(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 May 2017 (18.05.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/083351 A1

(51) International Patent Classification:
C12N 1/20 (2006.01) *A01N 41/08* (2006.01)
C12N 1/21 (2006.01) *C07C 403/24* (2006.01)
C12N 15/75 (2006.01) *A23L 33/175* (2016.01)
A23K 10/10 (2016.01) *A23L 5/44* (2016.01)

(21) International Application Number:
PCT/US2016/061081

(22) International Filing Date:
9 November 2016 (09.11.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/252,971 9 November 2015 (09.11.2015) US

(71) Applicant: **KNIPBIO, INC.** [US/US]; 110 Canal Street, Lowell, Massachusetts 01854 (US).

(72) Inventors: **FEINBERG, Lawrence F.**; 110 Canal Street, Lowell, Massachusetts 01854 (US). **MARX, Christopher J.**; 110 Canal Street, Lowell, Massachusetts 01854 (US). **WALL, Max A.**; 110 Canal Street, Lowell, Massachusetts 01854 (US). **SMITH, Daniel R.**; 110 Canal Street, Lowell, Massachusetts 01854 (US). **PUJOL-BAXLEY, Catherine J.**; 110 Canal Street, Lowell, Massachusetts 01854 (US). **MCAVOY, Bonnie D.**; 110 Canal Street, Lowell, Massachusetts 01854 (US).

(74) Agent: **JACOBSON, Jill A.**; FisherBroyles, LLP, 2784 Homestead Rd. #321, Santa Clara, California 95051 (US).

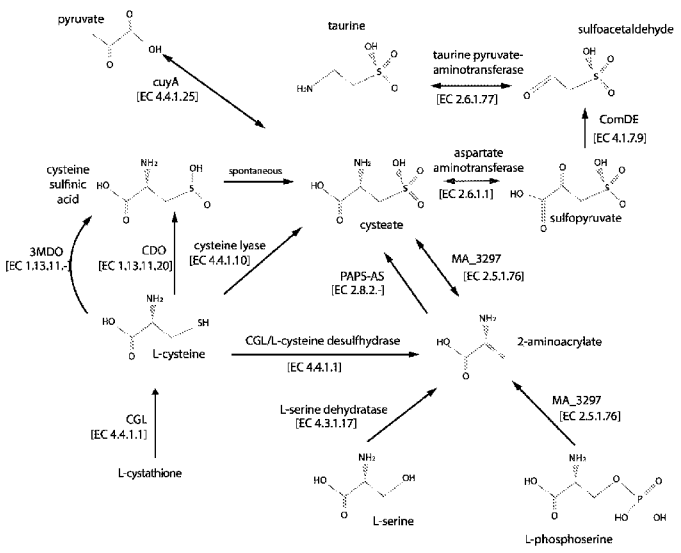
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS,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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,

[Continued on next page]

(54) Title: HETEROLOGOUS EXPRESSION OF TAURINE IN MICROORGANISMS

Figure 12



(57) Abstract: Non-naturally occurring microorganisms are provided that produce taurine and/or taurine precursors, e.g., hypotaurine, sulfoacetaldehyde, or cysteate, utilizing exogenously added enzyme activities. Methods of producing taurine and/or taurine precursors in microbial cultures, and feed and nutritional supplement compositions that include taurine and/or taurine precursors produced in the microbial cultures, such as taurine- and/or taurine precursor-containing biomass, are also provided.

WO 2017/083351 A1

SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

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

HETEROLOGOUS EXPRESSION OF TAURINE IN MICROORGANISMS

CROSS-REFERENCE TO RELATED APPLICATIONS

[01] This application claims the benefit of U.S. Provisional Application No. 62/252,971, filed November 9, 2015, which is incorporated herein by reference in its entirety.

FIELD OF THE INVENTION

[02] The invention relates to recombinant production of taurine in microbial organisms, and use in feed compositions, in particular for aquaculture, animal feeds and human nutrition.

BACKGROUND

[03] Taurine (2-aminoethanesulfonic acid) is a 2-carbon (β)-amino acid found in a broad swath of organisms, from birds to mammals, fish to plants, fungi and bacteria (McCusker et al. (2014) *J Nutr Sci* 3:e39). While some proteins contain taurine, it does not form peptide bonds due to the lack of a carboxyl group. In fact, taurine is the most abundant free amino acid (FAA) present in animal tissues, constituting 19% of the FAA in the brain, 50% in kidney, and 53% in muscle. (Brosnan et al. (2006) *J Nutr.* 136(6):16365 –16405)

[04] Taurine is critical in many basic cellular processes, including osmoregulation, membrane stabilization, and antioxidation. (Honjoh et al. (2010) *Amino Acids.* 38(4):1173–1183; Takeuchi et al. (2000) *Biochim Biophys Acta.* 1464(2):219–230) In addition, taurine participates in a variety of more complex physiological functions, such as bile conjugation and calcium signaling. (Salze et al. (2015) *Aquaculture* 437:215–229). Taurine and hypotaurine have also been shown to aid in protein folding (Warskulat et al. (2007) *Methods Enzymol* 428:439-58; Abe et al. (2015) *Amino acids* 47(5):909-15; Fujii et al. (2007) *J Biochem* 141(5):697-707).

[05] While taurine can be detected at high levels in a variety of fish species, some taurine has been suggested as a conditionally essential nutrient for many carnivorous fish species, from trout to snakehead, and its supplementation has been shown to increase their growth rate. (Gibson et al. (2007) *Aquaculture* 269(1-4):514–524; Wu et al. (2015) *Aquac Nutr.*

21(2):214–222) Furthermore, it appears that taurine supplementation can complement the reduction of fishmeal in the feedstock, a critical objective for achieving a more sustainable form of aquaculture.

[06] For cats, dietary taurine is often a necessary addition to feed. Inadequate levels of taurine may cause severe degenerative changes in the retina, visual cortex and brain development. Taurine has also been reported to have anti-epilepsy properties. (Ripps and Shen (2012) *Molecular Vision* 18:2673-2786)

[07] Chemical synthesis of taurine is undoubtedly the predominant means of production, via a plethora of known synthetic mechanisms. (Salze et al. (2015) *Aquaculture* 437:215–229) The biosynthesis of taurine in plant cells has also been described. (US2012/0222148 A1) Contrary to prior belief, a series of recent publications indicate that a large number of bacteria, fungi, and algae contain individual enzymes, or in some cases entire anabolic pathways, which are capable of taurine synthesis. (Tevatia et al. (2015) *Algal Res.* 9:21–26; Agnello et al. (2013) *ACS Chem Biol.* 8(10):2264–2271)

[08] There is a need for a combined protein/taurine feedstock to serve the fields of animal nutrition. The need is especially pressing in aquaculture during the larval stage of fish. Larval feed, which can sit in the water longer than adult feed, can result in significant loss of taurine due to dissipation. Chemically synthesized taurine in crystalline form is particularly susceptible to this process. Enriching rotifers with taurine is an effective solution (Matsunari et al. (2013) *Fish Sci.* 79(5):815–821), but a potentially uneconomical one, as live feeds tend to be expensive. An alternative strategy is to encapsulate taurine in microparticles, such as lipid-walled capsules (Langdon et al. (2003) *Aquaculture* 227(1-4):259–275). Plant-based production systems could achieve this objective by employing the cell membrane as a natural lipid-capsule. This approach is imperfect, however, as direct feeding with plant cells suffers from the anti-nutritional factors found in plant-based feeds. (Francis et al. (2001) *Aquaculture* 199(3-4):197-227) Therefore, there exists a need for an aquaculture feed that protects taurine from dissolving in water, while eschewing solutions involving plant-based biosynthesis or live feeds. The invention herein describes just such a solution.

BRIEF SUMMARY OF THE INVENTION

[09] In one aspect, non-naturally occurring microorganisms are provided that express one or more polynucleotide(s) expressing exogenous enzyme(s) for production of taurine. For example, a non-naturally occurring microorganism expresses the following enzyme activities: (a) cysteamine (2-aminoethanethiol) dioxygenase (ADO); (b) cysteine dioxygenase (CDO), and cysteine sulfinic acid decarboxylase (CSAD) or glutamate decarboxylase (GAD); (c) 3-mercaptopropionate dioxygenase (p3MDO), and CSAD or GAD; (d) L-serine dehydratase; sulfate adenyltransferase and adenylyl-sulfate kinase (APSK), and/or 3'-phosphoadenosine 5'-phosphosulfate synthase (PAPSS1)); 3'-phosphoadenylyl sulfate:2-aminoacrylate C-sulfotransferase (PAPS-AS), and CSAD or GAD; (e) cysteate synthase, optionally L-serine dehydratase, and CSAD or GAD; (f) L-cysteine desulfydrase (CD) activity; optionally cystathionine gamma-lyase (CGL) activity; sulfate adenyltransferase and APSK, and/or PAPSS1); PAPS-AS, and CSAD or GAD; (g) CD) activity; optionally CGL activity; cysteate synthase, and CSAD or GAD; (h) cysteate sulfo-lyase (CuyA), and CSAD or GAD; (i) phosphosulfolactate synthase (ComA), 2-phospho-3-sulfolactate phosphohydrolase (ComB), sulfolactate dehydrogenase (ComC), aspartate aminotransferase (AspAT), and CSAD or GAD; (j) sulfoacetaldehyde acetyltransferase (Xsc) and taurine-pyruvate aminotransferase (Tpa); (k) ComA, ComB, ComC, sulfopyruvate decarboxylase (ComDE), and Tpa; or (l) AspAT, ComDE and Tpa, wherein at least one of said enzyme activities is encoded by an exogenous polynucleotide that is expressed in the microorganism.

[10] In some embodiments, the non-naturally occurring microorganism is derived from a host cell from genera selected from *Methylobacterium*, *Methylomonas*, *Methylobacter*. *Methylococcus*, *Methylosinus*, *Methylocyctis*, *Methylomicrobium*, *Methylomonas*, *Methylpophilus*, *Methylobacillus*, *Methylobacterium*, *Hyphomicrobium*, *Xanthobacter*, *Bacillus*, *Paracoccus*, *Nocardia*, *Arthrobacter*, *Rhodopseudomonas*, *Pseudomonas*, *Candida*, *Hansenula, Pichia*, *Torulopsis*, *Rhodotorula*, *Escherichia*, and *Saccharomyces*. For example, the microorganism may be selected from *Methylobacterium*, *Escherichia*, *Saccharomyces*, and *Bacillus*. In some embodiments, the non-naturally occurring microorganism is a methylotrophic bacterium. For example, the non-naturally occurring

microorganism may be a *Methylobacterium* species, such as but not limited to, *Methylobacterium extorquens.*

[11] In some embodiments, the one or more exogenous polynucleotide(s) is/are codon optimized for expression in the microorganism. In some embodiments, the one or more exogenous polynucleotide(s) is/are operably linked to promoter(s) for expression in the microorganism.

[12] In some embodiments, the non-naturally occurring microorganism includes deletion of one or more genes that encode enzyme(s) that degrade taurine or the taurine precursor cysteate or sulfoacetaldehyde or modification of one or more genes that encode enzyme(s) that degrade taurine, cysteate, or sulfoacetaldehyde such that activity of the one or more enzyme(s) that degrade taurine, cysteate, or sulfoacetaldehyde is lower than in the microorganism parent strain from which the non-naturally microorganism is derived. In some embodiments, the one or more enzyme(s) that degrade taurine, cysteate, or sulfoacetaldehyde includes taurine dehydrogenase, taurine dioxygenase, Xsc, CuyA, Tpa, and/or gamma-glutamyltransferase.

[13] In some embodiments, the non-naturally occurring microorganism is genetically modified or artificially pre-selected to produce elevated levels of a carotenoid compound relative to the corresponding unmodified or unselected microorganism. For example, the microorganism may produce elevated levels of one or more carotenoid compound(s) selected from β-carotene, lycopene, rhodopsin, zeaxanthin, lutein, canthaxanthin, astaxanthin, and sprilloxanthin, in comparison to the host cell from which the carotenoid producing microorganism is derived.

[14] In some embodiments, the non-naturally occurring microorganism accumulates intracellular taurine and/or hypotaurine, wherein the taurine and/or hypotaurine aids in the folding of one or more native and/or heterologous protein(s), *e.g.*, for the purpose of increased enzymatic activity and/or protein yield in comparison to the parent microorganism from which the non-naturally occurring microorganism is derived, e.g., a parent microorganism that does not include the one or more exogenous polynucleotide(s).

[15] In another aspect, methods are provided for producing biomass that includes taurine and/or taurine precursors such as cysteate, sulfoacetaldehyde, and/or hypotaurine. The methods include culturing a non-naturally occurring microorganism as described herein in a

culture medium under conditions suitable for growth of the microorganism and expression of exogenous enzyme(s) for production of taurine and/or taurine precursors, wherein biomass comprising taurine and/or taurine precursors is produced in the culture.

[16] In another aspect, a feed or nutritional supplement is provided that includes taurine- and/or taurine precursor-containing biomass produced as described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[17] **Figure 1** depicts an embodiment of a biosynthetic pathway for production of taurine from cysteamine.

[18] **Figure 2** depicts an embodiment of a biosynthetic pathway for production of taurine from L-cysteine.

[19] **Figure 3** depicts an embodiment of a biosynthetic pathway for production of taurine from L-cysteine.

[20] **Figure 4** depicts an embodiment of a biosynthetic pathway for production of taurine from L-serine.

[21] **Figure 5** depicts an embodiment of a biosynthetic pathway for production of taurine from L-phosphoserine or L-serine.

[22] **Figure 6** depicts an embodiment of a biosynthetic pathway for production of taurine from L-cysteine.

[23] **Figure 7** depicts an embodiment of a biosynthetic pathway for production of taurine from L-cysteine.

[24] **Figure 8** depicts an embodiment of a biosynthetic pathway for production of taurine from pyruvate.

[25] **Figure 9** depicts an embodiment of a biosynthetic pathway for production of taurine from phosphoenolpyruvate.

[26] **Figure 10** depicts an embodiment of a biosynthetic pathway for production of taurine from acetyl phosphate.

[27] **Figure 11** depicts an embodiment of a biosynthetic pathway for production of taurine from phosphoenolpyruvate.

[28] **Figure 12** depicts an embodiment of a biosynthetic pathway for production of taurine through taurine-pyruvate aminotransferase.

[29] **Figure 13** depicts taurine and cysteate degradation pathways.

DETAILED DESCRIPTION

[30] The invention described herein addresses the dual challenge of producing taurine from an inexpensive feedstock, and encapsulating it to prevent dissolution in water for aquaculture. Microbial systems for taurine production are described as well as feed products in which the taurine is encapsulated in a natural lipid bilayer (microbial cells).

[31] Provided herein are non-naturally occurring microorganisms, *e.g.*, bacteria, yeast, Archaea, that are capable of producing taurine and/or the taurine precursor(s), *e.g.*, hypotaurine or cysteate. Also provided are methods of engineering and culturing such microorganisms, methods of using such microorganisms to produce taurine, and methods of producing taurine-containing compositions, such as feed compositions that contain the microorganisms or compositions that contain taurine recovered from such organisms.

[32] One aspect pertains to the field of aquaculture. Another aspect is the field of pet foods, for example, for cats and dogs. A further aspect is in the field of human nutrition and supplements. More specifically, aquaculture feeds, pet food, and nutritional supplement compositions are provided that include taurine-containing microbial biomass and a complete protein nutrition, that is, containing most or all amino acids necessary for healthy growth of the animal to which it is administered. In some embodiments, the aquaculture feed compositions herein contain one or more carotenoid(s) produced by the microorganism that produces taurine and/or taurine precursor(s), *e.g.*, hypotaurine or cysteate. The microbial biomass can be blended with other ingredients to form a portion or whole of a feed, or may be consumed directly as a protein-rich powder.

[33] Another aspect pertains to the field of industrial protein production. Osmolytes such as betaine, glycine, trimethylamine N-Oxide (TMAO), and taurine can aid in protein folding (Warskulat et al. (2007) *Methods Enzymol* 428:439-58; Abe et al. (2015) *Amino acids* 47(5):909-15; Fujii et al. (2007) *J Biochem* 141(5):697-707). Microorganisms engineered to accumulate intracellular taurine as a chemical chaperone could produce higher yields or more active proteins of interest. As antioxidants, taurine, hypotaurine, and their precursors also promote protein activity by limiting protein inactivation through oxidation (Oliveira et al. (2010) *Pharmacological Reports* 62:185-193; Aruoma et al. (1988) *Biochem J* 256:251-55; Bucolo et al. (2016) *Acta Ophthalmologic* 95(256); Patel et al. (2016) *Exp Toxic Pathol* 68(2-3):103-12; Fontana et al. (2004) *Neurochemical Research* 29(1):111-116).

Definitions

[34] Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Singleton, et al., Dictionary of Microbiology and Molecular Biology, second ed., John Wiley and Sons, New York (1994), and Hale & Markham, The Harper Collins Dictionary of Biology, Harper Perennial, NY (1991) provide one of skill with a general dictionary of many of the terms used in this invention. Any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention.

[35] The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology (including recombinant techniques), microbiology, cell biology, and biochemistry, which are within the skill of the art. Such techniques are explained fully in the literature, for example, Molecular Cloning: A Laboratory Manual, second edition (Sambrook et al., 1989); Oligonucleotide Synthesis (M. J. Gait, ed., 1984; Current Protocols in Molecular Biology (F. M. Ausubel et al., eds., 1994); PCR: The Polymerase Chain Reaction (Mullis et al., eds., 1994); and Gene Transfer and Expression: A Laboratory Manual (Kriegler, 1990).

[36] Numeric ranges provided herein are inclusive of the numbers defining the range.

[37] Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy orientation, respectively.

[38] "A," "an" and "the" include plural references unless the context clearly dictates otherwise.

[39] As used herein, the term "polynucleotide" refers to a polymeric form of nucleotides of any length and any three-dimensional structure and single- or multi-stranded (e.g., single-stranded, double-stranded, triple-helical, etc.), which contain deoxyribonucleotides, ribonucleotides, and/or analogs or modified forms of deoxyribonucleotides or ribonucleotides, including modified nucleotides or bases or their analogs. Because the genetic code is degenerate, more than one codon may be used to encode a particular amino acid, and the present invention encompasses polynucleotides which encode a particular amino acid sequence. Any type of modified nucleotide or nucleotide analog may be used, so long as the polynucleotide retains the desired

functionality under conditions of use, including modifications that increase nuclease resistance (e.g., deoxy, 2'-O-Me, phosphorothioates, etc.). Labels may also be incorporated for purposes of detection or capture, for example, radioactive or nonradioactive labels or anchors, e.g., biotin. The term polynucleotide also includes peptide nucleic acids (PNA). Polynucleotides may be naturally occurring or non-naturally occurring. The terms "polynucleotide," "nucleic acid," and "oligonucleotide" are used herein interchangeably. Polynucleotides may contain RNA, DNA, or both, and/or modified forms and/or analogs thereof. A sequence of nucleotides may be interrupted by non-nucleotide components. One or more phosphodiester linkages may be replaced by alternative linking groups. These alternative linking groups include, but are not limited to, embodiments wherein phosphate is replaced by P(O)S ("thioate"), P(S)S ("dithioate"), (O)NR.sub.2 ("amidate"), P(O)R, P(O)OR', CO or CH.sub.2 ("formacetal"), in which each R or R' is independently H or substituted or unsubstituted alkyl (1-20 C) optionally containing an ether (--O--) linkage, aryl, alkenyl, cycloalkyl, cycloalkenyl or araldyl. Not all linkages in a polynucleotide need be identical. Polynucleotides may be linear or circular or comprise a combination of linear and circular portions.

[40] As used herein, “polypeptide” refers to a composition comprised of amino acids and recognized as a protein by those of skill in the art. The conventional one-letter or three-letter code for amino acid residues is used herein. The terms “polypeptide” and “protein” are used interchangeably herein to refer to polymers of amino acids of any length. The polymer may be linear or branched, it may comprise modified amino acids, and it may be interrupted by non-amino acids. The terms also encompass an amino acid polymer that has been modified naturally or by intervention; for example, disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation or modification, such as conjugation with a labeling component. Also included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids, etc.), as well as other modifications known in the art.

[41] As used herein, a “vector” refers to a polynucleotide sequence designed to introduce nucleic acids into one or more cell types. Vectors include cloning vectors, expression vectors, shuttle vectors, plasmids, phage particles, cassettes and the like.

[42] As used herein, the term "expression" refers to the process by which a polypeptide is produced based on the nucleic acid sequence of a gene. The process includes both transcription and translation.

[43] As used herein, "expression vector" refers to a DNA construct containing a DNA coding sequence (e.g., gene sequence) that is operably linked to one or more suitable control sequence(s) capable of effecting expression of the coding sequence in a host. Such control sequences include a promoter to effect transcription, an optional operator sequence to control such transcription, a sequence encoding suitable mRNA ribosome binding sites, and sequences which control termination of transcription and translation. The vector may be a plasmid, a phage particle, or simply a potential genomic insert. Once transformed into a suitable host, the vector may replicate and function independently of the host genome, or may, in some instances, integrate into the genome itself. The plasmid is the most commonly used form of expression vector. However, the invention is intended to include such other forms of expression vectors that serve equivalent functions and which are, or become, known in the art.

[44] A "promoter" refers to a regulatory sequence that is involved in binding RNA polymerase to initiate transcription of a gene. A promoter may be an inducible promoter or a constitutive promoter. An "inducible promoter" is a promoter that is active under environmental or developmental regulatory conditions.

[45] The term "operably linked" refers to a juxtaposition or arrangement of specified elements that allows them to perform in concert to bring about an effect. For example, a promoter is operably linked to a coding sequence if it controls the transcription of the coding sequence.

[46] "Under transcriptional control" is a term well understood in the art that indicates that transcription of a polynucleotide sequence depends on its being operably linked to an element which contributes to the initiation of, or promotes transcription.

[47] "Under translational control" is a term well understood in the art that indicates a regulatory process which occurs after mRNA has been formed.

[48] A "gene" refers to a DNA segment that is involved in producing a polypeptide and includes regions preceding and following the coding regions as well as intervening sequences (introns) between individual coding segments (exons).

[49] As used herein, the term "host cell" refers to a cell or cell line into which a recombinant expression vector for production of a polypeptide may be transfected for expression of the polypeptide. Host cells include progeny of a single host cell, and the progeny may not necessarily be completely identical (in morphology or in total genomic DNA complement) to the original parent cell due to natural, accidental, or deliberate mutation. A host cell includes cells transfected or transformed *in vivo* with an expression vector.

[50] The term "recombinant," refers to genetic material (i.e., nucleic acids, the polypeptides they encode, and vectors and cells comprising such polynucleotides) that has been modified to alter its sequence or expression characteristics, such as by mutating the coding sequence to produce an altered polypeptide, fusing the coding sequence to that of another gene, placing a gene under the control of a different promoter, expressing a gene in a heterologous organism, expressing a gene at a decreased or elevated levels, expressing a gene conditionally or constitutively in manner different from its natural expression profile, and the like. Generally recombinant nucleic acids, polypeptides, and cells based thereon, have been manipulated by man such that they are not identical to related nucleic acids, polypeptides, and cells found in nature.

[51] A "signal sequence" refers to a sequence of amino acids bound to the N-terminal portion of a protein which facilitates the secretion of the mature form of the protein from the cell. The mature form of the extracellular protein lacks the signal sequence which is cleaved off during the secretion process.

[52] The term "selective marker" or "selectable marker" refers to a gene capable of expression in a host cell that allows for ease of selection of those hosts containing an introduced nucleic acid or vector. Examples of selectable markers include but are not limited to antimicrobial substances (*e.g.*, hygromycin, bleomycin, or chloramphenicol) and/or genes that confer a metabolic advantage, such as a nutritional advantage, on the host cell.

[53] The term "derived from" encompasses the terms "originated from," "obtained from," "obtainable from," "isolated from," and "created from," and generally indicates that one specified material finds its origin in another specified material or has features that can be described with reference to the another specified material.

[54] The term "culturing" refers to growing a population of cells, *e.g.*, microbial cells, under suitable conditions for growth, in a liquid or solid medium.

[55] The term "heterologous" or "exogenous," with reference to a polynucleotide or protein, refers to a polynucleotide or protein that does not naturally occur in a specified cell, *e.g.*, a host cell. It is intended that the term encompass proteins that are encoded by naturally occurring genes, mutated genes, and/or synthetic genes. In contrast, the term "homologous," with reference to a polynucleotide or protein, refers to a polynucleotide or protein that occurs naturally in the cell.

[56] The term "introduced," in the context of inserting a nucleic acid sequence into a cell, includes "transfection," "transformation," or "transduction" and refers to the incorporation of a nucleic acid sequence into a eukaryotic or prokaryotic cell wherein the nucleic acid sequence may be incorporated into the genome of the cell (*e.g.*, chromosome, plasmid, plastid, or mitochondrial DNA), converted into an autonomous replicon, or transiently expressed.

[57] "Transfection" or "transformation" refers to the insertion of an exogenous polynucleotide into a host cell. The exogenous polynucleotide may be maintained as a non-integrated vector, for example, a plasmid, or alternatively, may be integrated into the host cell genome. The term "transfecting" or "transfection" is intended to encompass all conventional techniques for introducing nucleic acid into host cells. Examples of transfection techniques include, but are not limited to, calcium phosphate precipitation, DEAE-dextran-mediated transfection, lipofection, electroporation, and microinjection.

[58] As used herein, the terms "transformed," "stably transformed," and "transgenic" refer to a cell that has a non-native (*e.g.*, heterologous) nucleic acid sequence integrated into its genome or as an episomal plasmid that is maintained through multiple generations.

[59] The terms "recovered," "isolated," "purified," and "separated" as used herein refer to a material (*e.g.*, a protein, nucleic acid, or cell) that is removed from at least one component with which it is naturally associated. For example, these terms may refer to a material which is substantially or essentially free from components which normally accompany it as found in its native state, such as, for example, an intact biological system.

[60] A "signal sequence" (also termed "presequence," "signal peptide," "leader sequence," or "leader peptide") refers to a sequence of amino acids at the amino terminus of a nascent polypeptide that targets the polypeptide to the secretory pathway and is cleaved

from the nascent polypeptide once it is translocated in the endoplasmic reticulum membrane.

[61] Related (and derivative) proteins encompass "variant" proteins. Variant proteins differ from a parent protein and/or from one another by a small number of amino acid residues. In some embodiments, the number of different amino acid residues is any of about 1, 2, 3, 4, 5, 10, 20, 25, 30, 35, 40, 45, or 50. In some embodiments, variants differ by about 1 to about 10 amino acids. Alternatively or additionally, variants may have a specified degree of sequence identity with a reference protein or nucleic acid, e.g., as determined using a sequence alignment tool, such as BLAST, ALIGN, and CLUSTAL (see, infra). For example, variant proteins or nucleic acid may have at least about 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or even 99.5% amino acid sequence identity with a reference sequence.

[62] As used herein, the term "analogous sequence" refers to a polypeptide sequence within a protein that provides a similar function, tertiary structure, and/or conserved residues with respect to a reference protein. For example, in epitope regions that contain an alpha helix or a beta sheet structure, replacement amino acid(s) in an analogous sequence maintain the same structural element. In some embodiments, analogous sequences are provided that result in a variant enzyme exhibiting a similar or improved function with respect to the parent protein from which the variant is derived.

[63] As used herein, "homologous protein" refers to a protein that has similar function and/or structure as a reference protein. Homologs may be from evolutionarily related or unrelated species. In some embodiments, a homolog has a quaternary, tertiary and/or primary structure similar to that of a reference protein, thereby potentially allowing for replacement of a segment or fragment in the reference protein with an analogous segment or fragment from the homolog, with reduced disruptiveness of structure and/or function of the reference protein in comparison with replacement of the segment or fragment with a sequence from a non-homologous protein.

[64] As used herein, "wild-type," "native," and "naturally-occurring" proteins are those found in nature. The terms "wild-type sequence" refers to an amino acid or nucleic acid sequence that is found in nature or naturally occurring. In some embodiments, a wild-type

sequence is the starting point of a protein engineering project, for example, production of variant proteins.

[65] The phrases "substantially similar" and "substantially identical" in the context of at least two nucleic acids or polypeptides typically means that a polynucleotide, polypeptide, or region or domain of a polypeptide that comprises a sequence that has at least about 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or even 99.5% sequence identity, in comparison with a reference (*e.g.*, wild-type) polynucleotide, polypeptide, or region or domain of a polypeptide. A region or domain of a polypeptide may contain, for example, at least about 20, 50, 100, or 200 amino acids within a longer polypeptide sequence. Sequence identity may be determined using known programs such as BLAST, ALIGN, and CLUSTAL using standard parameters. (See, e.g., Altshul et al. (1990) *J. Mol. Biol.* 215:403-410; Henikoff et al. (1989) *Proc. Natl. Acad. Sci.* 89:10915; Karin et al. (1993) *Proc. Natl. Acad. Sci.* 90:5873; and Higgins et al. (1988) *Gene* 73:237). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. Also, databases may be searched using FASTA (Person et al. (1988) Proc. Natl. Acad. Sci. 85:2444-2448.) In some embodiments, substantially identical polypeptides differ only by one or more conservative amino acid substitutions. In some embodiments, substantially identical polypeptides are immunologically cross-reactive. In some embodiments, substantially identical nucleic acid molecules hybridize to each other under stringent conditions (*e.g.*, within a range of medium to high stringency).

