway.

As used herein, the term “non-naturally occurring” when used in reference to a microbial organism or microorganism of the invention is intended to mean that the microbial organism has at least one genetic alteration not normally found in a naturally occurring strain of the referenced species, including wild-type strains of the referenced species. Genetic alterations include, for example, modifications introducing expressible nucleic acids encoding metabolic polypeptides, other nucleic acid additions, nucleic acid deletions and/or other functional disruption of the microbial organism’s genetic material. Such modifications include, for example, coding regions and functional fragments thereof, for heterologous, homologous or both heterologous and homologous polypeptides for the referenced species. Additional modifications include, for example, non-coding regulatory regions in which the modifications

alter expression of a gene or operon. Exemplary metabolic polypeptides include enzymes or proteins within an ethylene glycol biosynthetic pathway.

A metabolic modification refers to a biochemical reaction that is altered from its naturally occurring state. Therefore, non-naturally occurring microorganisms can have genetic
5 modifications to nucleic acids encoding metabolic polypeptides, or functional fragments thereof. Exemplary metabolic modifications are disclosed herein.

As used herein, the term “ethylene glycol,” having the molecular formula $C_2H_6O_2$ and a molecular mass of 62.068 g/mol (see Figures 1-3) (IUPAC name ethane-1,2-diol) is used interchangeably throughout with monoethylene glycol, MEG, and 1,2-ethanediol. In its pure
10 form, ethylene glycol is an odorless, colorless, syrupy, sweet-tasting liquid. Ethylene glycol is widely used as an antifreeze in automobiles, as a medium for convective heat transfer in cooling systems and as a precursor to polyester fibers and resins. For example, polyethylene terephthalate, which is used to make plastic bottles, is prepared from ethylene glycol. Other known uses for ethylene glycol include use as a desiccant, as a chemical intermediate in the
15 manufacture of capacitors, as an additive to prevent corrosion and as a protecting group for carbonyl groups in organic synthesis.

As used herein, the term “isolated” when used in reference to a microbial organism is intended to mean an organism that is substantially free of at least one component as the referenced microbial organism is found in nature. The term includes a microbial organism
20 that is removed from some or all components as it is found in its natural environment. The term also includes a microbial organism that is removed from some or all components as the microbial organism is found in non-naturally occurring environments. Therefore, an isolated microbial organism is partly or completely separated from other substances as it is found in nature or as it is grown, stored or subsisted in non-naturally occurring environments. Specific
25 examples of isolated microbial organisms include partially pure microbes, substantially pure microbes and microbes cultured in a medium that is non-naturally occurring.

As used herein, the terms “microbial,” “microbial organism” or “microorganism” are intended to mean any organism that exists as a microscopic cell that is included within the domains of archaea, bacteria or eukarya. Therefore, the term is intended to encompass
30 prokaryotic or eukaryotic cells or organisms having a microscopic size and includes bacteria, archaea and eubacteria of all species as well as eukaryotic microorganisms such as yeast and

fungi. The term also includes cell cultures of any species that can be cultured for the production of a biochemical.

As used herein, the term “CoA” or “coenzyme A” is intended to mean an organic cofactor or prosthetic group (nonprotein portion of an enzyme) whose presence is required for the activity of many enzymes (the apoenzyme) to form an active enzyme system. Coenzyme A functions in certain condensing enzymes, acts in acetyl or other acyl group transfer and in fatty acid synthesis and oxidation, pyruvate oxidation and in other acetylation.

As used herein, the term “substantially anaerobic” when used in reference to a culture or growth condition is intended to mean that the amount of oxygen is less than about 10% of saturation for dissolved oxygen in liquid media. The term also is intended to include sealed chambers of liquid or solid medium maintained with an atmosphere of less than about 1% oxygen.

“Exogenous” as it is used herein is intended to mean that the referenced molecule or the referenced activity is introduced into the host microbial organism. The molecule can be introduced, for example, by introduction of an encoding nucleic acid into the host genetic material such as by integration into a host chromosome or as non-chromosomal genetic material such as a plasmid. Therefore, the term as it is used in reference to expression of an encoding nucleic acid refers to introduction of the encoding nucleic acid in an expressible form into the microbial organism. When used in reference to a biosynthetic activity, the term refers to an activity that is introduced into the host reference organism. The source can be, for example, a homologous or heterologous encoding nucleic acid that expresses the referenced activity following introduction into the host microbial organism. Therefore, the term “endogenous” refers to a referenced molecule or activity that is present in the host. Similarly, the term when used in reference to expression of an encoding nucleic acid refers to expression of an encoding nucleic acid contained within the microbial organism. The term “heterologous” refers to a molecule or activity derived from a source other than the referenced species whereas “homologous” refers to a molecule or activity derived from the host microbial organism. Accordingly, exogenous expression of an encoding nucleic acid of the invention can utilize either or both a heterologous or homologous encoding nucleic acid.

It is understood that when more than one exogenous nucleic acid is included in a microbial organism that the more than one exogenous nucleic acids refers to the referenced encoding

nucleic acid or biosynthetic activity, as discussed above. It is further understood, as disclosed herein, that such more than one exogenous nucleic acids can be introduced into the host microbial organism on separate nucleic acid molecules, on polycistronic nucleic acid molecules, or a combination thereof, and still be considered as more than one exogenous nucleic acid. For example, as disclosed herein a microbial organism can be engineered to express two or more exogenous nucleic acids encoding a desired pathway enzyme or protein. In the case where two exogenous nucleic acids encoding a desired activity are introduced into a host microbial organism, it is understood that the two exogenous nucleic acids can be introduced as a single nucleic acid, for example, on a single plasmid, on separate plasmids, can be integrated into the host chromosome at a single site or multiple sites, and still be considered as two exogenous nucleic acids. Similarly, it is understood that more than two exogenous nucleic acids can be introduced into a host organism in any desired combination, for example, on a single plasmid, on separate plasmids, can be integrated into the host chromosome at a single site or multiple sites, and still be considered as two or more exogenous nucleic acids, for example three exogenous nucleic acids. Thus, the number of referenced exogenous nucleic acids or biosynthetic activities refers to the number of encoding nucleic acids or the number of biosynthetic activities, not the number of separate nucleic acids introduced into the host organism.

The non-naturally occurring microbial organisms of the invention can contain stable genetic alterations, which refers to microorganisms that can be cultured for greater than five generations without loss of the alteration. Generally, stable genetic alterations include modifications that persist greater than 10 generations, particularly stable modifications will persist more than about 25 generations, and more particularly, stable genetic modifications will be greater than 50 generations, including indefinitely.

Those skilled in the art will understand that the genetic alterations, including metabolic modifications exemplified herein, are described with reference to a suitable host organism such as *E. coli* and their corresponding metabolic reactions or a suitable source organism for desired genetic material such as genes for a desired metabolic pathway. However, given the complete genome sequencing of a wide variety of organisms and the high level of skill in the area of genomics, those skilled in the art will readily be able to apply the teachings and guidance provided herein to essentially all other organisms. For example, the *E. coli* metabolic alterations exemplified herein can readily be applied to other species by

incorporating the same or analogous encoding nucleic acid from species other than the referenced species. Such genetic alterations include, for example, genetic alterations of species homologs, in general, and in particular, orthologs, paralogs or nonorthologous gene displacements.

- 5 An ortholog is a gene or genes that are related by vertical descent and are responsible for substantially the same or identical functions in different organisms. For example, mouse epoxide hydrolase and human epoxide hydrolase can be considered orthologs for the biological function of hydrolysis of epoxides. Genes are related by vertical descent when, for example, they share sequence similarity of sufficient amount to indicate they are
- 10 homologous, or related by evolution from a common ancestor. Genes can also be considered orthologs if they share three-dimensional structure but not necessarily sequence similarity, of a sufficient amount to indicate that they have evolved from a common ancestor to the extent that the primary sequence similarity is not identifiable. Genes that are orthologous can encode proteins with sequence similarity of about 25% to 100% amino acid sequence
- 15 identity. Genes encoding proteins sharing an amino acid similarity less than 25% can also be considered to have arisen by vertical descent if their three-dimensional structure also shows similarities. Members of the serine protease family of enzymes, including tissue plasminogen activator and elastase, are considered to have arisen by vertical descent from a common ancestor.
- 20 Orthologs include genes or their encoded gene products that through, for example, evolution, have diverged in structure or overall activity. For example, where one species encodes a gene product exhibiting two functions and where such functions have been separated into distinct genes in a second species, the two genes and their corresponding products are considered to be orthologs. For the production of a biochemical product, those skilled in the
- 25 art will understand that the orthologous gene harboring the metabolic activity to be introduced or disrupted is to be chosen for construction of the non-naturally occurring microorganism. An example of orthologs exhibiting separable activities is where distinct activities have been separated into distinct gene products between two or more species or within a single species. A specific example is the separation of elastase proteolysis and
- 30 plasminogen proteolysis, two types of serine protease activity, into distinct molecules as plasminogen activator and elastase. A second example is the separation of mycoplasma 5'-3' exonuclease and Drosophila DNA polymerase III activity. The DNA polymerase from the

first species can be considered an ortholog to either or both of the exonuclease or the polymerase from the second species and vice versa.

In contrast, paralogs are homologs related by, for example, duplication followed by evolutionary divergence and have similar or common, but not identical functions. Paralogs can originate or derive from, for example, the same species or from a different species. For example, microsomal epoxide hydrolase (epoxide hydrolase I) and soluble epoxide hydrolase (epoxide hydrolase II) can be considered paralogs because they represent two distinct enzymes, co-evolved from a common ancestor, that catalyze distinct reactions and have distinct functions in the same species. Paralogs are proteins from the same species with significant sequence similarity to each other suggesting that they are homologous, or related through co-evolution from a common ancestor. Groups of paralogous protein families include HipA homologs, luciferase genes, peptidases, and others.

A nonorthologous gene displacement is a nonorthologous gene from one species that can substitute for a referenced gene function in a different species. Substitution includes, for example, being able to perform substantially the same or a similar function in the species of origin compared to the referenced function in the different species. Although generally, a nonorthologous gene displacement will be identifiable as structurally related to a known gene encoding the referenced function, less structurally related but functionally similar genes and their corresponding gene products nevertheless will still fall within the meaning of the term as it is used herein. Functional similarity requires, for example, at least some structural similarity in the active site or binding region of a nonorthologous gene product compared to a gene encoding the function sought to be substituted. Therefore, a nonorthologous gene includes, for example, a paralog or an unrelated gene.

Therefore, in identifying and constructing the non-naturally occurring microbial organisms of the invention having ethylene glycol biosynthetic capability, those skilled in the art will understand with applying the teaching and guidance provided herein to a particular species that the identification of metabolic modifications can include identification and inclusion or inactivation of orthologs. To the extent that paralogs and/or nonorthologous gene displacements are present in the referenced microorganism that encode an enzyme catalyzing a similar or substantially similar metabolic reaction, those skilled in the art also can utilize these evolutionally related genes.

Orthologs, paralogs and nonorthologous gene displacements can be determined by methods well known to those skilled in the art. For example, inspection of nucleic acid or amino acid sequences for two polypeptides will reveal sequence identity and similarities between the compared sequences. Based on such similarities, one skilled in the art can determine if the similarity is sufficiently high to indicate the proteins are related through evolution from a common ancestor. Algorithms well known to those skilled in the art, such as Align, BLAST, Clustal W and others compare and determine a raw sequence similarity or identity, and also determine the presence or significance of gaps in the sequence which can be assigned a weight or score. Such algorithms also are known in the art and are similarly applicable for determining nucleotide sequence similarity or identity. Parameters for sufficient similarity to determine relatedness are computed based on well known methods for calculating statistical similarity, or the chance of finding a similar match in a random polypeptide, and the significance of the match determined. A computer comparison of two or more sequences can, if desired, also be optimized visually by those skilled in the art. Related gene products or proteins can be expected to have a high similarity, for example, 25% to 100% sequence identity. Proteins that are unrelated can have an identity which is essentially the same as would be expected to occur by chance, if a database of sufficient size is scanned (about 5%). Sequences between 5% and 24% may or may not represent sufficient homology to conclude that the compared sequences are related. Additional statistical analysis to determine the significance of such matches given the size of the data set can be carried out to determine the relevance of these sequences.

Exemplary parameters for determining relatedness of two or more sequences using the BLAST algorithm, for example, can be as set forth below. Briefly, amino acid sequence alignments can be performed using BLASTP version 2.0.8 (Jan-05-1999) and the following parameters: Matrix: 0 BLOSUM62; gap open: 11; gap extension: 1; x_dropoff: 50; expect: 10.0; wordsize: 3; filter: on. Nucleic acid sequence alignments can be performed using BLASTN version 2.0.6 (Sept-16-1998) and the following parameters: Match: 1; mismatch: -2; gap open: 5; gap extension: 2; x_dropoff: 50; expect: 10.0; wordsize: 11; filter: off. Those skilled in the art will know what modifications can be made to the above parameters to either increase or decrease the stringency of the comparison, for example, and determine the relatedness of two or more sequences.

In some embodiments, the invention provides a non-naturally occurring microbial organism, including a microbial organism having an ethylene glycol pathway having at least one exogenous nucleic acid encoding an ethylene glycol pathway enzyme expressed in a sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a serine aminotransferase, a serine oxidoreductase (deaminating), a hydroxypyruvate decarboxylase, a glycolaldehyde reductase, a serine decarboxylase, an ethanolamine aminotransferase, an ethanolamine oxidoreductase (deaminating), a hydroxypyruvate reductase or a glycerate decarboxylase (see steps 1-9 of Figure 1). In one aspect, the non-naturally occurring microbial organism includes a microbial organism having an ethylene glycol pathway having at least one exogenous nucleic acid encoding ethylene glycol pathway enzymes expressed in a sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a serine aminotransferase or a serine oxidoreductase (deaminating); a hydroxypyruvate decarboxylase, and a glycolaldehyde reductase (see steps 1/2, 3 and 4 of Figure 1). In one aspect, the non-naturally occurring microbial organism includes a microbial organism having an ethylene glycol pathway having at least one exogenous nucleic acid encoding ethylene glycol pathway enzymes expressed in a sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a serine aminotransferase or a serine oxidoreductase (deaminating); a hydroxypyruvate reductase, and a glycerate decarboxylase (see steps 1/2, 8, and 9 of Figure 1). In one aspect, the non-naturally occurring microbial organism includes a microbial organism having an ethylene glycol pathway having at least one exogenous nucleic acid encoding ethylene glycol pathway enzymes expressed in a sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a serine decarboxylase; an ethanolamine aminotransferase or an ethanolamine oxidoreductase (deaminating), and a glycolaldehyde reductase (see steps 5, 6/7 and 4 of Figure 1).

In some embodiments, the invention provides a non-naturally occurring microbial organism, including a microbial organism having an ethylene glycol pathway having at least one exogenous nucleic acid encoding an ethylene glycol pathway enzyme expressed in a sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a hydroxypyruvate decarboxylase, glycolaldehyde reductase, a hydroxypyruvate reductase, a glycerate decarboxylase, a 3-phosphoglycerate phosphatase, a glycerate kinase, a 2-phosphoglycerate phosphatase, a glycerate-2-kinase or a glyceraldehyde dehydrogenase (see steps 3, 4, and 8-14 of Figure 1). In one aspect, the non-naturally occurring microbial organism includes a microbial organism having an ethylene glycol pathway having at least

one exogenous nucleic acid encoding ethylene glycol pathway enzymes expressed in a sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a hydroxypyruvate reductase; a hydroxypyruvate decarboxylase, and a glycolaldehyde reductase (see steps 8, 3 and 4 of Figure 1). In one aspect, the non-naturally occurring

5 microbial organism includes a microbial organism having an ethylene glycol pathway having at least one exogenous nucleic acid encoding ethylene glycol pathway enzymes expressed in a sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a 3-phosphoglycerate phosphatase or a glycerate kinase; a hydroxypyruvate reductase; a hydroxypyruvate decarboxylase, and a glycolaldehyde reductase (see steps 10/11, 8, 3 and 4
10 of Figure 1). In one aspect, the non-naturally occurring microbial organism includes a microbial organism having an ethylene glycol pathway having at least one exogenous nucleic acid encoding ethylene glycol pathway enzymes expressed in a sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a 2-phosphoglycerate phosphatase or a glycerate-2-kinase; a hydroxypyruvate reductase; a hydroxypyruvate decarboxylase, and a
15 glycolaldehyde reductase (see steps 12/13, 8, 3 and 4 of Figure 1). In one aspect, the non-naturally occurring microbial organism includes a microbial organism having an ethylene glycol pathway having at least one exogenous nucleic acid encoding ethylene glycol pathway enzymes expressed in a sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a glyceraldehyde dehydrogenase, a hydroxypyruvate reductase, a
20 hydroxypyruvate decarboxylase and a glycolaldehyde reductase (see steps 14, 8, 3 and 4 of Figure 1).

In some embodiments, the invention provides a non-naturally occurring microbial organism, including a microbial organism having an ethylene glycol pathway having at least one exogenous nucleic acid encoding an ethylene glycol pathway enzyme expressed in a
25 sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a glycerate decarboxylase (see step 9 of Figure 1). In one aspect, the non-naturally occurring microbial organism includes a microbial organism having an ethylene glycol pathway having at least one exogenous nucleic acid encoding ethylene glycol pathway enzymes expressed in a sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a 3-
30 phosphoglycerate phosphatase or a glycerate kinase and a glycerate decarboxylase (see steps 10/11 and 9 of Figure 1). In one aspect, the non-naturally occurring microbial organism includes a microbial organism having an ethylene glycol pathway having at least one exogenous nucleic acid encoding ethylene glycol pathway enzymes expressed in a sufficient

amount to produce ethylene glycol, the ethylene glycol pathway including a 2-phosphoglycerate phosphatase, a glycerate-2-kinase and a glycerate decarboxylase (see steps 12/13, 9 of Figure 1). In one aspect, the non-naturally occurring microbial organism includes a microbial organism having an ethylene glycol pathway having at least one exogenous
5 nucleic acid encoding ethylene glycol pathway enzymes expressed in a sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a, a glyceraldehyde dehydrogenase and a glycerate decarboxylase (see steps 14 and 9 of Figure 1).

In some embodiments, the invention provides a non-naturally occurring microbial organism, including a microbial organism having an ethylene glycol pathway having at least one
10 exogenous nucleic acid encoding an ethylene glycol pathway enzyme expressed in a sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a glyoxylate carboligase, a hydroxypyruvate isomerase, a hydroxypyruvate decarboxylase, a glycolaldehyde reductase or a glycerate dehydrogenase (see steps 1, 2, 3, 4 or 5 of Figure 2).

In one aspect, the non-naturally occurring microbial organism includes a microbial organism
15 having an ethylene glycol pathway having at least one exogenous nucleic acid encoding ethylene glycol pathway enzymes expressed in a sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a glyoxylate carboligase, a hydroxypyruvate isomerase, a hydroxypyruvate decarboxylase and a glycolaldehyde reductase (see steps 1, 2, 3 and 4 of Figure 2). In one aspect, the non-naturally occurring microbial organism includes

a microbial organism having an ethylene glycol pathway having at least one exogenous
20 nucleic acid encoding ethylene glycol pathway enzymes expressed in a sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a glycerate dehydrogenase, a hydroxypyruvate isomerase, a hydroxypyruvate decarboxylase and a glycolaldehyde reductase (see steps 5, 2, 3 and 4 of Figure 2). In one aspect, the non-naturally occurring

microbial organism includes a microbial organism having an ethylene glycol pathway having
25 at least one exogenous nucleic acid encoding ethylene glycol pathway enzymes expressed in a sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a 3-phosphoglycerate phosphatase or a glycerate kinase; a glycerate dehydrogenase; a hydroxypyruvate isomerase; a hydroxypyruvate decarboxylase; and a glycolaldehyde

reductase (see steps 10/11 of Figure 1 and steps 5, 2, 3 and 4 of Figure 2). In one aspect, the
30 non-naturally occurring microbial organism includes a microbial organism having an ethylene glycol pathway having at least one exogenous nucleic acid encoding ethylene glycol pathway enzymes expressed in a sufficient amount to produce ethylene glycol, the ethylene

glycol pathway including a 2-phosphoglycerate phosphatase or a glycerate-2-kinase; a glycerate dehydrogenase; a hydroxypyruvate isomerase; a hydroxypyruvate decarboxylase; and a glycolaldehyde reductase (see steps 12/13 of Figure 1 and steps 5, 2, 3 and 4 of Figure 2). In one aspect, the non-naturally occurring microbial organism includes a microbial
5 organism having an ethylene glycol pathway having at least one exogenous nucleic acid encoding ethylene glycol pathway enzymes expressed in a sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a glyceraldehyde dehydrogenase, a glycerate dehydrogenase, a hydroxypyruvate isomerase, a hydroxypyruvate decarboxylase and a glycolaldehyde reductase (see step 14 of Figure 1 and steps 5, 2, 3 and 4 of Figure 2).

10 In some embodiments, the invention provides a non-naturally occurring microbial organism, including a microbial organism having an ethylene glycol pathway having at least one exogenous nucleic acid encoding an ethylene glycol pathway enzyme expressed in a sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a glyoxylate reductase, a glycolyl-CoA transferase, a glycolyl-CoA synthetase, a glycolyl-CoA
15 reductase (aldehyde forming), a glycolaldehyde reductase, a glycolate reductase, a glycolate kinase, a phosphotransglycolylase, a glycolylphosphate reductase or a glycolyl-CoA reductase (alcohol forming) (see steps 1-10 of Figure 3). In one aspect, the non-naturally occurring microbial organism includes a microbial organism having an ethylene glycol pathway having at least one exogenous nucleic acid encoding ethylene glycol pathway
20 enzymes expressed in a sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a glyoxylate reductase; a glycolyl-CoA transferase or a glycolyl-CoA synthetase; a glycolyl-CoA reductase (aldehyde forming), and a glycolaldehyde reductase (see steps 1, 2/3, 4 and 5 of Figure 3). In one aspect, the non-naturally occurring microbial organism includes a microbial organism having an ethylene glycol pathway having at least
25 one exogenous nucleic acid encoding ethylene glycol pathway enzymes expressed in a sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a glyoxylate reductase; a glycolate reductase, and a glycolaldehyde reductase (see steps 1, 6 and 5 of Figure 3). In one aspect, the non-naturally occurring microbial organism includes a microbial organism having an ethylene glycol pathway having at least one exogenous nucleic
30 acid encoding ethylene glycol pathway enzymes expressed in a sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a glyoxylate reductase; a glycolyl-CoA transferase or a glycolyl-CoA synthetase, and a glycolyl-CoA reductase (alcohol forming) (see steps 1, 2/3 and 10 of Figure 3). In one aspect, the non-naturally occurring

microbial organism includes a microbial organism having an ethylene glycol pathway having at least one exogenous nucleic acid encoding ethylene glycol pathway enzymes expressed in a sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a glyoxylate reductase, a glycolate kinase, a phosphotransglycolylase, glycolyl-CoA reductase (aldehyde forming) and a glycolaldehyde reductase (see steps 1, 7, 8, 4 and 5 of Figure 3). In one aspect, the non-naturally occurring microbial organism includes a microbial organism having an ethylene glycol pathway having at least one exogenous nucleic acid encoding ethylene glycol pathway enzymes expressed in a sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a glyoxylate reductase, a glycolate kinase, a phosphotransglycolylase and a glycolyl-CoA reductase (alcohol forming) (see steps 1, 7, 8 and 10 of Figure 3). In one aspect, the non-naturally occurring microbial organism includes a microbial organism having an ethylene glycol pathway having at least one exogenous nucleic acid encoding ethylene glycol pathway enzymes expressed in a sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a glyoxylate reductase, glycolate kinase, a glycolylphosphate reductase and a glycolaldehyde reductase (see steps 1, 7, 9 and 5 of Figure 3).

In some embodiments, the invention provides a non-naturally occurring microbial organism, including a microbial organism having an ethylene glycol pathway and having at least one exogenous nucleic acid encoding an ethylene glycol pathway enzyme expressed in a sufficient amount to produce ethylene glycol, wherein the ethylene glycol pathway comprises a pathway selected from: (A) a 3-phosphohydroxypyruvate decarboxylase, a 2-oxoethyl-phosphate reductase, and a 2-hydroxyethyl-phosphate phosphatase or an ethylene glycol kinase (see steps A, B and C of Figure 14); and (B) a 3-phosphohydroxypyruvate decarboxylase, a 2-oxoethyl-phosphate phosphatase or a glycolaldehyde kinase, and a glycolaldehyde reductase (see steps A, D and E of Figure 14).

In an additional embodiment, the invention provides a non-naturally occurring microbial organism having an ethylene glycol pathway, wherein the non-naturally occurring microbial organism comprises at least one exogenous nucleic acid encoding an enzyme or protein that converts a substrate to a product selected from the group consisting of serine to hydroxypyruvate, hydroxypyruvate to glycolaldehyde, glycolaldehyde to ethylene glycol, serine to ethanolamine, ethanolamine to glycolaldehyde, 3-phosphoglycerate to glycerate; 2-phosphoglycerate to glycerate, glyceraldehyde to glycerate, glycerate to hydroxypyruvate,

hydroxypyruvate to glycerate, glycerate to ethylene glycol, glycerate to tartronate semialdehyde, glyoxylate to tartronate semialdehyde, tartronate semialdehyde to hydroxypyruvate, glyoxylate to glycolate, glycolate to glycolaldehyde, glycolate to glycolylphosphate, glycolate to glycolyl-CoA, glycolyl-CoA to ethylene glycol, glycolyl-CoA to glycolaldehyde, glycolylphosphate to glycolyl-CoA, glycolylphosphate to glycolaldehyde, 3-phosphohydroxypyruvate to 2-oxoethyl-phosphate, 2-oxoethyl-phosphate to 2-hydroxyethyl-phosphate, 2-oxoethyl-phosphate to glycolaldehyde, and 2-hydroxyethyl-phosphate to ethylene glycol. One skilled in the art will understand that these are merely exemplary and that any of the substrate-product pairs disclosed herein suitable to produce a desired product and for which an appropriate activity is available for the conversion of the substrate to the product can be readily determined by one skilled in the art based on the teachings herein. Thus, the invention provides a non-naturally occurring microbial organism containing at least one exogenous nucleic acid encoding an enzyme or protein, where the enzyme or protein converts the substrates and products of an ethylene glycol pathway, such as that shown in Figures 1-3 and 14.

While generally described herein as a microbial organism that contains an ethylene glycol pathway, it is understood that the invention additionally provides a non-naturally occurring microbial organism comprising at least one exogenous nucleic acid encoding an ethylene glycol pathway enzyme expressed in a sufficient amount to produce an intermediate of an ethylene glycol pathway. For example, as disclosed herein, an ethylene glycol pathway is exemplified in Figures 1-3 and 14. Therefore, in addition to a microbial organism containing an ethylene glycol pathway that produces ethylene glycol, the invention additionally provides a non-naturally occurring microbial organism comprising at least one exogenous nucleic acid encoding an ethylene glycol pathway enzyme, where the microbial organism produces an ethylene glycol pathway intermediate, for example, hydroxypyruvate, ethanolamine, glycolaldehyde, glycerate, tartronate semialdehyde, glycolate, glycolylphosphate, glycolyl-CoA, 2-oxoethyl-phosphate or 2-hydroxyethyl-phosphate.

It is understood that any of the pathways disclosed herein, as described in the Examples and exemplified in the Figures, including the pathways of Figures 1-3 and 14, can be utilized to generate a non-naturally occurring microbial organism that produces any pathway intermediate or product, as desired. As disclosed herein, such a microbial organism that produces an intermediate can be used in combination with another microbial organism

expressing downstream pathway enzymes to produce a desired product. However, it is understood that a non-naturally occurring microbial organism that produces an ethylene glycol pathway intermediate can be utilized to produce the intermediate as a desired product.

This invention is also directed, in part to engineered biosynthetic pathways to improve carbon flux through a central metabolism intermediate en route to ethylene glycol. The present invention provides non-naturally occurring microbial organisms having one or more exogenous genes encoding enzymes that can catalyze various enzymatic transformations en route to ethylene glycol. In some embodiments, these enzymatic transformations are part of the reductive tricarboxylic acid (RTCA) cycle and are used to improve product yields, including but not limited to, from carbohydrate-based carbon feedstock.

In numerous engineered pathways, realization of maximum product yields based on carbohydrate feedstock is hampered by insufficient reducing equivalents or by loss of reducing equivalents and/or carbon to byproducts. In accordance with some embodiments, the present invention increases the yields of ethylene glycol by (i) enhancing carbon fixation via the reductive TCA cycle, and/or (ii) accessing additional reducing equivalents from gaseous carbon sources and/or syngas components such as CO, CO₂, and/or H₂. In addition to syngas, other sources of such gases include, but are not limited to, the atmosphere, either as found in nature or generated.

The CO₂-fixing reductive tricarboxylic acid (RTCA) cycle is an endergenic anabolic pathway of CO₂ assimilation which uses reducing equivalents and ATP (Figure 4). One turn of the RTCA cycle assimilates two moles of CO₂ into one mole of acetyl-CoA, or four moles of CO₂ into one mole of oxaloacetate. This additional availability of acetyl-CoA improves the maximum theoretical yield of product molecules derived from carbohydrate-based carbon feedstock. Exemplary carbohydrates include but are not limited to glucose, sucrose, xylose, arabinose and glycerol.

In some embodiments, the reductive TCA cycle, coupled with carbon monoxide dehydrogenase and/or hydrogenase enzymes, can be employed to allow syngas, CO₂, CO, H₂, and/or other gaseous carbon source utilization by microorganisms. Synthesis gas (syngas), in particular is a mixture of primarily H₂ and CO, sometimes including some amounts of CO₂, that can be obtained via gasification of any organic feedstock, such as coal, coal oil, natural gas, biomass, or waste organic matter. Numerous gasification processes have

been developed, and most designs are based on partial oxidation, where limiting oxygen avoids full combustion, of organic materials at high temperatures (500-1500°C) to provide syngas as a 0.5:1-3:1 H₂/CO mixture. In addition to coal, biomass of many types has been used for syngas production and represents an inexpensive and flexible feedstock for the biological production of renewable chemicals and fuels. Carbon dioxide can be provided from the atmosphere or in condensed form, for example, from a tank cylinder, or via sublimation of solid CO₂. Similarly, CO and hydrogen gas can be provided in reagent form and/or mixed in any desired ratio. Other gaseous carbon forms can include, for example, methanol or similar volatile organic solvents.

The components of synthesis gas and/or other carbon sources can provide sufficient CO₂, reducing equivalents, and ATP for the reductive TCA cycle to operate. One turn of the RTCA cycle assimilates two moles of CO₂ into one mole of acetyl-CoA and requires 2 ATP and 4 reducing equivalents. CO and/or H₂ can provide reducing equivalents by means of carbon monoxide dehydrogenase and hydrogenase enzymes, respectively. Reducing equivalents can come in the form of NADH, NADPH, FADH, reduced quinones, reduced ferredoxins, and reduced flavodoxins. The reducing equivalents, particularly NADH, NADPH, reduced ferredoxin and thioredoxins, can serve as cofactors for the RTCA cycle enzymes, for example, malate dehydrogenase, fumarate reductase, alpha-ketoglutarate:ferredoxin oxidoreductase (alternatively known as 2-oxoglutarate:ferredoxin oxidoreductase, alpha-ketoglutarate synthase, or 2-oxoglutarate synthase), pyruvate:ferredoxin oxidoreductase and isocitrate dehydrogenase. The electrons from these reducing equivalents can alternatively pass through an ion-gradient producing electron transport chain where they are passed to an acceptor such as oxygen, nitrate, oxidized metal ions, protons, or an electrode. The ion-gradient can then be used for ATP generation via an ATP synthase or similar enzyme.

The reductive TCA cycle was first reported in the green sulfur photosynthetic bacterium *Chlorobium limicola* (Evans et al., Proc. Natl. Acad. Sci. U.S.A. 55:928-934 (1966)). Similar pathways have been characterized in some prokaryotes (proteobacteria, green sulfur bacteria and thermophilic Knallgas bacteria) and sulfur-dependent archaea (Hugler et al., J. Bacteriol. 187:3020-3027 (2005); Hugler et al., Environ. Microbiol. 9:81-92 (2007)). In some cases, reductive and oxidative (Krebs) TCA cycles are present in the same organism (Hugler et al., supra (2007); Siebers et al., J. Bacteriol. 186:2179-2194 (2004)). Some methanogens and

obligate anaerobes possess incomplete oxidative or reductive TCA cycles that may function to synthesize biosynthetic intermediates (Ekiel et al., J. Bacteriol. 162:905-908 (1985); Wood et al., FEMS Microbiol. Rev. 28:335-352 (2004)).

The key carbon-fixing enzymes of the reductive TCA cycle are alpha-

- 5 ketoglutarate:ferredoxin oxidoreductase, pyruvate:ferredoxin oxidoreductase and isocitrate dehydrogenase. Additional carbon may be fixed during the conversion of phosphoenolpyruvate to oxaloacetate by phosphoenolpyruvate carboxylase or phosphoenolpyruvate carboxykinase or by conversion of pyruvate to malate by malic enzyme.
- 10 Many of the enzymes in the TCA cycle are reversible and can catalyze reactions in the reductive and oxidative directions. However, some TCA cycle reactions are irreversible in vivo and thus different enzymes are used to catalyze these reactions in the directions required for the reverse TCA cycle. These reactions are: (1) conversion of citrate to oxaloacetate and acetyl-CoA, (2) conversion of fumarate to succinate, and (3) conversion of succinyl-CoA to
- 15 alpha-ketoglutarate. In the TCA cycle, citrate is formed from the condensation of oxaloacetate and acetyl-CoA. The reverse reaction, cleavage of citrate to oxaloacetate and acetyl-CoA, is ATP-dependent and catalyzed by ATP-citrate lyase, or citryl-CoA synthetase and citryl-CoA lyase. Alternatively, citrate lyase can be coupled to acetyl-CoA synthetase, an acetyl-CoA transferase, or phosphotransacetylase and acetate kinase to form acetyl-CoA
- 20 and oxaloacetate from citrate. The conversion of succinate to fumarate is catalyzed by succinate dehydrogenase while the reverse reaction is catalyzed by fumarate reductase. In the TCA cycle succinyl-CoA is formed from the NAD(P)⁺ dependent decarboxylation of alpha-ketoglutarate by the alpha-ketoglutarate dehydrogenase complex. The reverse reaction is catalyzed by alpha-ketoglutarate:ferredoxin oxidoreductase.
- 25 An organism capable of utilizing the reverse tricarboxylic acid cycle to enable production of acetyl-CoA-derived products on 1) CO, 2) CO₂ and H₂, 3) CO and CO₂, 4) synthesis gas comprising CO and H₂, and 5) synthesis gas or other gaseous carbon sources comprising CO, CO₂, and H₂ can include any of the following enzyme activities: ATP-citrate lyase, a citrate lyase, a citryl-CoA synthetase, a citryl-CoA lyase, aconitase, isocitrate dehydrogenase, alpha-
- 30 ketoglutarate:ferredoxin oxidoreductase, succinyl-CoA synthetase, succinyl-CoA transferase, fumarate reductase, fumarase, malate dehydrogenase, acetate kinase, phosphotransacetylase, acetyl-CoA synthetase, acetyl-CoA transferase, pyruvate:ferredoxin oxidoreductase,

NAD(P)H:ferredoxin oxidoreductase, carbon monoxide dehydrogenase, hydrogenase, and ferredoxin (see Figure 5). Enzymes and the corresponding genes required for these activities are described herein above.

Carbon from syngas or other gaseous carbon sources can be fixed via the reverse TCA cycle and components thereof. Specifically, the combination of certain carbon gas-utilization pathway components with the pathways for formation of ethylene glycol from acetyl-CoA results in high yields of these products by providing an efficient mechanism for fixing the carbon present in carbon dioxide, fed exogenously or produced endogenously from CO, into acetyl-CoA.

- 5 In some embodiments, an ethylene glycol pathway in a non-naturally occurring microbial organism of the invention can utilize any combination of (1) CO, (2) CO₂, (3) H₂, or mixtures thereof to enhance the yields of biosynthetic steps involving reduction, including addition to driving the reductive TCA cycle.

- 15 In some embodiments a non-naturally occurring microbial organism having an ethylene glycol pathway includes at least one exogenous nucleic acid encoding a reductive TCA pathway enzyme. The at least one exogenous nucleic acid is selected from an ATP-citrate lyase, a citrate lyase, a citryl-CoA synthetase, a citryl-CoA lyase, a fumarate reductase, and an alpha-ketoglutarate:ferredoxin oxidoreductase; and at least one exogenous enzyme selected from a carbon monoxide dehydrogenase, a hydrogenase, a NAD(P)H:ferredoxin oxidoreductase, and a ferredoxin, expressed in a sufficient amount to allow the utilization of (1) CO, (2) CO₂, (3) H₂, (4) CO₂ and H₂, (5) CO and CO₂, (6) CO and H₂, or (7) CO, CO₂, and H₂.

- 25 In some embodiments a method includes culturing a non-naturally occurring microbial organism having an ethylene glycol pathway also comprising at least one exogenous nucleic acid encoding a reductive TCA pathway enzyme. The at least one exogenous nucleic acid is selected from an ATP-citrate lyase, a citrate lyase, a citryl-CoA synthetase, a citryl-CoA lyase, a fumarate reductase, isocitrate dehydrogenase, aconitase, and an alpha-ketoglutarate:ferredoxin oxidoreductase. Additionally, such an organism can also include at least one exogenous enzyme selected from a carbon monoxide dehydrogenase, a hydrogenase, a NAD(P)H:ferredoxin oxidoreductase, and a ferredoxin, expressed in a
- 30

sufficient amount to allow the utilization of (1) CO, (2) CO₂, (3) H₂, (4) CO₂ and H₂, (5) CO and CO₂, (6) CO and H₂, or (7) CO, CO₂, and H₂ to produce a product.

In some embodiments a non-naturally occurring microbial organism having an ethylene glycol pathway further includes at least one exogenous nucleic acid encoding a reductive

5 TCA pathway enzyme expressed in a sufficient amount to enhance carbon flux through acetyl-CoA. The at least one exogenous nucleic acid is selected from an ATP-citrate lyase, a citrate lyase, a citryl-CoA synthetase, a citryl-CoA lyase, a fumarate reductase, a pyruvate:ferredoxin oxidoreductase, isocitrate dehydrogenase, aconitase, and an alpha-ketoglutarate:ferredoxin oxidoreductase.

10 In some embodiments a non-naturally occurring microbial organism having an ethylene glycol pathway includes at least one exogenous nucleic acid encoding an enzyme expressed in a sufficient amount to enhance the availability of reducing equivalents in the presence of carbon monoxide and/or hydrogen, thereby increasing the yield of redox-limited products via carbohydrate-based carbon feedstock. The at least one exogenous nucleic acid is selected
15 from a carbon monoxide dehydrogenase, a hydrogenase, an NAD(P)H:ferredoxin oxidoreductase, and a ferredoxin. In some embodiments, the present invention provides a method for enhancing the availability of reducing equivalents in the presence of carbon monoxide or hydrogen thereby increasing the yield of redox-limited products via carbohydrate-based carbon feedstock, such as sugars or gaseous carbon sources, the method
20 includes culturing this non-naturally occurring microbial organism under conditions and for a sufficient period of time to produce ethylene glycol.

In some embodiments, the non-naturally occurring microbial organism having an ethylene glycol pathway includes two exogenous nucleic acids, each encoding a reductive TCA

25 pathway enzyme. In some embodiments, the non-naturally occurring microbial organism having an ethylene glycol pathway includes three exogenous nucleic acids each encoding a reductive TCA pathway enzyme. In some embodiments, the non-naturally occurring microbial organism includes three exogenous nucleic acids encoding an ATP-citrate lyase, a fumarate reductase, and an alpha-ketoglutarate:ferredoxin oxidoreductase. In some
30 embodiments, the non-naturally occurring microbial organism includes three exogenous nucleic acids encoding a citrate lyase, a fumarate reductase, and an alpha-ketoglutarate:ferredoxin oxidoreductase. In some embodiments, the non-naturally occurring microbial organism includes four exogenous nucleic acids encoding a pyruvate:ferredoxin

oxidoreductase; a phosphoenolpyruvate carboxylase or a phosphoenolpyruvate carboxykinase, a CO dehydrogenase; and an H₂ hydrogenase. In some embodiments, the non-naturally occurring microbial organism includes two exogenous nucleic acids encoding a CO dehydrogenase and an H₂ hydrogenase.

- 5 In some embodiments, the non-naturally occurring microbial organisms having an ethylene glycol pathway further include an exogenous nucleic acid encoding an enzyme selected from a pyruvate:ferredoxin oxidoreductase, an aconitase, an isocitrate dehydrogenase, a succinyl-CoA synthetase, a succinyl-CoA transferase, a fumarase, a malate dehydrogenase, an acetate kinase, a phosphotransacetylase, an acetyl-CoA synthetase, an NAD(P)H:ferredoxin
10 oxidoreductase, and combinations thereof.

In some embodiments, the non-naturally occurring microbial organism having an ethylene glycol pathway further includes an exogenous nucleic acid encoding an enzyme selected from carbon monoxide dehydrogenase, acetyl-CoA synthase, ferredoxin, NAD(P)H:ferredoxin oxidoreductase and combinations thereof.

- 15 In some embodiments, the non-naturally occurring microbial organism having an ethylene glycol pathway utilizes a carbon feedstock selected from (1) CO, (2) CO₂, (3) CO₂ and H₂, (4) CO and H₂, or (5) CO, CO₂, and H₂. In some embodiments, the non-naturally occurring microbial organism having an ethylene glycol pathway utilizes hydrogen for reducing equivalents. In some embodiments, the non-naturally occurring microbial organism having
20 an ethylene glycol pathway utilizes CO for reducing equivalents. In some embodiments, the non-naturally occurring microbial organism having an ethylene glycol pathway utilizes combinations of CO and hydrogen for reducing equivalents.

- In some embodiments, the non-naturally occurring microbial organism having an ethylene glycol pathway further includes one or more nucleic acids encoding an enzyme selected from
25 a phosphoenolpyruvate carboxylase, a phosphoenolpyruvate carboxykinase, a pyruvate carboxylase, and a malic enzyme.

- In some embodiments, the non-naturally occurring microbial organism having an ethylene glycol pathway further includes one or more nucleic acids encoding an enzyme selected from a malate dehydrogenase, a fumarase, a fumarate reductase, a succinyl-CoA synthetase, and a
30 succinyl-CoA transferase.

In some embodiments, the non-naturally occurring microbial organism having an ethylene glycol pathway further includes at least one exogenous nucleic acid encoding a citrate lyase, an ATP-citrate lyase, a citryl-CoA synthetase, a citryl-CoA lyase, an aconitase, an isocitrate dehydrogenase, a succinyl-CoA synthetase, a succinyl-CoA transferase, a fumarase, a malate dehydrogenase, an acetate kinase, a phosphotransacetylase, an acetyl-CoA synthetase, and a ferredoxin.

It is understood by those skilled in the art that the above-described pathways for increasing product yield can be combined with any of the pathways disclosed herein, including those pathways depicted in the figures. One skilled in the art will understand that, depending on the pathway to a desired product and the precursors and intermediates of that pathway, a particular pathway for improving product yield, as discussed herein above and in the examples, or combination of such pathways, can be used in combination with a pathway to a desired product to increase the yield of that product or a pathway intermediate.

In one embodiment, the invention provides a non-naturally occurring microbial organism, comprising a microbial organism having an ethylene glycol pathway or pathway intermediate, including any of the pathways disclosed herein and shown in the figures. In one embodiment, the invention provides a microbial organism comprising at least one exogenous nucleic acid encoding an ethylene glycol pathway enzyme expressed in a sufficient amount to produce ethylene glycol. Such a non-naturally occurring microbial organism can further comprise (i) a reductive TCA pathway comprising at least one exogenous nucleic acid encoding a reductive TCA pathway enzyme, wherein the at least one exogenous nucleic acid is selected from an ATP-citrate lyase, a citrate lyase, a citryl-CoA synthetase, a citryl-CoA lyase, a fumarate reductase, and an alpha-ketoglutarate:ferredoxin oxidoreductase; (ii) a reductive TCA pathway comprising at least one exogenous nucleic acid encoding a reductive TCA pathway enzyme, wherein the at least one exogenous nucleic acid is selected from a pyruvate:ferredoxin oxidoreductase, a phosphoenolpyruvate carboxylase, a phosphoenolpyruvate carboxykinase, a CO dehydrogenase, and an H₂ hydrogenase; or (iii) at least one exogenous nucleic acid encodes an enzyme selected from a CO dehydrogenase, an H₂ hydrogenase, and combinations thereof. Such a microbial organism can comprise an ethylene glycol pathway selected, for example, from (A) a serine aminotransferase or a serine oxidoreductase (deaminating); a hydroxypyruvate decarboxylase, and a glycolaldehyde reductase; (B) a serine aminotransferase or a serine

oxidoreductase (deaminating); a hydroxypyruvate reductase, and a glycerate decarboxylase;
 (C) a serine decarboxylase; an ethanolamine aminotransferase or an ethanolamine
 oxidoreductase (deaminating), and a glycolaldehyde reductase; (D) a hydroxypyruvate
 reductase; a hydroxypyruvate decarboxylase, and a glycolaldehyde reductase; (E) a glycerate
 5 decarboxylase; (F) glycerate dehydrogenase, a hydroxypyruvate isomerase, a
 hydroxypyruvate decarboxylase and a glycolaldehyde reductase; (G) a glyoxylate
 carboligase, a hydroxypyruvate isomerase, a hydroxypyruvate decarboxylase and a
 glycolaldehyde reductase; (H) a glyoxylate reductase; a glycolyl-CoA transferase or a
 glycolyl-CoA synthetase; a glycolyl-CoA reductase (aldehyde forming), and a
 10 glycolaldehyde reductase; (I) a glyoxylate reductase; a glycolate reductase, and a
 glycolaldehyde reductase; (J) a glyoxylate reductase; a glycolyl-CoA transferase or a
 glycolyl-CoA synthetase, and a glycolyl-CoA reductase (alcohol forming); (K) a glyoxylate
 reductase, a glycolate kinase, a phosphotransglycolylase, glycolyl-CoA reductase (aldehyde
 forming) and a glycolaldehyde reductase; (L) a glyoxylate reductase, a glycolate kinase, a
 15 phosphotransglycolylase and a glycolyl-CoA reductase (alcohol forming); (M) a glyoxylate
 reductase, glycolate kinase, a glycolylphosphate reductase and a glycolaldehyde reductase;
 (N) a 3-phosphohydroxypyruvate decarboxylase, a 2-oxoethyl-phosphate reductase, and a 2-
 hydroxyethyl-phosphate phosphatase or an ethylene glycol kinase; and (O) a 3-
 phosphohydroxypyruvate decarboxylase, a 2-oxoethyl-phosphate phosphatase or a
 20 glycolaldehyde kinase, and a glycolaldehyde reductase.

In a particular embodiment, a microbial organism have the ethylene glycol pathway of (D),
 (E) and/or (F) can further comprise a 3-phosphoglycerate phosphatase, a glycerate kinase, a
 2-phosphoglycerate phosphatase, a glycerate-2-kinase, or a glyceraldehyde dehydrogenase.
 In another embodiment, a microbial organism comprising (i) can further comprise an
 25 exogenous nucleic acid encoding an enzyme selected from a pyruvate:ferredoxin
 oxidoreductase, an aconitase, an isocitrate dehydrogenase, a succinyl-CoA synthetase, a
 succinyl-CoA transferase, a fumarase, a malate dehydrogenase, an acetate kinase, a
 phosphotransacetylase, an acetyl-CoA synthetase, an NAD(P)H:ferredoxin oxidoreductase,
 ferredoxin, and combinations thereof. In still a further embodiment, a microbial organism
 30 comprising (ii) further comprises an exogenous nucleic acid encoding an enzyme selected
 from an aconitase, an isocitrate dehydrogenase, a succinyl-CoA synthetase, a succinyl-CoA
 transferase, a fumarase, a malate dehydrogenase, and combinations thereof.

In a particular embodiment, microbial organism can comprise exogenous nucleic acids encoding each of the enzymes selected from: (A) a serine aminotransferase or a serine oxidoreductase (deaminating); a hydroxypyruvate decarboxylase, and a glycolaldehyde reductase; (B) a serine aminotransferase or a serine oxidoreductase (deaminating); a hydroxypyruvate reductase, and a glycerate decarboxylase; (C) a serine decarboxylase; an ethanolamine aminotransferase or an ethanolamine oxidoreductase (deaminating), and a glycolaldehyde reductase; (D) a hydroxypyruvate reductase; a hydroxypyruvate decarboxylase, and a glycolaldehyde reductase; (E) a glycerate decarboxylase; (F) glycerate dehydrogenase, a hydroxypyruvate isomerase, a hydroxypyruvate decarboxylase and a glycolaldehyde reductase; (G) a glyoxylate carboligase, a hydroxypyruvate isomerase, a hydroxypyruvate decarboxylase and a glycolaldehyde reductase; (H) a glyoxylate reductase; a glycolyl-CoA transferase or a glycolyl-CoA synthetase; a glycolyl-CoA reductase (aldehyde forming), and a glycolaldehyde reductase; (I) a glyoxylate reductase; a glycolate reductase, and a glycolaldehyde reductase; (J) a glyoxylate reductase; a glycolyl-CoA transferase or a glycolyl-CoA synthetase, and a glycolyl-CoA reductase (alcohol forming); (K) a glyoxylate reductase, a glycolate kinase, a phosphotransglycolylase, glycolyl-CoA reductase (aldehyde forming) and a glycolaldehyde reductase; (L) a glyoxylate reductase, a glycolate kinase, a phosphotransglycolylase and a glycolyl-CoA reductase (alcohol forming); (M) a glyoxylate reductase, glycolate kinase, a glycolylphosphate reductase and a glycolaldehyde reductase; (N) a 3-phosphohydroxypyruvate decarboxylase, a 2-oxoethyl-phosphate reductase, and a 2-hydroxyethyl-phosphate phosphatase or an ethylene glycol kinase; and (O) a 3-phosphohydroxypyruvate decarboxylase, a 2-oxoethyl-phosphate phosphatase or a glycolaldehyde kinase, and a glycolaldehyde reductase. In one embodiment, such a microbial organism can comprise two, three, four or five exogenous nucleic acids each encoding enzymes of (i), (ii) or (iii).

For example, a microbial organism comprising (i) can comprise three exogenous nucleic acids encoding ATP-citrate lyase or citrate lyase; a fumarate reductase; and an alpha-ketoglutarate:ferredoxin oxidoreductase; microbial organism comprising (ii) can comprise four exogenous nucleic acids encoding a pyruvate:ferredoxin oxidoreductase; a phosphoenolpyruvate carboxylase or a phosphoenolpyruvate carboxykinase; a CO dehydrogenase; and an H₂ hydrogenase; or a microbial organism comprising (iii) can comprise two exogenous nucleic acids encoding CO dehydrogenase and H₂ hydrogenase. The invention further provides methods for producing ethylene glycol by culturing such non-

naturally occurring microbial organism under conditions and for a sufficient period of time to produce ethylene glycol.

In some embodiments, the carbon feedstock and other cellular uptake sources such as phosphate, ammonia, sulfate, chloride and other halogens can be chosen to alter the isotopic distribution of the atoms present in ethylene glycol or any ethylene glycol pathway intermediate. The various carbon feedstock and other uptake sources enumerated above will be referred to herein, collectively, as “uptake sources.” Uptake sources can provide isotopic enrichment for any atom present in the product ethylene glycol or ethylene glycol pathway intermediate, or for side products generated in reactions diverging away from an ethylene glycol pathway. Isotopic enrichment can be achieved for any target atom including, for example, carbon, hydrogen, oxygen, nitrogen, sulfur, phosphorus, chloride or other halogens.

In some embodiments, the uptake sources can be selected to alter the carbon-12, carbon-13, and carbon-14 ratios. In some embodiments, the uptake sources can be selected to alter the oxygen-16, oxygen-17, and oxygen-18 ratios. In some embodiments, the uptake sources can be selected to alter the hydrogen, deuterium, and tritium ratios. In some embodiments, the uptake sources can be selected to alter the nitrogen-14 and nitrogen-15 ratios. In some embodiments, the uptake sources can be selected to alter the sulfur-32, sulfur-33, sulfur-34, and sulfur-35 ratios. In some embodiments, the uptake sources can be selected to alter the phosphorus-31, phosphorus-32, and phosphorus-33 ratios. In some embodiments, the uptake sources can be selected to alter the chlorine-35, chlorine-36, and chlorine-37 ratios.

In some embodiments, the isotopic ratio of a target atom can be varied to a desired ratio by selecting one or more uptake sources. An uptake source can be derived from a natural source, as found in nature, or from a man-made source, and one skilled in the art can select a natural source, a man-made source, or a combination thereof, to achieve a desired isotopic ratio of a target atom. An example of a man-made uptake source includes, for example, an uptake source that is at least partially derived from a chemical synthetic reaction. Such isotopically enriched uptake sources can be purchased commercially or prepared in the laboratory and/or optionally mixed with a natural source of the uptake source to achieve a desired isotopic ratio. In some embodiments, a target atom isotopic ratio of an uptake source can be achieved by selecting a desired origin of the uptake source as found in nature. For example, as discussed herein, a natural source can be a biobased derived from or synthesized by a biological organism or a source such as petroleum-based products or the atmosphere. In

some such embodiments, a source of carbon, for example, can be selected from a fossil fuel-derived carbon source, which can be relatively depleted of carbon-14, or an environmental or atmospheric carbon source, such as CO₂, which can possess a larger amount of carbon-14 than its petroleum-derived counterpart.

- 5 The unstable carbon isotope carbon-14 or radiocarbon makes up for roughly 1 in 10¹² carbon atoms in the earth's atmosphere and has a half-life of about 5700 years. The stock of carbon is replenished in the upper atmosphere by a nuclear reaction involving cosmic rays and ordinary nitrogen (¹⁴N). Fossil fuels contain no carbon-14, as it decayed long ago. Burning of fossil fuels lowers the atmospheric carbon-14 fraction, the so-called "Suess effect".
- 10 Methods of determining the isotopic ratios of atoms in a compound are well known to those skilled in the art. Isotopic enrichment is readily assessed by mass spectrometry using techniques known in the art such as accelerated mass spectrometry (AMS), Stable Isotope Ratio Mass Spectrometry (SIRMS) and Site-Specific Natural Isotopic Fractionation by Nuclear Magnetic Resonance (SNIF-NMR). Such mass spectral techniques can be integrated
- 15 with separation techniques such as liquid chromatography (LC), high performance liquid chromatography (HPLC) and/or gas chromatography, and the like.

In the case of carbon, ASTM D6866 was developed in the United States as a standardized analytical method for determining the biobased content of solid, liquid, and gaseous samples using radiocarbon dating by the American Society for Testing and Materials (ASTM)

- 20 International. The standard is based on the use of radiocarbon dating for the determination of a product's biobased content. ASTM D6866 was first published in 2004, and the current active version of the standard is ASTM D6866-11 (effective April 1, 2011). Radiocarbon dating techniques are well known to those skilled in the art, including those described herein.

- The biobased content of a compound is estimated by the ratio of carbon-14 (¹⁴C) to carbon-12 (¹²C). Specifically, the Fraction Modern (Fm) is computed from the expression: Fm = (S-B)/(M-B), where B, S and M represent the ¹⁴C/¹²C ratios of the blank, the sample and the modern reference, respectively. Fraction Modern is a measurement of the deviation of the ¹⁴C/¹²C ratio of a sample from "Modern." Modern is defined as 95% of the radiocarbon concentration (in AD 1950) of National Bureau of Standards (NBS) Oxalic Acid I (i.e.,
- 25 (12C). Specifically, the Fraction Modern (Fm) is computed from the expression: Fm = (S-B)/(M-B), where B, S and M represent the ¹⁴C/¹²C ratios of the blank, the sample and the modern reference, respectively. Fraction Modern is a measurement of the deviation of the ¹⁴C/¹²C ratio of a sample from "Modern." Modern is defined as 95% of the radiocarbon concentration (in AD 1950) of National Bureau of Standards (NBS) Oxalic Acid I (i.e.,
- 30 standard reference materials (SRM) 4990b) normalized to δ¹³C_{VPDB} = -19 per mil (Olsson, *The use of Oxalic acid as a Standard*. in, Radiocarbon Variations and Absolute Chronology,

Nobel Symposium, 12th Proc., John Wiley & Sons, New York (1970)). Mass spectrometry results, for example, measured by ASM, are calculated using the internationally agreed upon definition of 0.95 times the specific activity of NBS Oxalic Acid I (SRM 4990b) normalized to $\delta^{13}\text{C}_{\text{VPDB}} = -19$ per mil. This is equivalent to an absolute (AD 1950) $^{14}\text{C}/^{12}\text{C}$ ratio of $1.176 \pm 0.010 \times 10^{-12}$ (Karlen et al., *Arkiv Geofysik*, 4:465-471 (1968)). The standard calculations take into account the differential uptake of one isotope with respect to another, for example, the preferential uptake in biological systems of C^{12} over C^{13} over C^{14} , and these corrections are reflected as a Fm corrected for δ^{13} .

