# MICROORGANISMS AND METHODS FOR THE BIOSYNTHESIS OF P-TOLUATE AND TEREPHTHALATE

## BACKGROUND OF THE INVENTION

[0001] The present invention relates generally to biosynthetic processes, and more specifically to organisms having p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate biosynthetic capability.

[0002] Terephthalate (also known as terephthalic acid and PTA) is the immediate precursor of polyethylene terepthalate (PET), used to make clothing, resins, plastic bottles and even as a poultry feed additive. Nearly all PTA is produced from *para*-xylene by oxidation in air in a process known as the Mid Century Process. This oxidation is conducted at high temperature in an acetic acid solvent with a catalyst composed of cobalt and/or manganese salts. Para-xylene is derived from petrochemical sources and is formed by high severity catalytic reforming of naphtha. Xylene is also obtained from the pyrolysis gasoline stream in a naphtha steam cracker and by toluene disproportion.

[0003] Cost-effective methods for generating renewable PTA have not yet been developed to date. PTA, toluene and other aromatic precursors are naturally degraded by some bacteria. However, these degradation pathways typically involve monooxygenases that operate irreversibly in the degradative direction. Hence, biosynthetic pathways for PTA are severely limited by the properties of known enzymes to date.

[0004] A promising precursor for PTA is p-toluate, also known as p-methylbenzoate. P-Toluate is an intermediate in some industrial processes for the oxidation of p-xylene to PTA. It is also an intermediate for polymer stabilizers, pesticides, light sensitive compounds, animal feed supplements and other organic chemicals. Only slightly soluble in aqueous solution, p-toluate is a solid at physiological temperatures, with a melting point of 275°C. Microbial catalysts for synthesizing this compound from sugar feedstocks have not been described to date.

[0005] Thus, there exists a need for alternative methods for effectively producing commercial quantities of compounds such as p-toluate or terephthalate. The present invention satisfies this need and provides related advantages as well.

## SUMMARY OF THE INVENTION

[0006] The disclosure provides non-naturally occurring microbial organisms having a (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway, p- toluate pathway, and/or terephthalate pathway. The disclosure additionally provides methods of using such organisms to produce (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway, p- toluate pathway or terephthalate pathway.

[0006a] In a first aspect, the present invention refers to a non-naturally occurring microbial organism having carbon monoxide dehydrogenase or hydrogenase activity and a reductive tricarboxylic acid (rTCA) pathway, wherein said microbial organism comprises (1) at least one exogenous nucleic acid encoding a rTCA pathway enzyme expressed in a sufficient amount to convert CO, $CO_2$ or $H_2$ to acetyl-CoA, wherein said rTCA enzyme is selected from the group consisting of an ATP citrate-lyase, a citrate lyase, an aconitase, an isocitrate dehydrogenase, an alpha-ketoglutarate:ferredoxin oxidoreductase, a succinyl-CoA synthetase, a succinyl-CoA transferase, a fumarate reductase, a fumarase, and a malate dehydrogenase; and (2) at least one exogenous nucleic acid encoding a biosynthetic product pathway enzyme expressed in a sufficient amount to produce said biosynthetic product, wherein the microbial organism does not require light for growth.

[0006b] In a second aspect, the present invention refers to a non-naturally occurring microbial organism comprising

a rTCA pathway comprising at least one exogenous nucleic acid encoding a rTCA pathway enzyme expressed in a sufficient amount to convert (i) CO, (ii) $CO_2$ and $H_2$, (iii) CO and $CO_2$, (iv) synthesis gas comprising CO and $H_2$, or (v) synthesis gas comprising CO, $CO_2$, and $H_2$; to acetyl-CoA; wherein said at rTCA pathway enzyme is selected from the group consisting of an ATP citrate-lyase, a citrate lyase, an aconitase, an isocitrate dehydrogenase, an alpha-ketoglutarate:ferredoxin oxidoreductase, a succinyl-CoA synthetase, a succinyl-CoA transferase, a fumarate reductase, a fumarase and a malate dehydrogenase;

a reducing equivalent producing pathway comprising at least one exogenous nucleic acid encoding a reducing equivalent producing enzyme selected from the group consisting of a carbon monoxide dehydrogenase, a hydrogenase and a NAD(P)H:ferredoxin oxidoreductase; and

at least one exogenous nucleic acid encoding a biosynthetic product pathway enzyme expressed in a sufficient amount to produce a biosynthetic product, wherein the microbial organism does not require acetyl-CoA synthase.

[0006c] In a third aspect, the present invention refers to a non-naturally occurring microbial organism comprising a reducing equivalent producing pathway comprising at least one exogenous nucleic acid encoding a reducing equivalent producing enzyme selected from the group consisting of a carbon monoxide dehydrogenase, a hydrogenase and a NAD(P)H:ferredoxin oxidoreductase; and

at least one exogenous nucleic acid encoding a biosynthetic product pathway enzyme expressed in a sufficient amount to produce said biosynthetic product, wherein the microbial organism does not require acetyl-CoA synthase.

## BRIEF DESCRIPTION OF THE DRAWINGS

[0007] Figure 1 shows a schematic depiction of an exemplary pathway to (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate (2H3M4OP) from glyceraldehyde-3-phosphate and pyruvate. G3P is glyceraldehyde-3-phosphate, DXP is 1-deoxy-D-xylulose-5-phosphate and 2ME4P is C-methyl-D-erythritol-4-phosphate. Enzymes are (A) DXP synthase; (B) DXP reductoisomerase; and (C) 2ME4P dehydratase.

[0008] Figure 2 shows a schematic depiction of an exemplary alternate shikimate pathway to p-toluate. Enzymes are: (A) 2-dehydro-3-deoxyphosphoheptonate synthase; (B) 3-dehydroquinate synthase; (C) 3-dehydroquinate dehydratase; (D) shikimate dehydrogenase; (E) Shikimate kinase; (F) 3-phosphoshikimate-2-carboxyvinyltransferase; (G) chorismate synthase; and (H) chorismate lyase. Compounds are: (1) (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate; (2) 2,4-dihydroxy-5-methyl-6-[(phosphonooxy)methyl]oxane-2-carboxylate; (3) 1,3-dihydroxy-4-methyl-5-oxocyclohexane-1-carboxylate; (4) 5-hydroxy-4-methyl-3-oxocyclohex-1-ene-1-carboxylate; (5) 3,5-dihydroxy-4-methylcyclohex-1-ene-1-carboxylate; (6) 5-hydroxy-4-methyl-3-(phosphonooxy)cyclohex-1-ene-1-carboxylate; (7) 5-[(1-carboxyeth-1-en-1-yl)oxy]-4-methyl-3-(phosphonooxy)cyclohex-1-ene-1-carboxylate; (8) 3-[(1-carboxyeth-1-en-1-yl)oxy]-4-methylcyclohexa-1,5-diene-1-carboxylate; and (9) p-toluate.

[0009] Figure 3 shows an exemplary pathway for conversion of p-toluate to terephthalic acid (PTA). Reactions A, B and C are catalyzed by p-toluate methyl-monooxygenase reductase, 4-carboxybenzyl alcohol dehydrogenase and 4-carboxybenzyl aldehyde dehydrogenase, respectively. The compounds shown are (1) p-toluic acid; (2) 4-carboxybenzyl alcohol; (3) 4-carboxybenzaldehyde and (4) terephthalic acid.

## DETAILED DESCRIPTION OF THE INVENTION

[0010] The present invention is directed to the design and production of cells and organisms having biosynthetic production capabilities for p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate. The results described herein indicate that metabolic pathways can be designed and recombinantly engineered to achieve the biosynthesis of p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate in *Escherichia coli* and other cells or organisms. Biosynthetic production of p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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 p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate biosynthesis, including under conditions approaching theoretical maximum growth.

[0011] The shikimate biosynthesis pathway in *E. coli* converts erythrose-4-phosphate to chorismate, an important intermediate that leads to the biosynthesis of many essential metabolites including 4-hydroxybenzoate. 4-Hydroxybenzoate is structurally similar to p-toluate, an industrial precursor of terephthalic acid. As disclosed herein, shikimate pathway enzymes are utilized to accept the alternate substrate, (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate (2H3M4OP) and transform it to p-toluate. In addition, a pathway is used to synthesize the 2H3M4OP precursor using enzymes from the non-mevalonate pathway for isoprenoid biosynthesis.

[0012] Disclosed herein are strategies for engineering a microorganism to produce renewable p-toluate or terephthalate (PTA) from carbohydrate feedstocks. First, glyceraldehyde-3-phosphate (G3P) and pyruvate are converted to 2-hydroxy-3-methyl-4-oxobutoxy)phosphonate (2H3M4OP) in three enzymatic steps (see Example I and Figure 1). The 2H3M4OP intermediate is subsequently transformed to p-toluate by enzymes in the shikimate pathway (see Example II and Figure 2). P-Toluate can be further converted to PTA by a microorganism (see Example III and Figure 3).

[0013] The conversion of G3P to p-toluate requires one ATP, two reducing equivalents (NAD(P)H), and two molecules of phosphoenolpyruvate, according to net reaction below.

$$G3P + 2\ PEP + ATP + 2\ NAD(P)H + 2\ H^+ \rightarrow p\text{-}Toluate + 4\ Pi + ADP + 2\ NAD(P)^+ + CO_2 + H_2O$$

[0014] An additional ATP is required to synthesize G3P from glucose. The maximum theoretical p-toluate yield is 0.67 mol/mol (0.51 g/g) from glucose minus carbon required for energy. Under the assumption that 2 ATPs are consumed per p-toluate molecule synthesized, the predicted p-toluate yield from glucose is 0.62 mol/mol (0.46 g/g) p-toluate.

[0015] If p-toluate is further converted to PTA by enzymes as described in Example III, the predicted PTA yield from glucose is 0.64 mol/mol (0.58 g/g). In this case, the oxidation of p-toluate to PTA generates an additional net reducing equivalent according to the net reaction:

$$p\text{-}toluate + O_2 + NAD^+ \rightarrow PTA + NADH + 2\ H^+$$

[0016] Enzyme candidates for catalyzing each step of the proposed pathways are described in the following sections.