[66] The term "carotenoid" is understood in the art to refer to a structurally diverse class of pigments derived from isoprenoid pathway intermediates. The commitment step in carotenoid biosynthesis is the formation of phytoene from geranylgeranyl pyrophosphate. Carotenoids can be acyclic or cyclic, and may or may not contain oxygen, so that the term carotenoids include both carotenes and xanthophylls. In general, carotenoids are hydrocarbon compounds having a conjugated polyene carbon skeleton formally derived from the five-carbon compound IPP, including triterpenes (C_{30} diapocarotenoids) and tetraterpenes (C_{40} carotenoids) as well as their oxygenated derivatives and other compounds that are, for example, C_{35}, C_{50}, C_{60}, C_{70}, C_{80} in length or other lengths. Many carotenoids have strong light absorbing properties and may range in length in excess of C_{200}- C_{30} diapocarotenoids typically consist of six isoprenoid units joined in such a manner that the

arrangement of isoprenoid units is reversed at the center of the molecule so that the two central methyl groups are in a 1 ,6-positional relationship and the remaining non-terminal methyl groups are in a 1 ,5-positional relationship. Such C_{30} carotenoids may be formally derived from the acyclic $C_{30}H_{42}$ structure, having a long central chain of conjugated double bonds, by: (i) hydrogenation (ii) dehydrogenation, (iii) cyclization, (iv) oxidation, (v) esterification/glycosylation, or any combination of these processes. C_{40} carotenoids typically consist of eight isoprenoid units joined in such a manner that the arrangement of isoprenoid units is reversed at the center of the molecule so that the two central methyl groups are in a 1 ,6-positional relationship and the remaining non-terminal methyl groups are in a 1 ,5-positional relationship. Such C_{40} carotenoids may be formally derived from the acyclic $C_{40}H_{56}$ structure, having a long central chain of conjugated double bonds, by (i) hydrogenation, (ii) dehydrogenation, (iii) cyclization, (iv) oxidation, (v) esterification/glycosylation, or any combination of these processes. The class of C_{40} carotenoids also includes certain compounds that arise from rearrangements of the carbon skeleton, or by the (formal) removal of part of this structure. More than 600 different carotenoids have been identified in nature. Carotenoids include but are not limited to: antheraxanthin, adonirubin, adonixanthin, astaxanthin, canthaxanthin, capsorubrin, β-cryptoxanthin, a-carotene, β-carotene, β,ψ-carotene, δ-carotene, ε-carotene, echinenone, 3-hydroxyechinenone, 3'-hydroxyechinenone, γ-carotene, ψ-carotene, 4-keto-Y-carotene, ζ-carotene, a-cryptoxanthin, deoxyflexixanthin, diatoxanthin, 7,8-didehydroastaxanthin, didehydrolycopene, fucoxanthin, fucoxanthinol, isorenieratene, β-isorenieratene, lactucaxanthin, lutein, lycopene, myxobactone, neoxanthin, neurosporene, hydroxyneurosporene, peridinin, phytoene, rhodopin, rhodopin glucoside, 4-keto-rubixanthin,siphonaxanthin, spheroidene, spheroidenone, spirilloxanthin, torulene, 4-keto-torulene, 3-hydroxy-4-keto-torulene, uriolide, uriolide acetate, violaxanthin, zeaxanthin-β-diglucoside, zeaxanthin, and C30 carotenoids. Additionally, carotenoid compounds include derivatives of these molecules, which may include hydroxy-, methoxy-, oxo-, epoxy-, carboxy-, or aldehydic functional groups. Further, included carotenoid compounds include ester (e.g., glycoside ester, fatty acid ester) and sulfate derivatives (e.g., esterified xanthophylls).

[67] The "isoprenoid pathway" is understood in the art to refer to a metabolic pathway that either produces or utilizes the five-carbon metabolite isopentyl pyrophosphate (IPP). As

discussed herein, two different pathways can produce the common isoprenoid precursor IPP— the "mevalonate pathway" and the "non-mevalonate pathway." The term "isoprenoid pathway" is sufficiently general to encompass both of these types of pathway. Biosynthesis of isoprenoids from IPP occurs by polymerization of several five-carbon isoprene subunits. Isoprenoid metabolites derived from IPP vary greatly in chemical structure, including both cyclic and acyclic molecules. Isoprenoid metabolites include, but are not limited to, monoterpenes, sesquiterpenes, diterpenes, sterols, and polyprenols such as carotenoids.

[68] The term "isoprenoid compound" refers to any compound which is derived via the pathway beginning with isopentenyl pyrophosphate (IPP) and formed by the head-to-tail condensation of isoprene units which may be of 5, 10, 15, 20, 30 or 40 carbons in length. There term "isoprenoid pigment" refers to a class of isoprenoid compounds which typically have strong light absorbing properties.

[69] The term "feed premix" refers to the crude mixture of aquaculture feed or animal/pet food components prior to processing, optionally at high temperature, into an aquaculture feed or animal or pet food composition that is in the form of pellets or flakes.

[70] An aquaculture feed composition is used in the production of an "aquaculture product," wherein the product is a harvestable aquacultured species (e.g., finfish, crustaceans), which is often sold for human consumption. For example, salmon are intensively produced in aquaculture and thus are aquaculture products. Aquaculture compositions may also be used as feed for aquaculture feed organisms such as small fish like krill, rotifers, and the like, that are food sources for larger aquaculture organisms such as carnivorous fish. In addition, aquaculture compositions described herein can be used as feed for ornamental fish, shrimp, hobbyist aquaculture, and the like, that are not intended as food for other organisms.

[71] The term "aquaculture meat product" refers to food products intended for human consumption comprising at least a portion of meat from an aquaculture product as defined above. An aquaculture meat product may be, for example, a whole fish or a filet cut from a fish, each of which may be consumed as food. In some embodiments, such a product can be referred to as a fish or seafood product.

[72] The term "biomass" refers to microbial cellular material. Biomass may be produced naturally, or may be produced from the fermentation of a native host or a recombinant production host. The biomass may be in the form of whole cells, whole cell lysates,

homogenized cells, partially hydrolyzed cellular material, and/or partially purified cellular material (*e.g.*, microbially produced oil).

[73] The term "processed biomass" refers to biomass that has been subjected to additional processing such as drying, pasteurization, disruption, etc., each of which is discussed in greater detail below.

[74] The term "C-1 carbon substrate" refers to any carbon-containing molecule that lacks a carbon-carbon bond. Examples are methane, methanol, formaldehyde, formic acid, formate, methylated amines (*e.g.*, mono-, di-, and tri- methyl amine), methylated thiols, and carbon dioxide.

[75] The term "C1 metabolizer" refers to a microorganism that has the ability to use a single carbon substrate as a sole source of energy and biomass. C1 metabolizers will typically be methylotrophs and/or methanotrophs capable of growth.

[76] The term "methylotroph" means an organism capable of oxidizing organic compounds which do not contain carbon-carbon bonds. Where the methylotroph is able to oxidize CH_4, the methylotroph is also a methanotroph.

[77] The term "methanotroph" means a prokaryote capable of utilizing methane as a substrate. Complete oxidation of methane to carbon dioxide occurs by aerobic degradation pathways. Typical examples of methanotrophs useful in the present invention include but are not limited to the genera *Methylomonas*, *Methylobacter*, *Methylococcus*, and *Methylosinus*.

[78] The term "high growth methanotrophic bacterial strain" refers to a bacterium capable of growth using methane as its sole carbon and energy source.

Microorganisms

[79] Non-naturally occurring microorganisms are provided for production of taurine or the taurine precursors hypotaurine, cysteate, or sulfoacetaldehyde. Non-naturally occurring, *e.g.*, recombinant, microorganisms herein include, *e.g.*, bacteria, yeast, Archaea, that have been engineered to express at least one (*i.e.*, one or more) enzyme(s) for biosynthesis of taurine or taurine precursors and that produce taurine or taurine precursors when cultured under conditions suitable for microbial growth and taurine production.

[80] Non-naturally occurring microorganisms as described herein include one or more exogenous polynucleotide(s) that encode and express one or more enzyme or enzyme activity for biosynthesis of taurine or the taurine precursors cysteate, sulfoacetaldehyde, or

hypotaurine. The exogenous polynucleotide(s) may include one or more coding sequence for one or more enzyme or enzyme activity for biosynthesis of taurine or taurine precursors, operably linked to one or more promoter for expression in the non-naturally occurring microorganism. Such promoters may include, but are not limited to, P_R (*e.g.*, SEQ ID NO:42), P_Lac (*e.g.*, SEQ ID NO:41), P_tac (*e.g.*, SEQ ID NO:39), P_tacA (*e.g.*, SEQ ID NO:40), PmxaF (*e.g.*, SEQ ID NO:43), P_rrnB, and P_T7. In some embodiments, the polynucleotide(s) are codon optimized for expression in the microorganism.

[81] In some embodiments, the non-naturally occurring microorganism includes one or more exogenous and/or endogenous polynucleotide(s) that encodes one or more enzymes or enzyme activities for taurine biosynthesis, as described herein, that has been modified for improved stability and/or activity relative to the stability and/or activity of the enzyme or enzyme activity in the host cell from which it is derived or relative to the wild-type stability and/or activity of the enzyme or enzyme activity. For example, the non-naturally occurring microorganism may express a variant of an enzyme of taurine biosynthesis that has greater stability and/or activity than the wild-type enzyme from which it is derived.

[82] In some embodiments, the host cell from which a non-naturally occurring microorganism as described herein is derived has one or more endogenous taurine, cysteate, or sulfoacetaldehyde degrading activity, for example, but not limited to, taurine dehydrogenase, Tpa, CuyA, gamma-glutamyltransferase, Xsc, and/or taurine dioxygenase. In some embodiments, the non-naturally occurring microorganism includes deletion of one or more genes that encode taurine, cysteate, or sulfoacetaldehyde degrading enzyme(s). In some embodiments, the host cell from which a non-naturally occurring microorganism includes modification of one or more genes that encode taurine, cysteate, or sulfoacetaldehyde degrading enzyme(s), such that the taurine, cysteate, or sulfoacetaldehyde degrading activity of the enzyme(s) is lower in the non-naturally occurring microorganism than in the host cell from which it is derived. In some embodiments, the host cell is *Methylobacterium extorquens* and the non-naturally occurring microorganism derived from the host cell includes deletion or modification of the gene that encodes gamma-glutamyltransferase in the host cell.

[83] In certain embodiments, the host cell comprises one or more of endogenous genes in the described pathway. In certain embodiments, the host cell is modified so that one or more genes producing enzymes that divert compounds and taurine precursors away from a

taurine biosynthetic pathway are blocked or deleted. In certain embodiments, the one or more blocked or deleted genes are selected from genes involved in the degradation of taurine, cysteate, or sulfoacetaldehyde. In certain embodiments, the host cell is a spontaneous mutant whose rate of growth is increased relative to a corresponding non-mutant. In certain embodiments, the host cell is cultured under stress conditions selected from light depletion, nutrient depletion, nitrogen depletion, high salt, or a chemical that inhibits growth of the host cell, wherein the stress conditions induce changes in gene expression leading to increased taurine or taurine precursor production.

[84] In some embodiments, the non-naturally occurring microorganism or the host cell from which the non-naturally occurring microorganism is derived is genetically modified or artificially pre-selected to produce elevated levels of one or more carotenoid compound(s) relative to the corresponding unmodified or unselected microorganism. The one or more carotenoid compound(s) may include, but are not limited to, β-carotene, lycopene, zeaxanthin, lutein, canthaxanthin, rhodopin, astaxanthin, and/or sprilloxanthin. Non-limiting examples of host cells that produce elevated levels of one or more carotenoid compound(s) and methods for producing such microorganisms are provided in WO2015/021352 A2.

[85] Non-limiting examples of genera from which the non-naturally occurring microorganism may be derived include *Methylobacterium*, *Methylomonas*, *Methylobacter*. *Methylococcus*, *Methylosinus*, *Methylocyctis*, *Methylomicrobium*, *Methylomonas*, *Methylpophilus*, *Methylobacillus*, *Methylobacterium*, *Hyphomicrobium*, *Xanthobacter*, *Bacillus*, *Paracoccus*, *Nocardia*, *Arthrobacter*, *Rhodopseudomonas*, *Pseudomonas*, *Candida*, *Hansenula, Pichia*, *Torulopsis*, *Rhodotorula*, *Escherichia*, and *Saccharomyces*. Non-limiting examples of microbial species from which the non-naturally occurring microorganism may be derived include *Methylobacterium extorquens* (*e.g.*, strains AM1, DM4, CM4, PA1, or BJ001 (formerly *Methylobacterium populi*)), *Methylobacterium radiotolerans*, *Methylobacterium nodulans*, *Methylobacterium* spp. 4-46, and *Escherichia coli*.

[86] In some embodiments, the non-naturally occurring microorganism is a methylotrophic bacterium.

Conversion of cysteamine to taurine

[87] In some embodiments, a non-naturally occurring microorganism is provided that expresses an exogenous enzyme activity of 2-aminoethanol (cysteamine) dioxygenase (ADO) (EC1.13.11.19), which converts cysteamine to hypotaurine, for biosynthesis of taurine, as shown in **Fig. 1**.

[88] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes ADO comprising or consisting of the amino acid sequence depicted in SEQ ID NO:44, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:44. In some embodiments, the polynucleotide that encodes ADO comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:45 or SEQ ID NO:57 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:45 or SEQ ID NO:57. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:57.

[89] In some embodiments, the non-naturally occurring microorganism is a *Methylobacterium*, *Escherichia*, *Saccharomyces*, or *Bacillus* microorganism that includes an exogenous polynucleotide that encodes ADO.

Conversion of cysteine to taurine via CDO

[90] In some embodiments, a non-naturally occurring microorganism is provided that expresses one or more exogenous enzyme activity/ies for the conversion of cysteine to taurine via the enzyme cysteine dioxygenase (CDO), for example, exogenous enzyme(s) of the CDO/CSAD or GAD pathway for biosynthesis of taurine. The CDO/CSAD or GAD pathway for taurine biosynthesis is shown schematically in **Fig. 2**.

[91] In some embodiments, the non-naturally occurring microorganism that expresses exogenous enzyme activities of the CDO/CSAD or GAD pathway is not of genera *Escherichia* or species *Escherichia coli*. In some embodiments, the non-naturally occurring microorganism is not of genera *Saccharomyces* or species *Saccharomyces cerevisiae.*

[92] Non-naturally occurring microorganisms are provided that include the following enzymes or enzyme activities: cysteine dioxygenase (CDO) (EC 1.13.11.20); and cysteine sulfinic acid decarboxylase (CSAD) (EC 4.1.1.29), wherein at least one of these enzymes or

enzyme activities is encoded by an exogenous polynucleotide with which the microorganism has been transformed. In some embodiments, CDO and CSAD enzymes or enzyme activities are encoded by one or more exogenous polynucleotide(s) (e.g., one or two exogenous polynucleotide(s)) in the microorganism. In some embodiments, one of CDO and CSAD enzymes or enzyme activities is encoded an exogenous polynucleotide in the microorganism and the remaining activity is endogenously expressed in the host cell from which the non-naturally occurring microorganism is derived, *i.e.*, not expressed from an exogenous polynucleotide. In one embodiment, the microorganism expresses an endogenous CSAD activity and CDO is encoded by an exogenous polynucleotide in the microorganism. In one embodiment, the microorganism expresses an endogenous CDO activity and CSAD is encoded by an exogenous polynucleotide in the microorganism.

[93] Non-naturally occurring microorganisms are provided that include the following enzymes or enzyme activities: cysteine dioxygenase (CDO) (EC 1.13.11.20); and glutamate decarboxylase (GAD) (EC 4.1.1.15), wherein at least one of these enzymes or enzyme activities is encoded by an exogenous polynucleotide with which the microorganism has been transformed. In some embodiments, CDO and GAD enzymes or enzyme activities are encoded by one or more exogenous polynucleotide(s) (e.g., one or two exogenous polynucleotide(s)) in the microorganism. In some embodiments, one of CDO and GAD enzymes or enzyme activities is encoded by an exogenous polynucleotide in the microorganism and the remaining activity is endogenously expressed in the host cell from which the non-naturally occurring microorganism is derived, *i.e.*, not expressed from an exogenous polynucleotide. In one embodiment, the microorganism expresses an endogenous GAD activity and CDO is encoded by an exogenous polynucleotide in the microorganism. In one embodiment, the microorganism expresses an endogenous CDO activity and GAD is encoded by an exogenous polynucleotide in the microorganism.

[94] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes CDO comprising or consisting of the amino acid sequence depicted in SEQ ID NO:15, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:15. In some embodiments, the polynucleotide that encodes CDO comprises or consists of the polynucleotide sequence

depicted in SEQ ID NO:16, SEQ ID NO:50, or SEQ ID NO:58 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:16, SEQ ID NO:50, or SEQ ID NO:58. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:58.

[95] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes CDO comprising or consisting of the amino acid sequence depicted in SEQ ID NO:35, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:35. In some embodiments, the polynucleotide that encodes CDO comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:36, SEQ ID NO:51, or SEQ ID NO:59 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:36, SEQ ID NO:51, or SEQ ID NO:59. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ. NO:51 or SEQ ID NO: 59.

[96] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes CSAD comprising or consisting of the amino acid sequence depicted in SEQ ID NO:11, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:11. In some embodiments, the polynucleotide that encodes CSAD comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:12, SEQ ID NO:53, or SEQ ID NO:54 or a polynucleotide having at least about 30, 35, 40,45,50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:12, SEQ ID NO:53, or SEQ ID NO:54. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:53 or SEQ ID NO:54.

[97] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes GAD comprising or consisting of the amino acid sequence depicted in SEQ ID NO:13, or a variant or homolog thereof. In some

embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:13. In some embodiments, the polynucleotide that encodes GAD comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:14 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:14. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism.

[98] In one embodiment, a non-naturally occurring *Methylobacterium* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding CDO and CSAD. In another embodiment, a non-naturally occurring *Methylobacterium* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding CDO and GAD.

[99] In one embodiment, a non-naturally occurring *Escherichia* microorganism is provided that includes an exogenous polynucleotide encoding CDO. In one embodiment, a non-naturally occurring *Escherichia* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding CDO and CSAD.

[100] In one embodiment, a non-naturally occurring *Saccharomyces* microorganism is provided that includes an exogenous polynucleotide encoding CDO. In one embodiment, a non-naturally occurring *Saccharomyces* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding CDO and CSAD.

[101] In one embodiment, a non-naturally occurring *Bacillus* microorganism is provided that includes an exogenous polynucleotide encoding CSAD.

Conversion of cysteine to taurine via p3MDO

[102] In some embodiments, a non-naturally occurring microorganism is provided that expresses one or more exogenous enzyme activity/ies for the conversion of cysteine to taurine via the enzyme 3-mercaptopropionate dioxygenase (p3MDO), for example, exogenous enzyme(s) of the p3MDO/CSAD or GAD pathway for biosynthesis of taurine. The p3MDO/CSAD or GAD pathway for taurine biosynthesis is shown schematically in **Fig. 3**.

[103] Non-naturally occurring microorganisms are provided that include the following enzymes or enzyme activities: 3-mercaptopropionate dioxygenase (MDO; p3MDO) (EC

1.13.11.-); and cysteine sulfinic acid decarboxylase (CSAD) (EC 4.1.1.29), wherein at least one of these enzymes or enzyme activities is encoded by an exogenous polynucleotide with which the microorganism has been transformed. In some embodiments, p3MDO and CSAD enzymes or enzyme activities are encoded by one or more exogenous polynucleotide(s) (e.g., one or two, exogenous polynucleotide(s)) in the microorganism. In some embodiments, one of p3MDO and CSAD enzymes or enzyme activities is encoded by an exogenous polynucleotide in the microorganism and the remaining activity is endogenously expressed in the host cell from which the non-naturally occurring microorganism is derived, *i.e.*, not expressed from an exogenous polynucleotide. In one embodiment, the microorganism expresses an endogenous CSAD activity and p3MDO is encoded by an exogenous polynucleotide in the microorganism. In one embodiment, the microorganism expresses an endogenous p3MDO activity and CSAD is encoded by an exogenous polynucleotide in the microorganism.

[104] Non-naturally occurring microorganisms are provided that include the following enzymes or enzyme activities: 3-mercaptopropionate dioxygenase (MDO; p3MDO) (EC 1.13.11.-); and glutamate decarboxylase (GAD) (EC 4.1.1.15), wherein at least one of these enzymes or enzyme activities is encoded by an exogenous polynucleotide with which the microorganism has been transformed. In some embodiments, p3MDO and GAD enzymes or enzyme activities are encoded by one or more exogenous polynucleotide(s) (e.g., one or two, exogenous polynucleotide(s)) in the microorganism. In some embodiments, one of p3MDO and GAD enzymes or enzyme activities is encoded an exogenous polynucleotide in the microorganism and the remaining activity is endogenously expressed in the host cell from which the non-naturally occurring microorganism is derived, *i.e.*, not expressed from an exogenous polynucleotide. In one embodiment, the microorganism expresses an endogenous GAD activity and p3MDO is encoded by an exogenous polynucleotide in the microorganism. In one embodiment, the microorganism expresses an endogenous p3MDO activity and GAD is encoded by an exogenous polynucleotide in the microorganism.

[105] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes p3MDO comprising or consisting of the amino acid sequence depicted in SEQ ID NO:33, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70,

75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:33. In some embodiments, the polynucleotide that encodes p3MDO comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:34 or SEQ ID NO:60 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:34 or SEQ ID NO:60. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:60.

[106] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes CSAD comprising or consisting of the amino acid sequence depicted in SEQ ID NO:11, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:11. In some embodiments, the polynucleotide that encodes CSAD comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:12, SEQ ID NO:53, or SEQ ID NO:54 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:12, SEQ ID NO:53, or SEQ ID NO:54. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.* SEQ ID NO: 53 or SEQ ID NO:54.

[107] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes GAD comprising or consisting of the amino acid sequence depicted in SEQ ID NO:13, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:13. In some embodiments, the polynucleotide that encodes GAD comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:14 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:14. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism.

[108] In one embodiment, a non-naturally occurring *Methylobacterium* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding p3MDO and CSAD. In another embodiment, a non-naturally occurring *Methylobacterium*

microorganism is provided that includes one or more exogenous polynucleotide(s) encoding p3MDO and GAD.

[109] In one embodiment, a non-naturally occurring *Escherichia* microorganism is provided that includes an exogenous polynucleotide encoding p3MDO. In one embodiment, a non-naturally occurring *Escherichia* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding p3MDO and CSAD.

[110] In one embodiment, a non-naturally occurring *Saccharomyces* microorganism is provided that includes an exogenous polynucleotide encoding p3MDO. In one embodiment, a non-naturally occurring *Saccharomyces* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding p3MDO and CSAD.

[111] In one embodiment, a non-naturally occurring *Bacillus* microorganism is provided that includes an exogenous polynucleotide encoding p3MDO. In one embodiment, a non-naturally occurring *Bacillus* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding p3MDO and CSAD.

Conversion of serine to taurine

[112] In some embodiments, a non-naturally occurring microorganism is provided that expresses one or more exogenous enzyme activity/ies of the serine/sulfate pathway for biosynthesis of taurine. The serine/sulfate pathway for taurine biosynthesis is shown schematically in **Fig. 4**.

[113] In some embodiments, the non-naturally occurring microorganism includes one or more mutations that cause accumulation of serine in the microorganism. For example, a methylotrophic strain that uses ribulose monophosphate (RuMP) for carbon assimilation from methanol may include a deletion or mutation in HprA (hydroxypyruvate reductase), which blocks the serine cycle from being completed, resulting in serine accumulation.

[114] Non-naturally occurring microorganisms are provided that include the following enzymes or enzyme activities: L-serine dehydratase (EC 4.3.1.17); sulfate adenyltransferase (EC 2.7.7.4) and adenylyl-sulfate kinase (APS kinase) (EC 2.7.1.25), and/or 3'-phosphoadenosine 5'-phosphosulfate synthase (PAPSS1) (EC 2.7.7.4/EC 2.7.1.25); 3'-phosphoadenylyl sulfate: 2-aminoacrylate C-sulfotransferase (PAPS-AS) (EC 2.8.2.-); and cysteine sulfinic acid decarboxylase (CSAD) (EC 4.1.1.29), wherein at least one of these enzymes or enzyme activities is encoded by an exogenous polynucleotide with which the microorganism has been transformed. In some embodiments, L-serine dehydratase; sulfate

adenyltransferase and APS kinase, and/or PAPSS1; PAPS-AS; and CSAD enzymes or enzyme activities are encoded by one or more exogenous polynucleotide(s) (e.g., one, two, three, four, five, six, or seven exogenous polynucleotide(s)) in the microorganism. In some embodiments, one, two, three, four, five, or six of L-serine dehydratase, sulfate adenyltrasnferase, APS kinase, PAPSS1, PAPS-AS, and CSAD enzymes or enzyme activities is encoded by one or more exogenous polynucleotide(s) in the microorganism and the remaining activity/ies is/are endogenously expressed in the host cell from which the non-naturally occurring microorganism is derived, *i.e.*, not expressed from exogenous polynucleotide(s). In one embodiment, the microorganism expresses an endogenous L-serine dehydratase activity, and sulfate adenyltransferase and APS kinase, and/or PAPSS1, PAPS-AS, and CSAD are expressed from exogenous polynucleotide(s). In one embodiment, the microorganism expresses endogenous L-serine dehydratase, sulfate adenyltransferase, and APS kinase activity, and PAPS-AS and CSAD are expressed from exogenous polynucleotide(s). In one embodiment, the microorganism expresses endogenous L-serine dehydratase, sulfate adenyltransferase, APS kinase, and CSAD activities, and PAPS-AS is expressed from an exogenous polynucleotide.

[115] Non-naturally occurring microorganisms are provided that include the following enzymes or enzyme activities: L-serine dehydratase (EC 4.3.1.17); sulfate adenyltransferase (EC 2.7.7.4) and adenylyl-sulfate kinase (APS kinase) (EC 2.7.1.25), and/or 3'-phosphoadenosine 5'-phosphosulfate synthase (PAPSS1) (EC 2.7.7.4/EC 2.7.1.25); 3'-phosphoadenylyl sulfate: 2-aminoacrylate C-sulfotransferase (PAPS-AS) (EC 2.8.2.-); and glutamate decarboxylase (GAD) (EC 4.1.1.15), wherein at least one of these enzymes or enzyme activities is encoded by an exogenous polynucleotide with which the microorganism has been transformed. In some embodiments, L-serine dehydratase; sulfate adenyltrasnferase and APS kinase, and/or PAPSS1; PAPS-AS; and GAD enzymes or enzyme activities are encoded by one or more exogenous polynucleotide(s) (e.g., one, two, three, four, five, six, or seven exogenous polynucleotide(s)) in the microorganism. In some embodiments, one, two, three, four, five, or six of L-serine dehydratase, sulfate adenyltransferase, APS kinase, PAPSS1, PAPS-AS, and GAD enzymes or enzyme activities is encoded by one or more exogenous polynucleotide(s) in the microorganism and the remaining activity/ies is/are endogenously expressed in the host cell from which the non-naturally occurring microorganism is derived, *i.e.*, not expressed from exogenous

polynucleotide(s). In one embodiment, the microorganism expresses an endogenous L-serine dehydratase activity, and sulfate adenyltransferase and APS kinase, and/or PAPSS1, PAPS-AS, and GAD are expressed from exogenous polynucleotide(s). In one embodiment, the microorganism expresses endogenous L-serine dehydratase, sulfate adenyltransferase, APS kinase activity, and PAPS-AS and GAD are expressed from exogenous polynucleotide(s). In one embodiment, the microorganism expresses endogenous L-serine dehydratase, sulfate adenyltransferase, APS kinase, and GAD activities, and PAPS-AS is expressed from an exogenous polynucleotide.

[116] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes L-serine dehydratase comprising or consisting of the amino acid sequence depicted in SEQ ID NO:1, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:1. In some embodiments, the polynucleotide that encodes L-serine dehydratase comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:2 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:2. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism.