An oxalic acid standard (SRM 4990b or HOx 1) was made from a crop of 1955 sugar beet. Although there were 1000 lbs made, this oxalic acid standard is no longer commercially available. The Oxalic Acid II standard (HOx 2; N.I.S.T designation SRM 4990 C) was made from a crop of 1977 French beet molasses. In the early 1980's, a group of 12 laboratories measured the ratios of the two standards. The ratio of the activity of Oxalic acid II to 1 is 1.2933 ± 0.001 (the weighted mean). The isotopic ratio of HOx II is -17.8 per mille. ASTM D6866-11 suggests use of the available Oxalic Acid II standard SRM 4990 C (Hox2) for the modern standard (see discussion of original vs. currently available oxalic acid standards in Mann, *Radiocarbon*, 25(2):519-527 (1983)). A Fm = 0% represents the entire lack of carbon-14 atoms in a material, thus indicating a fossil (for example, petroleum based) carbon source. A Fm = 100%, after correction for the post-1950 injection of carbon-14 into the atmosphere from nuclear bomb testing, indicates an entirely modern carbon source. As described herein, such a "modern" source includes biobased sources.

As described in ASTM D6866, the percent modern carbon (pMC) can be greater than 100% because of the continuing but diminishing effects of the 1950s nuclear testing programs, which resulted in a considerable enrichment of carbon-14 in the atmosphere as described in ASTM D6866-11. Because all sample carbon-14 activities are referenced to a "pre-bomb" standard, and because nearly all new biobased products are produced in a post-bomb environment, all pMC values (after correction for isotopic fraction) must be multiplied by 0.95 (as of 2010) to better reflect the true biobased content of the sample. A biobased content that is greater than 103% suggests that either an analytical error has occurred, or that the source of biobased carbon is more than several years old.

ASTM D6866 quantifies the biobased content relative to the material's total organic content and does not consider the inorganic carbon and other non-carbon containing substances

present. For example, a product that is 50% starch-based material and 50% water would be considered to have a Biobased Content = 100% (50% organic content that is 100% biobased) based on ASTM D6866. In another example, a product that is 50% starch-based material, 25% petroleum-based, and 25% water would have a Biobased Content = 66.7% (75% organic content but only 50% of the product is biobased). In another example, a product that is 50% organic carbon and is a petroleum-based product would be considered to have a Biobased Content = 0% (50% organic carbon but from fossil sources). Thus, based on the well known methods and known standards for determining the biobased content of a compound or material, one skilled in the art can readily determine the biobased content and/or prepared downstream products that utilize of the invention having a desired biobased content.

Applications of carbon-14 dating techniques to quantify bio-based content of materials are known in the art (Currie et al., *Nuclear Instruments and Methods in Physics Research B*, 172:281-287 (2000)). For example, carbon-14 dating has been used to quantify bio-based content in terephthalate-containing materials (Colonna et al., *Green Chemistry*, 13:2543-2548 (2011)). Notably, polypropylene terephthalate (PPT) polymers derived from renewable 1,3-propanediol and petroleum-derived terephthalic acid resulted in Fm values near 30% (i.e., since 3/11 of the polymeric carbon derives from renewable 1,3-propanediol and 8/11 from the fossil end member terephthalic acid) (Currie et al., *supra*, 2000). In contrast, polybutylene terephthalate polymer derived from both renewable 1,4-butanediol and renewable terephthalic acid resulted in bio-based content exceeding 90% (Colonna et al., *supra*, 2011).

Accordingly, in some embodiments, the present invention provides ethylene glycol or an ethylene glycol intermediate that has a carbon-12, carbon-13, and carbon-14 ratio that reflects an atmospheric carbon, also referred to as environmental carbon, uptake source. For example, in some aspects the ethylene glycol or an ethylene glycol intermediate can have an Fm value of at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98% or as much as 100%. In some such embodiments, the uptake source is CO₂. In some embodiments, the present invention provides ethylene glycol or an ethylene glycol intermediate that has a carbon-12, carbon-13, and carbon-14 ratio that reflects petroleum-based carbon uptake source. In this aspect, the ethylene glycol or an ethylene glycol intermediate can have an Fm

value of less than 95%, less than 90%, less than 85%, less than 80%, less than 75%, less than 70%, less than 65%, less than 60%, less than 55%, less than 50%, less than 45%, less than 40%, less than 35%, less than 30%, less than 25%, less than 20%, less than 15%, less than 10%, less than 5%, less than 2% or less than 1%. In some embodiments, the present

- 5 invention provides ethylene glycol or an ethylene glycol intermediate that has a carbon-12, carbon-13, and carbon-14 ratio that is obtained by a combination of an atmospheric carbon uptake source with a petroleum-based uptake source. Using such a combination of uptake sources is one way by which the carbon-12, carbon-13, and carbon-14 ratio can be varied, and the respective ratios would reflect the proportions of the uptake sources.
- 10 Further, the present invention relates to the biologically produced ethylene glycol or ethylene glycol intermediate as disclosed herein, and to the products derived therefrom, wherein the ethylene glycol or an ethylene glycol intermediate has a carbon-12, carbon-13, and carbon-14 isotope ratio of about the same value as the CO₂ that occurs in the environment. For example, in some aspects the invention provides: bioderived ethylene glycol or a bioderived
- 15 ethylene glycol intermediate having a carbon-12 versus carbon-13 versus carbon-14 isotope ratio of about the same value as the CO₂ that occurs in the environment, or any of the other ratios disclosed herein. It is understood, as disclosed herein, that a product can have a carbon-12 versus carbon-13 versus carbon-14 isotope ratio of about the same value as the CO₂ that occurs in the environment, or any of the ratios disclosed herein, wherein the product
- 20 is generated from bioderived bioderived ethylene glycol or a bioderived ethylene glycol intermediate as disclosed herein, wherein the bioderived product is chemically modified to generate a final product. Methods of chemically modifying a bioderived product of ethylene glycol, or an intermediate thereof, to generate a desired product are well known to those skilled in the art, as described herein. The invention further provides antifreezes, coolants,
- 25 polyester fibers, fiberglass, resins or films having a carbon-12 versus carbon-13 versus carbon-14 isotope ratio of about the same value as the CO₂ that occurs in the environment, wherein the antifreezes, coolants, polyester fibers, fiberglass, resins or films are generated directly from or in combination with bioderived ethylene glycol or a bioderived ethylene glycol intermediate as disclosed herein.
- 30 Ethylene glycol is a chemical commonly used in many commercial and industrial applications. Non-limiting examples of such applications include production of antifreezes and coolants. Moreover, ethylene glycol is also used as a raw material in the production of a

wide range of products including polyester fibers for clothes, upholstery, carpet and pillows; fiberglass used in products such as jet skis, bathtubs, and bowling balls; and polyethylene terephthalate resin used in packaging film and bottles. Around 82% of ethylene glycol consumed worldwide is used in the production of polyester fibers, resins and films. The

5 second largest market for ethylene glycol is in the production of antifreeze formulations. Accordingly, in some embodiments, the invention provides biobased antifreezes, coolants, polyester fibers, fiberglass, resins or films comprising one or more bioderived ethylene glycol or bioderived ethylene glycol intermediate produced by a non-naturally occurring microorganism of the invention or produced using a method disclosed herein.

10 As used herein, the term "bioderived" means derived from or synthesized by a biological organism and can be considered a renewable resource since it can be generated by a biological organism. Such a biological organism, in particular the microbial organisms of the invention disclosed herein, can utilize feedstock or biomass, such as, sugars or carbohydrates obtained from an agricultural, plant, bacterial, or animal source. Alternatively, the biological

15 organism can utilize atmospheric carbon. As used herein, the term "biobased" means a product as described above that is composed, in whole or in part, of a bioderived compound of the invention. A biobased or bioderived product is in contrast to a petroleum derived product, wherein such a product is derived from or synthesized from petroleum or a petrochemical feedstock.

20 In some embodiments, the invention provides an antifreeze, coolant, polyester fiber, fiberglass, resin or film comprising bioderived ethylene glycol or bioderived ethylene glycol intermediate, wherein the bioderived ethylene glycol or bioderived ethylene glycol intermediate includes all or part of the ethylene glycol or ethylene glycol intermediate used in the production of the antifreeze, coolant, polyester fiber, fiberglass, resin or film. Thus, in

25 some aspects, the invention provides a biobased antifreeze, coolant, polyester fiber, fiberglass, resin or film comprising at least 2%, at least 3%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98% or 100% bioderived ethylene glycol or bioderived ethylene glycol intermediate as disclosed herein. Additionally,

30 in some aspects, the invention provides a biobased antifreeze, coolant, polyester fiber, fiberglass, resin or film wherein the ethylene glycol or ethylene glycol intermediate used in its production is a combination of bioderived and petroleum derived ethylene glycol or

ethylene glycol intermediate. For example, a biobased antifreeze, coolant, polyester fiber, fiberglass, resin or film can be produced using 50% bioderived ethylene glycol and 50% petroleum derived ethylene glycol or other desired ratios such as 60%/40%, 70%/30%, 80%/20%, 90%/10%, 95%/5%, 100%/0%, 40%/60%, 30%/70%, 20%/80%, 10%/90% of bioderived/petroleum derived precursors, so long as at least a portion of the product comprises a bioderived product produced by the microbial organisms disclosed herein. It is understood that methods for producing antifreeze, coolant, polyester fiber, fiberglass, resin or film using the bioderived ethylene glycol or bioderived ethylene glycol intermediate of the invention are well known in the art.

The invention is described herein with general reference to the metabolic reaction, reactant or product thereof, or with specific reference to one or more nucleic acids or genes encoding an enzyme associated with or catalyzing, or a protein associated with, the referenced metabolic reaction, reactant or product. Unless otherwise expressly stated herein, those skilled in the art will understand that reference to a reaction also constitutes reference to the reactants and products of the reaction. Similarly, unless otherwise expressly stated herein, reference to a reactant or product also references the reaction, and reference to any of these metabolic constituents also references the gene or genes encoding the enzymes that catalyze or proteins involved in the referenced reaction, reactant or product. Likewise, given the well known fields of metabolic biochemistry, enzymology and genomics, reference herein to a gene or encoding nucleic acid also constitutes a reference to the corresponding encoded enzyme and the reaction it catalyzes or a protein associated with the reaction as well as the reactants and products of the reaction.

As disclosed herein, the intermediates glycerate, tartarate semialdehyde, hydroxypyruvate and glyoxylate, as well as other intermediates, are carboxylic acids, which can occur in various ionized forms, including fully protonated, partially protonated, and fully deprotonated forms. Accordingly, the suffix “-ate,” or the acid form, can be used interchangeably to describe both the free acid form as well as any deprotonated form, in particular since the ionized form is known to depend on the pH in which the compound is found. It is understood that carboxylate products or intermediates includes ester forms of carboxylate products or pathway intermediates, such as O-carboxylate and S-carboxylate esters. O- and S-carboxylates can include lower alkyl, that is C1 to C6, branched or straight chain carboxylates. Some such O- or S-carboxylates include, without limitation, methyl, ethyl, n-

propyl, n-butyl, i-propyl, sec-butyl, and tert-butyl, pentyl, hexyl O- or S-carboxylates, any of which can further possess an unsaturation, providing for example, propenyl, butenyl, pentyl, and hexenyl O- or S-carboxylates. O-carboxylates can be the product of a biosynthetic pathway. Exemplary O-carboxylates accessed via biosynthetic pathways can include, without limitation, methyl glycerate, ethyl glycerate, n-propyl glycerate, methyl tartarate semialdehyde, ethyl tartarate semialdehyde, n-propyl tartarate semialdehyde, methyl hydroxypyruvate, ethyl hydroxypyruvate, n-propyl hydroxypyruvate, methyl glyoxylate, ethyl glyoxylate, and n-propyl glyoxylate. Other biosynthetically accessible O-carboxylates can include medium to long chain groups, that is C7-C22, O-carboxylate esters derived from fatty alcohols, such as heptyl, octyl, nonyl, decyl, undecyl, lauryl, tridecyl, myristyl, pentadecyl, cetyl, palmitoyl, heptadecyl, stearyl, nonadecyl, arachidyl, heneicosyl, and behenyl alcohols, any one of which can be optionally branched and/or contain unsaturations. O-carboxylate esters can also be accessed via a biochemical or chemical process, such as esterification of a free carboxylic acid product or transesterification of an O- or S-carboxylate. S-carboxylates are exemplified by CoA S-esters, cysteinyl S-esters, alkylthioesters, and various aryl and heteroaryl thioesters.

The non-naturally occurring microbial organisms of the invention can be produced by introducing expressible nucleic acids encoding one or more of the enzymes or proteins participating in one or more ethylene glycol biosynthetic pathways. Depending on the host microbial organism chosen for biosynthesis, nucleic acids for some or all of a particular ethylene glycol biosynthetic pathway can be expressed. For example, if a chosen host is deficient in one or more enzymes or proteins for a desired biosynthetic pathway, then expressible nucleic acids for the deficient enzyme(s) or protein(s) are introduced into the host for subsequent exogenous expression. Alternatively, if the chosen host exhibits endogenous expression of some pathway genes, but is deficient in others, then an encoding nucleic acid is needed for the deficient enzyme(s) or protein(s) to achieve ethylene glycol biosynthesis. Thus, a non-naturally occurring microbial organism of the invention can be produced by introducing exogenous enzyme or protein activities to obtain a desired biosynthetic pathway or a desired biosynthetic pathway can be obtained by introducing one or more exogenous enzyme or protein activities that, together with one or more endogenous enzymes or proteins, produces a desired product such as ethylene glycol.

Host microbial organisms can be selected from, and the non-naturally occurring microbial organisms generated in, for example, bacteria, yeast, fungus or any of a variety of other microorganisms applicable to fermentation processes. Exemplary bacteria include species selected from *Escherichia coli*, *Klebsiella oxytoca*, *Anaerobiospirillum succiniciproducens*,
5 *Actinobacillus succinogenes*, *Mannheimia succiniciproducens*, *Rhizobium etli*, *Bacillus subtilis*, *Corynebacterium glutamicum*, *Gluconobacter oxydans*, *Zymomonas mobilis*, *Lactococcus lactis*, *Lactobacillus plantarum*, *Streptomyces coelicolor*, *Clostridium acetobutylicum*, *Pseudomonas fluorescens*, and *Pseudomonas putida*. Exemplary yeasts or fungi include species selected from *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*,
10 *Kluyveromyces lactis*, *Kluyveromyces marxianus*, *Aspergillus terreus*, *Aspergillus niger*, *Pichia pastoris*, *Rhizopus arrhizus*, *Rhizobus oryzae*, *Yarrowia lipolytica*, and the like. *E. coli* is a particularly useful host organism since it is a well characterized microbial organism suitable for genetic engineering. Other particularly useful host organisms include yeast such as *Saccharomyces cerevisiae*. It is understood that any suitable microbial host organism can
15 be used to introduce metabolic and/or genetic modifications to produce a desired product.

Depending on the ethylene glycol biosynthetic pathway constituents of a selected host microbial organism, the non-naturally occurring microbial organisms of the invention will include at least one exogenously expressed ethylene glycol pathway-encoding nucleic acid and up to all encoding nucleic acids for one or more ethylene glycol biosynthetic pathways.

20 For example, ethylene glycol biosynthesis can be established in a host deficient in a pathway enzyme or protein through exogenous expression of the corresponding encoding nucleic acid. In a host deficient in all enzymes or proteins of an ethylene glycol pathway, exogenous expression of all enzyme or proteins in the pathway can be included, although it is understood that all enzymes or proteins of a pathway can be expressed even if the host contains at least
25 one of the pathway enzymes or proteins. For example, exogenous expression of all enzymes or proteins in a pathway for production of ethylene glycol can be included, such as, a serine aminotransferase, a serine oxidoreductase (deaminating), a hydroxypyruvate decarboxylase, and a glycolaldehyde reductase.

Given the teachings and guidance provided herein, those skilled in the art will understand that
30 the number of encoding nucleic acids to introduce in an expressible form will, at least, parallel the ethylene glycol pathway deficiencies of the selected host microbial organism. Therefore, a non-naturally occurring microbial organism of the invention can have one, two,

three, four, five, six, seven, eight, nine or ten up to all nucleic acids encoding the enzymes or proteins constituting an ethylene glycol biosynthetic pathway disclosed herein. In some embodiments, the non-naturally occurring microbial organisms also can include other genetic modifications that facilitate or optimize ethylene glycol biosynthesis or that confer other useful functions onto the host microbial organism. One such other functionality can include, for example, augmentation of the synthesis of one or more of the ethylene glycol pathway precursors such as glycolaldehyde, hydroxypyruvate, ethanolamine, glycerate, tartarate semialdehyde, glycolate, glycolyl-CoA or glycolylphosphate.

Generally, a host microbial organism is selected such that it produces the precursor of an ethylene glycol pathway, either as a naturally produced molecule or as an engineered product that either provides de novo production of a desired precursor or increased production of a precursor naturally produced by the host microbial organism. For example, serine is produced naturally in a host organism such as *E. coli*. A host organism can be engineered to increase production of a precursor, as disclosed herein. In addition, a microbial organism that has been engineered to produce a desired precursor can be used as a host organism and further engineered to express enzymes or proteins of an ethylene glycol pathway.

In some embodiments, a non-naturally occurring microbial organism of the invention is generated from a host that contains the enzymatic capability to synthesize ethylene glycol. In this specific embodiment it can be useful to increase the synthesis or accumulation of an ethylene glycol pathway product to, for example, drive ethylene glycol pathway reactions toward ethylene glycol production. Increased synthesis or accumulation can be accomplished by, for example, overexpression of nucleic acids encoding one or more of the above-described ethylene glycol pathway enzymes or proteins. Overexpression of the enzyme or enzymes and/or protein or proteins of the ethylene glycol pathway can occur, for example, through exogenous expression of the endogenous gene or genes, or through exogenous expression of the heterologous gene or genes. Therefore, naturally occurring organisms can be readily generated to be non-naturally occurring microbial organisms of the invention, for example, producing ethylene glycol, through overexpression of one, two, three, four five, six, seven, eight, nine or 10, that is, up to all nucleic acids encoding ethylene glycol biosynthetic pathway enzymes or proteins. In addition, a non-naturally occurring organism can be generated by mutagenesis of an endogenous gene that results in an increase in activity of an enzyme in the ethylene glycol biosynthetic pathway.

In particularly useful embodiments, exogenous expression of the encoding nucleic acids is employed. Exogenous expression confers the ability to custom tailor the expression and/or regulatory elements to the host and application to achieve a desired expression level that is controlled by the user. However, endogenous expression also can be utilized in other
5 embodiments such as by removing a negative regulatory effector or induction of the gene's promoter when linked to an inducible promoter or other regulatory element. Thus, an endogenous gene having a naturally occurring inducible promoter can be up-regulated by providing the appropriate inducing agent, or the regulatory region of an endogenous gene can be engineered to incorporate an inducible regulatory element, thereby allowing the regulation
10 of increased expression of an endogenous gene at a desired time. Similarly, an inducible promoter can be included as a regulatory element for an exogenous gene introduced into a non-naturally occurring microbial organism.

It is understood that, in methods of the invention, any of the one or more exogenous nucleic acids can be introduced into a microbial organism to produce a non-naturally occurring
15 microbial organism of the invention. The nucleic acids can be introduced so as to confer, for example, an ethylene glycol biosynthetic pathway onto the microbial organism.

Alternatively, encoding nucleic acids can be introduced to produce an intermediate microbial organism having the biosynthetic capability to catalyze some of the required reactions to confer ethylene glycol biosynthetic capability. For example, a non-naturally occurring
20 microbial organism having an ethylene glycol biosynthetic pathway can comprise at least two exogenous nucleic acids encoding desired enzymes or proteins, such as the combination of a hydroxypyruvate decarboxylase and a glycolaldehyde reductase, or alternatively a serine decarboxylase and an ethanolamine oxidoreductase (deaminating), or alternatively a glyoxylate carboligase and a hydroxypyruvate isomerase, or alternatively a glycolyl-CoA
25 reductase (aldehyde forming) and a glycolaldehyde reductase, or alternatively 2-phosphoglycerate phosphatase and glycoaldehyde reductase and the like. Thus, it is understood that any combination of two or more enzymes or proteins of a biosynthetic pathway can be included in a non-naturally occurring microbial organism of the invention. Similarly, it is understood that any combination of three or more enzymes or proteins of a
30 biosynthetic pathway can be included in a non-naturally occurring microbial organism of the invention, for example, a serine oxidoreductase (deaminating), a hydroxypyruvate decarboxylase, and a glycolaldehyde reductase, or alternatively, a glycerate kinase; a hydroxypyruvate reductase and a hydroxypyruvate decarboxylase, or alternatively a 3-

phosphoglycerate phosphatase, a glycerate kinase and a glycerate decarboxylase, or alternatively a glyoxylate carboligase, a hydroxypyruvate isomerase and a hydroxypyruvate decarboxylase, or alternatively a glycolyl-CoA transferase, a glycolyl-CoA reductase (aldehyde forming) and a glycolaldehyde reductase, or alternatively a glyoxylate reductase, a glycolyl-CoA transferase and a glycolyl-CoA reductase (alcohol forming), or alternatively a 3-phosphohydroxypyruvate decarboxylase, a 2-oxoethyl-phosphate reductase, and a 2-hydroxyethyl-phosphate phosphatase, and so forth, as desired, so long as the combination of enzymes and/or proteins of the desired biosynthetic pathway results in production of the corresponding desired product. Similarly, any combination of four, a serine decarboxylase, an ethanolamine aminotransferase, an ethanolamine oxidoreductase (deaminating) and a glycolaldehyde reductase, or alternatively a 3-phosphoglycerate phosphatase, a hydroxypyruvate reductase, a hydroxypyruvate decarboxylase and a glycolaldehyde reductase, or alternatively a glyoxylate carboligase, a hydroxypyruvate isomerase, a hydroxypyruvate decarboxylase and a glycolaldehyde reductase, or alternatively a glyoxylate reductase, glycolate kinase, a glycolylphosphate reductase and a glycolaldehyde reductase, or alternatively a glyceraldehyde dehydrogenase, a glycerate dehydrogenase, and a hydroxypyruvate decarboxylase or more enzymes or proteins of a biosynthetic pathway as disclosed herein can be included in a non-naturally occurring microbial organism of the invention, as desired, so long as the combination of enzymes and/or proteins of the desired biosynthetic pathway results in production of the corresponding desired product.

In addition to the biosynthesis of ethylene glycol as described herein, the non-naturally occurring microbial organisms and methods of the invention also can be utilized in various combinations with each other and with other microbial organisms and methods well known in the art to achieve product biosynthesis by other routes. For example, one alternative to produce ethylene glycol other than use of the ethylene glycol producers is through addition of another microbial organism capable of converting an ethylene glycol pathway intermediate to ethylene glycol. One such procedure includes, for example, the fermentation of a microbial organism that produces an ethylene glycol pathway intermediate. The ethylene glycol pathway intermediate can then be used as a substrate for a second microbial organism that converts the ethylene glycol pathway intermediate to ethylene glycol. The ethylene glycol pathway intermediate can be added directly to another culture of the second organism or the original culture of the ethylene glycol pathway intermediate producers can be depleted of these microbial organisms by, for example, cell separation, and then subsequent addition of

the second organism to the fermentation broth can be utilized to produce the final product without intermediate purification steps.

In other embodiments, the non-naturally occurring microbial organisms and methods of the invention can be assembled in a wide variety of subpathways to achieve biosynthesis of, for example, ethylene glycol. In these embodiments, biosynthetic pathways for a desired product of the invention can be segregated into different microbial organisms, and the different microbial organisms can be co-cultured to produce the final product. In such a biosynthetic scheme, the product of one microbial organism is the substrate for a second microbial organism until the final product is synthesized. For example, the biosynthesis of ethylene glycol can be accomplished by constructing a microbial organism that contains biosynthetic pathways for conversion of one pathway intermediate to another pathway intermediate or the product. Alternatively, ethylene glycol also can be biosynthetically produced from microbial organisms through co-culture or co-fermentation using two organisms in the same vessel, where the first microbial organism produces an ethylene glycol intermediate and the second microbial organism converts the intermediate to ethylene glycol.

Given the teachings and guidance provided herein, those skilled in the art will understand that a wide variety of combinations and permutations exist for the non-naturally occurring microbial organisms and methods of the invention together with other microbial organisms, with the co-culture of other non-naturally occurring microbial organisms having subpathways and with combinations of other chemical and/or biochemical procedures well known in the art to produce ethylene glycol.

Sources of encoding nucleic acids for an ethylene glycol pathway enzyme or protein can include, for example, any species where the encoded gene product is capable of catalyzing the referenced reaction. Such species include both prokaryotic and eukaryotic organisms including, but not limited to, bacteria, including archaea and eubacteria, and eukaryotes, including yeast, plant, insect, animal, and mammal, including human. Exemplary species for such sources include, for example, *Escherichia coli*, *Rattus norvegicus*, *Homo sapiens*, *Drosophila melanogaster*, *Mus musculus*, *Sus scrofa*, *Arabidopsis thaliana*, *Oryza sativa*, *Hyphomicrobium methylovorum*, *Methylobacterium extorquens*, *Thermotoga maritima*, *Halobacterium salinarum*, *Lactococcus lactis*, *Saccharomyces cerevisiae*, *Zymomonas mobilis*, *Acinetobacter* sp. Strain M-1, *Brassica napus*, *Beta vulgaris*, *Geobacillus stearothermophilus*, *Agrobacterium tumefaciens*, *Acinetobacter calcoaceticus*, *Acinetobacter*

baylyi, *Achromobacter denitrificans*, *Streptococcus thermophilus*, *Bacillus brevis*, *Bacillus subtilis*, *Bacillus megaterium*, *Enterobacter aerogenes*, *Ralstonia eutropha*, *Salmonella enterica*, *Salmonella typhimurium*, *Burkholderia ambifaria*, *Acidaminococcus fermentans*, *Archaeoglobus fulgidus*, *Haloarcula marismortui*, *Pyrobaculum aerophilum* str. IM2,

5 *Pseudomonas putida*, *Pseudomonas* sp, *Rhizobium leguminosarum*, *Clostridium kluyveri*, *Clostridium saccharoperbutylacetonicum*, *Clostridium acetobutylicum*, *Clostridium beijerinckii*, *Porphyromonas gingivalis*, *Leuconostoc mesenteroides*, *Metallosphaera sedula*, *Sulfolobus tokodaii*, *Sulfolobus solfataricus*, *Sulfolobus acidocaldarius*, *Nocardia iowensis*, *Streptomyces griseus*, *Candida albicans*, *Schizosaccharomyces pombe*, *Penicillium*

10 *chrysogenum*, butyrate-producing bacterium L2-50, *Haemophilus influenzae*, *Mycobacterium tuberculosis*, *Vibrio cholera*, *Helicobacter pylori*, *Campylobacter jejuni*, *Leuconostoc mesenteroides*, *Chloroflexus aurantiacus*, *Roseiflexus castenholzii*, *Erythrobacter* sp. NAP1, marine gamma proteobacterium HTCC2080, *Simmondsia chinensis*, *Azospirillum brasilense*, *Bos Taurus*, *Clostridium kluyveri* DSM 555, *Geobacillus thermoglucosidasius*,

15 *Methanocaldococcus jannaschii*, *Oryctolagus cuniculus*, *Oryza sativa*, *Phaseolus vulgaris*, *Picrophilus torridus*, *Pseudomonas aeruginosa*, *Pyrococcus furiosus*, *Ralstonia eutropha* H16, *Staphylococcus aureus*, *Thermoproteus tenax*, *Thermus thermophilus*, *Zea mays*, *Acetobacter pasteurians*, *Aspergillus niger*, *Aspergillus terreus* NIH2624, *Bacteroides fragilis*, *Bradyrhizobium japonicum* USDA110, *Chlamydomonas reinhardtii*, *Enterobacter*

20 *cloacae*, *Kluyveromyces lactis*, *Mesorhizobium loti*, *Nicotiana tabacum*, *Pseudomonas fluorescens*, *Pseudomonas stutzeri*, *Streptomyces luridus*, *Streptomyces viridochromogenes*, *Streptomyces wedmorensis*, *Synechococcus* sp. PCC 7002, and *Thermoplasma acidophilum* as well as other exemplary species disclosed herein or available as source organisms for corresponding genes. However, with the complete genome sequence available for now more

25 than 550 species (with more than half of these available on public databases such as the NCBI), including 395 microorganism genomes and a variety of yeast, fungi, plant, and mammalian genomes, the identification of genes encoding the requisite ethylene glycol biosynthetic activity for one or more genes in related or distant species, including for example, homologues, orthologs, paralogs and nonorthologous gene displacements of known

30 genes, and the interchange of genetic alterations between organisms is routine and well known in the art. Accordingly, the metabolic alterations allowing biosynthesis of ethylene glycol described herein with reference to a particular organism such as *E. coli* can be readily applied to other microorganisms, including prokaryotic and eukaryotic organisms alike.

Given the teachings and guidance provided herein, those skilled in the art will know that a metabolic alteration exemplified in one organism can be applied equally to other organisms.

In some instances, such as when an alternative ethylene glycol biosynthetic pathway exists in an unrelated species, ethylene glycol biosynthesis can be conferred onto the host species by, for example, exogenous expression of a paralog or paralogs from the unrelated species that catalyzes a similar, yet non-identical metabolic reaction to replace the referenced reaction. Because certain differences among metabolic networks exist between different organisms, those skilled in the art will understand that the actual gene usage between different organisms may differ. However, given the teachings and guidance provided herein, those skilled in the art also will understand that the teachings and methods of the invention can be applied to all microbial organisms using the cognate metabolic alterations to those exemplified herein to construct a microbial organism in a species of interest that will synthesize ethylene glycol.

Methods for constructing and testing the expression levels of a non-naturally occurring ethylene glycol-producing host can be performed, for example, by recombinant and detection methods well known in the art. Such methods can be found described in, for example, Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Third Ed., Cold Spring Harbor Laboratory, New York (2001); and Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley and Sons, Baltimore, MD (1999).

Exogenous nucleic acid sequences involved in a pathway for production of ethylene glycol can be introduced stably or transiently into a host cell using techniques well known in the art including, but not limited to, conjugation, electroporation, chemical transformation, transduction, transfection, and ultrasound transformation. For exogenous expression in *E. coli* or other prokaryotic cells, some nucleic acid sequences in the genes or cDNAs of eukaryotic nucleic acids can encode targeting signals such as an N-terminal mitochondrial or other targeting signal, which can be removed before transformation into prokaryotic host cells, if desired. For example, removal of a mitochondrial leader sequence led to increased expression in *E. coli* (Hoffmeister et al., *J. Biol. Chem.* 280:4329-4338 (2005)). For exogenous expression in yeast or other eukaryotic cells, genes can be expressed in the cytosol without the addition of leader sequence, or can be targeted to mitochondrion or other organelles, or targeted for secretion, by the addition of a suitable targeting sequence such as a mitochondrial targeting or secretion signal suitable for the host cells. Thus, it is understood that appropriate modifications to a nucleic acid sequence to remove or include a targeting

sequence can be incorporated into an exogenous nucleic acid sequence to impart desirable properties. Furthermore, genes can be subjected to codon optimization with techniques well known in the art to achieve optimized expression of the proteins.

An expression vector or vectors can be constructed to include one or more ethylene glycol

5 biosynthetic pathway encoding nucleic acids as exemplified herein operably linked to expression control sequences functional in the host organism. Expression vectors applicable for use in the microbial host organisms of the invention include, for example, plasmids, phage vectors, viral vectors, episomes and artificial chromosomes, including vectors and selection sequences or markers operable for stable integration into a host chromosome.

10 Additionally, the expression vectors can include one or more selectable marker genes and appropriate expression control sequences. Selectable marker genes also can be included that, for example, provide resistance to antibiotics or toxins, complement auxotrophic deficiencies, or supply critical nutrients not in the culture media. Expression control sequences can include constitutive and inducible promoters, transcription enhancers, transcription

15 terminators, and the like which are well known in the art. When two or more exogenous encoding nucleic acids are to be co-expressed, both nucleic acids can be inserted, for example, into a single expression vector or in separate expression vectors. For single vector expression, the encoding nucleic acids can be operationally linked to one common expression control sequence or linked to different expression control sequences, such as one inducible

20 promoter and one constitutive promoter. The transformation of exogenous nucleic acid sequences involved in a metabolic or synthetic pathway can be confirmed using methods well known in the art. Such methods include, for example, nucleic acid analysis such as Northern blots or polymerase chain reaction (PCR) amplification of mRNA, or immunoblotting for expression of gene products, or other suitable analytical methods to test the expression of an

25 introduced nucleic acid sequence or its corresponding gene product. It is understood by those skilled in the art that the exogenous nucleic acid is expressed in a sufficient amount to produce the desired product, and it is further understood that expression levels can be optimized to obtain sufficient expression using methods well known in the art and as disclosed herein.

30 In some embodiments, the invention provides a method for producing ethylene glycol that includes culturing a non-naturally occurring microbial organism, including a microbial organism having an ethylene glycol pathway, the ethylene glycol pathway including at least

one exogenous nucleic acid encoding an ethylene glycol pathway enzyme expressed in a sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a serine aminotransferase, a serine oxidoreductase (deaminating), a hydroxypyruvate decarboxylase, a glycolaldehyde reductase, a serine decarboxylase, an ethanolamine aminotransferase, an ethanolamine oxidoreductase (deaminating), a hydroxypyruvate reductase or a glycerate decarboxylase (see steps 1-9 of Figure 1). In one aspect, the method includes a microbial organism having an ethylene glycol pathway including a serine aminotransferase or a serine oxidoreductase (deaminating); a hydroxypyruvate decarboxylase, and a glycolaldehyde reductase (see steps 1/2, 3 and 4 of Figure 1). In one aspect, the method includes a microbial organism having an ethylene glycol pathway including a serine aminotransferase or a serine oxidoreductase (deaminating); a hydroxypyruvate reductase, and a glycerate decarboxylase (see steps 1/2, 8, and 9 of Figure 1). In one aspect, the method includes a microbial organism having an ethylene glycol pathway including a serine decarboxylase; an ethanolamine aminotransferase or an ethanolamine oxidoreductase (deaminating), and a glycolaldehyde reductase (see steps 5, 6/7 and 4 of Figure 1).

In some embodiments, the invention provides a method for producing ethylene glycol that includes culturing a non-naturally occurring microbial organism, including a microbial organism having an ethylene glycol pathway, the ethylene glycol pathway including at least one exogenous nucleic acid encoding an ethylene glycol pathway enzyme expressed in a sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a hydroxypyruvate decarboxylase, glycolaldehyde reductase, a hydroxypyruvate reductase, a glycerate decarboxylase, a 3-phosphoglycerate phosphatase, a glycerate kinase, a 2-phosphoglycerate phosphatase, a glycerate-2-kinase or a glyceraldehyde dehydrogenase (see steps 3, 4, and 8-14 of Figure 1). In one aspect, the method includes a microbial organism having an ethylene glycol pathway including a hydroxypyruvate reductase, a hydroxypyruvate decarboxylase and a glycolaldehyde reductase (see steps 8, 3 and 4 of Figure 1). In one aspect, the method includes a microbial organism having an ethylene glycol pathway including a 3-phosphoglycerate phosphatase or a glycerate kinase; a hydroxypyruvate reductase; a hydroxypyruvate decarboxylase, and a glycolaldehyde reductase (see steps 10/11, 8, 3 and 4 of Figure 1). In one aspect, the method includes a microbial organism having an ethylene glycol pathway including a 2-phosphoglycerate phosphatase or a glycerate-2-kinase; a hydroxypyruvate reductase; a hydroxypyruvate decarboxylase, and a glycolaldehyde reductase (see steps 12/13, 8, 3 and 4 of Figure 1). In

one aspect, the method includes a microbial organism having an ethylene glycol pathway including a glyceraldehyde dehydrogenase, a hydroxypyruvate reductase; a hydroxypyruvate decarboxylase, and a glycolaldehyde reductase (see steps 14, 8, 3 and 4 of Figure 1).

5 In some embodiments, the invention provides a method for producing ethylene glycol that includes culturing a non-naturally occurring microbial organism, including a microbial organism having an ethylene glycol pathway, the ethylene glycol pathway including at least one exogenous nucleic acid encoding an ethylene glycol pathway enzyme expressed in a sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a glycerate decarboxylase (see step 9 of Figure 1). In one aspect, the method includes a
10 microbial organism having an ethylene glycol pathway including a 3-phosphoglycerate phosphatase or a glycerate kinase and a glycerate decarboxylase (see steps 10/11 and 9 of Figure 1). In one aspect, the method includes a microbial organism having an ethylene glycol pathway including a 2-phosphoglycerate phosphatase, a glycerate-2-kinase and a glycerate decarboxylase (see steps 12/13, 9 of Figure 1). In one aspect, the method includes a
15 microbial organism having an ethylene glycol pathway including a glyceraldehyde dehydrogenase and a glycerate decarboxylase (see steps 14 and 9 of Figure 1).

In some embodiments, the invention provides a method for producing ethylene glycol that includes culturing a non-naturally occurring microbial organism, including a microbial organism having an ethylene glycol pathway, the ethylene glycol pathway including at least
20 one exogenous nucleic acid encoding an ethylene glycol pathway enzyme expressed in a sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a glyoxylate carboligase, a hydroxypyruvate isomerase, a hydroxypyruvate decarboxylase, a glycolaldehyde reductase or a glycerate dehydrogenase (see steps 1, 2, 3, 4 or 5 of Figure 2). In one aspect, the method includes a microbial organism having an ethylene glycol pathway
25 including a glyoxylate carboligase, a hydroxypyruvate isomerase, a hydroxypyruvate decarboxylase and a glycolaldehyde reductase (see steps 1, 2, 3 and 4 of Figure 2). In one aspect, the method includes a microbial organism having an ethylene glycol pathway including a glycerate dehydrogenase, a hydroxypyruvate isomerase, a hydroxypyruvate decarboxylase and a glycolaldehyde reductase (see steps 5, 2, 3 and 4 of Figure 2). In one
30 aspect, the method includes a microbial organism having an ethylene glycol pathway including a 3-phosphoglycerate phosphatase or a glycerate kinase; a glycerate dehydrogenase; a hydroxypyruvate isomerase; a hydroxypyruvate decarboxylase; and a

glycolaldehyde reductase (see steps 10/11 of Figure 1 and steps 5, 2, 3 and 4 of Figure 2). In one aspect, the method includes a microbial organism having an ethylene glycol pathway including a 2-phosphoglycerate phosphatase or a glycerate-2-kinase; a glycerate

dehydrogenase; a hydroxypyruvate isomerase; a hydroxypyruvate decarboxylase; and a

- 5 glycolaldehyde reductase (see steps 12/13 of Figure 1 and steps 5, 2, 3 and 4 of Figure 2). In one aspect, the method includes a microbial organism having an ethylene glycol pathway including a glyceraldehyde dehydrogenase, a glycerate dehydrogenase, a hydroxypyruvate isomerase, a hydroxypyruvate decarboxylase and a glycolaldehyde reductase (see step 14 of Figure 1 and steps 5, 2, 3 and 4 of Figure 2).

- 10 In some embodiments, the invention provides a method for producing ethylene glycol that includes culturing a non-naturally occurring microbial organism, including a microbial organism having an ethylene glycol pathway, the ethylene glycol pathway including at least one exogenous nucleic acid encoding an ethylene glycol pathway enzyme expressed in a sufficient amount to produce ethylene glycol, the ethylene glycol pathway including a
- 15 glyoxylate reductase, a glycolyl-CoA transferase, a glycolyl-CoA synthetase, a glycolyl-CoA reductase (aldehyde forming), a glycolaldehyde reductase, a glycolate reductase, a glycolate kinase, a phosphotransglycolylase, a glycolylphosphate reductase or a glycolyl-CoA reductase (alcohol forming) (see steps 1-10 of Figure 3). In one aspect, the method includes a microbial organism having an ethylene glycol pathway including a glyoxylate reductase; a
- 20 glycolyl-CoA transferase or a glycolyl-CoA synthetase; a glycolyl-CoA reductase (aldehyde forming), and a glycolaldehyde reductase (see steps 1, 2/3, 4 and 5 of Figure 3). In one aspect, the method includes a microbial organism having an ethylene glycol pathway including a glyoxylate reductase; a glycolate reductase, and a glycolaldehyde reductase (see steps 1, 6 and 5 of Figure 3). In one aspect, the method includes a microbial organism having
- 25 an ethylene glycol pathway including a glyoxylate reductase; a glycolyl-CoA transferase or a glycolyl-CoA synthetase, and a glycolyl-CoA reductase (alcohol forming) (see steps 1, 2/3 and 10 of Figure 3). In one aspect, the method includes a microbial organism having an ethylene glycol pathway including a glyoxylate reductase, a glycolate kinase, a phosphotransglycolylase, glycolyl-CoA reductase (aldehyde forming) and a glycolaldehyde
- 30 reductase (see steps 1, 7, 8, 4 and 5 of Figure 3). In one aspect, the method includes a microbial organism having an ethylene glycol pathway including a glyoxylate reductase, a glycolate kinase, a phosphotransglycolylase and a glycolyl-CoA reductase (alcohol forming) (see steps 1, 7, 8 and 10 of Figure 3). In one aspect, the method includes a microbial

organism having an ethylene glycol pathway including a glyoxylate reductase, glycolate kinase, a glycolylphosphate reductase and a glycolaldehyde reductase (see steps 1, 7, 9 and 5 of Figure 3).

5 In some embodiments, the invention provides a method for producing ethylene glycol that includes culturing a non-naturally occurring microbial organism, including a microbial organism having an ethylene glycol pathway and having at least one exogenous nucleic acid encoding an ethylene glycol pathway enzyme expressed in a sufficient amount to produce ethylene glycol, wherein the ethylene glycol pathway comprises a pathway selected from: (A) a 3-phosphohydroxypyruvate decarboxylase, a 2-oxoethyl-phosphate reductase, and a 2-
10 hydroxyethyl-phosphate phosphatase or an ethylene glycol kinase (see steps A, B and C of Figure 14); and (B) a 3-phosphohydroxypyruvate decarboxylase, a 2-oxoethyl-phosphate phosphatase or a glycolaldehyde kinase, and a glycolaldehyde reductase (see steps A, D and E of Figure 14).

15 In some embodiments, the invention provides a method for producing ethylene glycol that includes culturing a non-naturally occurring microbial organism, including a microbial organism having at least one exogenous nucleic acid encoding an ethylene glycol pathway enzyme expressed in a sufficient amount to produce ethylene glycol, wherein the non-naturally occurring microbial organism includes (i) a reductive TCA pathway comprising at least one exogenous nucleic acid encoding a reductive TCA pathway enzyme, wherein the at
20 least one exogenous nucleic acid is selected from an ATP-citrate lyase, a citrate lyase, a citryl-CoA synthetase, a citryl-CoA lyase, a fumarate reductase, and an alpha-ketoglutarate:ferredoxin oxidoreductase; (ii) a reductive TCA pathway comprising at least one exogenous nucleic acid encoding a reductive TCA pathway enzyme, wherein the at least one exogenous nucleic acid is selected from a pyruvate:ferredoxin oxidoreductase, a
25 phosphoenolpyruvate carboxylase, a phosphoenolpyruvate carboxykinase, a CO dehydrogenase, and an H₂ hydrogenase; or (iii) at least one exogenous nucleic acid encodes an enzyme selected from a CO dehydrogenase, an H₂ hydrogenase, and combinations thereof. Such a microbial organism can comprise an ethylene glycol pathway selected, for example, from (A) a serine aminotransferase or a serine oxidoreductase (deaminating); a
30 hydroxypyruvate decarboxylase, and a glycolaldehyde reductase; (B) a serine aminotransferase or a serine oxidoreductase (deaminating); a hydroxypyruvate reductase, and a glycerate decarboxylase; (C) a serine decarboxylase; an ethanolamine aminotransferase or

an ethanolamine oxidoreductase (deaminating), and a glycolaldehyde reductase; (D) a hydroxypyruvate reductase; a hydroxypyruvate decarboxylase, and a glycolaldehyde reductase; (E) a glycerate decarboxylase; (F) glycerate dehydrogenase, a hydroxypyruvate isomerase, a hydroxypyruvate decarboxylase and a glycolaldehyde reductase; (G) a glyoxylate carboligase, a hydroxypyruvate isomerase, a hydroxypyruvate decarboxylase and a glycolaldehyde reductase; (H) a glyoxylate reductase; a glycolyl-CoA transferase or a glycolyl-CoA synthetase; a glycolyl-CoA reductase (aldehyde forming), and a glycolaldehyde reductase; (I) a glyoxylate reductase; a glycolate reductase, and a glycolaldehyde reductase; (J) a glyoxylate reductase; a glycolyl-CoA transferase or a glycolyl-CoA synthetase, and a glycolyl-CoA reductase (alcohol forming); (K) a glyoxylate reductase, a glycolate kinase, a phosphotransglycolylase, glycolyl-CoA reductase (aldehyde forming) and a glycolaldehyde reductase; (L) a glyoxylate reductase, a glycolate kinase, a phosphotransglycolylase and a glycolyl-CoA reductase (alcohol forming); (M) a glyoxylate reductase, glycolate kinase, a glycolylphosphate reductase and a glycolaldehyde reductase; (N) a 3-phosphohydroxypyruvate decarboxylase, a 2-oxoethyl-phosphate reductase, and a 2-hydroxyethyl-phosphate phosphatase or an ethylene glycol kinase; and (O) a 3-phosphohydroxypyruvate decarboxylase, a 2-oxoethyl-phosphate phosphatase or a glycolaldehyde kinase, and a glycolaldehyde reductase.

In a particular embodiment, the invention provides a method for producing ethylene glycol that includes culturing a non-naturally occurring microbial organism, wherein the microbial organism having the ethylene glycol pathway of (D), (E) and/or (F) can further include a 3-phosphoglycerate phosphatase, a glycerate kinase, a 2-phosphoglycerate phosphatase, a glycerate-2-kinase, or a glyceraldehyde dehydrogenase. In another embodiment, a microbial organism having (i) can further include an exogenous nucleic acid encoding an enzyme selected from a pyruvate:ferredoxin oxidoreductase, an aconitase, an isocitrate dehydrogenase, a succinyl-CoA synthetase, a succinyl-CoA transferase, a fumarase, a malate dehydrogenase, an acetate kinase, a phosphotransacetylase, an acetyl-CoA synthetase, an NAD(P)H:ferredoxin oxidoreductase, ferredoxin, and combinations thereof. In still a further embodiment, a microbial organism having (ii) can further include an exogenous nucleic acid encoding an enzyme selected from an aconitase, an isocitrate dehydrogenase, a succinyl-CoA synthetase, a succinyl-CoA transferase, a fumarase, a malate dehydrogenase, and combinations thereof.

In a particular embodiment, the invention provides a method for producing ethylene glycol that includes culturing a non-naturally occurring microbial organism, wherein the microbial organism can include exogenous nucleic acids encoding each of the enzymes selected from:

(A) a serine aminotransferase or a serine oxidoreductase (deaminating); a hydroxypyruvate decarboxylase, and a glycolaldehyde reductase; (B) a serine aminotransferase or a serine oxidoreductase (deaminating); a hydroxypyruvate reductase, and a glycerate decarboxylase; (C) a serine decarboxylase; an ethanolamine aminotransferase or an ethanolamine oxidoreductase (deaminating), and a glycolaldehyde reductase; (D) a hydroxypyruvate reductase; a hydroxypyruvate decarboxylase, and a glycolaldehyde reductase; (E) a glycerate decarboxylase; (F) glycerate dehydrogenase, a hydroxypyruvate isomerase, a hydroxypyruvate decarboxylase and a glycolaldehyde reductase; (G) a glyoxylate carboligase, a hydroxypyruvate isomerase, a hydroxypyruvate decarboxylase and a glycolaldehyde reductase; (H) a glyoxylate reductase; a glycolyl-CoA transferase or a glycolyl-CoA synthetase; a glycolyl-CoA reductase (aldehyde forming), and a glycolaldehyde reductase; (I) a glyoxylate reductase; a glycolate reductase, and a glycolaldehyde reductase; (J) a glyoxylate reductase; a glycolyl-CoA transferase or a glycolyl-CoA synthetase, and a glycolyl-CoA reductase (alcohol forming); (K) a glyoxylate reductase, a glycolate kinase, a phosphotransglycolylase, glycolyl-CoA reductase (aldehyde forming) and a glycolaldehyde reductase; (L) a glyoxylate reductase, a glycolate kinase, a phosphotransglycolylase and a glycolyl-CoA reductase (alcohol forming); (M) a glyoxylate reductase, glycolate kinase, a glycolylphosphate reductase and a glycolaldehyde reductase; (N) a 3-phosphohydroxypyruvate decarboxylase, a 2-oxoethyl-phosphate reductase, and a 2-hydroxyethyl-phosphate phosphatase or an ethylene glycol kinase; and (O) a 3-phosphohydroxypyruvate decarboxylase, a 2-oxoethyl-phosphate phosphatase or a glycolaldehyde kinase, and a glycolaldehyde reductase. In one embodiment, such a microbial organism can include two, three, four or five exogenous nucleic acids each encoding enzymes of (i), (ii) or (iii).

For example, a method for producing ethylene glycol can include culturing a microorganism wherein the microbial organism comprising (i) includes three exogenous nucleic acids encoding ATP-citrate lyase or citrate lyase; a fumarate reductase; and an alpha-ketoglutarate:ferredoxin oxidoreductase; microbial organism comprising (ii) includes four exogenous nucleic acids encoding a pyruvate:ferredoxin oxidoreductase; a phosphoenolpyruvate carboxylase or a phosphoenolpyruvate carboxykinase; a CO

dehydrogenase; and an H₂ hydrogenase; or a microbial organism comprising (iii) includes two exogenous nucleic acids encoding CO dehydrogenase and H₂ hydrogenase.

Suitable purification and/or assays to test for the production of ethylene glycol can be performed using well known methods. Suitable replicates such as triplicate cultures can be grown for each engineered strain to be tested. For example, product and byproduct formation in the engineered production host can be monitored. The final product and intermediates, and other organic compounds, can be analyzed by methods such as HPLC (High Performance Liquid Chromatography), GC-MS (Gas Chromatography-Mass Spectroscopy) and LC-MS (Liquid Chromatography-Mass Spectroscopy) or other suitable analytical methods using routine procedures well known in the art. The release of product in the fermentation broth can also be tested with the culture supernatant. Byproducts and residual glucose can be quantified by HPLC using, for example, a refractive index detector for glucose and alcohols, and a UV detector for organic acids (Lin et al., *Biotechnol. Bioeng.* 90:775-779 (2005)), or other suitable assay and detection methods well known in the art. The individual enzyme or protein activities from the exogenous DNA sequences can also be assayed using methods well known in the art. For example, glycolaldehyde reductase activity can be measured by its NADH-dependent glycolaldehyde reduction to ethylene glycol using a molar absorption coefficient of $6.22 \times 10^{-3} \text{ M}^{-1}$ at 340 nm.

The ethylene glycol can be separated from other components in the culture using a variety of methods well known in the art. Such separation methods include, for example, extraction procedures as well as methods that include continuous liquid-liquid extraction, pervaporation, membrane filtration, membrane separation, reverse osmosis, electrodialysis, distillation, crystallization, centrifugation, extractive filtration, ion exchange chromatography, size exclusion chromatography, adsorption chromatography, and ultrafiltration. All of the above methods are well known in the art.

Any of the non-naturally occurring microbial organisms described herein can be cultured to produce and/or secrete the biosynthetic products of the invention. For example, the ethylene glycol producers can be cultured for the biosynthetic production of ethylene glycol.

For the production of ethylene glycol, the recombinant strains are cultured in a medium with carbon source and other essential nutrients. It is sometimes desirable and can be highly desirable to maintain anaerobic conditions in the fermenter to reduce the cost of the overall

process. Such conditions can be obtained, for example, by first sparging the medium with nitrogen and then sealing the flasks with a septum and crimp-cap. For strains where growth is not observed anaerobically, microaerobic or substantially anaerobic conditions can be applied by perforating the septum with a small hole for limited aeration. Exemplary
5 anaerobic conditions have been described previously and are well-known in the art. Exemplary aerobic and anaerobic conditions are described, for example, in United State publication 2009/0047719, filed August 10, 2007. Fermentations can be performed in a batch, fed-batch or continuous manner, as disclosed herein.

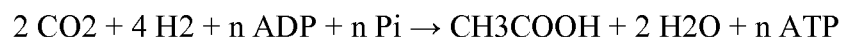
If desired, the pH of the medium can be maintained at a desired pH, in particular neutral pH,
10 such as a pH of around 7 by addition of a base, such as NaOH or other bases, or acid, as needed to maintain the culture medium at a desirable pH. The growth rate can be determined by measuring optical density using a spectrophotometer (600 nm), and the glucose uptake rate by monitoring carbon source depletion over time.

The growth medium can include, for example, any carbohydrate source which can supply a
15 source of carbon to the non-naturally occurring microorganism. Such sources include, for example, sugars such as glucose, xylose, arabinose, galactose, mannose, fructose, sucrose and starch. Other sources of carbohydrate include, for example, renewable feedstocks and biomass. Exemplary types of biomasses that can be used as feedstocks in the methods of the invention include cellulosic biomass, hemicellulosic biomass and lignin feedstocks or
20 portions of feedstocks. Such biomass feedstocks contain, for example, carbohydrate substrates useful as carbon sources such as glucose, xylose, arabinose, galactose, mannose, fructose and starch. Given the teachings and guidance provided herein, those skilled in the art will understand that renewable feedstocks and biomass other than those exemplified above also can be used for culturing the microbial organisms of the invention for the production of
25 ethylene glycol.

In addition to renewable feedstocks such as those exemplified above, the ethylene glycol microbial organisms of the invention also can be modified for growth on syngas as its source of carbon. In this specific embodiment, one or more proteins or enzymes are expressed in the ethylene glycol producing organisms to provide a metabolic pathway for utilization of syngas
30 or other gaseous carbon source.

Synthesis gas, also known as syngas or producer gas, is the major product of gasification of coal and of carbonaceous materials such as biomass materials, including agricultural crops and residues. Syngas is a mixture primarily of H₂ and CO and can be obtained from the gasification of any organic feedstock, including but not limited to coal, coal oil, natural gas, biomass, and waste organic matter. Gasification is generally carried out under a high fuel to oxygen ratio. Although largely H₂ and CO, syngas can also include CO₂ and other gases in smaller quantities. Thus, synthesis gas provides a cost effective source of gaseous carbon such as CO and, additionally, CO₂.

The Wood-Ljungdahl pathway catalyzes the conversion of CO and H₂ to acetyl-CoA and other products such as acetate. Organisms capable of utilizing CO and syngas also generally have the capability of utilizing CO₂ and CO₂/H₂ mixtures through the same basic set of enzymes and transformations encompassed by the Wood-Ljungdahl pathway. H₂-dependent conversion of CO₂ to acetate by microorganisms was recognized long before it was revealed that CO also could be used by the same organisms and that the same pathways were involved. Many acetogens have been shown to grow in the presence of CO₂ and produce compounds such as acetate as long as hydrogen is present to supply the necessary reducing equivalents (see for example, Drake, *Acetogenesis*, pp. 3-60 Chapman and Hall, New York, (1994)). This can be summarized by the following equation:



Hence, non-naturally occurring microorganisms possessing the Wood-Ljungdahl pathway can utilize CO₂ and H₂ mixtures as well for the production of acetyl-CoA and other desired products.