[0017] 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 a p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate biosynthetic pathway.

[0018] A metabolic modification refers to a biochemical reaction that is altered from its naturally occurring state. Therefore, non-naturally occurring microorganisms can have genetic

modifications to nucleic acids encoding metabolic polypeptides, or functional fragments thereof. Exemplary metabolic modifications are disclosed herein.

[0019] 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 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 examples of isolated microbial organisms include partially pure microbes, substantially pure microbes and microbes cultured in a medium that is non-naturally occurring.

[0020] 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 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.

[0021] 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.

[0022] As used herein, the term "(2-hydroxy-3-methyl-4-oxobutoxy)phosphonate," abbreviated herein as 2H3M4OP, has the chemical formula as shown in Figure 1. Such a compound can also be described as 3-hydroxy-2-methyl butanal-4-phosphate.

[0023] As used herein, the term "p-toluate," having the molecular formula $C_8H_7O_2^-$ (see Figure 2, compound 9)(IUPAC name 4-methylbenzoate) is the ionized form of p-toluic acid, and

it is understood that p-toluate and p-toluic acid can be used interchangeably throughout to refer to the compound in any of its neutral or ionized forms, including any salt forms thereof. It is understood by those skilled understand that the specific form will depend on the pH.

[0024] As used herein, the term "terephthalate," having the molecular formula $C_8H_4O_4^{-2}$ (see Figure 3, compound 4)(IUPAC name terephthalate) is the ionized form of terephthalic acid, also referred to as p-phthalic acid or PTA, and it is understood that terephthalate and terephthalic acid can be used interchangeably throughout to refer to the compound in any of its neutral or ionized forms, including any salt forms thereof. It is understood by those skilled understand that the specific form will depend on the pH.

[0025] 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.

[0026] "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.

[0027] 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.

[0028] 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.

[0029] 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.

[0030] 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 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 identity. Genes encoding proteins sharing an amino acid similarity less that 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.

[0031] 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 three genes and their corresponding products are considered to be orthologs. For the production of a biochemical product, those skilled in the 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 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.

[0032] 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.

[0033] 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.

[0034] Therefore, in identifying and constructing the non-naturally occurring microbial organisms of the invention having p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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.

[0035] 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.

[0036] 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.

[0037] The invention provides a non-naturally occurring microbial organism, comprising a microbial organism having a (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway comprising at least one exogenous nucleic acid encoding a (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway enzyme expressed in a sufficient amount to produce (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate, the (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway comprising 2-C-methyl-D-erythritol-4-phosphate dehydratase (see Example I and Figure 1, step C). A non-naturally occurring microbial organism comprising a (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway can further comprise 1-deoxyxylulose-5-phosphate synthase or 1-deoxy-D-xylulose-5-phosphate reductoisomerase (see Example I and Figure 1, steps A and B). Thus, a (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate can comprise 5 2-C-methyl-D-erythritol-4-phosphate dehydratase, 1-deoxyxylulose-5-phosphate synthase and 1-deoxy-D-xylulose-5-phosphate reductoisomerase.

[0038] The invention also provides a non-naturally occurring microbial organism, comprising a microbial organism having a p-toluate pathway comprising at least one exogenous nucleic acid encoding a p-toluate pathway enzyme expressed in a sufficient amount to produce p-toluate, the p-toluate pathway comprising 2-dehydro-3-deoxyphosphoheptonate synthase; 3-dehydroquinate synthase; 3-dehydroquinate dehydratase; shikimate dehydrogenase; shikimate kinase; 3-phosphoshikimate-2-carboxyvinyltransferase; chorismate synthase; or chorismate lyase (see Example II and Figure 2, steps A-H). A non-naturally occurring microbial organism having a p-toluate pathway can further comprise a (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway (Figure 1). A (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway can comprise, for example, 2-C-methyl-D-erythritol-4-phosphate dehydratase, 1-deoxyxylulose-5-phosphate synthase or 1-deoxy-D-xylulose-5-phosphate reductoisomerase (Figure 1).

[0039] The invention additionally provides a non-naturally occurring microbial organism, comprising a microbial organism having a terephthalate pathway comprising at least one exogenous nucleic acid encoding a terephthalate pathway enzyme expressed in a sufficient

amount to produce terephthalate, the terephthalate pathway comprising p-toluate methyl-monooxygenase reductase; 4-carboxybenzyl alcohol dehydrogenase; or 4-carboxybenzyl aldehyde dehydrogenase (see Example III and Figure 3). Such an organism containing a terephthalate pathway can additionally comprise a p-toluate pathway, wherein the p-toluate pathway comprises 2-dehydro-3-deoxyphosphoheptonate synthase; 3-dehydroquinate synthase; 3-dehydroquinate dehydratase; shikimate dehydrogenase; shikimate kinase; 3-phosphoshikimate-2-carboxyvinyltransferase; chorismate synthase; or chorismate lyase (see Examples II and III and Figures 2 and 3). Such a non-naturally occurring microbialorganism having a terephthalate pathway and a p-toluate pathway can further comprise a (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway (see Example I and Figure 1). A (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway can comprise, for example, 2-C-methyl-D-erythritol-4-phosphate dehydratase, 1-deoxyxylulose-5-phosphate synthase or 1-deoxy-D-xylulose-5-phosphate reductoisomerase (see Example I and Figure 1).

[0040] In an additional embodiment, the invention provides a non-naturally occurring microbial organism having a p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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. For example, in a (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway, the substrates and products can be selected from the group consisting of glyceraldehyde-3-phosphate and pyruvate to 1-deoxy-D-xylulose-5-phosphate; 1-deoxy-D-xylulose-5-phosphate to C-methyl-D-erythritol-4-phosphate; and C-methyl-D-erythritol-4-phosphate to (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate (see Example I and Figure 1). In another embodiment, a p-toluate pathway can comprise substrates and products selected from (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate to 2,4-dihydroxy-5-methyl-6-[(phosphonooxy)methyl]oxane-2-carboxylate; 2,4-dihydroxy-5-methyl-6-[(phosphonooxy)methyl]oxane-2-carboxylate to 1,3-dihydroxy-4-methyl-5-oxocyclohexane-1-carboxylate; 1,3-dihydroxy-4-methyl-5-oxocyclohexane-1-carboxylate to 5-hydroxy-4-methyl-3-oxocyclohex-1-ene-1-carboxylic acid; 5-hydroxy-4-methyl-3-oxocyclohex-1-ene-1-carboxylic acid to 3,5-dihydroxy-4-methylcyclohex-1-ene-1-carboxylate; 3,5-dihydroxy-4-methylcyclohex-1-ene-1-carboxylate to 5-hydroxy-4-methyl-3-(phosphonooxy)cyclohex-1-ene-1-carboxylate; 5-hydroxy-4-methyl-3-(phosphonooxy)cyclohex-1-ene-1-carboxylate to 5-[(1-carboxyeth-1-en-1-yl)oxy]-4-methyl-3-(phosphonooxy)cyclohex-1-ene-1-carboxylate; 5-[(1-carboxyeth-1-en-1-

yl)oxy]-4-methyl-3-(phosphonooxy)cyclohex-1-ene-1-carboxylate to 3-[(1-carboxyeth-1-en-1-yl)oxy]-4-methylcyclohexa-1,5-diene-1-carboxylate; and 3-[(1-carboxyeth-1-en-1-yl)oxy]-4-methylcyclohexa-1,5-diene-1-carboxylate to p-toluate (see Example II and Figure 2). In still another embodiment, a terephthalate pathway can comprise substrates and products selected from p-toluate to 4-carboxybenzyl alcohol; 4-carboxybenzyl alcohol to 4-carboxybenzaldehyde; and 4-carboxybenzaldehyde to and terephthalic acid (see Example III and Figure 3). 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 a p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway, such as that shown in Figures 1-3.

[0041] While generally described herein as a microbial organism that contains a p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway, it is understood that the invention additionally provides a non-naturally occurring microbial organism comprising at least one exogenous nucleic acid encoding a p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway enzyme expressed in a sufficient amount to produce an intermediate of a p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway. For example, as disclosed herein, a (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway is exemplified in Figure 1 (see Example I). Therefore, in addition to a microbial organism containing a (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway that produces (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate, the invention additionally provides a non-naturally occurring microbial organism comprising at least one exogenous nucleic acid encoding a (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway enzyme, where the microbial organism produces a (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway intermediate, for example, 1-deoxy-D-xylulose-5-phosphate or C-methyl-D-erythritol-4-phosphate. Similarly, the invention also provides a non-naturally occurring microbial organism containing a p-toluate pathway that produces p-toluate, wherein the non-naturally occurring microbial organism comprises at least one exogenous nucleic acid encoding a p-toluate pathway enzyme, where the microbial organism produces a p-toluate pathway intermediate, for example, 2,4-dihydroxy-5-

methyl-6-[(phosphonooxy)methyl]oxane-2-carboxylate, 1,3-dihydroxy-4-methyl-5-oxocyclohexane-1-carboxylate, 5-hydroxy-4-methyl-3-oxocyclohex-1-ene-1-carboxylate, 3,5-dihydroxy-4-methylcyclohex-1-ene-1-carboxylate, 5-hydroxy-4-methyl-3-(phosphonooxy)cyclohex-1-ene-1-carboxylate, 5-[(1-carboxyeth-1-en-1-yl)oxy]-4-methyl-3-(phosphonooxy)cyclohex-1-ene-1-carboxylate, or 3-[(1-carboxyeth-1-en-1-yl)oxy]-4-methylcyclohexa-1,5-diene-1-carboxylate. Further, the invention additionally provides a non-naturally occurring microbial organism containing a terephthalate pathway enzyme, where the microbial organism produces a terephthalate pathway intermediate, for example, 4-carboxybenzyl alcohol or 4-carboxybenzaldehyde.

[0042] 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, 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 a p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway intermediate can be utilized to produce the intermediate as a desired product.

[0043] 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.