[117] In some embodiments, the non-naturally occurring microorganism includes exogenous polynucleotides that encode sulfate adenyltransferase comprising or consisting of the amino acid sequences depicted in SEQ ID NO:3 and SEQ ID NO:5, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotides encode polypeptides comprising or consisting of amino acid sequences having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:3 or SEQ ID NO:5. In some embodiments, the polynucleotides that encode sulfate adenyltransferase comprise or consists of the polynucleotide sequences depicted in SEQ ID NO:4 and SEQ ID NO:6 or polynucleotides having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:4 or SEQ ID NO:6. In some embodiments, the polynucleotide sequences are codon optimized for expression in the microorganism.

[118] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes APS kinase comprising or consisting of the amino

acid sequence depicted in SEQ ID NO:7, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:7. In some embodiments, the polynucleotide that encodes APS kinase comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:8 or SEQ ID NO:62 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:8 or SEQ ID NO:62. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.* SEQ ID NO:62.

[119] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes PAPSS1 comprising or consisting of the amino acid sequence depicted in SEQ ID NO:31, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:31. In some embodiments, the polynucleotide that encodes PAPSS1 comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:32 or SEQ ID NO:63 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:32 or SEQ ID NO:63. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:63.

[120] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes PAPS-AS comprising or consisting of the amino acid sequence depicted in SEQ ID NO:9, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:9. In some embodiments, the polynucleotide that encodes PAPS-AS comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:10 or SEQ ID NO:61 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:10 or SEQ ID NO:61. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:61.

[121] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes CSAD comprising or consisting of the amino acid sequence depicted in SEQ ID NO:11, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:11. In some embodiments, the polynucleotide that encodes CSAD comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:12, SEQ ID NO:53, or SEQ ID NO:54 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:12, SEQ ID NO:53, or SEQ ID NO:54. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:53 or SEQ ID NO:54.

[122] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes GAD comprising or consisting of the amino acid sequence depicted in SEQ ID NO:13, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:13. In some embodiments, the polynucleotide that encodes GAD comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:14 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:14. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism.

[123] In one embodiment, a non-naturally occurring *Methylobacterium* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding: sulfate adenyltransferase and APS kinase, and/or PAPSS1; PAPS-AS; and CSAD. In one embodiment, a non-naturally occurring *Methylobacterium* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding PAPS-AS and CSAD. In another embodiment, a non-naturally occurring *Methylobacterium* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding: sulfate adenyltransferase and APS kinase, and/or PAPSS1; PAPS-AS; and GAD. In one

embodiment, a non-naturally occurring *Methylobacterium* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding PAPS-AS and GAD.

[124] In one embodiment, a non-naturally occurring *Escherichia* microorganism is provided that includes an exogenous polynucleotide encoding PAPS-AS. In one embodiment, a non-naturally occurring *Escherichia* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding PAPSS1 and PAPS-AS. In one embodiment, a non-naturally occurring *Escherichia* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding PAPSS1, PAPS-AS, and CSAD.

[125] In one embodiment, a non-naturally occurring *Saccharomyces* microorganism is provided that includes an exogenous polynucleotide encoding PAPS-AS. In one embodiment, a non-naturally occurring *Saccharomyces* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding PAPSS1 and PAPS-AS. In one embodiment, a non-naturally occurring *Saccharomyces* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding PAPSS1, PAPS-AS, and CSAD.

[126] In one embodiment, a non-naturally occurring *Bacillus* microorganism is provided that includes an exogenous polynucleotide encoding PAPS-AS. In one embodiment, a non-naturally occurring *Bacillus* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding PAPSS1 and PAPS-AS. In one embodiment, a non-naturally occurring *Bacillus* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding PAPSS1, PAPS-AS, and CSAD.

Conversion of phosphoserine or serine to taurine

[127] In some embodiments, a non-naturally occurring microorganism is provided that expresses one or more exogenous enzyme activity/ies for the conversion of phosphoserine to taurine via the enzyme cysteate synthase, for example, enzyme(s) of the cysteate synthase (*e.g.*, MA_3297)/CSAD or GAD pathway for biosynthesis of taurine, and/or expresses one or more exogenous enzyme activity/ies for the conversion of serine to taurine via the enzymes L-serine dehydratase, cysteate synthase (*e.g.*, MA_3297), and CSAD/GAD. The cysteate synthase (*e.g.*, MA_3297)/CSAD or GAD, and L-serine dehydratase cysteate synthase (*e.g.*, MA_3297)/CSAD or GAD pathways for taurine biosynthesis are shown schematically in **Fig. 5**.

[128] Non-naturally occurring microorganisms are provided that include the following enzymes or enzyme activities: cysteate synthase, *e.g.*, MA_3297 (EC 2.5.1.76); optionally L-serine dehydratase (EC 4.3.1.17); and cysteine sulfinic acid decarboxylase (CSAD) (EC 4.1.1.29), wherein at least one of these enzymes or enzyme activities is encoded by an exogenous polynucleotide with which the microorganism has been transformed. In some embodiments, cysteate synthase, optionally L-serine dehydratase, and CSAD enzymes or enzyme activities are encoded by one or more exogenous polynucleotide(s) (e.g., one, two, or three, exogenous polynucleotide(s)) in the microorganism. In some embodiments, one of cysteate synthase, L-serine dehydratase, or CSAD enzymes or enzyme activities is encoded by an exogenous polynucleotide in the microorganism and the remaining activity/ies is/are endogenously expressed in the host cell from which the non-naturally occurring microorganism is derived, *i.e.*, not expressed from exogenous polynucleotide(s). In one embodiment, the microorganism expresses endogenous CSAD and optionally L-serine dehydratase activity, and cysteate synthase (*e.g.*, MA_3297) is encoded by an exogenous polynucleotide in the microorganism. In one embodiment, the microorganism expresses an endogenous cysteate synthase and CSAD is encoded by an exogenous polynucleotide in the microorganism.

[129] Non-naturally occurring microorganisms are provided that include the following enzymes or enzyme activities: cysteate synthase, *e.g.*, MA_3297 (EC 2.5.1.76); optionally L-serine dehydratase (EC 4.3.1.17); and glutamate decarboxylase (GAD) (EC 4.1.1.15), wherein at least one of these enzymes or enzyme activities is encoded by an exogenous polynucleotide with which the microorganism has been transformed. In some embodiments, cysteate synthase, optionally L-serine dehydratase, and GAD enzymes or enzyme activities are encoded by one or more exogenous polynucleotide(s) (e.g., one, two, or three, exogenous polynucleotide(s)) in the microorganism. In some embodiments, one of cysteate synthase, L-serine dehydratase, or GAD enzymes or enzyme activities is encoded by an exogenous polynucleotide in the microorganism and the remaining activity/ies is/are endogenously expressed in the host cell from which the non-naturally occurring microorganism is derived, *i.e.*, not expressed from exogenous polynucleotide(s). In one embodiment, the microorganism expresses endogenous GAD and optionally L-serine dehydratase activity and cysteate synthase (*e.g.*, MA_3297) is encoded by an exogenous polynucleotide in the microorganism. In one embodiment, the microorganism expresses an

endogenous cysteate synthase and GAD is encoded by an exogenous polynucleotide in the microorganism.

[130] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes cysteate synthase comprising or consisting of the amino acid sequence depicted in SEQ ID NO:17, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:17. In some embodiments, the polynucleotide that encodes cysteate synthase comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:18, SEQ ID NO:52, or SEQ ID NO:64 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:18, SEQ ID NO:52, or SEQ ID NO:64. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:52 or SEQ ID NO:64.

[131] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes L-serine dehydratase comprising or consisting of the amino acid sequence depicted in SEQ ID NO:1, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:1. In some embodiments, the polynucleotide that encodes L-serine dehydratase comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:2 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:2. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism.

[132] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes CSAD comprising or consisting of the amino acid sequence depicted in SEQ ID NO:11, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:11. In some embodiments, the polynucleotide that encodes CSAD comprises or consists of the polynucleotide sequence

depicted in SEQ ID NO:12, SEQ ID NO:53, or SEQ ID NO:54 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:12, SEQ ID NO:53, or SEQ ID NO:54. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:53 or SEQ ID NO:54.

[133] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes GAD comprising or consisting of the amino acid sequence depicted in SEQ ID NO:13, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:13. In some embodiments, the polynucleotide that encodes GAD comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:14 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:14. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism.

[134] In one embodiment, a non-naturally occurring *Methylobacterium* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding cysteate synthase (*e.g.*, MA_3297) and CSAD. In another embodiment, a non-naturally occurring *Methylobacterium* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding cysteate synthase (*e.g.*, MA_3297) and GAD.

[135] In one embodiment, a non-naturally occurring *Escherichia* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding cysteate synthase (*e.g.*, MA_3297) and optionally L-serine dehydratase. In one embodiment, a non-naturally occurring *Escherichia* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding cysteate synthase (*e.g.*, MA_3297), optionally L-serine dehydratase, and CSAD.

[136] In one embodiment, a non-naturally occurring *Saccharomyces* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding cysteate synthase (*e.g.*, MA_3297) and optionally L-serine dehydratase. In one embodiment, a non-naturally occurring *Saccharomyces* microorganism is provided that includes one or more exogenous

polynucleotide(s) encoding cysteate synthase (*e.g.*, MA_3297), optionally L-serine dehydratase, and CSAD.

[137] In one embodiment, a non-naturally occurring *Bacillus* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding cysteate synthase (*e.g.*, MA_3297) and optionally L-serine dehydratase. In one embodiment, a non-naturally occurring *Bacillus* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding cysteate synthase (*e.g.*, MA_3297), optionally L-serine dehydratase, and CSAD.

Conversion of cysteine to taurine, via CGL/CD; PAPS-AS, and CSAD or GAD

[138] In some embodiments, a non-naturally occurring microorganism is provided that expresses one or more exogenous enzyme activity/ies for the conversion of cysteine to taurine via the enzymes cystathionine gamma-lyase (CGL)/ L-cysteine desulfhydrase (CD), 3'phosphoadenylyl sulfate:2-aminoacrylate C-sulfotransferase (PAPS-AS), and cysteine sulfinic acid decarboxylase (CSAD) or glutamate decarboxylase (GAD). This pathway for taurine biosynthesis is shown schematically in **Fig. 6**. Several proteins have been found to have L-cysteine desulfhydrase (CD) activity (EC 4.4.1.1) including cystathionine gamma-lyase (CGL), tryptophanase, cysteine synthases, and MalY (Awano et al. (2005) *Appl Environ Microbiol* 71(7):4149-52.). In some embodiments, a single enzyme includes both CGL and CD activities. In other embodiments, CGL and CD activities are provided by two separate enzymes. In some embodiments, CD activity is provided by an enzyme and CGL activity is absent. When CGL activity is present, it may provide greater flux via production of L-cysteine, which serves as a substrate for CD activity. In some embodiments, a first enzyme that includes both CGL and CD activities and a second enzyme that includes only CD activity are provided.

[139] Non-naturally occurring microorganisms are provided that include the following enzymes or enzyme activities: cystathionine gamma-lyase (CGL)/L-cysteine desulfhydrase (CD) (EC 4.4.1.1); sulfate adenyltransferase (EC 2.7.7.4) and adenylyl-sulfate kinase (APS kinase) (EC 2.7.1.25), and/or 3'-phosphoadenosine 5'-phosphosulfate synthase (PAPSS1) (EC 2.7.7.4/EC 2.7.1.25); 3'- phosphoadenylyl sulfate: 2-aminoacrylate C-sulfotransferase (PAPS-AS) (EC 2.8.2.-); and cysteine sulfinic acid decarboxylase (CSAD) (EC 4.1.1.29), wherein at least one of these enzymes or enzyme activities is encoded by an exogenous polynucleotide with which the microorganism has been transformed. In some

embodiments, CGL/CD; sulfate adenyltransferase and APS kinase, and/or PAPSS1; PAPS-AS; and CSAD enzymes or enzyme activities are encoded by one or more exogenous polynucleotide(s) (e.g., one, two, three, four, five, six, or seven exogenous polynucleotide(s)) in the microorganism. In some embodiments, one, two, three, four, five, or six of CGL/CD, sulfate adenyltransferase, APS kinase, PAPSS1, PAPS-AS, and CSAD enzymes or enzyme activities is encoded by one or more exogenous polynucleotide(s) in the microorganism and the remaining activity/ies is/are endogenously expressed in the host cell from which the non-naturally occurring microorganism is derived, *i.e.*, not expressed from exogenous polynucleotide(s). In one embodiment, the microorganism expresses an endogenous activity, and sulfate adenyltransferase and APS kinase and/or PAPSS1, PAPS-AS, and CSAD are expressed from exogenous polynucleotide(s). In one embodiment, the microorganism expresses endogenous CGL/CD and sulfate adenyltransferase, APS kinase activities, and PAPS-AS and CSAD activities are expressed from exogenous polynucleotide(s). In one embodiment, the microorganism expresses endogenous CGL/CD, sulfate adenyltransferase, APS kinase, and CSAD activities, and PAPS-AS is expressed from an exogenous polynucleotide.

[140] Non-naturally occurring microorganisms are provided that include the following enzymes or enzyme activities: cystathionine gamma-lyase (CGL)/ L-cysteine desulfhydrase (CD) (EC 4.4.1.1); sulfate adenyltransferase (EC 2.7.7.4) and adenylyl-sulfate kinase (APS kinase) (EC 2.7.1.25), and/or 3'-phosphoadenosine 5'-phosphosulfate synthase (PAPSS1) (EC 2.7.7.4/EC 2.7.1.25); 3'- phosphoadenylyl sulfate: 2-aminoacrylate C-sulfotransferase (PAPS-AS) (EC 2.8.2.-); and glutamate decarboxylase (GAD) (EC 4.1.1.15), wherein at least one of these enzymes or enzyme activities is encoded by an exogenous polynucleotide with which the microorganism has been transformed. In some embodiments, CGL/CD, sulfate adenyltransferase and APS kinase, and/or PAPSS1, PAPS-AS, and GAD enzymes or enzyme activities are encoded by one or more exogenous polynucleotide(s) (e.g., one, two, three, four, five, six, or seven exogenous polynucleotide(s)) in the microorganism. In some embodiments, one, two, three, four, five, or six of CGL/CD, sulfate adenyltransferase, APS kinase, PAPSS1, PAPS-AS, and GAD enzymes or enzyme activities is encoded by one or more exogenous polynucleotide(s) in the microorganism and the remaining activity/ies is/are endogenously expressed in the host cell from which the non-naturally occurring microorganism is derived, *i.e.*, not expressed from exogenous polynucleotide(s). In one

embodiment, the microorganism expresses an endogenous activity, and sulfate adenyltransferase and APS kinase, and/or PAPSS1, PAPS-AS, and GAD are expressed from exogenous polynucleotide(s). In one embodiment, the microorganism expresses endogenous CGL/CD, sulfate adenyltransferase, and APS kinase activities, and PAPS-AS and GAD activities are expressed from exogenous polynucleotide(s). In one embodiment, the microorganism expresses endogenous CGL/CD, sulfate adenyltransferase, APS kinase, and GAD activities, and PAPS-AS is expressed from an exogenous polynucleotide.

[141] In some embodiments, the non-naturally occurring microorganism includes one or more exogenous polynucleotide(s) that encode(s) CGL/CD comprising or consisting of the amino acid sequence(s) depicted in SEQ ID NO:46, SEQ ID NO:70, and/or SEQ ID NO:72 or variant(s) or homolog(s) thereof. In some embodiments, the exogenous polynucleotide(s) encode(s) polypeptide(s) comprising or consisting of amino acid sequence(s) having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:46, SEQ ID NO:70, and/or SEQ ID NO:72. In some embodiments, the polynucleotide(s) that encode(s) CGL/CD comprise or consist of the polynucleotide sequence(s) depicted in SEQ ID NO:47, SEQ ID NO:65, SEQ ID NO:71 and/or SEQ ID NO:73, or polynucleotide(s) having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:47, SEQ ID NO:65, SEQ ID NO:71, and/or SEQ ID NO:73. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:65.

[142] In some embodiments, the non-naturally occurring microorganism includes exogenous polynucleotides that encode sulfate adenyltransferase comprising or consisting of the amino acid sequences depicted in SEQ ID NO:3 and SEQ ID NO:5, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotides encode polypeptides comprising or consisting of an amino acid sequences having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:3 or SEQ ID NO:5. In some embodiments, the polynucleotides that encode sulfate adenyltransferase comprises or consists of the polynucleotide sequences depicted in SEQ ID NO:4 or SEQ ID NO:6 or polynucleotides having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:4 or SEQ ID NO:6. In some embodiments, the polynucleotide sequences are codon optimized for expression in the microorganism.

[143] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes APS kinase comprising or consisting of the amino acid sequence depicted in SEQ ID NO:7, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:7. In some embodiments, the polynucleotide that encodes APS kinase comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:8 or SEQ ID NO:62 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:8 or SEQ ID NO:62. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:62.

[144] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes PAPSS1 comprising or consisting of the amino acid sequence depicted in SEQ ID NO:31, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:31. In some embodiments, the polynucleotide that encodes PAPSS1 comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:32 or SEQ ID NO:63 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:32 or SEQ ID NO:63. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:63.

[145] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes PAPS-AS comprising or consisting of the amino acid sequence depicted in SEQ ID NO:9, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:9. In some embodiments, the polynucleotide that encodes PAPS-AS comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:10 or SEQ ID NO:61 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:10 or

SEQ ID NO:61. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:61.

[146] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes CSAD comprising or consisting of the amino acid sequence depicted in SEQ ID NO:11, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:11. In some embodiments, the polynucleotide that encodes CSAD comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:12, SEQ ID NO:53, or SEQ ID NO:54 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:12, SEQ ID NO:53, or SEQ ID NO:54. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:53 or SEQ ID NO:54.

[147] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes GAD comprising or consisting of the amino acid sequence depicted in SEQ ID NO:13, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:13. In some embodiments, the polynucleotide that encodes GAD comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:14 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:14. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism.

[148] In one embodiment, a non-naturally occurring *Methylobacterium* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding: CGL/CD, sulfate adenyltransferase and APS kinase, and/or PAPSS1; PAPS-AS, and CSAD. In one embodiment, a non-naturally occurring *Methylobacterium* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding CGL/CD, PAPS-AS and CSAD. In another embodiment, a non-naturally occurring *Methylobacterium* microorganism is provided that includes one or more exogenous polynucleotide(s)

encoding: CGL/CD; sulfate adenyltransferase and APS kinase, and/or PAPSS1, PAPS-AS, and GAD. In one embodiment, a non-naturally occurring *Methylobacterium* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding CGL/CD, PAPS-AS and GAD.

[149] In one embodiment, a non-naturally occurring *Escherichia* microorganism is provided that includes an exogenous polynucleotide encoding PAPS-AS. In one embodiment, a non-naturally occurring *Escherichia* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding PAPSS1 and PAPS-AS. In one embodiment, a non-naturally occurring *Escherichia* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding PAPSS1, PAPS-AS, and CSAD.

[150] In one embodiment, a non-naturally occurring *Saccharomyces* microorganism is provided that includes an exogenous polynucleotide encoding PAPS-AS. In one embodiment, a non-naturally occurring *Saccharomyces* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding PAPSS1 and PAPS-AS. In one embodiment, a non-naturally occurring *Saccharomyces* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding PAPSS1, PAPS-AS, and CSAD.

[151] In one embodiment, a non-naturally occurring *Bacillus* microorganism is provided that includes an exogenous polynucleotide encoding PAPS-AS. In one embodiment, a non-naturally occurring *Bacillus* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding PAPSS1 and PAPS-AS. In one embodiment, a non-naturally occurring *Bacillus* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding PAPSS1, PAPS-AS, and CSAD.

Conversion of cysteine to taurine via CGL/CD and cysteate synthase

[152] In some embodiments, a non-naturally occurring microorganism is provided that expresses one or more exogenous enzyme activity/ies for the conversion of cysteine to taurine via the enzymes cystathionine gamma-lyase/ L-cysteine desulfhydrase, cysteate synthase (*e.g.*, MA_3297), and cysteine sulfinic acid decarboxylase (CSAD) or glutamate decarboxylase (GAD). The CGL/CD, cysteate synthase (*e.g.*, MA_3297), CSAD or GAD pathway for taurine biosynthesis is shown schematically in **Fig. 7**. Several proteins have been found to have L-cysteine desulfhydrase (CD) activity (EC 4.4.1.1) including

cystathionine gamma-lyase (CGL), tryptophanase, cysteine synthases, and MalY (Awano et al. (2005) *Appl Environ Microbiol* 71(7):4149-52.). In some embodiments, a single enzyme includes both CGL and CD activities. In other embodiments, CGL and CD activities are provided by two separate enzymes. In some embodiments, CD activity is provided by an enzyme and CGL activity is absent. When CGL activity is present, it may provide greater flux via production of L-cysteine, which serves as a substrate for CD activity. In some embodiments, a first enzyme that includes both CGL and CD activities and a second enzyme that includes only CD activity are provided.

[153] Non-naturally occurring microorganisms are provided that include the following enzymes or enzyme activities: cystathionine gamma-lyase (CGL)/ L-cysteine desulfhydrase (CD) (EC4.4.4.1), cysteate synthase, *e.g.*, MA_3297 (EC 2.5.1.76); and cysteine sulfinic acid decarboxylase (CSAD) (EC 4.1.1.29), wherein at least one of these enzymes or enzyme activities is encoded by an exogenous polynucleotide with which the microorganism has been transformed. In some embodiments, CGL/CD, cysteate synthase, and CSAD enzymes or enzyme activities are encoded by one or more exogenous polynucleotide(s) (e.g., one, two, three, or four exogenous polynucleotide(s)) in the microorganism. In some embodiments, one or more of CGL/CD, cysteate synthase, and CSAD enzymes or enzyme activities is encoded by one or more exogenous polynucleotide(s) in the microorganism and the remaining activity/ies is/are endogenously expressed in the host cell from which the non-naturally occurring microorganism is derived, *i.e.*, not expressed from exogenous polynucleotide(s). In one embodiment, the microorganism expresses an endogenous CSAD activity and CGL/CD and cysteate synthase (*e.g.*, MA_3297) are encoded by exogenous polynucleotide(s) in the microorganism. In one embodiment, the microorganism expresses endogenous CGL/CD and CSAD activities and cysteate synthase (*e.g.*, MA_3297) is encoded by an exogenous polynucleotide in the microorganism.

[154] Non-naturally occurring microorganisms are provided that include the following enzymes or enzyme activities: cystathionine gamma-lyase (CGL)/ L-cysteine desulfhydrase (CD) (EC4.4.4.1), cysteate synthase, *e.g.*, MA_3297 (EC 2.5.1.76); and glutamate decarboxylase (GAD) (EC 4.1.1.15), wherein at least one of these enzymes or enzyme activities is encoded by an exogenous polynucleotide with which the microorganism has been transformed. In some embodiments, CGL/CD, cysteate synthase, and GAD enzymes

or enzyme activities are encoded by one or more exogenous polynucleotide(s) (e.g., one, two, three, or four exogenous polynucleotide(s)) in the microorganism. In some embodiments, one or more of CGL/CD, cysteate synthase, and GAD enzymes or enzyme activities is encoded by one or more exogenous polynucleotide(s) in the microorganism and the remaining activity/ies is/are endogenously expressed in the host cell from which the non-naturally occurring microorganism is derived, *i.e.*, not expressed from exogenous polynucleotide(s). In one embodiment, the microorganism expresses an endogenous GAD activity and CGL/CD and cysteate synthase (*e.g.*, MA_3297) are encoded by exogenous polynucleotide(s) in the microorganism. In one embodiment, the microorganism expresses endogenous CGL/CD and GAD activities and cysteate synthase (*e.g.*, MA_3297) is encoded by an exogenous polynucleotide in the microorganism.

[155] In some embodiments, the non-naturally occurring microorganism includes one or more exogenous polynucleotide(s) that encode(s) CGL/CD comprising or consisting of the amino acid sequence(s) depicted in SEQ ID NO:46, SEQ ID NO:70, and/or SEQ ID NO:72 or variant(s) or homolog(s) thereof. In some embodiments, the exogenous polynucleotide(s) encode(s) polypeptide(s) comprising or consisting of amino acid sequence(s) having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:46, SEQ ID NO:70, and/or SEQ ID NO:72. In some embodiments, the polynucleotide(s) that encode(s) CGL/CD comprise or consist of the polynucleotide sequence(s) depicted in SEQ ID NO:47, SEQ ID NO:65, SEQ ID NO:71, and/or SEQ ID NO:73 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:47, SEQ ID NO:65, SEQ ID NO:71, and/or SEQ ID NO:73. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:65.

[156] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes cysteate synthase comprising or consisting of the amino acid sequence depicted in SEQ ID NO:17, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:17. In some embodiments, the polynucleotide that encodes cysteate synthase comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:18, SEQ ID NO:52, or SEQ ID NO:64 or a

polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:18, SEQ ID NO:52, or SEQ ID NO:64. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:52 or SEQ ID NO:64.

[157] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes CSAD comprising or consisting of the amino acid sequence depicted in SEQ ID NO:11, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:11. In some embodiments, the polynucleotide that encodes CSAD comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:12, SEQ ID NO:53, or SEQ ID NO:54 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:12, SEQ ID NO:53, or SEQ ID NO:54. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:53 or SEQ ID NO:54.

[158] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes GAD comprising or consisting of the amino acid sequence depicted in SEQ ID NO:13, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:13. In some embodiments, the polynucleotide that encodes GAD comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:14 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:14. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism.

[159] In one embodiment, a non-naturally occurring *Methylobacterium* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding CGL/CD, cysteate synthase (*e.g.*, MA_3297), and CSAD. In another embodiment, a non-naturally occurring *Methylobacterium* microorganism is provided that includes one or more

exogenous polynucleotide(s) encoding CGL/CD, cysteate synthase (*e.g.*, MA_3297), and GAD.

[160] In one embodiment, a non-naturally occurring *Escherichia* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding CGL/CD and cysteate synthase (*e.g.*, MA_3297). In one embodiment, a non-naturally occurring *Escherichia* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding CGL/CD, cysteate synthase, and CSAD.

[161] In one embodiment, a non-naturally occurring *Saccharomyces* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding CGL/CD and cysteate synthase (*e.g.*, MA_3297). In one embodiment, a non-naturally occurring *Saccharomyces* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding CGL/CD, cysteate synthase, and CSAD.

[162] In one embodiment, a non-naturally occurring *Bacillus* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding cysteate synthase (*e.g.*, MA_3297) and CSAD.

Conversion of pyruvate to taurine

[163] In some embodiments, a non-naturally occurring microorganism is provided that expresses exogenous enzyme activity/ies for the conversion of pyruvate to taurine via the enzyme L-cysteate sulfo-lyase (cuyA), for example, exogenous enzyme(s) of the L-cysteate sulfo-lyase (cuyA)/CSAD or GAD pathway for biosynthesis of taurine. The cuyA/CSAD or GAD pathway for taurine biosynthesis is shown schematically in **Fig. 8**.

[164] Non-naturally occurring microorganisms are provided that include the following enzymes or enzyme activities: L-cysteate sulfo-lyase (cuyA) (EC 4.4.1.25); and cysteine sulfinic acid decarboxylase (CSAD) (EC 4.1.1.29), wherein at least one of these enzymes or enzyme activities is encoded by an exogenous polynucleotide with which the microorganism has been transformed. In some embodiments, cuyA and CSAD enzymes or enzyme activities are encoded by one or more exogenous polynucleotide(s) (e.g., one or two exogenous polynucleotide(s)) in the microorganism. In some embodiments, one of cuyA and CSAD enzymes or enzyme activities is encoded by an exogenous polynucleotide in the microorganism and the remaining activity is endogenously expressed in the host cell from which the non-naturally occurring microorganism is derived, *i.e.*, not expressed from an exogenous polynucleotide. In one embodiment, the microorganism expresses an

endogenous CSAD activity and cuyA is encoded by an exogenous polynucleotide in the microorganism. In one embodiment, the microorganism expresses an endogenous cuyA activity and CSAD is encoded by an exogenous polynucleotide in the microorganism.

[165] Non-naturally occurring microorganisms are provided that include the following enzymes or enzyme activities: L-cysteate sulfo-lyase (cuyA) (EC 4.4.1.25); and glutamate decarboxylase (GAD) (EC 4.1.1.15), wherein at least one of these enzymes or enzyme activities is encoded by an exogenous polynucleotide with which the microorganism has been transformed. In some embodiments, cuyA and GAD enzymes or enzyme activities are encoded by one or more exogenous polynucleotide(s) (e.g., one or two exogenous polynucleotide(s)) in the microorganism. In some embodiments, one of cuyA and GAD enzymes or enzyme activities is encoded by an exogenous polynucleotide in the microorganism and the remaining activity is endogenously expressed in the host cell from which the non-naturally occurring microorganism is derived, *i.e.*, not expressed from an exogenous polynucleotide. In one embodiment, the microorganism expresses an endogenous GAD activity and cuyA is encoded by an exogenous polynucleotide in the microorganism. In one embodiment, the microorganism expresses an endogenous cuyA activity and GAD is encoded by an exogenous polynucleotide in the microorganism.