The Wood-Ljungdahl pathway is well known in the art and consists of 12 reactions which can be separated into two branches: (1) methyl branch and (2) carbonyl branch. The methyl branch converts syngas to methyl-tetrahydrofolate (methyl-THF) whereas the carbonyl branch converts methyl-THF to acetyl-CoA. The reactions in the methyl branch are catalyzed in order by the following enzymes or proteins: ferredoxin oxidoreductase, formate dehydrogenase, formyltetrahydrofolate synthetase, methenyltetrahydrofolate cyclodehydratase, methylenetetrahydrofolate dehydrogenase and methylenetetrahydrofolate reductase. The reactions in the carbonyl branch are catalyzed in order by the following enzymes or proteins: methyltetrahydrofolate:corrinoid protein methyltransferase (for example, AcsE), corrinoid iron-sulfur protein, nickel-protein assembly protein (for example,

AcsF), ferredoxin, acetyl-CoA synthase, carbon monoxide dehydrogenase and nickel-protein assembly protein (for example, CooC). Following the teachings and guidance provided herein for introducing a sufficient number of encoding nucleic acids to generate an ethylene glycol pathway, those skilled in the art will understand that the same engineering design also
5 can be performed with respect to introducing at least the nucleic acids encoding the Wood-Ljungdahl enzymes or proteins absent in the host organism. Therefore, introduction of one or more encoding nucleic acids into the microbial organisms of the invention such that the modified organism contains the complete Wood-Ljungdahl pathway will confer syngas utilization ability.

10 Additionally, the reductive (reverse) tricarboxylic acid cycle coupled with carbon monoxide dehydrogenase and/or hydrogenase activities can also be used for the conversion of CO, CO₂ and/or H₂ to acetyl-CoA and other products such as acetate. Organisms capable of fixing carbon via the reductive TCA pathway can utilize one or more of the following enzymes: ATP citrate-lyase, citrate lyase, aconitase, isocitrate dehydrogenase, alpha-

15 ketoglutarate:ferredoxin oxidoreductase, succinyl-CoA synthetase, succinyl-CoA transferase, fumarate reductase, fumarase, malate dehydrogenase, NAD(P)H:ferredoxin oxidoreductase, carbon monoxide dehydrogenase, and hydrogenase. Specifically, the reducing equivalents extracted from CO and/or H₂ by carbon monoxide dehydrogenase and hydrogenase are utilized to fix CO₂ via the reductive TCA cycle into acetyl-CoA or acetate. Acetate can be

20 converted to acetyl-CoA by enzymes such as acetyl-CoA transferase, acetate kinase/phosphotransacetylase, and acetyl-CoA synthetase. Acetyl-CoA can be converted to several metabolic intermediates including serine, 3-phosphoglycerate, 2-phosphoglycerate, glyceraldehyde and glyoxylate precursors by common central metabolic reactions, and glyceraldehyde-3-phosphate, phosphoenolpyruvate, and pyruvate, by pyruvate:ferredoxin
25 oxidoreductase and the enzymes of gluconeogenesis. Following the teachings and guidance provided herein for introducing a sufficient number of encoding nucleic acids to generate a serine, 3-phosphoglycerate, 2-phosphoglycerate, glyceraldehyde or glyoxylate pathway, those skilled in the art will understand that the same engineering design also can be performed with respect to introducing at least the nucleic acids encoding the reductive TCA pathway
30 enzymes or proteins absent in the host organism. Therefore, introduction of one or more encoding nucleic acids into the microbial organisms of the invention such that the modified organism contains the complete reductive TCA pathway will confer syngas utilization ability.

Accordingly, given the teachings and guidance provided herein, those skilled in the art will understand that a non-naturally occurring microbial organism can be produced that secretes the biosynthesized compounds of the invention when grown on a carbon source such as a carbohydrate. Such compounds include, for example, ethylene glycol and any of the intermediate metabolites in the ethylene glycol pathway. All that is required is to engineer in one or more of the required enzyme or protein activities to achieve biosynthesis of the desired compound or intermediate including, for example, inclusion of some or all of the ethylene glycol biosynthetic pathways. Accordingly, the invention provides a non-naturally occurring microbial organism that produces and/or secretes ethylene glycol when grown on a carbohydrate or other carbon source and produces and/or secretes any of the intermediate metabolites shown in the ethylene glycol pathway when grown on a carbohydrate or other carbon source. The ethylene glycol producing microbial organisms of the invention can initiate synthesis from an intermediate, for example, hydroxypyruvate, ethanolamine, glycolaldehyde, glycerate, tartronate semialdehyde, glycolate, glycolylphosphate or glycolyl-CoA.

The non-naturally occurring microbial organisms of the invention are constructed using methods well known in the art as exemplified herein to exogenously express at least one nucleic acid encoding an ethylene glycol pathway enzyme or protein in sufficient amounts to produce ethylene glycol. It is understood that the microbial organisms of the invention are cultured under conditions sufficient to produce ethylene glycol. Following the teachings and guidance provided herein, the non-naturally occurring microbial organisms of the invention can achieve biosynthesis of ethylene glycol resulting in intracellular concentrations between about 0.1-2000 mM or more. Generally, the intracellular concentration of ethylene glycol is between about 3-1500 mM, particularly between about 5-1250 mM and more particularly between about 8-1000 mM, including about 100 mM, 200 mM, 500 mM, 800 mM, or more. Intracellular concentrations between and above each of these exemplary ranges also can be achieved from the non-naturally occurring microbial organisms of the invention.

In some embodiments, culture conditions include anaerobic or substantially anaerobic growth or maintenance conditions. Exemplary anaerobic conditions have been described previously and are well known in the art. Exemplary anaerobic conditions for fermentation processes are described herein and are described, for example, in U.S. publication 2009/0047719, filed August 10, 2007. Any of these conditions can be employed with the non-naturally occurring

microbial organisms as well as other anaerobic conditions well known in the art. Under such anaerobic or substantially anaerobic conditions, the ethylene glycol producers can synthesize ethylene glycol at intracellular concentrations of 5-10 mM or more as well as all other concentrations exemplified herein. It is understood that, even though the above description refers to intracellular concentrations, ethylene glycol producing microbial organisms can produce ethylene glycol intracellularly and/or secrete the product into the culture medium.

In addition to the culturing and fermentation conditions disclosed herein, growth condition for achieving biosynthesis of ethylene glycol can include the addition of an osmoprotectant to the culturing conditions. In certain embodiments, the non-naturally occurring microbial organisms of the invention can be sustained, cultured or fermented as described herein in the presence of an osmoprotectant. Briefly, an osmoprotectant refers to a compound that acts as an osmolyte and helps a microbial organism as described herein survive osmotic stress.

Osmoprotectants include, but are not limited to, betaines, amino acids, and the sugar trehalose. Non-limiting examples of such are glycine betaine, proline betaine, dimethylthetin, dimethylsulfoniopropionate, 3-dimethylsulfonio-2-methylpropionate, pipercolic acid, dimethylsulfonioacetate, choline, L-carnitine and ectoine. In one aspect, the osmoprotectant is glycine betaine. It is understood to one of ordinary skill in the art that the amount and type of osmoprotectant suitable for protecting a microbial organism described herein from osmotic stress will depend on the microbial organism used. The amount of osmoprotectant in the culturing conditions can be, for example, no more than about 0.1 mM, no more than about 0.5 mM, no more than about 1.0 mM, no more than about 1.5 mM, no more than about 2.0 mM, no more than about 2.5 mM, no more than about 3.0 mM, no more than about 5.0 mM, no more than about 7.0 mM, no more than about 10mM, no more than about 50mM, no more than about 100mM or no more than about 500mM.

The culture conditions can include, for example, liquid culture procedures as well as fermentation and other large scale culture procedures. As described herein, particularly useful yields of the biosynthetic products of the invention can be obtained under anaerobic or substantially anaerobic culture conditions.

As described herein, one exemplary growth condition for achieving biosynthesis of ethylene glycol includes anaerobic culture or fermentation conditions. In certain embodiments, the non-naturally occurring microbial organisms of the invention can be sustained, cultured or fermented under anaerobic or substantially anaerobic conditions. Briefly, anaerobic

conditions refers to an environment devoid of oxygen. Substantially anaerobic conditions include, for example, a culture, batch fermentation or continuous fermentation such that the dissolved oxygen concentration in the medium remains between 0 and 10% of saturation. Substantially anaerobic conditions also includes growing or resting cells in liquid medium or
5 on solid agar inside a sealed chamber maintained with an atmosphere of less than 1% oxygen. The percent of oxygen can be maintained by, for example, sparging the culture with an N₂/CO₂ mixture or other suitable non-oxygen gas or gases.

The culture conditions described herein can be scaled up and grown continuously for manufacturing of ethylene glycol. Exemplary growth procedures include, for example, fed-
10 batch fermentation and batch separation; fed-batch fermentation and continuous separation, or continuous fermentation and continuous separation. All of these processes are well known in the art. Fermentation procedures are particularly useful for the biosynthetic production of commercial quantities of ethylene glycol. Generally, and as with non-continuous culture procedures, the continuous and/or near-continuous production of ethylene glycol will include
15 culturing a non-naturally occurring ethylene glycol producing organism of the invention in sufficient nutrients and medium to sustain and/or nearly sustain growth in an exponential phase. Continuous culture under such conditions can include, for example, growth for 1 day, 2, 3, 4, 5, 6 or 7 days or more. Additionally, continuous culture can include longer time periods of 1 week, 2, 3, 4 or 5 or more weeks and up to several months. Alternatively,
20 organisms of the invention can be cultured for hours, if suitable for a particular application. It is to be understood that the continuous and/or near-continuous culture conditions also can include all time intervals in between these exemplary periods. It is further understood that the time of culturing the microbial organism of the invention is for a sufficient period of time to produce a sufficient amount of product for a desired purpose.

25 Fermentation procedures are well known in the art. Briefly, fermentation for the biosynthetic production of ethylene glycol can be utilized in, for example, fed-batch fermentation and batch separation; fed-batch fermentation and continuous separation, or continuous fermentation and continuous separation. Examples of batch and continuous fermentation procedures are well known in the art.

30 In addition to the above fermentation procedures using the ethylene glycol producers of the invention for continuous production of substantial quantities of ethylene glycol, the ethylene glycol producers also can be, for example, simultaneously subjected to chemical synthesis

procedures to convert the product to other compounds or the product can be separated from the fermentation culture and sequentially subjected to chemical or enzymatic conversion to convert the product to other compounds, if desired.

To generate better producers, metabolic modeling can be utilized to optimize growth conditions. Modeling can also be used to design gene knockouts that additionally optimize utilization of the pathway (see, for example, U.S. patent publications US 2002/0012939, US 2003/0224363, US 2004/0029149, US 2004/0072723, US 2003/0059792, US 2002/0168654 and US 2004/0009466, and U.S. Patent No. 7,127,379). Modeling analysis allows reliable predictions of the effects on cell growth of shifting the metabolism towards more efficient production of ethylene glycol.

One computational method for identifying and designing metabolic alterations favoring biosynthesis of a desired product is the OptKnock computational framework (Burgard et al., Biotechnol. Bioeng. 84:647-657 (2003)). OptKnock is a metabolic modeling and simulation program that suggests gene deletion or disruption strategies that result in genetically stable microorganisms which overproduce the target product. Specifically, the framework examines the complete metabolic and/or biochemical network of a microorganism in order to suggest genetic manipulations that force the desired biochemical to become an obligatory byproduct of cell growth. By coupling biochemical production with cell growth through strategically placed gene deletions or other functional gene disruption, the growth selection pressures imposed on the engineered strains after long periods of time in a bioreactor lead to improvements in performance as a result of the compulsory growth-coupled biochemical production. Lastly, when gene deletions are constructed there is a negligible possibility of the designed strains reverting to their wild-type states because the genes selected by OptKnock are to be completely removed from the genome. Therefore, this computational methodology can be used to either identify alternative pathways that lead to biosynthesis of a desired product or used in connection with the non-naturally occurring microbial organisms for further optimization of biosynthesis of a desired product.

Briefly, OptKnock is a term used herein to refer to a computational method and system for modeling cellular metabolism. The OptKnock program relates to a framework of models and methods that incorporate particular constraints into flux balance analysis (FBA) models. These constraints include, for example, qualitative kinetic information, qualitative regulatory information, and/or DNA microarray experimental data. OptKnock also computes solutions

to various metabolic problems by, for example, tightening the flux boundaries derived through flux balance models and subsequently probing the performance limits of metabolic networks in the presence of gene additions or deletions. OptKnock computational framework allows the construction of model formulations that allow an effective query of the

5 performance limits of metabolic networks and provides methods for solving the resulting mixed-integer linear programming problems. The metabolic modeling and simulation methods referred to herein as OptKnock are described in, for example, U.S. publication 2002/0168654, filed January 10, 2002, in International Patent No. PCT/US02/00660, filed January 10, 2002, and U.S. publication 2009/0047719, filed August 10, 2007.

10 Another computational method for identifying and designing metabolic alterations favoring biosynthetic production of a product is a metabolic modeling and simulation system termed SimPheny®. This computational method and system is described in, for example, U.S. publication 2003/0233218, filed June 14, 2002, and in International Patent Application No. PCT/US03/18838, filed June 13, 2003. SimPheny® is a computational system that can be
15 used to produce a network model in silico and to simulate the flux of mass, energy or charge through the chemical reactions of a biological system to define a solution space that contains any and all possible functionalities of the chemical reactions in the system, thereby determining a range of allowed activities for the biological system. This approach is referred to as constraints-based modeling because the solution space is defined by constraints such as
20 the known stoichiometry of the included reactions as well as reaction thermodynamic and capacity constraints associated with maximum fluxes through reactions. The space defined by these constraints can be interrogated to determine the phenotypic capabilities and behavior of the biological system or of its biochemical components.

These computational approaches are consistent with biological realities because biological
25 systems are flexible and can reach the same result in many different ways. Biological systems are designed through evolutionary mechanisms that have been restricted by fundamental constraints that all living systems must face. Therefore, constraints-based modeling strategy embraces these general realities. Further, the ability to continuously impose further restrictions on a network model via the tightening of constraints results in a
30 reduction in the size of the solution space, thereby enhancing the precision with which physiological performance or phenotype can be predicted.

Given the teachings and guidance provided herein, those skilled in the art will be able to apply various computational frameworks for metabolic modeling and simulation to design and implement biosynthesis of a desired compound in host microbial organisms. Such metabolic modeling and simulation methods include, for example, the computational systems exemplified above as SimPheny® and OptKnock. For illustration of the invention, some methods are described herein with reference to the OptKnock computation framework for modeling and simulation. Those skilled in the art will know how to apply the identification, design and implementation of the metabolic alterations using OptKnock to any of such other metabolic modeling and simulation computational frameworks and methods well known in the art.

The methods described above will provide one set of metabolic reactions to disrupt. Elimination of each reaction within the set or metabolic modification can result in a desired product as an obligatory product during the growth phase of the organism. Because the reactions are known, a solution to the bilevel OptKnock problem also will provide the associated gene or genes encoding one or more enzymes that catalyze each reaction within the set of reactions. Identification of a set of reactions and their corresponding genes encoding the enzymes participating in each reaction is generally an automated process, accomplished through correlation of the reactions with a reaction database having a relationship between enzymes and encoding genes.

Once identified, the set of reactions that are to be disrupted in order to achieve production of a desired product are implemented in the target cell or organism by functional disruption of at least one gene encoding each metabolic reaction within the set. One particularly useful means to achieve functional disruption of the reaction set is by deletion of each encoding gene. However, in some instances, it can be beneficial to disrupt the reaction by other genetic aberrations including, for example, mutation, deletion of regulatory regions such as promoters or cis binding sites for regulatory factors, or by truncation of the coding sequence at any of a number of locations. These latter aberrations, resulting in less than total deletion of the gene set can be useful, for example, when rapid assessments of the coupling of a product are desired or when genetic reversion is less likely to occur.

To identify additional productive solutions to the above described bilevel OptKnock problem which lead to further sets of reactions to disrupt or metabolic modifications that can result in the biosynthesis, including growth-coupled biosynthesis of a desired product, an optimization

method, termed integer cuts, can be implemented. This method proceeds by iteratively solving the OptKnock problem exemplified above with the incorporation of an additional constraint referred to as an integer cut at each iteration. Integer cut constraints effectively prevent the solution procedure from choosing the exact same set of reactions identified in any previous iteration that obligatorily couples product biosynthesis to growth. For example, if a previously identified growth-coupled metabolic modification specifies reactions 1, 2, and 3 for disruption, then the following constraint prevents the same reactions from being simultaneously considered in subsequent solutions. The integer cut method is well known in the art and can be found described in, for example, Burgard et al., *Biotechnol. Prog.* 17:791-797 (2001). As with all methods described herein with reference to their use in combination with the OptKnock computational framework for metabolic modeling and simulation, the integer cut method of reducing redundancy in iterative computational analysis also can be applied with other computational frameworks well known in the art including, for example, SimPheny®.

The methods exemplified herein allow the construction of cells and organisms that biosynthetically produce a desired product, including the obligatory coupling of production of a target biochemical product to growth of the cell or organism engineered to harbor the identified genetic alterations. Therefore, the computational methods described herein allow the identification and implementation of metabolic modifications that are identified by an in silico method selected from OptKnock or SimPheny®. The set of metabolic modifications can include, for example, addition of one or more biosynthetic pathway enzymes and/or functional disruption of one or more metabolic reactions including, for example, disruption by gene deletion.

As discussed above, the OptKnock methodology was developed on the premise that mutant microbial networks can be evolved towards their computationally predicted maximum-growth phenotypes when subjected to long periods of growth selection. In other words, the approach leverages an organism's ability to self-optimize under selective pressures. The OptKnock framework allows for the exhaustive enumeration of gene deletion combinations that force a coupling between biochemical production and cell growth based on network stoichiometry. The identification of optimal gene/reaction knockouts requires the solution of a bilevel optimization problem that chooses the set of active reactions such that an optimal

growth solution for the resulting network overproduces the biochemical of interest (Burgard et al., *Biotechnol. Bioeng.* 84:647-657 (2003)).

An in silico stoichiometric model of *E. coli* metabolism can be employed to identify essential genes for metabolic pathways as exemplified previously and described in, for example, U.S. patent publications US 2002/0012939, US 2003/0224363, US 2004/0029149, US 2004/0072723, US 2003/0059792, US 2002/0168654 and US 2004/0009466, and in U.S. Patent No. 7,127,379. As disclosed herein, the OptKnock mathematical framework can be applied to pinpoint gene deletions leading to the growth-coupled production of a desired product. Further, the solution of the bilevel OptKnock problem provides only one set of deletions. To enumerate all meaningful solutions, that is, all sets of knockouts leading to growth-coupled production formation, an optimization technique, termed integer cuts, can be implemented. This entails iteratively solving the OptKnock problem with the incorporation of an additional constraint referred to as an integer cut at each iteration, as discussed above.

As disclosed herein, a nucleic acid encoding a desired activity of an ethylene glycol pathway can be introduced into a host organism. In some cases, it can be desirable to modify an activity of an ethylene glycol pathway enzyme or protein to increase production of ethylene glycol. For example, known mutations that increase the activity of a protein or enzyme can be introduced into an encoding nucleic acid molecule. Additionally, optimization methods can be applied to increase the activity of an enzyme or protein and/or decrease an inhibitory activity, for example, decrease the activity of a negative regulator.

One such optimization method is directed evolution. Directed evolution is a powerful approach that involves the introduction of mutations targeted to a specific gene in order to improve and/or alter the properties of an enzyme. Improved and/or altered enzymes can be identified through the development and implementation of sensitive high-throughput screening assays that allow the automated screening of many enzyme variants (for example, >104). Iterative rounds of mutagenesis and screening typically are performed to afford an enzyme with optimized properties. Computational algorithms that can help to identify areas of the gene for mutagenesis also have been developed and can significantly reduce the number of enzyme variants that need to be generated and screened. Numerous directed evolution technologies have been developed (for reviews, see Hibbert et al., *Biomol. Eng.* 22:11-19 (2005); Huisman and Lalonde, In *Biocatalysis in the pharmaceutical and biotechnology industries* pgs. 717-742 (2007), Patel (ed.), CRC Press; Otten and Quax.

Biomol.Eng 22:1-9 (2005).; and Sen et al., Appl Biochem.Biotechnol 143:212-223 (2007)) to be effective at creating diverse variant libraries, and these methods have been successfully applied to the improvement of a wide range of properties across many enzyme classes.

Enzyme characteristics that have been improved and/or altered by directed evolution

5 technologies include, for example: selectivity/specificity, for conversion of non-natural substrates; temperature stability, for robust high temperature processing; pH stability, for bioprocessing under lower or higher pH conditions; substrate or product tolerance, so that high product titers can be achieved; binding (K_m), including broadening substrate binding to include non-natural substrates; inhibition (K_i), to remove inhibition by products, substrates,
10 or key intermediates; activity (kcat), to increase enzymatic reaction rates to achieve desired flux; expression levels, to increase protein yields and overall pathway flux; oxygen stability, for operation of air sensitive enzymes under aerobic conditions; and anaerobic activity, for operation of an aerobic enzyme in the absence of oxygen.

A number of exemplary methods have been developed for the mutagenesis and

15 diversification of genes to target desired properties of specific enzymes. Such methods are well known to those skilled in the art. Any of these can be used to alter and/or optimize the activity of an ethylene glycol pathway enzyme or protein. Such methods include, but are not limited to EpPCR, which introduces random point mutations by reducing the fidelity of DNA polymerase in PCR reactions (Pritchard et al., J Theor.Biol. 234:497-509 (2005)); Error-prone Rolling Circle Amplification (epRCA), which is similar to epPCR except a whole
20 circular plasmid is used as the template and random 6-mers with exonuclease resistant thiophosphate linkages on the last 2 nucleotides are used to amplify the plasmid followed by transformation into cells in which the plasmid is re-circularized at tandem repeats (Fujii et al., Nucleic Acids Res. 32:e145 (2004); and Fujii et al., Nat. Protoc. 1:2493-2497 (2006)); DNA
25 or Family Shuffling, which typically involves digestion of two or more variant genes with nucleases such as Dnase I or EndoV to generate a pool of random fragments that are reassembled by cycles of annealing and extension in the presence of DNA polymerase to create a library of chimeric genes (Stemmer, Proc Natl Acad Sci USA 91:10747-10751 (1994); and Stemmer, Nature 370:389-391 (1994)); Staggered Extension (StEP), which
30 entails template priming followed by repeated cycles of 2 step PCR with denaturation and very short duration of annealing/extension (as short as 5 sec) (Zhao et al., Nat. Biotechnol. 16:258-261 (1998)); Random Priming Recombination (RPR), in which random sequence

primers are used to generate many short DNA fragments complementary to different segments of the template (Shao et al., *Nucleic Acids Res* 26:681-683 (1998)).

Additional methods include Heteroduplex Recombination, in which linearized plasmid DNA is used to form heteroduplexes that are repaired by mismatch repair (Volkov et al, *Nucleic*
5 *Acids Res.* 27:e18 (1999); and Volkov et al., *Methods Enzymol.* 328:456-463 (2000)); Random Chimera-genesis on Transient Templates (RACHITT), which employs Dnase I fragmentation and size fractionation of single stranded DNA (ssDNA) (Coco et al., *Nat. Biotechnol.* 19:354-359 (2001)); Recombined Extension on Truncated templates (RETT), which entails template switching of unidirectionally growing strands from primers in the
10 presence of unidirectional ssDNA fragments used as a pool of templates (Lee et al., *J. Molec. Catalysis* 26:119-129 (2003)); Degenerate Oligonucleotide Gene Shuffling (DOGS), in which degenerate primers are used to control recombination between molecules; (Bergquist and Gibbs, *Methods Mol.Biol* 352:191-204 (2007); Bergquist et al., *Biomol.Eng* 22:63-72 (2005); Gibbs et al., *Gene* 271:13-20 (2001)); Incremental Truncation for the Creation of Hybrid
15 Enzymes (ITCHY), which creates a combinatorial library with 1 base pair deletions of a gene or gene fragment of interest (Ostermeier et al., *Proc. Natl. Acad. Sci. USA* 96:3562-3567 (1999); and Ostermeier et al., *Nat. Biotechnol.* 17:1205-1209 (1999)); Thio-Incremental Truncation for the Creation of Hybrid Enzymes (THIO-ITCHY), which is similar to ITCHY except that phosphothioate dNTPs are used to generate truncations (Lutz et al., *Nucleic Acids*
20 *Res* 29:E16 (2001)); SCRATCHY, which combines two methods for recombining genes, ITCHY and DNA shuffling (Lutz et al., *Proc. Natl. Acad. Sci. USA* 98:11248-11253 (2001)); Random Drift Mutagenesis (RNDM), in which mutations made via epPCR are followed by screening/selection for those retaining usable activity (Bergquist et al., *Biomol. Eng.* 22:63-72 (2005)); Sequence Saturation Mutagenesis (SeSaM), a random mutagenesis method that
25 generates a pool of random length fragments using random incorporation of a phosphothioate nucleotide and cleavage, which is used as a template to extend in the presence of “universal” bases such as inosine, and replication of an inosine-containing complement gives random base incorporation and, consequently, mutagenesis (Wong et al., *Biotechnol. J.* 3:74-82 (2008); Wong et al., *Nucleic Acids Res.* 32:e26 (2004); and Wong et al., *Anal. Biochem.*
30 341:187-189 (2005)); Synthetic Shuffling, which uses overlapping oligonucleotides designed to encode “all genetic diversity in targets” and allows a very high diversity for the shuffled progeny (Ness et al., *Nat. Biotechnol.* 20:1251-1255 (2002)); Nucleotide Exchange and Excision Technology NexT, which exploits a combination of dUTP incorporation followed

by treatment with uracil DNA glycosylase and then piperidine to perform endpoint DNA fragmentation (Muller et al., *Nucleic Acids Res.* 33:e117 (2005)).

Further methods include Sequence Homology-Independent Protein Recombination (SHIPREC), in which a linker is used to facilitate fusion between two distantly related or unrelated genes, and a range of chimeras is generated between the two genes, resulting in libraries of single-crossover hybrids (Sieber et al., *Nat. Biotechnol.* 19:456-460 (2001)); Gene Site Saturation Mutagenesis™ (GSSM™), in which the starting materials include a supercoiled double stranded DNA (dsDNA) plasmid containing an insert and two primers which are degenerate at the desired site of mutations (Kretz et al., *Methods Enzymol.* 388:3-11 (2004)); Combinatorial Cassette Mutagenesis (CCM), which involves the use of short oligonucleotide cassettes to replace limited regions with a large number of possible amino acid sequence alterations (Reidhaar-Olson et al. *Methods Enzymol.* 208:564-586 (1991); and Reidhaar-Olson et al. *Science* 241:53-57 (1988)); Combinatorial Multiple Cassette Mutagenesis (CMCM), which is essentially similar to CCM and uses epPCR at high mutation rate to identify hot spots and hot regions and then extension by CMCM to cover a defined region of protein sequence space (Reetz et al., *Angew. Chem. Int. Ed Engl.* 40:3589-3591 (2001)); the Mutator Strains technique, in which conditional ts mutator plasmids, utilizing the mutD5 gene, which encodes a mutant subunit of DNA polymerase III, to allow increases of 20 to 4000-X in random and natural mutation frequency during selection and block accumulation of deleterious mutations when selection is not required (Selifonova et al., *Appl. Environ. Microbiol.* 67:3645-3649 (2001)); Low et al., *J. Mol. Biol.* 260:359-3680 (1996)).

Additional exemplary methods include Look-Through Mutagenesis (LTM), which is a multidimensional mutagenesis method that assesses and optimizes combinatorial mutations of selected amino acids (Rajpal et al., *Proc. Natl. Acad. Sci. USA* 102:8466-8471 (2005)); Gene Reassembly, which is a DNA shuffling method that can be applied to multiple genes at one time or to create a large library of chimeras (multiple mutations) of a single gene (Tunable GeneReassembly™ (TGR™) Technology supplied by Verenum Corporation), in Silico Protein Design Automation (PDA), which is an optimization algorithm that anchors the structurally defined protein backbone possessing a particular fold, and searches sequence space for amino acid substitutions that can stabilize the fold and overall protein energetics, and generally works most effectively on proteins with known three-dimensional structures (Hayes et al., *Proc. Natl. Acad. Sci. USA* 99:15926-15931 (2002)); and Iterative Saturation

Mutagenesis (ISM), which involves using knowledge of structure/function to choose a likely site for enzyme improvement, performing saturation mutagenesis at chosen site using a mutagenesis method such as Stratagene QuikChange (Stratagene; San Diego CA), screening/selecting for desired properties, and, using improved clone(s), starting over at another site and continue repeating until a desired activity is achieved (Reetz et al., Nat. Protoc. 2:891-903 (2007); and Reetz et al., Angew. Chem. Int. Ed Engl. 45:7745-7751 (2006)).

Any of the aforementioned methods for mutagenesis can be used alone or in any combination. Additionally, any one or combination of the directed evolution methods can be used in conjunction with adaptive evolution techniques, as described herein.

It is understood that modifications which do not substantially affect the activity of the various embodiments of this invention are also provided within the definition of the invention provided herein. Accordingly, the following examples are intended to illustrate but not limit the present invention.

EXAMPLE I

Pathways for producing ethylene glycol from serine

Several pathways are shown in Figure 1 for synthesis of MEG from serine. In one embodiment serine is converted to hydroxypyruvate by a serine-hydroxypyruvate aminotransferase or a serine oxidoreductase (deaminating) (Figure 1, Steps 1 or 2). Hydroxypyruvate is subsequently decarboxylated to glycolaldehyde by hydroxypyruvate decarboxylase (Figure 1, Step 3). Finally, glycolaldehyde is reduced to MEG by an aldehyde reductase (Figure 1, Step 4). In an alternate route, the hydroxypyruvate intermediate is reduced to glycerate by hydroxypyruvate reductase, and subsequently decarboxylated yielding ethylene glycol (Figure 1, Steps 8 and 9). In yet another pathway, serine is first decarboxylated to ethanolamine (Figure 1, Step 5). This compound is subsequently converted to glycolaldehyde by a serine aminotransferase or oxidoreductase (deaminating) (Figure 1, Steps 6 or 7). Exemplary enzyme candidates for serine pathway enzymes (Steps 1-9 of Figure 1) are described below.

The conversion of serine to hydroxypyruvate (Figure 1, Step 1) is catalyzed by an enzyme with serine aminotransferase activity. Exemplary enzymes include serine:pyruvate aminotransferase (EC 2.6.1.510), alanine:glyoxylate aminotransferase (EC 2.6.1.44) and

serine:glyoxylate aminotransferase (EC 2.6.1.45). Serine:pyruvate aminotransferase participates in serine metabolism and glyoxylate detoxification in mammals. These enzymes have been shown to utilize a variety of alternate oxo donors such as pyruvate, phenylpyruvate and glyoxylate; and amino acceptors including alanine, glycine and phenylalanine (Ichiyama et al., Mol.Urol. 4:333-340 (2000)). The rat mitochondria serine:pyruvate aminotransferase, encoded by *agxt*, is also active as an alanine-glyoxylate aminotransferase. This enzyme was heterologously expressed in *E. coli* (Oda et al., J Biochem. 106:460-467 (1989)). Similar enzymes have been characterized in humans and flies (Oda et al., Biochem.Biophys.Res.Comm. 228:341-346 (1996)). The human enzyme, encoded by *agxt*, functions as a serine:pyruvate aminotransferase, an alanine:glyoxylate aminotransferase and a serine:glyoxylate aminotransferase (Nagata et al., Biomed.Res. 30:295-301 (2009)). The fly enzyme is encoded by *spat* (Han et al., FEBS Lett. 527:199-204 (2002)). An exemplary alanine:glyoxylate aminotransferase is encoded by *AGT1* of *Arabidopsis thaliana*. In addition to the alanine:glyoxylate activity, the purified, recombinant *AGT1* expressed in *E. coli* also catalyzed serine:glyoxylate and serine:pyruvate aminotransferase activities (Liepman et al., Plant J 25:487-498 (2001)). In several organisms serine:glyoxylate aminotransferase enzymes (EC 2.6.1.45) also exhibit reduced but detectable serine:pyruvate aminotransferase activity. Exemplary enzymes are found in *Phaseolus vulgaris*, *Pisum sativum*, *Secale cereal* and *Spinacia oleracea*. Serine:glyoxylate aminotransferase enzymes interconvert serine and hydroxypyruvate and utilize glyoxylate as an amino acceptor. The serine:glyoxylate aminotransferase from the obligate methylotroph *Hyphomicrobium methylovorum* GM2 has been functionally expressed in *E. coli* and characterized (Hagishita et al., Eur.J Biochem. 241:1-5 (1996)).

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>Agxt</i>	NP_085914.1	13470096	<i>Rattus norvegicus</i>
<i>Agxt</i>	NP_085914.1	13470096	<i>Rattus norvegicus</i>
<i>Agxt</i>	NP_000021.1	4557289	<i>Homo sapiens</i>
<i>Spat</i>	NP_511062.1	17530823	<i>Drosophila melanogaster</i>
<i>AGT1</i>	NP_849951.1	30678921	<i>Arabidopsis thaliana</i>
<i>D86125.1:914..2131</i>	BAA19919.1	2081618	<i>Hyphomicrobium methylovorum</i>

The conversion of serine to hydroxypyruvate (Figure 1, Step 1) is alternately catalyzed by serine oxidoreductase (deaminating). One enzyme with this functionality is serine oxidase, which utilizes oxygen as an electron acceptor, converting serine, O₂ and water to ammonia, hydrogen peroxide and hydroxypyruvate (Chumakov, et al., Proc. Nat. Acad. Sci.,

99(21):13675-13680; Verral et al., *Eur J Neurosci.*, 26(6) 1657-1669 (2007)). Some amino oxidases are specific for the D-amino acid (Dixon and Kleppe, *Biochim Biophys Acta*, 96: 368-382 (1965)) and L-serine can be converted to D-serine by serine racemase (Miranda, et al., *Gene*, 256:183-188 (2000)). Enzymes in the EC class 1.4.1 catalyze the oxidative

5 deamination of alpha-amino acids with NAD⁺, NADP⁺ or FAD as acceptor, and the reactions are typically reversible. Exemplary enzymes with serine oxidoreductase (deaminating) activity include serine dehydrogenase (EC 1.4.1.7), L-amino acid dehydrogenase (EC 1.4.1.5) and glutamate dehydrogenase (EC 1.4.1.2). An enzyme with serine dehydrogenase activity from *Petroselinum crispum* was purified and characterized

10 although the gene associated with the enzyme has not been identified to date (Kretovich et al., *Izv.Akad.Nauk SSSR Ser.Biol.* 2:295-301 (1966)). Serine dehydrogenase activity attributed to L-amino-acid dehydrogenase was identified in soil bacteria isolates, but specific genes were not identified (Mohammadi et al., *Iran Biomed.J* 11:131-135 (2007)). The glutamate dehydrogenase from *Vigna unguiculata* accepts serine as an alternate substrate.

15 The gene associated with this enzyme has not been identified to date. Other glutamate dehydrogenase enzymes are encoded by *gdhA* in *Escherichia coli* (Korber et al., *J Mol.Biol.* 234:1270-1273 (1993); McPherson et al., *Nucleic Acids Res.* 11:5257-5266 (1983)), *gdh* from *Thermotoga maritime* (Kort et al., *Extremophiles.* 1:52-60 (1997); Lebbink et al., *J Mol.Biol.* 280:287-296 (1998); Lebbink et al., *J Mol.Biol.* 289:357-369 (1999)), and *gdhA1* from

20 *Halobacterium salinarum* (Ingoldsby et al., *Gene* 349:237-244 (2005)).

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>gdhA</i>	118547	P00370	<i>Escherichia coli</i>
<i>gdh</i>	6226595	P96110.4	<i>Thermotoga maritima</i>
<i>gdhA1</i>	15789827	NP_279651.1	<i>Halobacterium salinarum</i>

Decarboxylation of hydroxypyruvate to glycolaldehyde (Figure 1, Step 3 and Figure 2, Step 3) is catalyzed by hydroxypyruvate decarboxylase (EC 4.1.1.40), an enzyme found in many mammals (Hendrick et al., *Arch.Biochem.Biophys.* 105:261-269 (1964)). The enzyme activity

25 has been studied in the context of hydroxypyruvate metabolism to oxalate in rat mitochondria, although the activity is not associated with a gene to date (Rofe et al., *Biochem.Med.Metab Biol.* 36:141-150 (1986)). Other keto-acid decarboxylases include pyruvate decarboxylase (EC 4.1.1.1), benzoylformate decarboxylase (EC 4.1.1.7), alpha-ketoglutarate decarboxylase and branched-chain alpha-ketoacid decarboxylase. Several keto-

30 acid decarboxylase enzymes have been shown to accept hydroxypyruvate as an alternate

substrate, including the *kivd* gene product of *Lactococcus lactis* (de la Plaza et al., *FEMS Microbiol Lett.* 238:367-374 (2004)) and the *pdcl* gene product of *Saccharomyces cerevisiae* (Cusa et al., *J Bacteriol.* 181:7479-7484 (1999)). The *S. cerevisiae* enzyme has been extensively studied, engineered for altered activity, and functionally expressed in *E. coli* (Killenberg-Jabs et al., *Eur.J.Biochem.* 268:1698-1704 (2001); Li et al., *Biochemistry.* 38:10004-10012 (1999); ter Schure et al., *Appl.Environ.Microbiol.* 64:1303-1307 (1998)). The PDC from *Zymomonas mobilis*, encoded by *pdc*, also has a broad substrate range and has been a subject of directed engineering studies to alter the affinity for different substrates (Siegert et al., *Protein Eng Des Sel* 18:345-357 (2005)). An additional candidate is the *kdcA* gene product of *Lactococcus lactis*, which decarboxylates a variety of branched and linear ketoacid substrates including 2-oxobutanoate, 2-oxohexanoate, 2-oxopentanoate, 3-methyl-2-oxobutanoate, 4-methyl-2-oxobutanoate and isocaproate (Smit et al., *Appl Environ Microbiol* 71:303-311 (2005)). Additional decarboxylase enzyme candidates are provided in Example XI.

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>kivd</i>	CAG34226.1	51870502	<i>Lactococcus lactis</i>
<i>pdcl</i>	P06169	30923172	<i>Saccharomyces cerevisiae</i>
<i>pdc</i>	P06672.1	118391	<i>Zymomonas mobilis</i>
<i>kdcA</i>	AAS49166.1	44921617	<i>Lactococcus lactis</i>

The reduction of glycolaldehyde to ethylene glycol (all Figures) is catalyzed by glycolaldehyde reductase. The iron-activated 1,2-PDO oxidoreductase (EC 1.1.1.77) *E. coli* encoded by *fucO* efficiently catalyzes the reduction of glycolaldehyde (Obradors et al., *Eur.J Biochem.* 258:207-213 (1998); Boronat et al., *J Bacteriol.* 153:134-139 (1983)). Other aldehyde reductase enzyme candidates include *alrA* from *Acinetobacter* sp. Strain M-1 encoding a medium-chain alcohol dehydrogenase for C2-C14 (Tani et al., *Appl.Environ.Microbiol.* 66:5231-5235 (2000)), ADH2 from *Saccharomyces cerevisiae* (Atsumi et al., *Nature* 451:86-89 (2008)) and the *adhA* gene product from *Zymomonas mobilis*, which was demonstrated to have activity on a number of aldehydes including formaldehyde, acetaldehyde, propionaldehyde, butyraldehyde, and acrolein (Kinoshita et al., *Appl Microbiol Biotechnol* 22:249-254 (1985)). Additional aldehyde reductase enzyme candidates are provided in Example XI.

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
-----------------------	--------------------------	-------------------------	------------------------

Protein	GenBank ID	GI Number	Organism
<i>fucO</i>	AAA23825.1	146045	<i>Escherichia coli</i>
<i>alrA</i>	BAB12273.1	9967138	<i>Acinetobacter</i> sp. Strain M-1
<i>ADH2</i>	NP_014032.1	6323961	<i>Saccharomyces cerevisiae</i>
<i>adhA</i>	YP_162971.1	56552132	<i>Zymomonas mobilis</i>

Serine decarboxylase (EC 4.1.1.-) catalyzes the decarboxylation of serine to ethanolamine (Figure 1, Step 5). Enzymes with this activity have been characterized in plants such as *Spinacia oleracea*, *Arabidopsis thaliana* and *Brassica napus* in the context of choline

- 5 biosynthesis. The *A. thaliana* serine decarboxylase encoded by *AtSDC* is a soluble tetramer and was characterized by heterologous expression in *E. coli* and ability to complement a yeast mutant deficient in ethanolamine biosynthesis (Rontein et al., *J Biol.Chem.* 276:35523-35529 (2001)). The *Brassica napus* serine decarboxylase was identified and characterized in the same study. A similar enzyme is found in *Spinacia oleracea* although the gene has not
10 been identified to date (Summers et al., *Plant Physiol* 103:1269-1276 (1993)). Other serine decarboxylase candidates can be identified by sequence homology to the *Arabidopsis* or *Brassica* enzymes. A candidate with high homology is the putative serine decarboxylase from *Beta vulgaris*.

Protein	GenBank ID	GI Number	Organism
<i>AtSDC</i>	AAK77493.1	15011302	<i>Arabidopsis thaliana</i>
<i>BnSDC</i>	BAA78331.1	4996105	<i>Brassica napus</i>
<i>BvSDC1</i>	BAE07183.1	71000475	<i>Beta vulgaris</i>

- 15 The conversion of ethanolamine to glycolaldehyde is catalyzed by an enzyme with ethanolamine aminotransferase activity. Such an enzyme activity has not been demonstrated to date. Exemplary candidates are aminotransferases with broad substrate specificity that convert terminal amines to aldehydes, such as gamma-aminobutyrate GABA transaminase (EC 2.6.1.19), diamine aminotransferase (EC 2.6.1.29) and putrescine aminotransferase (EC
20 2.6.1.82). GABA aminotransferase naturally interconverts succinic semialdehyde and glutamate to 4-aminobutyrate and alpha-ketoglutarate and is known to have a broad substrate range (Schulz et al., 56:1-6 (1990); Liu et al., 43:10896-10905 (2004)). The two GABA transaminases in *E. coli* are encoded by *gabT* (Bartsch et al., *J Bacteriol.* 172:7035-7042 (1990)) and *puuE* (Kurihara et al., *J.Biol.Chem.* 280:4602-4608 (2005)). GABA
25 transaminases in *Mus musculus* and *Sus scrofa* have also been shown to react with a range of alternate substrates (Cooper, *Methods Enzymol.* 113:80-82 (1985)). Additional enzyme candidates for interconverting ethanolamine and glycolaldehyde are putrescine

aminotransferases and other diamine aminotransferases. The *E. coli* putrescine aminotransferase is encoded by the *ygiG* gene and the purified enzyme also was able to transaminate cadaverine and spermidine (Samsonova et al., *BMC.Microbiol* 3:2 (2003)). In addition, activity of this enzyme on 1,7-diaminoheptane and with amino acceptors other than 2-oxoglutarate (e.g., pyruvate, 2-oxobutanoate) has been reported (Samsonova et al., *BMC.Microbiol* 3:2 (2003); Kim, *J Biol.Chem.* 239:783-786 (1964)).

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>gabT</i>	NP_417148.1	16130576	<i>Escherichia coli</i>
<i>puuE</i>	NP_415818.1	16129263	<i>Escherichia coli</i>
<i>abat</i>	NP_766549.2	37202121	<i>Mus musculus</i>
<i>abat</i>	NP_999428.1	47523600	<i>Sus scrofa</i>
<i>ygiG</i>	NP_417544	145698310	<i>Escherichia coli</i>

The oxidative deamination of ethanolamine to glycolaldehyde is catalyzed by ethanolamine oxidoreductase (deaminating). One enzyme with this functionality is ethanolamine oxidase (EC 1.4.3.8), which utilizes oxygen as an electron acceptor, converting ethanolamine, O₂ and water to ammonia, hydrogen peroxide and glycolaldehyde (Schomburg et al., *Springer Handbook of Enzymes*. 320-323 (2005)). Ethanolamine oxidase has been characterized in *Pseudomonas* sp and *Phormia regina*; however, the enzyme activity has not been associated with a gene to date. Alternately, the oxidative deamination of ethanolamine can be catalyzed by a deaminating oxidoreductase that utilizes NAD⁺, NADP⁺ or FAD as acceptor. An exemplary enzyme for catalyzing the conversion of a primary amine to an aldehyde is lysine 6-dehydrogenase (EC 1.4.1.18), encoded by the *lysDH* genes. This enzyme catalyzes the oxidative deamination of the 6-amino group of L-lysine to form 2-amino adipate-6-semialdehyde (Misono et al., *J Bacteriol.* 150:398-401 (1982)). Additional enzyme candidates are found in *Geobacillus stearothermophilus* (Heydari et al., *Appl Environ.Microbiol* 70:937-942 (2004)), *Agrobacterium tumefaciens* (Hashimoto et al., *J Biochem.* 106:76-80 (1989); Misono and Nagasaki, *J Bacteriol.* 150:398-401 (1982)), and *Achromobacter denitrificans* (Ruldeekulthamrong et al., *BMB.Rep.* 41:790-795 (2008)).

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>lysDH</i>	BAB39707	13429872	<i>Geobacillus stearothermophilus</i>
<i>lysDH</i>	NP_353966	15888285	<i>Agrobacterium tumefaciens</i>
<i>lysDH</i>	AAZ94428	74026644	<i>Achromobacter denitrificans</i>

Hydroxypyruvate reductase (EC 1.1.1.29 and EC 1.1.1.81), also called glycerate dehydrogenase, catalyzes the reversible NAD(P)H-dependent reduction of hydroxypyruvate

to glycerate (Figure 1, Step 8). The *ghrA* and *ghrB* genes of *E. coli* encode enzymes with hydroxypyruvate reductase activity (Nunez et al., *Biochem.J* 354:707-715 (2001)). Both gene products also catalyze the reduction of glyoxylate to glycolate and the *ghrB* gene product prefers hydroxypyruvate as a substrate. Hydroxypyruvate reductase participates in the serine cycle in methylotrophic bacterium such as *Methylobacterium extorquens* AM1 and *Hyphomicrobium methylovorum* (Chistoserdova et al., *J Bacteriol.* 185:2980-2987 (2003)). Hydroxypyruvate reductase enzymes from *Hyphomicrobium methylovorum* and *Methylobacterium* sp. MB200 have been cloned and heterologously expressed in *E. coli* (Yoshida et al., *Eur.J Biochem.* 223:727-732 (1994)). The *Methylobacterium* sp. MB200 HPR has not been assigned a GenBank identifier to date but the sequence is available in the literature and bears 98% identity to the sequence of the *M. extorquens hprA* gene product, which uses both NADH and NADPH as cofactors (Chistoserdova et al., *J Bacteriol.* 173:7228-7232 (1991)). Bifunctional enzymes with hydroxypyruvate reductase and glyoxylate reductase activities (GRHPR) are found in mammals including *Homo sapiens* and *Mus musculus*. Recombinant NADPH-dependent GRHPR enzymes from these organisms were heterologously expressed in *E. coli* (Booth et al., *J Mol.Biol.* 360:178-189 (2006)).

Protein	GenBank ID	GI Number	Organism
<i>ghrA</i>	NP_415551.2	90111205	<i>Escherichia coli</i>
<i>ghrB</i>	NP_418009.2	90111614	<i>Escherichia coli</i>
D31857.1:286..1254	BAA06662.1	1304133	<i>Hyphomicrobium methylovorum</i>
<i>hprA</i>	ACS39571.1	240008345	<i>Methylobacterium extorquens</i>
<i>GRHPR</i>	NP_036335.1	6912396	<i>Homo sapiens</i>
<i>GRHPR</i>	NP_525028.1	17933768	<i>Mus musculus</i>

An enzyme with glycerate decarboxylase activity is required to convert glycerate to ethylene glycol (Figure 1, Step 9). Such an enzyme has not been characterized to date. However, a similar alpha,beta-hydroxyacid decarboxylation reaction is catalyzed by tartrate decarboxylase (EC 4.1.1.73). The enzyme, characterized in *Pseudomonas* sp. group Ve-2, is NAD⁺ dependent and catalyzes a coupled oxidation-reduction reaction that proceeds through an oxaloglycolate intermediate (Furuyoshi et al., *J Biochem.* 110:520-525 (1991)). A side reaction catalyzed by this enzyme is the NAD⁺ dependent oxidation of tartrate (1% of activity). Glycerate was not reactive as a substrate for this enzyme and was instead an inhibitor, so enzyme engineering or directed evolution is likely required for this enzyme to function in the desired context. A gene has not been associated with this enzyme activity to date.

An additional candidate glycerate decarboxylase is acetolactate decarboxylase (EC 4.1.1.5) which participates in citrate catabolism and branched-chain amino acid biosynthesis, converting the 2-hydroxyacid 2-acetolactate to acetoin. In *Lactococcus lactis* the enzyme is a hexamer encoded by gene *aldB*, and is activated by valine, leucine and isoleucine (Goupil-Feuillerat et al., J.Bacteriol. 182:5399-5408 (2000); Goupil et al., Appl.Environ.Microbiol. 62:2636-2640 (1996)). This enzyme has been overexpressed and characterized in *E. coli* (Phalip et al., FEBS Lett. 351:95-99 (1994)). In other organisms the enzyme is a dimer, encoded by *aldC* in *Streptococcus thermophilus* (Monnet et al., Lett.Appl.Microbiol. 36:399-405 (2003)), *aldB* in *Bacillus brevis* (Najmudin et al., Acta Crystallogr.D.Biol.Crystallogr. 59:1073-1075 (2003); Diderichsen et al., J.Bacteriol. 172:4315-4321 (1990)) and *budA* from *Enterobacter aerogenes* (Diderichsen et al., J.Bacteriol. 172:4315-4321 (1990)). The enzyme from *Bacillus brevis* was cloned and overexpressed in *Bacillus subtilis* and structurally characterized (Najmudin et al., Acta Crystallogr.D.Biol.Crystallogr. 59:1073-1075 (2003)). A similar enzyme from *Leuconostoc lactis* has been purified and characterized but the gene has not been isolated to date (O'Sullivan et al., FEMS Microbiol.Lett. 194:245-249 (2001)).

Protein	GenBank ID	GI Number	Organism
<i>aldB</i>	NP_267384.1	15673210	<i>Lactococcus lactis</i>
<i>aldC</i>	Q8L208	75401480	<i>Streptococcus thermophilus</i>
<i>aldB</i>	P23616.1	113592	<i>Bacillus brevis</i>
<i>budA</i>	P05361.1	113593	<i>Enterobacter aerogenes</i>

EXAMPLE II

Pathways for producing ethylene glycol from 3-phosphoglycerate

Also shown in Figure 1 are pathways to convert 3-phosphoglycerate (3PG) to ethylene glycol. In these pathways, 3-phosphoglycerate is first converted to glycerate by either a 3PG phosphatase or a glycerate kinase enzyme operating in the glycerate-generating direction (Figure 1, Steps 10 or 11). Glycerate is then directly decarboxylated to ethylene glycol (Figure 1, Step 9). Alternately, glycerate is oxidized to hydroxypyruvate (Figure 1, Step 8), which is subsequently converted to ethylene glycol by the combined actions of hydroxypyruvate decarboxylase and glycolaldehyde reductase as described previously. Enzyme candidates for steps 10-11 of Figure 1 are provided below.

3-Phosphoglycerate phosphatase (EC 3.1.3.38) catalyzes the hydrolysis of 3PG to glycerate, releasing pyrophosphate (Figure 1, Step 10). The enzyme is found in plants and has a broad substrate range that includes phosphoenolpyruvate, ribulose-1,5-bisphosphate,

dihydroxyacetone phosphate and glucose-6-phosphate (Randall et al., Plant Physiol 48:488-492 (1971); Randall et al., J Biol.Chem. 246:5510-5517 (1971)). Purified enzyme from various plant sources has been characterized but a gene has not been associated with this enzyme to date. Another enzyme with 3-phosphoglycerate phosphatase activity is the
 5 phosphoglycerate phosphatase (EC 3.1.3.20) from pig liver (Fallon et al., Biochim.Biophys.Acta 105:43-53 (1965)). The gene associated with this enzyme is not available.

The enzyme alkaline phosphatase (EC 3.1.3.1) hydrolyses a broad range of phosphorylated substrates to their corresponding alcohols. These enzymes are typically secreted into the
 10 periplasm in bacteria, where they play a role in phosphate transport and metabolism. The *E. coli* *phoA* gene encodes a periplasmic zinc-dependent alkaline phosphatase active under conditions of phosphate starvation (Coleman Annu. Rev. Biophys. Biomol. Struct. 21:441-83 (1992)). Similar enzymes have been characterized in *Campylobacter jejuni* (van Mourik et al., Microbiol. 154:584-92 (2008)), *Saccharomyces cerevisiae* (Oshima et al., Gene 179:171-
 15 7 (1996)) and *Staphylococcus aureus* (Shah and Blobel, J. Bacteriol. 94:780-1 (1967)). Enzyme engineering and/or removal of targeting sequences may be required for alkaline phosphatase enzymes to function in the cytoplasm. Additional phosphatase enzymes are described in Example XI.

Protein	GenBank ID	GI Number	Organism
<i>phoA</i>	NP_414917.2	49176017	<i>Escherichia coli</i>
<i>phoX</i>	ZP_01072054.1	86153851	<i>Campylobacter jejuni</i>
PHO8	AAA34871.1	172164	<i>Saccharomyces cerevisiae</i>
SaurJH1_2706	YP_001317815.1	150395140	<i>Staphylococcus aureus</i>

20 The interconversion of 3-phosphoglycerate and glycerate (Figure 1, Step 11) is also catalyzed by glycerate kinase (EC 2.7.1.31). This enzyme naturally operates in the ATP-consuming phosphorylation direction and has not been shown to function in the ATP-generating direction. Three classes of glycerate kinase have been identified. Enzymes in class I and II produce glycerate-2-phosphate, whereas the class III enzymes found in plants and yeast
 25 produce glycerate-3-phosphate (Bartsch et al., FEBS Lett. 582:3025-3028 (2008)). In a recent study, class III glycerate kinase enzymes from *Saccharomyces cerevisiae*, *Oryza sativa* and *Arabidopsis thaliana* were heterologously expressed in *E. coli* and characterized (Bartsch et al., FEBS Lett. 582:3025-3028 (2008)). This study also assayed the *glxK* gene product of *E. coli* for ability to form glycerate-3-phosphate and found that the enzyme can only catalyze the

formation of glycerate-2-phosphate, in contrast to previous work (Doughty et al., *J Biol.Chem.* 241:568-572 (1966)).

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>glxK</i>	AAC73616.1	1786724	<i>Escherichia coli</i>
YGR205W	AAS56599.1	45270436	<i>Saccharomyces cerevisiae</i>
<i>Os01g0682500</i>	BAF05800.1	113533417	<i>Oryza sativa</i>
<i>Atlg80380</i>	BAH57057.1	227204411	<i>Arabidopsis thaliana</i>

EXAMPLE III

5 Pathways for producing ethylene glycol from glyoxylate via tartronate semialdehyde

Figure 2 shows a pathway for producing ethylene glycol from glyoxylate via a tartrate semialdehyde intermediate. The glyoxylate precursor may be derived from central metabolites such as isocitrate, via isocitrate lyase, or glycine, via one of several
 10 aminotransferase enzymes that utilize glycine as an amino donor such as serine:glyoxylate aminotransferase or glycine aminotransferase. In the proposed pathway, two equivalents of glyoxylate are joined by glyoxylate carboligase to form one equivalent of tartronate semialdehyde (Figure 2, Step 1). Tartronate semialdehyde is subsequently isomerized to form hydroxypyruvate by hydroxypyruvate isomerase (Figure 2, Step 2). The decarboxylation and
 15 reduction of hydroxypyruvate yield ethylene glycol as described previously (Figure 2, Steps 3 and 4). Enzyme candidates for steps 1 and 2 of Figure 2 are provided below.

Glyoxylate carboligase (EC 4.1.1.47), also known as tartrate semialdehyde synthase, catalyzes the condensation of two molecules of glyoxylate to form tartronate semialdehyde (Figure 2, Step 2). The *E. coli* enzyme, encoded by *gcl*, is active under anaerobic conditions
 20 and requires FAD for activity although no net redox reaction takes place (Chang et al., *J Biol.Chem.* 268:3911-3919 (1993)). Glyoxylate carboligase activity has also been detected in *Ralstonia eutropha*, where it is encoded by *h16_A3598* (Eschmann et al., *Arch.Microbiol.* 125:29-34 (1980)). Additional candidate glyoxylate carboligase enzyme candidates can be identified by sequence homology. Two exemplary candidates with high homology to the *E.*
 25 *coli* enzyme are found in *Salmonella enterica* and *Burkholderia ambifaria*.