[0044] 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 p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate biosynthetic pathways. Depending on the host microbial organism chosen for biosynthesis, nucleic acids for some or all of a particular p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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 p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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 p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate.

[0045] 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, 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, Kluyveromyces lactis, Kluyveromyces marxianus, Aspergillus terreus, Aspergillus niger*, *Pichia pastoris, Rhizopus arrhizus, Rhizobus oryzae*, and the like. *E. coli* is a particularly useful host organisms 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 be used to introduce metabolic and/or genetic modifications to produce a desired product.

[0046] Depending on the p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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 p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway-encoding nucleic acid and up to all encoding nucleic acids for one or more p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate biosynthetic pathways. For example, p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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 a p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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 one of the pathway enzymes or proteins. For example, exogenous expression of all enzymes or proteins in a pathway for production of p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate can be included. For example, all enzymes in a p-toluate pathway can be included, such as 2-dehydro-3-deoxyphosphoheptonate synthase; 3-dehydroquinate synthase; 3-dehydroquinate dehydratase; shikimate dehydrogenase; shikimate kinase; 3-phosphoshikimate-2-carboxyvinyltransferase; chorismate synthase; and chorismate lyase. In addition, all enzymes in a terephthalate pathway can be included, such as p-toluate methyl-monooxygenase reductase; 4-carboxybenzyl alcohol dehydrogenase; and 4-carboxybenzyl aldehyde dehydrogenase. Furthermore, all enzymes in a (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway can be included, such as 2-C-methyl-D-erythritol-4-phosphate dehydratase, 1-deoxyxylulose-5-phosphate synthase and 1-deoxy-D-xylulose-5-phosphate reductoisomerase.

[0047] Given the teachings and guidance provided herein, those skilled in the art will understand that the number of encoding nucleic acids to introduce in an expressible form will, at least, parallel the p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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, or eight, depending on the particular pathway, that is, up to all nucleic acids encoding the enzymes or proteins constituting a p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate biosynthetic pathway disclosed herein. In some embodiments, the non-

naturally occurring microbial organisms also can include other genetic modifications that facilitate or optimize p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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 p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway precursors such as glyceraldehyde-3-phosphate, pyruvate, (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate or p-toluate. Furthermore, as disclosed herein, multiple pathways can be included in a single organism such as the pathway to produce p-toluate (Figure 2), terephthalate (Figure 3) and (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate (Figure 1), as desired.

[0048] Generally, a host microbial organism is selected such that it produces the precursor of a p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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, glyceraldehyde-3-phosphate and phosphoenolpyruvate are 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 a p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway.

[0049] In some embodiments, a non-naturally occurring microbial organism of the invention is generated from a host that contains the enzymatic capability to synthesize p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate. In this specific embodiment it can be useful to increase the synthesis or accumulation of a p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway product to, for example, drive p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway reactions toward p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate production. Increased synthesis or accumulation can be accomplished by, for example, overexpression of nucleic acids encoding one or more of the above-described p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway enzymes or proteins. Over expression the enzyme or enzymes and/or protein or proteins of the p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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 p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate, through overexpression of one, two, three, four, five, and so forth, that is, up to all nucleic acids encoding p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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 p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate biosynthetic pathway.

[0050] 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 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 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.

[0051] 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 microbial organism of the invention. The nucleic acids can be introduced so as to confer, for example, a p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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 p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate biosynthetic capability. For example, a non-naturally occurring microbial organism having a p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate biosynthetic pathway can comprise at least two exogenous nucleic acids

encoding desired enzymes or proteins. For example, in a (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway, a combination of the enzymes expressed can be a combination of 2-C-methyl-D-erythritol-4-phosphate dehydratase and 1-deoxyxylulose-5-phosphate synthase, or 2-C-methyl-D-erythritol-4-phosphate dehydratase and 1-deoxy-D-xylulose-5-phosphate reductoisomerase. In a p-toluate pathway, a combination of the enzymes expressed can be a combination of 2-dehydro-3-deoxyphosphoheptonate synthase and 3-dehydroquinate dehydratase; shikimate kinase and 3-phosphoshikimate-2-carboxyvinyltransferase; shikimate kinase and shikimate dehydrogenase and, and the like. Similarly, in a terephthalate pathway, a combination of the expressed enzymes can be p-toluate methyl-monooxygenase reductase and 4-carboxybenzyl alcohol dehydrogenase; or 4-carboxybenzyl alcohol dehydrogenase and 4-carboxybenzyl aldehyde dehydrogenase, 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 biosynthetic pathway can be included in a non-naturally occurring microbial organism of the invention, for example, 3-dehydroquinate synthase, shikimate dehydrogenase and shikimate kinase; shikimate kinase, chorismate synthase and chorismate lyase; 3-dehydroquinate dehydratase, chorismate synthase and chorismate lyase, 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, five, six, seven or more enzymes or proteins of a biosynthetic pathway, depending on the 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.

[0052] In addition to the biosynthesis of p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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 p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate other than use of the p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate producers is through addition of another microbial organism capable of converting a p-toluate, terephthalate or (2-hydroxy-3-methyl-4-

oxobutoxy)phosphonate pathway intermediate to p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate. One such procedure includes, for example, the fermentation of a microbial organism that produces a p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway intermediate. The p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway intermediate can then be used as a substrate for a second microbial organism that converts the p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway intermediate to p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate. The p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway intermediate can be added directly to another culture of the second organism or the original culture of the p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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.

[0053] 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, p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate. 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 p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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, p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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 a p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate intermediate and the second microbial organism converts the intermediate to p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate.

[0054] 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 p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate.

[0055] Sources of encoding nucleic acids for a p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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, Mycobacterium tuberculosis, Agrobacterium tumefaciens, Bacillus subtilis, Synechocystis* species, *Arabidopsis thaliana, Zymomonas mobilis, Klebsiella oxytoca, Salmonella typhimurium, Salmonella typhi, Lactobacullus collinoides, Klebsiella pneumoniae, Clostridium pasteuranum, Citrobacter freundii, Clostridium butyricum, Roseburia inulinivorans, Sulfolobus solfataricus, Neurospora crassa, Sinorhizobium fredii, Helicobacter pylori, Pyrococcus furiosus, Haemophilus influenzae, Erwinia chrysanthemi, Staphylococcus aureus, Dunaliella salina, Streptococcus pneumoniae, Saccharomyces cerevisiae, Aspergillus nidulans, Pneumocystis carinii, Streptomyces coelicolor,* species from the genera *Burkholderia*, *Alcaligenes*, *Pseudomonas*, *Shingomonas* and *Comamonas*, for example, *Comamonas testosteroni*, 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 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 p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate biosynthetic activity for one or more genes in related or distant species, including for example, homologues, orthologs, paralogs and nonorthologous gene displacements of known genes, and the interchange of genetic alterations between organisms is routine and well known in the art. Accordingly, the metabolic alterations allowing biosynthesis of p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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.

[0056] In some instances, such as when an alternative p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate biosynthetic pathway exists in an unrelated species, p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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 p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate.

[0057] Methods for constructing and testing the expression levels of a non-naturally occurring p-toluate-, terephthalate- or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate-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).

[0058] Exogenous nucleic acid sequences involved in a pathway for production of p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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.

[0059] An expression vector or vectors can be constructed to include one or more p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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. 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 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 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 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.

[0060] The invention additionally provides a method for producing (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate, comprising culturing the non-naturally occurring microbial organism containing a (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway under conditions and for a sufficient period of time to produce (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate. Such a microbial organism can have a (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway comprising at least one exogenous nucleic acid encoding a (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway enzyme expressed in a sufficient amount to produce (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate, the (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway comprising 2-C-methyl-D-erythritol-4-phosphate dehydratase (see Example I and Figure 1, step C). A (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway can optionally further comprise 1-deoxyxylulose-5-phosphate synthase and/or 1-deoxy-D-xylulose-5-phosphate reductoisomerase (see Example I and Figure 1, steps A and B).

[0061] In another embodiment, the invention provides a method for producing p-toluate, comprising culturing the non-naturally occurring microbial organism comprising a p-toluate pathway under conditions and for a sufficient period of time to produce p-toluate. A p-toluate pathway can comprise at least one exogenous nucleic acid encoding a p-toluate pathway enzyme expressed in a sufficient amount to produce p-toluate, the p-toluate pathway comprising 2-dehydro-3-deoxyphosphoheptonate synthase; 3-dehydroquinate synthase; 3-dehydroquinate dehydratase; shikimate dehydrogenase; shikimate kinase; 3-phosphoshikimate-2-carboxyvinyltransferase; chorismate synthase; and/or chorismate lyase (see Example II and Figure 2, steps A-H). In another embodiment, a method of the invention can utilize a non-naturally occurring microbial organism that further comprises a (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway (see Example I and Figure 1). Such a (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway can comprise 2-C-methyl-D-erythritol-4-phosphate dehydratase, 1-deoxyxylulose-5-phosphate synthase and/or 1-deoxy-D-xylulose-5-phosphate reductoisomerase (see Example I and Figure 1).

[0062] The invention further provides a method for producing terephthalate, comprising culturing a non-naturally occurring microbial organism containing a terephthalate pathway under

conditions and for a sufficient period of time to produce terephthalate. Such a terephthalate pathway can comprise at least one exogenous nucleic acid encoding a terephthalate pathway enzyme expressed in a sufficient amount to produce terephthalate, the terephthalate pathway comprising p-toluate methyl-monooxygenase reductase; 4-carboxybenzyl alcohol dehydrogenase; and/or 4-carboxybenzyl aldehyde dehydrogenase. Such a microbial organism can further comprise a p-toluate pathway, wherein the p-toluate pathway comprises 2-dehydro-3-deoxyphosphoheptonate synthase; 3-dehydroquinate synthase; 3-dehydroquinate dehydratase; shikimate dehydrogenase; shikimate kinase; 3-phosphoshikimate-2-carboxyvinyltransferase; chorismate synthase; and/or chorismate lyase (see Examples 2 and 3 and Figures 2 and 3). In another embodiment, the non-naturally occurring microbial organism can further comprise a (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway (see Example I and Figure 1). Thus, in a particular embodiment, the invention provides a non-naturally occurring microbial organism and methods of use, in which the microbial organism contains p-toluate, terephthalate and (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathways.