[166] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes cuyA comprising or consisting of the amino acid sequence depicted in SEQ ID NO:37, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:37. In some embodiments, the polynucleotide that encodes cuyA comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:38 or SEQ ID NO:66 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:38 or SEQ ID NO:66. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:66.

[167] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes CSAD comprising or consisting of the amino acid sequence depicted in SEQ ID NO:11, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or

consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:11. In some embodiments, the polynucleotide that encodes CSAD comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:12, SEQ ID NO:53, or SEQ ID NO:54 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:12, SEQ ID NO:53, or SEQ ID NO:54. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:53 or SEQ ID NO:54.

[168] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes GAD comprising or consisting of the amino acid sequence depicted in SEQ ID NO:13, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:13. In some embodiments, the polynucleotide that encodes GAD comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:14 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:14. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism.

[169] In one embodiment, a non-naturally occurring *Methylobacterium* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding cuyA and CSAD. In another embodiment, a non-naturally occurring *Methylobacterium* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding cuyA and GAD.

[170] In one embodiment, a non-naturally occurring *Escherichia* microorganism is provided that includes an exogenous polynucleotide encoding cuyA. In one embodiment, a non-naturally occurring *Escherichia* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding cuyA and CSAD.

[171] In one embodiment, a non-naturally occurring *Saccharomyces* microorganism is provided that includes an exogenous polynucleotide encoding cuyA. In one embodiment, a non-naturally occurring *Saccharomyces* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding cuyA and CSAD.

[172] In one embodiment, a non-naturally occurring *Bacillus* microorganism is provided that includes an exogenous polynucleotide encoding cuyA. In one embodiment, a non-naturally occurring *Bacillus* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding cuyA and CSAD.

Conversion of phosphoenolpyruvate to taurine via CSAD or GAD

[173] In some embodiments, a non-naturally occurring microorganism is provided that expresses one or more exogenous enzyme activity/ies for the conversion of phosphoenolpyruvate to taurine via the enzymes phosphosulfolactate synthase (ComA), 2-phsopho-e-sulfolactate dehydrogenase (ComB), sulfolactate dehydrogenase (ComC), and aspartate aminotransferase (AspAT), for example, one or more exogenous enzyme(s) of the ComA/ComB/ComC/AspAT/CSAD or GAD pathway for biosynthesis of taurine. The ComA/ComB/ComC/AspAT/CSAD or GAD pathway for taurine biosynthesis is shown schematically in **Fig. 9**.

[174] Non-naturally occurring microorganisms are provided that include the following enzymes or enzyme activities: phosphosulfolactate synthase (ComA) (EC 4.4.1.19), 2-phospho-3-sulfolactate phosphohydrolase (ComB) (EC 3.1.3.71), sulfolactate dehydrogenase (ComC) (EC 1.1.1.337), aspartate aminotransferase (AspAT) (EC 2.6.1.1), and cysteine sulfinic acid decarboxylase (CSAD) (EC 4.1.1.29), wherein at least one of these enzymes or enzyme activities is encoded by an exogenous polynucleotide with which the microorganism has been transformed. In some embodiments, ComA, ComB, ComC, AspAT, and CSAD enzymes or enzyme activities are encoded by one or more exogenous polynucleotide(s) (e.g., one, two, three, four, or five exogenous polynucleotide(s)) in the microorganism. In some embodiments, one or more of ComA, ComB, ComC, AspAT, and CSAD enzymes or enzyme activities is encoded by one or more exogenous polynucleotide(s) in the microorganism and the remaining activity/ies is/are endogenously expressed in the host cell from which the non-naturally occurring microorganism is derived, *i.e.*, not expressed from exogenous polynucleotide(s). In one embodiment, the microorganism expresses endogenous CSAD and AspAT activities, and ComA, ComB, and ComC are encoded by exogenous polynucleotide(s) in the microorganism. In one embodiment, the microorganism expresses endogenous AspAT activity, and ComA, ComB, ComC, and CSAD are encoded by exogenous polynucleotide(s) in the microorganism.

[175] Non-naturally occurring microorganisms are provided that include the following enzymes or enzyme activities: phosphosulfolactate synthase (ComA) (EC 4.4.1.19), 2-phospho-3-sulfolactate phosphohydrolase (ComB) (EC 3.1.3.71), sulfolactate dehydrogenase (ComC) (EC 1.1.1.337), aspartate aminotransferase (AspAT) (EC 2.6.1.1), and glutamate decarboxylase (GAD) (EC 4.1.1.15), wherein at least one of these enzymes or enzyme activities is encoded by an exogenous polynucleotide with which the microorganism has been transformed. In some embodiments, ComA, ComB, ComC, AspAT, and GAD enzymes or enzyme activities are encoded by one or more exogenous polynucleotide(s) (e.g., one, two, three, four, or five exogenous polynucleotide(s)) in the microorganism. In some embodiments, one or more of ComA, ComB, ComC, AspAT, and GAD enzymes or enzyme activities is encoded by one or more exogenous polynucleotide(s) in the microorganism and the remaining activity/ies is/are endogenously expressed in the host cell from which the non-naturally occurring microorganism is derived, *i.e.*, not expressed from exogenous polynucleotide(s). In one embodiment, the microorganism expresses endogenous GAD and AsPAT activities, and ComA, ComB, and ComC are encoded by exogenous polynucleotide(s) in the microorganism. In one embodiment, the microorganism expresses endogenous AsPAT activity, and ComA, ComB, ComC, and GAD are encoded by exogenous polynucleotide(s) in the microorganism.

[176] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes ComA comprising or consisting of the amino acid sequence depicted in SEQ ID NO:19, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:19. In some embodiments, the polynucleotide that encodes ComA comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:20 or SEQ ID NO:67 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:20 or SEQ ID NO:67. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:67.

[177] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes ComB comprising or consisting of the amino acid sequence depicted in SEQ ID NO:21, or a variant or homolog thereof. In some

embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:21. In some embodiments, the polynucleotide that encodes ComB comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:22 or SEQ ID NO:68 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:22 or SEQ ID NO:68. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:68.

[178] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes ComC comprising or consisting of the amino acid sequence depicted in SEQ ID NO:23, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:23. In some embodiments, the polynucleotide that encodes ComC comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:24 or SEQ ID NO:69 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:24 or SEQ ID NO:69. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.* SEQ ID NO:69.

[179] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes CSAD comprising or consisting of the amino acid sequence depicted in SEQ ID NO:11, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:11. In some embodiments, the polynucleotide that encodes CSAD comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:12, SEQ ID NO:53, or SEQ ID NO:54 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:12, SEQ ID NO:53, or SEQ ID NO:54. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:53 or SEQ ID NO:54.

[180] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes GAD comprising or consisting of the amino acid sequence depicted in SEQ ID NO:13, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:13. In some embodiments, the polynucleotide that encodes GAD comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:14 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:14. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism.

[181] In one embodiment, a non-naturally occurring *Methylobacterium* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding ComA, ComB, ComC, and CSAD. In another embodiment, a non-naturally occurring *Methylobacterium* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding ComA, ComB, ComC, and GAD.

[182] In one embodiment, a non-naturally occurring *Escherichia* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding ComA, ComB, and ComC. In one embodiment, a non-naturally occurring *Escherichia* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding ComA, ComB, ComC, and CSAD.

[183] In one embodiment, a non-naturally occurring *Saccharomyces* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding ComA, ComB, and ComC. In one embodiment, a non-naturally occurring *Saccharomyces* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding ComA, ComB, ComC, and CSAD.

[184] In one embodiment, a non-naturally occurring *Bacillus* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding ComA, ComB, and ComC. In one embodiment, a non-naturally occurring *Bacillus* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding ComA, ComB, ComC, and CSAD.

Conversion of acetyl phosphate to taurine

[185] In some embodiments, a non-naturally occurring microorganism is provided that expresses one or more exogenous enzyme activity/ies for the conversion of acetyl phosphate to taurine, for example, enzymes of the Xsc/Tpa pathway for biosynthesis of taurine. The Xsc/Tpa pathway of taurine biosynthesis is shown schematically in **Fig. 10**.

[186] Non-naturally occurring microorganisms are provided that include the following enzymes or enzyme activities: sulfoacetaldehyde acetyltransferase (Xsc) (EC 2.3.3.15); and taurine-pyruvate aminotransferase (Tpa) (EC 2.6.1.77), wherein at least one of these enzymes or enzyme activities is encoded by an exogenous polynucleotide with which the microorganism has been transformed. In some embodiments, Xsc and Tpa enzymes or enzyme activities are encoded by one or more exogenous polynucleotides in the microorganism (e.g., one or two exogenous polynucleotide(s). In some embodiments, one of Xsc and Tpa enzymes or enzyme activities is encoded an exogenous polynucleotide in the microorganism and the remaining activity is endogenously expressed in the host cell from which the non-naturally occurring microorganism is derived, *i.e.*, not expressed from an exogenous polynucleotide.

[187] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes Xsc comprising or consisting of the amino acid sequence depicted in SEQ ID NO:48, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:48. In some embodiments, the polynucleotide that encodes Xsc comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:49 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:49. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism.

[188] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes Tpa comprising or consisting of the amino acid sequence depicted in SEQ ID NO:27, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70,

75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:27. In some embodiments, the polynucleotide that encodes Tpa comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:28 or SEQ ID NO:56 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:28 or SEQ ID NO:56. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:56.

[189] In one embodiment, a non-naturally occurring *Methylobacterium* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding Xsc and Tpa.

[190] In one embodiment, a non-naturally occurring *Escherichia* microorganism is provided that includes an exogenous polynucleotide encoding Xsc and Tpa.

[191] In one embodiment, a non-naturally occurring *Saccharomyces* microorganism is provided that includes an exogenous polynucleotide encoding Xsc and Tpa.

[192] In one embodiment, a non-naturally occurring *Bacillus* microorganism is provided that includes an exogenous polynucleotide encoding Xsc and Tpa.

Conversion of phosphoenolpyruvate to taurine via Tpa

[193] In some embodiments, a non-naturally occurring microorganism is provided that expresses one or more exogenous enzyme activity/ies for the conversion of phosphoenolpyruvate to taurine via the enzymes phosphosulfolactate synthase (ComA), 2-phsopho-e-sulfolactate dehydrogenase (ComB), sulfolactate dehydrogenase (ComC), sulfopyruvate decarboxylase (ComDE), and taurine-pyruvate aminotransferase (Tpa), for example, one or more exogenous enzyme(s) of the ComA/ComB/ComC, ComDE/Tpa pathway for biosynthesis of taurine. This pathway for taurine biosynthesis is shown schematically in **Fig. 11**.

[194] Non-naturally occurring microorganisms are provided that include the following enzymes or enzyme activities: phosphosulfolactate synthase (ComA) (EC 4.4.1.19), 2-phospho-3-sulfolactate phosphohydrolase (ComB) (EC 3.1.3.71), sulfolactate dehydrogenase (ComC) (EC 1.1.1.337), sulfopyruvate decarboxylase (ComDE) (EC 4.1.7.9), and taurine-pyruvate aminotransferase (Tpa) (EC 2.6.1.77), wherein at least one of these enzymes or enzyme activities is encoded by an exogenous polynucleotide with which the microorganism has been transformed. In some embodiments, ComA, ComB, ComC, ComDE, and Tpa enzymes or enzyme activities are encoded by one or more exogenous polynucleotide(s) (e.g., one, two, three, four, or five exogenous polynucleotide(s)) in the

microorganism. In some embodiments, one or more of ComA, ComB, ComC, ComDE, and Tpa enzymes or enzyme activities is encoded by one or more exogenous polynucleotide(s) in the microorganism and the remaining activity/ies is/are endogenously expressed in the host cell from which the non-naturally occurring microorganism is derived, *i.e.*, not expressed from exogenous polynucleotide(s).

[195] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes ComA comprising or consisting of the amino acid sequence depicted in SEQ ID NO:19, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:19. In some embodiments, the polynucleotide that encodes ComA comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:20 or SEQ ID NO:67 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:20 or SEQ ID NO:67. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:67.

[196] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes ComB comprising or consisting of the amino acid sequence depicted in SEQ ID NO:21, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:21. In some embodiments, the polynucleotide that encodes ComB comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:22 or SEQ ID NO:68 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:22 or SEQ ID NO:68. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:68.

[197] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes ComC comprising or consisting of the amino acid sequence depicted in SEQ ID NO:23, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70,

75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:23. In some embodiments, the polynucleotide that encodes ComC comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:24 or SEQ ID NO:69 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:24 or SEQ ID NO:69. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:69.

[198] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes ComDE comprising or consisting of the amino acid sequence depicted in SEQ ID NO:25, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:25. In some embodiments, the polynucleotide that encodes ComDE comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:26 or SEQ ID NO:55 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:26 or SEQ ID NO:55. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:55.

[199] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes Tpa comprising or consisting of the amino acid sequence depicted in SEQ ID NO:27, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:27. In some embodiments, the polynucleotide that encodes Tpa comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:28 or SEQ ID NO:56 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:28 or SEQ ID NO:56. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:56.

[200] In one embodiment, a non-naturally occurring *Methylobacterium* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding ComA, ComB, ComC, and ComDE, and Tpa.

[201] In one embodiment, a non-naturally occurring *Escherichia* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding ComA, ComB, ComC, ComDE, and Tpa.

[202] In one embodiment, a non-naturally occurring *Saccharomyces* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding ComA, ComB, ComC, ComDE, and Tpa.

[203] In one embodiment, a non-naturally occurring *Bacillus* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding ComA, ComB, ComC, ComDE, and Tpa.

Conversion of cysteate to taurine via AspAT, ComDE, and Tpa

[204] Cysteate, which is an intermediate in some of the biosynthetic pathways described above (see **Figs. 4**, **5**, **6**, **7**, and **8**) be converted to taurine via aspartate aminotransferase (AspAT) (EC 2.6.1.1), which converts cysteate to sulfopyruvate, sulfopyruvate decarboxylase (ComDE) (EC4.1.7.9), which converts sulfopyruvate to sulfoacetaldehyde, and taurine-pyruvate aminotransferase (Tpa) (EC 2.67.1.77), which converts sulfoacetaldehyde to taurine, instead of or in addition to CSAD or GAD. This is shown schematically in **Fig. 12**.

[205] Non-naturally occurring microorganisms are provided that include the following enzymes or enzyme activities: aspartate aminotransferase (AspAT) (EC 2.6.1.1); sulfopyruvate decarboxylase (ComDE) (EC4.1.7.9); and taurine-pyruvate aminotransferase (Tpa) (EC 2.67.1.77), wherein at least one of these enzymes or enzyme activities is encoded by an exogenous polynucleotide with which the microorganism has been transformed, and optionally other enzyme activities for production of cysteate, as shown in **Fig. 12**, either endogenous to the microorganism or encoded by exogenous polynucleotide(s) with which the microorganism has been transformed.

[206] In some embodiments, one or more of AspAT, ComDE, and Tpa enzymes or enzyme activities is encoded by exogenous polynucleotide(s) in the microorganism and the remaining activity/ies is/are endogenously expressed in the host cell from which the non-naturally occurring microorganism is derived, *i.e.*, not expressed from an exogenous polynucleotide. In one embodiment, the microorganism expresses an endogenous AspAT activity and ComDE and Tpa are encoded by exogenous polynucleotide(s) in the microorganism.

[207] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes ComDE comprising or consisting of the amino acid sequence depicted in SEQ ID NO:25, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:25. In some embodiments, the polynucleotide that encodes ComDE comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:26 or SEQ ID NO:55 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:26 or SEQ ID NO:55. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:55.

[208] In some embodiments, the non-naturally occurring microorganism includes an exogenous polynucleotide that encodes Tpa comprising or consisting of the amino acid sequence depicted in SEQ ID NO:27, or a variant or homolog thereof. In some embodiments, the exogenous polynucleotide encodes a polypeptide comprising or consisting of an amino acid sequence having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:27. In some embodiments, the polynucleotide that encodes Tpa comprises or consists of the polynucleotide sequence depicted in SEQ ID NO:28 or SEQ ID NO:56 or a polynucleotide having at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO:28 or SEQ ID NO:56. In some embodiments, the polynucleotide sequence is codon optimized for expression in the microorganism, *e.g.*, SEQ ID NO:56.

[209] In one embodiment, a non-naturally occurring *Methylobacterium* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding ComDE and Tpa.

[210] In one embodiment, a non-naturally occurring *Escherichia* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding ComDE and Tpa.

[211] In one embodiment, a non-naturally occurring *Saccharomyces* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding ComDE and Tpa.

[212] In one embodiment, a non-naturally occurring *Bacillus* microorganism is provided that includes one or more exogenous polynucleotide(s) encoding ComDE and Tpa.

Mutations to enhance accumulation of taurine or an intermediate in taurine biosynthesis

[213] In some embodiments, a non-naturally occurring microorganism is provided that produces taurine, as described above, and that further includes a mutation in a pathway for degradation of taurine and/or in a pathway for degradation of an intermediate of taurine biosynthesis (*e.g.*, a cysteate degradation pathway). In some embodiments, the microorganism includes deletion of one or more endogenous gene sequence that encodes an enzyme that degrades taurine or an intermediate in taurine biosynthesis (*e.g.*, cysteate or sulfoacetaldehyde), thus enhancing accumulation of taurine in the microorganism. Taurine and cysteate degradation pathways are shown schematically in **Fig. 13**. Examples of enzymes in taurine degradation pathways include, but are not limited to, taurine dehydrogenase (TDH), taurine dioxygenase (TDO/TauD), gamma-glutamyltransferase, and taurine-pyruvate aminotransferase (Tpa). A non-limiting example of an enzyme in a cysteate degradation pathway is cysteate sulfo-lyase (CuyA). A non-limiting example of an enzyme in a sulfoacetaldehyde degradation pathway is sulfoacetaldehyde acetyltransferase (Xsc).

[214] In some embodiments, the non-naturally occurring microorganism includes one or more mutation(s) or deletion of a gene sequence that encodes taurine dehydrogenase (TDH) (EC 1.4.99.2), thus reducing or eliminating activity of TDH in the microorganism in comparison to the host cell from which the microorganism was derived.

[215] In some embodiments, the non-naturally occurring microorganism includes one or more mutation(s) or deletion of a gene sequence that encodes taurine dioxygenase (TDO)/TauD (EC 1.14.11.17), thus reducing or eliminating activity of TDO/TauD in the microorganism in comparison the host cell from which the microorganism was derived. In one embodiment, an *Escherichia* microorganism is provided with one or more mutation(s) or deletion of a gene sequence that encodes TDO/TauD, thus reducing or eliminating activity of this enzyme in the microorganism. In one embodiment, a *Saccharomyces* microorganism is provided with one or more mutation(s) or deletion of a gene sequence that encodes TDO/TauD, thus reducing or eliminating activity of this enzyme in the microorganism in comparison to the host cell from which the microorganism was derived. In one embodiment, a *Bacillus* microorganism is provided with one or more mutation(s) or deletion of a gene sequence that encodes TDO/TauD, thus reducing or eliminating activity

of this enzyme in the microorganism in comparison to the host cell from which the microorganism was derived.

[216] In some embodiments, the non-naturally occurring microorganism includes one or more mutation(s) or deletion of a gene sequence that encodes cysteate sulfo-lyase (CuyA) (EC 4.4.1.25), thus reducing or eliminating activity of CuyA in the microorganism in comparison to the host cell from which the microorganism was derived.

[217] In some embodiments, the non-naturally occurring microorganism includes one or more mutation(s) or deletion of a gene sequence that encodes gamma-glutamyltransferase (EC 2.3.2.2), thus reducing or eliminating activity of gamma-glutamyltransferase in the microorganism in comparison the host cell from which the microorganism was derived. In one embodiment, a *Methylobacterium* microorganism is provided with one or more mutation(s) or deletion of a gene sequence that encodes gamma-glutamyltransferase, thus reducing or eliminating activity of this enzyme in the microorganism. In one embodiment, an *Escherichia* microorganism is provided with one or more mutation(s) or deletion of a gene sequence that encodes gamma-glutamyltransferase, thus reducing or eliminating activity of this enzyme in the microorganism in comparison to the host cell from which the microorganism was derived. In one embodiment, a *Bacillus* microorganism is provided with one or more mutation(s) or deletion of a gene sequence that encodes gamma-glutamyltransferase, thus reducing or eliminating activity of this enzyme in the microorganism in comparison to the host cell from which the microorganism was derived.

[218] In some embodiments, the non-naturally occurring microorganism includes one or more mutations(s) or deletion of a gene sequence that encodes a taurine-pyruvate aminotransferase (Tpa) (EC 2.6.1.77), thus reducing or eliminating activity of taurine-pyruvate aminotransferase in the microorganism in comparison the host cell from which the microorganism was derived.

[219] In some embodiments, the non-naturally occurring microorganism includes one or more mutations(s) or deletion of a gene sequence that encodes a sulfoacetaldehyde acetyltransferase (Xsc) (EC 2.3.3.15), thus reducing or eliminating activity of sulfoacetaldehyde acetyltransferase in the microorganism in comparison the host cell from which the microorganism was derived.

Transformation of microorganisms

[220] Numerous transformation protocols and constructs for introducing and expressing exogenous polynucleotides in host cells are known in the art.

[221] In certain embodiments, genetic modifications will take advantage of freely replicating plasmid vectors for cloning. These may include small IncP vectors developed for use in *Methylobacterium*. These vectors may include pCM62, pCM66, or pHC41 for cloning. (Marx, C. J. and M. E. Lidstrom *Microbiology* (2001) 147: 2065-2075; Chou, H.-H. *et al. PLoS Genetics* (2009) 5: e1000652)

[222] In certain embodiments, genetic modifications will take advantage of freely replicating expression plasmids such as pCM80, pCM160, pHC90, or pHC91. (Marx, C. J. and M. E. Lidstrom *Microbiology* (2001) 147: 2065-2075; Chou, H.-H. *et al. PLoS Genetics* (2009) 5: e1000652)

[223] In certain embodiments, genetic modifications will utilize freely replicating expression plasmids that have the ability to respond to levels of inducing molecules such as cumate or anhydrotetracycline. These include pHC115, pLC 290, pLC291. (Chou, H.-H. *et al. PLoS Genetics* (2009) 5: e1000652; Chubiz, L. M. et al. *BMC Research Notes* (2013) 6: 183)

[224] In certain embodiments, genetic modifications will utilize recyclable antibiotic marker systems such as the *cre-lox* system. This may include use of the pCM157, pCM158, pCM184, pCM351 series of plasmids developed for use in *M. extorquens.* (Marx, C. J. and M. E. Lidstrom *BioTechniques* (2002) 33: 1062-1067)

[225] In certain embodiments, genetic modifications will utilize recyclable antibiotic marker systems such as the *cre-lox* system. This may include use of the pCM157, pCM158, pCM184, pCM351 series of plasmids developed for use in *M. extorquens* (Marx, C. J. and M. E. Lidstrom *BioTechniques* (2002) 33: 1062-1067).

[226] In certain embodiments, genetic modifications will utilize transposon mutagenesis. This may include mini-Tn*5* delivery systems such as pCM639 (D'Argenio, D. A. *et al. Journal of Bacteriology* (2001) 183: 1466-1471) demonstrated in *M. extorquens.* (Marx, C. J. *et al. Journal of Bacteriology* (2003) 185: 669-673)

[227] In certain embodiments, genetic modifications will utilize expression systems introduced directly into a chromosomal locus. This may include pCM168, pCM172, and pHC01 plasmids developed for *M. extorquens* AM1. (Marx, C. J. and M. E. Lidstrom

Microbiology (2001) 147: 2065-2075; Lee, M.-C. *et al. Evolution* (2009) 63: 2813-2830)

[228] In certain embodiments, genetic modifications will utilize a *sacB*-based system for unmarked exchange of alleles due to the sucrose sensitivity provided by *sacB* expression. This may include the pCM433 vector originally tested with *M. extorquens.* (Marx, C. J. et al. *BMC Research Notes* (2008) 1: 1)

Microbial cultures

[229] Methods for producing taurine and/or taurine precursors are provided. The methods include culturing a non-naturally occurring microorganism as described herein in a culture medium under conditions suitable for growth of the microorganism and expression of enzymes for taurine biosynthesis as described herein, wherein biomass that includes taurine and/or taurine precursors is produced in the culture. In embodiments in which the microorganism also produces one or more carotenoid compound(s) (*e.g.*, a microorganism that has been genetically modified or artificially pre-selected to produce elevated levels of one or more carotenoid compound(s)), biomass that includes taurine and/or taurine precursors and the one or more carotenoid compound(s) is produced.

[230] The culture medium includes carbon source(s), nitrogen source(s), inorganic substances (*e.g.*, inorganic salts), and any other substances required for the growth of the microorganism (*e.g.*, vitamins, amino acids, etc.).

[231] The carbon source may include sugars, such as glucose, sucrose, lactose, fructose, trehalose, mannose, mannitol, and maltose; organic acids, such as acetic acid, fumaric acid, citric acid, propionic acid, malic acid, pyruvic acid, malonic acid, and ascorbic acid; alcohols, such as ethanol, propanol, butanol, pentanol, hexanol, isobutanol, and glycerol; oil or fat, such as soybean oil, rice bran oil, olive oil, corn oil, sesame oil, linseed oil, and the like. The amount of the carbon source added varies according to the kind of the carbon source, for example, about 1 to about 100 g, or about 2 to about 50 g per liter of medium.

[232] In some embodiments, a C1 carbon substrate is provided to a microorganism that is capable of converting such a substrate to organic products (*e.g.*, microorganisms of the genera *Methylobacterium*, *Methylomonas*, *Methylobacter*. *Methylococcus*, *Methylosinus*, *Methylocyctis*, *Methylomicrobium*). In certain embodiments, the C1 carbon substrate is selected from methane, methanol, formaldehyde, formic acid, methylated amines, methylated thiols, and carbon dioxide. In certain embodiments, the C1 carbon substrate is

selected from methanol, formaldehyde, and methylated amines. In certain embodiments, the C1 carbon substrate is methanol.

[233] The nitrogen source may include potassium nitrate, ammonium nitrate, ammonium chloride, ammonium sulfate, ammonium phosphate, ammonia, urea, and the like, alone or in combination. Amount of the nitrogen source added varies according to the kind of the nitrogen source, for example, about 0.1 to about 30 g, or about 1 to about 10 g per liter of medium.

[234] Inorganic salts may include potassium dihydrogen phosphate, dipotassium hydrogen phosphate, disodium hydrogen phosphate, magnesium sulfate, magnesium chloride, ferric sulfate, ferrous sulfate, ferric chloride, ferrous chloride, manganous sulfate, manganous chloride, zinc sulfate, zinc, chloride, cupric sulfate, calcium chloride, calcium carbonate, sodium carbonate, sodium sulfate, and the like, alone or in combination. Amount of inorganic salt varies according to the kind of the inorganic salt, for example, about 0.001 to about 10 g per liter of medium.

[235] Special required substances, for example, vitamins, nucleic acids, yeast extract, peptone, meat extract, malt extract, corn steep liquor, soybean meal, dried yeast etc., may be included alone or in combination. Amount of the special required substance used varies according to the kind of the substance, for example, about 0.2 g to about 200 g, or about 3 to about 10 g per liter of medium.

[236] In some embodiments, the pH of the culture medium is adjusted to pH about 2 to about 12, or about 6 to about 9. The medium may further include one or more buffer(s) to maintain the culture at the desired pH. Numerous buffers are known in the art and include phosphate, carbonate, acetate, PIPES, HEPES, and Tris buffers. A suitable buffer for a given microorganism can easily be determined by one of ordinary skill in the art. For *Methylobacterium*, a common medium, described by Lee, et al. (2009) *Evolution* 63:2813-2830, is a phosphate buffered medium that consists of 1 mL of trace metal solution (to 1 liter of deionized water the following are added in this order: 12.738 g of EDTA disodium salt dihydrate, 4.4 g of $ZnS0\text{-}7H_20$, 1 .466 g of $CaCI_2\text{-}2H_2O$, 1 .012 g of $MnCI_2\text{-}4H_20$, 0.22 g of $(NH_4)_6Mo_7O_{24}\text{-}4H_20$, 0.314 g of $CuSO_4\text{-}5H_20$, 0.322 g of $CoCl_2\text{-}6H_20$, and 0.998 g of $Fe_3(SO_4)_2\text{-}7H_20$; pH 5.0 is maintained after every addition), 100 mL of phosphate buffer (25.3 g of K_2HPO_4 and 22.5 g of NaH_2PO_4 in 1 liter of deionized water), 100 mL of sulfate solution (5 g of $(NH_4)_2(SO_4)$ and 0.98 g of $Mg(SO_4)_2$ in 1 liter of deionized water), and 799

mL of deionized water. All components are heat sterilized separately and then pooled together. An alternative medium recently developed for use with *Methylobacterium extorquens* takes advantage of an organic buffer and has a citrate-chelated trace metal mix. Culturing is carried out at temperature of 15° to 40°C, and preferably 20° to 35°C, usually for 1 to 20 days, and preferably 1 to 4 days, under aerobic conditions provided by shaking or aeration/agitation. Common practice with *Methylobacterium* is at 30°C. The protocol for making M-PIPES medium is described in Table S1 of Delaney et al. (2013) PLoS One (8:e62957). Figure 2 in USSN 61/863,701 shows an exemplary recipe for medium optimized for use with *M. extorquens*.