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>gcl</i>	AAC73609.1	1786717	<i>Escherichia coli</i>
<i>h16_A3598</i>	YP_728024.1	113869535	<i>Ralstonia eutropha</i>
<i>SentesTy_010100014274</i>	ZP_03378393.1	213648340	<i>Salmonella enterica</i>
<i>Bamb_1815</i>	YP_773705.1	115351866	<i>Burkholderia ambifaria</i>

Hydroxypyruvate isomerase catalyzes the reversible isomerization of hydroxypyruvate and tartronate semialdehyde. The *E. coli* enzyme, encoded by *hyi*, is cotranscribed with glyoxylate carboligase (*gcl*) in a glyoxylate utilization operon (Ashiuchi et al., Biochim.Biophys.Acta 1435:153-159 (1999); Cusa et al., J Bacteriol. 181:7479-7484 (1999)).

5 This enzyme has also been purified and characterized in *Bacillus fastidiosus*, although the associated gene is not known (de Windt et al., Biochim.Biophys.Acta 613:556-562 (1980)).

Hydroxypyruvate isomerase enzyme candidates in other organisms such as *Ralstonia eutropha* and *Burkholderia ambifaria* can be identified by sequence homology to the *E. coli* gene product. Note that the predicted hydroxypyruvate isomerase gene candidates in these
10 organisms are also co-localized with genes predicted to encode glyoxylate carboligase.

Protein	GenBank ID	GI Number	Organism
<i>hyi</i>	AAC73610.1	1786718	<i>Escherichia coli</i>
<i>h16 A3599</i>	YP_728025.1	113869536	<i>Ralstonia eutropha</i>
<i>Bamb 1816</i>	YP_773706.1	115351867	<i>Burkholderia ambifaria</i>

EXAMPLE IV

Pathways for producing ethylene glycol from glyoxylate via glycolate

15 Additional pathways from glyoxylate to ethylene glycol proceed through the intermediate glycolate as shown in Figure 3. In the first step of all pathways, glyoxylate is converted to glycolate by glyoxylate reductase (Figure 3, Step 1). Glycolate is then converted to ethylene glycol by one of several routes. In one route, glycolate is converted to glycolyl-CoA by a
20 CoA transferase or synthetase (Figure 3, Step 2/3). Glycolyl-CoA is then reductively deacylated to glycolaldehyde by glycolyl-CoA reductase (aldehyde forming) (Figure 3, Step 4). Glycolaldehyde is converted to ethylene glycol by glycolaldehyde reductase as described previously (Figure 3, Step 5). Alternately, glycolyl-CoA is directly converted to ethylene glycol by a bifunctional enzyme with CoA reductase (alcohol forming) activity (Figure 3,
25 Step 10). In an alternative route, glycolate is directly converted to glycolaldehyde by a carboxylic acid reductase enzyme with glycolate reductase activity (Figure 3, Step 6). In yet another route, glycolate is converted to glycolaldehyde via a glycolylphosphate intermediate by the enzymes glycolate kinase and glycolylphosphate reductase (Figure 3, Steps 7 and 9). Alternatively, the glycolylphosphate intermediate is converted to glycolyl-CoA by
30 phosphotransglycolylase (Figure 3, Step 8). Enzyme candidates for each of these steps are provided below.

The reduction of glyoxylate to glycolate is catalyzed by glyoxylate reductase (EC 1.1.1.79 and EC 1.1.1.26). In *E. coli* this reaction is catalyzed by the products of two genes, *ghrB* and *ghrA* (Nunez et al., *Biochem.J* 354:707-715 (2001)). Both gene products utilize NADPH and also catalyze the reduction of hydroxypyruvate and the *ghrA* gene product prefers glycolate as a substrate. Eukaryotic glyoxylate reductase enzyme candidates that have been expressed in *E. coli* include the NADPH or NADH dependent *YNL274C* gene product from *S. cerevisiae* (Rintala et al., *Yeast* 24:129-136 (2007)) and GR1 of *Arabidopsis thaliana* (Hoover et al., *Can.J.Bot.* 85:896-902 (2007); Allan et al., *J Exp.Bot.* 59:2555-2564 (2008)). The yeast enzyme also catalyzes the reduction of hydroxypyruvate.

Protein	GenBank ID	GI Number	Organism
<i>ghrA</i>	NP_415551.2	90111205	<i>Escherichia coli</i>
<i>ghrB</i>	NP_418009.2	90111614	<i>Escherichia coli</i>
<i>YNL274C</i>	AAT92679.1	51012771	<i>Saccharomyces cerevisiae</i>
<i>GRI</i>	NP_566768.1	18404556	<i>Arabidopsis thaliana</i>

The activation of glycolate to glycolyl-CoA is catalyzed by an enzyme with glycolyl-CoA transferase activity. Such an enzyme has not been characterized to date. Glutaconyl-CoA transferase (EC 2.8.3.12) catalyzes the transfer of the 2-hydroxyacid, 2-hydroxyglutarate, to CoA. The glutaconyl-CoA-transferase (EC 2.8.3.12) enzyme from anaerobic bacterium *Acidaminococcus fermentans* reacts with a range of substrates including 2-hydroxyglutarate, glutarate, crotonate, adipate and acrylate (Buckel et al., *Eur.J Biochem.* 118:315-321 (1981); Mack et al., *Eur. J. Biochem.* 226:41-51 (1994)). The genes encoding this enzyme, *gctA* and *gctB*, have been cloned and functionally expressed in *E. coli* (Mack et al., *supra*).

Protein	GenBank ID	GI Number	Organism
<i>gctA</i>	559392	CAA57199.1	<i>Acidaminococcus fermentans</i>
<i>gctB</i>	559393	CAA57200.1	<i>Acidaminococcus fermentans</i>

The ATP-dependent acylation of glycolylate to glycolyl-CoA (Figure 3, Step 3) is catalyzed by a glycolyl-CoA synthetase or acid-thiol ligase. Enzymes catalyzing this exact transformation have not been characterized to date; however, several enzymes with broad substrate specificities have been described in the literature. ADP-forming acetyl-CoA synthetase (ACD, EC 6.2.1.13) is an enzyme that couples the conversion of acids to their corresponding acyl-CoA esters with the concomitant consumption of ATP. ACD I from *Archaeoglobus fulgidus*, encoded by AF1211, was shown to operate on a variety of linear and branched-chain substrates including isobutyrate, isopentanoate, and fumarate (Musfeldt et al.,

J Bacteriol. 184:636-644 (2002)). A second reversible ACD in *Archaeoglobus fulgidus*, encoded by AF1983, was also shown to have a broad substrate range with high activity on cyclic compounds phenylacetate and indoleacetate (Musfeldt et al., *supra*). The enzyme from *Haloarcula marismortui*, annotated as a succinyl-CoA synthetase, accepts propionate, butyrate, and branched-chain acids (isovalerate and isobutyrate) as substrates, and was shown to operate in the forward and reverse directions (Brasen et al., *Arch. Microbiol* 182:277-287 (2004)). The ACD encoded by PAE3250 from hyperthermophilic crenarchaeon *Pyrobaculum aerophilum* showed the broadest substrate range of all characterized ACDs, reacting with acetyl-CoA, isobutyryl-CoA (preferred substrate) and phenylacetyl-CoA (Brasen and Schonheit, *Arch. Microbiol* 182:277-287 (2004)). Directed evolution or engineering can be used to modify this enzyme to operate at the physiological temperature of the host organism. The enzymes from *A. fulgidus*, *H. marismortui* and *P. aerophilum* have all been cloned, functionally expressed, and characterized in *E. coli* (Brasen and Schonheit, *Arch. Microbiol* 182:277-287 (2004); Musfeldt and Schonheit, *J Bacteriol.* 184:636-644 (2002)). An additional candidate is the enzyme encoded by *sucCD* in *E. coli*, which naturally catalyzes the formation of succinyl-CoA from succinate with the concomitant consumption of one ATP, a reaction which is reversible *in vivo* (Buck et al., *Biochemistry* 24:6245-6252 (1985)). The acyl CoA ligase from *Pseudomonas putida* has been demonstrated to work on several aliphatic substrates including acetic, propionic, butyric, valeric, hexanoic, heptanoic, and octanoic acids and on aromatic compounds such as phenylacetic and phenoxyacetic acids (Fernandez-Valverde et al., *Appl. Environ. Microbiol.* 59:1149-1154 (1993)). A related enzyme, malonyl CoA synthetase (6.3.4.9) from *Rhizobium leguminosarum* could convert several diacids, namely, ethyl-, propyl-, allyl-, isopropyl-, dimethyl-, cyclopropyl-, cyclopropylmethylene-, cyclobutyl-, and benzyl-malonate into their corresponding monothioesters (Pohl et al., *J. Am. Chem. Soc.* 123:5822-5823 (2001)).

Protein	GenBank ID	GI Number	Organism
AF1211	NP_070039.1	11498810	<i>Archaeoglobus fulgidus</i>
AF1983	NP_070807.1	11499565	<i>Archaeoglobus fulgidus</i>
<i>scs</i>	YP_135572.1	55377722	<i>Haloarcula marismortui</i>
PAE3250	NP_560604.1	18313937	<i>Pyrobaculum aerophilum</i> str. IM2
<i>sucC</i>	NP_415256.1	16128703	<i>Escherichia coli</i>
<i>sucD</i>	AAC73823.1	1786949	<i>Escherichia coli</i>
<i>paaF</i>	AAC24333.2	22711873	<i>Pseudomonas putida</i>
<i>matB</i>	AAC83455.1	3982573	<i>Rhizobium leguminosarum</i>

Several acyl-CoA dehydrogenases are capable of reducing an acyl-CoA to its corresponding aldehyde and can be used for catalyzing the glycolyl-CoA reductase (aldehyde forming) activity. Exemplary genes that encode such enzymes include the *Acinetobacter calcoaceticus* *acrI* encoding a fatty acyl-CoA reductase (Reiser et al., *J. Bacteriol.* 179:2969-2975 (1997)), the *Acinetobacter sp. M-1* fatty acyl-CoA reductase (Ishige et al., *Appl. Environ. Microbiol.* 68:1192-1195 (2002)), and a CoA- and NADP- dependent succinate semialdehyde dehydrogenase encoded by the *sucD* gene in *Clostridium kluyveri* (Sohling et al., *J. Bacteriol.* 178:871-880 (1996)). *SucD* of *P. gingivalis* is another succinate semialdehyde dehydrogenase (Takahashi et al., *J. Bacteriol.* 182:4704-4710 (2000)). The enzyme acylating acetaldehyde dehydrogenase in *Pseudomonas sp.*, encoded by *bphG*, is yet another as it has been demonstrated to oxidize and acylate acetaldehyde, propionaldehyde, butyraldehyde, isobutyraldehyde and formaldehyde (Powlowski et al., 175:377-385 (1993)). In addition to reducing acetyl-CoA to ethanol, the enzyme encoded by *adhE* in *Leuconostoc mesenteroides* has been shown to oxidize the branched chain compound isobutyraldehyde to isobutyryl-CoA (Koo et al., *Biotechnol Lett.* 27:505-510 (2005)). Butyraldehyde dehydrogenase catalyzes a similar reaction, conversion of butyryl-CoA to butyraldehyde, in solventogenic organisms such as *Clostridium saccharoperbutylacetonicum* (Kosaka et al., *Biosci. Biotechnol Biochem.* 71:58-68 (2007)).

Protein	GenBank ID	GI Number	Organism
<i>acrI</i>	YP_047869.1	50086359	<i>Acinetobacter calcoaceticus</i>
<i>acrI</i>	AAC45217	1684886	<i>Acinetobacter baylyi</i>
<i>acrI</i>	BAB85476.1	18857901	<i>Acinetobacter sp.</i> Strain M-1
<i>sucD</i>	P38947.1	172046062	<i>Clostridium kluyveri</i>
<i>sucD</i>	NP_904963.1	34540484	<i>Porphyromonas gingivalis</i>
<i>bphG</i>	BAA03892.1	425213	<i>Pseudomonas sp</i>
<i>adhE</i>	AAV66076.1	55818563	<i>Leuconostoc mesenteroides</i>
<i>bld</i>	AAP42563.1	31075383	<i>Clostridium saccharoperbutylacetonicum</i>

An additional enzyme type that converts an acyl-CoA to its corresponding aldehyde is malonyl-CoA reductase which transforms malonyl-CoA to malonic semialdehyde. Malonyl-CoA reductase is a key enzyme in autotrophic carbon fixation via the 3-hydroxypropionate cycle in thermoacidophilic archaeal bacteria (Berg et al., *Science* 318:1782-1786 (2007); Thauer, *Science* 318:1732-1733 (2007)). The enzyme utilizes NADPH as a cofactor and has been characterized in *Metallosphaera* and *Sulfolobus* spp (Alber et al., *J. Bacteriol.* 188:8551-8559 (2006); Hugler et al., *J. Bacteriol.* 184:2404-2410 (2002)). The enzyme is

encoded by *Msed_0709* in *Metallosphaera sedula* (Alber et al., *supra*; Berg et al., *supra*). A gene encoding a malonyl-CoA reductase from *Sulfolobus tokodaii* was cloned and heterologously expressed in *E. coli* (Alber et al., *supra*). This enzyme has also been shown to catalyze the conversion of methylmalonyl-CoA to its corresponding aldehyde

(WO/2007/141208). Although the aldehyde dehydrogenase functionality of these enzymes is similar to the bifunctional dehydrogenase from *Chloroflexus aurantiacus*, there is little sequence similarity. Both malonyl-CoA reductase enzyme candidates have high sequence similarity to aspartate-semialdehyde dehydrogenase, an enzyme catalyzing the reduction and concurrent dephosphorylation of aspartyl-4-phosphate to aspartate semialdehyde. Additional gene candidates can be found by sequence homology to proteins in other organisms including *Sulfolobus solfataricus* and *Sulfolobus acidocaldarius*. Yet another acyl-CoA reductase (aldehyde forming) candidate is the *ald* gene from *Clostridium beijerinckii* (Toth et al., *Appl Environ. Microbiol* 65:4973-4980 (1999)). This enzyme has been reported to reduce acetyl-CoA and butyryl-CoA to their corresponding aldehydes. This gene is very similar to *eutE* that encodes acetaldehyde dehydrogenase of *Salmonella typhimurium* and *E. coli* (Toth et al., *supra*).

Protein	GenBank ID	GI Number	Organism
<i>MSED_0709</i>	YP_001190808.1	146303492	<i>Metallosphaera sedula</i>
<i>mcr</i>	NP_378167.1	15922498	<i>Sulfolobus tokodaii</i>
<i>asd-2</i>	NP_343563.1	15898958	<i>Sulfolobus solfataricus</i>
<i>Saci_2370</i>	YP_256941.1	70608071	<i>Sulfolobus acidocaldarius</i>
<i>Ald</i>	AAT66436	9473535	<i>Clostridium beijerinckii</i>
<i>eutE</i>	AAA80209	687645	<i>Salmonella typhimurium</i>
<i>eutE</i>	P77445	2498347	<i>Escherichia coli</i>

Direct conversion of glycolate to glycolaldehyde is catalyzed by an acid reductase enzyme with glycolate reductase activity. Exemplary enzymes include carboxylic acid reductase, alpha-aminoacidate reductase and retinoic acid reductase. Carboxylic acid reductase (CAR), found in *Nocardia iowensis*, catalyzes the magnesium, ATP and NADPH-dependent reduction of carboxylic acids to their corresponding aldehydes (Venkitasubramanian et al., *J Biol.Chem.* 282:478-485 (2007)). The natural substrate of this enzyme is vanillic acid and the enzyme exhibits broad acceptance of aromatic and aliphatic substrates (Venkitasubramanian et al., 425-440 (2006)). This enzyme, encoded by *car*, was cloned and functionally expressed in *E. coli* (Venkitasubramanian et al., *J Biol.Chem.* 282:478-485 (2007)). CAR requires post-translational activation by a phosphopantetheine transferase (PPTase) that converts the inactive apo-enzyme to the active holo-enzyme (Hansen et al., *Appl.Environ.Microbiol*

75:2765-2774 (2009)). Expression of the *npt* gene, encoding a specific PPTase, product improved activity of the enzyme. An additional enzyme candidate found in *Streptomyces griseus* is encoded by the *griC* and *griD* genes. This enzyme is believed to convert 3-amino-4-hydroxybenzoic acid to 3-amino-4-hydroxybenzaldehyde as deletion of either *griC* or *griD* led to accumulation of extracellular 3-acetylamin-4-hydroxybenzoic acid, a shunt product of 3-amino-4-hydroxybenzoic acid metabolism (Suzuki, et al., *J. Antibiot.* 60(6):380-387 (2007)). Co-expression of *griC* and *griD* with SGR_665, an enzyme similar in sequence to the *Nocardia iowensis npt*, can be beneficial.

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>car</i>	AAR91681.1	40796035	<i>Nocardia iowensis</i>
<i>npt</i>	ABI83656.1	114848891	<i>Nocardia iowensis</i>
<i>griC</i>	YP_001825755.1	182438036	<i>Streptomyces griseus</i>
<i>griD</i>	YP_001825756.1	182438037	<i>Streptomyces griseus</i>

10 An enzyme with similar characteristics, alpha-aminoadipate reductase (AAR, EC 1.2.1.31), participates in lysine biosynthesis pathways in some fungal species. This enzyme naturally reduces alpha-aminoadipate to alpha-aminoadipate semialdehyde. The carboxyl group is first activated through the ATP-dependent formation of an adenylate that is then reduced by NAD(P)H to yield the aldehyde and AMP. Like CAR, this enzyme utilizes magnesium and requires activation by a PPTase. Enzyme candidates for AAR and its corresponding PPTase are found in *Saccharomyces cerevisiae* (Morris et al., *Gene* 98:141-145 (1991)), *Candida albicans* (Guo et al., *Mol.Genet.Genomics* 269:271-279 (2003)), and *Schizosaccharomyces pombe* (Ford et al., *Curr.Genet.* 28:131-137 (1995)). The AAR from *S. pombe* exhibited significant activity when expressed in *E. coli* (Guo et al., *Yeast* 21:1279-1288 (2004)). The AAR from *Penicillium chrysogenum* accepts S-carboxymethyl-L-cysteine as an alternate substrate, but did not react with adipate, L-glutamate or diaminopimelate (Hijarrubia et al., *J Biol.Chem.* 278:8250-8256 (2003)). The gene encoding the *P. chrysogenum* PPTase has not been identified to date and no high-confidence hits were identified by sequence comparison homology searching.

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>LYS2</i>	AAA34747.1	171867	<i>Saccharomyces cerevisiae</i>
<i>LYS5</i>	P50113.1	1708896	<i>Saccharomyces cerevisiae</i>
<i>LYS2</i>	AAC02241.1	2853226	<i>Candida albicans</i>
<i>LYS5</i>	AAO26020.1	28136195	<i>Candida albicans</i>

Protein	GenBank ID	GI Number	Organism
<i>LysIp</i>	P40976.3	13124791	<i>Schizosaccharomyces pombe</i>
<i>Lys7p</i>	Q10474.1	1723561	<i>Schizosaccharomyces pombe</i>
<i>Lys2</i>	CAA74300.1	3282044	<i>Penicillium chrysogenum</i>

Kinase or phosphotransferase enzymes transform carboxylic acids to phosphonic acids with concurrent hydrolysis of one ATP. Such an enzyme with glycolate kinase activity is required to convert glycolate to glycoylphosphate (Figure 3, Step 7). This exact transformation has

not been demonstrated to date. Exemplary enzyme candidates include butyrate kinase (EC 2.7.2.7), isobutyrate kinase (EC 2.7.2.14), aspartokinase (EC 2.7.2.4), acetate kinase (EC 2.7.2.1) and gamma-glutamyl kinase (EC 2.7.2.11). Butyrate kinase catalyzes the reversible conversion of butyryl-phosphate to butyrate during acidogenesis in Clostridial species (Cary et al., Appl.Environ.Microbiol 56:1576-1583 (1990)). The Clostridium acetobutylicum

enzyme is encoded by either of the two buk gene products (Huang et al., J Mol.Microbiol Biotechnol 2:33-38 (2000)). Other butyrate kinase enzymes are found in C. butyricum and C. tetanomorphum (TWAROG et al., J Bacteriol. 86:112-117 (1963)). A related enzyme, isobutyrate kinase from Thermotoga maritima, was expressed in E. coli and crystallized (Diao et al., J Bacteriol. 191:2521-2529 (2009); Diao et al., Acta

Crystallogr.D.Biol.Crystallogr. 59:1100-1102 (2003)). Aspartokinase catalyzes the ATP-dependent phosphorylation of aspartate and participates in the synthesis of several amino acids. The aspartokinase III enzyme in E. coli, encoded by lysC, has a broad substrate range and the catalytic residues involved in substrate specificity have been elucidated (Keng et al., Arch.Biochem.Biophys. 335:73-81 (1996)). Two additional kinases in E. coli are also good candidates: acetate kinase and gamma-glutamyl kinase. The E. coli acetate kinase, encoded by ackA karstedt et al., J.Biol.Chem. 251:6775-6783 (1976)), phosphorylates propionate in addition to acetate (Hesslinger et al., Mol.Microbiol 27:477-492 (1998)). The E. coli gamma-glutamyl kinase, encoded by proB (Smith et al., J.Bacteriol. 157:545-551 (1984a)), phosphorylates the gamma carbonic acid group of glutamate.

Protein	GenBank ID	GI Number	Organism
<i>buk1</i>	NP_349675	15896326	<i>Clostridium acetobutylicum</i>
<i>buk2</i>	Q97III	20137415	<i>Clostridium acetobutylicum</i>
<i>buk2</i>	Q9X278.1	6685256	<i>Thermotoga maritima</i>
<i>lysC</i>	NP_418448.1	16131850	<i>Escherichia coli</i>
<i>ackA</i>	NP_416799.1	16130231	<i>Escherichia coli</i>
<i>proB</i>	NP_414777.1	16128228	<i>Escherichia coli</i>

An enzyme with phosphotransglycolylase activity is required to convert glycolylphosphate to glycolyl-CoA (Figure 3, Step 8). Exemplary phosphate-transferring acyltransferases include phosphotransacetylase (EC 2.3.1.8) and phosphotransbutyrylase (EC 2.3.1.19). The *pta* gene from *E. coli* encodes a phosphotransacetylase that reversibly converts acetyl-CoA into acetyl-phosphate (Suzuki, *Biochim.Biophys.Acta* 191:559-569 (1969)). This enzyme can also utilize propionyl-CoA as a substrate, forming propionate in the process (Hesslinger et al., *Mol.Microbiol* 27:477-492 (1998)). Other phosphate acetyltransferases that exhibit activity on propionyl-CoA are found in *Bacillus subtilis* (Rado et al., *Biochim.Biophys.Acta* 321:114-125 (1973)), *Clostridium kluyveri* (Stadtman, 1:596-599 (1955)), and *Thermotoga maritima* (Bock et al., *J Bacteriol.* 181:1861-1867 (1999)). Similarly, the *ptb* gene from *C. acetobutylicum* encodes phosphotransbutyrylase, an enzyme that reversibly converts butyryl-CoA into butyryl-phosphate (Wiesenborn et al., *Appl Environ.Microbiol* 55:317-322 (1989); Walter et al., *Gene* 134:107-111 (1993)). Additional *ptb* genes are found in butyrate-producing bacterium L2-50 (Louis et al., *J.Bacteriol.* 186:2099-2106 (2004)) and *Bacillus megaterium* (Vazquez et al., *Curr.Microbiol* 42:345-349 (2001)).

Protein	GenBank ID	GI Number	Organism
<i>pta</i>	NP_416800.1	71152910	<i>Escherichia coli</i>
<i>pta</i>	P39646	730415	<i>Bacillus subtilis</i>
<i>pta</i>	A5N801	146346896	<i>Clostridium kluyveri</i>
<i>pta</i>	Q9X0L4	6685776	<i>Thermotoga maritima</i>
<i>ptb</i>	NP_349676	34540484	<i>Clostridium acetobutylicum</i>
<i>ptb</i>	AAR19757.1	38425288	butyrate-producing bacterium L2-50
<i>ptb</i>	CAC07932.1	10046659	<i>Bacillus megaterium</i>

The conversion of glycolylphosphate to glycolaldehyde is catalyzed by glycolylphosphate reductase (Figure 3, Step 9). Although an enzyme catalyzing this conversion has not been identified to date, similar transformations catalyzed by glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12), aspartate-semialdehyde dehydrogenase (EC 1.2.1.11) acetylglutamylphosphate reductase (EC 1.2.1.38) and glutamate-5-semialdehyde dehydrogenase (EC 1.2.1.) are well documented. Aspartate semialdehyde dehydrogenase (ASD, EC 1.2.1.11) catalyzes the NADPH-dependent reduction of 4-aspartyl phosphate to aspartate-4-semialdehyde. ASD participates in amino acid biosynthesis and recently has been studied as an antimicrobial target (Hadfield et al., *Biochemistry* 40:14475-14483 (2001)). The *E. coli* ASD structure has been solved (Hadfield et al., *J Mol.Biol.* 289:991-1002 (1999)) and the enzyme has been shown to accept the alternate substrate beta-3-methylaspartyl phosphate

(Shames et al., *J Biol.Chem.* 259:15331-15339 (1984)). The *Haemophilus influenzae* enzyme has been the subject of enzyme engineering studies to alter substrate binding affinities at the active site (Blanco et al., *Acta Crystallogr.D.Biol.Crystallogr.* 60:1388-1395 (2004)). Other ASD candidates are found in *Mycobacterium tuberculosis* (Shafiani et al., *J Appl Microbiol* 5 98:832-838 (2005)), *Methanococcus jannaschii* (Faehnle et al., *J Mol.Biol.* 353:1055-1068 (2005)), and the infectious microorganisms *Vibrio cholera* and *Helicobacter pylori* (Moore et al., *Protein Expr.Purif.* 25:189-194 (2002)). A related enzyme candidate is acetylglutamylphosphate reductase (EC 1.2.1.38), an enzyme that naturally reduces acetylglutamylphosphate to acetylglutamate-5-semialdehyde, found in *S. cerevisiae* (Pauwels 10 et al., *Eur.J Biochem.* 270:1014-1024 (2003)), *B. subtilis* (O'Reilly et al., *Microbiology* 140 (Pt 5):1023-1025 (1994)), *E. coli* (Parsot et al., *Gene.* 68:275-283 (1988)), and other organisms. Additional phosphate reductase enzymes of *E. coli* include glyceraldehyde 3-phosphate dehydrogenase encoded by *gapA* (Branlant et al., *Eur.J.Biochem.* 150:61-66 (1985)) and glutamate-5-semialdehyde dehydrogenase encoded by *proA* (Smith et al., 15 *J.Bacteriol.* 157:545-551 (1984b)). Genes encoding glutamate-5-semialdehyde dehydrogenase enzymes from *Salmonella typhimurium* (Mahan et al., *J Bacteriol.* 156:1249-1262 (1983)) and *Campylobacter jejuni* (Louie et al., *Mol.Gen.Genet.* 240:29-35 (1993)) were cloned and expressed in *E. coli*.

Protein	GenBank ID	GI Number	Organism
<i>asd</i>	NP_417891.1	16131307	<i>Escherichia coli</i>
<i>asd</i>	YP_248335.1	68249223	<i>Haemophilus influenzae</i>
<i>asd</i>	AAB49996	1899206	<i>Mycobacterium tuberculosis</i>
VC2036	NP_231670	15642038	<i>Vibrio cholera</i>
<i>asd</i>	YP_002301787.1	210135348	<i>Helicobacter pylori</i>
<i>ARG5,6</i>	NP_010992.1	6320913	<i>Saccharomyces cerevisiae</i>
<i>argC</i>	NP_389001.1	16078184	<i>Bacillus subtilis</i>
<i>argC</i>	NP_418393.1	16131796	<i>Escherichia coli</i>
<i>gapA</i>	P0A9B2.2	71159358	<i>Escherichia coli</i>
<i>proA</i>	NP_414778.1	16128229	<i>Escherichia coli</i>
<i>proA</i>	NP_459319.1	16763704	<i>Salmonella typhimurium</i>
<i>proA</i>	P53000.2	9087222	<i>Campylobacter jejuni</i>

20 The direct formation of ethylene glycol from glycolyl-CoA is catalyzed by a bifunctional enzyme with glycolyl-CoA reductase (alcohol forming) activity (Figure 3, Step 10). Exemplary bifunctional oxidoreductases that convert acyl-CoA molecules to their corresponding alcohols include enzymes that transform substrates such as acetyl-CoA to

ethanol (e.g., *adhE* from *E. coli* (Kessler et al., *FEBS.Lett.* 281:59-63 (1991))), butyryl-CoA to butanol (e.g. *adhE2* from *C. acetobutylicum* (Fontaine et al., *J.Bacteriol.* 184:821-830 (2002))) and malonyl-CoA to 3-hydroxypropanoate (e.g. *mcr* from *Chloroflexus aurantiacus* (Hugler et al., *J.Bacteriol.* 184:2404-2410 (2002))). In addition to reducing acetyl-CoA to ethanol, the enzyme encoded by *adhE* in *Leuconostoc mesenteroides* has been shown to oxidize the branched chain compound isobutyraldehyde to isobutyryl-CoA (Kazahaya et al., *J.Gen.Appl.Microbiol.* 18:43-55 (1972); Koo et al., *supra*). The NADPH-dependent alcohol-forming malonyl-CoA reductase of *Chloroflexus aurantiacus* participates in the 3-hydroxypropionate cycle (Hugler et al., 184:2404-2410 (2002); Strauss et al., *Eur.J.Biochem.* 215:633-643 (1993)). This enzyme, with a mass of 300 kDa, is highly substrate-specific and shows little sequence similarity to other known oxidoreductases (Hugler et al., *supra*). No enzymes in other organisms have been shown to catalyze this specific reaction; however there is bioinformatic evidence that other organisms may have similar pathways (Klatt et al., *supra*). Enzyme candidates in other organisms including *Roseiflexus castenholzii*, *Erythrobacter sp. NAP1* and marine gamma proteobacterium HTCC2080 can be inferred by sequence similarity.

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>adhE</i>	NP_415757.1	16129202	<i>Escherichia coli</i>
<i>adhE2</i>	AAK09379.1	12958626	<i>Clostridium acetobutylicum</i>
<i>adhE</i>	AAV66076.1	55818563	<i>Leuconostoc mesenteroides</i>
<i>mcr</i>	AAS20429.1	42561982	<i>Chloroflexus aurantiacus</i>
<i>Rcas 2929</i>	YP_001433009.1	156742880	<i>Roseiflexus castenholzii</i>
<i>NAP1_02720</i>	ZP_01039179.1	85708113	<i>Erythrobacter sp. NAP1</i>
<i>MGP2080_00535</i>	ZP_01626393.1	119504313	marine gamma proteobacterium <i>HTCC2080</i>

Longer chain acyl-CoA molecules can be reduced to their corresponding alcohols by enzymes such as the jojoba (*Simmondsia chinensis*) *FAR* which encodes an alcohol-forming fatty acyl-CoA reductase. Its overexpression in *E. coli* resulted in *FAR* activity and the accumulation of fatty alcohol (Metz et al., *Plant Physiol.* 122:635-644 (2000)).

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>FAR</i>	AAD38039.1	5020215	<i>Simmondsia chinensis</i>

EXAMPLE V

Pathways for converting glycerate to ethylene glycol

Pathways for converting glycerate to ethylene glycol are shown in Figure 1. Glycerate is a common metabolic intermediate in diverse metabolic biosynthetic and degradation pathways including the non-phosphorylative Entner-Doudoroff pathway, the serine pathway of formaldehyde assimilation and glyoxylate degradation. Glycerate can also be formed by oxidation of glyceraldehyde by glyceraldehyde dehydrogenase or glyceraldehyde oxidase (Step 14 of Figure 1) or dephosphorylation of 3-phosphoglycerate or 2-phosphoglycerate by either a phosphatase (Steps 10 and 12 of Figure 1) or a kinase operating in the reverse direction (Steps 13 and 11 of Figure 1). The glycerate is then converted to ethylene glycol by one of several pathways. In one pathway, glycerate is directly converted to ethylene glycol by a decarboxylase (Step 9 of Figure 1). Candidate enzymes for this decarboxylase were presented in Example I. In an alternate route, glycerate is oxidized to hydroxypyruvate by hydroxypyruvate reductase (Step 8 of Figure 1). The hydroxypyruvate intermediate is then decarboxylated and reduced to ethylene glycol by enzymes described in Example I (Figure 1, Steps 3, 4). In yet another route, glycerate is converted to hydroxypyruvate in two steps: oxidation to tartronate semialdehyde by glycerate dehydrogenase, followed by isomerization to hydroxypyruvate by hydroxypyruvate isomerase (Steps 5 and 2 of Figure 2). Enzyme candidates for hydroxypyruvate isomerase were described in Example III. Enzyme candidates for 2-phosphoglycerate phosphatase, glycerate-2-kinase, glyceraldehyde dehydrogenase, glyceraldehyde oxidase and glycerate dehydrogenase are provided below.

2-Phosphoglycerate phosphatase (EC 3.1.3.20) catalyzes the hydrolysis of 2PG to glycerate, releasing pyrophosphate (Figure 1, Step 12). This enzyme was purified from cell extracts of *Veillonella alcalescens* (Pestka et al., *Can. J Microbiol* 27:808-814 (1981)), where it is thought to participate in a serine biosynthetic pathway. A similar enzyme was also characterized in beef liver (Fallon et al., *Biochim Biophys Acta* 105:43-53 (1965)). However, genes have not been associated with either enzyme to date. Additional 2PG phosphatase enzyme candidates are alkaline phosphatase (EC 3.1.3.1) and acid phosphatase (EC 3.1.3.2). Both enzymes hydrolyze a broad range of phosphorylated substrates to their corresponding alcohols. Alkaline phosphatase enzymes are typically secreted into the periplasm in bacteria, where they play a role in phosphate transport and metabolism. The *E. coli phoA* gene encodes a periplasmic zinc-dependent alkaline phosphatase active under conditions of phosphate starvation (Coleman *Annu. Rev. Biophys. Biomol. Struct.* 21:441-83 (1992)). Similar enzymes

have been characterized in *Campylobacter jejuni* (van Mourik et al., *Microbiol.* 154:584-92 (2008)), *Saccharomyces cerevisiae* (Oshima et al., *Gene* 179:171-7 (1996)) and *Staphylococcus aureus* (Shah and Blobel, *J. Bacteriol.* 94:780-1 (1967)). Enzyme engineering and/or removal of targeting sequences may be required for alkaline phosphatase enzymes to function in the cytoplasm. Acid phosphatase enzymes from *Brassica nigra*, *Lupinus luteus* and *Phaseolus vulgaris* have been shown to catalyze the hydrolysis of 2PG to glycerate (Yoneyama et al., *J Biol Chem* 279:37477-37484 (2004); Olczak et al., *Biochim Biophys Acta* 1341:14-25 (1997); Duff et al., *Arch.Biochem.Biophys* 286:226-232 (1991)). Only the *P. vulgaris* enzyme has been associated with a gene to date.

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>phoA</i>	NP_414917.2	49176017	<i>Escherichia coli</i>
<i>phoX</i>	ZP_01072054.1	86153851	<i>Campylobacter jejuni</i>
PHO8	AAA34871.1	172164	<i>Saccharomyces cerevisiae</i>
SaurJH1_2706	YP_001317815.1	150395140	<i>Staphylococcus aureus</i>
<i>KeACP</i>	BAD05167.1	40217508	<i>Phaseolus vulgaris</i>

The ATP-dependent interconversion of 2-phosphoglycerate and glycerate (Figure 1, Step 13) is catalyzed by glycerate-2-kinase (EC 2.7.1.165). This enzyme naturally operates in the phosphorylating ATP-consuming direction and has not been shown to function in the reverse ATP-generating direction. Glycerate-2-kinase enzymes have been studied in animals, methylotrophic bacteria and organisms which utilize a branched Entner-Doudoroff pathway. Exemplary gene candidates include *glxK* of *E. coli* (Bartsch et al., *FEBS Lett.* 582:3025-3028 (2008)), *ST2037* of *Sulfolobus tolodaii*, *garK* of *Thermoproteus tenax* and *Pto1442* from *Picrophilus torridus* (Liu et al., *Biotechnol Lett.* 31:1937-1941 (2009); Kehrer et al., *BMC Genomics* 8:301 (2007); Reher et al., *FEMS Microbiol Lett.* 259:113-119 (2006)). The thermostable enzymes of *S. tolodaii* and *T. tenax* have been cloned and characterized in *E. coli*. Several enzymes in this class are inhibited by ADP, so removal or attenuation of this inhibition may be necessary for the enzyme to operate in the desired direction.

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>glxK</i>	AAC73616.1	1786724	<i>Escherichia coli</i>
<i>ST2037</i>	NP_378024.1	15922355	<i>Sulfolobus tolodaii</i>
<i>garK</i>	AJ621354	41033736	<i>Thermoproteus tenax</i>
<i>Pto1442</i>	YP_024220.1	48478514	<i>Picrophilus torridus</i>

Glyceraldehyde dehydrogenase catalyzes the oxidation of glyceraldehyde to glycerate. This reaction can be catalyzed by many NAD(P)⁺-dependent oxidoreductases in the EC class

1.2.1. Exemplary enzymes that catalyze this conversion include the glutarate semialdehyde dehydrogenase (EC 1.2.1.20) of *Pseudomonas putida*, lactate dehydrogenase (EC 1.2.1.22) of *Methanocaldococcus jannaschii*, the betaine-aldehyde dehydrogenase (EC 1.2.1.8) of *E. coli* (Gruez et al., *J Mol Biol* 343:29-41 (2004)) and the succinate semialdehyde dehydrogenase (EC 1.2.1.24) of *Azospirillum brasilense* (Watanabe et al., *J Biol Chem* 281:28876-28888 (2006); Grochowski et al., *J Bacteriol.* 188:2836-2844 (2006); Chang et al., *J Biol Chem* 252:7979-7986 (1977)). The NAD⁺- and NADP⁺-dependent aldehyde dehydrogenase enzymes (EC 1.2.1.3 and EC 1.2.1.4 and EC 1.2.1.5) are also suitable candidates. Some gene products with activity on glyceraldehyde include the NADP⁺ dependent enzyme from *Acetobacter aceti* and ALDH from *Saccharomyces cerevisiae* (Vandecasteele et al., *Methods Enzymol.* 89 Pt D:484-490 (1982); Tamaki et al., *J Biochem.* 82:73-79 (1977)). The NAD⁺-dependent aldehyde dehydrogenases (EC 1.2.1.3) found in human liver, ALDH-1 and ALDH-2, have broad substrate ranges for a variety of aliphatic, aromatic and polycyclic aldehydes (Klyosov, *Biochemistry* 35:4457-4467 (1996)). Active ALDH-2 has been efficiently expressed in *E. coli* using the GroEL proteins as chaperonins (Lee et al., *Biochem.Biophys.Res.Commun.* 298:216-224 (2002)). The rat mitochondrial aldehyde dehydrogenase also has a broad substrate range (Siew et al., *Arch.Biochem.Biophys.* 176:638-649 (1976)). The *E. coli* gene *astD* also encodes an NAD⁺-dependent aldehyde dehydrogenase (Kuznetsova et al., *FEMS Microbiol Rev* 29:263-279 (2005)).

Protein	GenBank ID	GI Number	Organism
<i>MJI411</i>	NP_248414.1	15669601	<i>Methanocaldococcus jannaschii</i>
<i>araE</i>	BAE94276.1	95102056	<i>Azospirillum brasilense</i>
<i>ydcW</i>	NP_415961.1	16129403	<i>Escherichia coli</i>
<i>ALDH</i>	AAA34419.1	171048	<i>Saccharomyces cerevisiae</i>
<i>ALDH-2</i>	P05091.2	118504	<i>Homo sapiens</i>
<i>ALDH-2</i>	NP_115792.1	14192933	<i>Rattus norvegicus</i>
<i>astD</i>	P76217.1	3913108	<i>Escherichia coli</i>
<i>PF0346</i>	NP_578075.1	18976718	<i>Pyrococcus furiosus</i>

Aldehyde oxidase enzymes (EC 1.2.3.1) can also catalyze the conversion of glyceraldehyde, water and oxygen to glycerate and hydrogen peroxide. Aldehyde oxidase enzymes in organisms such as *Streptomyces moderatus*, *Pseudomonas* sp. and *Methylobacillus* sp. catalyze the oxidation of a wide range of aldehydes including formaldehyde, aromatic, and aliphatic aldehydes including glyceraldehyde (Koshiba et al., *Plant Physiol* 110:781-789 (1996)). Although the genes associated with these enzymes are not known to date, the *zmAO*-

1 and *zmAO-2* genes of *Zea mays* encode flavin- and molybdenum- containing aldehyde oxidase isozymes (Sekimoto et al., *J Biol.Chem.* 272:15280-15285 (1997)). Additional glyceraldehyde oxidase candidates can be inferred by sequence homology to the *Z. mays* genes and are shown below.

Gene	GenBank Accession No.	GI No.	Organism
<i>zmAO-1</i>	NP_001105308.1	162458742	<i>Zea mays</i>
<i>zmAO-2</i>	BAA23227.1	2589164	<i>Zea mays</i>
<i>Aox1</i>	O54754.2	20978408	<i>Mus musculus</i>
<i>ALDO1_ORYSJ</i>	Q7XH05.1	75296231	<i>Oryza sativa</i>
<i>AAO3</i>	BAA82672.1	5672672	<i>Arabidopsis thaliana</i>
<i>XDH</i>	DAA24801.1	296482686	<i>Bos taurus</i>

5

Oxidation of glycerate to tartronate semialdehyde is catalyzed by glycerate dehydrogenase (EC 1.1.1.60). Two isozymes of this enzyme are encoded by the genes *garR* and *glxR* of *E. coli* (Cusa et al., *J Bacteriol.* 181:7479-7484 (1999); Monterrubio et al., *J Bacteriol.* 182:2672-2674 (2000); Njau et al., *J Biol Chem* 275:38780-38786 (2000)). The glycerate dehydrogenase encoded by *garR* of *Salmonella typhimurium* subsp. *enterica* serovar Typhimurium was recently crystallized (Osipiuk et al., *J Struct.Funct.Genomics* 10:249-253 (2009)).

10

Gene	GenBank Accession No.	GI No.	Organism
<i>garR</i>	AAC76159.3	145693186	<i>Escherichia coli</i>
<i>glxR</i>	AAC73611.1	1786719	<i>Escherichia coli</i>
<i>garR</i>	NP_462161.1	16766546	<i>Salmonella typhimurium</i>

Additional candidate alcohol dehydrogenases for converting glycerate to tartronate semialdehyde include medium-chain alcohol dehydrogenase, 4-hydroxybutyrate dehydrogenase and 3-hydroxyisobutyrate dehydrogenase. Exemplary genes encoding medium-chain alcohol dehydrogenase enzymes that catalyze the conversion of an alcohol to an aldehyde include *alrA* encoding a medium-chain alcohol dehydrogenase for C2-C14 (Tani et al., *Appl.Environ.Microbiol.* 66:5231-5235 (2000)), *ADH2* from *Saccharomyces cerevisiae* (Atsumi et al., *Nature* 451:86-89 (2008)), *yqhD* from *E. coli* which has preference for molecules longer than C(3) (Sulzenbacher et al., 342:489-502 (2004)), and *bdh I* and *bdh II* from *C. acetobutylicum* which converts butyrylaldehyde into butanol (Walter et al., 174:7149-7158 (1992)). The gene product of *yqhD* catalyzes the reduction of acetaldehyde, malondialdehyde, propionaldehyde, butyraldehyde, and acrolein using NADPH as the

15

20

cofactor (Perez et al., *J Biol.Chem.* 283:7346-7353 (2008a); Perez et al., *J Biol.Chem.* 283:7346-7353 (2008b)). The *adhA* gene product from *Zymomonas mobilis* has been demonstrated to have activity on a number of aldehydes including formaldehyde, acetaldehyde, propionaldehyde, butyraldehyde, and acrolein (Kinoshita et al., *Appl Microbiol* 5 *Biotechnol* 22:249-254 (1985)). Additional alcohol dehydrogenase candidates are encoded by *bdh* in *C. saccharoperbutylacetonicum* and *Cbei_1722*, *Cbei_2181* and *Cbei_2421* in *C. beijerinckii*.

Gene	GenBank Accession No.	GI No.	Organism
<i>alrA</i>	BAB12273.1	9967138	<i>Acinetobacter</i> sp. strain M-1
<i>ADH2</i>	NP_014032.1	6323961	<i>Saccharomyces cerevisiae</i>
<i>yqhD</i>	NP_417484.1	16130909	<i>Escherichia coli</i>
<i>bdh I</i>	NP_349892.1	15896543	<i>Clostridium acetobutylicum</i>
<i>bdh II</i>	NP_349891.1	15896542	<i>Clostridium acetobutylicum</i>
<i>adhA</i>	YP_162971.1	56552132	<i>Zymomonas mobilis</i>
<i>bdh</i>	BAF45463.1	124221917	<i>Clostridium saccharoperbutylacetonicum</i>
<i>Cbei_1722</i>	YP_001308850	150016596	<i>Clostridium beijerinckii</i>
<i>Cbei_2181</i>	YP_001309304	150017050	<i>Clostridium beijerinckii</i>
<i>Cbei_2421</i>	YP_001309535	150017281	<i>Clostridium beijerinckii</i>

Enzymes exhibiting 4-hydroxybutyrate dehydrogenase activity (EC 1.1.1.61) have been characterized in *Ralstonia eutropha* (Bravo et al., *J.Forensic Sci.* 49:379-387 (2004)), *Clostridium kluyveri* (Wolff et al., *Protein Expr.Purif.* 6:206-212 (1995)) and *Arabidopsis thaliana* (Breitkreuz et al., *J.Biol.Chem.* 278:41552-41556 (2003)). The *A. thaliana* enzyme was cloned and characterized in yeast (Breitkreuz et al., *J.Biol.Chem.* 278:41552-41556 (2003)). Yet another gene is the alcohol dehydrogenase *adhI* from *Geobacillus thermoglucosidasius* (Jeon et al., *J Biotechnol* 135:127-133 (2008)).

Gene	GenBank Accession No.	GI No.	Organism
<i>4hbd</i>	YP_726053.1	113867564	<i>Ralstonia eutropha</i> H16
<i>4hbd</i>	L21902.1	146348486	<i>Clostridium kluyveri</i> DSM 555
<i>4hbd</i>	Q94B07	75249805	<i>Arabidopsis thaliana</i>
<i>adhI</i>	AAR91477.1	40795502	<i>Geobacillus thermoglucosidasius</i>

3-Hydroxyisobutyrate dehydrogenase (EC 1.1.1.31) catalyzes the reversible oxidation of 3-hydroxyisobutyrate to methylmalonate semialdehyde. This enzyme participates in valine, leucine and isoleucine degradation and has been identified in bacteria, eukaryotes, and mammals. The enzyme encoded by *P84067* from *Thermus thermophilus* HB8 has been structurally characterized (Lokanath et al., 352:905-17 (2005)). Additional genes encoding this enzyme include *3hidh* in *Homo sapiens* (Hawes et al., 324:218-228 (2000)) and

Oryctolagus cuniculus (Hawes et al., *Methods Enzymol.* 324:218-228 (2000); Chowdhury et al., *Biosci.Biotechnol Biochem.* 60:2043-2047 (1996)), *mmsB* in *Pseudomonas aeruginosa* and *Pseudomonas putida*, and *dhat* in *Pseudomonas putida* (Aberhart et al., *J Chem.Soc.[Perkin 1]* 6:1404-1406 (1979); Chowdhury et al., *Biosci.Biotechnol Biochem.* 60:2043-2047 (1996); Chowdhury et al., *Biosci.Biotechnol Biochem.* 67:438-441 (2003)).

Gene	GenBank Accession No.	GI No.	Organism
<i>P84067</i>	P84067	75345323	<i>Thermus thermophilus</i>
<i>3hidh</i>	P31937.2	12643395	<i>Homo sapiens</i>
<i>3hidh</i>	P32185.1	416872	<i>Oryctolagus cuniculus</i>
<i>mmsB</i>	NP_746775.1	26991350	<i>Pseudomonas putida</i>
<i>mmsB</i>	P28811.1	127211	<i>Pseudomonas aeruginosa</i>
<i>dhat</i>	Q59477.1	2842618	<i>Pseudomonas putida</i>

EXAMPLE VI

Exemplary Hydrogenase and CO Dehydrogenase Enzymes for Extracting Reducing Equivalents from Syngas and Exemplary Reductive TCA Cycle Enzymes

Enzymes of the reductive TCA cycle useful in the non-naturally occurring microbial organisms of the present invention include one or more of ATP-citrate lyase and three CO₂-fixing enzymes: isocitrate dehydrogenase, alpha-ketoglutarate:ferredoxin oxidoreductase, pyruvate:ferredoxin oxidoreductase. The presence of ATP-citrate lyase or citrate lyase and alpha-ketoglutarate:ferredoxin oxidoreductase indicates the presence of an active reductive TCA cycle in an organism. Enzymes for each step of the reductive TCA cycle are shown below.

ATP-citrate lyase (ACL, EC 2.3.3.8), also called ATP citrate synthase, catalyzes the ATP-dependent cleavage of citrate to oxaloacetate and acetyl-CoA. ACL is an enzyme of the RTCA cycle that has been studied in green sulfur bacteria *Chlorobium limicola* and *Chlorobium tepidum*. The alpha(4)beta(4) heteromeric enzyme from *Chlorobium limicola* was cloned and characterized in *E. coli* (Kanao et al., *Eur. J. Biochem.* 269:3409-3416 (2002)). The *C. limicola* enzyme, encoded by *aclAB*, is irreversible and activity of the enzyme is regulated by the ratio of ADP/ATP. A recombinant ACL from *Chlorobium tepidum* was also expressed in *E. coli* and the holoenzyme was reconstituted *in vitro*, in a study elucidating the role of the alpha and beta subunits in the catalytic mechanism (Kim and Tabita, *J. Bacteriol.* 188:6544-6552 (2006)). ACL enzymes have also been identified in *Balnearium lithotrophicum*, *Sulfurihydrogenibium subterraneum* and other members of the

bacterial phylum *Aquificae* (Hugler et al., *Environ. Microbiol.* 9:81-92 (2007)). This activity has been reported in some fungi as well. Exemplary organisms include *Sordaria macrospora* (Nowrousian et al., *Curr. Genet.* 37:189-93 (2000)), *Aspergillus nidulans*, *Yarrowia lipolytica* (Hynes and Murray, *Eukaryotic Cell*, July: 1039-1048, (2010)) and *Aspergillus niger* (Meijer et al. *J. Ind. Microbiol. Biotechnol.* 36:1275-1280 (2009)). Other candidates can be found based on sequence homology. Information related to these enzymes is tabulated below:

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>aclA</i>	BAB21376.1	12407237	<i>Chlorobium limicola</i>
<i>aclB</i>	BAB21375.1	12407235	<i>Chlorobium limicola</i>
<i>aclA</i>	AAM72321.1	21647054	<i>Chlorobium tepidum</i>
<i>aclB</i>	AAM72322.1	21647055	<i>Chlorobium tepidum</i>
<i>aclA</i>	ABI50076.1	114054981	<i>Balnearium lithotrophicum</i>
<i>aclB</i>	ABI50075.1	114054980	<i>Balnearium lithotrophicum</i>
<i>aclA</i>	ABI50085.1	114055040	<i>Sulfurihydrogenibium subterraneum</i>
<i>aclB</i>	ABI50084.1	114055039	<i>Sulfurihydrogenibium subterraneum</i>
<i>aclA</i>	AAX76834.1	62199504	<i>Sulfurimonas denitrificans</i>
<i>aclB</i>	AAX76835.1	62199506	<i>Sulfurimonas denitrificans</i>
<i>acl1</i>	XP_504787.1	50554757	<i>Yarrowia lipolytica</i>
<i>acl2</i>	XP_503231.1	50551515	<i>Yarrowia lipolytica</i>
<i>SPBC1703.07</i>	NP_596202.1	19112994	<i>Schizosaccharomyces pombe</i>
<i>SPAC22A12.16</i>	NP_593246.1	19114158	<i>Schizosaccharomyces pombe</i>
<i>acl1</i>	CAB76165.1	7160185	<i>Sordaria macrospora</i>
<i>acl2</i>	CAB76164.1	7160184	<i>Sordaria macrospora</i>
<i>aclA</i>	CBF86850.1	259487849	<i>Aspergillus nidulans</i>
<i>aclB</i>	CBF86848	259487848	<i>Aspergillus nidulans</i>

In some organisms the conversion of citrate to oxaloacetate and acetyl-CoA proceeds through a citryl-CoA intermediate and is catalyzed by two separate enzymes, citryl-CoA synthetase (EC 6.2.1.18) and citryl-CoA lyase (EC 4.1.3.34) (Aoshima, M., *Appl. Microbiol. Biotechnol.* 75:249-255 (2007)). Citryl-CoA synthetase catalyzes the activation of citrate to citryl-CoA. The *Hydrogenobacter thermophilus* enzyme is composed of large and small subunits encoded by *ccsA* and *ccsB*, respectively (Aoshima et al., *Mol. Microbiol.* 52:751-761 (2004)). The citryl-CoA synthetase of *Aquifex aeolicus* is composed of alpha and beta subunits encoded by *sucC1* and *sucD1* (Hugler et al., *Environ. Microbiol.* 9:81-92 (2007)). Citryl-CoA lyase splits citryl-CoA into oxaloacetate and acetyl-CoA. This enzyme is a homotrimer encoded by *ccl* in *Hydrogenobacter thermophilus* (Aoshima et al., *Mol. Microbiol.* 52:763-770 (2004)) and *aq_150* in *Aquifex aeolicus* (Hugler et al., *supra* (2007)). The genes for this mechanism of

converting citrate to oxaloacetate and citryl-CoA have also been reported recently in *Chlorobium tepidum* (Eisen et al., PNAS 99(14): 9509-14 (2002)).

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>ccsA</i>	BAD17844.1	46849514	<i>Hydrogenobacter thermophilus</i>
<i>ccsB</i>	BAD17846.1	46849517	<i>Hydrogenobacter thermophilus</i>
<i>sucC1</i>	AAC07285	2983723	<i>Aquifex aeolicus</i>
<i>sucD1</i>	AAC07686	2984152	<i>Aquifex aeolicus</i>
<i>ccl</i>	BAD17841.1	46849510	<i>Hydrogenobacter thermophilus</i>
<i>aq_150</i>	AAC06486	2982866	<i>Aquifex aeolicus</i>
<i>CT0380</i>	NP_661284	21673219	<i>Chlorobium tepidum</i>
<i>CT0269</i>	NP_661173.1	21673108	<i>Chlorobium tepidum</i>
<i>CT1834</i>	AAM73055.1	21647851	<i>Chlorobium tepidum</i>

Oxaloacetate is converted into malate by malate dehydrogenase (EC 1.1.1.37), an enzyme

- 5 which functions in both the forward and reverse direction. *S. cerevisiae* possesses three copies of malate dehydrogenase, *MDH1* (McAlister-Henn and Thompson, *J. Bacteriol.* 169:5157-5166 (1987), *MDH2* (Minard and McAlister-Henn, *Mol. Cell. Biol.* 11:370-380 (1991); Gibson and McAlister-Henn, *J. Biol. Chem.* 278:25628-25636 (2003)), and *MDH3* (Steffan and McAlister-Henn, *J. Biol. Chem.* 267:24708-24715 (1992)), which localize to the
- 10 mitochondrion, cytosol, and peroxisome, respectively. *E. coli* is known to have an active malate dehydrogenase encoded by *mdh*.

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>MDH1</i>	NP_012838	6322765	<i>Saccharomyces cerevisiae</i>
<i>MDH2</i>	NP_014515	116006499	<i>Saccharomyces cerevisiae</i>
<i>MDH3</i>	NP_010205	6320125	<i>Saccharomyces cerevisiae</i>
<i>Mdh</i>	NP_417703.1	16131126	<i>Escherichia coli</i>

Fumarate hydratase (EC 4.2.1.2) catalyzes the reversible hydration of fumarate to malate.