[0063] Suitable purification and/or assays to test for the production of p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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), LC-MS (Liquid Chromatography-Mass Spectroscopy), and UV-visible 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, p-toluate methyl-monooxygenase activity can be assayed by incubating purified enzyme with NADH, $FeSO_4$ and the p-toluate substrate in a water bath, stopping the reaction by precipitation of the proteins, and analysis of the products in the supernatant by HPLC (Locher et al., *J. Bacteriol.* 173:3741-3748 (1991)).

[0064] The p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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.

[0065] 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 p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate producers can be cultured for the biosynthetic production of p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate.

[0066] For the production of p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate, the recombinant strains are cultured in a medium with carbon source and other essential nutrients. It is sometimes 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 conditions can be applied by perforating the septum with a small hole for limited aeration. Exemplary 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.

[0067] If desired, the pH of the medium can be maintained at a desired pH, in particular neutral pH, 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.

[0068] The growth medium can include, for example, any carbohydrate source which can supply a 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 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 p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate.

[0069] In addition to renewable feedstocks such as those exemplified above, the p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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 p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate producing organisms to provide a metabolic pathway for utilization of syngas or other gaseous carbon source.

[0070] 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 $H_2$ 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 $H_2$ and CO, syngas can also include $CO_2$ and other gases in smaller quantities. Thus, synthesis gas provides a cost effective source of gaseous carbon such as CO and, additionally, $CO_2$.

[0071] The Wood-Ljungdahl pathway catalyzes the conversion of CO and $H_2$ to acetyl-CoA and other products such as acetate. Organisms capable of utilizing CO and syngas also generally have the capability of utilizing $CO_2$ and $CO_2/H_2$ mixtures through the same basic set of enzymes and transformations encompassed by the Wood-Ljungdahl pathway. $H_2$-dependent conversion of $CO_2$ 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 $CO_2$ 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:

$$2\ CO_2 + 4\ H_2 + n\ ADP + n\ Pi \rightarrow CH_3COOH + 2\ H_2O + n\ ATP$$

Hence, non-naturally occurring microorganisms possessing the Wood-Ljungdahl pathway can utilize $CO_2$ and $H_2$ mixtures as well for the production of acetyl-CoA and other desired products.

[0072] 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 a p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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 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.

[0073] The reductive tricarboxylic acid cycle coupled with carbon monoxide dehydrogenase and/or hydrogenase activities can also allow the conversion of CO, $CO_2$ and/or $H_2$ 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-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 $H_2$ by carbon monoxide dehydrogenase and hydrogenase are utilized to fix $CO_2$ via the reductive TCA cycle into acetyl-CoA or acetate. Acetate can be converted to acetyl-CoA by enzymes such as acetyl-CoA transferase, acetate kinase/phosphotransacetylase, and acetyl-CoA synthetase. Acetyl-CoA can be converted to the p-toluate, terepathalate, or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate precursors, glyceraldehyde-3-phosphate, phosphoenolpyruvate, and pyruvate, by pyruvate:ferredoxin 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 p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway, those skilled in the art will understand that a similar engineering design also can be performed with respect to introducing at least the nucleic acids encoding the reductive TCA pathway 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.

[0074] 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, p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate and any of the intermediate metabolites in the p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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 p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate biosynthetic pathways. Accordingly, the invention provides a non-naturally occurring microbial organism that produces and/or secretes p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate when grown on a carbohydrate or other carbon source and produces and/or secretes any of the intermediate metabolites shown in the p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway when grown on a carbohydrate or other carbon source. The p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate producing microbial

organisms of the invention can initiate synthesis from an intermediate. For example, a (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway intermediate can be 1-deoxy-D-xylulose-5-phosphate or C-methyl-D-erythritol-4-phosphate (see Example I and Figure 1). A p-toluate pathway intermediate can be, for example, 2,4-dihydroxy-5-methyl-6-[(phosphonooxy)methyl]oxane-2-carboxylate, 1,3-dihydroxy-4-methyl-5-oxocyclohexane-1-carboxylate, 5-hydroxy-4-methyl-3-oxocyclohex-1-ene-1-carboxylate, 3,5-dihydroxy-4-methylcyclohex-1-ene-1-carboxylate, 5-hydroxy-4-methyl-3-(phosphonooxy)cyclohex-1-ene-1-carboxylate, 5-[(1-carboxyeth-1-en-1-yl)oxy]-4-methyl-3-(phosphonooxy)cyclohex-1-ene-1-carboxylate, or 3-[(1-carboxyeth-1-en-1-yl)oxy]-4-methylcyclohexa-1,5-diene-1-carboxylate (see Example II and Figure 2). A terephthalate intermediate can be, for example, 4-carboxybenzyl alcohol or 4-carboxybenzaldehyde (see Example III and Figure 3).

[0075] 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 a p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway enzyme or protein in sufficient amounts to produce p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate. It is understood that the microbial organisms of the invention are cultured under conditions sufficient to produce p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate. Following the teachings and guidance provided herein, the non-naturally occurring microbial organisms of the invention can achieve biosynthesis of p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate resulting in intracellular concentrations between about 0.1-200 mM or more. Generally, the intracellular concentration of p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate is between about 3-150 mM, particularly between about 5-125 mM and more particularly between about 8-100 mM, including about 10 mM, 20 mM, 50 mM, 80 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.

[0076] 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. The p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate producers can synthesize p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate at intracellular concentrations of 5-10 mM or more as well as all other concentrations exemplified herein under substantially anaerobic conditions. It is understood that, even though the above description refers to intracellular concentrations, p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate producing microbial organisms can produce p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate intracellularly and/or secrete the product into the culture medium.

[0077] In addition to the culturing and fermentation conditions disclosed herein, growth conditions for achieving biosynthesis of p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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, praline betaine, dimethylthetin, dimethylslfonioproprionate, 3-dimethylsulfonio-2-methylproprionate, pipecolic 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.

[0078] 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.

[0079] As described herein, one exemplary growth condition for achieving biosynthesis of p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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 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 $N_2/CO_2$ mixture or other suitable non-oxygen gas or gases.

[0080] The culture conditions described herein can be scaled up and grown continuously for manufacturing of p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate. Exemplary growth procedures include, for example, fed-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 p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate. Generally, and as with non-continuous culture procedures, the continuous and/or near-continuous production of p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate will include culturing a non-naturally occurring p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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 be 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, 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.

[0081] Fermentation procedures are well known in the art. Briefly, fermentation for the biosynthetic production of p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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.

[0082] In addition to the above fermentation procedures using the p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate producers of the invention for continuous production of substantial quantities of p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate, the p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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 conversion to convert the product to other compounds, if desired.

[0083] 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 p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate.

[0084] 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.

[0085] 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 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.

[0086] 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 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 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.

[0087] These computational approaches are consistent with biological realities because biological 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 reduction in the size of the solution space, thereby enhancing the precision with which physiological performance or phenotype can be predicted.

[0088] 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.

[0089] 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.

[0090] 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.

[0091] 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®.

[0092] 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.

[0093] 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)).

[0094] 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.

[0095] As disclosed herein, a nucleic acid encoding a desired activity of a p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway can be introduced into

a host organism. In some cases, it can be desirable to modify an activity of a p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate pathway enzyme or protein to increase production of p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate. 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.

[0096] 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, $>10^4$). 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 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, or key intermediates; activity (kcat), to increases 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.

[0097] A number of exemplary methods have been developed for the mutagenesis and 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 a p-toluate, terephthalate or (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate 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 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 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 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)).

[0098] 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 Acids Res.* 27:e18 (1999); and Volkov et al., *Methods Enzymol.* 328:456-463 (2000)); Random Chimeragenesis 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 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 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 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 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.* 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)).

[0100] 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)).

[0101] 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 Verenium 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)).

[0102] 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.

[0103] 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
## Exemplary Pathway for Producing (2-Hydroxy-3-methyl-4-oxobutoxy)phosphonate

[0104] This example describes an exemplary pathway for producing the terephthalic acid (PTA) precursor (2-hydroxy-3-methyl-4-oxobutoxy)phosphonate (2H3M4OP).

[0105] The precursor to the p-toluate and PTA pathways is 2H3M4OP. This chemical can be derived from central metabolites glyceraldehyde-3-phosphate (G3P) and pyruvate in three enzymatic steps as shown in Figure 1. The first two steps are native to *E. coli* and other organisms that utilize the methylerythritol phosphate (non-mevalonate) pathway for isoprenoid biosynthesis. Pyruvate and G3P are first condensed to form 1-deoxy-D-xylulose 5-phosphate (DXP) by DXP synthase. Subsequent reduction and rearrangement of the carbon backbone is catalyzed by DXP reductoisomerase. Finally, a novel diol dehydratase transforms 2-C-methyl-D-erythritol-4-phosphate to the p-toluate precursor 2H3M4OP.