[237] In order to generate dense cultures of microorganisms, such as *Methylobacterium*, it may be advantageous to use a fed-batch method. Methanol can be tolerated well at 0.5-1 % v/v (~120-240 mM), and thus this step size of addition can be used repeatedly. Critically, pH levels drop during culturing on methanol, such that the use of a base such as KOH or NaOH would be important to maintain the pH around 6.5. Aeration can be achieved via physical agitation, such as an impeller, via bubbling of filtered air or pure oxygen, or in combination. In order to reduce production costs, the buffer can be replaced from phosphates or PIPES to a carbonate-buffered medium.

[238] Microbial cells may be separated from the culture, for example, by a conventional means such as centrifugation or filtration. The cells may be isolated whole, or may be lysed to release their contents for extraction or further processing. The cells or the medium may be subjected to an extraction with a suitable solvent.

Intracellular taurine as a molecular chaperone and antioxidant.

[239] Microbial cells engineered to produce high levels of taurine or hypotaurine have increased levels of an important osmolyte known to promote protein folding and decrease oxidation (Warskulat et al. (2007) *Methods Enzymol* 428:439-58; Abe et al. (2015) *Amino acids* 47(5):909-15; Fujii et al. (2007) *J Biochem* 141(5):697-707); Oliveira et al. (2010) Pharmacological Reports 62:185-193; Aruoma et al. (1988) *Biochem J* 256:251-55; Bucolo et al. (2016) *Acta Ophthalmologic* 95(256); Patel et al. (2016) *Exp Toxic Pathol* 68(2-3):103-12; Fontana et al. (2004) *Neurochemical Research* 29(1):111-116). When microbial cells are used to express a protein of interest, intracellular taurine or hypotaurine could aid in increasing protein folding or decreasing protein inactivation through oxidation. Thus

the use of microorganisms engineered to accumulate intracellular taurine or hypotaurine could be used to increase the yield and/or specific activity of proteins of interest.

[240] Production of intracellular taurine and/or hypotaurine to aid in protein folding has potential benefits, both for cost and effectiveness. *In vivo* production of taurine or hypotaurine should be less expensive than when taurine or hypotaurine are added externally. Intracellular protein production may also be more effective if it simultaneously allows for higher levels of taurine or hypotaurine. Transport from the cellular medium into cells generally requires a higher concentration of these substrates in the medium and/or requires cellular energy for active transport.

Compositions containing taurine and taurine precursors

[241] Feed compositions are provided for use in aquaculture, or as animal feed, or as human nutritional supplements containing processed or unprocessed biomass from non-naturally occurring microorganism cells as described herein, as are methods of preparation of the feed compositions.

[242] The feed compositions or nutritional supplements include taurine and/or one or more taurine precursor(s), *e.g.*, cysteate, sulfoacetaldehyde, and/or hypotaurine, produced by the non-naturally occurring microorganism. In some embodiments, taurine and/or taurine precursor(s) produced by the microorganism is encapsulated in the microorganism in the feed composition or supplement, *e.g.*, encapsulated in the lipid bilayer of the cell membrane of the microorganism. In some embodiments, taurine and/or taurine precursor(s) produced by the microbial biocatalyst is/are excreted into the culture medium and further purified, for example, using chromatographic or other separation and purification procedures. In some embodiments, taurine and/or taurine precursor(s) is/are chemically extracted from the producing microorganism.

[243] Taurine and/or taurine precursor(s) can be accumulated and encapsulated by the microorganism or can be exported outside the cell. Conditions required for export may be continuous during microbial growth or can be stimulated by limitation of nutrients, *e.g.*, biotin, or by the presence of an inhibitor of microbial growth, such as an antibiotic or surfactant.

[244] In some embodiments, methods for separating and purifying taurine and/or taurine precursors from a culture containing microbial cells and microbially produced taurine may

deploy ion exchange, *e.g.*, ion exchange resins. In some embodiments, microbial cells may be separated by centrifugation, condensed, or filtered, and taurine and/or taurine precursors concentrated to, for example, at least about 80% purity.

[245] In certain embodiments, biomass that is incorporated into a feed or nutritional supplement composition can be in a dry, or substantially dry, form, *e.g.*, containing less than about 20%, 10%, 5%, or 2% of moisture. In certain embodiments, the cultures are isolated by removing substantially all supernatant, such as by filtering, sedimentation, or centrifugation. In certain embodiments, the collection of cultures and further processing of biomass excludes a bacterial lysis step, *e.g.*, by use of detergents or ultrasound. In certain embodiments, the processed microbial cells maintain substantially whole cell membranes. In some embodiments, a substantial portion (*e.g.*, more than about 5%, 10%, 20%, 30%, 50%, or 80%) of bacterial cells may maintain viability in the processed biomass.

[246] The feed composition may contain at least about 1% of the biomass by weight. In certain embodiments, the feed composition is optimized for consumption by fish, seafood, humans, or other animals. For example, the feed may include one or more of EPA, DHA, and one or more essential amino acids.

[247] Methods for preparing a feed composition are also provided. In some embodiments, the method includes: (a) culturing in an appropriate medium at least one non-naturally occurring microorganism as described above; (b) concentrating the medium to provide a biomass; (c) optionally providing additional feed components; and (d) producing the feed composition from the biomass. In certain embodiments, step (b) includes centrifugation. In certain embodiments, step (b) includes allowing the biomass to settle. In certain embodiments, step (b) includes filtration. In certain embodiments, the method further includes a pre-treatment of the biomass after step (a) with a chemical agent (*e.g.*, a surfactant or solvent) to disrupt the cell membranes of the biomass. In certain embodiments, the method further includes mechanical disruption of the cell membranes of the biomass after step (a).

[248] Examples of feedstuffs into which single cell protein enriched with taurine and/or taurine precursors, produced as described herein, may be incorporated include, for example, pet foods, such as cat foods, dog foods and the like, feeds for aquarium fish, cultured fish or crustaceans, etc., feed for farm-raised animals (including livestock and further including fish or crustaceans raised in aquaculture). The state of the biomass can be in whole cell,

lysed or partially processed. The taurine and/or taurine precursors and/or other caloric or nutritional supplements produced in described herein can also be incorporated into food or vitamin supplements for human consumption. Food or feed material into which taurine and/or taurine precursors produced as described herein is incorporated is preferably palatable to the organism that is the intended recipient. This food or feed material may have any physical properties currently known for a food material (*e.g.*, solid, liquid, soft). In some embodiments,, feed produced as described herein will undergo a pelletization process, *e.g.*, through a hot or cold extrusion process at an inclusion rate of less than about 1%, 5%, 10%, 20%, 25%, 30%, 40%, 50%, 60%, or 75%. In other scenarios, the taurine and/or taurine precursors-enriched protein can be consumed directly at 100% or combined with another substance in the form of liquid, baked goods or other to form, including but not limited to, various types of tablets, capsules, drinkable agents, gargles, etc.

[249] Methods of producing fish or seafood are also provided, including farming fish or seafood, and providing a diet, which includes a feed composition as described herein, to the fish or seafood.

[250] The following examples are intended to illustrate, but not limit the invention.

EXAMPLES

Example 1

Methods:

[251] Expression plasmids were constructed utilizing standard molecular cloning techniques and codon optimized, synthetically-derived DNA (see Table 1). These plasmids were transformed into *Methylobacterium extorquens* or *Escherichia coli* BL21 (DE3).

[252] *M. extorquens* strains were grown in a minimal media based on Choi et al(1989) *Appl Microbiol Bioeng* 17:392-6). This media was amended with 0.5% methanol, 10ug/mL trimethylprim, and 50ug/mL Kanamycin. For expression, a saturated *M. extorquens* culture was diluted 100 fold into 25mL of fresh media in a 250mL Erlenmeyer flask and shaken at 200rpm at 30°C. At 24 hours and 36 hours, cultures were fed an additional 0.5% methanol and induced with 0.0125-0.05ng/uL anhydrotetracycline (ATC). *M. extorquens* cultures were harvested between 48 and 52 hours. Following centrifugation, the bacterial pellets were washed once with 1/20X phosphate buffered saline (PBS) and frozen at -20°C.

[253] *E. coli* cultures were grown in LB (10g Tryptone, 10g NaCl, 5g Yeast extract per

liter) amended with 100ug/mL carbenicillin and 125uM isopropyl β-D-1-thiogalactopyranoside (IPTG). Following a 100-500 fold dilution, *E. coli* cultures were shaken at 200rpm at 30°C for 12-24 hours. Following centrifugation, the bacterial pellets were washed once with 1/20X PBS and frozen at -20°C.

[254] To induce chaperones to aid in protein folding, Betaine (Bet) or Benzyl alcohol (BA) were added at 5-10mm to cultures of *E. coli* or *M. extorquens*, as described in Marco et al. (2005) *Cell Stress & Chaperones* 10(4):329-339.

[255] For extraction of intracellular free amino acids, frozen bacterial pellets were resuspended in 1:1 methanol:water and subjected to 3 to 4 freeze thaw cycles using dry-ice/ethanol slurry and a bath sonicator. Following centrifugation, the extraction supernatants were derivatized with the Waters AccQ-Tag Ultra Chemistry kit (176001235) utilizing the provided protocols. Derivatized samples were analyzed on a Waters Acquity H-Class UPLC equipped with a 3100 Mass spectrophotometer. Samples were compared to the included amino acid standard amended with taurine, hypotaurine, and L-cysteate. The presence of taurine or hypotaurine was confirmed by the presence of mass spec ions matching the correct derivatized amino acid mass at the same retention time as in the standard samples. Results are in Table 2.

Results:

Table 1. Expression constructs for taurine and hypotaurine production

Plasmid Name	Vector	Genes present in plasmid	SEQ ID NOs
E2bA	pUC19	CDO_Bacillus, CSAD	50, 53
E2rA	pUC19	CDO_Rat, CSAD	51, 53
E5A	pUC19	MA_3297, CSAD	52, 53
M1A	pLC291	ADO, CSAD	57, 54
M1B	pLC291	ADO, ComDE, TPA	57, 55, 56
M2bA	pLC291	CDO_Bacillus, CSAD	58, 54
M2bB	pLC291	CDO_Bacillus, ComDE, TPA	58, 55, 56
M2rA	pLC291	CDO_Rat, CSAD	59, 54
M2rB	pLC291	CDO_Rat, ComDE, TPA	59, 55, 56
M3A	pLC291	3MDO, CSAD	60, 54
M3B	pLC291	3MDO, ComDE, TPA	60, 55, 56
M4A	pLC291	PAPS-AS, APSK, PAPSSS1, CSAD	61, 62, 63, 54
M4B	pLC291	PAPS-AS, APSK, PAPSSS1, ComDE, TPA	61, 62, 63, 55, 56
M5A	pLC291	MA_3297, CSAD	64, 54
M5B	pLC291	MA_3297, ComDE, TPA	64, 55, 56
M6A	pLC291	CGL/CD, PAP-AS, CSAD	65, 61, 54

M6B	pLC291	CGL/CD, PAP-AS, ComDE, TPA	65, 51, 55, 56
M7A	pLC291	CGL/CD, MA3297, CSAD	65, 64, 54
M7B	pLC291	CGL/CD, MA3297, ComDE, TPA	65, 64, 55, 56
M8A	pLC291	CuyA, CSAD	66, 54
M8B	pLC291	CuyA, ComDE, TPA	66, 55, 56
M9A	pLC291	ComA, ComB, ComC, CSAD	67, 68, 69, 54
M9B	pLC291	ComA, ComB, ComC, ComDE, TPA	67, 68, 69, 55, 56

Table 2. Detection and concentration of intracellular taurine or hypotaurine

Plasmid Name	Organism	Taurine in cells (ppm)	Taurine in Media (ng/mL)	Hypotaurine in cells (ppm)	Hypotaurine in Media (ng/mL)
E2bA	*E. coli* BL21		6		54
E2bA + Bet	*E. coli* BL21			2.7	9
E2bA + BA	*E. coli* BL21		12	1.4	22
E2rA	*E. coli* BL21		19	4.6	264
E2rA + Bet	*E. coli* BL21	1.0	10	64.3	218
E2rA + BA	*E. coli* BL21		8	29.3	209
E5A	*E. coli* BL21	24.8	122		39
E5A + Bet	*E. coli* BL21	9.8	88		10
E5A + BA	*E. coli* BL21	8.7	25		48
M2bA	*M. extorquens*	Peak in MS	5	0.5	56
M2bA + Bet	*M. extorquens*				419
M2bB	*M. extorquens*	Peak in MS		Peak in MS	
M2rA	*M. extorquens*				9
M3A	*M. extorquens*	0.2			
M5A	*M. extorquens*	3.9	55		
M5A + BA	*M. extorquens*	6.7	60		
M5B	*M. extorquens*	0.8			
M5B + BA	*M. extorquens*	1.2			
M7A	*M. extorquens*	12.4	144		
M7A + BA	*M. extorquens*	6.0	75		
M7B	*M. extorquens*	0.6			
M7B + BA	*M. extorquens*	2.2			

Amino Acid and Nucleotide Sequences

SEQ ID NO:1
Description: L-serine dehydratase
Alias: Mext_3740, A9VXE2
Length: 453
Type: Protein
Organism: *Methylobacterium extorquens* PA1
>MISTFDLFKIGIGPSSSHTVGPMIAGRRFRETVLARGGIARISAEIYGSLAWTGRGHGTD
VAILLGLMGHAPSTIDPDRTAPLADELRRTGDLGIPGVHFEPERDLVFNFKDILPLHTNG
MRFRAYDAGDAPIEDQIFYSVGGGFVVTAAEAEAAAAGHAECVPPPLAFGSGRELLDLTL
RTGLTIPQIQLANELTLRPRDEIDAGLDAIRDAMFACIERGLRMDGELPGGLRVRRRAKR
LYESLEATKLANSRPAHEIMDWISLYALAVNEENASGGRVVTAPTNGAAGIVPAVLRYTR
DFCPDWSDERGREFLLTAAAIGGLIKARASISGAEVGCQGEVGSAAAMAAAGLTAVLGGS
AFQIENAAEIAMEHHLGMTCDPIAGLVQVPCIERNAFGANKAVVAASLSLRGDGQHRVSL
DEVIETMRQTGHDMQAKYKETSLGGLAVNVAAC

SEQ ID NO:2
Description: L-serine dehydratase
Alias: Mext_3740, A9VXE2
Length: 1362
Type: DNA
Organism: *Methylobacterium extorquens* PA1
>atgATCAGCACCTTCGATCTGTTCAAGATCGGGATCGGTCCGTCGAGCTCCCACACCGTCGGGCC
GATGATCGCCGGGCGCCGGTTCCGCGAGACCGTACTCGCCCGCGGCGGCATCGCCCGCATCAGC
GCCGAGATCTACGGCTCGCTCGCCTGGACCGGGCGCGGCCACGGCACCGACGTGGCGATCCTGC
TCGGGCTCATGGGCCACGCGCCCTCCACCATCGACCCGGATCGGACGGCGCCGCTCGCCGACGA
ACTGCGCCGCACCGGCGATCTCGGCATTCCCGGCGTCCATTTCGAGCCCGAGCGCGACCTCGTCT
TCAACTTCAAGGACATCCTGCCGCTGCACACCAACGGCATGCGCTTTCGCGCCTACGATGCCGGG
GACGCGCCGATCGAGGACCAGATCTTCTACTCGGTCGGCGGCGGCTTCGTCGTCACCGCCGCCG
AGGCGGAAGCTGCCGCGGCGGGTCATGCGGAGTGCGTGCCACCCCCGCTCGCCTTCGGCAGCGG
GCGTGAACTCCTCGACCTGACGCTACGCACCGGGCTGACGATCCCGCAGATCCAGCTCGCCAAC
GAGCTGACCCTGCGCCCGCGCGACGAGATCGATGCCGGCCTCGACGCGATCCGCGATGCGATGT
TCGCCTGCATCGAGCGCGGCCTGCGCATGGACGGCGAATTGCCCGGCGGCCTGCGGGTGCGGCG
GCGGGCCAAGCGGCTCTACGAGTCGCTGGAGGCGACGAAGCTCGCCAACAGCCGCCCGGCCCAC
GAGATCATGGATTGGATCAGCCTCTACGCGCTCGCCGTCAACGAGGAGAACGCCTCGGGCGGCC
GGGTGGTGACGGCGCCGACCAACGGCGCGGCCGGCATCGTCCCGGCGGTGCTGCGCTACACCCG
CGATTTCTGCCCCGATTGGAGCGACGAGCGCGGGCGCGAGTTCCTGCTCACCGCCGCCGCCATCG
GCGGGCTGATCAAGGCCCGTGCCTCGATCTCGGGGGCCGGAGGTCGGCTGCCAGGGCGAGGTCGG
CTCGGCCGCGGCGATGGCGGCGGCGGGGCTGACCGCCGTGCTCGGCGGCTCGGCCTTCCAGATC
GAGAACGCCGCCGAGATCGCCATGGAGCACCATCTAGGCATGACCTGCGATCCGATCGCCGGCC
TCGTGCAAGTGCCCTGCATCGAGCGCAACGCCTTCGGCGCCAACAAGGCAGTGGTGGCGGCCTC
GCTGTCGCTCCGCGGCGACGGCCAGCACCGGGTGAGCCTGGACGAGGTGATCGAGACCATGCGC
CAGACCGGCCACGACATGCAGGCCAAGTACAAGGAAACCTCGCTCGGGGGGCTAGCCGTCAAC
GTCGCCGCCTGCtga

SEQ ID NO:3
Description: sulfate adenylyltransferase subunit 1
Alias: Mext_2232
Length: 469
Type: DNA
Organism: *Methylobacterium extorquens* PA1
>MTIHQSPEAFGYDAFLRQHQNKEVLRFITCGSVDDGKSTLIGRLLHDTKQIFDDQVTALQRDSRKHG
TQGGEVDLALLVDGLQAEREQGITIDVAYRFFSTDRRSFIVADTPGHEQYTRNMATGASTADLAVIL
VDARHGLTRQSRRHALLVSLLGIRRVALAINKMDLVGWSQDKFEAIVSGFQAFAAPLNFTEVRAIPLS

AKNGDNVVLPGTAATWYTDVPLLRYLEEVPVKSEERAAAFRMPVQWVNRPNSDFRGFSGLIASGSV
APGDAVTVAPSGKTSTIARIFTADGDLERASEGQSVTLVLADEVDASRGAVIATSDAPLTLTDSLDVR
LFWAAESDLVPGANLWAKVGTQTVNAVVKAVHRRIDPETGQAGPADKLAVNDIGDVTLTLDRQIA
VDPYAENRDTGSLILIDRETTDTAALGLVQRVVASSKVAPAPTASVTASAEPARSGGLLAGLKRLFGG

SEQ ID NO:4
Description: sulfate adenylyltransferase subunit 1
Alias: Mext_2232
Length: 1410
Type: DNA
Organism: *Methylobacterium extorquens* PA1
>atgACCATCCATCAGTCTCCGGAAGCGTTCGGCTACGACGCCTTCCTGCGTCAGCACCAG
AACAAGGAAGTCCTGCGCTTCATCACCTGCGGCTCGGTCGATGACGGCAAGTCCACCCTG
ATCGGGCGGCTCCTGCACGACACCAAGCAGATCTTCGACGATCAGGTGACGGCGCTCCAG
CGCGATTCGCGCAAGCACGGCACGCAGGGCGGCGAGGTCGATCTCGCCCTTCTGGTTGAC
GGACTCCAGGCCGAGCGCGAGCAGGGCATCACCATCGATGTCGCCTACCGCTTCTTCTCG
ACCGACCGGCGCTCCTTCATCGTCGCCGACACCCCCGGCCACGAGCAGTACACCCGCAAC
ATGGCGACCGGCGCCTCGACCGCCGACCTCGCCGTGATCCTGGTGGACGCCCGCCACGGG
CTGACCCGCCAGAGCCGGCGCCACGCGCTGCTGGTCTCGCTGCTCGGCATCCGCCGCGTC
GCGCTCGCCATCAACAAGATGGACCTCGTCGGCTGGTCGCAGGACAAGTTCGAGGCGATC
GTCTCCGGCTTCCAGGCCTTTGCCGCGCCGCTGAACTTCACCGAGGTGCGGGCGATCCCG
CTCTCGGCCAAGAACGGCGACAACGTCGTCCTGCCGGGCACCGCCGCGACCTGGTACACG
GACGTTCCGCTGCTGCGCTATCTCGAAGAGGTGCCGGTGAAGTCGGAGGAGCGCGCCGCC
GCCTTCCGCATGCCGGTGCAGTGGGTGAACCGCCCGAATTCCGACTTCCGCGGCTTCTCG
GGGCTGATCGCCTCGGGCTCCGTCGCGCCGGGCGATGCCGTCACCGTCGCGCCTTCCGGC
AAGACCTCGACGATCGCCCGCATCTTCACCGCCGACGGCGATCTGGAACGGGCGAGCGAG
GGCCAGTCGGTGACGCTGGTGCTGGCCGACGAAGTCGATGCCTCGCGCGGCGCGGTGATC
GCGACCTCGGACGCACCGTTGACGCTGACCGACAGCCTCGACGTGCGCCTGTTCTGGGCC
GCCGAATCCGATCTCGTTCCCGGCGCCAACCTGTGGGCGAAGGTCGGCACGCAGACCGTC
AACGCGGTGGTGAAGGCGGTGCACCGCCGGATCGATCCGGAGACGGGACAGGCCGGTCCG
GCCGACAAGCTCGCGGTCAACGACATCGGCGACGTGACGCTGACCCTCGACCGGCAGATC
GCGGTCGATCCCTATGCCGAGAACCGCGACACCGGCAGCCTGATCCTGATCGACCGTGAG
ACGACCGACACGGCCGCGCTCGGCCTCGTGCAGAGGGTCGTTGCGTCGAGCAAGGTCGCT
CCGGCGCCGACCGCGTCTGTGACGGCTTCGGCGGAGCCCGCACGTAGCGGCGGTTTGCTG
GCCGGCCTCAAGCGGCTGTTCGGCGGAtaa

SEQ ID NO:5
Description: sulfate adenylyltransferase subunit 2
Alias: Mext_2233
Length: 309
Type: Protein
Organism: *Methylobacterium extorquens* PA1
>MSAAVAAPARTRLTHLQRLEAESIHIFREAVAEAENPVMLYSIGKDSSVLLHLALKAFAP
GRLPFPLMHIDTTWKFREMIAFRDRRAKELGLELIVHTNQDGLAKGVGPVSHGSEVHTDV
MKTQALRQALDKYKYDVAFGGARRDEEASRAKERIVSLRNGQHRWDPKRQRAEPWHLYNF
KKRRGESFRVFPLSNWTELDIWLYIEQENIPIVPLYFAAERPVVERDGQLIMVDDERFPL
EPGETPQQRQVRFRTLGCYPLTGAVESPAATLPEIIGETLAARTSERQGRVIDKDGAGAM
ERKKQEGYF

SEQ ID NO:6
Description: sulfate adenylyltransferase subunit 2
Alias: Mext_2233
Length: 930
Type: DNA
Organism: *Methylobacterium extorquens* PA1
>atgAGCGCTGCCGTCGCCGCGCCCGCGCGCACCCGCCTGACGCATCTCCAGCGTCTCGAG
GCCGAGAGCATCCACATCTTCCGGGAGGCCGTCGCCGAGGCCGAGAACCCGGTGATGCTC

TACTCGATCGGCAAGGATTCGTCGGTGCTGCTGCACCTGGCGCTGAAGGCCTTCGCGCCG
GGGCGCCTCCCGTTCCCCCTGATGCACATCGACACGACCTGGAAGTTCCGCGAGATGATC
GCCTTCCGCGATCGGCGAGCCAAGGAGCTCGGGCTCGAACTCATCGTGCACACGAATCAG
GACGGGCTTGCCAAGGGCGTCGGCCCGGTCAGCCACGGCTCGGAAGTGCATACCGACGTG
ATGAAGACGCAGGCCCTGCGGCAGGCGCTCGACAAGTACAAGTATGACGTGGCCTTCGGC
GGCGCCCGCCGGGACGAGGAGGCCAGCCGCGCCAAGGAGCGCATCGTGAGCCTGCGCAAC
GGCCAGCACCGCTGGGACCCGAAGCGCCAGCGCGCCGAGCCGTGGCACCTCTACAATTTC
AAGAAGCGGCGCGGCGAGAGTTTTCGCGTGTTCCCGCTATCCAACTGGACCGAATTGGAT
ATCTGGCTCTACATCGAGCAGGAAAATATTCCGATCGTCCCGCTCTACTTCGCCGCCGAG
CGCCCGGTGGTGGAGCGCGACGGCCAGCTCATCATGGTCGATGACGAGCGCTTTCCGCTG
GAGCCGGGCGAGACCCCACAACAGCGGCAGGTCCGGTTCCGCACGCTCGGCTGCTACCCG
CTGACCGGCGCGGTCGAGAGCCCGGCCGCGACCCTGCCGGAGATCATCGGCGAGACGCTG
GCCGCCCGAACCTCGGAGCGCCAGGGCCGGGTCATCGACAAGGACGGCGCCGGCGCCATG
GAGCGCAAGAAGCAGGAGGGCTATTTCtga

SEQ ID NO:7
Description: Adenylyl-Sulfate Kinase
Alias: cysC, NC_000913.3
Length: 204
Type: Protein
Organism: *Escherichia coli K-12*
>MALHDENVVWHSHPVTVQQRELHHGHRGVVLWFTGLSGSGKSTVAGALEEALHKLGVSTY
LLDGDNVRHGLCSDLGFSDADRKENIRRVGEVANLMVEAGLVVLTAFISPHRAERQMVRE
RVGEGRFIEVFVDTPLAICEARDPKGLYKKARAGELRNFTGIDSVYEAPESAEIHLNGEQ
LVTNLVQQLLDLLRQNDIIRS

SEQ ID NO:8
Description: Adenylyl-Sulfate Kinase
Alias: cysC, NC_000913.3
Length: 606
Type: DNA
Organism: *Escherichia coli K-12*
>atgGCGCTGCATGACGAAAACGTCGTCTGGCATAGCCATCCGGTCACTGTGCAACAACGCGAGCT
ACACCACGGTCATCGTGGTGTAGTGCTGTGGTTTACCGGCCTCTCCGGGTCCGGTAAATCAACGG
TCGCCGGGGCGCTGGAGGAGGCGTTACATAAACTCGGCGTCAGTACGTATCTGCTGGATGGCGA
CAATGTTCGCCACGGATTATGCAGCGATCTCGGTTTTAGCGATGCCGATCGTAAAGAGAATATCC
GTCGCGTCGGTGAAGTGGCGAATTTGATGGTTGAAGCCGGACTGGTGGTGCTGACCGCATTTATC
TCGCCACACCGCGCCGAACGCCAGATGGTTCGCGAACGCGTAGGAGAAGGGCGCTTTATCGAAG
TGTTTGTCGATACGCCGCTGGCGATTTGCGAAGCCCGCGATCCCAAAGGCTTATATAAGAAAGC
GCGTGCCGGTGAACTGCGCAACTTTACGGGAATAGATTCCGTTTACGAAGCGCCTGAATCGGCA
GAAATTCATCTCAATGGTGAACAATTAGTAACAAATTTGGTACAGCAATTATTAGATCTGTTGAG
ACAGAACGATATTATCAGATCCtga

SEQ ID NO:9
Description: PAPS-AS
Alias: QO19P5
Length: 322
Type: Protein
Organism: *Ostreococcus tauri*
>MPRGWTKTRAYDSHHFDADAWSVVTPRAGDVIIATAYKSGTTWMQQIVSQLVFEGAAPAALGELS
PWVDLRVPPREVKRGMIEGLPSPRILKTHLPTTGLEYDENAKYIYVARDGRDAFMSLMNHYKNGNE
AFYGALNGPGLKGAPLPTWEEACEGEGDEKLRALFDKWLNTPWGQHPWEEDGWPFWSLFYNMKT
WWDARESKNIIFVHFSDLKKDLKGQMRRIAKFLNAPIDESKFDAQVTACTFESMKGNAASVAPLGG
ALWKGGAETFINKGTNGRWRNVLTKEQVKQYEQVAEKRLGKDCAKWLANGGDMNGRGCVIM