- The three fumarases of *E. coli*, encoded by *fumA*, *fumB* and *fumC*, are regulated under
- 15 different conditions of oxygen availability. FumB is oxygen sensitive and is active under anaerobic conditions. FumA is active under microanaerobic conditions, and FumC is active under aerobic growth conditions (Tseng et al., *J. Bacteriol.* 183:461-467 (2001); Woods et al., *Biochim. Biophys. Acta* 954:14-26 (1988); Guest et al., *J. Gen. Microbiol.* 131:2971-2984 (1985)). *S. cerevisiae* contains one copy of a fumarase-encoding gene, *FUM1*, whose
- 20 product localizes to both the cytosol and mitochondrion (Sass et al., *J. Biol. Chem.* 278:45109-45116 (2003)). Additional fumarase enzymes are found in *Campylobacter jejuni*

(Smith et al., *Int. J. Biochem. Cell. Biol.* 31:961-975 (1999)), *Thermus thermophilus* (Mizobata et al., *Arch. Biochem. Biophys.* 355:49-55 (1998)) and *Rattus norvegicus* (Kobayashi et al., *J. Biochem.* 89:1923-1931 (1981)). Similar enzymes with high sequence homology include *fumI* from *Arabidopsis thaliana* and *fumC* from *Corynebacterium glutamicum*. The *MmcBC* fumarase from *Pelotomaculum thermopropionicum* is another class of fumarase with two subunits (Shimoyama et al., *FEMS Microbiol. Lett.* 270:207-213 (2007)).

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>fumA</i>	NP_416129.1	16129570	<i>Escherichia coli</i>
<i>fumB</i>	NP_418546.1	16131948	<i>Escherichia coli</i>
<i>fumC</i>	NP_416128.1	16129569	<i>Escherichia coli</i>
<i>FUMI</i>	NP_015061	6324993	<i>Saccharomyces cerevisiae</i>
<i>fumC</i>	Q8NRN8.1	39931596	<i>Corynebacterium glutamicum</i>
<i>fumC</i>	O69294.1	9789756	<i>Campylobacter jejuni</i>
<i>fumC</i>	P84127	75427690	<i>Thermus thermophilus</i>
<i>fumH</i>	P14408.1	120605	<i>Rattus norvegicus</i>
<i>MmcB</i>	YP_001211906	147677691	<i>Pelotomaculum thermopropionicum</i>
<i>MmcC</i>	YP_001211907	147677692	<i>Pelotomaculum thermopropionicum</i>

Fumarate reductase catalyzes the reduction of fumarate to succinate. The fumarate reductase of *E. coli*, composed of four subunits encoded by *frdABCD*, is membrane-bound and active under anaerobic conditions. The electron donor for this reaction is menaquinone and the two protons produced in this reaction do not contribute to the proton gradient (Iverson et al., *Science* 284:1961-1966 (1999)). The yeast genome encodes two soluble fumarate reductase isozymes encoded by *FRDS1* (Enomoto et al., *DNA Res.* 3:263-267 (1996)) and *FRDS2* (Muratsubaki et al., *Arch. Biochem. Biophys.* 352:175-181 (1998)), which localize to the cytosol and promitochondrion, respectively, and are used during anaerobic growth on glucose (Arikawa et al., *FEMS Microbiol. Lett.* 165:111-116 (1998)).

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>FRDS1</i>	P32614	418423	<i>Saccharomyces cerevisiae</i>
<i>FRDS2</i>	NP_012585	6322511	<i>Saccharomyces cerevisiae</i>
<i>frdA</i>	NP_418578.1	16131979	<i>Escherichia coli</i>
<i>frdB</i>	NP_418577.1	16131978	<i>Escherichia coli</i>
<i>frdC</i>	NP_418576.1	16131977	<i>Escherichia coli</i>
<i>frdD</i>	NP_418475.1	16131877	<i>Escherichia coli</i>

The ATP-dependent acylation of succinate to succinyl-CoA is catalyzed by succinyl-CoA synthetase (EC 6.2.1.5). The product of the *LSC1* and *LSC2* genes of *S. cerevisiae* and the *sucC* and *sucD* genes of *E. coli* naturally form a succinyl-CoA synthetase complex that catalyzes the formation of succinyl-CoA from succinate with the concomitant consumption of one ATP, a reaction which is reversible *in vivo* (Buck et al., *Biochemistry* 24:6245-6252 (1985)). These proteins are identified below:

Protein	GenBank ID	GI Number	Organism
<i>LSC1</i>	NP_014785	6324716	<i>Saccharomyces cerevisiae</i>
<i>LSC2</i>	NP_011760	6321683	<i>Saccharomyces cerevisiae</i>
<i>sucC</i>	NP_415256.1	16128703	<i>Escherichia coli</i>
<i>sucD</i>	AAC73823.1	1786949	<i>Escherichia coli</i>

Alpha-ketoglutarate:ferredoxin oxidoreductase (EC 1.2.7.3), also known as 2-oxoglutarate synthase or 2-oxoglutarate:ferredoxin oxidoreductase (OFOR), forms alpha-ketoglutarate from CO₂ and succinyl-CoA with concurrent consumption of two reduced ferredoxin equivalents. OFOR and pyruvate:ferredoxin oxidoreductase (PFOR) are members of a diverse family of 2-oxoacid:ferredoxin (flavodoxin) oxidoreductases which utilize thiamine pyrophosphate, CoA and iron-sulfur clusters as cofactors and ferredoxin, flavodoxin and FAD as electron carriers (Adams et al., *Archaea. Adv. Protein Chem.* 48:101-180 (1996)).

Enzymes in this class are reversible and function in the carboxylation direction in organisms that fix carbon by the RTCA cycle such as *Hydrogenobacter thermophilus*, *Desulfobacter hydrogenophilus* and *Chlorobium species* (Shiba et al. 1985; Evans et al., *Proc. Natl. Acad. Sci. U.S.A.* 55:92934 (1966); Buchanan, 1971). The two-subunit enzyme from *H. thermophilus*, encoded by *korAB*, has been cloned and expressed in *E. coli* (Yun et al., *Biochem. Biophys. Res. Commun.* 282:589-594 (2001)). A five subunit OFOR from the same organism with strict substrate specificity for succinyl-CoA, encoded by *forDABGE*, was recently identified and expressed in *E. coli* (Yun et al. 2002). The kinetics of CO₂ fixation of both *H. thermophilus* OFOR enzymes have been characterized (Yamamoto et al., *Extremophiles* 14:79-85 (2010)). A CO₂-fixing OFOR from *Chlorobium thiosulfatophilum* has been purified and characterized but the genes encoding this enzyme have not been identified to date. Enzyme candidates in *Chlorobium* species can be inferred by sequence similarity to the *H. thermophilus* genes. For example, the *Chlorobium limicola* genome encodes two similar proteins. Acetogenic bacteria such as *Moorella thermoacetica* are predicted to encode two OFOR enzymes. The enzyme encoded by *Moth_0034* is predicted to

function in the CO₂-assimilating direction. The genes associated with this enzyme, *Moth_0034* have not been experimentally validated to date but can be inferred by sequence similarity to known OFOR enzymes.

OFOR enzymes that function in the decarboxylation direction under physiological conditions can also catalyze the reverse reaction. The OFOR from the thermoacidophilic archaeon *Sulfolobus* sp. strain 7, encoded by ST2300, has been extensively studied (Zhang et al. 1996. A plasmid-based expression system has been developed for efficiently expressing this protein in *E. coli* (Fukuda et al., Eur. J. Biochem. 268:5639-5646 (2001)) and residues involved in substrate specificity were determined (Fukuda and Wakagi, Biochim. Biophys. Acta 1597:74-80 (2002)). The OFOR encoded by *Ape1472/Ape1473* from *Aeropyrum pernix* str. K1 was recently cloned into *E. coli*, characterized, and found to react with 2-oxoglutarate and a broad range of 2-oxoacids (Nishizawa et al., FEBS Lett. 579:2319-2322 (2005)). Another exemplary OFOR is encoded by *oorDABC* in *Helicobacter pylori* (Hughes et al. 1998). An enzyme specific to alpha-ketoglutarate has been reported in *Thauera aromatica* (Dorner and Boll, J. Bacteriol. 184 (14), 3975-83 (2002). A similar enzyme can be found in *Rhodospirillum rubrum* by sequence homology. A two subunit enzyme has also been identified in *Chlorobium tepidum* (Eisen et al., PNAS 99(14): 9509-14 (2002)).

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>korA</i>	BAB21494	12583691	<i>Hydrogenobacter thermophilus</i>
<i>korB</i>	BAB21495	12583692	<i>Hydrogenobacter thermophilus</i>
<i>forD</i>	BAB62132.1	14970994	<i>Hydrogenobacter thermophilus</i>
<i>forA</i>	BAB62133.1	14970995	<i>Hydrogenobacter thermophilus</i>
<i>forB</i>	BAB62134.1	14970996	<i>Hydrogenobacter thermophilus</i>
<i>forG</i>	BAB62135.1	14970997	<i>Hydrogenobacter thermophilus</i>
<i>forE</i>	BAB62136.1	14970998	<i>Hydrogenobacter thermophilus</i>
<i>Clim_0204</i>	ACD89303.1	189339900	<i>Chlorobium limicola</i>
<i>Clim_0205</i>	ACD89302.1	189339899	<i>Chlorobium limicola</i>
<i>Clim_1123</i>	ACD90192.1	189340789	<i>Chlorobium limicola</i>
<i>Clim_1124</i>	ACD90193.1	189340790	<i>Chlorobium limicola</i>
<i>Moth_1984</i>	YP_430825.1	83590816	<i>Moorella thermoacetica</i>
<i>Moth_1985</i>	YP_430826.1	83590817	<i>Moorella thermoacetica</i>
<i>Moth_0034</i>	YP_428917.1	83588908	<i>Moorella thermoacetica</i>
<i>ST2300</i>	NP_378302.1	15922633	<i>Sulfolobus</i> sp. strain 7
<i>Ape1472</i>	BAA80470.1	5105156	<i>Aeropyrum pernix</i>
<i>Ape1473</i>	BAA80471.2	116062794	<i>Aeropyrum pernix</i>
<i>oorD</i>	NP_207383.1	15645213	<i>Helicobacter pylori</i>

<i>oorA</i>	NP_207384.1	15645214	<i>Helicobacter pylori</i>
<i>oorB</i>	NP_207385.1	15645215	<i>Helicobacter pylori</i>
<i>oorC</i>	NP_207386.1	15645216	<i>Helicobacter pylori</i>
<i>CT0163</i>	NP_661069.1	21673004	<i>Chlorobium tepidum</i>
<i>CT0162</i>	NP_661068.1	21673003	<i>Chlorobium tepidum</i>
<i>korA</i>	CAA12243.2	19571179	<i>Thauera aromatica</i>
<i>korB</i>	CAD27440.1	19571178	<i>Thauera aromatica</i>
<i>Rru_A2721</i>	YP_427805.1	83594053	<i>Rhodospirillum rubrum</i>
<i>Rru_A2722</i>	YP_427806.1	83594054	<i>Rhodospirillum rubrum</i>

Isocitrate dehydrogenase catalyzes the reversible decarboxylation of isocitrate to 2-oxoglutarate coupled to the reduction of NAD(P)⁺. IDH enzymes in *Saccharomyces cerevisiae* and *Escherichia coli* are encoded by *IDP1* and *icd*, respectively (Haselbeck and McAlister-Henn, *J. Biol. Chem.* 266:2339-2345 (1991); Nimmo, H.G., *Biochem. J.* 234:317-2332 (1986)). The reverse reaction in the reductive TCA cycle, the reductive carboxylation of 2-oxoglutarate to isocitrate, is favored by the NADPH-dependent CO₂-fixing IDH from *Chlorobium limicola* and was functionally expressed in *E. coli* (Kanao et al., *Eur. J. Biochem.* 269:1926-1931 (2002)). A similar enzyme with 95% sequence identity is found in the *C. tepidum* genome in addition to some other candidates listed below.

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>Icd</i>	ACI84720.1	209772816	<i>Escherichia coli</i>
<i>IDP1</i>	AAA34703.1	171749	<i>Saccharomyces cerevisiae</i>
<i>Idh</i>	BAC00856.1	21396513	<i>Chlorobium limicola</i>
<i>Icd</i>	AAM71597.1	21646271	<i>Chlorobium tepidum</i>
<i>icd</i>	NP_952516.1	39996565	<i>Geobacter sulfurreducens</i>
<i>icd</i>	YP_393560.	78777245	<i>Sulfurimonas denitrificans</i>

In *H. thermophilus* the reductive carboxylation of 2-oxoglutarate to isocitrate is catalyzed by two enzymes: 2-oxoglutarate carboxylase and oxalosuccinate reductase. 2-Oxoglutarate carboxylase (EC 6.4.1.7) catalyzes the ATP-dependent carboxylation of alpha-ketoglutarate to oxalosuccinate (Aoshima and Igarashi, *Mol. Microbiol.* 62:748-759 (2006)). This enzyme is a large complex composed of two subunits. Biotinylation of the large (A) subunit is required for enzyme function (Aoshima et al., *Mol. Microbiol.* 51:791-798 (2004)). Oxalosuccinate reductase (EC 1.1.1.-) catalyzes the NAD-dependent conversion of oxalosuccinate to D-threo-isocitrate. The enzyme is a homodimer encoded by *icd* in *H.*

thermophilus. The kinetic parameters of this enzyme indicate that the enzyme only operates in the reductive carboxylation direction *in vivo*, in contrast to isocitrate dehydrogenase enzymes in other organisms (Aoshima and Igarashi, *J. Bacteriol.* 190:2050-2055 (2008)).

Based on sequence homology, gene candidates have also been found in *Thiobacillus*

5 *denitrificans* and *Thermocrinis albus*.

Protein	GenBank ID	GI Number	Organism
<i>cfiA</i>	BAF34932.1	116234991	<i>Hydrogenobacter thermophilus</i>
<i>cifB</i>	BAF34931.1	116234990	<i>Hydrogenobacter thermophilus</i>
<i>Icd</i>	BAD02487.1	38602676	<i>Hydrogenobacter thermophilus</i>
<i>Tbd_1556</i>	YP_315314	74317574	<i>Thiobacillus denitrificans</i>
<i>Tbd_1555</i>	YP_315313	74317573	<i>Thiobacillus denitrificans</i>
<i>Tbd_0854</i>	YP_314612	74316872	<i>Thiobacillus denitrificans</i>
<i>Thal_0268</i>	YP_003473030	289548042	<i>Thermocrinis albus</i>
<i>Thal_0267</i>	YP_003473029	289548041	<i>Thermocrinis albus</i>
<i>Thal_0646</i>	YP_003473406	289548418	<i>Thermocrinis albus</i>

Aconitase (EC 4.2.1.3) is an iron-sulfur-containing protein catalyzing the reversible isomerization of citrate and iso-citrate via the intermediate *cis*-aconitate. Two aconitase enzymes are encoded in the *E. coli* genome by *acnA* and *acnB*. AcnB is the main catabolic enzyme, while AcnA is more stable and appears to be active under conditions of oxidative or acid stress (Cunningham et al., *Microbiology* 143 (Pt 12):3795-3805 (1997)). Two isozymes of aconitase in *Salmonella typhimurium* are encoded by *acnA* and *acnB* (Horswill and Escalante-Semerena, *Biochemistry* 40:4703-4713 (2001)). The *S. cerevisiae* aconitase, encoded by *ACO1*, is localized to the mitochondria where it participates in the TCA cycle (Gangloff et al., *Mol. Cell. Biol.* 10:3551-3561 (1990)) and the cytosol where it participates in the glyoxylate shunt (Regev-Rudzki et al., *Mol. Biol. Cell.* 16:4163-4171 (2005)).

Protein	GenBank ID	GI Number	Organism
<i>acnA</i>	AAC7438.1	1787531	<i>Escherichia coli</i>
<i>acnB</i>	AAC73229.1	2367097	<i>Escherichia coli</i>
<i>acnA</i>	NP_460671.1	16765056	<i>Salmonella typhimurium</i>
<i>HP0779</i>	NP_207572.1	15645398	<i>Helicobacter pylori</i> 26695
<i>H16_B0568</i>	CAJ95365.1	113529018	<i>Ralstonia eutropha</i>
<i>DesfrDRAFT_3783</i>	ZP_07335307.1	303249064	<i>Desulfovibrio fructosovorans</i> JJ
<i>Suden_1040 (acnB)</i>	ABB44318.1	78497778	<i>Sulfurimonas denitrificans</i>
<i>Hydth_0755</i>	ADO45152.1	308751669	<i>Hydrogenobacter thermophilus</i>
<i>CT0543 (acn)</i>	AAM71785.1	21646475	<i>Chlorobium tepidum</i>
<i>Clim_2436</i>	YP_001944436.1	189347907	<i>Chlorobium limicola</i>
<i>Clim_0515</i>	ACD89607.1	189340204	<i>Chlorobium limicola</i>
<i>acnB</i>	NP_459163.1	16763548	<i>Salmonella typhimurium</i>

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>ACO1</i>	AAA34389.1	170982	<i>Saccharomyces cerevisiae</i>

Pyruvate:ferredoxin oxidoreductase (PFOR) catalyzes the reversible oxidation of pyruvate to form acetyl-CoA. The PFOR from *Desulfovibrio africanus* has been cloned and expressed in *E. coli* resulting in an active recombinant enzyme that was stable for several days in the presence of oxygen (Pieulle et al., *J. Bacteriol.* 179:5684-5692 (1997)). Oxygen stability is relatively uncommon in PFORs and is believed to be conferred by a 60 residue extension in the polypeptide chain of the *D. africanus* enzyme. Two cysteine residues in this enzyme form a disulfide bond that protects it against inactivation in the form of oxygen. This disulfide bond and the stability in the presence of oxygen has been found in other *Desulfovibrio* species also (Vita et al., *Biochemistry*, 47: 957-64 (2008)). The *M. thermoacetica* PFOR is also well characterized (Menon and Ragsdale, *Biochemistry* 36:8484-8494 (1997)) and was shown to have high activity in the direction of pyruvate synthesis during autotrophic growth (Furdui and Ragsdale, *J. Biol. Chem.* 275:28494-28499 (2000)). Further, *E. coli* possesses an uncharacterized open reading frame, *ydbK*, encoding a protein that is 51% identical to the *M. thermoacetica* PFOR. Evidence for pyruvate oxidoreductase activity in *E. coli* has been described (Blaschkowski et al., *Eur. J. Biochem.* 123:563-569 (1982)). PFORs have also been described in other organisms, including *Rhodobacter capsulatus* (Yakunin and Hallenbeck, *Biochimica et Biophysica Acta* 1409 (1998) 39-49 (1998)) and *Choloboum tepidum* (Eisen et al., *PNAS* 99(14): 9509-14 (2002)). The five subunit PFOR from *H. thermophilus*, encoded by *porEDABG*, was cloned into *E. coli* and shown to function in both the decarboxylating and CO₂-assimilating directions (Ikeda et al. 2006; Yamamoto et al., *Extremophiles* 14:79-85 (2010)). Homologs also exist in *C. carboxidivorans* P7. Several additional PFOR enzymes are described in the following review (Ragsdale, S.W., *Chem. Rev.* 103:2333-2346 (2003)). Finally, flavodoxin reductases (e.g., *fqrB* from *Helicobacter pylori* or *Campylobacter jejuni*) (St Maurice et al., *J. Bacteriol.* 189:4764-4773 (2007)) or Rnf-type proteins (Seedorf et al., *Proc. Natl. Acad. Sci. U.S.A.* 105:2128-2133 (2008); and Herrmann, *J. Bacteriol* 190:784-791 (2008)) provide a means to generate NADH or NADPH from the reduced ferredoxin generated by PFOR. These proteins are identified below.

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>DesfrDRAFT 0121</i>	ZP_07331646.1	303245362	<i>Desulfovibrio fructosovorans JJ</i>
Por	CAA70873.1	1770208	<i>Desulfovibrio africanus</i>
por	YP_012236.1	46581428	<i>Desulfovibrio vulgaris str.</i>

Protein	GenBank ID	GI Number	Organism
			<i>Hildenborough</i>
Dde_3237	ABB40031.1	78220682	<i>DesulfoVibrio desulfuricans</i> G20
Ddes_0298	YP_002478891.1	220903579	<i>Desulfovibrio desulfuricans</i> subsp. <i>desulfuricans</i> str. ATCC 27774
Por	YP_428946.1	83588937	<i>Moorella thermoacetica</i>
YdbK	NP_415896.1	16129339	<i>Escherichia coli</i>
nifJ (CT1628)	NP_662511.1	21674446	<i>Chlorobium tepidum</i>
CJE1649	YP_179630.1	57238499	<i>Campylobacter jejuni</i>
nifJ	ADE85473.1	294476085	<i>Rhodobacter capsulatus</i>
<i>porE</i>	BAA95603.1	7768912	<i>Hydrogenobacter thermophilus</i>
<i>porD</i>	BAA95604.1	7768913	<i>Hydrogenobacter thermophilus</i>
<i>porA</i>	BAA95605.1	7768914	<i>Hydrogenobacter thermophilus</i>
<i>porB</i>	BAA95606.1	776891	<i>Hydrogenobacter thermophilus</i>
<i>porG</i>	BAA95607.1	7768916	<i>Hydrogenobacter thermophilus</i>
FqrB	YP_001482096.1	157414840	<i>Campylobacter jejuni</i>
HP1164	NP_207955.1	15645778	<i>Helicobacter pylori</i>
RnfC	EDK33306.1	146346770	<i>Clostridium kluyveri</i>
RnfD	EDK33307.1	146346771	<i>Clostridium kluyveri</i>
RnfG	EDK33308.1	146346772	<i>Clostridium kluyveri</i>
RnfE	EDK33309.1	146346773	<i>Clostridium kluyveri</i>
RnfA	EDK33310.1	146346774	<i>Clostridium kluyveri</i>
RnfB	EDK33311.1	146346775	<i>Clostridium kluyveri</i>

The conversion of pyruvate into acetyl-CoA can be catalyzed by several other enzymes or their combinations thereof. For example, pyruvate dehydrogenase can transform pyruvate into acetyl-CoA with the concomitant reduction of a molecule of NAD into NADH. It is a multi-enzyme complex that catalyzes a series of partial reactions which results in acylating oxidative decarboxylation of pyruvate. The enzyme comprises of three subunits: the pyruvate decarboxylase (E1), dihydrolipoamide acyltransferase (E2) and dihydrolipoamide dehydrogenase (E3). This enzyme is naturally present in several organisms, including *E. coli* and *S. cerevisiae*. In the *E. coli* enzyme, specific residues in the E1 component are responsible for substrate specificity (Bisswanger, H., J. Biol. Chem. 256:815-82 (1981); Bremer, J., Eur. J. Biochem. 8:535-540 (1969); Gong et al., J. Biol. Chem. 275:13645-13653 (2000)). Enzyme engineering efforts have improved the *E. coli* PDH enzyme activity under anaerobic conditions (Kim et al., J. Bacteriol. 190:3851-3858 (2008); Kim et al., Appl. Environ. Microbiol. 73:1766-1771 (2007); Zhou et al., Biotechnol. Lett. 30:335-342 (2008)). In contrast to the *E. coli* PDH, the *B. subtilis* complex is active and required for growth under

anaerobic conditions (Nakano et al., J. Bacteriol. 179:6749-6755 (1997)). The *Klebsiella pneumoniae* PDH, characterized during growth on glycerol, is also active under anaerobic conditions (5). Crystal structures of the enzyme complex from bovine kidney (18) and the E2 catalytic domain from *Azotobacter vinelandii* are available (4). Yet another enzyme that can

5 catalyze this conversion is pyruvate formate lyase. This enzyme catalyzes the conversion of pyruvate and CoA into acetyl-CoA and formate. Pyruvate formate lyase is a common enzyme in prokaryotic organisms that is used to help modulate anaerobic redox balance. Exemplary enzymes can be found in *Escherichia coli* encoded by *pflB* (Knappe and Sawers, FEMS.Microbiol Rev. 6:383-398 (1990)), *Lactococcus lactis* (Melchiorson et al., Appl

10 Microbiol Biotechnol 58:338-344 (2002)), and *Streptococcus mutans* (Takahashi-Abbe et al., Oral.Microbiol Immunol. 18:293-297 (2003)). *E. coli* possesses an additional pyruvate formate lyase, encoded by *tdcE*, that catalyzes the conversion of pyruvate or 2-oxobutanoate to acetyl-CoA or propionyl-CoA, respectively (Hesslinger et al., Mol. Microbiol 27:477-492 (1998)). Both *pflB* and *tdcE* from *E. coli* require the presence of pyruvate formate lyase

15 activating enzyme, encoded by *pflA*. Further, a short protein encoded by *yfiD* in *E. coli* can associate with and restore activity to oxygen-cleaved pyruvate formate lyase (Vey et al., Proc.Natl. Acad. Sci. U.S.A. 105:16137-16141 (2008)). Note that *pflA* and *pflB* from *E. coli* were expressed in *S. cerevisiae* as a means to increase cytosolic acetyl-CoA for butanol production as described in WO/2008/080124]. Additional pyruvate formate lyase and

20 activating enzyme candidates, encoded by *pfl* and *act*, respectively, are found in *Clostridium pasteurianum* (Weidner et al., J Bacteriol. 178:2440-2444 (1996)).

Further, different enzymes can be used in combination to convert pyruvate into acetyl-CoA. For example, in *S. cerevisiae*, acetyl-CoA is obtained in the cytosol by first decarboxylating pyruvate to form acetaldehyde; the latter is oxidized to acetate by acetaldehyde

25 dehydrogenase and subsequently activated to form acetyl-CoA by acetyl-CoA synthetase. Acetyl-CoA synthetase is a native enzyme in several other organisms including *E. coli* (Kumari et al., J. Bacteriol. 177:2878-2886 (1995)), *Salmonella enterica* (Starai et al., Microbiology 151:3793-3801 (2005); Starai et al., J. Biol. Chem. 280:26200-26205 (2005)), and *Moorella thermoacetica* (described already). Alternatively, acetate can be activated to

30 form acetyl-CoA by acetate kinase and phosphotransacetylase. Acetate kinase first converts acetate into acetyl-phosphate with the accompanying use of an ATP molecule. Acetyl-phosphate and CoA are next converted into acetyl-CoA with the release of one phosphate by

phosphotransacetylase. Both acetate kinase and phosphotransacetylase are well-studied enzymes in several *Clostridia* and *Methanosarcina thermophila*.

Yet another way of converting pyruvate to acetyl-CoA is via pyruvate oxidase. Pyruvate oxidase converts pyruvate into acetate, using ubiquinone as the electron acceptor. In *E. coli*, this activity is encoded by *poxB*. *PoxB* has similarity to pyruvate decarboxylase of *S. cerevisiae* and *Zymomonas mobilis*. The enzyme has a thiamin pyrophosphate cofactor (Koland and Gennis, *Biochemistry* 21:4438-4442 (1982)); O'Brien et al., *Biochemistry* 16:3105-3109 (1977); O'Brien and Gennis, *J. Biol. Chem.* 255:3302-3307 (1980)) and a flavin adenine dinucleotide (FAD) cofactor. Acetate can then be converted into acetyl-CoA by either acetyl-CoA synthetase or by acetate kinase and phosphotransacetylase, as described earlier. Some of these enzymes can also catalyze the reverse reaction from acetyl-CoA to pyruvate.

For enzymes that use reducing equivalents in the form of NADH or NADPH, these reduced carriers can be generated by transferring electrons from reduced ferredoxin. Two enzymes catalyze the reversible transfer of electrons from reduced ferredoxins to NAD(P)⁺, ferredoxin:NAD⁺ oxidoreductase (EC 1.18.1.3) and ferredoxin:NADP⁺ oxidoreductase (FNR, EC 1.18.1.2). Ferredoxin:NADP⁺ oxidoreductase (FNR, EC 1.18.1.2) has a noncovalently bound FAD cofactor that facilitates the reversible transfer of electrons from NADPH to low-potential acceptors such as ferredoxins or flavodoxins (Blaschkowski et al., *Eur. J. Biochem.* 123:563-569 (1982); Fujii et al., 1977). The *Helicobacter pylori* FNR, encoded by HP1164 (*fqrB*), is coupled to the activity of pyruvate:ferredoxin oxidoreductase (PFOR) resulting in the pyruvate-dependent production of NADPH (St et al. 2007). An analogous enzyme is found in *Campylobacter jejuni* (St et al. 2007). A ferredoxin:NADP⁺ oxidoreductase enzyme is encoded in the *E. coli* genome by *fpr* (Bianchi et al. 1993). Ferredoxin:NAD⁺ oxidoreductase utilizes reduced ferredoxin to generate NADH from NAD⁺. In several organisms, including *E. coli*, this enzyme is a component of multifunctional dioxygenase enzyme complexes. The ferredoxin:NAD⁺ oxidoreductase of *E. coli*, encoded by *hcaD*, is a component of the 3-phenylproppionate dioxygenase system involved in aromatic acid utilization (Diaz et al. 1998). NADH:ferredoxin reductase activity was detected in cell extracts of *Hydrogenobacter thermophilus* strain TK-6, although a gene with this activity has not yet been indicated (Yoon et al. 2006). NADP oxidoreductase of *C. kluyveri*, encoded by *nfnAB*, catalyzes the concomitant reduction of

ferredoxin and NAD⁺ with two equivalents of NADPH (Wang et al, *J Bacteriol* 192: 5115-5123 (2010)). Finally, the energy-conserving membrane-associated Rnf-type proteins (Seedorf et al., *Proc. Natl. Acad. Sci. U.S.A.* 105:2128-2133 (2008); Herrmann et al., *J. Bacteriol.* 190:784-791 (2008)) provide a means to generate NADH or NADPH from reduced

5 ferredoxin. Additional ferredoxin:NAD(P)⁺ oxidoreductases have been annotated in *Clostridium carboxidivorans* P7 and *Clostridium ljungdahli*.

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>HP1164</i>	NP_207955.1	15645778	<i>Helicobacter pylori</i>
<i>RPA3954</i>	CAE29395.1	39650872	<i>Rhodopseudomonas palustris</i>
<i>fpr</i>	BAH29712.1	225320633	<i>Hydrogenobacter thermophilus</i>
<i>yumC</i>	NP_391091.2	255767736	<i>Bacillus subtilis</i>
<i>CJE0663</i>	AAW35824.1	57167045	<i>Campylobacter jejuni</i>
<i>fpr</i>	P28861.4	399486	<i>Escherichia coli</i>
<i>hcaD</i>	AAC75595.1	1788892	<i>Escherichia coli</i>
<i>LOC100282643</i>	NP_001149023.1	226497434	<i>Zea mays</i>
<i>NfnA</i>	YP_001393861.1	153953096	<i>Clostridium kluyveri</i>
<i>NfnB</i>	YP_001393862.1	153953097	<i>Clostridium kluyveri</i>
<i>RnfC</i>	EDK33306.1	146346770	<i>Clostridium kluyveri</i>
<i>RnfD</i>	EDK33307.1	146346771	<i>Clostridium kluyveri</i>
<i>RnfG</i>	EDK33308.1	146346772	<i>Clostridium kluyveri</i>
<i>RnfE</i>	EDK33309.1	146346773	<i>Clostridium kluyveri</i>
<i>RnfA</i>	EDK33310.1	146346774	<i>Clostridium kluyveri</i>
<i>RnfB</i>	EDK33311.1	146346775	<i>Clostridium kluyveri</i>
<i>CcarbDRAFT_2639</i>	ZP_05392639.1	255525707	<i>Clostridium carboxidivorans</i> P7
<i>CcarbDRAFT_2638</i>	ZP_05392638.1	255525706	<i>Clostridium carboxidivorans</i> P7
<i>CcarbDRAFT_2636</i>	ZP_05392636.1	255525704	<i>Clostridium carboxidivorans</i> P7
<i>CcarbDRAFT_5060</i>	ZP_05395060.1	255528241	<i>Clostridium carboxidivorans</i> P7
<i>CcarbDRAFT_2450</i>	ZP_05392450.1	255525514	<i>Clostridium carboxidivorans</i> P7
<i>CcarbDRAFT_1084</i>	ZP_05391084.1	255524124	<i>Clostridium carboxidivorans</i> P7
<i>CLJU_c11410</i> (<i>RnfB</i>)	ADK14209.1	300434442	<i>Clostridium ljungdahli</i>
<i>CLJU_c11400</i> (<i>RnfA</i>)	ADK14208.1	300434441	<i>Clostridium ljungdahli</i>
<i>CLJU_c11390</i> (<i>RnfE</i>)	ADK14207.1	300434440	<i>Clostridium ljungdahli</i>
<i>CLJU_c11380</i> (<i>RnfG</i>)	ADK14206.1	300434439	<i>Clostridium ljungdahli</i>
<i>CLJU_c11370</i> (<i>RnfD</i>)	ADK14205.1	300434438	<i>Clostridium ljungdahli</i>
<i>CLJU_c11360</i> (<i>RnfC</i>)	ADK14204.1	300434437	<i>Clostridium ljungdahli</i>

Ferredoxins are small acidic proteins containing one or more iron-sulfur clusters that function as intracellular electron carriers with a low reduction potential. Reduced ferredoxins donate electrons to Fe-dependent enzymes such as ferredoxin-NADP⁺ oxidoreductase, pyruvate:ferredoxin oxidoreductase (PFOR) and 2-oxoglutarate:ferredoxin oxidoreductase (OFOR). The *H. thermophilus* gene *fdx1* encodes a [4Fe-4S]-type ferredoxin that is required for the reversible carboxylation of 2-oxoglutarate and pyruvate by OFOR and PFOR, respectively (Yamamoto et al., *Extremophiles* 14:79-85 (2010)). The ferredoxin associated with the *Sulfolobus solfataricus* 2-oxoacid:ferredoxin reductase is a monomeric dicluster [3Fe-4S][4Fe-4S] type ferredoxin (Park et al. 2006). While the gene associated with this protein has not been fully sequenced, the N-terminal domain shares 93% homology with the *zfx* ferredoxin from *S. acidocaldarius*. The *E. coli* genome encodes a soluble ferredoxin of unknown physiological function, *fdx*. Some evidence indicates that this protein can function in iron-sulfur cluster assembly (Takahashi and Nakamura, 1999). Additional ferredoxin proteins have been characterized in *Helicobacter pylori* (Mukhopadhyay et al. 2003) and *Campylobacter jejuni* (van Vliet et al. 2001). A 2Fe-2S ferredoxin from *Clostridium pasteurianum* has been cloned and expressed in *E. coli* (Fujinaga and Meyer, Biochemical and Biophysical Research Communications, 192(3): (1993)). Acetogenic bacteria such as *Moorella thermoacetica*, *Clostridium carboxidivorans* P7, *Clostridium ljungdahlii* and *Rhodospirillum rubrum* are predicted to encode several ferredoxins, listed in the table below.

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>fdx1</i>	BAE02673.1	68163284	<i>Hydrogenobacter thermophilus</i>
M11214.1	AAA83524.1	144806	<i>Clostridium pasteurianum</i>
<i>Zfx</i>	AAY79867.1	68566938	<i>Sulfolobus acidocalarius</i>
<i>Fdx</i>	AAC75578.1	1788874	<i>Escherichia coli</i>
<i>hp_0277</i>	AAD07340.1	2313367	<i>Helicobacter pylori</i>
<i>fdxA</i>	CAL34484.1	112359698	<i>Campylobacter jejuni</i>
<i>Moth_0061</i>	ABC18400.1	83571848	<i>Moorella thermoacetica</i>
<i>Moth_1200</i>	ABC19514.1	83572962	<i>Moorella thermoacetica</i>
<i>Moth_1888</i>	ABC20188.1	83573636	<i>Moorella thermoacetica</i>
<i>Moth_2112</i>	ABC20404.1	83573852	<i>Moorella thermoacetica</i>
<i>Moth_1037</i>	ABC19351.1	83572799	<i>Moorella thermoacetica</i>
CcarbDRAFT_4383	ZP_05394383.1	255527515	<i>Clostridium carboxidivorans</i> P7
CcarbDRAFT_2958	ZP_05392958.1	255526034	<i>Clostridium carboxidivorans</i> P7
CcarbDRAFT_2281	ZP_05392281.1	255525342	<i>Clostridium carboxidivorans</i> P7
CcarbDRAFT_5296	ZP_05395295.1	255528511	<i>Clostridium carboxidivorans</i> P7
CcarbDRAFT_1615	ZP_05391615.1	255524662	<i>Clostridium carboxidivorans</i> P7
CcarbDRAFT_1304	ZP_05391304.1	255524347	<i>Clostridium carboxidivorans</i> P7

cooF	AAG29808.1	11095245	<i>Carboxydotherrnus hydrogenoformans</i>
fdxN	CAA35699.1	46143	<i>Rhodobacter capsulatus</i>
Rru_A2264	ABC23064.1	83576513	<i>Rhodospirillum rubrum</i>
Rru_A1916	ABC22716.1	83576165	<i>Rhodospirillum rubrum</i>
Rru_A2026	ABC22826.1	83576275	<i>Rhodospirillum rubrum</i>
cooF	AAC45122.1	1498747	<i>Rhodospirillum rubrum</i>
fdxN	AAA26460.1	152605	<i>Rhodospirillum rubrum</i>
Alvin_2884	ADC63789.1	288897953	<i>Allochromatium vinosum</i> DSM 180
fdx	YP_002801146.1	226946073	<i>Azotobacter vinelandii</i> DJ
CKL_3790	YP_001397146.1	153956381	<i>Clostridium kluyveri</i> DSM 555
fer1	NP_949965.1	39937689	<i>Rhodopseudomonas palustris</i> CGA009
fdx	CAA12251.1	3724172	<i>Thauera aromatica</i>
CHY_2405	YP_361202.1	78044690	<i>Carboxydotherrnus hydrogenoformans</i>
fer	YP_359966.1	78045103	<i>Carboxydotherrnus hydrogenoformans</i>
fer	AAC83945.1	1146198	<i>Bacillus subtilis</i>
fdx1	NP_249053.1	15595559	<i>Pseudomonas aeruginosa</i> PA01
yfhL	AP_003148.1	89109368	<i>Escherichia coli</i> K-12
CLJU_c00930	ADK13195.1	300433428	<i>Clostridium ljungdahli</i>
CLJU_c00010	ADK13115.1	300433348	<i>Clostridium ljungdahli</i>
CLJU_c01820	ADK13272.1	300433505	<i>Clostridium ljungdahli</i>
CLJU_c17980	ADK14861.1	300435094	<i>Clostridium ljungdahli</i>
CLJU_c17970	ADK14860.1	300435093	<i>Clostridium ljungdahli</i>
CLJU_c22510	ADK15311.1	300435544	<i>Clostridium ljungdahli</i>
CLJU_c26680	ADK15726.1	300435959	<i>Clostridium ljungdahli</i>
CLJU_c29400	ADK15988.1	300436221	<i>Clostridium ljungdahli</i>

Succinyl-CoA transferase catalyzes the conversion of succinyl-CoA to succinate while transferring the CoA moiety to a CoA acceptor molecule. Many transferases have broad specificity and can utilize CoA acceptors as diverse as acetate, succinate, propionate, butyrate, 2-methylacetoacetate, 3-ketohexanoate, 3-ketopentanoate, valerate, crotonate, 3-mercaptopropionate, propionate, vinylacetate, and butyrate, among others.

The conversion of succinate to succinyl-CoA can be carried by a transferase which does not require the direct consumption of an ATP or GTP. This type of reaction is common in a number of organisms. The conversion of succinate to succinyl-CoA can also be catalyzed by succinyl-CoA:Acetyl-CoA transferase. The gene product of *cat1* of *Clostridium kluyveri* has been shown to exhibit succinyl-CoA: acetyl-CoA transferase activity (Sohling and Gottschalk, J. Bacteriol. 178:871-880 (1996)). In addition, the activity is present in *Trichomonas vaginalis* (van Grinsven et al. 2008) and *Trypanosoma brucei* (Riviere et al.

2004). The succinyl-CoA:acetate CoA-transferase from *Acetobacter aceti*, encoded by *aarC*, replaces succinyl-CoA synthetase in a variant TCA cycle (Mullins et al. 2008). Similar succinyl-CoA transferase activities are also present in *Trichomonas vaginalis* (van Grinsven et al. 2008), *Trypanosoma brucei* (Riviere et al. 2004) and *Clostridium kluyveri* (Sohling and Gottschalk, 1996c). The beta-ketoadipate:succinyl-CoA transferase encoded by *pcaI* and *pcaJ* in *Pseudomonas putida* is yet another candidate (Kaschabek et al. 2002). The aforementioned proteins are identified below.

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>catI</i>	P38946.1	729048	<i>Clostridium kluyveri</i>
<i>TVAG_395550</i>	XP_001330176	123975034	<i>Trichomonas vaginalis</i> G3
<i>Tb11.02.0290</i>	XP_828352	71754875	<i>Trypanosoma brucei</i>
<i>pcaI</i>	AAN69545.1	24985644	<i>Pseudomonas putida</i>
<i>pcaJ</i>	NP_746082.1	26990657	<i>Pseudomonas putida</i>
<i>aarC</i>	ACD85596.1	189233555	<i>Acetobacter aceti</i>

An additional exemplary transferase that converts succinate to succinyl-CoA while converting a 3-ketoacyl-CoA to a 3-ketoacid is succinyl-CoA:3-ketoacid-CoA transferase (EC 2.8.3.5). Exemplary succinyl-CoA:3-ketoacid-CoA transferases are present in *Helicobacter pylori* (Corthesy-Theulaz et al. 1997), *Bacillus subtilis*, and *Homo sapiens* (Fukao et al. 2000; Tanaka et al. 2002). The aforementioned proteins are identified below.

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>HPAG1_0676</i>	YP_627417	108563101	<i>Helicobacter pylori</i>
<i>HPAG1_0677</i>	YP_627418	108563102	<i>Helicobacter pylori</i>
<i>ScoA</i>	NP_391778	16080950	<i>Bacillus subtilis</i>
<i>ScoB</i>	NP_391777	16080949	<i>Bacillus subtilis</i>
<i>OXCT1</i>	NP_000427	4557817	<i>Homo sapiens</i>
<i>OXCT2</i>	NP_071403	11545841	<i>Homo sapiens</i>

Converting succinate to succinyl-CoA by succinyl-CoA:3-ketoacid-CoA transferase requires the simultaneous conversion of a 3-ketoacyl-CoA such as acetoacetyl-CoA to a 3-ketoacid such as acetoacetate. Conversion of a 3-ketoacid back to a 3-ketoacyl-CoA can be catalyzed by an acetoacetyl-CoA:acetate:CoA transferase. Acetoacetyl-CoA:acetate:CoA transferase converts acetoacetyl-CoA and acetate to acetoacetate and acetyl-CoA, or vice versa. Exemplary enzymes include the gene products of *atoAD* from *E. coli* (Hanai et al., Appl Environ Microbiol 73:7814-7818 (2007)), *ctfAB* from *C. acetobutylicum* (Jojima et al., Appl Microbiol Biotechnol 77:1219-1224 (2008)), and *ctfAB* from *Clostridium*

saccharoperbutylaceticum (Kosaka et al., *Biosci. Biotechnol. Biochem.* 71:58-68 (2007)) are shown below.

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>AtoA</i>	NP_416726.1	2492994	<i>Escherichia coli</i>
<i>AtoD</i>	NP_416725.1	2492990	<i>Escherichia coli</i>
<i>CtfA</i>	NP_149326.1	15004866	<i>Clostridium acetobutylicum</i>
<i>CtfB</i>	NP_149327.1	15004867	<i>Clostridium acetobutylicum</i>
<i>CtfA</i>	AAP42564.1	31075384	<i>Clostridium saccharoperbutylaceticum</i>
<i>CtfB</i>	AAP42565.1	31075385	<i>Clostridium saccharoperbutylaceticum</i>

Yet another possible CoA acceptor is benzylsuccinate. Succinyl-CoA:(R)-Benzylsuccinate

- 5 CoA-Transferase functions as part of an anaerobic degradation pathway for toluene in organisms such as *Thauera aromatica* (Leutwein and Heider, *J. Bact.* 183(14) 4288-4295 (2001)). Homologs can be found in *Azoarcus* sp. T, *Aromatoleum aromaticum* EbN1, and *Geobacter metallireducens* GS-15. The aforementioned proteins are identified below.

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>bbsE</i>	AAF89840	9622535	<i>Thauera aromatic</i>
<i>Bbsf</i>	AAF89841	9622536	<i>Thauera aromatic</i>
<i>bbsE</i>	AAU45405.1	52421824	<i>Azoarcus</i> sp. T
<i>bbsF</i>	AAU45406.1	52421825	<i>Azoarcus</i> sp. T
<i>bbsE</i>	YP_158075.1	56476486	<i>Aromatoleum aromaticum</i> EbN1
<i>bbsF</i>	YP_158074.1	56476485	<i>Aromatoleum aromaticum</i> EbN1
<i>Gmet_1521</i>	YP_384480.1	78222733	<i>Geobacter metallireducens</i> GS-15
<i>Gmet_1522</i>	YP_384481.1	78222734	<i>Geobacter metallireducens</i> GS-15

- 10 Additionally, *ygfH* encodes a propionyl CoA:succinate CoA transferase in *E. coli* (Haller et al., *Biochemistry*, 39(16) 4622-4629). Close homologs can be found in, for example, *Citrobacter youngae* ATCC 29220, *Salmonella enterica* subsp. *arizonae* serovar, and *Yersinia intermedia* ATCC 29909. The aforementioned proteins are identified below.

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>ygfH</i>	NP_417395.1	16130821	<i>Escherichia coli</i> str. K-12 substr. MG1655
<i>CIT292_04485</i>	ZP_03838384.1	227334728	<i>Citrobacter youngae</i> ATCC 29220
<i>SARI_04582</i>	YP_001573497.1	161506385	<i>Salmonella enterica</i> subsp.

			<i>arizonae</i> serovar
yinte0001 14430	ZP_04635364.1	238791727	<i>Yersinia intermedia</i> ATCC 29909

Citrate lyase (EC 4.1.3.6) catalyzes a series of reactions resulting in the cleavage of citrate to acetate and oxaloacetate. The enzyme is active under anaerobic conditions and is composed of three subunits: an acyl-carrier protein (ACP, gamma), an ACP transferase (alpha), and an acyl lyase (beta). Enzyme activation uses covalent binding and acetylation of an unusual prosthetic group, 2'-(5''-phosphoribosyl)-3'-dephospho-CoA, which is similar in structure to acetyl-CoA. Acylation is catalyzed by CitC, a citrate lyase synthetase. Two additional proteins, CitG and CitX, are used to convert the apo enzyme into the active holo enzyme (Schneider et al., *Biochemistry* 39:9438-9450 (2000)). Wild type *E. coli* does not have citrate lyase activity; however, mutants deficient in molybdenum cofactor synthesis have an active citrate lyase (Clark, *FEMS Microbiol. Lett.* 55:245-249 (1990)). The *E. coli* enzyme is encoded by *citEFD* and the citrate lyase synthetase is encoded by *citC* (Nilekani and SivaRaman, *Biochemistry* 22:4657-4663 (1983)). The *Leuconostoc mesenteroides* citrate lyase has been cloned, characterized and expressed in *E. coli* (Bekal et al., *J. Bacteriol.* 180:647-654 (1998)). Citrate lyase enzymes have also been identified in enterobacteria that utilize citrate as a carbon and energy source, including *Salmonella typhimurium* and *Klebsiella pneumoniae* (Bott, *Arch. Microbiol.* 167: 78-88 (1997); Bott and Dimroth, *Mol. Microbiol.* 14:347-356 (1994)). The aforementioned proteins are tabulated below.

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>citF</i>	AAC73716.1	1786832	<i>Escherichia coli</i>
<i>Cite</i>	AAC73717.2	87081764	<i>Escherichia coli</i>
<i>citD</i>	AAC73718.1	1786834	<i>Escherichia coli</i>
<i>citC</i>	AAC73719.2	87081765	<i>Escherichia coli</i>
<i>citG</i>	AAC73714.1	1786830	<i>Escherichia coli</i>
<i>citX</i>	AAC73715.1	1786831	<i>Escherichia coli</i>
<i>citF</i>	CAA71633.1	2842397	<i>Leuconostoc mesenteroides</i>
<i>Cite</i>	CAA71632.1	2842396	<i>Leuconostoc mesenteroides</i>
<i>citD</i>	CAA71635.1	2842395	<i>Leuconostoc mesenteroides</i>
<i>citC</i>	CAA71636.1	3413797	<i>Leuconostoc mesenteroides</i>
<i>citG</i>	CAA71634.1	2842398	<i>Leuconostoc mesenteroides</i>
<i>citX</i>	CAA71634.1	2842398	<i>Leuconostoc mesenteroides</i>
<i>citF</i>	NP_459613.1	16763998	<i>Salmonella typhimurium</i>
<i>cite</i>	AAL19573.1	16419133	<i>Salmonella typhimurium</i>
<i>citD</i>	NP_459064.1	16763449	<i>Salmonella typhimurium</i>
<i>citC</i>	NP_459616.1	16764001	<i>Salmonella typhimurium</i>
<i>citG</i>	NP_459611.1	16763996	<i>Salmonella typhimurium</i>
<i>citX</i>	NP_459612.1	16763997	<i>Salmonella typhimurium</i>
<i>citF</i>	CAA56217.1	565619	<i>Klebsiella pneumoniae</i>

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>cite</i>	CAA56216.1	565618	<i>Klebsiella pneumoniae</i>
<i>citD</i>	CAA56215.1	565617	<i>Klebsiella pneumoniae</i>
<i>citC</i>	BAH66541.1	238774045	<i>Klebsiella pneumoniae</i>
<i>citG</i>	CAA56218.1	565620	<i>Klebsiella pneumoniae</i>
<i>citX</i>	AAL60463.1	18140907	<i>Klebsiella pneumoniae</i>

Acetate kinase (EC 2.7.2.1) catalyzes the reversible ATP-dependent phosphorylation of acetate to acetylphosphate. Exemplary acetate kinase enzymes have been characterized in many organisms including *E. coli*, *Clostridium acetobutylicum* and *Methanosarcina*

- 5 *thermophila* (Ingram-Smith et al., *J. Bacteriol.* 187:2386-2394 (2005); Fox and Roseman, *J. Biol. Chem.* 261:13487-13497 (1986); Winzer et al., *Microbiology* 143 (Pt 10):3279-3286 (1997)). Acetate kinase activity has also been demonstrated in the gene product of *E. coli purT* (Marolewski et al., *Biochemistry* 33:2531-2537 (1994). Some butyrate kinase enzymes (EC 2.7.2.7), for example *buk1* and *buk2* from *Clostridium acetobutylicum*, also accept
- 10 acetate as a substrate (Hartmanis, M.G., *J. Biol. Chem.* 262:617-621 (1987)).

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>ackA</i>	NP_416799.1	16130231	<i>Escherichia coli</i>
<i>Ack</i>	AAB18301.1	1491790	<i>Clostridium acetobutylicum</i>
<i>Ack</i>	AAA72042.1	349834	<i>Methanosarcina thermophila</i>
<i>purT</i>	AAC74919.1	1788155	<i>Escherichia coli</i>
<i>buk1</i>	NP_349675	15896326	<i>Clostridium acetobutylicum</i>
<i>buk2</i>	Q97III	20137415	<i>Clostridium acetobutylicum</i>

The formation of acetyl-CoA from acetylphosphate is catalyzed by phosphotransacetylase (EC 2.3.1.8). The *pta* gene from *E. coli* encodes an enzyme that reversibly converts acetyl-CoA into acetyl-phosphate (Suzuki, T., *Biochim. Biophys. Acta* 191:559-569 (1969)).

- 15 Additional acetyltransferase enzymes have been characterized in *Bacillus subtilis* (Rado and Hoch, *Biochim. Biophys. Acta* 321:114-125 (1973), *Clostridium kluyveri* (Stadtman, E., *Methods Enzymol.* 1:5896-599 (1955), and *Thermotoga maritima* (Bock et al., *J. Bacteriol.* 181:1861-1867 (1999)). This reaction is also catalyzed by some phosphotranbutyrylase enzymes (EC 2.3.1.19) including the *ptb* gene products from *Clostridium acetobutylicum*
- 20 (Wiesenborn et al., *App. Environ. Microbiol.* 55:317-322 (1989); Walter et al., *Gene* 134:107-111 (1993)). Additional *ptb* genes are found in butyrate-producing bacterium L2-50 (Louis et al., *J. Bacteriol.* 186:2099-2106 (2004) and *Bacillus megaterium* (Vazquez et al., *Curr. Microbiol.* 42:345-349 (2001)).

Protein	GenBank ID	GI Number	Organism
<i>Pta</i>	NP_416800.1	71152910	<i>Escherichia coli</i>
<i>Pta</i>	P39646	730415	<i>Bacillus subtilis</i>
<i>Pta</i>	A5N801	146346896	<i>Clostridium kluyveri</i>
<i>Pta</i>	Q9X0L4	6685776	<i>Thermotoga maritima</i>
<i>Ptb</i>	NP_349676	34540484	<i>Clostridium acetobutylicum</i>
<i>Ptb</i>	AAR19757.1	38425288	butyrate-producing bacterium L2-50
<i>Ptb</i>	CAC07932.1	10046659	<i>Bacillus megaterium</i>

The acylation of acetate to acetyl-CoA is catalyzed by enzymes with acetyl-CoA synthetase activity. Two enzymes that catalyze this reaction are AMP-forming acetyl-CoA synthetase (EC 6.2.1.1) and ADP-forming acetyl-CoA synthetase (EC 6.2.1.13). AMP-forming acetyl-CoA synthetase (ACS) is the predominant enzyme for activation of acetate to acetyl-CoA. Exemplary ACS enzymes are found in *E. coli* (Brown et al., *J. Gen. Microbiol.* 102:327-336 (1977)), *Ralstonia eutropha* (Priefert and Steinbuchel, *J. Bacteriol.* 174:6590-6599 (1992)), *Methanothermobacter thermautotrophicus* (Ingram-Smith and Smith, *Archaea* 2:95-107 (2007)), *Salmonella enterica* (Gulick et al., *Biochemistry* 42:2866-2873 (2003)) and *Saccharomyces cerevisiae* (Jogl and Tong, *Biochemistry* 43:1425-1431 (2004)). ADP-forming acetyl-CoA synthetases are reversible enzymes with a generally broad substrate range (Musfeldt and Schonheit, *J. Bacteriol.* 184:636-644 (2002)). Two isozymes of ADP-forming acetyl-CoA synthetases are encoded in the *Archaeoglobus fulgidus* genome by are encoded by AF1211 and AF1983 (Musfeldt and Schonheit, *supra* (2002)). The enzyme from *Haloarcula marismortui* (annotated as a succinyl-CoA synthetase) also accepts acetate as a substrate and reversibility of the enzyme was demonstrated (Brasen and Schonheit, *Arch. Microbiol.* 182:277-287 (2004)). The ACD encoded by PAE3250 from hyperthermophilic crenarchaeon *Pyrobaculum aerophilum* showed the broadest substrate range of all characterized ACDs, reacting with acetate, isobutyryl-CoA (preferred substrate) and phenylacetyl-CoA (Brasen and Schonheit, *supra* (2004)). Directed evolution or engineering can be used to modify this enzyme to operate at the physiological temperature of the host organism. The enzymes from *A. fulgidus*, *H. marismortui* and *P. aerophilum* have all been cloned, functionally expressed, and characterized in *E. coli* (Brasen and Schonheit, *supra* (2004); Musfeldt and Schonheit, *supra* (2002)). Additional candidates include the succinyl-CoA synthetase encoded by *sucCD* in *E. coli* (Buck et al., *Biochemistry* 24:6245-6252 (1985)) and the acyl-CoA ligase from *Pseudomonas putida* (Fernandez-Valverde et al., *Appl. Environ. Microbiol.* 59:1149-1154 (1993)). The aforementioned proteins are tabulated below.