[0106] **A. 1- Deoxyxylulose-5-phosphate (DXP) synthase.** Pyruvate and G3P are condensed to form DXP by DXP synthase (EC 2.2.1.7). This enzyme catalyzes the first step in the non-mevalonate pathway of isoprenoid biosynthesis. The enzyme requires thiamine diphosphate as a cofactor, and also requires reduced FAD, although there is no net redox change. A crystal structure of the *E. coli* enzyme is available (Xiang et al., *J. Biol. Chem.* 282:2676-2682 (2007)(doi:M610235200, pii;10.1074/jbc.M610235200 doi). Other enzymes have been cloned and characterized in *M. tuberculosis* (Bailey et al., *Glycobiology* 12:813-820 (2002) and *Agrobacterium tumefaciens* (Lee et al., *J. Biotechnol.* 128:555-566 (2007)(doi:S0168-1656(06)00966-7, pii;10.1016/j.jbiotec.2006.11.009, doi). DXP synthase enzymes from *B. subtilis* and *Synechocystis* sp. PCC 6803 were cloned into *E. coli* (Harker and Bramley, *FEBS Lett.* 448:115-119 (1999)(doi:S0014-5793(99)00360-9, pii).

| Gene | GenBank Accession No. | GI No. | Organism |
|---|---|---|---|
| *dxs* | AAC73523.1 | 1786622 | *Escherichia coli* |
| *dxs* | P0A554.1 | 61222979 | *M. tuberculosis* |
| *dxs11* | AAP56243.1 | 37903541 | *Agrobacterium tumefaciens* |
| *dxs* | P54523.1 | 1731052 | *Bacillus subtilis* |
| *sll1945* | BAA17089.1 | 1652165 | *Synechocystis* sp. PCC 6803 |

[0107] B. 1-Deoxy-D-xylulose-5-phosphate reductoisomerase (EC 1.1.1.267). The NAD(P)H-dependent reduction and rearrangement of 1-deoxy-D-xylulose-5-phosphate (DXP) to 2-C-methyl-D-erythritol-4-phosphate is catalyzed by DXP reductoisomerase (DXR, EC 1.1.1.267) in the second step of the non-mevalonate pathway for isoprenoid biosynthesis. The NADPH-dependent *E. coli* enzyme is encoded by *dxr* (Takahashi et al., *Proc. Natl. Acad. Sci. USA* 95:9879-9884 (1998)). A recombinant enzyme from *Arabidopsis thaliana* was functionally expressed in *E. coli* (Carretero-Paulet et al., *Plant Physiol.* 129:1581-1591 (2002)(doi:10.1104/pp.003798 (doi). DXR enzymes from *Zymomonas mobilis* and *Mycobacterium tuberculosis* have been characterized and crystal structures are available (Grolle et al., *FEMS Microbiol. Lett.* 191:131-137 (2000)(doi:S0378-1097(00)00382-7, pii); Henriksson et al., *Acta Crystallogr. D. Biol. Crystallogr.* 62:807-813 (2006)(doi:S0907444906019196, pii;10.1107/S0907444906019196, doi). Most characterized DXR enzymes are strictly NADPH dependent, but the enzymes from *A. thaliana* and *M. tuberculosis* react with NADH at a reduced rate (Argyrou and Blanchard, *Biochemistry* 43:4375-4384 (2004)(doi:10.1021/bi049974k, doi); Rohdich et al., *FEBS J.* 273:4446-4458 (2006)(doi:EJB5446, pii;10.1111/j.1742-4658.2006.05446.x, doi.

| Gene | GenBank Accession No. | GI No. | Organism |
|---|---|---|---|
| *dxr* | AAC73284.1 | 1786369 | *Escherichia coli* |
| *dxr* | AAF73140.1 | 8131928 | *Arabisopsis thaliana* |
| *dxr* | CAB60758.1 | 6434139 | *Zymomonas mobilis* |
| *dxr* | NP_217386.2 | 57117032 | *Mycobacterium tuberculosis* |

[0108] C. 2-C-Methyl-D-erythritol-4-phosphate dehydratase. A diol dehydratase is required to convert 2-C-methyl-D-erythritol-4-phosphate into the p-toluate precursor (Altmiller and

Wagner, *Arch. Biochem. Biophys.* 138:160-170 (1970)). Although this transformation has not been demonstrated experimentally, several enzymes catalyze similar transformations including dihydroxy-acid dehydratase (EC 4.2.1.9), propanediol dehydratase (EC 4.2.1.28), glycerol dehydratase (EC 4.2.1.30) and myo-inositose dehydratase (EC 4.2.1.44).

[0109] Diol dehydratase or propanediol dehydratase enzymes (EC 4.2.1.28) capable of converting the secondary diol 2,3-butanediol to 2-butanone are excellent candidates for this transformation. Adenosylcobalamin-dependent diol dehydratases contain alpha, beta and gamma subunits, which are all required for enzyme function. Exemplary gene candidates are found in *Klebsiella pneumoniae* (Tobimatsu et al., *Biosci. Biotechnol. Biochem.* 62:1774-1777 (1998); Toraya et al.,. *Biochem. Biophys. Res. Commun.* 69:475-480 (1976)), *Salmonella typhimurium* (Bobik et al., *J. Bacteriol.* 179:6633-6639 (1997)), *Klebsiella oxytoca* (Tobimatsu et al., *J. Biol. Chem.* 270:7142-7148 (1995)) and *Lactobacillus collinoides* (Sauvageot et al., *FEMS Microbiol. Lett.* 209:69-74 (2002)). Methods for isolating diol dehydratase gene candidates in other organisms are well known in the art (see, for example, U.S. Patent No. 5,686,276).

| Gene | GenBank Accession No. | GI No. | Organism |
|---|---|---|---|
| *pddA* | BAA08099.1 | 868006 | *Klebsiella oxytoca* |
| *pddB* | BAA08100.1 | 868007 | *Klebsiella oxytoca* |
| *pddC* | BAA08101.1 | 868008 | *Klebsiella oxytoca* |
| *pduC* | AAB84102.1 | 2587029 | *Salmonella typhimurium* |
| *pduD* | AAB84103.1 | 2587030 | *Salmonella typhimurium* |
| *pduE* | AAB84104.1 | 2587031 | *Salmonella typhimurium* |
| *pduC* | CAC82541.1 | 18857678 | *Lactobacullus collinoides* |
| *pduD* | CAC82542.1 | 18857679 | *Lactobacullus collinoides* |
| *pduE* | CAD01091.1 | 18857680 | *Lactobacullus collinoides* |
| *pddA* | AAC98384.1 | 4063702 | *Klebsiella pneumoniae* |
| *pddB* | AAC98385.1 | 4063703 | *Klebsiella pneumoniae* |
| *pddC* | AAC98386.1 | 4063704 | *Klebsiella pneumoniae* |

[0110] Enzymes in the glycerol dehydratase family (EC 4.2.1.30) can also be used to dehydrate 2-C-methyl-D-erythritol-4-phosphate. Exemplary gene candidates encoded by *gldABC* and *dhaB123* in *Klebsiella pneumoniae* (WO 2008/137403) and (Toraya et al., *Biochem. Biophys. Res. Commun.* 69:475-480 (1976)), *dhaBCE* in *Clostridium pasteuranum* (Macis et al.,

*FEMS Microbiol Lett.* 164:21-28 (1998)) and *dhaBCE* in *Citrobacter freundii* (Seyfried et al., *J. Bacteriol.* 178:5793-5796 (1996)). Variants of the B12-dependent diol dehydratase from *K. pneumoniae* with 80- to 336-fold enhanced activity were recently engineered by introducing mutations in two residues of the beta subunit (Qi et al., *J. Biotechnol.* 144:43-50 (2009)(doi:S0168-1656(09)00258-2, pii;10.1016/j.jbiotec.2009.06.015, doi). Diol dehydratase enzymes with reduced inactivation kinetics were developed by DuPont using error-prone PCR (WO 2004/056963).

| Gene | GenBank Accession No. | GI No. | Organism |
|---|---|---|---|
| *gldA* | AAB96343.1 | 1778022 | *Klebsiella pneumoniae* |
| *gldB* | AAB96344.1 | 1778023 | *Klebsiella pneumoniae* |
| *gldC* | AAB96345.1 | 1778024 | *Klebsiella pneumoniae* |
| *dhaB1* | ABR78884.1 | 150956854 | *Klebsiella pneumoniae* |
| *dhaB2* | ABR78883.1 | 150956853 | *Klebsiella pneumoniae* |
| *dhaB3* | ABR78882.1 | 150956852 | *Klebsiella pneumoniae* |
| *dhaB* | AAC27922.1 | 3360389 | *Clostridium pasteuranum* |
| *dhaC* | AAC27923.1 | 3360390 | *Clostridium pasteuranum* |
| *dhaE* | AAC27924.1 | 3360391 | *Clostridium pasteuranum* |
| *dhaB* | P45514.1 | 1169287 | *Citrobacter freundii* |
| *dhaC* | AAB48851.1 | 1229154 | *Citrobacter freundii* |
| *dhaE* | AAB48852.1 | 1229155 | *Citrobacter freundii* |

[0111] If a B12-dependent diol dehydratase is utilized, heterologous expression of the corresponding reactivating factor is recommended. B12-dependent diol dehydratases are subject to mechanism-based suicide activation by substrates and some downstream products. Inactivation, caused by a tight association with inactive cobalamin, can be partially overcome by diol dehydratase reactivating factors in an ATP-dependent process. Regeneration of the B12 cofactor requires an additional ATP. Diol dehydratase regenerating factors are two-subunit proteins. Exemplary candidates are found in *Klebsiella oxytoca* (Mori et al., *J. Biol. Chem.* 272:32034-32041 (1997)), *Salmonella typhimurium* (Bobik et al., *J. Bacteriol.* 179:6633-6639 (1997); Chen et al., *J. Bacteriol.* 176:5474-5482 (1994)), *Lactobacillus collinoides* (Sauvageot et al., *FEMS Microbiol. Lett.* 209:69-74 (2002)), and *Klebsiella pneumonia* (WO 2008/137403).

| Gene | GenBank Accession No. | GI No. | Organism |
|---|---|---|---|
| *ddrA* | AAC15871.1 | 3115376 | *Klebsiella oxytoca* |
| *ddrB* | AAC15872.1 | 3115377 | *Klebsiella oxytoca* |
| *pduG* | AAL20947.1 | 16420573 | *Salmonella typhimurium* |
| *pduH* | AAL20948.1 | 16420574 | *Salmonella typhimurium* |
| *pduG* | YP_002236779 | 206579698 | *Klebsiella pneumonia* |
| *pduH* | YP_002236778 | 206579863 | *Klebsiella pneumonia* |
| *pduG* | CAD01092 | 29335724 | *Lactobacillus collinoides* |
| *pduH* | CAD01093 | 29335725 | *Lactobacillus collinoides* |