SEQ ID NO:10
Description: PAPS-AS

Alias: OT_ostta05g01260
Length: 696
Type: DNA
Organism: *Ostreococcus tauri*
>atgCCGCGCGGATGGACGAAGACGCGCGCGTACGACTCGCATCACTTTGACGCCGACGCGTGGT
CGGTGGTGACGCCTCGAGCGGGTGACGTCATCATCGCCACCGCGTATAAATCTGGCACGACGTG
GATGCAACAGATCGTGTCGCAACTCGTGTTCGAGGGCGCGGCCCCGGCGGCGTTGGGGGAGCTC
TCGCCGTGGGTGGATCTGCGCGTGCCCCCGCGGGAGGTGAAGCGAGGGATGATCGAGGGATTGC
CCTCGCCCCGGATCTTGAAGACGCATCTTCCGACGACGGGGTTGGAATACGACGAAAACGCGAA
GTACATTTACGTCGCGCGGGACGGCCGCGACGCGTTCATGTCTTTGATGAACCACTATAAGAACG
GTAATGAAGCGTTTTACGGCGCGCTGAACGGCCCTGGGTTAAAGGGCGCACCTTTGCCTACGTG
GGAAGAGGCGTGCGAAGGCGAGGGCGACGAGAAACTTCGCGCGCTTTTTGACAAGTGGCTCAAC
ACGCCGTGGGGCCAGCACCCGTGGGAAGAAGACGGGTGGCCTTTCTGGTCTCTGTTCTATAACAT
GAAGACGTGGTGGGACGCGCGCGAATCCAAGAACATCATCTTCGTGCATTTTTCGGATTTGAAG
AAGGATTTGAAGGGTCAGATGCGACGCATTGCGAAGTTTTTGAACGCCCCGATCGATGAAAGCA
AATTCGATGCGCAAGTCACAGCGTGCACGTTCGAGAGCATGAAGGGTAACGCCGCGAGCGTCGC
GCCTCTCGGTGGCGCGCTGTGGAAGGGCGGTGCGGAGACGTTCATTAACAAAGGTACCAACGGC
CGGTGGAGGAACGTTCTAACCAAGGAACAAGTCAAGCAGTACGAGCAGGTGGCTGAGAAACGG
CTGGGTAAGGACTGCGCAAAGTGGCTCGCCAACGGCGGCGATATGAACGGCCGTGGGTGCGTGA
TCATGtga

SEQ ID NO:11
Description: CSAD
Alias: ref|WP_006454033.1
Length: 488
Type: Protein
Organism: *Synechoccocus* sp. PCC7335
>MFKASKYYNLLQQLENFFSTANSSSLLTKPIDPNVLKSQLSLDLPNEGKPVEELRTEITSYLNNALKT
AHPSYFNQLWGGFNSACFMGDMLASATNTSMYTYEVAPAATLIEQALVTKMSGILGFKSADGQFTT
GGSNGNLMAMAIARHHVLPTVKQDGMTSGPKLVAFVSREAHYSFDKAAHILGLGTEQLWKVPVDS
DGRMKPEALSELVDRARVQGSIPFFVAGTAGTTVRGAFDPFEEISAIAHQENLWFHIDGAWGASVSLS
ATHRQLMAGANQADSLVWDAHKMMGMTLMCSLLLVKQRGQMLRTFSTAGTDYLFHDEVSAGEV
PTESSTSSTELPIEELPTDFGPATMHCGRRVDALKLWLAWRHLGDRGWERLIDSYFELAQRAETIIDK
HPSLELVSSRQSVNLCFRYLPQNKQQADELTLKVRQALWETGTAMVNYAQVEGKTVFRLVICNNQT
RSEDIERFFEALVAIARRLEQEMC

SEQ ID NO:12
Description: CSAD
Alias: ref|WP_006454033.1
Length: 1467
Type: DNA
Organism: *Synechoccocus* sp. PCC7335
>ATGTTCAAAGCCTCCAAATACTACAACTTGTTGCAGCAGCTTGAAAATTTCTTTTCGACAGCTAATTCGTCG
AGTCTGCTTACTAAACCAATAGATCCTAACGTTTTGAAATCTCAACTTTCTTTGGATTTACCAAATGAGGGTA
AACCTGTAGAAGAACTGCGAACGGAGATTACTAGCTATTTGAATAACGCGCTGAAGACAGCTCATCCTAGCTA
TTTTAATCAGCTGTGGGGCGGTTTCAACTCAGCCTGTTTCATGGGTGATATGCTTGCGAGTGCGACAAATACC
TCGATGTATACCTACGAGGTGGCGCCGGCTGCTACTTTAATCGAGCAGGCGCTAGTTACTAAGATGTCTGGCA
TCTTAGGGTTTAAGAGTGCCGATGGGCAGTTTACAACCGGAGGGAGTAACGGAAATTTGATGGCGATGGCGAT
CGCTCGCCATCATGTTCTACCGACTGTTAAGCAGGACGGTATGACCAGCGGCCCCAAACTAGTTGCTTTTGTC
TCTAGAGAGGCGCACTATTCTTTTGATAAAGCTGCTCATATATTGGGATTAGGAACAGAGCAGCTATGGAAAG
TTCCTGTAGACAGCGATGGCAGAATGAAGCCGGAGGCATTATCTGAGCTAGTAGATAGAGCGCGTGTACAAGG
CTCTATTCCTTTCTTTGTTGCCGGAACTGCTGGAACAACTGTAAGAGGTGCCTTCGATCCGTTTGAAGAGATT
AGCGCGATCGCCCACCAGGAAAACCTGTGGTTTCATATCGATGGAGCTTGGGGTGCTAGCGTATCGCTGAGCG
CTACTCATCGACAGCTAATGGCTGGGGCAAACCAAGCAGACTCTCTGGTGTGGGACGCACACAAAATGATGGG
GATGACGCTGATGTGTTCTTTGCTGTTGGTCAAGCAGCGTGGTCAAATGTTAAGGACTTTCTCTACTGCAGGC
ACCGACTATCTATTCCACGATGAAGTCTCTGCTGGGGAAGTGCCTACAGAATCATCAACATCATCAACAGAAT
TGCCCATAGAAGAACTACCAACAGACTTTGGCCCTGCAACTATGCACTGCGGTCGGCGTGTGGATGCACTCAA
GCTTTGGCTAGCCTGGCGGCACCTAGGCGATCGCGGCTGGGAAAGGCTAATCGACAGCTACTTTGAGCTGGCT
CAGCGAGCAGAAACTATCATCGATAAGCATCCTTCGCTGGAGCTAGTGTCTTCGAGACAGTCGGTGAACCTAT
GCTTTCGGTATCTACCTCAGAACAAACAGCAGGCCGATGAGCTGACGCTGAAAGTGCGACAGGCGCTGTGGGA
AACCGGAACTGCGATGGTGAACTACGCTCAAGTAGAAGGCAAAACGGTTTTTCGTTTGGTCATTTGCAACAAT

CAAACCCGCTCTGAGGACATCGAGCGTTTTTTCGAGGCTTTAGTAGCGATCGCCCGGCGGTTAGAGCAGGAGA
TGTGCTGA

SEQ ID NO:13
Description: GadA
Length: 466
Type: Protein
Organism: *Escherichia coli str k12 subs. MG1655*
>MDQKLLTDFRSELLDSRFGAKAISTIAESKRFPLHEMRDDVAFQIINDELYLDGNARQNLATFCQTW
DDENVHKLMDLSINKNWIDKEEYPQSAAIDLRCVNMVADLWHAPAPKNGQAVGTNTIGSSEACML
GGMAMKWRWRKRMEAAGKPTDKPNLVCGPVQICWHKFARYWDVELREIPMRPGQLFMDPKRMIE
ACDENTIGVVPTFGVTYTGNYEFPQPLHDALDKFQADTGIDIDMHIDAASGGFLAPFVAPDIVWDFRL
PRVKSISASGHKFGLAPLGCGWVIWRDEEALPQELVFNVDYLGGQIGTFAINFSRPAGQVIAQYYEFL
RLGREGYTKVQNASYQVAAYLADEIAKLGPYEFICTGRPDEGIPAVCFKLKDGEDPGYTLYDLSERL
RLRGWQVPAFTLGGEATDIVVMRIMCRRGFEMDFAELLLEDYKASLKYLSDHPKLQGIAQQNSFKH
T

SEQ ID NO:14
Description: GadA
Length: 1401
Type: DNA
Organism: *Escherichia coli str k12 subs. MG1655*
>atgGACCAGAAGCTGTTAACGGATTTCCGCTCAGAACTACTCGATTCACGTTTTGGCGCAAAGGC
CATTTCTACTATCGCGGAGTCAAAACGATTTCCGCTGCACGAAATGCGCGATGATGTCGCATTTC
AGATTATCAATGATGAATTATATCTTGATGGCAACGCTCGTCAGAACCTGGCCACTTTCTGCCAG
ACCTGGGACGACGAAAACGTCCATAAATTGATGGATTTGTCGATCAATAAAAACTGGATCGACA
AAGAAGAATATCCGCAATCCGCAGCCATCGACCTGCGTTGCGTAAATATGGTTGCCGATCTGTG
GCATGCGCCTGCGCCGAAAAATGGTCAGGCCGTTGGCACCAACACCATTGGTTCTTCCGAGGCCT
GTATGCTCGGCGGGATGGCGATGAAATGGCGTTGGCGCAAGCGTATGGAAGCTGCAGGCAAACC
AACGGATAAACCAAACCTGGTGTGCGGTCCGGTACAAATCTGCTGGCATAAATTCGCCCGCTAC
TGGGATGTGGAGCTGCGTGAGATCCCTATGCGCCCCGGTCAGTTGTTTATGGACCCGAAACGCAT
GATTGAAGCCTGTGACGAAAACACCATCGGCGTGGTGCCGACTTTCGGCGTGACCTACACCGGT
AACTATGAGTTCCCACAACCGCTGCACGATGCGCTGGATAAATTCCAGGCCGACACCGGTATCG
ACATCGACATGCACATCGACGCTGCCAGCGGTGGCTTCCTGGCACCGTTCGTCGCCCCGGATATC
GTCTGGGACTTCCGCCTGCCGCGTGTGAAATCGATCAGTGCTTCAGGCCATAAATTCGGTCTGGC
TCCGCTGGGCTGCGGCTGGGTTATCTGGCGTGACGAAGAAGCGCTGCCGCAGGAACTGGTGTTC
AACGTTGACTACCTGGGTGGTCAAATTGGTACTTTTGCCATCAACTTCTCCCGCCCGGCGGGTCA
GGTAATTGCACAGTACTATGAATTCCTGCGCCTCGGTCGTGAAGGCTATACCAAAGTACAGAAC
GCCTCTTACCAGGTTGCCGCTTATCTGGCGGATGAAATCGCCAAACTGGGGCCGTATGAGTTCAT
CTGTACGGGTCGCCCGGACGAAGGCATCCCGGCGGTTTGCTTCAAACTGAAAGATGGTGAAGAT
CCGGGATACACCCTGTACGACCTCTCTGAACGTCTGCGTCTGCGCGGCTGGCAGGTTCCGGCCTT
CACTCTCGGCGGTGAAGCCACCGACATCGTGGTGATGCGCATTATGTGTCGTCGCGGCTTCGAAA
TGGACTTTGCTGAACTGTTGCTGGAAGACTACAAAGCCTCCCTGAAATATCTCAGCGATCACCCG
AAACTGCAGGGTATTGCCCAGCAGAACAGCTTTAAACACACCtga

SEQ ID NO:15
Description: CDO
Alias: cdoA, BSU31140, O32085, CDO_Bacillus
Length: 160
Type: Protein
Organism: Bacillus subtilis
>MELYECIQDIFGGLKNPSVKDLATSLKQIPNAAKLSQPYIKEPDQYAYGRNAIYRNNELEIIVINIPPN
KETTVHDHGQSIGCAMVLEGKLLNSIYRSTGEHAELSNSYFVHEGECLISTKGLIHKMSNPTSERMVS
LHVYSPPLEDMTVFEEQKEVLENS

SEQ ID NO:16
Description: CDO
Alias: cdoA, BSU31140, O32085, CDO_Bacillus

Length: 486
Type: DNA
Organism: *Bacillus subtilis*
>atgGAACTGTATGAGTGTATCCAAGACATTTTTGGCGGCTTGAAAAATCCATCGGTTAAAGATTT
AGCAACCGTCTTTAAAACAAATTCCAAACGCAGCAAAATTGAGTCAACCGTATATTAAGGAACCA
GACCAGTACGCTTACGGCCGAAATGCCATCTATCGAAATAATGAATTGGAAATTATCGTGATTA
ACATTCCGCCAAACAAGGAGACAACAGTACACGATCATGGTCAATCCATTGGTTGTGCAATGGT
GTTAGAAGGAAAGCTTCTTAATTCTATTTATCGTTCAACCGGCGAACACGCAGAACTCTCCAATT
CATACTTTGTCCACGAAGGAGAATGCCTTATTTCAACCAAAGGTTTAATTCACAAAATGTCCAAT
CCAACATCTGAACGAATGGTGTCTCTTCATGTCTACTCCCCTCCTTTGGAAGACATGACGGTCTTT
GAGGAACAAAAGGAGGTATTGGAAAATTCAtga

SEQ ID NO:17
Description: MA_3297
Length: 416
Type: Protein
Organism: *Methanosarcina acetivorans* str. C2A
>MGRFILKCLKCGREYSQEYRLTCENDDSFLRAEYLEKKLELRKQ
PGIGRFHSWLPVQEELTTEAGPITYKSEALARELGLSNLYIGFSGYWPEKGAFIKTCS
FKELEAHPTMQLLKESGGKAIVLASAGNTGRAFAHVSALTGTDVYIVVPDSGIPKLWL
PEEPTDSIHLISMTPGNDYTDAINLAGRIAKLPGMVPEGGARNVARREGMGTVMLDAA
VTIGKMPDHYFQAVGSGTGGISAWEASLRLREDGRFGSKLPKLQLTQNLPFVPMYNAWQEGRRDIIP
EIDMKDAKKRIEETYATVLTNRAPPYSVTGGLYDALVDTDGIMYAVSKE
EALDAKALFESLEGIDILPPSAVAAASLLKAVEAGNVGKDDTILLNIAGGGFKRLKED
FTLFQIEPEITVSNPDVPLEELKL

SEQ ID NO:18
Description: MA_3297
Length: 1251
Type: DNA
Organism: *Methanosarcina acetivorans* str. C2A
>atgGGAAGATTCATATTAAAATGTCTGAAATGCGGCAGAGAATACAGCCAGGAATACAGGCTGA
CCTGCGAGAATGACGACTCCTTTTTGCGGGCGGAATACCTTGAAAAAAAACTTGAGCTGAGAAA
GCAGCCAGGGATAGGAAGATTTCACTCATGGCTTCCGGTTCAGGAAGAGCTTACTACCGAAGCC
GGGCCCATCACGTACAAAAGCGAAGCTCTTGCGAGGGAACTTGGGCTTTCGAATCTGTACATAG
GGTTCAGCGGGTACTGGCCCGAGAAAGGAGCTTTTATCAAGACCTGCAGTTTCAAAGAACTCGA
AGCCCATCCTACGATGCAGCTTCTCAAGGAATCCGGGGGAAAAGCCATAGTCCTTGCCTCTGCA
GGGAATACGGGGAGGGCTTTTGCACATGTTTCGGCACTTACCGGAACCGATGTTTATATCGTGGT
TCCCGACTCAGGCATCCCTAAACTCTGGCTGCCTGAAGAACCGACCGATTCCATTCACCTTATCA
GCATGACTCCGGGGAACGATTACACCGATGCTATCAACCTTGCAGGAAGAATTGCAAAGCTTCC
TGGAATGGTCCCTGAAGGAGGAGCCAGAAACGTTGCCAGAAGAGAAGGAATGGGTACTGTAAT
GCTTGATGCAGCCGTAACCATAGGAAAGATGCCTGATCACTACTTCCAGGCTGTCGGAAGCGGG
ACGGGAGGAATCTCAGCCTGGGAAGCTTCTCTGCGCCTCAGAGAGGACGGGCGTTTTGGTTCCA
AACTTCCAAAGCTCCAGCTTACCCAGAATCTCCCCTTCGTTCCCATGTATAATGCATGGCAAGAA
GGCAGGAGGGATATAATTCCCGAAATTGACATGAAAGATGCAAAGAAGCGGATCGAAGAGACC
TACGCCACTGTACTTACCAACCGAGCACCACCTTACTCCGTGACAGGCGGGCTCTATGACGCACT
TGTCGATACGGACGGGATAATGTATGCAGTAAGCAAAGAAGAAGCCCTTGACGCAAAAGCGCTT
TTTGAGTCCCTTGAAGGAATAGATATCCTTCCCCCATCTGCCGTTGCTGCTGCTTCCCTCTTAAAA
GCCGTGGAAGCCGGAAATGTCGGAAAGGACGACACTATCCTCCTGAACATTGCAGGCGGAGGTT
TCAAACGGCTGAAGGAAGACTTCACACTATTCCAGATTGAACCTGAAATTACTGTCTCGAACCCG
GATGTGCCGCTTGAGGAACTGAAGCTCtga

SEQ ID NO:19
Description: ComA
Alias: phosphosulfolactate synthase
Length: 252
Type: Protein
Organism: *Methanosphaera stadtmanae* DSM 3091

>MNAFKFLDEIGPVNTNTMVLDKALGYKTVEDMLTISGNYFNLLK
YGWGTSILYDEEIIKDKNELYHSYNIRTYTGGTLFELANKQNKIDEYFNEIDRLGFNA
VEISDGSTTIDSDRRAQLINKSKELGFYTLSEIGKKNPQKDSEYTTQQRIDLINTDIE
AGSDMVIIEGRESGKNIGIYDDKGNVKKDDLTSIYENTPKEKVLWEAPQKNQQVELIL
TLSNDVNLGNINSNEIVSLETLRRGLRGDTLGKL

SEQ ID NO:20
Description: ComA
Alias: phosphosulfolactate synthase
Length: 759
Type: DNA
Organism: *Methanosphaera stadtmanae* DSM 3091
>atgAACGCTTTTAAGTTTCTAGATGAAATTGGACCAGTAAATACCAATACCATGGTTCTTGATAA
GGCATTAGGATACAAAACAGTTGAAGATATGTTAACAATTAGTGGAAACTATTTTAATCTATTGA
AGTATGGATGGGGAACTTCAATATTATATGATGAAGAAATAATAAAAGATAAAAATGAATTATA
TCACTCATATAATATTAGAACATATACTGGTGGAACTTTATTTGAATTAGCAAATAAACAAAATA
AAATAGATGAATATTTTAATGAAATTGACAGATTAGGATTTAATGCTGTGGAAATATCTGATGGA
TCAACTACCATTGACAGTGATAGACGTGCACAGTTAATTAATAAATCAAAAGAATTAGGTTTCTA
CACTTTGAGTGAAATAGGTAAGAAAAATCCACAAAAAGATTCTGAATATACAACACAACAACGT
ATAGATCTTATAAATACAGATATTGAAGCAGGTTCTGATATGGTTATTATTGAAGGACGTGAAAG
TGGTAAAAATATTGGTATATACGATGATAAAGGTAATGTAAAAAAAGATGATTTAACTTCAATC
TATGAAAATACACCTAAAGAAAAAGTATTGTGGGAAGCTCCACAGAAAAATCAACAAGTAGAA
TTAATACTTACATTAAGTAATGATGTAAATCTTGGAAACATTAATTCTAATGAAATAGTCTCCCT
TGAAACATTACGTCGTGGATTAAGAGGAGACACTCTTGGAAAATTAtaa

SEQ ID NO:21
Description: ComB1
Length: 342
Type: Protein
Organism: *Methanosphaera stadtmanae* DSM 3091
>MKINVSLYNSRTNDLAIVIDLLRASTTISVALNTFKRIVPINDI
DEAIKLKEKHNAILAGEIKSSDFDVSNSPVQISNYAGDTLILKTTNGTKVLENIKQRN
SEVNILVGASINAKTVAQKALDIADNEIELVMAGRHQRFTIEDCIGAGIIINEIVNIA
KEKNIYLELSESAKASKIISNNSNIIKQLINTSHSADKLRYLGFGEDIEICSLINKID
TVPIYKNNYIVSLD

SEQ ID NO:22
Description: ComB1
Length: 342
Type: DNA
Organism: *Methanosphaera stadtmanae* DSM 3091
>atgAAAATTAATGTAAGTTTATATAATTCACGAACCAATGATTTAGCTATAGTAATTGATTTATTA
AGGGCAAGTACAACAATAAGTGTAGCATTAAATACTTTTAAAAGAATTGTTCCGATTAATGATAT
AGATGAAGCTATTAAATTAAAAGAAAAACATAATGCAATATTGGCAGGTGAAATTAAATCATCA
GATTTTGATGTTTCAAATTCACCAGTTCAAATATCAAATTATGCTGGTGATACATTAATTTTGAAA
ACAACAAATGGTACAAAGGTATTAGAAAATATAAAACAAAGAAATTCAGAAGTAAATATATTG
GTTGGAGCATCAATAAATGCAAAAACAGTAGCACAAAAGGCATTAGATATTGCAGATAATGAAA
TTGAATTAGTTATGGCAGGAAGACATCAAAGATTTACAATAGAGGATTGTATTGGTGCAGGAAT
AATTATTAATGAAATAGTAAACATAGCTAAAGAAAAAAATATATACTTAGAACTTTCAGAATCA
GCAAAAGCATCAAAAATAATATCAAATAATTCTAATATAATAAAACAATTAATAAATACTTCAC
ACAGTGCAGATAAATTACGTTATCTTGGATTTGGTGAAGATATTGAAATATGTAGTTTAATTAAC
AAGATAGATACAGTTCCAATCTATAAGAATAATTACATAGTCTCATTAGATtaa

SEQ ID NO:23
Description: ComC
Alias: Sulfolactate dehydrogenase
Length: 342
Type: Protein

Organism: *Methanobacterium* sp. MB1
>MNITPEQELSLIIDILTKFDVPEDQASIIAEVTLDGDLKGFSSHGIGRFPQYIKGLECGHIKPHTEIVVEK
ETAATALINGNHGFGHVVTYQAMKMAIEKAKEVGIGLVGIHNSNHFGVAGYYSDMALMEDIIGIVT
ANTEPAVAPIGGKEPILGTNPLAIGIPSGSHYLSVDMATSASARGKLMEAKRLGEPIPENVALDSDGNP
TTDPAEALKGSILPFGAHKGYALSLMIEVIAGPLVRASYGKGVTGTADPEVPCTKGDLIAAIDPSKFV
DIDQFKEEVDDLISELKSTPNVMIPGDFEVLNVKRHQKEGIALDETLVQQLREIASNVDVDVSDILGD

SEQ ID NO:24
Description: ComC
Alias: Sulfolactate dehydrogenase
Length: 342
Type: DNA
Organism: *Methanobacterium* sp. MB1
>atgAACATTACTCCAGAACAGGAATTATCCCTGATCATCGATATTTTAACTAAATTTGACGTACCT
GAAGACCAAGCATCCATCATTGCCGAAGTGACACTAGACGGTGATCTTAAGGGTTTCTCATCTCA
TGGAATTGGTAGATTCCCCCAGTACATTAAGGGATTGGAATGTGGTCATATCAAGCCCCACACA
GAAATAGTTGTGGAGAAAGAAACTGCAGCCACCGCTCTGATAAATGGTAACCATGGTTTTGGAC
ATGTAGTAACCTACCAGGCCATGAAAATGGCCATAGAGAAAGCTAAAGAAGTAGGTATTGGTTT
AGTGGGTATCCATAACTCCAACCACTTTGGAGTGGCTGGTTATTACTCCGACATGGCATTGATGG
AAGATATCATTGGCATTGTAACTGCCAACACTGAACCAGCCGTGGCCCCTATTGGAGGGAAAGA
ACCAATACTGGGTACTAATCCCCTGGCCATAGGAATACCTTCCGGTAGCCACTATCTCTCCGTGG
ACATGGCCACATCAGCTTCCGCCCGTGGAAAACTCATGGAAGCCAAACGTCTTGGTGAACCCAT
ACCAGAAAATGTGGCCCTGGATTCCGATGGAAATCCCACCACCGACCCAGCAGAAGCACTCAAA
GGATCAATCCTCCCCTTCGGAGCCCATAAAGGATATGCCTTATCCCTTATGATTGAAGTTATAGC
CGGCCCACTGGTACGTGCCTCCTATGGTAAGGGAGTTACTGGAACAGCTGACCCCGAGGTTCCCT
GCACCAAAGGAGACCTGATTGCCGCCATTGACCCCTCCAAATTTGTGGATATAGACCAGTTTAAG
GAAGAGGTGGATGATCTTATAAGTGAATTAAAATCCACTCCTAATGTAATGATACCCGGAGATTT
TGAAGTCTTAAATGTGAAACGTCACCAGAAAGAAGGAATAGCTCTGGATGAAACCCTTGTACAG
CAGTTAAGGGAAATCGCCAGCAATGTAGATGTGGATGTATCAGATATACTGGGAGATtaa

SEQ ID NO:25
Description: ComDE
Alias: sulfopyruvate decarboxylase
Length: 397
Type: Protein
Organism: *Methanosarcina acetivorans* str. C2A
>MYVVNPEEKVIEIMKQTGIDLAATLPCDRIKNLLPLVSENFPEI
KLTREENGVGICAGIYLAGGKPMMLIQSTGLGNMINALESLNVTCKIPLPILASWRGV
YKEGIEAQVPLGAHLPSILEGAGLTYTIIGETEKLPLLENVILDAFENSRPHIALVSP
KVWEASECCAWQAAGMPIKPEIMERTCRFSLTSGTLKPFMLRNDAICTLASELDDEIT
VTNLGVPCKELYACRDRELNFYMFGSMGLVSSIGLGLALRSEKTVITFDGDGSLLMNP
NALLEIAKEAPKNLIIIALDNGAYGSTGSQETCALRYIDLEIFANACGIQNTAKVNSK
EGVIEAFRKFKAMRELSFIHVILKPGNTNAPNIPMSPEEATKRFKETLDVKKF

SEQ ID NO:26
Description: ComDE
Alias: sulfopyruvate decarboxylase
Length: 1164
Type: DNA
Organism: *Methanosarcina acetivorans* str. C2A
>atgTACGTGGTAAACCCGGAAGAAAAAGTAATAGAAATCATGAAACAAACAGGTATTGATCTTG
CTGCAACGCTTCCCTGCGACAGGATCAAGAACCTGCTTCCCCTGGTCTCGGAAAATTTTCCAGAA
ATCAAATTGACAAGGGAAGAAAACGGAGTGGGGATCTGTGCAGGCATCTACCTTGCAGGCGGA
AAGCCAATGATGCTTATCCAGAGTACGGGGCTCGGGAATATGATCAATGCCCTTGAATCCCTGA
ACGTAACCTGTAAAATCCCCCTTCCGATCCTGGCTAGCTGGCGCGGTGTATATAAAGAAGGCATC
GAAGCTCAGGTTCCCCTGGGAGCCCACCTCCCTTCCATCCTTGAAGGGGCCGGACTTACATACAC
AATAATTGGCGAAACTGAAAAGCTTCCTCTTCTTGAAAATGTAATTCTTGACGCCTTTGAAAACT
CGAGACCCCATATTGCCCTGGTCTCCCCTAAAGTTTGGGAAGCTTCGGAATGCTGTGCTTGGCAG

GCTGCAGGGATGCCGATAAAGCCGGAAATTATGGAAAGGACCTGCAGGTTTTCCCTCACAAGCG
GGACTCTCAAGCCTTTTATGCTCAGAAACGATGCAATCTGCACCTTAGCCTCCGAGCTTGATGAC
GAAATTACCGTGACAAACCTCGGAGTCCCCTGCAAGGAGCTTTACGCCTGCAGGGACAGGGAAC
TCAACTTCTATATGTTCGGCTCCATGGGGCTTGTTTCTTCAATAGGGCTTGGTCTTGCCCTGCGCT
CGGAAAAGACAGTTATCACTTTTGACGGGGACGGGAGCCTTTTAATGAACCCAAATGCCCTCCTT
GAAATTGCAAAAGAAGCCCCGAAAAACCTCATAATCATTGCCCTTGACAACGGCGCCTATGGTT
CTACAGGTTCTCAGGAGACCTGCGCCCTCCGCTACATTGACCTTGAAATCTTTGCAAACGCCTGC
GGGATTCAGAACACCGCCAAAGTGAACAGCAAAGAAGGGGTGATAGAAGCTTTCAGGAAATTC
AAAGCCATGAGAGAGCTCTCCTTTATCCATGTGATCCTGAAACCCGGGAACACAAATGCTCCCA
ATATTCCTATGAGCCCTGAAGAAGCAACAAAACGCTTCAAAGAAACACTGGATGTAAAAAAGTT
Ttaa