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>acs</i>	AAC77039.1	1790505	<i>Escherichia coli</i>
<i>acoE</i>	AAA21945.1	141890	<i>Ralstonia eutropha</i>
<i>acsI</i>	ABC87079.1	86169671	<i>Methanothermobacter thermautotrophicus</i>
<i>acsI</i>	AAL23099.1	16422835	<i>Salmonella enterica</i>
<i>ACSI</i>	Q01574.2	257050994	<i>Saccharomyces cerevisiae</i>
<i>AF1211</i>	NP_070039.1	11498810	<i>Archaeoglobus fulgidus</i>
<i>AF1983</i>	NP_070807.1	11499565	<i>Archaeoglobus fulgidus</i>
<i>scs</i>	YP_135572.1	55377722	<i>Haloarcula marismortui</i>
<i>PAE3250</i>	NP_560604.1	18313937	<i>Pyrobaculum aerophilum str. IM2</i>
<i>sucC</i>	NP_415256.1	16128703	<i>Escherichia coli</i>
<i>sucD</i>	AAC73823.1	1786949	<i>Escherichia coli</i>
<i>paaF</i>	AAC24333.2	22711873	<i>Pseudomonas putida</i>

The product yields per C-mol of substrate of microbial cells synthesizing reduced fermentation products such as ethylene glycol, are limited by insufficient reducing equivalents in the carbohydrate feedstock. Reducing equivalents, or electrons, can be extracted from synthesis gas components such as CO and H₂ using carbon monoxide dehydrogenase (CODH) and hydrogenase enzymes, respectively. The reducing equivalents are then passed to acceptors such as oxidized ferredoxins, oxidized quinones, oxidized cytochromes, NAD(P)⁺, water, or hydrogen peroxide to form reduced ferredoxin, reduced quinones, reduced cytochromes, NAD(P)H, H₂, or water, respectively. Reduced ferredoxin and NAD(P)H are particularly useful as they can serve as redox carriers for various Wood-Ljungdahl pathway and reductive TCA cycle enzymes.

When both feedstocks of sugar and syngas are available, the syngas components CO and H₂ can be utilized to generate reducing equivalents by employing the hydrogenase and CO dehydrogenase. The reducing equivalents generated from syngas components will be utilized to power the glucose to ethylene glycol production pathways. Theoretically, all carbons in glucose will be conserved, thus resulting in a maximal theoretical yield to produce ethylene glycol from glucose.

As described above, a combined feedstock strategy where syngas is combined with a sugar-based feedstock or other carbon substrate can greatly improve the theoretical yields. In this co-feeding approach, syngas components H₂ and CO can be utilized by the hydrogenase and CO dehydrogenase to generate reducing equivalents, that can be used to power chemical

production pathways in which the carbons from sugar or other carbon substrates will be maximally conserved and the theoretical yields improved for production of ethylene glycol.

Herein below the enzymes and the corresponding genes used for extracting redox from synags components are described. CODH is a reversible enzyme that interconverts CO and CO₂ at the expense or gain of electrons. The natural physiological role of the CODH in ACS/CODH complexes is to convert CO₂ to CO for incorporation into acetyl-CoA by acetyl-CoA synthase. Nevertheless, such CODH enzymes are suitable for the extraction of reducing equivalents from CO due to the reversible nature of such enzymes. Expressing such CODH enzymes in the absence of ACS allows them to operate in the direction opposite to their natural physiological role (i.e., CO oxidation).

In *M. thermoacetica*, *C. hydrogenoformans*, *C. carboxidivorans* P7, and several other organisms, additional CODH encoding genes are located outside of the ACS/CODH operons. These enzymes provide a means for extracting electrons (or reducing equivalents) from the conversion of carbon monoxide to carbon dioxide. The *M. thermoacetica* gene (Genbank Accession Number: YP_430813) is expressed by itself in an operon and is believed to transfer electrons from CO to an external mediator like ferredoxin in a “Ping-pong” reaction. The reduced mediator then couples to other reduced nicotinamide adenine dinucleotide phosphate (NAD(P)H) carriers or ferredoxin-dependent cellular processes (Ragsdale, *Annals of the New York Academy of Sciences* 1125: 129–136 (2008)). The genes encoding the *C. hydrogenoformans* CODH-II and CooF, a neighboring protein, were cloned and sequenced (Gonzalez and Robb, *FEMS Microbiol Lett.* 191:243-247 (2000)). The resulting complex was membrane-bound, although cytoplasmic fractions of CODH-II were shown to catalyze the formation of NADPH suggesting an anabolic role (Svetlitchnyi et al., *J Bacteriol.* 183:5134-5144 (2001)). The crystal structure of the CODH-II is also available (Dobbek et al., *Science* 293:1281-1285 (2001)). Similar ACS-free CODH enzymes can be found in a diverse array of organisms including *Geobacter metallireducens* GS-15, *Chlorobium phaeobacteroides* DSM 266, *Clostridium cellulolyticum* H10, *Desulfovibrio desulfuricans* subsp. *desulfuricans* str. ATCC 27774, *Pelobacter carbinolicus* DSM 2380, *C. ljungdahli* and *Campylobacter curvus* 525.92.

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
CODH (putative)	YP_430813	83590804	<i>Moorella thermoacetica</i>
CODH-II (CooS-II)	YP_358957	78044574	<i>Carboxydotherrmus hydrogenoformans</i>

Protein	GenBank ID	GI Number	Organism
CooF	YP_358958	78045112	<i>Carboxydotherrmus hydrogenoformans</i>
<u>CODH (putative)</u>	ZP_05390164.1	255523193	<i>Clostridium carboxidivorans P7</i>
CcarbDRAFT_0341	ZP_05390341.1	255523371	<i>Clostridium carboxidivorans P7</i>
CcarbDRAFT_1756	ZP_05391756.1	255524806	<i>Clostridium carboxidivorans P7</i>
CcarbDRAFT_2944	ZP_05392944.1	255526020	<i>Clostridium carboxidivorans P7</i>
CODH	YP_384856.1	78223109	<i>Geobacter metallireducens GS-15</i>
<i>Cpha266_0148</i> (cytochrome c)	YP_910642.1	119355998	<i>Chlorobium phaeobacteroides DSM 266</i>
<i>Cpha266_0149</i> (CODH)	YP_910643.1	119355999	<i>Chlorobium phaeobacteroides DSM 266</i>
<i>Ccel_0438</i>	YP_002504800.1	220927891	<i>Clostridium cellulolyticum H10</i>
<i>Ddes_0382</i> (CODH)	YP_002478973.1	220903661	<i>Desulfovibrio desulfuricans</i> subsp. <i>desulfuricans</i> str. ATCC 27774
<i>Ddes_0381</i> (CooC)	YP_002478972.1	220903660	<i>Desulfovibrio desulfuricans</i> subsp. <i>desulfuricans</i> str. ATCC 27774
<i>Pcar_0057</i> (CODH)	YP_355490.1	7791767	<i>Pelobacter carbinolicus DSM 2380</i>
<i>Pcar_0058</i> (CooC)	YP_355491.1	7791766	<i>Pelobacter carbinolicus DSM 2380</i>
<i>Pcar_0058</i> (HypA)	YP_355492.1	7791765	<i>Pelobacter carbinolicus DSM 2380</i>
<i>CooS</i> (CODH)	YP_001407343.1	154175407	<i>Campylobacter curvus 525.92</i>
<i>CLJU_c09110</i>	ADK13979.1	300434212	<i>Clostridium ljungdahli</i>
<i>CLJU_c09100</i>	ADK13978.1	300434211	<i>Clostridium ljungdahli</i>
<i>CLJU_c09090</i>	ADK13977.1	300434210	<i>Clostridium ljungdahli</i>

In some cases, hydrogenase encoding genes are located adjacent to a CODH. In *Rhodospirillum rubrum*, the encoded CODH/hydrogenase proteins form a membrane-bound enzyme complex that has been indicated to be a site where energy, in the form of a proton gradient, is generated from the conversion of CO and H₂O to CO₂ and H₂ (Fox et al., *J Bacteriol.* 178:6200-6208 (1996)). The CODH-I of *C. hydrogenoformans* and its adjacent genes have been proposed to catalyze a similar functional role based on their similarity to the *R. rubrum* CODH/hydrogenase gene cluster (Wu et al., *PLoS Genet.* 1:e65 (2005)). The *C. hydrogenoformans* CODH-I was also shown to exhibit intense CO oxidation and CO₂ reduction activities when linked to an electrode (Parkin et al., *J Am.Chem.Soc.* 129:10328-10329 (2007)). The protein sequences of exemplary CODH and hydrogenase genes can be identified by the following GenBank accession numbers.

Protein	GenBank ID	GI Number	Organism
CODH-I (CooS-I)	YP_360644	78043418	<i>Carboxydotherrmus hydrogenoformans</i>
CooF	YP_360645	78044791	<i>Carboxydotherrmus hydrogenoformans</i>
HypA	YP_360646	78044340	<i>Carboxydotherrmus hydrogenoformans</i>
CooH	YP_360647	78043871	<i>Carboxydotherrmus hydrogenoformans</i>
CooU	YP_360648	78044023	<i>Carboxydotherrmus hydrogenoformans</i>
CooX	YP_360649	78043124	<i>Carboxydotherrmus hydrogenoformans</i>
CooL	YP_360650	78043938	<i>Carboxydotherrmus hydrogenoformans</i>
CooK	YP_360651	78044700	<i>Carboxydotherrmus hydrogenoformans</i>
CooM	YP_360652	78043942	<i>Carboxydotherrmus hydrogenoformans</i>
CooC	YP_360654.1	78043296	<i>Carboxydotherrmus hydrogenoformans</i>
CooA-1	YP_360655.1	78044021	<i>Carboxydotherrmus hydrogenoformans</i>
CooL	AAC45118	1515468	<i>Rhodospirillum rubrum</i>
CooX	AAC45119	1515469	<i>Rhodospirillum rubrum</i>
CooU	AAC45120	1515470	<i>Rhodospirillum rubrum</i>
CooH	AAC45121	1498746	<i>Rhodospirillum rubrum</i>
CooF	AAC45122	1498747	<i>Rhodospirillum rubrum</i>
CODH (CooS)	AAC45123	1498748	<i>Rhodospirillum rubrum</i>
CooC	AAC45124	1498749	<i>Rhodospirillum rubrum</i>
CooT	AAC45125	1498750	<i>Rhodospirillum rubrum</i>
CooJ	AAC45126	1498751	<i>Rhodospirillum rubrum</i>

Native to *E. coli* and other enteric bacteria are multiple genes encoding up to four hydrogenases (Sawers, G., *Antonie Van Leeuwenhoek* 66:57-88 (1994); Sawers et al., *J Bacteriol.* 164:1324-1331 (1985); Sawers and Boxer, *Eur.J Biochem.* 156:265-275 (1986);

5 Sawers et al., *J Bacteriol.* 168:398-404 (1986)). Given the multiplicity of enzyme activities, *E. coli* or another host organism can provide sufficient hydrogenase activity to split incoming molecular hydrogen and reduce the corresponding acceptor. *E. coli* possesses two uptake hydrogenases, Hyd-1 and Hyd-2, encoded by the *hyaABCDEF* and *hybOABCDEFG* gene clusters, respectively (Lukey et al., How *E. coli* is equipped to oxidize hydrogen under

10 different redox conditions, *J Biol Chem* published online Nov 16, 2009). Hyd-1 is oxygen-tolerant, irreversible, and is coupled to quinone reduction via the *hyaC* cytochrome. Hyd-2 is sensitive to O₂, reversible, and transfers electrons to the periplasmic ferredoxin *hybA* which, in turn, reduces a quinone via the *hybB* integral membrane protein. Reduced quinones can serve as the source of electrons for fumarate reductase in the reductive branch of the TCA

15 cycle. Reduced ferredoxins can be used by enzymes such as NAD(P)H:ferredoxin oxidoreductases to generate NADPH or NADH. They can alternatively be used as the

electron donor for reactions such as pyruvate ferredoxin oxidoreductase, AKG ferredoxin oxidoreductase, and 5,10-methylene-H4folate reductase.

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
HyaA	AAC74057.1	1787206	<i>Escherichia coli</i>
HyaB	AAC74058.1	1787207	<i>Escherichia coli</i>
HyaC	AAC74059.1	1787208	<i>Escherichia coli</i>
HyaD	AAC74060.1	1787209	<i>Escherichia coli</i>
HyaE	AAC74061.1	1787210	<i>Escherichia coli</i>
HyaF	AAC74062.1	1787211	<i>Escherichia coli</i>
HybO	AAC76033.1	1789371	<i>Escherichia coli</i>
HybA	AAC76032.1	1789370	<i>Escherichia coli</i>
HybB	AAC76031.1	2367183	<i>Escherichia coli</i>
HybC	AAC76030.1	1789368	<i>Escherichia coli</i>
HybD	AAC76029.1	1789367	<i>Escherichia coli</i>
HybE	AAC76028.1	1789366	<i>Escherichia coli</i>
HybF	AAC76027.1	1789365	<i>Escherichia coli</i>
HybG	AAC76026.1	1789364	<i>Escherichia coli</i>

The hydrogen-lyase systems of *E. coli* include hydrogenase 3, a membrane-bound enzyme complex using ferredoxin as an acceptor, and hydrogenase 4 that also uses a ferredoxin acceptor. Hydrogenase 3 and 4 are encoded by the *hyc* and *hyf* gene clusters, respectively. Hydrogenase 3 has been shown to be a reversible enzyme (Maeda et al., *Appl Microbiol Biotechnol* 76(5):1035-42 (2007)). Hydrogenase activity in *E. coli* is also dependent upon the expression of the *hyp* genes whose corresponding proteins are involved in the assembly of the hydrogenase complexes (Jacobi et al., *Arch. Microbiol* 158:444-451 (1992); Rangarajan et al., *J. Bacteriol.* 190:1447-1458 (2008)).

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
HycA	NP_417205	16130632	<i>Escherichia coli</i>
HycB	NP_417204	16130631	<i>Escherichia coli</i>
HycC	NP_417203	16130630	<i>Escherichia coli</i>
HycD	NP_417202	16130629	<i>Escherichia coli</i>
HycE	NP_417201	16130628	<i>Escherichia coli</i>
HycF	NP_417200	16130627	<i>Escherichia coli</i>
HycG	NP_417199	16130626	<i>Escherichia coli</i>
HycH	NP_417198	16130625	<i>Escherichia coli</i>
HycI	NP_417197	16130624	<i>Escherichia coli</i>
HyfA	NP_416976	90111444	<i>Escherichia coli</i>
HyfB	NP_416977	16130407	<i>Escherichia coli</i>
HyfC	NP_416978	90111445	<i>Escherichia coli</i>
HyfD	NP_416979	16130409	<i>Escherichia coli</i>
HyfE	NP_416980	16130410	<i>Escherichia coli</i>

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
HyfF	NP_416981	16130411	<i>Escherichia coli</i>
HyfG	NP_416982	16130412	<i>Escherichia coli</i>
HyfH	NP_416983	16130413	<i>Escherichia coli</i>
HyfI	NP_416984	16130414	<i>Escherichia coli</i>
HyfJ	NP_416985	90111446	<i>Escherichia coli</i>
HyfR	NP_416986	90111447	<i>Escherichia coli</i>
HypA	NP_417206	16130633	<i>Escherichia coli</i>
HypB	NP_417207	16130634	<i>Escherichia coli</i>
HypC	NP_417208	16130635	<i>Escherichia coli</i>
HypD	NP_417209	16130636	<i>Escherichia coli</i>
HypE	NP_417210	226524740	<i>Escherichia coli</i>
HypF	NP_417192	16130619	<i>Escherichia coli</i>

The *M. thermoacetica* hydrogenases are suitable for a host that lacks sufficient endogenous hydrogenase activity. *M. thermoacetica* can grow with CO₂ as the exclusive carbon source indicating that reducing equivalents are extracted from H₂ to enable acetyl-CoA synthesis via the Wood-Ljungdahl pathway (Drake, H. L., *J. Bacteriol.* 150:702-709 (1982); Drake and Daniel, *Res. Microbiol.* 155:869-883 (2004); Kellum and Drake, *J. Bacteriol.* 160:466-469 (1984)) (see Figure 4). *M. thermoacetica* has homologs to several *hyp*, *hyc*, and *hyf* genes from *E. coli*. The protein sequences encoded for by these genes are identified by the following GenBank accession numbers.

- 10 Proteins in *M. thermoacetica* whose genes are homologous to the *E. coli hyp* genes are shown below.

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
Moth_2175	YP_431007	83590998	<i>Moorella thermoacetica</i>
Moth_2176	YP_431008	83590999	<i>Moorella thermoacetica</i>
Moth_2177	YP_431009	83591000	<i>Moorella thermoacetica</i>
Moth_2178	YP_431010	83591001	<i>Moorella thermoacetica</i>
Moth_2179	YP_431011	83591002	<i>Moorella thermoacetica</i>
Moth_2180	YP_431012	83591003	<i>Moorella thermoacetica</i>
Moth_2181	YP_431013	83591004	<i>Moorella thermoacetica</i>

Proteins in *M. thermoacetica* that are homologous to the *E. coli* Hydrogenase 3 and/or 4 proteins are listed in the following table.

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
Moth_2182	YP_431014	83591005	<i>Moorella thermoacetica</i>
Moth_2183	YP_431015	83591006	<i>Moorella thermoacetica</i>
Moth_2184	YP_431016	83591007	<i>Moorella thermoacetica</i>

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
Moth_2185	YP_431017	83591008	<i>Moorella thermoacetica</i>
Moth_2186	YP_431018	83591009	<i>Moorella thermoacetica</i>
Moth_2187	YP_431019	83591010	<i>Moorella thermoacetica</i>
Moth_2188	YP_431020	83591011	<i>Moorella thermoacetica</i>
Moth_2189	YP_431021	83591012	<i>Moorella thermoacetica</i>
Moth_2190	YP_431022	83591013	<i>Moorella thermoacetica</i>
Moth_2191	YP_431023	83591014	<i>Moorella thermoacetica</i>
Moth_2192	YP_431024	83591015	<i>Moorella thermoacetica</i>

In addition, several gene clusters encoding hydrogenase functionality are present in *M. thermoacetica* and their corresponding protein sequences are provided below.

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
Moth_0439	YP_429313	83589304	<i>Moorella thermoacetica</i>
Moth_0440	YP_429314	83589305	<i>Moorella thermoacetica</i>
Moth_0441	YP_429315	83589306	<i>Moorella thermoacetica</i>
Moth_0442	YP_429316	83589307	<i>Moorella thermoacetica</i>
Moth_0809	YP_429670	83589661	<i>Moorella thermoacetica</i>
Moth_0810	YP_429671	83589662	<i>Moorella thermoacetica</i>
Moth_0811	YP_429672	83589663	<i>Moorella thermoacetica</i>
Moth_0812	YP_429673	83589664	<i>Moorella thermoacetica</i>
Moth_0814	YP_429674	83589665	<i>Moorella thermoacetica</i>
Moth_0815	YP_429675	83589666	<i>Moorella thermoacetica</i>
Moth_0816	YP_429676	83589667	<i>Moorella thermoacetica</i>
Moth_1193	YP_430050	83590041	<i>Moorella thermoacetica</i>
Moth_1194	YP_430051	83590042	<i>Moorella thermoacetica</i>
Moth_1195	YP_430052	83590043	<i>Moorella thermoacetica</i>
Moth_1196	YP_430053	83590044	<i>Moorella thermoacetica</i>
Moth_1717	YP_430562	83590553	<i>Moorella thermoacetica</i>
Moth_1718	YP_430563	83590554	<i>Moorella thermoacetica</i>
Moth_1719	YP_430564	83590555	<i>Moorella thermoacetica</i>
Moth_1883	YP_430726	83590717	<i>Moorella thermoacetica</i>
Moth_1884	YP_430727	83590718	<i>Moorella thermoacetica</i>
Moth_1885	YP_430728	83590719	<i>Moorella thermoacetica</i>
Moth_1886	YP_430729	83590720	<i>Moorella thermoacetica</i>
Moth_1887	YP_430730	83590721	<i>Moorella thermoacetica</i>
Moth_1888	YP_430731	83590722	<i>Moorella thermoacetica</i>
Moth_1452	YP_430305	83590296	<i>Moorella thermoacetica</i>
Moth_1453	YP_430306	83590297	<i>Moorella thermoacetica</i>
Moth_1454	YP_430307	83590298	<i>Moorella thermoacetica</i>

- 5 *Ralstonia eutropha* H16 uses hydrogen as an energy source with oxygen as a terminal electron acceptor. Its membrane-bound uptake [NiFe]-hydrogenase is an ‘‘O₂-tolerant’’ hydrogenase (Cracknell, et al. Proc Nat Acad Sci, 106(49) 20681-20686 (2009)) that is

periplasmically-oriented and connected to the respiratory chain via a b-type cytochrome (Schink and Schlegel, *Biochim. Biophys. Acta*, 567, 315-324 (1979); Bernhard et al., *Eur. J. Biochem.* 248, 179–186 (1997)). *R. eutropha* also contains an O₂-tolerant soluble hydrogenase encoded by the *Hox* operon which is cytoplasmic and directly reduces NAD⁺ at the expense of hydrogen (Schneider and Schlegel, *Biochim. Biophys. Acta* 452, 66–80 (1976); Burgdorf, *J. Bact.* 187(9) 3122-3132(2005)). Soluble hydrogenase enzymes are additionally present in several other organisms including *Geobacter sulfurreducens* (Coppi, *Microbiology* 151, 1239–1254 (2005)), *Synechocystis* str. PCC 6803 (Germer, *J. Biol. Chem.*, 284(52), 36462–36472 (2009)), and *Thiocapsa roseopersicina* (Rakhely, *Appl. Environ. Microbiol.* 70(2) 722–728 (2004)). The *Synechocystis* enzyme is capable of generating NADPH from hydrogen. Overexpression of both the *Hox* operon from *Synechocystis* str. PCC 6803 and the accessory genes encoded by the *Hyp* operon from *Nostoc* sp. PCC 7120 led to increased hydrogenase activity compared to expression of the *Hox* genes alone (Germer, *J. Biol. Chem.* 284(52), 36462–36472 (2009)).

Protein	GenBank ID	GI Number	Organism
HoxF	NP_942727.1	38637753	<i>Ralstonia eutropha</i> H16
HoxU	NP_942728.1	38637754	<i>Ralstonia eutropha</i> H16
HoxY	NP_942729.1	38637755	<i>Ralstonia eutropha</i> H16
HoxH	NP_942730.1	38637756	<i>Ralstonia eutropha</i> H16
HoxW	NP_942731.1	38637757	<i>Ralstonia eutropha</i> H16
HoxI	NP_942732.1	38637758	<i>Ralstonia eutropha</i> H16
HoxE	NP_953767.1	39997816	<i>Geobacter sulfurreducens</i>
HoxF	NP_953766.1	39997815	<i>Geobacter sulfurreducens</i>
HoxU	NP_953765.1	39997814	<i>Geobacter sulfurreducens</i>
HoxY	NP_953764.1	39997813	<i>Geobacter sulfurreducens</i>
HoxH	NP_953763.1	39997812	<i>Geobacter sulfurreducens</i>
GSU2717	NP_953762.1	39997811	<i>Geobacter sulfurreducens</i>
HoxE	NP_441418.1	16330690	<i>Synechocystis</i> str. PCC 6803
HoxF	NP_441417.1	16330689	<i>Synechocystis</i> str. PCC 6803
Unknown function	NP_441416.1	16330688	<i>Synechocystis</i> str. PCC 6803
HoxU	NP_441415.1	16330687	<i>Synechocystis</i> str. PCC 6803
HoxY	NP_441414.1	16330686	<i>Synechocystis</i> str. PCC 6803
Unknown function	NP_441413.1	16330685	<i>Synechocystis</i> str. PCC 6803
Unknown function	NP_441412.1	16330684	<i>Synechocystis</i> str. PCC 6803
HoxH	NP_441411.1	16330683	<i>Synechocystis</i> str. PCC 6803
HypF	NP_484737.1	17228189	<i>Nostoc</i> sp. PCC 7120
HypC	NP_484738.1	17228190	<i>Nostoc</i> sp. PCC 7120

Protein	GenBank ID	GI Number	Organism
HypD	NP_484739.1	17228191	<i>Nostoc</i> sp. PCC 7120
Unknown function	NP_484740.1	17228192	<i>Nostoc</i> sp. PCC 7120
HypE	NP_484741.1	17228193	<i>Nostoc</i> sp. PCC 7120
HypA	NP_484742.1	17228194	<i>Nostoc</i> sp. PCC 7120
HypB	NP_484743.1	17228195	<i>Nostoc</i> sp. PCC 7120
Hox1E	AAP50519.1	37787351	<i>Thiocapsa roseopersicina</i>
Hox1F	AAP50520.1	37787352	<i>Thiocapsa roseopersicina</i>
Hox1U	AAP50521.1	37787353	<i>Thiocapsa roseopersicina</i>
Hox1Y	AAP50522.1	37787354	<i>Thiocapsa roseopersicina</i>
Hox1H	AAP50523.1	37787355	<i>Thiocapsa roseopersicina</i>

Genes encoding hydrogenase enzymes from *C. ljungdahli* are shown below.

Protein	GenBank ID	GI Number	Organism
<i>CLJU_c20290</i>	ADK15091.1	300435324	<i>Clostridium ljungdahli</i>
<i>CLJU_c07030</i>	ADK13773.1	300434006	<i>Clostridium ljungdahli</i>
<i>CLJU_c07040</i>	ADK13774.1	300434007	<i>Clostridium ljungdahli</i>
<i>CLJU_c07050</i>	ADK13775.1	300434008	<i>Clostridium ljungdahli</i>
<i>CLJU_c07060</i>	ADK13776.1	300434009	<i>Clostridium ljungdahli</i>
<i>CLJU_c07070</i>	ADK13777.1	300434010	<i>Clostridium ljungdahli</i>
<i>CLJU_c07080</i>	ADK13778.1	300434011	<i>Clostridium ljungdahli</i>
<i>CLJU_c14730</i>	ADK14541.1	300434774	<i>Clostridium ljungdahli</i>
<i>CLJU_c14720</i>	ADK14540.1	300434773	<i>Clostridium ljungdahli</i>
<i>CLJU_c14710</i>	ADK14539.1	300434772	<i>Clostridium ljungdahli</i>
<i>CLJU_c14700</i>	ADK14538.1	300434771	<i>Clostridium ljungdahli</i>
<i>CLJU_c28670</i>	ADK15915.1	300436148	<i>Clostridium ljungdahli</i>
<i>CLJU_c28660</i>	ADK15914.1	300436147	<i>Clostridium ljungdahli</i>
<i>CLJU_c28650</i>	ADK15913.1	300436146	<i>Clostridium ljungdahli</i>
<i>CLJU_c28640</i>	ADK15912.1	300436145	<i>Clostridium ljungdahli</i>

Several enzymes and the corresponding genes used for fixing carbon dioxide to either pyruvate or phosphoenolpyruvate to form the TCA cycle intermediates, oxaloacetate or malate are described below.

Carboxylation of phosphoenolpyruvate to oxaloacetate is catalyzed by phosphoenolpyruvate carboxylase. Exemplary PEP carboxylase enzymes are encoded by *ppc* in *E. coli* (Kai et al., *Arch. Biochem. Biophys.* 414:170-179 (2003)), *ppcA* in *Methylobacterium extorquens* AM1 (Arps et al., *J. Bacteriol.* 175:3776-3783 (1993)), and *ppc* in *Corynebacterium glutamicum* (Eikmanns et al., *Mol. Gen. Genet.* 218:330-339 (1989)).

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>Ppc</i>	NP_418391	16131794	<i>Escherichia coli</i>
<i>ppcA</i>	AAB58883	28572162	<i>Methylobacterium extorquens</i>
<i>Ppc</i>	ABB53270	80973080	<i>Corynebacterium glutamicum</i>

An alternative enzyme for converting phosphoenolpyruvate to oxaloacetate is PEP carboxykinase, which simultaneously forms an ATP while carboxylating PEP. In most organisms PEP carboxykinase serves a gluconeogenic function and converts oxaloacetate to PEP at the expense of one ATP. *S. cerevisiae* is one such organism whose native PEP carboxykinase, *PCK1*, serves a gluconeogenic role (Valdes-Hevia et al., *FEBS Lett.* 258:313-316 (1989). *E. coli* is another such organism, as the role of PEP carboxykinase in producing oxaloacetate is believed to be minor when compared to PEP carboxylase, which does not form ATP, possibly due to the higher K_m for bicarbonate of PEP carboxykinase (Kim et al., *Appl. Environ. Microbiol.* 70:1238-1241 (2004)). Nevertheless, activity of the native *E. coli* PEP carboxykinase from PEP towards oxaloacetate has been recently demonstrated in *ppc* mutants of *E. coli* K-12 (Kwon et al., *J. Microbiol. Biotechnol.* 16:1448-1452 (2006)). These strains exhibited no growth defects and had increased succinate production at high NaHCO_3 concentrations. Mutant strains of *E. coli* can adopt Pck as the dominant CO_2 -fixing enzyme following adaptive evolution (Zhang et al. 2009). In some organisms, particularly rumen bacteria, PEP carboxykinase is quite efficient in producing oxaloacetate from PEP and generating ATP. Examples of PEP carboxykinase genes that have been cloned into *E. coli* include those from *Mannheimia succiniciproducens* (Lee et al., *Biotechnol. Bioprocess Eng.* 7:95-99 (2002)), *Anaerobiospirillum succiniciproducens* (Laivenieks et al., *Appl. Environ. Microbiol.* 63:2273-2280 (1997)), and *Actinobacillus succinogenes* (Kim et al. *supra*). The PEP carboxykinase enzyme encoded by *Haemophilus influenza* is effective at forming oxaloacetate from PEP.

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>PCK1</i>	NP_013023	6322950	<i>Saccharomyces cerevisiae</i>
<i>pck</i>	NP_417862.1	16131280	<i>Escherichia coli</i>
<i>pckA</i>	YP_089485.1	52426348	<i>Mannheimia succiniciproducens</i>
<i>pckA</i>	O09460.1	3122621	<i>Anaerobiospirillum succiniciproducens</i>
<i>pckA</i>	Q6W6X5	75440571	<i>Actinobacillus succinogenes</i>
<i>pckA</i>	P43923.1	1172573	<i>Haemophilus influenza</i>

Pyruvate carboxylase (EC 6.4.1.1) directly converts pyruvate to oxaloacetate at the cost of one ATP. Pyruvate carboxylase enzymes are encoded by *PYC1* (Walker et al., *Biochem. Biophys. Res. Commun.* 176:1210-1217 (1991) and *PYC2* (Walker et al., *supra*) in *Saccharomyces cerevisiae*, and *pyc* in *Mycobacterium smegmatis* (Mukhopadhyay and

5 Purwantini, *Biochim. Biophys. Acta* 1475:191-206 (2000)).

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>PYC1</i>	NP_011453	6321376	<i>Saccharomyces cerevisiae</i>
<i>PYC2</i>	NP_009777	6319695	<i>Saccharomyces cerevisiae</i>
<i>Pyc</i>	YP_890857.1	118470447	<i>Mycobacterium smegmatis</i>

Malic enzyme can be applied to convert CO₂ and pyruvate to malate at the expense of one reducing equivalent. Malic enzymes for this purpose can include, without limitation, malic enzyme (NAD-dependent) and malic enzyme (NADP-dependent). For example, one of the

10 *E. coli* malic enzymes (Takeo, *J. Biochem.* 66:379-387 (1969)) or a similar enzyme with higher activity can be expressed to enable the conversion of pyruvate and CO₂ to malate. By fixing carbon to pyruvate as opposed to PEP, malic enzyme allows the high-energy phosphate bond from PEP to be conserved by pyruvate kinase whereby ATP is generated in the formation of pyruvate or by the phosphotransferase system for glucose transport. Although

15 malic enzyme is typically assumed to operate in the direction of pyruvate formation from malate, overexpression of the NAD-dependent enzyme, encoded by *maeA*, has been demonstrated to increase succinate production in *E. coli* while restoring the lethal Δ pfl- Δ ldhA phenotype under anaerobic conditions by operating in the carbon-fixing direction (Stols and Donnelly, *Appl. Environ. Microbiol.* 63(7) 2695-2701 (1997)). A similar observation was

20 made upon overexpressing the malic enzyme from *Ascaris suum* in *E. coli* (Stols et al., *Appl. Biochem. Biotechnol.* 63-65(1), 153-158 (1997)). The second *E. coli* malic enzyme, encoded by *maeB*, is NADP-dependent and also decarboxylates oxaloacetate and other alpha-keto acids (Iwakura et al., *J. Biochem.* 85(5):1355-65 (1979)).

<u>Protein</u>	<u>GenBank ID</u>	<u>GI Number</u>	<u>Organism</u>
<i>maeA</i>	NP_415996	90111281	<i>Escherichia coli</i>
<i>maeB</i>	NP_416958	16130388	<i>Escherichia coli</i>
<i>NAD-ME</i>	P27443	126732	<i>Ascaris suum</i>

25 The enzymes used for converting oxaloacetate (formed from, for example, PEP carboxylase, PEP carboxykinase, or pyruvate carboxylase) or malate (formed from, for example, malic

enzyme or malate dehydrogenase) to succinyl-CoA via the reductive branch of the TCA cycle are malate dehydrogenase, fumarate hydratase (fumarase), fumarate reductase, and succinyl-CoA transferase. The genes for each of the enzymes are described herein above.

Enzymes, genes and methods for engineering pathways from succinyl-CoA to various products into a microorganism are now known in the art. The additional reducing equivalents obtained from CO and/or H₂, as disclosed herein, improve the yields of ethylene glycol when utilizing carbohydrate-based feedstock.

Enzymes, genes and methods for engineering pathways from glycolysis intermediates to various products into a microorganism are known in the art. The additional reducing equivalents obtained from CO and H₂, as described herein, improve the yields of all these products on carbohydrates.

EXAMPLE VII

Methods for Handling CO and Anaerobic Cultures

This example describes methods used in handling CO and anaerobic cultures.

A. Handling of CO in small quantities for assays and small cultures. CO is an odorless, colorless and tasteless gas that is a poison. Therefore, cultures and assays that utilized CO required special handling. Several assays, including CO oxidation, acetyl-CoA synthesis, CO concentration using myoglobin, and CO tolerance/utilization in small batch cultures, called for small quantities of the CO gas that were dispensed and handled within a fume hood. Biochemical assays called for saturating very small quantities (<2 mL) of the biochemical assay medium or buffer with CO and then performing the assay. All of the CO handling steps were performed in a fume hood with the sash set at the proper height and blower turned on; CO was dispensed from a compressed gas cylinder and the regulator connected to a Schlenk line. The latter ensures that equal concentrations of CO were dispensed to each of several possible cuvettes or vials. The Schlenk line was set up containing an oxygen scrubber on the input side and an oil pressure release bubbler and vent on the other side. Assay cuvettes were both anaerobic and CO-containing. Therefore, the assay cuvettes were tightly sealed with a rubber stopper and reagents were added or removed using gas-tight needles and syringes. Secondly, small (~50 mL) cultures were grown with saturating CO in tightly stoppered serum bottles. As with the biochemical assays, the CO-saturated microbial cultures were equilibrated in the fume hood using the Schlenk line setup. Both the biochemical assays

and microbial cultures were in portable, sealed containers and in small volumes making for safe handling outside of the fume hood. The compressed CO tank was adjacent to the fume hood.

Typically, a Schlenk line was used to dispense CO to cuvettes, each vented. Rubber stoppers on the cuvettes were pierced with 19 or 20 gage disposable syringe needles and were vented with the same. An oil bubbler was used with a CO tank and oxygen scrubber. The glass or quartz spectrophotometer cuvettes have a circular hole on top into which a Kontes stopper sleeve, Sz7 774250-0007 was fitted. The CO detector unit was positioned proximal to the fume hood.

B. Handling of CO in larger quantities fed to large-scale cultures. Fermentation cultures are fed either CO or a mixture of CO and H₂ to simulate syngas as a feedstock in fermentative production. Therefore, quantities of cells ranging from 1 liter to several liters can include the addition of CO gas to increase the dissolved concentration of CO in the medium. In these circumstances, fairly large and continuously administered quantities of CO gas are added to the cultures. At different points, the cultures are harvested or samples removed. Alternatively, cells are harvested with an integrated continuous flow centrifuge that is part of the fermenter.

The fermentative processes are carried out under anaerobic conditions. In some cases, it is uneconomical to pump oxygen or air into fermenters to ensure adequate oxygen saturation to provide a respiratory environment. In addition, the reducing power generated during anaerobic fermentation may be needed in product formation rather than respiration.

Furthermore, many of the enzymes for various pathways are oxygen-sensitive to varying degrees. Classic acetogens such as *M. thermoacetica* are obligate anaerobes and the enzymes in the Wood-Ljungdahl pathway are highly sensitive to irreversible inactivation by molecular oxygen. While there are oxygen-tolerant acetogens, the repertoire of enzymes in the Wood-Ljungdahl pathway might be incompatible in the presence of oxygen because most are metallo-enzymes, key components are ferredoxins, and regulation can divert metabolism away from the Wood-Ljungdahl pathway to maximize energy acquisition. At the same time, cells in culture act as oxygen scavengers that moderate the need for extreme measures in the presence of large cell growth.

C. Anaerobic chamber and conditions. Exemplary anaerobic chambers are available commercially (see, for example, Vacuum Atmospheres Company, Hawthorne CA; MBraun, Newburyport MA). Conditions included an O₂ concentration of 1 ppm or less and 1 atm pure N₂. In one example, 3 oxygen scrubbers/catalyst regenerators were used, and the chamber
5 included an O₂ electrode (such as Teledyne; City of Industry CA). Nearly all items and reagents were cycled four times in the airlock of the chamber prior to opening the inner chamber door. Reagents with a volume >5mL were sparged with pure N₂ prior to introduction into the chamber. Gloves are changed twice/yr and the catalyst containers were regenerated periodically when the chamber displays increasingly sluggish response to
10 changes in oxygen levels. The chamber's pressure was controlled through one-way valves activated by solenoids. This feature allowed setting the chamber pressure at a level higher than the surroundings to allow transfer of very small tubes through the purge valve.

The anaerobic chambers achieved levels of O₂ that were consistently very low and were needed for highly oxygen sensitive anaerobic conditions. However, growth and handling of
15 cells does not usually require such precautions. In an alternative anaerobic chamber configuration, platinum or palladium can be used as a catalyst that requires some hydrogen gas in the mix. Instead of using solenoid valves, pressure release can be controlled by a bubbler. Instead of using instrument-based O₂ monitoring, test strips can be used instead.

D. Anaerobic microbiology. Small cultures were handled as described above for CO
20 handling. In particular, serum or media bottles are fitted with thick rubber stoppers and aluminum crimps are employed to seal the bottle. Medium, such as Terrific Broth, is made in a conventional manner and dispensed to an appropriately sized serum bottle. The bottles are sparged with nitrogen for ~30 min of moderate bubbling. This removes most of the oxygen from the medium and, after this step, each bottle is capped with a rubber stopper (such as
25 Bellco 20 mm septum stoppers; Bellco, Vineland, NJ) and crimp-sealed (Bellco 20 mm). Then the bottles of medium are autoclaved using a slow (liquid) exhaust cycle. At least sometimes a needle can be poked through the stopper to provide exhaust during autoclaving; the needle needs to be removed immediately upon removal from the autoclave. The sterile medium has the remaining medium components, for example buffer or antibiotics, added via
30 syringe and needle. Prior to addition of reducing agents, the bottles are equilibrated for 30 - 60 minutes with nitrogen (or CO depending upon use). A reducing agent such as a 100 x 150 mM sodium sulfide, 200 mM cysteine-HCl is added. This is made by weighing the sodium

sulfide into a dry beaker and the cysteine into a serum bottle, bringing both into the anaerobic chamber, dissolving the sodium sulfide into anaerobic water, then adding this to the cysteine in the serum bottle. The bottle is stoppered immediately as the sodium sulfide solution generates hydrogen sulfide gas upon contact with the cysteine. When injecting into the culture, a syringe filter is used to sterilize the solution. Other components are added through syringe needles, such as B12 (10 μ M cyanocobalamin), nickel chloride (NiCl_2 , 20 μ M final concentration from a 40 mM stock made in anaerobic water in the chamber and sterilized by autoclaving or by using a syringe filter upon injection into the culture), and ferrous ammonium sulfate (final concentration needed is 100 μ M—made as 100-1000x stock solution in anaerobic water in the chamber and sterilized by autoclaving or by using a syringe filter upon injection into the culture). To facilitate faster growth under anaerobic conditions, the 1 liter bottles were inoculated with 50 mL of a preculture grown anaerobically. Induction of the pA1-lacO1 promoter in the vectors was performed by addition of isopropyl β -D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.2 mM and was carried out for about 3 hrs.

Large cultures can be grown in larger bottles using continuous gas addition while bubbling. A rubber stopper with a metal bubbler is placed in the bottle after medium addition and sparged with nitrogen for 30 minutes or more prior to setting up the rest of the bottle. Each bottle is put together such that a sterile filter will sterilize the gas bubbled in and the hoses on the bottles are compressible with small C clamps. Medium and cells are stirred with magnetic stir bars. Once all medium components and cells are added, the bottles are incubated in an incubator in room air but with continuous nitrogen sparging into the bottles.

EXAMPLE VIII

CO oxidation (CODH) Assay

This example describes assay methods for measuring CO oxidation (CO dehydrogenase; CODH).

The 7 gene CODH/ACS operon of *Moorella thermoacetica* was cloned into *E. coli* expression vectors. The intact ~10 kbp DNA fragment was cloned, and it is likely that some of the genes in this region are expressed from their own endogenous promoters and all contain endogenous ribosomal binding sites. These clones were assayed for CO oxidation, using an assay that quantitatively measures CODH activity. Antisera to the *M. thermoacetica* gene products was used for Western blots to estimate specific activity. *M. thermoacetica* is

Gram positive, and ribosome binding site elements are expected to work well in *E. coli*. This activity, described below in more detail, was estimated to be ~1/50th of the *M. thermoacetica* specific activity. It is possible that CODH activity of recombinant *E. coli* cells could be limited by the fact that *M. thermoacetica* enzymes have temperature optima around 55°C.

5 Therefore, a mesophilic CODH/ACS pathway could be advantageous such as the close relative of *Moorella* that is mesophilic and does have an apparently intact CODH/ACS operon and a Wood-Ljungdahl pathway, *Desulfitobacterium hafniense*. Acetogens as potential host organisms include, but are not limited to, *Rhodospirillum rubrum*, *Moorella thermoacetica* and *Desulfitobacterium hafniense*.

10 CO oxidation is both the most sensitive and most robust of the CODH/ACS assays. It is likely that an *E. coli*-based syngas using system will ultimately need to be about as anaerobic as *Clostridial* (i.e., *Moorella*) systems, especially for maximal activity. Improvement in CODH should be possible but will ultimately be limited by the solubility of CO gas in water.

Initially, each of the genes was cloned individually into expression vectors. Combined
15 expression units for multiple subunits/1 complex were generated. Expression in *E. coli* at the protein level was determined. Both combined *M. thermoacetica* CODH/ACS operons and individual expression clones were made.

CO oxidation assay. This assay is one of the simpler, reliable, and more versatile assays of enzymatic activities within the Wood-Ljungdahl pathway and tests CODH (Seravalli et al.,
20 *Biochemistry* 43:3944-3955 (2004)). A typical activity of *M. thermoacetica* CODH specific activity is 500 U at 55°C or ~60U at 25°C. This assay employs reduction of methyl viologen in the presence of CO. This is measured at 578 nm in stoppered, anaerobic, glass cuvettes.

In more detail, glass rubber stoppered cuvettes were prepared after first washing the cuvette four times in deionized water and one time with acetone. A small amount of vacuum grease
25 was smeared on the top of the rubber gasket. The cuvette was gassed with CO, dried 10 min with a 22 Ga. needle plus an exhaust needle. A volume of 0.98 mL of reaction buffer (50 mM Hepes, pH 8.5, 2mM dithiothreitol (DTT) was added using a 22 Ga. needle, with exhaust needed, and 100%CO. Methyl viologen (CH₃ viologen) stock was 1 M in water. Each assay used 20 microliters for 20 mM final concentration. When methyl viologen was added, an 18
30 Ga needle (partial) was used as a jacket to facilitate use of a Hamilton syringe to withdraw the CH₃ viologen. 4 -5 aliquots were drawn up and discarded to wash and gas equilibrate the

syringe. A small amount of sodium dithionite (0.1 M stock) was added when making up the CH₃ viologen stock to slightly reduce the CH₃ viologen. The temperature was equilibrated to 55°C in a heated Olis spectrophotometer (Bogart GA). A blank reaction (CH₃ viologen + buffer) was run first to measure the base rate of CH₃ viologen reduction. Crude *E. coli* cell extracts of ACS90 and ACS91 (CODH-ACS operon of *M. thermoacetica* with and without, respectively, the first *cooC*). 10 microliters of extract were added at a time, mixed and assayed. Reduced CH₃ viologen turns purple. The results of an assay are shown in Table I.

Table I. Crude extract CO Oxidation Activities.

ACS90	7.7 mg/ml	ACS91	11.8 mg/ml	
Mta98	9.8 mg/ml	Mta99	11.2 mg/ml	
Extract	Vol	OD/	U/ml	U/mg
ACS90	10 microliters	0.073	0.376	0.049
ACS91	10 microliters	0.096	0.494	0.042
Mta99	10 microliters	0.0031	0.016	0.0014
ACS90	10 microliters	0.099	0.51	0.066
Mta99	25 microliters	0.012	0.025	0.0022
ACS91	25 microliters	0.215	0.443	0.037
Mta98	25 microliters	0.019	0.039	0.004
ACS91	10 microliters	0.129	0.66	0.056
Averages				
ACS90	0.057 U/mg			
ACS91	0.045 U/mg			
Mta99	0.0018 U/mg			

Mta98/Mta99 are *E. coli* MG1655 strains that express methanol methyltransferase genes from *M. thermoacetia* and, therefore, are negative controls for the ACS90 ACS91 *E. coli* strains that contain *M. thermoacetica* CODH operons.

If ~ 1% of the cellular protein is CODH, then these figures would be approximately 100X less than the 500 U/mg activity of pure *M. thermoacetica* CODH. Actual estimates based on Western blots are 0.5% of the cellular protein, so the activity is about 50X less than for *M. thermoacetica* CODH. Nevertheless, this experiment demonstrates CO oxidation activity in recombinant *E. coli* with a much smaller amount in the negative controls. The small amount of CO oxidation (CH₃ viologen reduction) seen in the negative controls indicates that *E. coli* may have a limited ability to reduce CH₃ viologen.

To estimate the final concentrations of CODH and Mtr proteins, SDS-PAGE followed by Western blot analyses were performed on the same cell extracts used in the CO oxidation, ACS, methyltransferase, and corrinoid Fe-S assays. The antisera used were polyclonal to purified *M. thermoacetica* CODH-ACS and Mtr proteins and were visualized using an alkaline phosphatase-linked goat-anti-rabbit secondary antibody. The Westerns were performed and results are shown in Figure 9. The amounts of CODH in ACS90 and ACS91 were estimated at 50 ng by comparison to the control lanes. Expression of CODH-ACS operon genes including 2 CODH subunits and the methyltransferase were confirmed via Western blot analysis. Therefore, the recombinant *E. coli* cells express multiple components of a 7 gene operon. In addition, both the methyltransferase and corrinoid iron sulfur protein were active in the same recombinant *E. coli* cells. These proteins are part of the same operon cloned into the same cells.

The CO oxidation assays were repeated using extracts of *Moorella thermoacetica* cells for the positive controls. Though CODH activity in *E. coli* ACS90 and ACS91 was measurable, it was at about 130 – 150 X lower than the *M. thermoacetica* control. The results of the assay are shown in Figure 10. Briefly, cells (*M. thermoacetica* or *E. coli* with the CODH/ACS operon; ACS90 or ACS91 or empty vector: pZA33S) were grown and extracts prepared as described above. Assays were performed as described above at 55°C at various times on the day the extracts were prepared. Reduction of methylviologen was followed at 578 nm over a 120 sec time course.

These results describe the CO oxidation (CODH) assay and results. Recombinant *E. coli* cells expressed CO oxidation activity as measured by the methyl viologen reduction assay.

EXAMPLE IX

E. coli CO Tolerance Experiment and CO Concentration Assay (myoglobin assay)

This example describes the tolerance of *E. coli* for high concentrations of CO.

To test whether or not *E. coli* can grow anaerobically in the presence of saturating amounts of CO, cultures were set up in 120 ml serum bottles with 50 ml of Terrific Broth medium (plus reducing solution, NiCl₂, Fe(II)NH₄SO₄, cyanocobalamin, IPTG, and chloramphenicol) as described above for anaerobic microbiology in small volumes. One half of these bottles were equilibrated with nitrogen gas for 30 min. and one half was equilibrated with CO gas for 30 min. An empty vector (pZA33) was used as a control, and cultures containing the pZA33

empty vector as well as both ACS90 and ACS91 were tested with both N₂ and CO. All were inoculated and grown for 36 hrs with shaking (250 rpm) at 37°C. At the end of the 36 hour period, examination of the flasks showed high amounts of growth in all. The bulk of the observed growth occurred overnight with a long lag.

- 5 Given that all cultures appeared to grow well in the presence of CO, the final CO concentrations were confirmed. This was performed using an assay of the spectral shift of myoglobin upon exposure to CO. Myoglobin reduced with sodium dithionite has an absorbance peak at 435 nm; this peak is shifted to 423 nm with CO. Due to the low wavelength and need to record a whole spectrum from 300 nm on upwards, quartz cuvettes
10 must be used. CO concentration is measured against a standard curve and depends upon the Henry's Law constant for CO of maximum water solubility = 970 micromolar at 20°C and 1 atm.

- For the myoglobin test of CO concentration, cuvettes were washed 10X with water, 1X with acetone, and then stoppered as with the CODH assay. N₂ was blown into the cuvettes for ~10
15 min. A volume of 1 ml of anaerobic buffer (HEPES, pH 8.0, 2mM DTT) was added to the blank (not equilibrated with CO) with a Hamilton syringe. A volume of 10 microliter myoglobin (~1 mM—can be varied, just need a fairly large amount) and 1 microliter dithionite (20 mM stock) were added. A CO standard curve was made using CO saturated buffer added at 1 microliter increments. Peak height and shift was recorded for each
20 increment. The cultures tested were pZA33/CO, ACS90/CO, and ACS91/CO. Each of these was added in 1 microliter increments to the same cuvette. Midway through the experiment a second cuvette was set up and used. The results are shown in Table II.

Table II. Carbon Monoxide Concentrations, 36 hrs.

Strain and Growth Conditions	Final CO concentration (micromolar)
pZA33-CO	930
ACS90-CO	638
	494
	734
	883
ave	687
SD	164
ACS91-CO	728
	812
	760
	611
ave.	728
SD	85

The results shown in Table II indicate that the cultures grew whether or not a strain was cultured in the presence of CO or not. These results indicate that *E. coli* can tolerate exposure to CO under anaerobic conditions and that *E. coli* cells expressing the CODH-ACS operon can metabolize some of the CO.

- 5 These results demonstrate that *E. coli* cells, whether expressing CODH/ACS or not, were able to grow in the presence of saturating amounts of CO. Furthermore, these grew equally well as the controls in nitrogen in place of CO. This experiment demonstrated that laboratory strains of *E. coli* are insensitive to CO at the levels achievable in a syngas project performed at normal atmospheric pressure. In addition, preliminary experiments indicated that the
- 10 recombinant *E. coli* cells expressing CODH/ACS actually consumed some CO, probably by oxidation to carbon dioxide.

EXAMPLE X

Exemplary Carboxylic Acid Reductases

- 15 This example describes the use of carboxylic acid reductases to carry out the conversion of a carboxylic acid to an aldehyde.

1.2.1.e Acid reductase. The conversion of unactivated acids to aldehydes can be carried out by an acid reductase. Examples of such conversions include, but are not limited, the conversion of 4-hydroxybutyrate, succinate, alpha-ketoglutarate, and 4-aminobutyrate to 4-

20 hydroxybutanal, succinate semialdehyde, 2,5-dioxopentanoate, and 4-aminobutanal, respectively. One notable carboxylic acid reductase can be found in *Nocardia iowensis* which catalyzes the magnesium, ATP and NADPH-dependent reduction of carboxylic acids to their corresponding aldehydes (Venkitasubramanian et al., *J. Biol. Chem.* 282:478-485 (2007)). This enzyme is encoded by the *car* gene and was cloned and functionally expressed

25 in *E. coli* (Venkitasubramanian et al., *J. Biol. Chem.* 282:478-485 (2007)). Expression of the *npt* gene product improved activity of the enzyme via post-transcriptional modification. The *npt* gene encodes a specific phosphopantetheine transferase (PPTase) that converts the inactive apo-enzyme to the active holo-enzyme. The natural substrate of this enzyme is vanillic acid, and the enzyme exhibits broad acceptance of aromatic and aliphatic substrates

30 (Venkitasubramanian et al., in *Biocatalysis in the Pharmaceutical and Biotechnology Industries*, ed. R.N. Patel, Chapter 15, pp. 425-440, CRC Press LLC, Boca Raton, FL. (2006)).

Gene	Accession No.	GI No.	Organism
<i>car</i>	AAR91681.1	40796035	<i>Nocardia iowensis</i> (sp. NRRL 5646)
<i>npt</i>	ABI83656.1	114848891	<i>Nocardia iowensis</i> (sp. NRRL 5646)

Additional *car* and *npt* genes can be identified based on sequence homology.

Gene	Accession No.	GI No.	Organism
<i>fadD9</i>	YP_978699.1	121638475	<i>Mycobacterium bovis</i> BCG
<i>BCG_2812c</i>	YP_978898.1	121638674	<i>Mycobacterium bovis</i> BCG
<i>nfa20150</i>	YP_118225.1	54023983	<i>Nocardia farcinica</i> IFM 10152
<i>nfa40540</i>	YP_120266.1	54026024	<i>Nocardia farcinica</i> IFM 10152
<i>SGR_6790</i>	YP_001828302.1	182440583	<i>Streptomyces griseus</i> subsp. griseus NBRC 13350
<i>SGR_665</i>	YP_001822177.1	182434458	<i>Streptomyces griseus</i> subsp. griseus NBRC 13350

An additional enzyme candidate found in *Streptomyces griseus* is encoded by the *griC* and *griD* genes. This enzyme is believed to convert 3-amino-4-hydroxybenzoic acid to 3-amino-4-hydroxybenzaldehyde as deletion of either *griC* or *griD* led to accumulation of extracellular 3-acetylamino-4-hydroxybenzoic acid, a shunt product of 3-amino-4-hydroxybenzoic acid metabolism (Suzuki, et al., *J. Antibiot.* 60(6):380-387 (2007)). Co-expression of *griC* and *griD* with *SGR_665*, an enzyme similar in sequence to the *Nocardia iowensis npt*, can be beneficial.

Gene	Accession No.	GI No.	Organism
<i>griC</i>	182438036	YP_001825755.1	<i>Streptomyces griseus</i> subsp. griseus NBRC 13350
<i>griD</i>	182438037	YP_001825756.1	<i>Streptomyces griseus</i> subsp. griseus NBRC 13350
MSMEG_2956	YP_887275.1	YP_887275.1	<i>Mycobacterium smegmatis</i> MC2 155
MSMEG_5739	YP_889972.1	118469671	<i>Mycobacterium smegmatis</i> MC2 155
MSMEG_2648	YP_886985.1	118471293	<i>Mycobacterium smegmatis</i> MC2 155
MAP1040c	NP_959974.1	41407138	<i>Mycobacterium avium</i> subsp. paratuberculosis K-10
MAP2899c	NP_961833.1	41408997	<i>Mycobacterium avium</i> subsp. paratuberculosis K-10

Gene	Accession No.	GI No.	Organism
MMAR_2117	YP_001850422.1	183982131	Mycobacterium marinum M
MMAR_2936	YP_001851230.1	183982939	Mycobacterium marinum M
MMAR_1916	YP_001850220.1	183981929	Mycobacterium marinum M
TpauDRAFT_33060	ZP_04027864.1	227980601	Tsukamurella paurometabola DSM 20162
TpauDRAFT_20920	ZP_04026660.1	227979396	Tsukamurella paurometabola DSM 20162
CPCC7001_1320	ZP_05045132.1	254431429	Cyanobium PCC7001
DDBDRAFT_0187729	XP_636931.1	66806417	Dictyostelium discoideum AX4

An enzyme with similar characteristics, alpha-aminoadipate reductase (AAR, EC 1.2.1.31), participates in lysine biosynthesis pathways in some fungal species. This enzyme naturally reduces alpha-aminoadipate to alpha-aminoadipate semialdehyde. The carboxyl group is first

5 activated through the ATP-dependent formation of an adenylate that is then reduced by NAD(P)H to yield the aldehyde and AMP. Like CAR, this enzyme utilizes magnesium and requires activation by a PPTase. Enzyme candidates for AAR and its corresponding PPTase are found in *Saccharomyces cerevisiae* (Morris et al., *Gene* 98:141-145 (1991)), *Candida albicans* (Guo et al., *Mol. Genet. Genomics* 269:271-279 (2003)), and *Schizosaccharomyces*

10 *pombe* (Ford et al., *Curr. Genet.* 28:131-137 (1995)). The AAR from *S. pombe* exhibited significant activity when expressed in *E. coli* (Guo et al., *Yeast* 21:1279-1288 (2004)). The AAR from *Penicillium chrysogenum* accepts S-carboxymethyl-L-cysteine as an alternate substrate, but did not react with adipate, L-glutamate or diaminopimelate (Hijarrubia et al., *J. Biol. Chem.* 278:8250-8256 (2003)). The gene encoding the *P. chrysogenum* PPTase has not

15 been identified to date.