[0112] B12-independent diol dehydratase enzymes utilize S-adenosylmethionine (SAM) as a cofactor, function under strictly anaerobic conditions, and require activation by a specific activating enzyme (Frey et al., *Chem. Rev.* 103:2129-2148 (2003)). The glycerol dehydrogenase and corresponding activating factor of *Clostridium butyricum*, encoded by *dhaB1* and *dhaB2*, have been well-characterized (O'Brien et al., *Biochemistry* 43:4635-4645 (2004); Raynaud et al., *Proc. Natl. Acad. Sci USA* 100:5010-5015 (2003)). This enzyme was recently employed in a 1,3-propanediol overproducing strain of *E. coli* and was able to achieve very high titers of product (Tang et al., *Appl. Environ. Microbiol.* 75:1628-1634 (2009)(doi:AEM.02376-08, pii;10.1128/AEM.02376-08, doi). An additional B12-independent diol dehydratase enzyme and activating factor from *Roseburia inulinivorans* was shown to catalyze the conversion of 2,3-butanediol to 2-butanone (US publication 2009/09155870).

| Gene | GenBank Accession No. | GI No. | Organism |
|---|---|---|---|
| *dhaB1* | AAM54728.1 | 27461255 | *Clostridium butyricum* |
| *dhaB2* | AAM54729.1 | 27461256 | *Clostridium butyricum* |
| *rdhtA* | ABC25539.1 | 83596382 | *Roseburia inulinivorans* |
| *rdhtB* | ABC25540.1 | 83596383 | *Roseburia inulinivorans* |

[0113] Dihydroxy-acid dehydratase (DHAD, EC 4.2.1.9) is a B12-independent enzyme participating in branched-chain amino acid biosynthesis. In its native role, it converts 2,3-dihydroxy-3-methylvalerate to 2-keto-3-methyl-valerate, a precursor of isoleucine. In valine biosynthesis, the enzyme catalyzes the dehydration of 2,3-dihydroxy-isovalerate to 2-oxoisovalerate. The DHAD from *Sulfolobus solfataricus* has a broad substrate range, and activity of a recombinant enzyme expressed in *E. coli* was demonstrated on a variety of aldonic

acids (Kim and Lee, *J. Biochem.* 139:591-596 (2006)(doi:139/3/591, pii;10.1093/jb/mvj057, doi). The *S. solfataricus* enzyme is tolerant of oxygen unlike many diol dehydratase enzymes. The *E. coli* enzyme, encoded by *ilvD,* is sensitive to oxygen, which inactivates its iron-sulfur cluster (Flint et al., *J. Biol. Chem.* 268:14732-14742 (1993)). Similar enzymes have been characterized in *Neurospora crassa* (Altmiller and Wagner, *Arch. Biochem. Biophys.* 138:160-170 (1970)) and *Salmonella typhimurium* (Armstrong et al., *Biochim. Biophys. Acta* 498:282-293 (1977)).

| Gene | GenBank Accession No. | GI No. | Organism |
|---|---|---|---|
| *ilvD* | NP_344419.1 | 15899814 | *Sulfolobus solfataricus* |
| *ilvD* | AAT48208.1 | 48994964 | *Escherichia coli* |
| *ilvD* | NP_462795.1 | 16767180 | *Salmonella typhimurium* |
| *ilvD* | XP_958280.1 | 85090149 | *Neurospora crassa* |

[0114] The diol dehydratase myo-inosose-2-dehydratase (EC 4.2.1.44) is another exemplary candidate. Myo-inosose is a six-membered ring containing adjacent alcohol groups. A purified enzyme encoding myo-inosose-2-dehydratase functionality has been studied in *Klebsiella aerogenes* in the context of myo-inositol degradation (Berman and Magasanik, *J. Biol. Chem.* 241:800-806 (1966)), but has not been associated with a gene to date. The myo-inosose-2-dehydratase of *Sinorhizobium fredii* was cloned and functionally expressed in *E. coli* (Yoshida et al., *Biosci .Biotechnol. Biochem.* 70:2957-2964 (2006)(doi:JST.JSTAGE/bbb/60362, pii). A similar enzyme from *B. subtilis*, encoded by *iolE*, has also been studied (Yoshida et al., *Microbiology* 150:571-580 (2004)).

| Gene | GenBank Accession No. | GI No. | Organism |
|---|---|---|---|
| *iolE* | P42416.1 | 1176989 | *Bacillus subtilis* |
| *iolE* | AAX24114.1 | 60549621 | *Sinorhizobium fredii* |

## EXAMPLE II
## Exemplary Pathway for Synthesis of p-Toluate from (2-Hydroxy-3-methyl-4-oxobutoxy)phosphonate by Shikimate Pathway Enzymes

[0115] This example describes exemplary pathways for synthesis of p-toluate using shikimate pathway enzymes.

[0116] The chemical structure of p-toluate closely resembles p-hydroxybenzoate, a precursor of the electron carrier ubiquinone. 4-Hydroxybenzoate is synthesized from central metabolic precursors by enzymes in the shikimate pathway, found in bacteria, plants and fungi. The shikimate pathway is comprised of seven enzymatic steps that transform D-erythrose-4-phosphate (E4P) and phosphoenolpyruvate (PEP) to chorismate. Pathway enzymes include 2-dehydro-3-deoxyphosphoheptonate (DAHP) synthase, dehydroquinate (DHQ) synthase, DHQ dehydratase, shikimate dehydrogenase, shikimate kinase, 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase and chorismate synthase. In the first step of the pathway, D-erythrose-4-phosphate and phosphoenolpyruvate are joined by DAHP synthase to form 3-deoxy-D-arabino-heptulosonate-7-phosphate. This compound is then dephosphorylated, dehydrated and reduced to form shikimate. Shikimate is converted to chorismate by the actions of three enzymes: shikimate kinase, 3-phosphoshikimate-2-carboxyvinyltransferase and chorismate synthase. Subsequent conversion of chorismate to 4-hydroxybenzoate is catalyzed by chorismate lyase.

[0117] The synthesis of p-toluate proceeds in an analogous manner as shown in Figure 2. The pathway originates with PEP and 2H3M4OP, a compound analogous to E4P with a methyl group in place of the 3-hydroxyl group of E4P. The hydroxyl group of E4P does not directly participate in the chemistry of the shikimate pathway reactions, so the methyl-substituted 2H3M4OP precursor is expected to react as an alternate substrate. Directed or adaptive evolution can be used to improve preference for 2H3M4OP and downstream derivatives as substrates. Such methods are well-known in the art.

[0118] Strain engineering strategies for improving the efficiency of flux through shikimate pathway enzymes are also applicable here. The availability of the pathway precursor PEP can be increased by altering glucose transport systems (Yi et al., *Biotechnol. Prog.* 19:1450-1459 (2003)(doi:10.1021/bp0340584, doi). 4-Hydroxybenzoate-overproducing strains were engineered to improve flux through the shikimate pathway by means of overexpression of a feedback-insensitive isozyme of 3-deoxy-D-arabinoheptulosonic acid-7-phosphate synthase (Barker and Frost, *Biotechnol. Bioeng.* 76:376-390 (2001)(doi:10.1002/bit.10160, pii). Additionally, expression levels of shikimate pathway enzymes and chorismate lyase were enhanced. Similar strategies can be employed in a strain for overproducing p-toluate.

[0119] A. 2-Dehydro-3-deoxyphosphoheptonate synthase (EC 2.5.1.54). The condensation of D-erythrose-4-phosphate and phosphoenolpyruvate is catalyzed by 2-dehydro-3-deoxyphosphoheptonate (DAHP) synthase ( EC 2.5.1.54). Three isozymes of this enzyme are encoded in the *E. coli* genome by *aroG*, *aroF* and *aroH* and are subject to feedback inhibition by phenylalanine, tyrosine and tryptophan, respectively. In wild-type cells grown on minimal medium, the *aroG*, *aroF* and *aroH* gene products contributed 80%, 20% and 1% of DAHP synthase activity, respectively (Hudson and Davidson, *J. Mol. Biol.* 180:1023-1051 (1984)(doi:0022-2836(84)90269-9, pii). Two residues of AroG were found to relieve inhibition by phenylalanine (Kikuchi et al., *Appl. Environ. Microbiol.* 63:761-762 (1997)). The feedback inhibition of AroF by tyrosine was removed by a single base-pair change (Weaver and Herrmann, *J. Bacteriol.* 172:6581-6584 (1990)). The tyrosine-insensitive DAHP synthase was overexpressed in a 4-hydroxybenzoate-overproducing strain of *E. coli* (Barker and Frost, *Biotechnol. Bioeng.* 76:376-390 (2001)(doi:10.1002/bit.10160, pii). The *aroG* gene product was shown to accept a variety of alternate 4- and 5-carbon length substrates (Sheflyan et al., *J. Am. Chem. Soc.* 120(43):11027-11032 (1998); Williamson et al., *Bioorg. Med. Chem. Lett.* 15:2339-2342 (2005)(doi:S0960-894X(05)00273-8, pii;10.1016/j.bmcl.2005.02.080, doi). The enzyme reacts efficiently with (3S)-2-deoxyerythrose-4-phosphate, a substrate analogous to D-erythrose-4-phosphate but lacking the alcohol at the 2-position ( Williamson et al., *supra* 2005). Enzymes from *Helicobacter pylori* and *Pyrococcus furiosus* also accept this alternate substrate (Schofield et al., *Biochemistry* 44:11950-11962 (2005)(doi:10.1021/bi050577z, doi; Webby et al., *Biochem. J.* 390:223-230 2005)(doi:BJ20050259, pii;10.1042/BJ20050259, doi) and have been expressed in *E. coli*. An evolved variant of DAHP synthase, differing from the wild type *E. coli* AroG enzyme by 7 amino acids, was shown to exhibit a 60-fold improvement in Kcat/$K_M$ (Ran and Frost, *J. Am. Chem. Soc.* 129:6130-6139 (2007)(doi:10.1021/ja067330p, doi).

| Gene | GenBank Accession No. | GI No. | Organism |
|---|---|---|---|
| *aroG* | AAC73841.1 | 1786969 | *Escherichia coli* |
| *aroF* | AAC75650.1 | 1788953 | *Escherichia coli* |
| *aroH* | AAC74774.1 | 1787996 | *Escherichia coli* |
| *aroF* | Q9ZMU5 | 81555637 | *Helicobacter pylori* |
| *PF1690* | NP_579419.1 | 18978062 | *Pyrococcus furiosus* |