SEQ ID NO:27
Description: Taurine-pyruvate aminotransferase (Tpa)
Length: 463
Type: Protein
Organism: *Rhodococcus opacus*
>MVVDVTELRARARRHLGPHFTRKDTWESDFPVFVRGEGSYLIDTEGDRFLDGLAGLFCVNIGHGRD
DIAKAASEQIGTLAYASNWGSAHIPAIEASALIADLAPGDLGTTFFVNSGSEAVETAVKFARQYHRSQ
GNPQRTKIISREMAYHGTTLGALSVTQLPKIKDPFGPLLPGVRSVPNTLGYLGDCGPANELDCIAAIEA
VIEEEGAETIAAVFAEPVQNGRGALVPPDGYWAALRALCDKHGILLVSDEVICSFGRLGHWFGHGLT
GVVPDMITFAKGSTSGYAPLGGLIVREQLVRELYDSPKGGVFTHGATWGGHPVSTAVAVANITAMR
DENVLGNVSARGPKLRSALDSLMSSHRCVKDVRGTGFFYAIELMADSDSGREFTEQESLTVLRKVLP
EAFARTKVILRGDDRGATMLMISPPLVADDEVLSELLHGIDSMLTDIEKAIQP

SEQ ID NO:28
Description: Taurine-pyruvate aminotransferase (Tpa)
Length: 1393
Type: DNA
Organism: *Rhodococcus opacus*
>atgGTCGTGGACGTCACCGAATTGCGAGCACGGGCCCGCCGGCACCTCGGACCTCATTTCACCCG
TAAGGACACCTGGGAAAGCGACTTTCCGGTGTTCGTTCGTGGCGAGGGAAGCTATCTGATCGAC
ACCGAGGGGGACCGTTTCCTCGACGGTCTGGCAGGCCTGTTCTGTGTGAACATCGGTCACGGCCG
CGACGACATCGCCAAGGCGGCGAGCGAGCAGATCGGGACGCTGGCGTACGCCTCCAACTGGGG
CAGCGCCCACATTCCCGCGATCGAGGCGTCCGCGCTCATCGCGGACCTGGCGCCCGGTGATCTCG
GGACGACCTTCTTCGTCAACTCGGGTTCCGAGGCCGTGGAGACGGCCGTCAAGTTCGCCCGGCA
GTACCACCGCAGCCAGGGCAACCCGCAGCGCACCAAGATCATCAGCCGCGAGATGGCGTATCAC
GGAACCACTCTCGGCGCCCTCTCGGTGACACAGCTGCCCAAGATCAAAGACCCGTTCGGACCGC
TGCTGCCCGGGGTCCGCTCCGTACCCAACACCCTCGGTTACCTCGGCGACTGCGGCCCGGCGAAC
GAGCTCGACTGCATCGCCGCGATCGAAGCCGTCATCGAGGAAGAGGGCGCCGAGACCATCGCCG
CCGTGTTCGCCGAGCCGGTTCAGAACGGGCGCGGCGCCCTCGTCCCGCCGGACGGATACTGGGC
CGCGCTGCGCGCGCTGTGCGACAAGCACGGGATCCTGCTGGTCTCCGACGAGGTGATCTGCTCGT
TCGGCCGCCTCGGACACTGGTTCGGGCACGGGCTGACCGGTGTGGTTCCCGACATGATCACGTTC
GCGAAGGGCTCCACGTCCGGATACGCGCCGCTCGGCGGCCTGATCGTGCGTGAGCAGCTGGTTC
GCGAGCTCTACGACTCGCCCAAGGGCGGCGTGTTCACGCACGGCGCGACGTGGGGCGGACACCC
GGTGTCGACTGCGGTGGCGGTCGCGAACATCACCGCGATGCGCGACGAGAAGCGTGCTGGGCAAC
GTCTCCGCGCGCGGCCCGAAGTTGCGGTCGGCACTCGACTCGCTGATGAGCTCGCACCGCTGCGT
CAAGGACGTGCGCGGCACCGGCTTCTTCTACGCGATCGAGTTGATGGCCGACAGCGACAGCGGC
CGCGAGTTCACCGAGCAGGAGTCGCTGACGGTGTTGCGCAAGGTGCTGCCGGAGGCGTTCGCCC
GCACCAAGGTGATCCTCCGCGGCGACGACCGCGGTGCCACGATGCTGATGATTTCGCCGCCACT
CGTCGCCGACGACGAGGTGCTCTCGGAACTGCTCCACGGAATCGACAGCATGCTCACCGACATC
GAAAAGGCAATCCAGCCGtag

SEQ ID NO:29
Description: gamma-glutamyltransferase / glutathione hydrolase / gamma-glutamyl-transpeptidase
Alias: MEXT_1030
Length: 1393
Type: Protein

Organism: *Methylobacterium extorquens* PA1
>MSSRPHRRSSFSATFAKRQRRHPEPFSACGKSARLRRILSAHPGPSAILREPVARSRNAG
GARWRGARQLPFAPTRGPDPASRPVRSQVSSESVMPDTPVFAHAAVAAPHALAASAGQNVLAQGG
NAIEAMVAMAAAIAVVYPHMNGIGGDGFWLIRERNGRVRGIEACGPAGQLATRARYREKELDAIPS
RGPDAAVTVAGTVGGWRLALDMARAFGGRLPLDTILADAIRHARAGCPV
SASEARYVPKELDTLHDAPNFAATYLDDGKPYAAGAIRAQPKLADTLAQLAHAGLDDFYR
GDIGREIASDLERLGAPVTRADLTAYAAKERAPLTLRRRDATLYNFPPPTQGLAALIILG
IFDRLNIAEPESTAHYHGLIEATKRAFAIRDRFVTDFDRLKGDPAAFLDPRRLDREAALI
DMRRAASIPVRSGEGDTVWMGAIDNDGMAVSFIQSVYWEYGSGTVLPGTGICWQNRGMSFSLDAN
AVNPLEPGRRPFHTLIPALAAFDDGRVMSYGSMGGDGQPQFQAQIFTRYADYGMSVADAVDAPRLL
YGRTWGAESLSVKVEDRFDPACIAALRRLGHDIEELGGAYIDSLGHAGMLVRHVKDGRIEATHDPRS
DGGAAGL

SEQ ID NO:30
Description: gamma-glutamyltransferase / glutathione hydrolase / gamma-glutamyl-transpeptidase
Alias: MEXT_1030
Length: 1878
Type: DNA
Organism: *Methylobacterium extorquens* PA1
>atgTCATCCCGCCCGCACCGGCGCAGTTCCTTTTCTGCAACATTTGCAAAAAGGCAGAGG
CGCCACCCGGAACCATTTTCGGCTTGTGGGAAATCCGCACGTCTCCGACGCATCCTGAGC
GCGCATCCAGGGCCATCTGCGATCCTGCGGGAGCCGGTCGCGCGATCGCGGAATGCCGGG
GGTGCGCGGTGGCGGGGAGCACGGCAGTTGCCCTTCGCGCCGACGCGTGGTCCTGATCCT
GCCTCTCGGCCCGTCCGATCTCAGGTTTCGTCAGAGTCCGTCATGCCCGACACGCCCGTC
TTCGCCCATGCGGCCGTTGCCGCCCCCCACGCGCTGGCGGCTTCGGCCGGTCAGAACGTA
CTGGCGCAGGGCGGCAACGCCATCGAGGCGATGGTCGCGATGGCCGCCGCCATCGCGGTG
GTCTACCCGCACATGAACGGCATCGGCGGCGACGGCTTCTGGCTGATCCGCGAGCGGAAC
GGCCGCGTGCGCGGCATCGAGGCCTGCGGACCGGCCGGGCAGCTCGCGACCCGCGCCCGC
TACCGGGAGAAGGAGCTCGACGCGATCCCCTCCCGCGGCCCCGACGCGGCAGTGACGGTG
GCGGGCACCGTCGGCGGCTGGCGCCTCGCGCTCGACATGGCGCGCGCCTTCGGCGGCCGG
CTCCCCCTCGATACGATTCTGGCCGACGCCATCCGCCACGCTCGCGCAGGCTGCCCGGTC
TCGGCCTCGGAAGCGCGCTACGTGCCAAAGGAACTCGACACGCTGCACGACGCGCCGAAT
TTCGCTGCGACCTATCTCGATGACGGCAAGCCCTACGCGGCGGGCGCGATCCGGGCGCAG
CCCAAGCTCGCCGACACCCTGGCCCAGCTCGCCCATGCCGGGCTCGACGACTTCTACCGC
GGCGATATCGGCCGCGAGATCGCCAGCGATCTGGAACGTCTCGGCGCCCCCGTTACCCGC
GCCGACCTCACCGCCTACGCGGCCAAGGAGCGGGCACCGCTGACCCTGCGGCGGCGCGAC
GCCACGCTCTACAACTTCCCGCCGCCGACCCAGGGCCTCGCGGCGCTGATCATCCTCGGG
ATCTTCGACCGGCTGAACATCGCCGAGCCGGAGAGCACCGCGCATTATCACGGGCTGATC
GAGGCGACGAAGCGCGCCTTCGCCATCCGCGACCGCTTCGTCACCGATTTCGACCGCCTG
AAGGGCGACCCCGCCGCCTTCCTCGATCCGAGGCGCCTCGACCGCGAGGCGGCCCTGATC
GACATGCGGCGTGCCGCGAGCATCCCGGTCCGCTCGGGCGAGGGCGACACCGTCTGGATG
GGCGCGATCGACAACGACGGCATGGCCGTCTCCTTCATCCAGTCGGTCTACTGGGAGTAC
GGCTCCGGCACGGTGCTGCCGGGAACCGGCATCTGCTGGCAGAACCGCGGCATGTCGTTC
TCGCTCGACGCGAACGCGGTGAACCCGCTGGAACCGGGCCGGCGCCCGTTCCACACCCTG
ATCCCGGCGCTGGCCGCCTTCGATGACGGCCGGGTCATGTCCTACGGCTCCATGGGCGGT
GACGGGCAGCCGCAGTTCCAGGCGCAGATCTTCACCCGCTACGCCGATTACGGGATGTCG
GTGGCCGATGCGGTGGACGCGCCGCGCCTGCTCTACGGCCGCACCTGGGGCGCCGAGTCG
CTGAGTGTGAAGGTCGAGGACCGCTTCGATCCGGCCTGCATCGCGGCGCTCCGGCGCCTG
GGCCACGACATCGAGGAGCTGGGCGGCGCCTATATCGACTCGCTGGGCCATGCCGGCATG
CTGGTGCGCCATGTCAAAGACGGGCGGATCGAAGCGACGCACGATCCGCGCTCCGATGGC
GGCGCGGCGGGGCTTtga

SEQ ID NO:31
Description: PAPSS1 - Bifunctional 3'-phosphoadenosine 5'-phosphosulfate synthase 1
Alias: E1C8P2
Length: 624
Type: Protein

Organism: *Gallus gallus*
>MELPESQCKKAKLSNRVPNWGMQRATNVTYQAHHVSRNKRGQVVGTRSGFRGCTVWLTGL
SGAGKTTVSMALEEYLVCHGIPCYTLDGDNIRQGLNKNLGFTPEDREENVRRIAEVAKLF
ADAGLVCITSFISPYAQDRNNARRIHEGASLPFFEVFVDAPLHVCEQRDVKGLYKKARAG
EIKGFTGIDSEYEKPEAPELVLKTDSCDVNDCVQQVVELLQERDIVPVDASYEVKELYVP
ENKLKLAKTDAESLLTLEINKVDMQWVQVLAEGWATPLSGFMREREYLQCLHFDCLLDGG
VINLSVPIVLTATQEDKERLDGCTAIALVYEGRRVAILRNPEFYEHRKEERCARQWGTTC
KDHPYIKMVMEQGNWLVGGDLQVLDRIYWNDGLDQYRLTPAELRQKFKEMNADAVFAFQL
RNPVHNGHALLMQDTHKQLLERGYRRPVLLLHPLGGWTKEDDVPLMWRMKQHAAVLEEGV
LNPETTVVAIFPSPMMYAGPTEVQWHCRSRMVAGANFYIVGRDPAGMPHPGTGKDLYEPT
HGAKVLTMAPGLRALEIVPFRVAAYNKKKKSMDYYDSEHHEDFEFISGTHMRKLAREGQN
PPEGFMAPKAWTVLTEYYKSLEKA

SEQ ID NO:32
Description: PAPSS1 - Bifunctional 3'-phosphoadenosine 5'-phosphosulfate synthase 1
Alias: E1C8P2
Length: 1875
Type: DNA
Organism: *Gallus gallus*
>ATGGAGCTGCCTGAGAGCCAGTGCAAGAAAGCGAAGCTGAGCAACAGGGTGCCGAACTGGGG
AATGCAGAGGGCAACCAATGTTACCTACCAAGCTCATCATGTCAGCCGAAATAAGAGAGGCCAA
GTGGTAGGAACAAGAAGTGGTTTCCGTGGATGCACAGTCTGGTTAACAGGTCTATCTGGTGCTG
GGAAGACCACAGTTAGCATGGCCCTGGAGGAGTATTTAGTATGCCATGGCATTCCATGCTACAC
GTTGGATGGTGACAATATTCGCCAAGGCCTTAATAAGAATCTGGGTTTCACTCCAGAAGATAGA
GAAGAAAACGTCCGTCGGATTGCTGAGGTTGCTAAACTGTTTGCAGATGCTGGTTTGGTGTGCAT
CACTAGTTTCATCTCTCCTTATGCTCAGGATCGTAATAATGCTAGACGAATTCATGAAGGGGCCA
GCTTGCCTTTTTTTGAAGTATTTGTGGATGCTCCTTTGCATGTCTGTGAACAAAGAGATGTTAAGG
GACTGTATAAGAAAGCCAGAGCTGGAGAAATTAAAGGCTTTACTGGGATTGACTCTGAGTATGA
AAAACCAGAAGCCCCAGAGCTTGTGCTGAAAACTGATTCCTGTGATGTGAACGATTGTGTACAA
CAAGTTGTGGAACTTCTTCAAGAGAGGGACATCGTACCAGTAGATGCCTCGTATGAGGTGAAAG
AGCTTTATGTGCCAGAAAACAAACTGAAGTTGGCTAAAACTGATGCTGAGTCTCTGTTAACCTTG
GAAATAAATAAGGTGGATATGCAGTGGGTGCAAGTGTTGGCAGAAGGCTGGGCAACACCTCTGA
GTGGCTTTATGAGAGAGAGAGAATACCTGCAGTGCCTTCACTTTGACTGTCTCCTTGATGGGGGA
GTCATTAATCTTTCAGTGCCTATAGTGCTAACAGCTACACAGGAAGACAAGGAAAGACTGGATG
GTTGTACAGCAATTGCATTAGTGTACGAGGGTCGCCGTGTGGCCATTCTCCGTAATCCAGAATTC
TATGAGCATAGGAAAGAGGAACGCTGTGCGAGGCAGTGGGGAACAACATGCAAGGATCATCCT
TACATAAAGATGGTTATGGAGCAAGGGAACTGGCTTGTAGGTGGAGATTTACAGGTCCTTGATC
GTATTTATTGGAATGATGGACTTGATCAGTACCGTCTCACTCCAGCTGAACTAAGACAGAAGTTC
AAGGAAATGAATGCTGATGCTGTCTTTGCATTCCAGTTACGCAACCCAGTGCACAATGGGCACG
CACTTTTAATGCAGGATACTCATAAGCAGCTTTTGGAACGTGGCTACAGGCGTCCAGTTTTGCTC
TTGCATCCACTTGGAGGCTGGACAAAGGAGGACGACGTTCCTCTCATGTGGCGCATGAAACAGC
ATGCTGCAGTACTGGAGGAGGGAGTCTTGAATCCAGAAACAACGGTAGTGGCTATATTCCCCTC
CCCCATGATGTATGCTGGACCAACGGAGGTTCAGTGGCACTGCAGATCACGGATGGTTGCAGGT
GCTAACTTCTACATTGTGGGGCGAGATCCTGCAGGGATGCCGCACCCTGGTACTGGGAAAGATC
TGTATGAACCAACCCATGGTGCCAAAGTGTTGACAATGGCCCCAGGCCTCCGAGCACTGGAAAT
TGTACCTTTCAGGGTTGCGGCTTATAACAAGAAAAAGAAGTCCATGGACTACTATGACTCTGAGC
ACCATGAAGACTTTGAATTTATATCGGGGACCCACATGCGCAAGCTGGCTCGAGAAGGACAAAA
CCCACCGGAAGGCTTCATGGCTCCTAAGGCTTGGACTGTGCTGACAGAATACTACAAATCCTTGG
AGAAGGCTTAG

SEQ ID NO:33
Description: p3MDO
Alias: Q9I0N5, PA2602
Length: 201
Type: Protein
Organism: *Pseudomonas aeruginosa* PAO1
>MSSILRLDRLRQFIGELATLLDSRPDESTLLAQAHPLLAELVHQDDWLPEDCARPDPQRY
QQYLLHVDSRQRFSVVSFVWGPGQITPVHDHRVWGLIGMLRGAEYSQPYAFDAGGRPHPS

GARRRLEPGEVEALSPRIGDVHQVSNAFSDRTSISIHVYGANIGAVRRAVFSAEGEEKPF
ISGYSNSRLPNIWDLSKENPA

SEQ ID NO:34
Description: p3MDO
Alias: Q9I0N5, PA2602
Length: 606
Type: DNA
Organism: *Pseudomonas aeruginosa* PAO1
>atgTCATCCATCCTGCGCCTCGACCGCCTGCGCCAGTTCATCGGCGAGCTGGCGACACTGCTCGA
CAGCCGTCCCGACGAATCCACCCTGCTCGCCCAAGCCCACCCCCTGCTGGCCGAGCTGGTGCACC
AGGACGACTGGCTGCCGGAAGACTGCGCCCGCCCCGATCCACAGCGCTACCAACAGTACCTGCT
GCATGTCGACTCACGGCAGCGCTTCTCGGTGGTCAGCTTCGTCTGGGGGCCGGGCCAGATCACAC
CGGTACACGATCATCGGGTCTGGGGCCTGATCGGCATGCTCCGCGGGGCCGAATACTCGCAGCC
GTACGCCTTCGATGCGGGGGGGCGTCCGCATCCCAGCGGAGCCCGTCGACGCCTGGAGCCCGGC
GAGGTCGAAGCGCTGTCGCCACGCATTGGCGACGTGCACCAGGTGAGCAACGCCTTCAGCGACC
GCACATCCATCAGTATCCACGTCTACGGCGCCAATATCGGTGCGGTACGGCGTGCCGTGTTCAGC
GCCGAAGGTGAGGAAAAACCCTTCATTTCCGGCTATTCCAACAGCCGCTTGCCCAATATCTGGGA
CCTGTCGAAAGAGAACCCCGCAtga

SEQ ID NO:35
Description: Mammalian CDO
Alias: P21816, CDO_Rat
Length: 200
Type: Protein
Organism: *Rattus norvegicus*
MERTELLKPRTLADLIRILHELFAGDEVNVEEVQAVLEAYESNPAEWALYAKFDQYRYTR
NLVDQGNGKFNLMILCWGEGHGSSIHDHTDSHCFLKLLQGNLKETLFDWPDKKSNEMIKK
SERTLRENQCAYINDSIGLHRVENVSHTEPAVSLHLYSPPFDTCHAFDQRTGHKNKVTMT
FHSKFGIRTPFTTSGSLENN

SEQ ID NO:36
Description: Mammalian CDO
Alias: P21816, M35266.1, CDO_Rat
Length: 603
Type: DNA
Organism: *Rattus norvegicus*
>ATGGAACGGACCGAGCTGCTGAAGCCCCGGACCCTGGCCGACCTCATCCGAATCTTGCATGAGCTCTTCGCC
GGGGACGAAGTCAATGTGGAGGAGGTGCAGGCTGTGCTGGAAGCCTACGAGAGCAATCCTGCCGAGTGGGCTT
TGTATGCCAAATTCGATCAATACAGGTATACCCGAAACCTTGTGGATCAAGGAAATGGGAAGTTTAATCTGAT
GATTCTGTGCTGGGGTGAAGGGCATGGCAGCAGTATTCACGATCACACGGACTCCCACTGCTTTTTGAAGCTG
CTGCAAGGAAATCTAAAGGAGACATTGTTTGACTGGCCTGACAAGAAATCCAACGAGATGATCAAGAAGTCTG
AAAGAACTTTGAGGGAAAATCAGTGTGCCTACATTAATGATTCTATTGGCTTACATCGAGTAGAGAACGTCAG
CCACACAGAGCCTGCTGTGAGCCTTCACTTGTACAGTCCACCTTTCGATACATGCCATGCCTTTGACCAACGA
ACAGGGCATAAAAACAAAGTCACCATGACATTCCACAGCAAATTTGGAATCAGAACTCCATTTACAACTTCAG
GTTCACTGGAGAACAACTAA

SEQ ID NO:37
Description: cuyA
Length: 339
Type: Protein
Organism: *Ruegeria pomeroyi*
>MHLARYPRRFIAHLPTPLERLDRLTAELGGPEIWIKRDDCTGLSTGGNKTRKLEFLMAEA
ELQGADMVMTQGATQSNHARQTAAFAAKLGMDCHILLEDRTGSNNANYNNNGNVLLDHLHGATT
EKRPGSGLDMNAEMEKVAEKFRADGRKVYTIPGGGSNPTGALGYVNCAFEMLNQFNERGLKVDHI
VHATGSAGTQAGLITGLQAMNAQIPLLGIGVRAPKPKQEENVYNLACATAE
KLGCPGVVAREDVVANTDYVGEGYGIPTESGLEAIRMFAELEAILLDPVYSAKGAAGFID
LIRKGHFKKGERVVFLHTGGAVALFGYDNAFDYSGRWVA

SEQ ID NO:38
Description: cuyA
Length: 1020
Type: DNA
Organism: *Ruegeria pomeroyi*
>atgCATCTTGCCCGCTATCCCCGCCGCTTCATCGCCCATCTGCCGACGCCGCTGGAACGGCTGGA
CCGGCTGACCGCCGAACTGGGCGGGCCCGAGATCTGGATCAAGCGCGACGACTGCACCGGCCTG
TCCACCGGCGGCAACAAGACCCGCAAGCTGGAATTCCTGATGGCCGAGGCCGAGCTGCAAGGCG
CTGACATGGTGATGACGCAGGGCGCGACCCAGTCCAACCATGCCCGCCAGACCGCCGCATTCGC
CGCCAAGCTGGGCATGGATTGCCATATCCTGCTCGAGGACCGGACCGGCTCGAACAACGCCAAC
TACAACAACAACGGCAACGTTCTGCTCGACCATCTGCATGGCGCCACCACTGAAAAGCGCCCCG
GCAGCGGTCTGGACATGAATGCCGAGATGGAAAAGGTGGCCGAGAAGTTCCGCGCCGACGGGC
GCAAGGTCTATACCATCCCCGGCGGCGGCTCGAACCCGACCGGCGCGCTGGGATATGTCAACTG
CGCTTTCGAGATGCTGAACCAGTTCAATGAGCGCGGGCTGAAGGTGGACCATATCGTGCATGCC
ACCGGCAGCGCGGGCACCCAGGCAGGGCTGATCACCGGGCTTCAGGCGATGAACGCTCAGATCC
CGCTCTTGGGCATCGGCGTGCGTGCGCCCAAGCCCAAGCAGGAAGAGAATGTCTATAACCTGGC
CTGCGCCACCGCCGAGAAGCTGGGTTGCCCCGGTGTCGTCGCGCGCGAGGACGTGGTGGCCAAT
ACCGACTATGTCGGCGAAGGCTATGGCATCCCGACCGAAAGCGGGCTGGAGGCGATCCGCATGT
TCGCCGAGCTTGAGGCGATCCTGCTTGACCCGGTCTATTCGGCCAAGGGCGCGGCTGGCTTCATC
GACCTGATCCGCAAGGGTCATTTCAAAAAGGGCGAGCGGGTGGTGTTCCTGCATACCGGCGGCG
CTGTGGCGCTGTTCGGCTATGACAACGCCTTTGACTATTCGGGACGCTGGGTGGCCtaa

SEQ ID NO:39
Description: promoter P_tac
Length: 74
Type: DNA
Organism: *Methylobacterium extorquens*
>GGTCGACTCTAGTTCTGAAATGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAGG
CCTCATATGT

SEQ ID NO:40
Description: promoter P_tacA
Length: 80
Type: DNA
Organism: *Methylobacterium extorquens*
>GGTCGACTCTAGTAAGAAATCTGAAATGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTG
TGGAGGCCTCATATGT

SEQ ID NO:41
Description: promoter P_Lac
Length: 33
Type: DNA
Organism: *Methylobacterium extorquens*
>TTTACACTTTATGCTTCCGGCTCGTATGTTGTG

SEQ ID NO:42
Description: promoter P_R
Length: 109
Type: DNA
Organism: Bacteriophage 16-3
>CAACAACTTATACCATGGCCTACAAAAAGGCAAACAATGGTACTTGACGACTCATCACAACAA
TTGTAGTTGTAGATTGTAAAGATCTAGGGAGAGACCCCGAGGTACC

SEQ ID NO:43
Description: promoter PmxaF
Length: 109

Type: DNA
Organism: *Methylobacterium extorquens*
>CGACACTACGCCTTGGCACTTTTAGAATTGCCTTATCGTCCTGATAAGAAATGTCCGACCAGCT
AAAGACATCGCGTCCAATCAAAGCCTAGAAAATATAG

SEQ ID NO:44
Description: ADO (2-aminoethanol dioxygenase)
Alias: Gm237, NP_001005419.2
Length: 256
Type: Protein
Organism: *Mus musculus*
>MPRDNMASLIQRIARQACLTFRGSSTGSEGPAPGFPENLSLLKSLLTQVRAEDLNIAPRKALPQPLPR
NLPPVTYMHIYETEGFSLGVFLLKSGTCIPLHDHPGMHGMLKVLYGTVRISCMDKLDTGAGHRRPPP
EQQFEPPLQPLEREAVRPGVLRSRAEYTEASGPCVLTPHRDNLHQIDAVDGPAAFLDILAPPYDPEDG
RDCHYYRVVEPIRPKEASGSACDLPREVWLLETPQADDFWCEGEPYPGPKVLP

SEQ ID NO:45
Description: ADO (2-aminoethanol dioxygenase)
Alias: Gm237, NP_001005419.2
Length: 771
Type: DNA
Organism: *Mus musculus*
>ATGCCCCGCGACAACATGGCCTCCCTGATCCAGCGCATCGCTCGCCAGGCGTGTCTCACC
TTCCGCGGCAGCTCGACGGGCTCCGAAGGGCCGGCGCCGGGCTTCCCGGAGAACCTGAGC
CTGCTCAAGAGCCTGCTGACCCAGGTGCGCGCCGAGGACCTCAACATCGCGCCGCGCAAG
GCGCTGCCGCAGCCGCTGCCCCGCAACCTCCCGCCGGTCACCTACATGCACATCTACGAG
ACGGAGGGCTTCAGCCTGGGCGTGTTCCTGCTCAAGAGCGGCACGTGCATCCCGCTGCAC
GACCACCCGGGCATGCACGGTATGCTCAAGGTGCTGTACGGCACGGTCCGCATCAGCTGC
ATGGACAAGCTGGACACGGGGGCCGGGCATCGGCGGCCGCCGCCAGAGCAGCAGTTCGAG
CCCCCGCTGCAGCCCTTGGAGCGGGAGGCCGTGCGACCGGGCGTGCTGCGTTCCCGGGCC
GAGTACACCGAGGCCAGTGGGCCCTGCGTGCTCACTCCACACCGGGACAACCTGCACCAG
ATTGATGCCGTGGACGGGCCAGCTGCCTTCCTGGACATCCTGGCCCCACCCTACGACCCG
GAGGACGGCCGGGACTGCCACTATTACCGTGTAGTGGAGCCCATCAGACCCAAGGAGGCT
TCCGGCTCTGCCTGCGACCTTCCCCGAGAAGTGTGGCTCCTGGAGACACCACAGGCCGAC
GACTTCTGGTGCGAGGGAGAGCCCTATCCAGGCCCCAAGGTCCTACCTTGA