Gene	Accession No.	GI No.	Organism
<i>LYS2</i>	<i>AAA34747.1</i>	<i>171867</i>	<i>Saccharomyces cerevisiae</i>
<i>LYS5</i>	P50113.1	1708896	<i>Saccharomyces cerevisiae</i>
<i>LYS2</i>	AAC02241.1	2853226	<i>Candida albicans</i>
<i>LYS5</i>	AAO26020.1	28136195	<i>Candida albicans</i>
<i>Lys1p</i>	P40976.3	13124791	<i>Schizosaccharomyces pombe</i>
<i>Lys7p</i>	Q10474.1	1723561	<i>Schizosaccharomyces pombe</i>
<i>Lys2</i>	CAA74300.1	3282044	<i>Penicillium chrysogenum</i>

Cloning and Expression of Carboxylic Acid Reductase. *Escherichia coli* is used as a target organism to engineer the pathway for ethylene glycol. *E. coli* provides a good host for generating a non-naturally occurring microorganism capable of producing ethylene glycol. *E. coli* is amenable to genetic manipulation and is known to be capable of producing various intermediates and products effectively under various oxygenation conditions.

To generate a microbial organism strain such as an *E. coli* strain engineered to produce ethylene glycol, nucleic acids encoding a carboxylic acid reductase and phosphopantetheine transferase are expressed in *E. coli* using well known molecular biology techniques (see, for example, Sambrook, supra, 2001; Ausubel supra, 1999). In particular, *car* genes from *Nocardia iowensis* (designated 720), *Mycobacterium smegmatis* mc(2)155 (designated 890), *Mycobacterium avium* subspecies *paratuberculosis* K-10 (designated 891) and *Mycobacterium marinum* M (designated 892) were cloned into pZS*13 vectors (Expressys, Ruelzheim, Germany) under control of PA1/lacO promoters. The *npt* (ABI83656.1) gene (i.e., 721) was cloned into the pKJL33S vector, a derivative of the original mini-F plasmid vector PML31 under control of promoters and ribosomal binding sites similar to those used in pZS*13.

The *car* gene (GNM_720) was cloned by PCR from *Nocardia* genomic DNA. Its nucleic acid and protein sequences are shown in Figures 8A and 8B, respectively. A codon-optimized version of the *npt* gene (GNM_721) was synthesized by GeneArt (Regensburg, Germany). Its nucleic acid and protein sequences are shown in Figures 9A and 9B, respectively. The nucleic acid and protein sequences for the *Mycobacterium smegmatis* mc(2)155 (designated 890), *Mycobacterium avium* subspecies *paratuberculosis* K-10 (designated 891) and *Mycobacterium marinum* M (designated 892) genes and enzymes can be found in Figures 10, 11, and 12, respectively. The plasmids are transformed into a host cell to express the proteins and enzymes required for ethylene glycol production.

Additional CAR variants were generated. A codon optimized version of CAR 891 was generated and designated 891GA. The nucleic acid and amino acid sequences of CAR 891GA are shown in Figures 13A and 13B, respectively. Over 2000 CAR variants were generated. In particular, all 20 amino acid combinations were made at positions V295, M296, G297, G391, G421, D413, G414, Y415, G416, and S417, and additional variants were tested as well. Exemplary CAR variants include: E16K; Q95L; L100M; A1011T; K823E; T941S; H15Q; D198E; G446C; S392N; F699L; V883I; F467S; T987S; R12H; V295G;

- V295A; V295S; V295T; V295C; V295V; V295L; V295I; V295M; V295P; V295F; V295Y; V295W; V295D; V295E; V295N; V295Q; V295H; V295K; V295R; M296G; M296A; M296S; M296T; M296C; M296V; M296L; M296I; M296M; M296P; M296F; M296Y; M296W; M296D; M296E; M296N; M296Q; M296H; M296K; M296R; G297G; G297A; G297S; G297T; G297C; G297V; G297L; G297I; G297M; G297P; G297F; G297Y; G297W; G297D; G297E; G297N; G297Q; G297H; G297K; G297R; G391G; G391A; G391S; G391T; G391C; G391V; G391L; G391I; G391M; G391P; G391F; G391Y; G391W; G391D; G391E; G391N; G391Q; G391H; G391K; G391R; G421G; G421A; G421S; G421T; G421C; G421V; G421L; G421I; G421M; G421P; G421F; G421Y; G421W; G421D; G421E; G421N; G421Q; G421H; G421K; G421R; D413G; D413A; D413S; D413T; D413C; D413V; D413L; D413I; D413M; D413P; D413F; D413Y; D413W; D413D; D413E; D413N; D413Q; D413H; D413K; D413R; G414G; G414A; G414S; G414T; G414C; G414V; G414L; G414I; G414M; G414P; G414F; G414Y; G414W; G414D; G414E; G414N; G414Q; G414H; G414K; G414R; Y415G; Y415A; Y415S; Y415T; Y415C; Y415V; Y415L; Y415I; Y415M; Y415P; Y415F; Y415Y; Y415W; Y415D; Y415E; Y415N; Y415Q; Y415H; Y415K; Y415R; G416G; G416A; G416S; G416T; G416C; G416V; G416L; G416I; G416M; G416P; G416F; G416Y; G416W; G416D; G416E; G416N; G416Q; G416H; G416K; G416R; S417G; S417A; S417S; S417T; S417C; S417V; S417L; S417I; S417M; S417P; S417F; S417Y; S417W; S417D; S417E; S417N; S417Q; S417H; S417K; and S417R.
- The CAR variants were screened for activity, and numerous CAR variants were found to exhibit CAR activity.

This example describes the use of CAR for converting carboxylic acids to aldehydes.

EXAMPLE XI

Pathways for producing ethylene glycol from 3-phosphohydroxypyruvate

- Figure 14 shows exemplary pathways for converting 3-phosphohydroxypyruvate to ethylene glycol. 3-Phosphohydroxypyruvate is an intermediate in serine biosynthesis. The oxidation of 3-phosphoglycerate to 3-phosphohydroxypyruvate by 3-phosphoglycerate dehydrogenase is the first committed step of serine biosynthesis. The EG pathways shown in Figure 14 involve decarboxylation of 3-phosphohydroxypyruvate to 2-oxoethyl-phosphate (step 14A) by a keto-acid decarboxylase. EG is then formed by the reduction and dephosphorylation of 2-oxoethyl-phosphate via 2-hydroxyethyl-phosphate (steps 14B/14C) or glycolaldehyde (steps 14D/14E). Enzyme candidates for steps 14A-14E are described below.

The maximum theoretical yield of ethylene glycol from glucose is 2.4 mol/mol (0.835 g/g). The maximum yield of ethylene glycol from glucose via the 3-phosphohydroxypyruvate pathways shown in Figure 14 is 2 mol/mol glucose. The yield of this pathway may be limited due to the pathway not being redox balanced. For every equivalent of glucose converted to 2 equivalents of EG, an excess of 4 reducing equivalents is generated. In an organism with an active CO₂ fixation pathway such as the reductive TCA cycle, these excess reducing equivalents can be utilized to fix additional carbon. Non-naturally occurring organisms of the present invention employing an ethylene glycol biosynthetic pathway in conjunction with a carbon fixation pathway such as the reverse TCA cycle are thereby capable of achieving the maximum theoretical EG yield.

Enzyme candidates are described below.

EC class	Description	Transformation
1.1.1.a	Oxidoreductase (oxo to alcohol)	14B, 14E
2.7.1.a	Kinase	14C, 14D
3.1.3.a	Phosphatase	14C, 14D
4.1.1.a	Decarboxylase	14A

1.1.1.a

Several transformations in Figure 14 require the reduction of an aldehyde to an alcohol. The reduction of glycolaldehyde to ethylene glycol (Step 14E) is catalyzed by glycolaldehyde reductase. Step 14B is catalyzed by an enzyme with 2-oxoethyl-phosphate reductase activity. Enzymes catalyzing step 14E have been demonstrated in the literature and are described below. Enzymes catalyzing step 14B have not been identified to date. However, alcohol dehydrogenase enzymes typically accept a broad range of substrates. Exemplary oxidoreductase enzymes in EC class 1.1.1 are described below.

The iron-activated 1,2-PDO oxidoreductase (EC 1.1.1.77) *E. coli* encoded by *fucO* efficiently catalyzes the reduction of glycolaldehyde to ethylene glycol (Obradors et al., *Eur.J Biochem.* 258:207-213 (1998); Boronat et al., *J Bacteriol.* 153:134-139 (1983)). Other aldehyde reductase enzyme candidates include *alrA* from *Acinetobacter* sp. Strain M-1 encoding a medium-chain alcohol dehydrogenase for C2-C14 (Tani et al., *Appl.Environ.Microbiol.* 66:5231-5235 (2000)), ADH2 from *Saccharomyces cerevisiae* (Atsumi et al., *Nature* 451:86-89 (2008)) and the *adhA* gene product from *Zymomonas mobilis*, which was demonstrated to have activity on a number of aldehydes including formaldehyde, acetaldehyde,

propionaldehyde, butyraldehyde, and acrolein (Kinoshita et al., *Appl Microbiol Biotechnol* 22:249-254 (1985)).

Protein	GenBank ID	GI Number	Organism
fucO	AAA23825.1	146045	Escherichia coli
alrA	BAB12273.1	9967138	Acinetobacter sp. Strain M-1
ADH2	NP_014032.1	6323961	Saccharomyces cerevisiae
adhA	YP_162971.1	56552132	Zymomonas mobilis

- Other aldehyde reductase (alcohol dehydrogenase) enzymes include *alrA* encoding a medium-chain alcohol dehydrogenase for C2-C14 (Tani et al., *Appl. Environ. Microbiol.* 66:5231-5235 (2000)), *yqhD* and *fucO* from *E. coli* (Sulzenbacher et al., 342:489-502 (2004)), and *bdh* I and *bdh* II from *C. acetobutylicum* which convert butyrylaldehyde into butanol (Walter et al., 174:7149-7158 (1992)). The *adhA* gene product from *Zymomonas mobilis* has been demonstrated to have activity on a number of aldehydes including formaldehyde, acetaldehyde, propionaldehyde, butyraldehyde, and acrolein (Kinoshita et al., *Appl Microbiol Biotechnol* 22:249-254 (1985)). Additional aldehyde reductase candidates are encoded by *bdh* in *C. saccharoperbutylacetonicum* and *Cbei_1722*, *Cbei_2181* and *Cbei_2421* in *C. beijerinckii*. Aldehyde reductase gene candidates in *Saccharomyces cerevisiae* include the aldehyde reductases GRE3, ALD2-6 and HFD1, glyoxylate reductases GOR1 and YPL113C and glycerol dehydrogenase GCY1 (WO 2011/022651A1; Atsumi et al., *Nature* 451:86-89 (2008)).

Protein	GenBank ID	GI Number	Organism
alrA	BAB12273.1	9967138	Acinetobacter sp. strain M-1
ADH2	NP_014032.1	6323961	Saccharomyces cerevisiae
fucO	NP_417279.1	16130706	Escherichia coli
bdh I	NP_349892.1	15896543	Clostridium acetobutylicum
bdh II	NP_349891.1	15896542	Clostridium acetobutylicum
adhA	YP_162971.1	56552132	Zymomonas mobilis
bdh	BAF45463.1	124221917	Clostridium saccharoperbutylacetonicum
Cbei_1722	YP_001308850	150016596	Clostridium beijerinckii
Cbei_2181	YP_001309304	150017050	Clostridium beijerinckii
Cbei_2421	YP_001309535	150017281	Clostridium beijerinckii
GRE3	P38715.1	731691	Saccharomyces cerevisiae
ALD2	CAA89806.1	825575	Saccharomyces cerevisiae
ALD3	NP_013892.1	6323821	Saccharomyces cerevisiae
ALD4	NP_015019.1	6324950	Saccharomyces cerevisiae
ALD5	NP_010996.2	330443526	Saccharomyces cerevisiae
ALD6	ABX39192.1	160415767	Saccharomyces cerevisiae

Protein	GenBank ID	GI Number	Organism
HFD1	Q04458.1	2494079	Saccharomyces cerevisiae
GOR1	NP_014125.1	6324055	Saccharomyces cerevisiae
YPL113C	AAB68248.1	1163100	Saccharomyces cerevisiae
GCY1	CAA99318.1	1420317	Saccharomyces cerevisiae

Other exemplary aldehyde reductase enzyme candidates exhibit 4-hydroxybutyraldehyde reductase activity (EC 1.1.1.61). Such enzymes have been characterized in *Ralstonia eutropha* (Bravo et al., *J Forens Sci*, 49:379-387 (2004)), *Clostridium kluyveri* (Wolff et al., *Protein Expr.Purif.* 6:206-212 (1995)) and *Arabidopsis thaliana* (Breitkreuz et al., *J Biol Chem*, 278:41552-41556 (2003)). The *A. thaliana* enzyme was cloned and characterized in yeast. This enzyme also catalyzes the reduction of glutarate semialdehyde to 5-hydroxyvalerate (WO 2010/068953A2). Yet another gene is the alcohol dehydrogenase *adhI* from *Geobacillus thermoglucosidasius* (Jeon et al., *J Biotechnol* 135:127-133 (2008)). An enzyme with similar activity is the glutarate semialdehyde reductase enzyme of *Aspergillus terreus*, encoded by *ATEG_00539* (WO 2010/068953A2).

Protein	GenBank ID	GI number	Organism
4hbd	YP_726053.1	113867564	Ralstonia eutropha H16
4hbd	L21902.1	146348486	Clostridium kluyveri DSM 555
4hbd	Q94B07	75249805	Arabidopsis thaliana
adhI	AAR91477.1	40795502	Geobacillus thermoglucosidasius
ATEG_00539	XP_001210625.1	115491995	Aspergillus terreus NIH2624

2.7.1.a Kinase

The conversion of 2-hydroxyethyl-phosphate to ethylene glycol (step 14B) and 2-oxoethyl-phosphate to glycolaldehyde (step 14D) can be catalyzed by a kinase enzyme in EC class 2.7.1 operating in the dephosphorylating direction. Kinase enzyme candidates are described above in Examples II, III and V.

3.1.3.a Phosphatase

The hydrolysis of 2-hydroxyethyl-phosphate to ethylene glycol (step 14B) and 2-oxoethyl-phosphate to glycolaldehyde (step 14D) can be catalyzed by a phosphatase enzyme in EC class 3.1.3. The enzymatic hydrolysis of 2-hydroxyethyl-phosphate to ethylene glycol has been demonstrated in the literature; however, the hydrolysis of 2-oxoethyl-phosphate has not

been demonstrated to date. Exemplary phosphatase enzymes for catalyzing these reactions include phosphoglycolate phosphatase (EC 3.1.3.18), glycerol-1-phosphatase (EC 3.1.3.21), 3-phosphoglycerate phosphatase (3.1.3.38), phosphoglycerate phosphatase (EC 3.1.3.20) and alkaline phosphatase (EC 3.1.3.1).

- 5 The hydrolysis of 2-hydroxyethyl-phosphate to ethylene glycol can be catalyzed by the phosphoglycolate phosphatase of *Nicotiana tabacum* (Christeller et al, J Biol Chem 253(6):1791-8 (1978). Phosphoglycolate phosphatase enzymes from *Homo sapiens* and *Spinacia oleracea* hydrolyze a similar substrate, ethyl-phosphate, at high velocity (Rose et al, J Biol Chem 261:10996-1002 (1986)). Genes associated with these enzymes have not been
- 10 identified to date. Several acid phosphatase enzymes from *Nicotiana tabacum* have been characterized (Kaida et al, *Biochim Biophys Acta* 1625:134-40 (2003)) and are shown in the table below. Phosphoglycolate phosphatase enzymes have been characterized in *Chlamydomonas reinhardtii* (Mamedov et al, *J Biol Chem* 276:45573-79 (2001)), *E. coli* (Pellicer et al, *J Bacteriol* 185:5815-21 (2003)) and *Thermoplasma acidophilum* (Kim et al, *J*
- 15 *Biol Chem* 279:517-26 (2003)). The nitrophenyl phosphate phosphatase of *Saccharomyces cerevisiae* accepts a range of substrates including glycerol-1-phosphate, glycerol-2-phosphate and phosphoglycolate (Yakunin and Proudfoot, Private Communication 10: 1-1 (2007)).

Protein	GenBank ID	GI Number	Organism
AB084124.1:1..1413	BAC55157.1	27597239	<i>Nicotiana tabacum</i>
AB017968.1:27..1433	BAC55156.1	27597231	<i>Nicotiana tabacum</i>
AB017967.1:292..1704	BAC55155.1	27597229	<i>Nicotiana tabacum</i>
AB017966.1:44..1429	BAC55154.1	27597227	<i>Nicotiana tabacum</i>
Pgp1	AB052169	21928193	<i>Chlamydomonas reinhardtii</i>
gph	AAC76410.1	1789787	<i>Escherichia coli</i>
TA0175	NP_393653.1	16081332	<i>Thermoplasma acidophilum</i>
PHO13	X51611.1	4142	<i>Saccharomyces cerevisiae</i>

- Alkaline phosphatase (EC 3.1.3.1) hydrolyses a broad range of phosphorylated substrates to
- 20 their corresponding alcohols. These enzymes are typically secreted into the periplasm in bacteria, where they play a role in phosphate transport and metabolism. The *E. coli phoA* gene encodes a periplasmic zinc-dependent alkaline phosphatase active under conditions of phosphate starvation (Coleman *Annu. Rev. Biophys. Biomol. Struct.* 21:441-83 (1992)). This phosphatase hydrolyzes substrates similar to 2-hydroxyethyl phosphate including ethyl
- 25 phosphate, methyl phosphate, and phosphoethanolamine. Similar enzymes have been characterized in *Campylobacter jejuni* (van Mourik et al., *Microbiol.* 154:584-92 (2008)),

Saccharomyces cerevisiae (Oshima et al., *Gene* 179:171-7 (1996)) and *Staphylococcus aureus* (Shah and Blobel, *J. Bacteriol.* 94:780-1 (1967)). Enzyme engineering and/or removal of targeting sequences may be required for alkaline phosphatase enzymes to function in the cytoplasm.

Protein	GenBank ID	GI Number	Organism
phoA	NP_414917.2	49176017	Escherichia coli
phoX	ZP_01072054.1	86153851	Campylobacter jejuni
PHO8	AAA34871.1	172164	Saccharomyces cerevisiae
SaurJH1_2706	YP_001317815.1	150395140	Staphylococcus aureus

5

Glycerol phosphatase enzymes have been characterized in *Saccharomyces cerevisiae* (*GPP1* and *GPP2*), *Candida albicans* and *Dunaleilla parva* (Popp et al, *Biotechnol Bioeng* 100:497-505 (2008); Fan et al, *FEMS Microbiol Lett* 245:107-16 (2005)). The *D. parva* gene has not been identified to date. The acid phosphatase of *Xenopus laevis* was shown to catalyze the hydrolysis of a wide range of phosphorylated substrates including dihydroxyacetone phosphate (Filburn, *Arch Biochem Biophys* 159:683-93 (1973)). The gene associated with this enzyme is not known.

10

Protein	GenBank ID	GI Number	Organism
GPP1	DAA08494.1	285812595	Saccharomyces cerevisiae
GPP2	NP_010984.1	6320905	Saccharomyces cerevisiae
GPP1	XP_717809.1	68476319	Candida albicans
KLLA0C08217g	XP_452565.1	50305213	Kluyveromyces lactis
KLLA0C11143g	XP_452697.1	50305475	Kluyveromyces lactis
ANI_1_380074	XP_001392369.1	145239445	Aspergillus niger
ANI_1_444054	XP_001390913.2	317029125	Aspergillus niger

3-phosphoglycerate phosphatase (EC 3.1.3.38) catalyzes the hydrolysis of 3PG to glycerate.

15

The enzyme is found in plants and has a broad substrate range that includes phosphoenolpyruvate, ribulose-1,5-bisphosphate, dihydroxyacetone phosphate and glucose-6-phosphate (Randall et al., *Plant Physiol* 48:488-492 (1971); Randall et al., *J Biol.Chem.* 246:5510-5517 (1971)). Purified enzyme from various plant sources has been characterized but a gene has not been associated with this enzyme to date. Another enzyme with 3-phosphoglycerate phosphatase activity is the phosphoglycerate phosphatase (EC 3.1.3.20) from pig liver (Fallon et al., *Biochim.Biophys.Acta* 105:43-53 (1965)). The gene associated with this enzyme is not available.

20

4.1.1.a

An enzyme with 3-phosphohydroxypyruvate decarboxylase activity can be used to convert 3-phosphohydroxypyruvate to 2-oxoethyl-phosphate (step 14A). An enzyme with this activity has not been demonstrated in the literature to date. Keto-acid decarboxylase enzymes in the EC class 4.1.1 represent suitable candidates. Several exemplary keto acid decarboxylases are listed in the table below.

EC number	Name
4.1.1.1	Pyruvate decarboxylase
4.1.1.7	Benzoylformate decarboxylase
4.1.1.40	Hydroxypyruvate decarboxylase
4.1.1.43	Ketophenylpyruvate decarboxylase
4.1.1.71	Alpha-ketoglutarate decarboxylase
4.1.1.72	Branched chain keto-acid decarboxylase
4.1.1.74	Indolepyruvate decarboxylase
4.1.1.75	2-Ketoarginine decarboxylase
4.1.1.79	Sulfopyruvate decarboxylase
4.1.1.80	Hydroxyphenylpyruvate decarboxylase
4.1.1.82	Phosphonopyruvate decarboxylase

Particularly good enzyme candidates are 3-phosphonopyruvate decarboxylase (EC 4.1.1.82) and hydroxypyruvate decarboxylase (EC 4.1.1.40). 3-Phosphonopyruvate decarboxylase catalyzes the decarboxylation of 3-phosphonopyruvate to 2-phosphonoacetaldehyde.

Exemplary phosphonopyruvate decarboxylase enzymes are encoded by *dhpF* of *Streptomyces luridus*, *ppd* of *Streptomyces viridochromogenes*, *fom2* of *Streptomyces wedmorensis* and *bcpC* of *Streptomyces hygroscopicus* (Circello et al, *Chem Biol* 17:402-11 (2010); Blodgett et al, *FEMS Microbiol Lett* 163:149-57 (2005); Hidaka et al, *Mol Gen Genet* 249:274-80 (1995); Nakashita et al, *Biochim Biophys Acta* 1490:159-62 (2000)). The *Bacteroides fragilis* enzyme, encoded by *aepY*, also decarboxylates pyruvate and sulfopyruvate (Zhang et al, *J Biol Chem* 278:41302-8 (2003)).

Protein	GenBank ID	GI Number	Organism
dhpF	ACZ13457.1	268628095	<i>Streptomyces luridus</i>
Ppd	CAJ14045.1	68697716	<i>Streptomyces viridochromogenes</i>
Fom2	BAA32496.1	1061008	<i>Streptomyces wedmorensis</i>
aepY	AAG26466.1	11023509	<i>Bacteroides fragilis</i>

Hydroxypyruvate decarboxylase catalyzes the decarboxylation of hydroxypyruvate to glycolaldehyde. This enzyme activity is found in many mammals (Hendrick et al.,

Arch.Biochem.Biophys. 105:261-269 (1964)) and has been studied in rat mitochondria in the context of hydroxypyruvate metabolism to oxalate. However, the enzyme activity has not been associated with a gene in mammalian systems to date (Rofe et al., *Biochem.Med.Metab Biol.* 36:141-150 (1986)). Several keto-acid decarboxylase enzymes have been shown to accept hydroxypyruvate as an alternate substrate, including the *kivD* gene product of *Lactococcus lactis* (*kdcA*; GenBank AAS49166.1; GI 44921617) (de la Plaza et al., *FEMS Microbiol Lett.* 238:367-374 (2004); WO 2010/076324) and the *pdcl* gene product of *Saccharomyces cerevisiae* (GenBank P06169; GI 30923172) (Cusa et al., *J Bacteriol.* 181:7479-7484 (1999)). The *S. cerevisiae* enzyme has been extensively studied, engineered for altered activity, and functionally expressed in *E. coli* (Killenberg-Jabs et al., *Eur.J.Biochem.* 268:1698-1704 (2001); Li et al., *Biochemistry.* 38:10004-10012 (1999); ter Schure et al., *Appl.Environ.Microbiol.* 64:1303-1307 (1998)).

The decarboxylation of keto-acids is also catalyzed by pyruvate decarboxylase, benzoylformate decarboxylase, alpha-ketoglutarate decarboxylase, branched-chain alpha-ketoacid decarboxylase and phenylpyruvate decarboxylase. Pyruvate decarboxylase (PDC), also termed keto-acid decarboxylase, is a key enzyme in alcoholic fermentation, catalyzing the decarboxylation of pyruvate to acetaldehyde. The PDC1 enzyme from *Saccharomyces cerevisiae* has a broad substrate range for aliphatic 2-keto acids including 2-ketobutyrate, 2-ketovalerate, 3-hydroxypyruvate and 2-phenylpyruvate (22). This enzyme has been extensively studied, engineered for altered activity, and functionally expressed in *E. coli* (Killenberg-Jabs et al., *Eur.J.Biochem.* 268:1698-1704 (2001); Li et al., *Biochemistry.* 38:10004-10012 (1999); ter Schure et al., *Appl.Environ.Microbiol.* 64:1303-1307 (1998)). The PDC from *Zymomonas mobilis*, encoded by *pdc*, also has a broad substrate range and has been a subject of directed engineering studies to alter the affinity for different substrates (Siegert et al., *Protein Eng Des Sel* 18:345-357 (2005)). The crystal structure of this enzyme is available (Killenberg-Jabs et al., *Eur.J.Biochem.* 268:1698-1704 (2001)). Other well-characterized PDC candidates include the enzymes from *Acetobacter pasteurians* (Chandra et al., 176:443-451 (2001)) and *Kluyveromyces lactis* (Krieger et al., 269:3256-3263 (2002)).

Protein	GenBank ID	GI Number	Organism
pdc	P06672.1	118391	Zymomonas mobilis
pdcl	P06169	30923172	Saccharomyces cerevisiae
pdc	Q8L388	20385191	Acetobacter pasteurians
pdcl	Q12629	52788279	Kluyveromyces lactis

Like PDC, benzoylformate decarboxylase (EC 4.1.1.7) has a broad substrate range and has been the target of enzyme engineering studies. The enzyme from *Pseudomonas putida* has been extensively studied and crystal structures of this enzyme are available (Polovnikova et al., 42:1820-1830 (2003); Hasson et al., 37:9918-9930 (1998)). Site-directed mutagenesis of two residues in the active site of the *Pseudomonas putida* enzyme altered the affinity (Km) of naturally and non-naturally occurring substrates (Siegert et al., *Protein Eng Des Sel* 18:345-357 (2005)). The properties of this enzyme have been further modified by directed engineering (Lingen et al., *Chembiochem.* 4:721-726 (2003); Lingen et al., *Protein Eng* 15:585-593 (2002)). The enzyme from *Pseudomonas aeruginosa*, encoded by *mdlC*, has also been characterized experimentally (Barrowman et al., 34:57-60 (1986)). Additional gene candidates from *Pseudomonas stutzeri*, *Pseudomonas fluorescens* and other organisms can be inferred by sequence homology or identified using a growth selection system developed in *Pseudomonas putida* (Henning et al., *Appl. Environ. Microbiol.* 72:7510-7517 (2006)).

Protein	GenBank ID	GI Number	Organism
mdlC	P20906.2	3915757	<i>Pseudomonas putida</i>
mdlC	Q9HUR2.1	81539678	<i>Pseudomonas aeruginosa</i>
dpgB	ABN80423.1	126202187	<i>Pseudomonas stutzeri</i>
ilvB-1	YP_260581.1	70730840	<i>Pseudomonas fluorescens</i>

A third enzyme capable of decarboxylating 2-oxoacids is alpha-ketoglutarate decarboxylase (KGD, EC 4.1.1.71). The substrate range of this class of enzymes has not been studied to date. An exemplarily KDC is encoded by *kad* in *Mycobacterium tuberculosis* (Tian et al., *PNAS* 102:10670-10675 (2005)). KDC enzyme activity has also been detected in several species of rhizobia including *Bradyrhizobium japonicum* and *Mesorhizobium loti* (Green et al., *J Bacteriol* 182:2838-2844 (2000)). Although the KDC-encoding gene(s) have not been isolated in these organisms, the genome sequences are available and several genes in each genome are annotated as putative KDCs. A KDC from *Euglena gracilis* has also been characterized but the gene associated with this activity has not been identified to date (Shigeoka et al., *Arch. Biochem. Biophys.* 288:22-28 (1991)). The first twenty amino acids starting from the N-terminus were sequenced MTYKAPVKDVKFLLDKVFKV (Shigeoka and Nakano, *Arch. Biochem. Biophys.* 288:22-28 (1991)). The gene could be identified by testing candidate genes containing this N-terminal sequence for KDC activity. A novel class of AKG decarboxylase enzymes has recently been identified in cyanobacteria such as *Synechococcus* sp. *PCC 7002* and homologs (Zhang and Bryant, *Science* 334:1551-3 (2011)).

Protein	GenBank ID	GI Number	Organism
kgd	O50463.4	160395583	Mycobacterium tuberculosis
kgd	NP_767092.1	27375563	Bradyrhizobium japonicum USDA110
kgd	NP_105204.1	13473636	Mesorhizobium loti
ilvB	ACB00744.1	169887030	Synechococcus sp. PCC 7002

A fourth candidate enzyme for catalyzing this reaction is branched chain alpha-ketoacid decarboxylase (BCKA). This class of enzyme has been shown to act on a variety of compounds varying in chain length from 3 to 6 carbons (Oku et al., *J Biol Chem.* 263:18386-18396 (1988); Smit et al., *Appl Environ Microbiol* 71:303-311 (2005)). The enzyme in *Lactococcus lactis* has been characterized on a variety of branched and linear substrates including 2-oxobutanoate, 2-oxohexanoate, 2-oxopentanoate, 3-methyl-2-oxobutanoate, 4-methyl-2-oxobutanoate and isocaproate (Smit et al., *Appl Environ Microbiol* 71:303-311 (2005)). The enzyme has been structurally characterized (Berg et al., *Science.* 318:1782-1786 (2007)). Sequence alignments between the *Lactococcus lactis* enzyme and the pyruvate decarboxylase of *Zymomonas mobilis* indicate that the catalytic and substrate recognition residues are nearly identical (Siegert et al., *Protein Eng Des Sel* 18:345-357 (2005)), so this enzyme would be a promising candidate for directed engineering. Several ketoacid decarboxylases of *Saccharomyces cerevisiae* catalyze the decarboxylation of branched substrates, including ARO10, PDC6, PDC5, PDC1 and THI3 (Dickenson et al., *J Biol Chem* 275:10937-42 (2000)). Yet another BCKAD enzyme is encoded by rv0853c of *Mycobacterium tuberculosis* (Werther et al., *J Biol Chem* 283:5344-54 (2008)). This enzyme is subject to allosteric activation by alpha-ketoacid substrates. Decarboxylation of alpha-ketoglutarate by a BCKA was detected in *Bacillus subtilis*; however, this activity was low (5%) relative to activity on other branched-chain substrates (Oku and Kaneda, *J Biol Chem.* 263:18386-18396 (1988)) and the gene encoding this enzyme has not been identified to date. Additional BCKA gene candidates can be identified by homology to the *Lactococcus lactis* protein sequence. Many of the high-scoring BLASTp hits to this enzyme are annotated as indolepyruvate decarboxylases (EC 4.1.1.74). Indolepyruvate decarboxylase (IPDA) is an enzyme that catalyzes the decarboxylation of indolepyruvate to indoleacetaldehyde in plants and plant bacteria. Recombinant branched chain alpha-keto acid decarboxylase enzymes derived from the E1 subunits of the mitochondrial branched-chain keto acid dehydrogenase complex from *Homo sapiens* and *Bos taurus* have been cloned and functionally expressed in *E. coli* (Davie et al., *J.Biol.Chem.* 267:16601-16606 (1992); Wynn et al., *J.Biol.Chem.* 267:12400-12403 (1992); Wynn et al., *J.Biol.Chem.* 267:1881-1887 (1992)). In these studies,

the authors found that co-expression of chaperonins GroEL and GroES enhanced the specific activity of the decarboxylase by 500-fold (Wynn et al., *J.Biol.Chem.* 267:12400-12403 (1992)). These enzymes are composed of two alpha and two beta subunits.

Protein	GenBank ID	GI Number	Organism
kdcA	AAS49166.1	44921617	Lactococcus lactis
PDC6	NP_010366.1	6320286	Saccharomyces cerevisiae
PDC5	NP_013235.1	6323163	Saccharomyces cerevisiae
PDC1	P06169	30923172	Saccharomyces cerevisiae
ARO10	NP_010668.1	6320588	Saccharomyces cerevisiae
THI3	NP_010203.1	6320123	Saccharomyces cerevisiae
rv0853c	O53865.1	81343167	Mycobacterium tuberculosis
BCKDHB	NP_898871.1	34101272	Homo sapiens
BCKDHA	NP_000700.1	11386135	Homo sapiens
BCKDHB	P21839	115502434	Bos taurus
BCKDHA	P11178	129030	Bos taurus

- 5 Phenylpyruvate decarboxylase enzymes (EC 4.1.1.43) also accept a range of substrates and are suitable candidates. Other phenylpyruvate decarboxylase enzymes have been characterized in *Azospirillum brasilense* (Spaepen et al, *J Bacteriol* 189:7626-33 (2007)), *Saccharomyces cerevisiae* (Kneen et al, *FEBS J* 278 (11) 1842-53 (2011)), *Thauera aromatica*, *Nocardia* sp. 239 and others. The *S. cerevisiae* enzyme catalyzes the
- 10 decarboxylation of numerous aromatic and aliphatic keto-acids. A phenylpyruvate decarboxylase from *Acinetobacter* sp. has similar properties; however a gene has not been associated with this enzyme to date (Fujioka et al, *Methods Enzymol* 17:585-96 (1970)). Another candidate is the indole-3-pyruvate decarboxylase from *Enterobacter cloacae* (Schultz et al, *Eur J Biochem* 270:2312-21 (2003)).

Protein	GenBank ID	GI Number	Organism
ARO10	NP_010668.1	6320588	Saccharomyces cerevisiae
ipdC	CCC97185.1	356876418	Azospirillum brasilense
idpC	P23234.1	118333	Enterobacter cloacae

15

Throughout this application various publications have been referenced. The disclosures of these publications in their entireties, including GenBank and GI number publications, are hereby incorporated by reference in this application in order to more fully describe the state of the art to which this invention pertains. Although the invention has been described with

20 reference to the examples provided above, it should be understood that various modifications can be made without departing from the spirit of the invention.

What is claimed is:

1. A non-naturally occurring microbial organism, comprising a microbial organism having an ethylene glycol pathway comprising at least one exogenous nucleic acid encoding an ethylene glycol pathway enzyme expressed in a sufficient amount to produce ethylene glycol; said non-naturally occurring microbial organism further comprising:

(i) a reductive TCA pathway comprising at least one exogenous nucleic acid encoding a reductive TCA pathway enzyme, wherein said at least one exogenous nucleic acid is selected from an ATP-citrate lyase, a citrate lyase, a citryl-CoA synthetase, a citryl-CoA lyase, a fumarate reductase, and an alpha-ketoglutarate:ferredoxin oxidoreductase;

(ii) a reductive TCA pathway comprising at least one exogenous nucleic acid encoding a reductive TCA pathway enzyme, wherein said at least one exogenous nucleic acid is selected from a pyruvate:ferredoxin oxidoreductase, a phosphoenolpyruvate carboxylase, a phosphoenolpyruvate carboxykinase, a CO dehydrogenase, and an H₂ hydrogenase; or

(iii) at least one exogenous nucleic acid encodes an enzyme selected from a CO dehydrogenase, an H₂ hydrogenase, and combinations thereof;

wherein said ethylene glycol pathway comprises a pathway selected from:

(A) a serine aminotransferase or a serine oxidoreductase (deaminating); a hydroxypyruvate decarboxylase, and a glycolaldehyde reductase;

(B) a serine aminotransferase or a serine oxidoreductase (deaminating); a hydroxypyruvate reductase, and a glycerate decarboxylase;

(C) a serine decarboxylase; an ethanolamine aminotransferase or an ethanolamine oxidoreductase (deaminating), and a glycolaldehyde reductase;

(D) a hydroxypyruvate reductase; a hydroxypyruvate decarboxylase, and a glycolaldehyde reductase;

(E) a glycerate decarboxylase;

(F) glycerate dehydrogenase, a hydroxypyruvate isomerase, a hydroxypyruvate decarboxylase and a glycolaldehyde reductase;

(G) a glyoxylate carboligase, a hydroxypyruvate isomerase, a hydroxypyruvate decarboxylase and a glycolaldehyde reductase;

(H) a glyoxylate reductase; a glycolyl-CoA transferase or a glycolyl-CoA synthetase; a glycolyl-CoA reductase (aldehyde forming), and a glycolaldehyde reductase;

5 (I) a glyoxylate reductase; a glycolate reductase, and a glycolaldehyde reductase;

(J) a glyoxylate reductase; a glycolyl-CoA transferase or a glycolyl-CoA synthetase, and a glycolyl-CoA reductase (alcohol forming);

(K) a glyoxylate reductase, a glycolate kinase, a phosphotransglycolylase, glycolyl-CoA reductase (aldehyde forming) and a glycolaldehyde reductase;

10 (L) a glyoxylate reductase, a glycolate kinase, a phosphotransglycolylase and a glycolyl-CoA reductase (alcohol forming);

(M) a glyoxylate reductase, glycolate kinase, a glycolylphosphate reductase and a glycolaldehyde reductase;

15 (N) a 3-phosphohydroxypyruvate decarboxylase, a 2-oxoethyl-phosphate reductase, and a 2-hydroxyethyl-phosphate phosphatase or an ethylene glycol kinase; and

(O) a 3-phosphohydroxypyruvate decarboxylase, a 2-oxoethyl-phosphate phosphatase or a glycolaldehyde kinase, and a glycolaldehyde reductase.

2. The non-naturally occurring microbial organism of claim 1, wherein the ethylene glycol pathway of (D), (E) and/or (F) further comprises a 3-phosphoglycerate phosphatase, a
20 glycerate kinase, a 2-phosphoglycerate phosphatase, a glycerate-2-kinase, or a glyceraldehyde dehydrogenase.

3. The non-naturally occurring microbial organism of claim 1 or claim 2, wherein said microbial organism comprising (i) further comprises an exogenous nucleic acid encoding an enzyme selected from a pyruvate:ferredoxin oxidoreductase, an aconitase, an isocitrate
25 dehydrogenase, a succinyl-CoA synthetase, a succinyl-CoA transferase, a fumarase, a malate dehydrogenase, an acetate kinase, a phosphotransacetylase, an acetyl-CoA synthetase, an NAD(P)H:ferredoxin oxidoreductase, ferredoxin, and combinations thereof.

4. The non-naturally occurring microbial organism of claim 1 or claim 2, wherein said microbial organism comprising (ii) further comprises an exogenous nucleic acid encoding an enzyme selected from an aconitase, an isocitrate dehydrogenase, a succinyl-CoA synthetase, a succinyl-CoA transferase, a fumarase, a malate dehydrogenase, and

5 combinations thereof.

5. The non-naturally occurring microbial organism of claim 1, wherein said microbial organism comprises two, three, four, five, six or seven exogenous nucleic acids each encoding a ethylene glycol pathway enzyme.

6. The non-naturally occurring microbial organism of claim 5, wherein said
10 microbial organism comprises exogenous nucleic acids encoding each of the enzymes selected from:

(A) a serine aminotransferase or a serine oxidoreductase (deaminating); a hydroxypyruvate decarboxylase, and a glycolaldehyde reductase;

(B) a serine aminotransferase or a serine oxidoreductase (deaminating); a
15 hydroxypyruvate reductase, and a glycerate decarboxylase;

(C) a serine decarboxylase; an ethanolamine aminotransferase or an ethanolamine oxidoreductase (deaminating), and a glycolaldehyde reductase;

(D) a hydroxypyruvate reductase; a hydroxypyruvate decarboxylase, and a glycolaldehyde reductase;

20 (E) a glycerate decarboxylase;

(F) glycerate dehydrogenase, a hydroxypyruvate isomerase, a hydroxypyruvate decarboxylase and a glycolaldehyde reductase;

(G) a glyoxylate carboligase, a hydroxypyruvate isomerase, a hydroxypyruvate decarboxylase and a glycolaldehyde reductase;

25 (H) a glyoxylate reductase; a glycolyl-CoA transferase or a glycolyl-CoA synthetase; a glycolyl-CoA reductase (aldehyde forming), and a glycolaldehyde reductase;

(I) a glyoxylate reductase; a glycolate reductase, and a glycolaldehyde reductase;

(J) a glyoxylate reductase; a glycolyl-CoA transferase or a glycolyl-CoA synthetase, and a glycolyl-CoA reductase (alcohol forming);

(K) a glyoxylate reductase, a glycolate kinase, a phosphotransglycolylase, glycolyl-CoA reductase (aldehyde forming) and a glycolaldehyde reductase;

5 (L) a glyoxylate reductase, a glycolate kinase, a phosphotransglycolylase and a glycolyl-CoA reductase (alcohol forming);

(M) a glyoxylate reductase, glycolate kinase, a glycolylphosphate reductase and a glycolaldehyde reductase;

10 (N) a 3-phosphohydroxypyruvate decarboxylase, a 2-oxoethyl-phosphate reductase, and a 2-hydroxyethyl-phosphate phosphatase or an ethylene glycol kinase; and

(O) a 3-phosphohydroxypyruvate decarboxylase, a 2-oxoethyl-phosphate phosphatase or a glycolaldehyde kinase, and a glycolaldehyde reductase.

15 6. The non-naturally occurring microbial organism of claim 1, wherein said microbial organism comprises two, three, four or five exogenous nucleic acids each encoding enzymes of (i), (ii) or (iii).

7. The non-naturally occurring microbial organism of claim 6, wherein said microbial organism comprising (i) comprises three exogenous nucleic acids encoding an ATP-citrate lyase or a citrate lyase; a fumarate reductase; and an alpha-ketoglutarate:ferredoxin oxidoreductase;

20 wherein said microbial organism comprising (ii) comprises four exogenous nucleic acids encoding a pyruvate:ferredoxin oxidoreductase; a phosphoenolpyruvate carboxylase or a phosphoenolpyruvate carboxykinase; a CO dehydrogenase; and an H₂ hydrogenase; or

wherein said microbial organism comprising (iii) comprises two exogenous nucleic acids encoding CO dehydrogenase and H₂ hydrogenase.

25 8. The non-naturally occurring microbial organism of claim 1, wherein said at least one exogenous nucleic acid is a heterologous nucleic acid.

9. The non-naturally occurring microbial organism of claim 1, wherein said non-naturally occurring microbial organism is in a substantially anaerobic culture medium.

10. A non-naturally occurring microbial organism, said microbial organism having an ethylene glycol pathway and comprising at least one exogenous nucleic acid encoding an ethylene glycol pathway enzyme expressed in a sufficient amount to produce ethylene glycol, wherein said ethylene glycol pathway comprises a pathway selected from:

(A) a 3-phosphohydroxypyruvate decarboxylase, a 2-oxoethyl-phosphate reductase, and a 2-hydroxyethyl-phosphate phosphatase or an ethylene glycol kinase; and

(B) a 3-phosphohydroxypyruvate decarboxylase, a 2-oxoethyl-phosphate phosphatase or a glycolaldehyde kinase, and a glycolaldehyde reductase.

11. The non-naturally occurring microbial organism of claim 10, wherein said microbial organism comprises two or three exogenous nucleic acids each encoding an ethylene glycol pathway enzyme.

12. The non-naturally occurring microbial organism of claim 10, wherein said microbial organism comprises exogenous nucleic acids encoding each of the enzymes of at least one of the pathways selected from (A) and (B).

13. The non-naturally occurring microbial organism of claim 10, wherein said at least one exogenous nucleic acid is a heterologous nucleic acid.

14. The non-naturally occurring microbial organism of claim 10, wherein said non-naturally occurring microbial organism is in a substantially anaerobic culture medium.

15. A method for producing ethylene glycol, comprising culturing the non-naturally occurring microbial organism of any of claims 1-14 under conditions and for a sufficient period of time to produce ethylene glycol.