[0120] B. 3-Dehydroquinate synthase (EC 4.2.3.4). The dephosphorylation of substrate (2)(2,4-dihydroxy-5-methyl-6-[(phosphonooxy)methyl]oxane-2-carboxylate) to substrate (3)(1,3-dihydroxy-4-methylcylohex-1-ene-1-carboxylate) as shown in Figure 2 is analogous to the dephosphorylation of 3-deoxy-arabino-heptulonate-7-phosphate by 3-dehydroquinate synthase. The enzyme has been characterized in *E. coli* (Mehdi et al., *Methods Enzymol.* 142:306-314 (1987), *B. subtilis* (Hasan and Nester, *J. Biol. Chem.* 253:4999-5004 (1978)) and *Mycobacterium tuberculosis H37Rv* (de Mendonca et al., *J. Bacteriol.* 189:6246-6252 (2007)(doi:JB.00425-07, pii;10.1128/JB.00425-07, doi). The *E. coli* enzyme is subject to inhibition by L-tyrosine (Barker and Frost, *Biotechnol. Bioeng.* 76:376-390 2001)(doi:10.1002/bit.10160, pii).

| Gene | GenBank Accession No. | GI No. | Organism |
|---|---|---|---|
| *aroB* | AAC76414.1 | 1789791 | *Escherichia coli* |
| *aroB* | NP_390151.1 | 16079327 | *Bacillus subtilis* |
| *aroB* | CAB06200.1 | 1781064 | *Mycobacterium tuberculosis* |

[0121] C. 3-Dehydroquinate dehydratase (EC 4.2.1.10). 3-Dehydroquinate dehydratase, also termed 3-dehydroquinase (DHQase), naturally catalyzes the dehydration of 3-dehydroquinate to 3-dehydroshikimate, analogous to step C in the p-toluate pathway of Figure 2. DHQase enzymes can be divided into two classes based on mechanism, stereochemistry and sequence homology (Gourley et al., *Nat. Struct. Biol.* 6:521-525. (1999)(doi:10.1038/9287, doi). Generally the type 1 enzymes are involved in biosynthesis, while the type 2 enzymes operate in the reverse (degradative) direction. Type 1 enzymes from *E. coli* (Kinghorn et al., *Gene* 14:73-80. 1981)(doi:0378-1119(81)90149-9, pii), *Salmonella typhi* (Kinghorn et al., *supra* 1981; Servos et al., *J. Gen. Microbiol.* 137:147-152 (1991)) and *B. subtilis* (Warburg et al., *Gene* 32:57-66 1984)(doi:0378-1119(84)90032-5, pii) have been cloned and characterized. Exemplary type II 3-dehydroquinate dehydratase enzymes are found in *Mycobacterium tuberculosis*, *Streptomyces coelicolor* (Evans et al., *FEBS Lett.* 530:24-30 (2002)) and *Helicobacter pylori* (Lee et al., *Proteins* 51:616-7 (2003)).

| Gene | GenBank Accession No. | GI No. | Organism |
|---|---|---|---|
| *aroD* | AAC74763.1 | 1787984 | *Escherichia coli* |
| *aroD* | P24670.2 | 17433709 | *Salmonella typhi* |
| *aroC* | NP_390189.1 | 16079365 | *Bacillus subtilis* |
| *aroD* | P0A4Z6.2 | 61219243 | *Mycobacterium tuberculosis* |
| *aroQ* | P15474.3 | 8039781 | *Streptomyces coelicolor* |
| *aroQ* | Q48255.2 | 2492957 | *Helicobacter pylori* |

[0122] D. Shikimate dehydrogenase (EC 1.1.1.25). Shikimate dehydrogenase catalyzes the NAD(P)H dependent reduction of 3-dehydroshikimate to shikimate, analogous to Step D of Figure 2. The *E. coli* genome encodes two shikimate dehydrogenase paralogs with different cofactor specificities. The enzyme encoded by *aroE* is NADPH specific, whereas the *ydiB* gene product is a quinate/shikimate dehydrogenase which can utilize NADH (preferred) or NADPH as a cofactor (Michel et al., *J. Biol. Chem.* 278:19463-19472 (2003)(doi:10.1074/jbc.M300794200, doi;M300794200, pii). NADPH-dependent enzymes from *Mycobacterium tuberculosis* (Zhang et al., *J. Biochem. Mol. Biol.* 38:624-631 (2005)), *Haemophilus influenzae* (Ye et al., *J. Bacteriol.* 185:4144-4151 (2003)) and *Helicobacter pylori* (Han et al., *FEBS J.* 273:4682-4692 (2006)(doi:EJB5469, pii;10.1111/j.1742-4658.2006.05469.x, doi) have been functionally expressed in *E. coli*.

| Gene | GenBank Accession No. | GI No. | Organism |
|---|---|---|---|
| *aroE* | AAC76306.1 | 1789675 | *Escherichia coli* |
| *ydiB* | AAC74762.1 | 1787983 | *Escherichia coli* |
| *aroE* | NP_217068.1 | 15609689 | *Mycobacterium tuberculosis* |
| *aroE* | P43876.1 | 1168510 | *Haemophilus influenzae* |
| *aroE* | AAW22052.1 | 56684731 | *Helicobacter pylori* |

[0123] E. Shikimate kinase (EC 2.7.1.71). Shikimate kinase catalyzes the ATP-dependent phosphorylation of the 3-hydroxyl group of shikimate analogous to Step E of Figure 2. Two shikimate kinase enzymes are encoded by *aroK* (SK1) and *aroL* (SK2) in *E. coli* (DeFeyter and Pittard, J. *Bacteriol.* 165:331-333 (1986); Lobner-Olesen and Marinus, *J. Bacteriol.* 174:525-529 (1992)). The Km of SK2, encoded by *aroL*, is 100-fold lower than that of SK1, indicating that this enzyme is responsible for aromatic biosynthesis (DeFeyter et al., *supra* 1986). Additional

shikimate kinase enzymes from *Mycobacterium tuberculosis* (Gu et al., *J. Mol. Biol.* 319:779-789 (2002)(doi:10.1016/S0022-2836(02)00339-X, doi;S0022-2836(02)00339-X, pii); Oliveira et al., *Protein Expr. Purif.* 22:430-435 (2001)(doi:10.1006/prep.2001.1457, doi;S1046-5928(01)91457-3, pii), *Helicobacter pylori* (Cheng et al., *J. Bacteriol.* 187:8156-8163 (2005)(doi:187/23/8156, pii;10.1128/JB.187.23.8156-8163.2005, doi) and *Erwinia chrysanthemi* (Krell et al., *Protein Sci.* 10:1137-1149 (2001)(doi:10.1110/ps.52501, doi) have been cloned in *E. coli*.

| Gene | GenBank Accession No. | GI No. | Organism |
|---|---|---|---|
| *aroK* | YP_026215.2 | 90111581 | *Escherichia coli* |
| *aroL* | NP_414922.1 | 16128373 | *Escherichia coli* |
| *aroK* | CAB06199.1 | 1781063 | *Mycobacterium tuberculosis* |
| *aroK* | NP_206956.1 | 15644786 | *Helicobacter pylori* |
| SK | CAA32883.1 | 42966 | *Erwinia chrysanthemi* |

[0124] F. 3-Phosphoshikimate-2-carboxyvinyltransferase (EC 2.5.1.19). 3-Phosphoshikimate-2-carboxyvinyltransferase, also known as 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), catalyzes the transfer of the enolpyruvyl moiety of phosphoenolpyruvate to the 5-hydroxyl of shikimate-3-phosphate. The enzyme is encoded by *aroA* in *E. coli* (Anderson et al., *Biochemistry* 27:1604-1610 (1988)). EPSPS enzymes from *Mycobacterium tuberculosis* (Oliveira et al., *Protein Expr. Purif.* 22:430-435 (2001)(doi:10.1006/prep.2001.1457, doi;S1046-5928(01)91457-3, pii), *Dunaliella salina* (Yi et al., *J. Microbiol.* 45:153-157 (2007)(doi:2519, pii) and *Staphylococcus aureus* (Priestman et al., *FEBS Lett.* 579:728-732 (2005)(doi:S0014-5793(05)00012-8, pii;10.1016/j.febslet.2004.12.057, doi) have been cloned and functionally expressed in *E. coli*.

| Gene | GenBank Accession No. | GI No. | Organism |
|---|---|---|---|
| *aroA* | AAC73994.1 | 1787137 | *Escherichia coli* |
| *aroA* | AAA25356.1 | 149928 | *Mycobacterium tuberculosis* |
| *aroA* | AAA71897.1 | 152956 | *Staphylococcus aureus* |
| *aroA* | ABM68632.1 | 122937807 | *Dunaliella salina* |

[0125] G. Chorismate synthase (EC 4.2.3.5). Chorismate synthase is the seventh enzyme in the shikimate pathway, catalyzing the transformation of 5-enolpyruvylshikimate-3-phosphate to

chorismate. The enzyme requires reduced flavin mononucleotide (FMN) as a cofactor, although the net reaction of the enzyme does not involve a redox change. In contrast to the enzyme found in plants and bacteria, the chorismate synthase in fungi is also able to reduce FMN at the expense of NADPH (Macheroux et al., *Planta* 207:325-334 (1999)). Representative monofunctional enzymes are encoded by *aroC* of *E. coli* (White et al., *Biochem. J.* 251:313-322 (1988)) and *Streptococcus pneumoniae* (Maclean and Ali, *Structure* 11:1499-1511 (2003)(doi:S0969212603002648, pii). Bifunctional fungal enzymes are found in *Neurospora crassa* (Kitzing et al., *J. Biol. Chem.* 276:42658-42666 (2001)(doi:10.1074/jbc.M107249200, doi;M107249200, pii) and *Saccharomyces cerevisiae* (Jones et al., *Mol. Microbiol.* 5:2143-2152 (1991)).