1/20

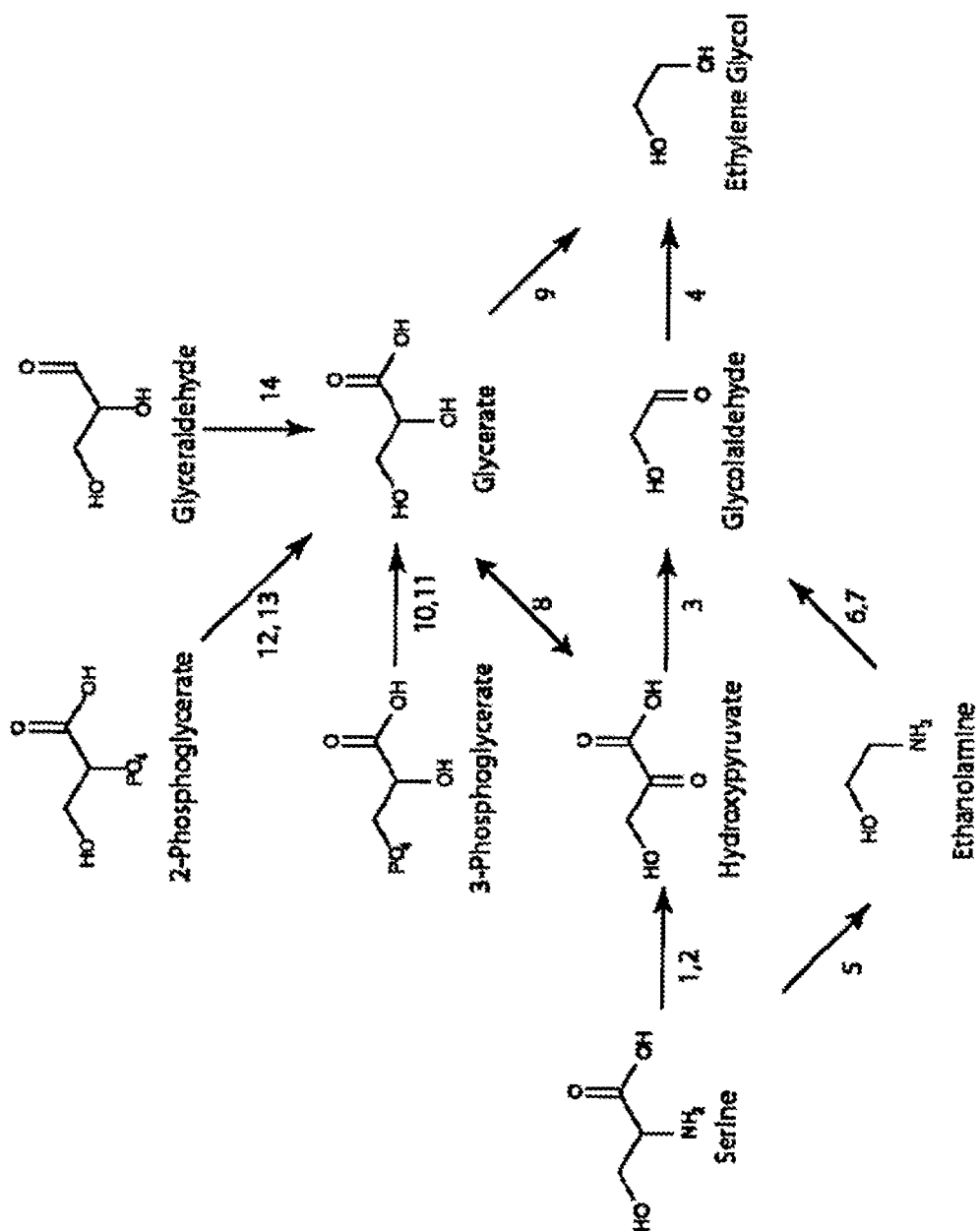


Figure 1

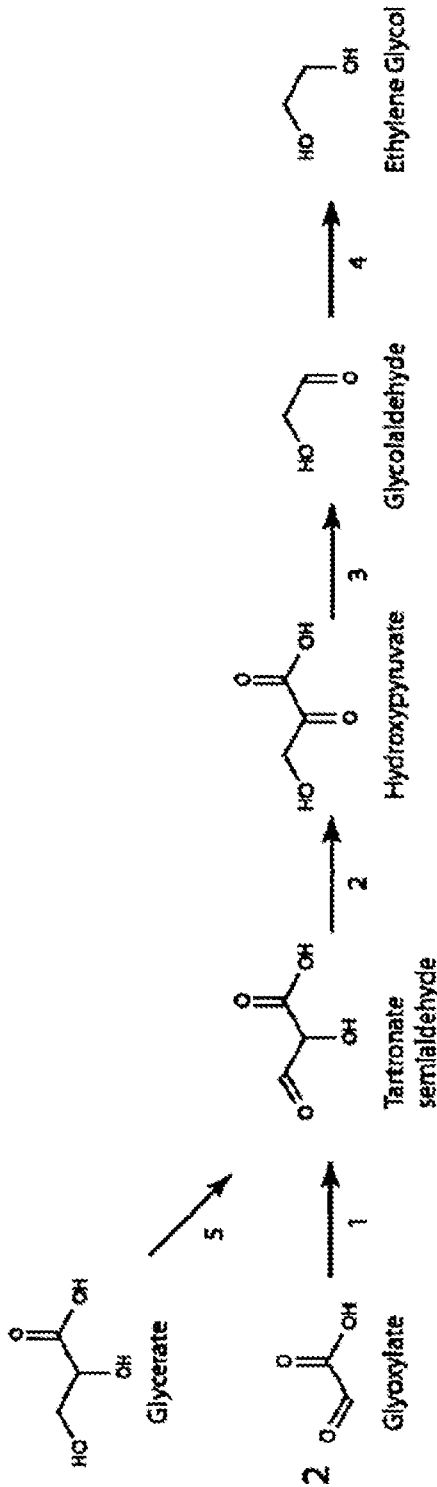


Figure 2

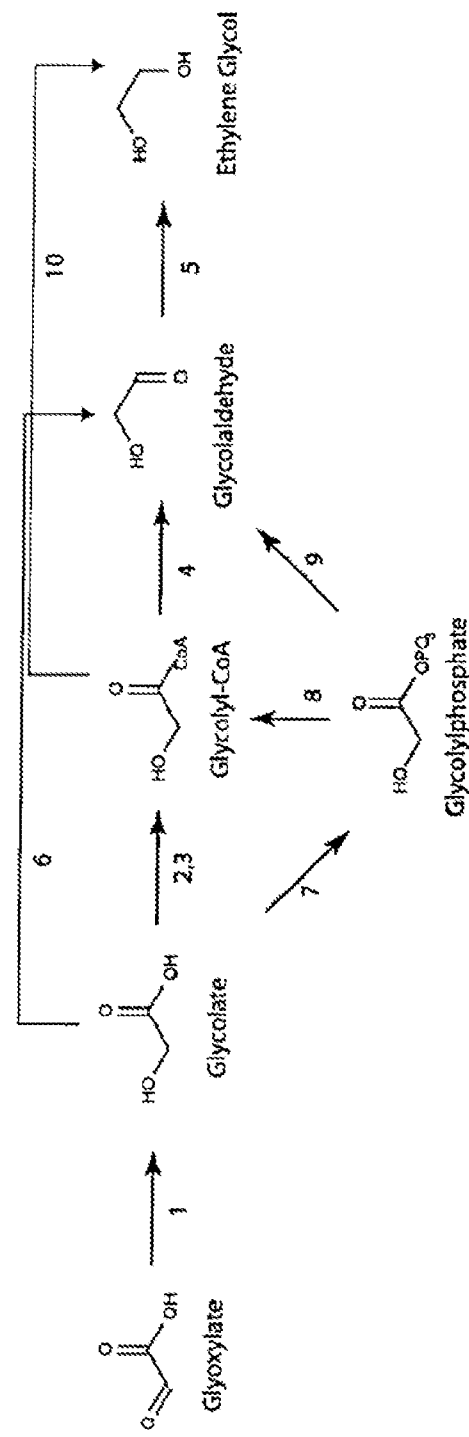


Figure 3

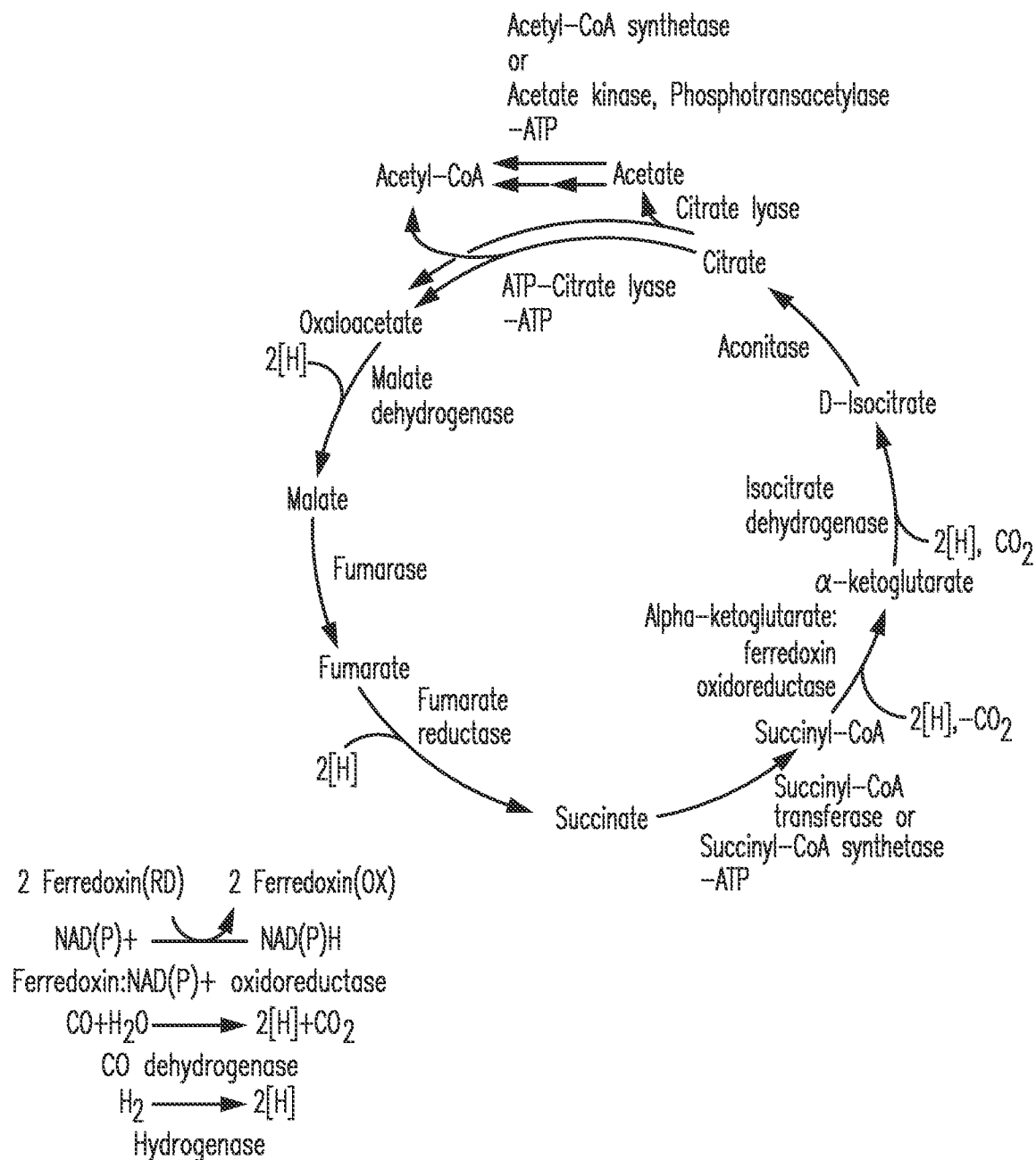


FIG. 4

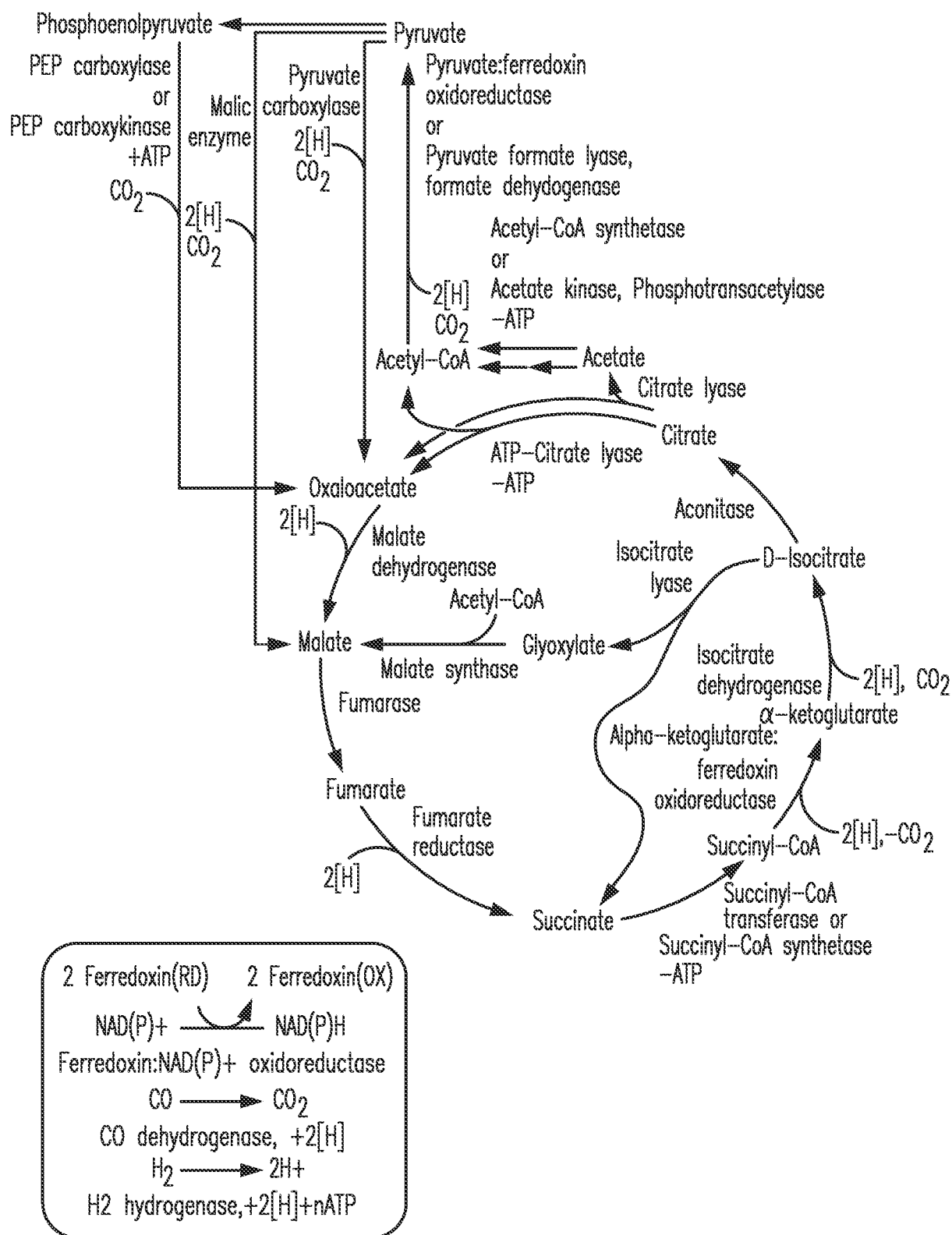


FIG. 5

6/20

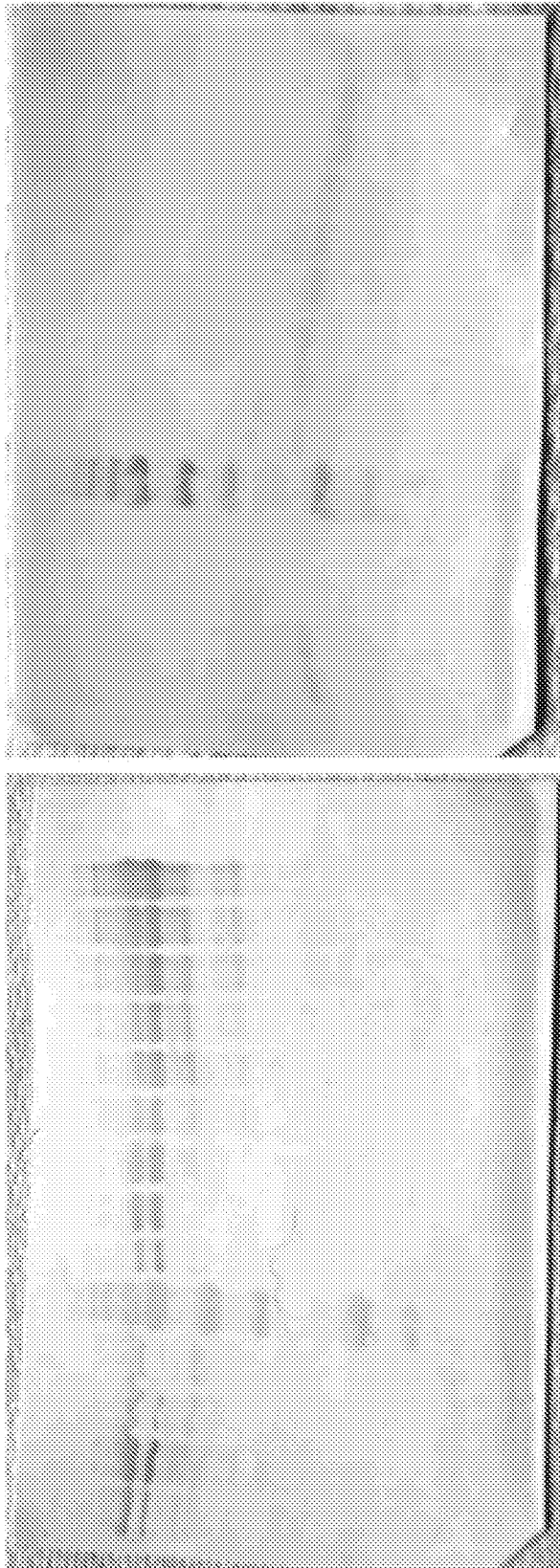


Figure 6

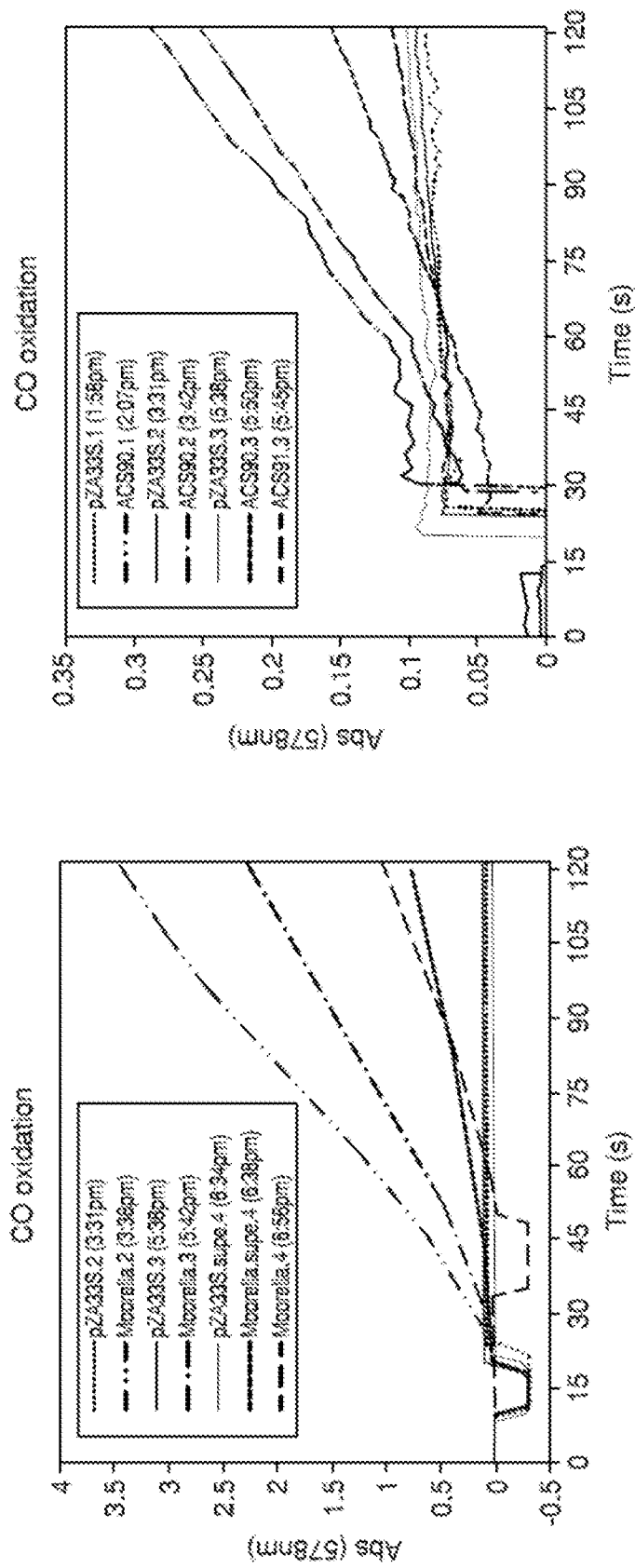


Figure 7

ATGGCAGTGGATTACCGGATGAGCGGCTACAGCGCCGCATTGCACAGTTGTTTGCAGAAGATG
AGCAGGTCAAGGCCGCACGTCCGCTCGAAGCGGTGAGCGCGCCGGTGAGCGCGCCCGGTATGCG
GCTGGCGCAGATCGCCGCCACTGTTATGGCGGGTTACGCGACCGCCCGGCCCGCGGCAGCGT
GCGTTTCGAACCTGAACACCGACGACGCGACGGGCCGCACCTCGCTGCGGTTACTTCCCGATTTCG
AGACCATCACCTATCGCGAACTGTGGCAGCGAGTCGGCGAGGTTGCCCGGCCCTGGCATCATGA
TCCCGAGAACCCCTTGC GCGCAGGTGATTCGTGCGCCCTGCTCGGCTTCACCAGCATCGACTAC
GCCACCCTCGACCTGGCCGATATCCACCTCGGCGCGGTTACCGTGCCGTTGCAGGCCAGCGCGG
CGGTGTCCAGCTGATCGCTATCCTCACCGAGACTTCGCCCGCGGCTGCTCGCCTCGACCCCGGA
GCACCTCGATGCGGCGGTGAGTGCTACTCGCGGGCACCACACCGGAACGACTGGTGGTCTTC
GACTACCACCCCGAGGACGACGACGACGAGCGTGCGGCCCTTGAATCCGCCCGCGCGCCCTTGCCG
ACGCGGGCAGCTTGGTGATCGTTCGAAACGCTCGATGCCGTGCGTGCCCGGGGCCGCGACTTACC
GGCCCGCGCCACTGTTTCGTTCCCGACACCGACGACGACCGCGCTGGCCCTGCTGATCTACACCTCC
GGCAGCACCGGAACGCCGAAGGGCGCGATGTACACCAATCGGTTGGCCCGCCACGATGTGGCAGG
GGAACCTCGATGCTGCAGGGGAACTCGCAACGGGTGCGGATCAATCTCAACTACATGCCGATGAG
CCACATCGCCGGTTCGCATATCGCTGTTTCGGCGTGCTCGCTCGCGGTGGCACCGCATACTTCGCG
GCCAAGAGCGACATGTGACACTGTTTCGAAGACATCGGCTTGGTACGTCCACCGAGATCTTCT
TCGTCCCGCGCGTGTGCGACATGGTCTTCCAGCGCTATCAGAGCGAGCTGGACCGGCGCTCGGT
GGCGGGCGCCGACCTGGACACGCTCGATCGGGAAGTGAAAGCCGACCTCCGGCAGAACTACCTC
GGTGGGCGCTTCTGGTGGCGGTGCTCGGCAGCGCGCCGCTGGCCGCGGAGATGAAGACGTTCA
TGGAGTCCGTCTCTCGATCTGCCACTGCACGACCGGTACGGGTGCGACCGAGGCGGGCGCAAGCGT
GCTGCTCGACAACAGATCCAGCGGCCCGCGGTGCTCGATTACAAGCTCGTCGACGTGCCCGAA
CTGGGTTACTTCCGCACCGACCGGCCGATCCGCGCGGTGAGCTGTTGTTGAAGGCGGAGACCA
CGATTCCGGGCTACTACAAGCGGCCCGAGGTACCGCGGAGATCTTCGACGAGGACGGCTTCTA
CAAGACCGGCGATATCGTGCCGAGCTCGAGCACGATCGGCTGGTCTATGTCGACCGTCGCAAC
AATGTGCTCAAACCTGTGCGAGGGCGAGTTCGTGACCGTCCGCCATCTCGAGGCGGTGTTTCGCCA
GCAGCCCGCTGATCCGGCAGATCTTCATCTACGGCAGCAGCGAACGTTCTATCTGCTCGCGGT
GATCGTCCCCACCGACGACGCGCTGCGCGGCCCGGACACCGCCACCTTGAAATCGGCACCTGGCC
GAATCGATTACGCGCATCGCCAAGGACGCGAACCTGCAGCCCTACGAGATTCCGCGCGGATTTC
TGATCGAGACCGAGCCGTTCAACATCGCCAACGGACTGCTCTCCGGCATCGCGAAGCTGCTGCG
CCCCAATCTGAAGGAACGCTACGGCGCTCAGCTGGAGCAGATGTACACCGATCTCGCGACAGGC
CAGGCCGATGAGCTGCTCGCCCTGCGCCGCGAAGCCGCGACCTGCCGGTGTCTCGAAACCGTCA
GCCGGGCAGCGAAAGCGATGCTCGGCGTCGCTCCGCCGATATGCGTCCCGACGCGCACTTCAC
CGACCTGGGCGGCGATTCCCTTTCCGCGCTGTGCTTCTCGAACCTGCTGCACGAGATCTTCGGG
GTCGAGGTGCCGTTGGGTGTGCTCGTCAGCCCGCGGAACGAGCTGCGCGATCTGGCGAATTACA
TTGAGGCGGAACGCAACTCGGGCGCGAAGCGTCCACCTTCACCTCGGTGCACGGCGGCGGTTTC
CGAGATCCGCGCCCGCGATCTGACCTCGACAAGTTTCATCGATGCCCGCACCTTGCCCGCCGCC
GACAGCATTCGCGACGCGCCGTTGCCAGCGCAGACGGTGCTGCTGACCGGCGCGAACGGCTACC
TCGGCCGTTCTCTGTGCTGGAATGGCTGGAGCGGCTGGACAAGACGGGTGGCACGCTGATCTG
CGTCGTGCGCGGTAGTGACGCGGCCCGCGCCCGTAAACGGCTGGACTCGGCGTTTCGACAGCGGC
GATCCCGGCCTGCTCGAGCACTACCAGCAACTGGCCGCGACGACCTTGAAGTCTTCGCCGGTG
ATATCGGCGACCCGAATCTCGGTCTGGACGACGCGACTTGGCAGCGGTTGGCCGAAACCGTCGA
CCTGATCGTCCATCCCGCCGCGTTGGTCAACCACGTCTTCCCTACACCCAGCTGTTGGCCCC

FIGURE 8A

9/20

AATGTCGTCGGCACCGCCGAAATCGTCCGGTTGGCGATCACGGCGCGCGCAAGCCGGTCACCT
 ACCTGTCGACCGTCGGAGTGGCCGACCAGGTGACCCGGCGGAGTATCAGGAGGACAGCGACGT
 CCGCGAGATGAGCGCGGTGCGCGTCGTGCGCGAGAGTTACGCCAACGGCTACGGCAACAGCAAG
 TGGGCGGGGGAGGTCTTGCTGCGCGAAGCACACGATCTGTGTGGCTTGCCGGTCGCGGTGTTCC
 GTTCGGACATGATCCTGGCGCACAGCCGGTACGCGGGTCAGCTCAACGTCCAGGACGTGTTTAC
 CCGGCTGATCCTCAGCCTGGTCGCCACCGGCATCGCGCCGTA CTGTTCTACCGAACCGACGCG
 GACGGCAACCGCGAGCGGGCCCACTATGACGGCTTGCCGGCGGACTTCACGGCGGCGGCGATCA
 CCGCGCTCGGCATCCAAGCCACCGAAGGCTTCCGGACCTACGACGTGCTCAATCCGTACGACGA
 TGGCATCTCCCTCGATGAATTCGTGCGACTGGCTCGTGAATCCGGCCACCCGATCCAGCGCATC
 ACCGACTACAGCGACTGGTTCCACCGTTTCGAGACGGCGATCCGCGCGCTGCCGGAAAAGCAAC
 GCCAGGCCTCGGTGCTGCCGTTGCTGGACGCCCTACCGCAACCCCTGCCCGGCGGTCCGCGCGCGC
 GATACTCCCGCCCAAGGAGTTCCAAGCGCGGTGCAAAACAGCCAAAATCGGTCCGGAACAGGAC
 ATCCCGCATTGTCCGCGCCACTGATCGATAAGTACGTACGCGATCTGGAAGTGTTCAGCTGC
 TCTAA

FIGURE 8A CONT.

mavdspderlqrriaqlfaedeqvkaarpleavsaavsapgmrlaqlaatvmagyadrpaagqr
 afelntddatgrtslrlprfetityrelwqrvgevaawhhdpenplragdfvallgftsdy
 atldladihlgaavtvpqlqasaavsqliailtetsprllastpehladaavecllagttperlvvf
 dyhpeddddqraafesarrrladagslvivetldavrargrdlpaaplfpvptdddplalliys
 gstgtgpkamytnrlaatmwqgnsmllqgnsqrvginlnympmshiagrislfgvlarggtayfa
 aksdmstlfediglvrpteiffvprvcdmvfqrygseldrrsvagadldtldrevkadlrqnyl
 ggrflvavvgsaplaaemktfmesvldlplhdgygsteagavlldnqigrppvldyklvdvpe
 lgyfrtdrphprgelllkaettipgyykrpevtaeifdedgfyktgdivaelehdrlvyvdrn
 nvlklsqgefvtvahleavfassplirqifiygssersyllavivptddalrgrdtatlksala
 esiqriakdanlqpyeiprdflietepftiangllsgiakllrpnlkerygaqleqmytdlatg
 qadellalrreaadlpvletvsraakamlgvasadmrpdahftdlggdsalsalsfsnllheifg
 vevpvgvvspanelrdlanyiearnsgakrptftsvhgggseiraadltldkfidartlaaa
 dsiphapvpagtvlltgangylgrflclewlerldktggtlicvvrgsdaaaarkrldsafds
 dpgllehyqqlaartlevlagdigdpnlglldatwqrlaetvdlivhpaalvnhvlpvtqlfgp
 nvvgtaeivrlaitarrkpvtylstvgvadqvdpaeyqedsdvremsavrsvresyangygnsk
 wagevllreahdlcglpvavfredmilahsryagqlnvqdvftrllslvatgiapysfyrt
 tdysdwfhrfetairalpekqrqasvplldayrnpavrgailpakefqaavqtakigpeq
 iphlsaplidkyvsdlellql1*

FIGURE 8B

10/20

FIGURE 9A

ATGATTGAAACCAATTCTGCCTGCAGGCGTTGAAAGCGCAGAACTGCTGGAATATCCGGAAGATC
TGAAAGCACATCCGGCAGAAGAACATCTGATTGCCAAAAGCGTTGAAAAACGTCGTCGTGATTT
TATTGGTGCACGTCATTGTGCACGTCTGGCACTGGCAGAACTGGGTGAACCTCCGGTTGCAATT
GGTAAAGGTGAACGTGGTGCACCGATTTGGCCTCGTGGTGGTGGTGGTAGCCTGACCCATTGTG
ATGGTTATCGTGCAGCAGCAGTTCACATAAAATGCGCTTTCGCAGCATTGGTATTGATGCAGA
ACCGCATGCAACCCTGCCGGAAGGTGTTCTGGATAGCGTTAGCCTGCCGCCGGAACGTGAATGG
CTGAAAACCACCGATAGCGCACTGCATCTGGATCGTCTGCTGTTTTGTGCAAAGAAGCCACCT
ATAAAGCCTGGTGGCCGCTGACAGCACGTTGGCTGGGTTTTGAAGAAGCCCATATTACCTTTGA
AATTGAAGATGGTAGCGCAGATAGCGTAATGGCACCTTTCATAGCGAACTGCTGGTTCCGGGT
CAGACCAATGATGGTGGTACACCGCTGCTGAGCTTTGATGGTCGTTGGCTGATTGCAGATGGTT
TTATTCTGACCGCAATTGCCTATGCCTAA

FIGURE 9B

mietilpagvesaelleypedlkahpaeenhliaksvekrerrdfigarhcarlalaaelgeppvai
gkgergapiwprgvvgslthcdgyraaavahkmrfrsigidaephatlpegvldsvslpperew
lkttdsalhldrllfcakeatykawwpltarwlgfeeahitfeiedgsadsgngtfhsellvpg
qtn dggtpllsfdgrwliadgfiltaiaya*

[illegible]

12/20

FIGURE 10B

MTSDVHDATDGVTTETALDDEQSTRRIAELYATDPEFAAAAPLPAVVDAAHKPGRLAEILQTLFTCYGDRPALGYRA
RELATDECGRTVTRELLPREFTLTYAQVMSRVQAVAAALRHNFAPQIYPGDAVATIGFASPDYLTLDLVCAYLGLVSV
PLQHNAPVSR LAPILAEVEPRILITVSAEYLDLAVESVRDVNSVSQLVVFDDHHEVDDHRDALARAREQLAGKGLAVT
TLDAIADGAGLPAEPIYTADHDQRLAMILYTSGSTGAPKGMYTEAMVARLWTMSFITGDPPTVINVMPLNHLG
GRIPISTAVQNGCSTSYFVPESDMSTLFEDLALVEPTLGLVPRVADMLYQHHLATVDRLVTQCADELTAEKQAGAE
REQVLGCRVITGEVSTAPLAAEMRAFLDITLGAHIVDGYLTETGAVTRDGVIVRPPVIDYKLIDVPELGYFESTDKP
YPRGELLVRSQTLTPGYKKRPEVTASVFDRLDGYHTGDVMAETA PDHLVYVDRENNVLKLAQCEFFVAVANLEAVFSG
AALVRQIPVYGNRSRFLAVVPTPEALEQYDEPAALKAALADSLQRTARDAELQSYEVPADFI VETETPFSAAANGLL
SGVGKLLRPNLKDRYQORLEQMYADIAATQANQLRELHRAAATQPV IDTLTQAAATILGTGSEVASDAHFTDLGGDS
LSALTLSNLLSDFRGFEVPVGTIVNPA TNLAQLAQHIEAQR TAGDRRPSFTTVHGADATEIRASELTLDKFI DAETL
RAAPGLPKVTTTEPRITVLLSGANGWLGREFLTQWLERLAPVGGTLITIVRGRDDAAARARLTQAYDTPPELSRRFAEL
ADRHLRVVAGDIGDPNLGLTPEIWHRLAAEVDLVVHPAALVNHVLPYRQLFGPNVVGTA EVIKLALTERIKPVTYLS
TVSVAMGIPDFEEDGDIRTVSPVRPLDGGYANGYCN SKWAGEVLLREAHDL CGLPVATFRSDMILAHPRYRGQVNV
DMFTKLLLSLLITGVAPRSFYIGDGERPRAHY PCLTVDFVAEAVTTLGAQQREGYVSVDVMNPHDEGISLSDVFVDWL
IRAGHPIDRVDDYDDWVRRFETALTALPEKRRATVLP LLLHAFRAPQAPLEGAPEPTEVFHAAVRTAKVGP CDIPHL
DEALIDKYIRD LREFGLI

atgtcgactgccacccatgacgaacgactcgacccgtcgcggtccacgaactcatcgccaccgaccccgcaatttcgcccgc
cgcccaaccgcgaccccgccgatcaacgcgcgcctcgaaacagcccgccggtcgccggtgcgcgagatcatccgcaccgtgc
tcgacggctacgcgcgacccggccggcgctgggacagcgcggtggtagagttcgtcacggacgcgaagacccggcgcaog
tcggcgcaagctgctcccccgcttcgagaccatcacgtacagcggaagtagcgcgagcgtgtttcggcgctgggcgcgcgc
cctgtccgacgacgcgggtgcaccccgccgacccgggtgtggtgctggggttcaacagcgtcgactacgcccacatcg
acatggcgctgggcgccatcggcgcgcgtctcggtgcccgtgcagaccagcgccggaatcagctcgctgcagccgatc
gtggcccgagaccgagcccccctgatcgcggtccagcgtgaaccagctgtccgacgcgggtgcagctgatcaccggcgcc
cgagcaggcgcccaaccggctgggtgggtgttcgactaccacccgcaggtcgacgcgacgagcgaggccgctccaggaca
ccgcggcgcggtgtccagacacggcgctggccgtccagacgctggccgagctgctggagcgcggaaggacacctgccc
ccgctcgcgaggcgccgcgcgcgacgaggaactcgctggccctgctgatctacacctccgggtccaccggcgcccccacaa
ggggcgcgatgtaccacagagcaacgtcggaagatgtggcgccgcggcagcaagaactgggtcggcgagagcgccgc
cgtcgatcacctgaacttcatgccgatgagccacgtgatggcccgaaagcatcctctacggcacgctgggcaacggc
ggcaccgcctacttcggcgcccgccgacgcacctgtccacctgcttgaggacctcgagctgggtgcggcccaaccgagct
caacttcgtcccgcggtatctgggagacgctgtacggcgcaatccagcgtcaggtcgagcgggcggtctccgagggccg
gggacgcgcgcgaacgtcgcgccgtcgagggccgaggtgctggccgagcagcgccagtaacctgctgggcggcggttcc
accttcgcgatgacgggctcggcgcccatctcgccggagctgcgcaactgggtcgagtcgctgctcgaaatgcaacct
gatggacgggtacgggtccaccgagggccggaatgggtgttgttcgacggggagattcagcgcccgccgggtgatcgact
acaagctgggtcgacgtgccggacctgggctacttcagcacccgacccggccgcctccgcgcggcgagctgctgctgcgc
accgagaacatgttcccgggctactacaagcgggcccgaaccacccgcggcgctcttcgacgaggacggctactaccg
caccggcgagctgttcgcgcgagatcgccccggacggctgggtctacgtcgacccgcgcaacaacgtgctcaagctgg
cgaggggcgaaatcgtcacgctggcccaagctggagcggtgttcggcaacagcccgctgatccgcgagatctacgtc
tacgggaacacagcgccccagccctacctgctggcggtcgctgggtgcccacccgagggagggcctggccctcggggtgaccccgga
gacgctcaagcccaagatcgccgactcgctgcagcaggtcgccaaaggagggccggcctgcagtcctacgaggtgcgcg
gcgacttcattcatcgagaccaccccggttcagcctggaaaaacggtctgctgacccgggatccggaagctggcgtggccg
aaactgaagcagcactacgggggaacggctggagcagatgtacgcgcacctggcccgccggacaggcccaacgagctggc
cgagctgcgcgcgcaacgggtgccccagggcgccgggtgttcgagacggtgagccgcgcgcgcggggcgccatgctgggttcgg
ccgctccgacctgtcccccgacgcccacttcaccgatctggcgggagactcggtgtcgcggttgacattcgccaac
ctgctgcgcgagatcttcgacgtcgacgtgccggtagggcgtgatcgctcagcccgcccaacgacctggcgggccatcg
gagctacatcgaggccgagcgccgagggcagggcagcaagcgcccgacgttcgcctcggtgcacggccggggagcggaacgtgg
tcgcgcgcgcgcacctgacgctggacaagtctcctcgacgcgcgagacgctggccgcgcgcgcgaacctgcccaagccg
gcccaccgaggtgcgcacccgtgctgctgacccggcgccacccggcttccctgggcgcgtacctggccctggaaatgggtgga
gcggtaggacatgggtggaacgaacgggtcatcgccctggttcggggcccgcttcgacgaggagggcagcgcgcgcggtgg
acaagaccttcgacagcgccgcaaccgaaactgctcgcgcaactaccagcagctggccgcgcatcacctggaggtcatc
gcccggcgacaaggggcgaggccaatctgggctggggccaaagacgtttgggaacgactggccgacacgggtcgacgtgat
cgtcgaccccgccgcgcgtgggtcaaccacgtgttgcggtacagcgagctgttcggggcccaacgcctgggcaccgcgg
agctgatccggctggcgctgacgtccaagcagaagccgtacacctacgtgtccacctcggcggtgggcgacccagatc
gagccggggcaagttcgtcgagaacgcgcgacatccggcgagatgagcgccaccccgggcgatcaacgacagctacgccaa
cggctatggcaacagcaagtgggcgggcgaggtgctgctgcgcgagggcgcaacgacctgtgcgggtgcgcgtcgcg
tgttcgcgtgcgacatgatcctggccgacaccacgtatgcggggcagctcaacctgccggacatgttcacccggctg
atgctgagcctgggtggccaccgggatcgcgcccggtcgttctacgagctcgacgcgcgacggcaaccgggcagcgggc
gcactacgacggcctgcccgtcgagttcatcgccgcggcgatctcgacgctgggttcgcgagatcaccgacagcgaca
ccggcttcagacctaccacgtgatgaacctacgatgacggcgctcggtctggacaggtacgtcgattggctgggtg
gacggcggtatttcgactcgagcggtatcgccgacactccgaatggctcgggcggttcgagacactcgcgagggccct
cgccggcgacgcgcagtaactcgctgctgcgcgttcgcaacactacggcagcgaggagaagccgatcaacgggtg
cgatagctcccacgcagctgttcggggcagcggtgcaggagcggaataatcgcccccgaacaagacattccgcacgtg
tcgcgcgcgggtcatcgtaagttacatcacggacctgcagctgctcgggctgctctaa

FIGURE 11A

FIGURE 11B

MSTATHDERLDRRVHLLIATDPQFAAAQDPFAITAALEQPGRLRPQIIRTVLDGYADRRPALGQRVVVEFVTDKTRT
SAQLLPRFETITYSEVAQRVSALGRALSDDAVHFGDRVCVLGFNSVDYATIDMALGAIGAVSVPLQTSAAISSLQPI
VAETEPTLIASSVNLSDAVQLITGAEQAPTRLVFDYHPQVDDQREAVQDAAARLSSTGVAVQTLAELLERCKDLF
AVAEPADDESLALLIYTSGSTCAPKGCAMPQSNVCKMWRGSKNWFGEAASITLNFMPMSHVNGRSILYCTLNG
GTAYFAARSDLSLLELVRPTELNFVPRIWETLYGEFQQRQVERLSEAGDAGERRAVEAEVIAEQRYLLGGRF
TFAMTGSAPI SPELRNWVESLLEHLMDDGYGSTEAGMVLFDGEIQPPVVDYKLVDPDLGYFSTDRPHPRGELLR
TENMFPGYYKRAETTAGVFDEDDGYRTGDVFAEIA PDLRVVDRNNVLKLAQGEFVTLAKLEAVFGNSPLIRQIYV
YGENSAQPYLLAVVPTBEALASGDPETLKPKIADSLQCVAKELQSYEVPRDFI IETTPEFSENGLLITGIRKLAWP
KLKQHYGELEQMYADLAAGQANELAELENGAQAPVLQTVSRAAGAMLGSAASDLSPDAHFTDLGGDSLSALTCGN
LLREIFDVDVPVGVIVSPANDLAAIASYTEAERQGSKEPTFASVHGRDATVVRADLTLDKFLDAETLAAAPNLPKP
ATEVRTVLLTGATGFLGRYLALEWLERMDMVDGKVIALVARSDEEARLKDXTFDSGDPKLLAHYQQLAADHLEVI
AGDKGEANLGLQDVWQRLADTVDVIVDPAALVNHVLPYSELFGPNALGTAEILRLALTSKQKPYTYVSTIGVGDQI
EPCKEVENADIRQMSATRAINDSYANGYGNKSWAGEVLLREAHDLGCLPVAVFRCDMI LADTTYAGQLNLPDMFTRL
MLSIVATGCIAPGSPYELDADGNRQRAHYDGLPVEFIAAAISTLGSQITDSDTGFQTYHVMNPPYDDGVGLDEYVDWLV
DAGYSIERIADYSEWLRFETSLRALPDRQRQYSLPLLLHNVRTPEKPINGSIAPTDVFRAAVQEAKEKIGPKDIPHV
SPFVIVKYITDLQLLGLL

atgtcgccaatcacgcgtgaagagcggctcgagcgccgcacccaggacctctacgccaaacgacccgcagttcgccgc
cgccaaaccccgccacggcgatcacccgcagcaatcgagcggcgccgggtctaccgctaccccagatcatcgagaccgtca
tgaccgggatacgccgatcgccgggtctctcgctcagcgctcggtcgaaattcgtagccgaacgcggccacggccacac
acgctgcgactgctccccacttcgaaaccatcagctacggcgagctttgggacccgcacagcgccactggccgacgt
gctcagcaccgaacagacgggtgaaacccggcgacccgggtctgcttgggttgggttcaacagcgctcgactacgccaaga
tcgacatgactttggcgcggtggcgcggtggcggtaccactgcagaccagcgcgcgataacccagctgcagccg
atcgctcgccgagacccagccaccatgatcgccggccagcgctcgacgcactcgctgacgccaccgaattgggtctgtc
cggtcagaccgctacccagagtcctgggtgttcgaccaccacccggcagggttgacgcacacccgcgcagcggtcgaaatccg
cccgggagcgctggcgcggtcgcgcggtcgctcgaaacccctggcgagggccatcgcgcgcgcgcgacgtgcccccggt
gggtcgccggctcgcgcgcccgccacccgatgtgtccgacgactcgctcgcgctactgatctacaccccgggcgagcac
gggtgcgccaaggcgcgatgtaccccgacgcaacgttgcgaccttctggcgcaagcgcgacctgggtcgaggcg
gctacgagcgctcgatcacgctgaacttcatgccaatgagccacgtcatggcgccgcaaatcctgtacggcacgctg
tgcaatggcgcgccacgcctacttctggcgcaaaagcgatctctccaccttggtcgaagacctggcgctgggtgcggcc
caccgagctgaccttctgtcccgcggtgtgggacatgggtgttcgacgagtttcagagtgagggtcgaccgcccgtgg
tcgacggcgccgacccgggtcgcgctcgaaagcccggtcaaggccgagatacgcaacgacgtgctcggtggacgggtat
accagcgactgacgggtccgccccctatctccgacgagatgaaggcggtgggtcgaggagctgctcgacatgcatct
gggtcgagggtacgggtccacccgagggccgggtatgatcctgatcgacggagccatccggcgcccgccgggtactcgact
acaagctggctgatgttcccgacctgggttacttctcgacgacccggccacatccgcggggcgagttgctgggtcaag
accgatagtttgttcccgggctactaccagcgagccgaagtccacccgacgctgttcgatgctgacggcttctaccg
gacggcgacatcatggcgaggtcgggcccgaaagttcggtgtaacctcgacggcgccgaacacgtgttgaagctgt
cgaggggcgagttcgtaacgctctccaaactcgaaagcggtgttggcgacagcccaactggtagcgcgagatctacatc
tacggcaacagcgcccggtgacctacctgttggcggtgatcgctccccaccagggcgctggacggcggtgctgtcga
ggagctcaaggcgcggtggcgcgactcgctgcaagaggtcgcaaaaggccgcccggctgcagtcctacgagatcccg
gcgaacttcatcatcgaaacacacccatggacgctggagaaacggcctgctcacggcgcacggcaagtggccagggcg
cagctgaaaaagcattacggcgagcttctcgagcagatctacacggacctggcacacggccagggccgacgaactgcg
ctcgctgcgcaaaagcggtgacgatggcgcggtgctgggtgacgggtgtgacgtggcgcgcccgcgctgttggggcgga
gcgctctgacgtccagcccgatgcgcacttcaccgatttggcgcgcgactcgctgtcgggcgctgtcggttaccacac
ctgctgcacgagatcttcgacatcgaaagtgcgggtggcggtcatcgctcagcccccgccaaacgacttgcaggccctggc
cgactacgtcgaggcggtcgcaaaacccggctcgtaacggcggaacctcgctcggtccacggcgccctcgaaatgggc
aggtcaccgaggtgcatgcgggtgacctgtccctggacaaattcatcgatgcgcgaacctggccgaagctccccgg
ctgcccccgcaaacacccaaagtgcgcaccggtgctgctgacggcgccacccggcttctcgggcgctacctggccct
ggaatgggtggagcggtggacctgggtcgacggcaaacgtgatctgctgggtccggggccaaagtcggacaccgaagcac
ggcgcggtgggacaagacgttcgacagcgcgaccccgaaactgctggcccaactaccgcgcactggccggcgaccac
ctcgaggtgctcgccggtgacaaggcggaagccgacctcgactggacggcgagacctggcaacgcctggccgacac
gggtcgacctgatcgctcgaccccgcgccctgggtcaaccacgtactgccatcacgacagctgttcggggcccaacgcgc
tgggcaccccgagctgctgcgggtggcgctcacctccaagatcaagccctacagctacacctcgacaatcggtgtc
gccgaccagatcccgccgtcgcggttcacccgaggaagccgacatccgggtcatcagcgccaccccgcgcggtcgacga
cagctacgccaatgggtactcgaaacagcaagtggcgccggcgaggtgctgttgcgcgagggcgcatgacctgtgtggc
tgccggttgcggtgttccgctgcgacatgatcctggcgacaccacatggcggggacagctcaatgtgcgggacatg
ttcaccggatgatcctgagcctggcgccacccgggtatcgcgccgggttcgttctatgagcttgccggccgacggcgcc
ccggcaacgcgcccactatgacgggtctgcccgtcgagttcatcgccgagggcgatttcgactttgggtgcgcagagcc
aggatgggttccacacgtatcacgtgatgaacccctacgacgacggcatcgactcgacgagttcgtcgactggctc
aacgagtcgggttgcgccatccagcgcatcgctgactatggcgactggctgcagcgcttcgaaacccgactgcgcgc
actgcccgatcggcagcgccacagctcaetgctgcgctgttgcacaactatcggcagccggagcgcccgctcgcg
ggctgatcgccctaccgatcgcttcggggcagcggtgcaagaggccaaagatcgcccccgaacaaagacattccgcac
gtcgcgcgccgatcatcgtagagtacgtcagcgacctgcgctactcgccctgctctaa

FIGURE 12A

FIGURE 12B

MSPI TREERLERRIQDLYANDPQFAAAKPATAITAAIERPGLPLPQLIETVMUGYADRPALAQRSVEFVTDAGTGHT
TLRLLPHFETISYGELWDRISALADVLSTEQIVKPGDRVCLLGFNSVDYATILMILARLGAVAVPLQTSAAITQLQP
IVAETQPTMIAASVDALADATELALSGQTATRVLVFDHHRQVDAHRAAVESARERLAGSAVETLAEAIARGDVPFRG
ASAGSAPGTDVSDDSLALLIYTSGSTGAPKGMYPERNVATFWRKRTWPEGGYEPSITLNFMPMSHVMGRQILYGTLL
CNGGTAYFVAKSDLSLTFEDUALVRPTELTFFVPRVWDMVFDEFQSEVDRAALVDGADRVALEAQVKAIEIENDVLGGRY
TSALTGSAFISDEMKAWEELLDMLHVEGYGSTAAGMILIDGAIIRFPFVLDYKLVDPDLGYFLTDRPHPRGELLVVK
TDSLFCGYQRAEVTADVPDADGFYRTGDIMAEVGPEQFVYLDRENNVLKLSQGEFVTVSKLEAVFGDSPVLRQIYI
YGNSARAYLLAVIVPTQEALDAVPVEELKARLGDLSQEVAKAAGLSQSYEIPRDFI IETTPWTLENGLLTGIRKLARP
QLKHYGELLEQIYTDLAHQADELSLRQSGADAPVLVTVCFRAAAALLGGSASDVQPDHFTDLGGDSLSALSFTN
LLHEIFDIEVPVGVIVSPANDLQALADYVEAARKPGSSRPFTFASVHGASNGQVTEVHAGDLSLDKPIDAATLAEAPR
LPAANTQVRTVLLTGATGFLGRYLALWLERMDLVDGKLI CLVRAKSUTEARARLDKTFDSGDPELLAHYFALAGDH
LEVLAGDKGEADLGLDERQTWQRLADTVDLIVDPAALVNHVLPYSQLFCFPAALGTAEELLRLALTSKINPYSYTTICV
ADQIPPSAFTEDADIRVISATRAVDDSYANGYSNSKWAGEVLLREAHDLCCGLPVAVFRCMDILADTTWAGQLNVFDM
FTRMILSLAATGIAPCSFYELAADGARQRAHYDGLPVEPIAEAI STLGAQSQDGEHTYHVMNPYDDGIGLDEFPVDWL
NESGCPIQRIADYGDWLQRFETALRALPDRQRHSSLLPLLNHYRQPERPVRGSIAPTRFRAAVQEA KIGPDKDIPH
VGAPIIVKYVSDLRLLGLL

atgagcaccgcaacccatgatgaacgtctggatcgtcgtgttcatgaactgattgcaaccgatc
cgcagtttgcagcagcacagccggtatcctgcaattaccgcagcactggaacagccctggtctgcg
tctgccgcagattattcgtaccgttctggatgggttatgcagatcgtccggcactgggtcagcgt
gttgttgaatttgttaccgatgcaaaaaccggctcgtaccagcgcacagctgctgcctcgttttg
aaaccattacctatagcgaagtgcacagcgtgttagcgcactgggtcgtgcactgagtgatga
tgcagttcatccgggtgatcgtgttctgtgttctgggttttaatagcgttgattatgccaccatt
gatatggcactgggtgcaattgggtgcagtttagcgttccgctgcagaccagcgcagcaattagca
gcctgcagccgattgttgcagaaaaccgaaccgaccctgattgcaagcagcgttaatacagctgtc
agatgcagttcagctgattaccgggtgcagaaacaggcaccgaccctcgtgtgttgttttgattat
catccgcaggttgatgatcagcgtgaagcagttcaggatgcagcagcagcgtctgagcagcaccg
gtgttgcagttcagaccctggcagaactgctggaacgtggtaaaagatctgcctgcagttgcaga
accgctgcagatgaagatagcctggcactgctgatttataccagcggtagcacaggtgcaccg
aaaggtgcaatgtatccgcagagcaatgttggtaaaatgtggcgtcgtggtagcaaaaattggt
ttggtgaaagcgcagcaagcattaccctgaatttcatgccgatgagccatgttatgggtcgtag
cattctgtatggcaccctgggtaatgggtggcaccgcataattttgcagcacgtagcgtatctgagc
accctgctggaagatctggaactggttctgcgcaccgaactgaattttgttccgcgtattttggg
aaaccctgtatggtgaatttcagcgtcaggttgaaacgtcgtctgagcgaagctggcgatgccgg
tgaaacgtcgtgcagttgaagcagaagtctcggcagaacagcgtcagtatctgctgggtggctcgt
tttacctttgcaatgaccggtagcgcaccgattagtcgcgaactgcgtaatgggttgaaagcc
tgctggaaatgcatctgatggatggctatggtagcaccgaagcaggtatgggtctgtttgatgg
cgaaattcagcgtccgcctgtgattgattataaactggttgatgttccggatctgggttatttt
agcaccgatcgtccgcacccgcgtgggtgaactgctgctgcgtaccgaaaatatgtttccgggtt
attataaacgtgcagaaaccaccgcagggcgtttttgatgaagatgggtattatcgtaccgggtga
tgtgtttgcagaaattgcaccggatcgtctgggtttatgttgatcgtcgtataaatgttctgaaa
ctggcacaggggtgaatttgtgaccctggccaaactggaagcagtttttggtaatagtcgcgtga
ttcgtcagatttatgtgtatggtaaatagcgcacagccgtatctgctggcagttgttgttccgac
cgaagaggcactggcaagcggatgatccggaaaccctgaaaccgaaaattgcagatagcctgcag
caggttgcaaaagaagcaggtctgcagagctatgaagttccgcgtgattttattattgaaacca
ccccgttttagcctggaaaatgggtctgctgaccgggtattcgtaaaactggcatggccgaaactgaa
acagcattatgggtgaacgcctggaacaaatgtatgcagatctggcagcaggtcaggcaaatgaa
ctggccgaactgcgtcgtaatgggtgcacaggcaccgggtctgcagaccgttagccgtgcagccg
gtgcaatgctgggtagcgcagccagcgtatctgagtcggatgcacattttaccgatctgggtgg
tgatagcctgagcgcactgacctttggtaactctgctgcgtgaaatttttgatgtttgatgtgccg
gttgggtgttattgttagtcgggctaattgatctggcagccatttgcaagctatatattgaagcagaac
gtcagggtagcaaacgtccgacctttgcaagcgttcatgggtcgtgatgcaaccgttgttctgtgc
agcagatctgaccctggataaatttctggatgcagaaaccctggcagcagcaccgaatctgccg
aaaccggcaaccgaagtctgtaccgtgctgctgacaggtgcaaccgggttttctgggtcgttatc
tggcactggaatggctggaacgtatggatatggttgatggtaaaagtattgcactgggtcgtgc
ccgtagtgtatgaagaagcacgcgcacgtctggataaaaacctttgatagtgggtgatccgaaactg
ctggcacattatcagcagctgggtgcagatcatctggaagtatttgccgggtgataaagggtgaag
caaatctgggtctgggtcaggatgtttggcagcgtctggcagataaccgttgatgttattgtgga
tccggcagcactgggttaatatcatgttctgccgtatagcgaactgtttgggtccgaatgcactgggc

FIGURE 13A

18/20

accgcagaactgattcgtctggcactgaccagcaaacagaaaccgtatacctatgttagcacca
 ttgggtgttggcgatcagattgaaccgggttaaattttgttgaaaatgccgatattcgtcagatgag
 cgcaaccocgtgcaattaatgatagctatgcaaattggctacggcaatagcaaattgggcaggcgaa
 gttctgctgcgcgaagcacatgatctgtgtggctctgccggttgacagtttttcgtttgtgatatga
 ttctggccgataaccacctatgcagggtcagctgaatctgccggatatgtttaccocgtctgatgct
 gagcctgggttgcaaccgggtattgcaccgggttagctttttatgaactggatgcagatggtaatcgt
 cagcgtgcacattatgatggcctgccggttgaaattttattgcagcagccattagcacccctgggtt
 cacagattaccgatagcgataaccgggttttcagacctatcatgttatgaaccocgtatgatgatgg
 tgttggctctggatgaatatgttgattggctgggttgatgccggttatagcattgaacgtattgca
 gattatagcgaatggctgcgtcgctttgaaacctcactgcgtgcactgccggatcgtcagcgc
 agtatagcctgctgccgctgctgcacaattatcgtaaccggaaaaaccgattaatggtagcat
 tgcaccgaccgatgttttttcgtgcagccgttcaagaagccaaaattgggtccggataaagatatt
 ccgcatgttagccctccggtgattgttaaatatattaccgatctgcagctgctgggtctgctgt
 aa

FIGURE 13A cont.

MSTATHDERLDRRVHELIAIDPQFAAAQPDPAITAALQPGRLRPQIIRTVLDGYADRPALGQR
 VVEFVTDAKTGRTSAQLLPRFETITYSEVAQRVSALGRALSDDAVHFGDRVCVLGFNSVDYATI
 DMALGAIGAVSVPLQTSAAISSLQPIVAETEPTLIASSVNQLSDAVQLITGAEQAPTELVVFDY
 HPQVDDQREAVQDAAARLSSTGVAVQTLAELLERKDLPAVAEPPADEDSLALLIYTSGSTGAP
 KGAMY PQSNVGKMWRRGSKNWFGEAASITLNFMPMSHVMGRSILYGTGLNGGTAYFAARSDLS
 TLEDLELVRPTELNFPRIWETLYGEFQRQVERRLSEAGDAGERRAVEAEVLAEQRQYLLGGR
 FTFAMTGSAPISPELRNWVESLLEMHLMDGYGSTEAGMVLFDGEIQRPVIDYKLVDVPDLGYF
 STDRPHPRGELLRLTENMFPGYYKRAETTAGVFEDDGYRTGDVFAEIPDRLVYVDRRNNVLK
 LAQGEFVTLAKLEAVFGNSPLIRQIYVYGNSAQPYLLAVVVPTEEALASGDPETLKPKIADSLQ
 QVAKEAGLQSYEVPRDFI IETTPFSLENGLLTGIRKLAWPKLKQHYGERLEQMYADLAAGQANE
 LAELRRNGAQAPVLQTVSRAAGAMLGSAASDLSPDAHFTDLGGDSLSALTFGNLLREIFDVP
 VGVIVSPANDLAAIASYIEAERQGSKRPTFASVHGRDATVVRAADLTLDKFLDAETLAAAPNLP
 KPATEVRTVLLTGATGFLGRYLALEWLERMDMVDGKVIALVRARSDEEARARLDKTFDSGDPKL
 LAHYQQLAADHLEVIAGDKGEANLGLGQDVWQRLADTVDVIVDPAALVNHVLPYSELFPGNALG
 TAEIIRLALTSKQKPYTYVSTIGVGQIEPGKFVENADIRQMSATRAINDSYANGYGNKSWAGE
 VLLREAHDLCLPVAVFRCMILADTTYAGQLNLPDMFTRLMLSLVATGIAPGSFYELDADGNR
 QRAHYDGLPVEFIAAAISTLGSQITDSDTGFQTYHVMNPYDDGVGLDEYVDWLVDAGYSIERIA
 DYSEWLRRFETSLRALPDRQRQYSLPLLNHYRTPEKPIINGSIAPTDFRAAVQEAKIGPDKDI
 PHVSPPVIVKYITDLQLLGLL

FIGURE 13B

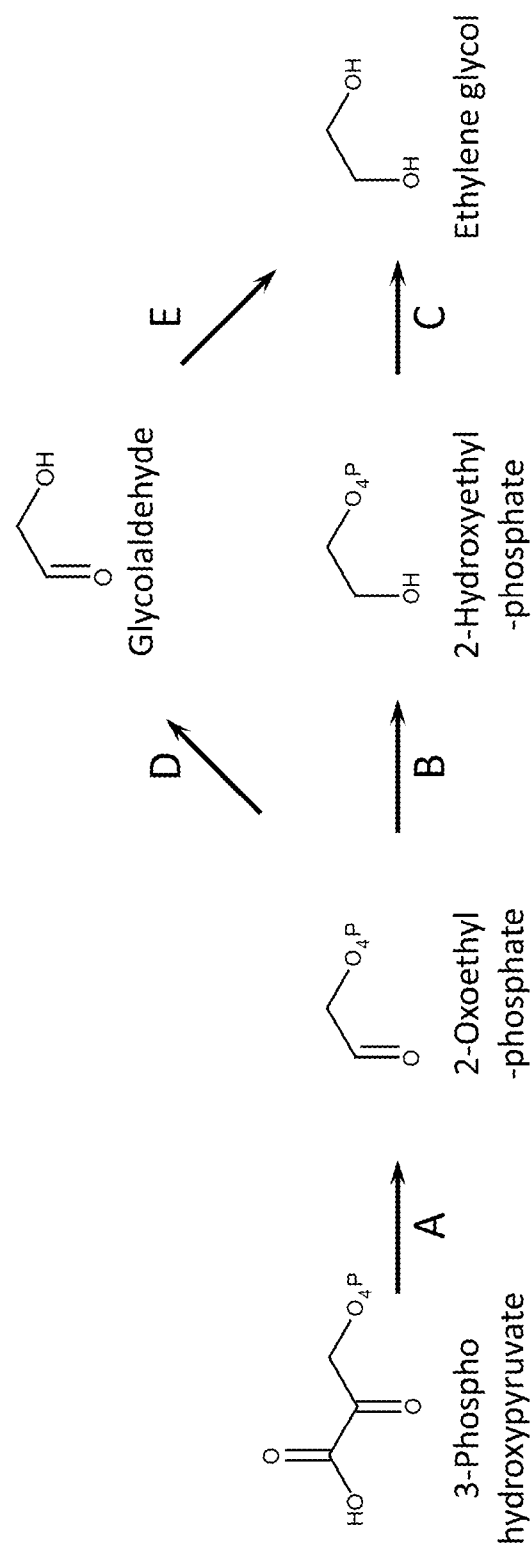


Figure 14

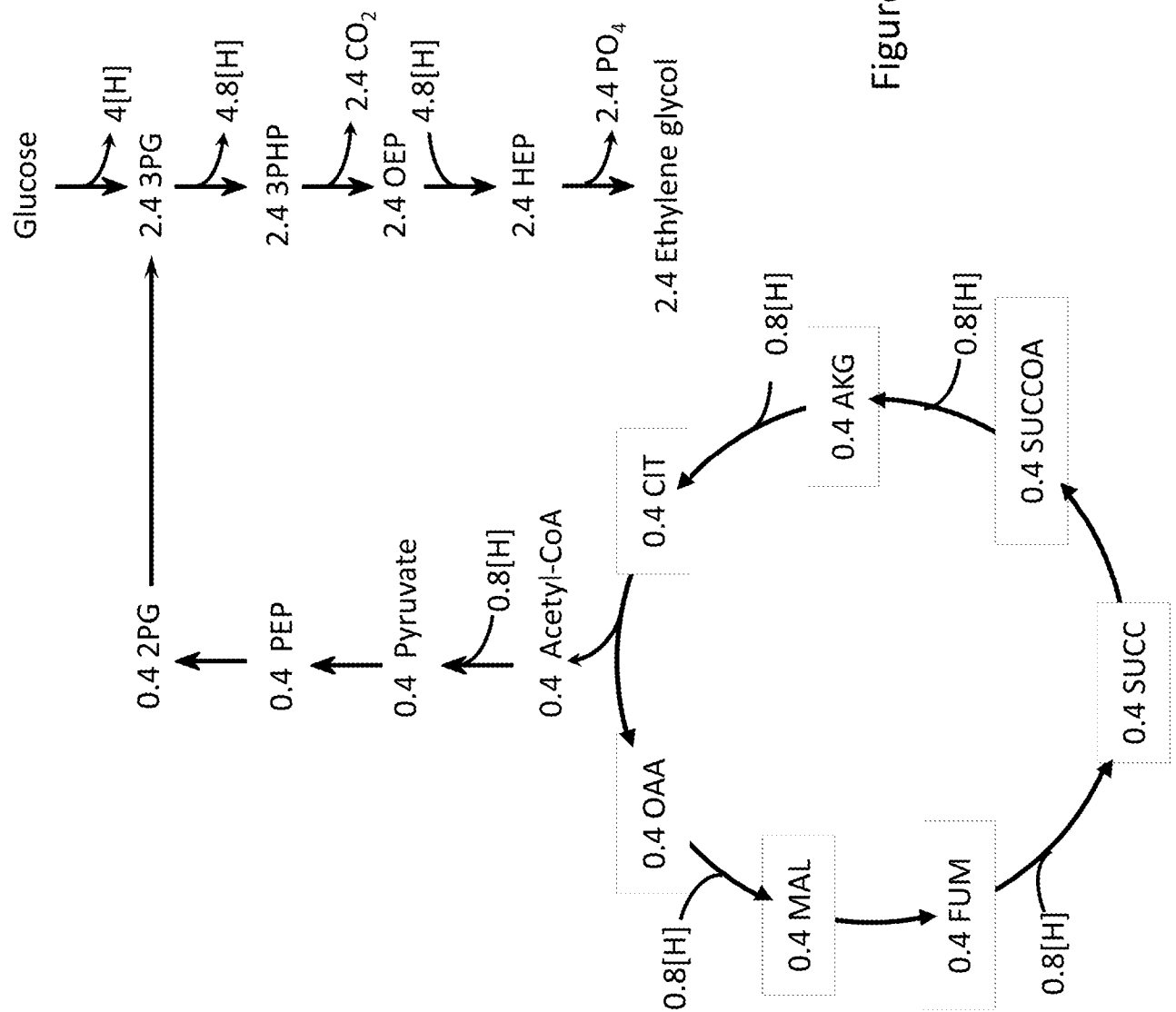


Figure 15