| Gene | GenBank Accession No. | GI No. | Organism |
|---|---|---|---|
| *aroC* | NP_416832.1 | 16130264 | *Escherichia coli* |
| *aroC* | ACH47980.1 | 197205483 | *Streptococcus pneumoniae* |
| *U25818.1:19..1317* | AAC49056.1 | 976375 | *Neurospora crassa* |
| *ARO2* | CAA42745.1 | 3387 | *Saccharomyces cerevisiae* |

[0126] H. Chorismate lyase (EC 4.1.3.40). Chorismate lyase catalyzes the first committed step in ubiquinone biosynthesis: the removal of pyruvate from chorismate to form 4-hydroxybenzoate. The enzymatic reaction is rate-limited by the slow release of the 4-hydroxybenzoate product (Gallagher et al., *Proteins* 44:304-311 (2001)(doi:10.1002/prot.1095, pii), which is thought to play a role in delivery of 4-hydroxybenzoate to downstream membrane-bound enzymes. The chorismate lyase of *E. coli* was cloned and characterized and the enzyme has been crystallized (Gallagher et al., *supra* 2001; Siebert et al., *FEBS Lett.* 307:347-350 (1992)(doi:0014-5793(92)80710-X, pii). Structural studies implicate the G90 residue as contributing to product inhibition (Smith et al., *Arch. Biochem. Biophys.* 445:72-80 (2006)(doi:S0003-9861(05)00446-7, pii;10.1016/j.abb.2005.10.026, doi). Modification of two surface-active cysteine residues reduced protein aggregation (Holden et al., *Biochim. Biophys. Acta* 1594:160-167 (2002)(doi:S0167483801003028, pii). A recombinant form of the *Mycobacterium tuberculosis* chorismate lyase was cloned and characterized in *E. coli* (Stadthagen et al., *J. Biol. Chem.* 280:40699-40706 2005)(doi:M508332200, pii;10.1074/jbc.M508332200, doi).

| Gene | GenBank Accession No. | GI No. | Organism |
|---|---|---|---|
| *ubiC* | AAC77009.2 | 87082361 | *Escherichia coli* |
| *Rv2949c* | NP_217465.1 | 15610086 | *Mycobacterium tuberculosis* |

[0127] B-F. Multifunctional AROM protein. In most bacteria, the enzymes of the shikimate pathway are encoded by separate polypeptides. In microbial eukaryotes, five enzymatic functions are catalyzed by a polyfunctional protein encoded by a pentafunctional supergene (Campbell et al., *Int. J. Parasitol.* 34:5-13 (2004)(doi:S0020751903003102, pii). The multifunctional AROM protein complex catalyzes reactions analogous to reactions B-F of Figure 2. The AROM protein complex has been characterized in fungi including *Aspergillus nidulans*, *Neurospora crassa*, *Saccharomyces cerevisiae* and *Pneumocystis carinii* (Banerji et al., *J. Gen. Microbiol.* 139:2901-2914 (1993); Charles et al., *Nucleic Acids Res.* 14:2201-2213 (1986); Coggins et al., *Methods Enzymol.* 142:325-341 (1987); Duncan, K., *Biochem. J.* 246:375-386 (1987)). Several components of AROM have been shown to function independently as individual polypeptides. For example, dehydroquinate synthase (DHQS) forms the amino-terminal domain of AROM, and can function independently when cloned into *E. coli* (Moore et al., *Biochem. J.* 301 ( Pt 1):297-304 (1994)). Several crystal structures of AROM components from *Aspergillus nidulans* provide insight into the catalytic mechanism (Carpenter et al., *Nature* 394:299-302 (1998)(doi:10.1038/28431, doi).

| Gene | GenBank Accession No. | GI No. | Organism |
|---|---|---|---|
| *AROM* | P07547.3 | 238054389 | *Aspergillus nidulans* |
| *AROM* | P08566.1 | 114166 | *Saccharomyces cerevisiae* |
| *AROM* | P07547.3 | 238054389 | *Aspergillus nidulans* |
| *AROM* | Q12659.1 | 2492977 | *Pneumocystis carinii* |

## EXAMPLE III

### Exemplary Pathway for Enzymatic Transformation of p-Toluate to Terephthalic Acid

[0128] This example describes exemplary pathways for conversion of p-toluate to terephthalic acid (PTA).

[0129] P-toluate can be further transformed to PTA by oxidation of the methyl group to an acid in three enzymatic steps as shown in Figure 3. The pathway is comprised of a p-toluate methyl-monooxygenase reductase, a 4-carboxybenzyl alcohol dehydrogenase and a 4-

carboxybenzyl aldehyde dehydrogenase. In the first step, p-toluate methyl-monooxyngenase oxidizes p-toluate to 4-carboxybenzyl alcohol in the presence of $O_2$. The *Comamonas testosteroni* enzyme (*tsaBM*), which also reacts with 4-toluene sulfonate as a substrate, has been purified and characterized (Locher et al., *J. Bacteriol.* 173:3741-3748 (1991)). 4-Carboxybenzyl alcohol is subsequently converted to an aldehyde by 4-carboxybenzyl alcohol dehydrogenase (*tsaC*). The aldehyde to acid transformation is catalyzed by 4-carboxybenzaldehyde dehydrogenase (*tsaD*). Enzymes catalyzing these reactions are found in *Comamonas testosteroni* T-2, an organism capable of utilizing p-toluate as the sole source of carbon and energy (Junker et al., *J. Bacteriol.* 179:919-927 (1997)). Additional genes to transform p-toluate to PTA can be found by sequence homology, in particular to proteobacteria in the genera *Burkholderia*, *Alcaligenes*, *Pseudomonas*, *Shingomonas* and *Comamonas* (US Patent No. 6,187,569 and US publication 2003/0170836). Genbank identifiers associated with the *Comamonas testosteroni* enzymes are listed below.

| Gene | GenBank Accession No. | GI No. | Organism |
|---|---|---|---|
| *tsaB* | AAC44805.1 | 1790868 | *Comamonas testosteroni* |
| *tsaM* | AAC44804.1 | 1790867 | *Comamonas testosteroni* |
| *tsaC* | AAC44807.1 | 1790870 | *Comamonas testosteroni* |
| *tsaD* | AAC44808.1 | 1790871 | *Comamonas testosteroni* |

[0130] 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 reference to the examples provided above, it should be understood that various modifications can be made without departing from the spirit of the invention.

## CLAIMS

1. A non-naturally occurring microbial organism having carbon monoxide dehydrogenase or hydrogenase activity and a reductive tricarboxylic acid (rTCA) pathway, wherein said microbial organism comprises (1) at least one exogenous nucleic acid encoding a rTCA pathway enzyme expressed in a sufficient amount to convert CO, $CO_2$ or $H_2$ to acetyl-CoA, wherein said rTCA enzyme is selected from the group consisting of an ATP citrate-lyase, a citrate lyase, an aconitase, an isocitrate dehydrogenase, an alpha-ketoglutarate:ferredoxin oxidoreductase, a succinyl-CoA synthetase, a succinyl-CoA transferase, a fumarate reductase, a fumarase, and a malate dehydrogenase; and (2) at least one exogenous nucleic acid encoding a biosynthetic product pathway enzyme expressed in a sufficient amount to produce said biosynthetic product, wherein the microbial organism does not require light for growth.

2. A non-naturally occurring microbial organism comprising

a rTCA pathway comprising at least one exogenous nucleic acid encoding a rTCA pathway enzyme expressed in a sufficient amount to convert (i) CO, (ii) $CO_2$ and $H_2$, (iii) CO and $CO_2$, (iv) synthesis gas comprising CO and $H_2$, or (v) synthesis gas comprising CO, $CO_2$, and $H_2$; to acetyl-CoA; wherein said at rTCA pathway enzyme is selected from the group consisting of an ATP citrate-lyase, a citrate lyase, an aconitase, an isocitrate dehydrogenase, an alpha-ketoglutarate:ferredoxin oxidoreductase, a succinyl-CoA synthetase, a succinyl-CoA transferase, a fumarate reductase, a fumarase and a malate dehydrogenase;

a reducing equivalent producing pathway comprising at least one exogenous nucleic acid encoding a reducing equivalent producing enzyme selected from the group consisting of a carbon monoxide dehydrogenase, a hydrogenase and a NAD(P)H:ferredoxin oxidoreductase; and

at least one exogenous nucleic acid encoding a biosynthetic product pathway enzyme expressed in a sufficient amount to produce a biosynthetic product, wherein the microbial organism does not require acetyl-CoA synthase.

3. The non-naturally occurring microbial organism of claim 1 or 2, wherein said microbial organism further comprises a ferredoxin.

4. The non-naturally occurring microbial organism of claim 1 or 2, wherein said rTCA pathway enzyme is selected from the group consisting of an ATP citrate-lyase, a citrate lyase, an alpha-ketoglutarate:ferredoxin oxidoreductase, and a fumarate reductase.

5. The non-naturally occurring microbial organism of claim 1 or 2, wherein said rTCA pathway enzyme is an aconitase or an isocitrate dehydrogenase.

6. The non-naturally occurring microbial organism of any one of claims 1 to 5, wherein said microbial organism is a species of bacteria, yeast or fungus.

7. A non-naturally occurring microbial organism comprising a reducing equivalent producing pathway comprising at least one exogenous nucleic acid encoding a reducing equivalent producing enzyme selected from the group consisting of a carbon monoxide dehydrogenase, a hydrogenase and a NAD(P)H:ferredoxin oxidoreductase; and

   at least one exogenous nucleic acid encoding a biosynthetic product pathway enzyme expressed in a sufficient amount to produce said biosynthetic product, wherein the microbial organism does not require acetyl-CoA synthase.

8. The non-naturally occurring microbial organism of any one of claims 2 to 7 comprising at least one exogenous nucleic acid encoding a reducing equivalent producing enzyme carbon monoxide dehydrogenase.

9. The non-naturally occurring microbial organism of any one of claims 2 to 8 comprising at least one exogenous nucleic acid encoding a reducing equivalent producing enzyme carbon monoxide dehydrogenase and a ferredoxin.