SEQ ID NO:46
Description: cystathionine gamma-lyase
Alias: mccB, BSU27250
Length: 379
Type: Protein
Organism: *Bacillus subtilis* 168
>MKKKTLMIHGGITGDEKTGAVSVPIYQVSTYKQPKAGQHTGYEYSRTANPTRTALEALVTELESGE
AGYAFSSGMAAITAVMMLFNSGDHVVLTDDVYGGTYRVMTKVLNRLGIESTFVDTSSREEVEKAIR
PNTKAIYIETPTNPLLKITDLTLMADIAKKAGVLLIVDNTFNTPYFQQPLTLGADIVLHSATKYLGGHS
DVVGGLVVTASKELGEELHFVQNSTGGVLGPQDSWLLMRGIKTLGLRMEAIDQNARKIASFLENHP
AVQTLYYPGSSNHPGHELAKTQGAGFGGMISFDIGSEERVDAFLGNLKLFTIAESLGAVESLISVPAR
MTHASIPRERRLELGITDGLIRISVGIEDAEDLLEDIGQALENI

SEQ ID NO:47
Description: cystathionine gamma-lyase
Alias: mccB, BSU27250
Length: 1140
Type: DNA
Organism: *Bacillus subtilis* 168
>atgAAGAAAAAAAACATTGATGATACATGGCGGAATCACAGGTGATGAGAAAACAGGCGCAGTTT
CCGTGCCGATTTATCAAGTAAGCACGTACAAGCAGCCGAAAGCAGGGCAGCATACAGGCTACGA
GTATTCAAGAACGGCCAATCCGACTCGAACCGCTCTCGAAGCACTTGTGACAGAACTGGAAAGC
GGGGAAGCAGGCTATGCGTTCAGCTCAGGAATGGCTGCCATTACAGCGGTTATGATGCTGTTTA
ACAGCGGAGATCATGTCGTGTTGACTGATGATGTGTACGGCGGAACATATCGCGTGATGACAAA

GGTGCTTAACCGTCTTGGCATTGAATCAACATTTGTTGATACGAGCAGCAGGGAAGAAGTTGAA
AAAGCGATTCGCCCTAATACAAAAGCAATTTATATTGAAACACCGACAAACCCGTTGCTCAAAA
TCACCGACCTGACGCTCATGGCTGATATCGCAAAAAAAGCGGGTGTTCTGCTTATCGTAGACAAT
ACCTTTAATACTCCTTATTTTCAACAGCCGCTTACTTTAGGCGCTGATATCGTACTGCACAGTGCG
ACAAAATATCTTGGCGGACACAGTGATGTCGTCGGAGGTTTAGTTGTGACAGCTTCGAAAGAGC
TTGGAGAAGAGCTGCATTTTGTGCAAAACTCCACAGGCGGCGTGCTCGGCCCTCAAGATTCCTGG
CTGTTAATGAGAGGAATCAAAACGTTGGGACTCAGAATGGAAGCGATCGATCAAAATGCGCGGA
AAATCGCAAGCTTTCTTGAGAATCACCCTGCTGTCCAAACGTTATATTACCCTGGTTCTTCAAATC
ATCCCGGACATGAGCTTGCAAAAACGCAAGGAGCGGGCTTCGGCGGCATGATCTCCTTTGATAT
TGGCAGTGAAGAACGGGTTGATGCGTTTTTAGGAAATCTGAAACTGTTTACCATTGCTGAAAGCC
TGGGGGCGGTTGAAAGCTTAATTTCTGTTCCTGCAAGAATGACACATGCCTCTATTCCGAGAGAA
CGCCGGCTTGAGCTCGGCATTACGGACGGCTTGATCAGAATTTCTGTAGGAATTGAAGATGCGG
AAGACTTGTTGGAAGATATCGGCCAAGCGCTTGAAAATATAtaa

SEQ ID NO:48
Description: Sulfoacetaldehyde acetyltransferase
Alias: Xsc
Length: 593
Type: Protein
Organism: *Paracoccus denitrificans*
>MRMTTEESFVKTLQLHGIEHAFGIIGSAMMPVSDLFPRAGITFWDCAHETNAGMMADGFTRSTGR
MSMAIAQNGPGVTGFVTPVKTAYWNHTPLLLVTPQAANRTIGQGGFQEMEQMRIFADCVCYQEEVR
DPSRIPEVLNRVIMQAWRNSAPAQINIPRDFWTQVIDVDLPQVVGFERPAGGERAVAEAARLLSEARF
PVILSGAGVVLSGAIPDLVGLAERLDAPVCSNYQHNDSFPGSHPLAMGPLGYNGSKAAMEIIARADV
VLALGTRLNPFSTLPGYGIDYWPKDARIIQVDINADRIGLTKKVAVGIQGDAAKVARGILAQLAPAAG
DAGRQERRDLVAQTRSRWAQELSSLDHEEDDPGTEWNEQARARDAGLMSPRQAWRAIMQAVPKE
AIVSSDIGNNCAIGNAYPSFEAGRKYLAPGLFGPCGYGFPAILGAKIGNPEVPVIGFAGDGAFGISMNE
MTACGREDWPAITMVIFRNYQWGAEKRNTTLWYDNNFVGTELDRDTSYAKIAQACGLVGVQVRSQ
EELTAALHDAVERQMQGRETTFIEVLLNQELGEPFRRDAMKKPVAVAGIDPADMRPQQGAA

SEQ ID NO:49
Description: Sulfoacetaldehyde acetyltransferase
Alias: Xsc
Length: 1782
Type: DNA
Organism: *Paracoccus denitrificans*
>ATGCGAATGACGACTGAGGAGTCTTTTGTCAAAACCCTTCAATTGCACGGGATCGAGCATGCCT
TTGGCATTATCGGCTCTGCGATGATGCCTGTTTCGGACCTGTTTCCGCGGGCTGGGATCACGTTCT
GGGACTGTGCGCATGAGACGAATGCCGGGATGATGGCGGACGGTTTCACGCGCTCGACGGGGCG
GATGTCGATGGCGATCGCGCAGAACGGTCCCGGGGTGACGGGGTTCGTGACGCCGGTCAAGACG
GCTTACTGGAACCACACGCCCTTGCTGCTGGTGACGCCGCAGGCGGCGAACCGGACCATCGGGC
AGGGCGGTTTCCAGGAGATGGAGCAGATGCGCATCTTCGCCGATTGCGTCTGCTACCAGGAGGA
GGTGCGCGACCCGAGCCGCATCCCCGAGGTTCTGAACCGGGTGATCATGCAGGCCTGGCGCAAC
TCGGCGCCGGCGCAGATCAACATCCCGCGCGATTTCTGGACCCAGGTGATCGACGTGGATCTGC
CGCAGGTGGTGGGCTTCGAGCGGCCGGCGGGCGGCGAGCGGGCGGTGGCCGAGGCGGCCAGGC
TGCTCTCCGAGGCGCGGTTCCCGGTGATCCTGTCGGGCGCCGGCGTGGTGCTGTCGGGCGCGATC
CCGGACCTGGTCGGGCTGGCCGAGCGGCTGGATGCGCCGGTCTGCTCGAACTACCAGCACAATG
ACAGCTTCCCGGGCAGCCACCCGCTGGCCATGGGGCCGCTGGGCTACAACGGCTCGAAGGCGGC
GATGGAGATCATCGCCCGGGCCGACGTGGTGCTGGCGCTGGGGACGCGGCTCAATCCGTTCTCG
ACCCTGCCGGGCTACGGCATCGACTACTGGCCGAAGGATGCCAGGATCATCCAGGTCGACATCA
ATGCCGACCGCATCGGGCTGACCAAGAAGGTGGCGGTGGGCATCCAGGGCGATGCGGCCAAGG
TGGCGCGCGGCATCCTGGCGCAGCTGGCCCCGGCCGCCGGCGATGCCGGGCGGCAGGAGCGCCG
CGACCTGGTGGCGCAGACCCGGTCCCGCTGGGCGCAGGAACTGTCGAGCCTGGACCACGAGGAG
GACGATCCCGGCACCGAATGGAACGAGCAGGCGCGGGCCCGCGACGCCGGTCTGATGAGCCCG
CGCCAGGCCTGGCGGGCGATCATGCAGGCGGTGCCGAAGGAGGCGATCGTCAGCTCGGAC

SEQ ID NO:50
Description: CDO

Alias: cdoA, BSU31140, O32085, CDO_Bacillus
Length: 486
Type: DNA
Codon Optimization: *E. coli*
>ATGGAACTGTATGAATGTATTCAGGATATTTTTGGTGGTCTGAAAAATCCGAGCGTTAAAGATC
TGGCAACCAGCCTGAAACAGATTCCGAATGCAGCAAAACTGAGCCAGCCGTATATTAAAGAACC
GGATCAGTATGCATATGGTCGTAATGCAATTTATCGTAATAATGAACTGGAAATTATTGTTATTA
ATATTCCGCCGAATAAAGAAACCACCGTTCATGATCATGGTCAGAGCATTGGTTGTGCAATGGTT
CTGGAAGGTAAACTGCTGAATAGCATTTATCGTAGCACCGGTGAACATGCAGAACTGAGCAATA
GCTATTTTGTTCATGAAGGTGAATGTCTGATTAGCACCAAAGGTCTGATTCATAAAATGAGCAAT
CCGACCAGCGAACGTATGGTTAGCCTGCATGTTTATAGCCCGCCGCTGGAAGATATGACCGTTTT
TGAAGAACAGAAAGAAGTTCTGGAAAATAGCTGA

SEQ ID NO:51
Description: Mammalian CDO
Alias: P21816, M35266.1, CDO_Rat
Length: 1482
Type: DNA
Codon Optimization: *E. coli*
>ATGGAACGTACCGAACTGCTGAAACCGCGTACCCTGGCAGATCTGATTCGTATTCTGCATGAAC
TGTTTGCCGGTGATGAAGTTAATGTTGAAGAAGTTCAGGCAGTTCTGGAAGCATATGAAAGCAA
TCCGGCAGAATGGGCACTGTATGCAAAATTTGATCAGTATCGTTATACCCGTAATCTGGTTGATC
AGGGTAATGGTAAATTTAATCTGATGATTCTGTGTTGGGGTGAAGGTCATGGTAGCAGCATTCAT
GATCATACCGATAGCCATTGTTTTCTGAAACTGCTGCAGGGTAATCTGAAAGAAACCCTGTTTGA
TTGGCCGGATAAAAAAAGCAATGAAATGATTAAAAAAAGCGAACGTACCCTGCGTGAAAATCA
GTGTGCATATATTAATGATAGCATTGGTCTGCATCGTGTTGAAAATGTTAGCCATACCGAACCGG
CAGTTAGCCTGCATCTGTATAGCCCGCCGTTTGATACCTGTCATGCATTTGATCAGCGTACCGGT
CATAAAAATAAAGTTACCATGACCTTTCATAGCAAATTTGGTATTCGTACCCCGTTTACCACCAG
CGGTAGCCTGGAAAATAATTAA

SEQ ID NO:52
Description: MA_3297
Length: 1251
Type: DNA
Codon Optimization: *E. coli*
>ATGGGTCGTTTTATTCTGAAATGTCTGAAATGTGGTCGTGAATATAGCCAGGAATATCGTCTGA
CCTGTGAAAATGATGATAGCTTTCTGCGTGCAGAATATCTGGAAAAAAAACTGGAACTGCGTAA
ACAGCCGGGTATTGGTCGTTTTCATAGCTGGCTGCCGGTTCAGGAAGAACTGACCACCGAAGCA
GGTCCGATTACCTATAAAAGCGAAGCACTGGCACGTGAACTGGGTCTGAGCAATCTGTATATTG
GTTTTAGCGGTTATTGGCCGGAAAAAGGTGCATTTATTAAAACCTGTAGCTTTAAAGAACTGGAA
GCACATCCGACCATGCAGCTGCTGAAAGAAAGCGGTGGTAAAGCAATTGTTCTGGCAAGCGCAG
GTAATACCGGTCGTGCATTTGCACATGTTAGCGCACTGACCGGTACCGATGTTTATATTGTTGTTC
CGGATAGCGGTATTCCGAAACTGTGGCTGCCGGAAGAACCGACCGATAGCATTCATCTGATTAG
CATGACCCCGGGTAATGATTATACCGATGCAATTAATCTGGCAGGTCGTATTGCAAAACTGCCGG
GTATGGTTCCGGAAGGTGGTGCACGTAATGTTGCACGTCGTGAAGGTATGGGTACCGTTATGCTG
GATGCAGCAGTTACCATTGGTAAAATGCCGGATCATTATTTTCAGGCAGTTGGTAGCGGTACCGG
TGGTATTAGCGCATGGGAAGCAAGCCTGCGTCTGCGTGAAGATGGTCGTTTTGGTAGCAAACTG
CCGAAACTGCAGCTGACCCAGAATCTGCCGTTTGTTCCGATGTATAATGCATGGCAGGAAGGTC
GTCGTGATATTATTCCGGAAATTGATATGAAAGATGCAAAAAAACGTATTGAAGAAACCTATGC
AACCGTTCTGACCAATCGTGCACCGCCGTATAGCGTTACCGGTGGTCTGTATGATGCACTGGTTG
ATACCGATGGTATTATGTATGCAGTTAGCAAAGAAGAAGCACTGGATGCAAAAGCACTGTTTGA
AAGCCTGGAAGGTATTGATATTCTGCCGCCGAGCGCAGTTGCAGCAGCAAGCCTGCTGAAAGCA
GTTGAAGCAGGTAATGTTGGTAAAGATGATACCATTCTGCTGAATATTGCCGGTGGTGGTTTTAA
ACGTCTGAAAGAAGATTTTACCCTGTTTCAGATTGAACCGGAAATTACCGTTAGCAATCCGGATG
TTCCGCTGGAAGAACTGAAACTGTGA

SEQ ID NO:53
Description: CSAD

Alias: ref|WP_006454033.1
Length: 1467
Type: DNA
Codon Optimization: *E. coli*
>ATGTTTAAAGCAAGCAAATATTATAATCTGCTGCAGCAGCTGGAAAATTTTTTTAGCACCGCAA
ATAGCAGCAGCCTGCTGACCAAACCGATTGATCCGAATGTTCTGAAAAGCCAGCTGAGCCTGGA
TCTGCCGAATGAAGGTAAACCGGTTGAAGAACTGCGTACCGAAATTACCAGCTATCTGAATAAT
GCACTGAAAACCGCACATCCGAGCTATTTTAATCAGCTGTGGGGTGGTTTTAATAGCGCATGTTT
TATGGGTGATATGCTGGCAAGCGCAACCAATACCAGCATGTATACCTATGAAGTTGCACCGGCA
GCAACCCTGATTGAACAGGCACTGGTTACCAAAATGAGCGGTATTCTGGGTTTTAAAAGCGCAG
ATGGTCAGTTTACCACCGGTGGTAGCAATGGTAATCTGATGGCAATGGCAATTGCACGTCATCAT
GTTCTGCCGACCGTTAAACAGGATGGTATGACCAGCGGTCCGAAACTGGTTGCATTTGTTAGCCG
TGAAGCACATTATAGCTTTGATAAAGCAGCACATATTCTGGGTCTGGGTACCGAACAGCTGTGG
AAAGTTCCGGTTGATAGCGATGGTCGTATGAAACCGGAAGCACTGAGCGAACTGGTTGATCGTG
CACGTGTTCAGGGTAGCATTCCGTTTTTTGTTGCAGGTACCGCAGGTACCACCGTTCGTGGTGCA
TTTGATCCGTTTGAAGAAATTAGCGCAATTGCACATCAGGAAAATCTGTGGTTTCATATTGATGG
TGCATGGGGTGCAAGCGTTAGCCTGAGCGCAACCCATCGTCAGCTGATGGCCGGTGCAAATCAG
GCAGATAGCCTGGTTTGGGATGCACATAAAATGATGGGTATGACCCTGATGTGTAGCCTGCTGCT
GGTTAAACAGCGTGGTCAGATGCTGCGTACCTTTAGCACCGCAGGTACCGATTATCTGTTTCATG
ATGAAGTTAGCGCCGGTGAAGTTCCGACCGAAAGCAGCACCAGCAGCACCGAACTGCCGATTGA
AGAACTGCCGACCGATTTTGGTCCGGCAACCATGCATTGTGGTCGTCGTGTTGATGCACTGAAAC
TGTGGCTGGCATGGCGTCATCTGGGTGATCGTGGTTGGGAACGTCTGATTGATAGCTATTTTGAA
CTGGCACAGCGTGCAGAAACCATTATTGATAAACATCCGAGCCTGGAACTGGTTAGCAGCCGTC
AGAGCGTTAATCTGTGTTTTCGTTATCTGCCGCAGAATAAACAGCAGGCAGATGAACTGACCCTG
AAAGTTCGTCAGGCACTGTGGGAAACCGGTACCGCAATGGTTAATTATGCACAGGTTGAAGGTA
AAACCGTTTTTCGTCTGGTTATTTGTAATAATCAGACCCGTAGCGAAGATATTGAACGTTTTTTTG
AAGCACTGGTTGCAATTGCACGTCGTCTGGAACAGGAAATGTGTTGA

SEQ ID NO:54
Description: CSAD
Alias: ref|WP_006454033.1
Length: 1467
Type: DNA
Codon Optimization: *M. extorquens*
>ATGTTCAAGGCCTCGAAGTACTACAACCTCCTCCAGCAGCTCGAGAACTTCTTCTCGACCGCCA
ACTCGTCGTCGCTCCTCACCAAGCCGATCGACCCGAACGTCCTCAAGTCGCAGCTTTCGCTCGAC
CTCCCGAACGAGGGCAAGCCGGTCGAGGAGCTCCGCACCGAGATCACCTCGTACCTCAACAACG
CCCTCAAGACCGCCCACCCGTCGTACTTCAACCAGCTCTGGGGCGGCTTCAACTCGGCCTGCTTC
ATGGGCGATATGCTCGCCTCGGCCACCAATACCTCGATGTACACCTACGAGGTCGCCCCGGCCGC
CACCCTCATCGAACAGGCCCTCGTCACCAAGATGTCGGGCATCCTCGGCTTCAAGTCGGCTGATG
GCCAGTTTACCACCGGCGGTTCGAACGGCAACCTCATGGCCATGGCCATCGCCCGCCACCACGTT
CTCCCGACCGTCAAGCAGGATGGTATGACCTCGGGCCCGAAGCTCGTCGCCTTTGTCTCGCGCGA
AGCCCATTACTCGTTCGACAAGGCCGCCCACATCCTCGGCCTCGGCACCGAGCAGCTTTGGAAG
GTCCCGGTCGACTCGGATGGCCGCATGAAGCCGGAAGCTCTTTCGGAGCTCGTTGACCGCGCCA
GAGTCCAAGGCTCGATCCCGTTTTTCGTCGCTGGCACCGCCGGCACCACCGTCCGTGGTGCCTTC
GATCCGTTCGAGGAGATCTCGGCCATTGCCCACCAGGAGAACCTCTGGTTCCACATTGATGGCGC
CTGGGGCGCCAGCGTCTCGCTTTCGGCCACCCACCGCCAACTCATGGCTGGTGCCAACCAGGCCG
ATTCGCTTGTCTGGGATGCCCACAAGATGATGGGCATGACCCTCATGTGCTCGCTCCTCCTCGTC
AAGCAGCGTGGCCAGATGCTCCGCACCTTCTCGACCGCTGGCACCGACTACCTCTTCCACGACGA
GGTCAGTGCTGGCGAGGTCCCGACCGAATCGTCGACCAGTTCGACCGAACTCCCGATCGAAGAG
CTTCCGACCGACTTCGGCCCGGCCACCATGCATTGCGGTCGTCGCGTCGATGCTCTTAAACTTTG
GCTCGCCTGGCGCCACCTCGGTGATCGTGGCTGGGAGCGCCTCATCGACTCGTACTTCGAGCTCG
CCCAGCGTGCCGAAACCATCATCGACAAGCACCCGTCGCTCGAGCTCGTCTCGTCGCGCCAGTCG
GTCAACCTCTGCTTCCGCTACCTCCCGCAGAACAAGCAACAGGCCGACGAGCTCACCCTTAAGGT
CCGCCAGGCCCTCTGGGAGACGGGCACCGCCATGGTCAACTACGCCCAGGTCGAAGGCAAGACC
GTTTTCCGCCTCGTCATCTGCAACAATCAGACCCGCTCGGAGGACATCGAGCGCTTCTTCGAGGC
CCTCGTCGCCATCGCCCGCCGCCTCGAGCAGGAGATGTGCTGA

SEQ ID NO:55
Description: ComDE
Alias: sulfopyruvate decarboxylase
Length: 1164
Type: DNA
Codon Optimization: *M. extorquens*
>ATGTACGTCGTCAACCCGGAGGAGAAGGTCATCGAGATCATGAAGCAGACCGGCATCGACCTC
GCCGCCACCCTCCCGTGCGACCGCATCAAGAACCTCCTCCCGCTCGTCTCGGAGAACTTCCCGGA
GATTAAGCTCACCCGCGAGGAGAACGGTGTCGGCATCTGCGCCGGTATCTACCTCGCCGGCGGC
AAGCCGATGATGCTCATCCAGTCGACCGGCCTCGGCAACATGATCAACGCCCTCGAGTCGCTCA
ACGTGACCTGCAAGATCCCGCTCCCGATCCTTGCCTCGTGGCGCGGCGTCTATAAGGAAGGCATC
GAAGCCCAGGTCCCGCTCGGTGCCCACCTTCCTTCGATCCTTGAGGGTGCCGGCCTCACCTACAC
CATCATCGGCGAGACGGAGAAGCTCCCGCTCCTCGAGAACGTCATCCTTGACGCCTTCGAGAAC
TCGCGTCCGCATATCGCCCTCGTTTCGCCGAAGGTCTGGGAAGCCTCGGAATGCTGCGCCTGGCA
GGCCGCTGGCATGCCGATCAAGCCGGAGATTATGGAGCGCACGTGCCGTTTCTCGCTCACCTCGG
GCACCCTCAAGCCGTTCATGCTCCGCAACGATGCCATCTGCACCCTCGCCTCGGAGCTCGACGAC
GAGATCACCGTCACCAACCTCGGCGTCCCGTGTAAGGAGCTCTACGCCTGCCGCGACCGCGAAC
TCAACTTCTACATGTTCGGCTCGATGGGCCTCGTCTCGTCGATCGGCCTCGGCCTCGCCCTCCGCT
CGGAAAAGACCGTCATCACCTTCGATGGCGACGGCTCGCTTCTCATGAACCCGAACGCCCTCCTC
GAGATCGCCAAGGAGGCCCCGAAGAACCTCATCATCATCGCCCTCGACAACGGCGCCTATGGCT
CGACCGGCTCGCAGGAAACCTGCGCCCTCCGCTACATCGATCTCGAGATCTTCGCCAACGCCTGC
GGCATCCAGAACACCGCCAAGGTCAACTCGAAGGAGGGCGTCATCGAGGCCTTCCGCAAGTTCA
AGGCCATGCGCGAGCTCTCGTTCATCCACGTCATCCTCAAGCCGGGCAACACCAACGCCCCGAA
CATCCCGATGTCGCCGGAGGAGGCCACCAAGCGCTTCAAGGAAACCCTCGACGTCAAGAAGTTC
TAA

SEQ ID NO:56
Description: Taurine-pyruvate aminotransferase (Tpa)
Length: 1393
Type: DNA
Codon Optimization: *M. extorquens*
>ATGGTCGTCGATGTCACCGAGTTACGTGCCCGCGCCCGCCGCCACCTCGGCCCGCATTTCACCC
GCAAGGATACCTGGGAATCGGATTTCCCGGTGTTCGTCCGCGGTGAAGGTTCGTACCTCATCGAC
ACCGAAGGCGATCGCTTCCTCGACGGCCTCGCCGGTCTGTTCTGCGTCAACATCGGCCATGGTCG
CGACGACATCGCCAAGGCCGCCTCGGAACAGATTGGCACCCTTGCCTATGCCTCGAACTGGGGC
TCGGCCCACATCCCGGCTATTGAGGCCTCGGCTCTCATCGCCGATCTTGCCCCGGGCGATCTCGG
CACTACCTTCTTCGTCAACTCGGGTTCGGAGGCCGTCGAAACCGCCGTCAAGTTCGCCCGCCAGT
ACCACCGCTCGCAGGGTAATCCGCAGCGCACCAAAATCATCTCGCGCGAGATGGCCTACCATGG
CACCACCCTCGGCGCCCTCAGTGTCACCCAGCTCCCTAAGATCAAGGACCCGTTCGGTCCGCTTC
TTCCGGGCGTTCGTTCGGTCCCGAATACCCTCGGCTACCTCGGTGATTGCGGTCCGGCCAACGAG
CTCGATTGCATCGCCGCCATCGAGGCCGTCATCGAGGAGGAGGGTGCCGAAACCATCGCTGCTG
TGTTCGCCGAACCGGTCCAAAATGGCCGCGGTGCCCTTGTCCCTCCTGATGGTTACTGGGCCGCT
CTCCGCGCCCTCTGCGACAAGCATGGCATCCTCCTCGTCTCGGACGAAGTCATCTGCTCGTTCGG
TCGCCTCGGCCACTGGTTCGGTCATGGCCTTACCGGCGTCGTCCCGGACATGATCACCTTCGCCA
AGGGCTCGACCTCGGGCTATGCTCCTCTCGGTGGCTTGATCGTCCGCGAGCAGCTCGTCCGTGAG
CTCTATGATTCGCCGAAGGGTGGCGTGTTCACTCACGGCGCTACCTGGGGTGGCCATCCTGTCTC
GACCGCCGTCGCTGTCGCCAACATCACCGCCATGCGCGATGAAAACGTCCTTGGCAACGTCAGT
GCCCGCGGCCCGAAGCTCCGCAGTGCTCTTGATTCGCTCATGTCGTCGCATCGCTGCGTCAAGGA
CGTCCGTGGCACCGGCTTCTTCTATGCCATCGAGCTCATGGCCGACTCGGATAGTGGCCGCGAGT
TCACCGAGCAGGAGTCGCTCACCGTCCTCCGCAAAGTTCTCCCGGAGGCCTTCGCCCGCACCAAG
GTCATCCTCCGTGGTGATGATCGTGGCGCCACCATGCTCATGATCTCGCCGCCGCTCGTCGCTGA
CGACGAAGTCCTCTCGGAGCTCCTCCACGGCATCGACTCGATGCTCACCGACATCGAGAAGGCC
ATCCAGCCGTAG

SEQ ID NO:57
Description: ADO (2-aminoethanol dioxygenase)
Alias: Gm237, NP_001005419.2
Length: 771

Type: DNA
Codon Optimization: *M. extorquens*
>ATGCCGCGTGATAACATGGCCTCGCTTATCCAGCGCATTGCCCGCCAAGCCTGCCTCACCTTCC
GCGGTTCGAGTACCGGCTCGGAGGGCCCGGCTCCGGGCTTCCCGGAAAACCTCTCGCTCCTCAA
GTCGCTTCTCACCCAGGTCCGTGCCGAGGATCTTAACATCGCCCCGCGTAAGGCCCTCCCGCAGC
CGCTCCCGCGCAACCTCCCGCCGGTCACCTACATGCACATCTACGAAACCGAGGGCTTCTCGCTC
GGCGTGTTCCTCCTCAAGTCGGGCACGTGCATCCCGCTCCACGACCACCCGGGCATGCACGGCAT
GCTCAAGGTCCTCTACGGCACCGTCCGCATCTCGTGCATGGACAAGCTCGACACCGGTGCCGGCC
ATAGACGTCCGCCTCCGGAACAGCAGTTCGAGCCTCCGCTTCAGCCGCTCGAACGCGAAGCCGT
TCGCCCGGGCGTCCTTAGAAGTCGCGCCGAATACACCGAGGCCAGTGGTCCGTGCGTCCTCACCC
CGCACCGTGATAACCTCCATCAGATCGATGCCGTCGACGGCCCGGCCGCCTTCCTCGATATCCTC
GCCCCGCCGTACGACCCGGAGGATGGCCGCGATTGCCATTATTATCGCGTCGTCGAGCCGATCCG
CCCGAAGGAAGCCTCGGGTTCGGCCTGTGATCTCCCGCGCGAGGTCTGGCTCCTCGAAACCCCGC
AGGCCGACGACTTTTGGTGCGAGGGTGAACCGTACCCGGGCCCGAAGGTCCTCCCGTGA

SEQ ID NO:58
Description: CDO
Alias: cdoA, BSU31140, O32085, CDO_Bacillus
Length: 486
Type: DNA
Codon Optimization: *M. extorquens*
>ATGGAGCTCTACGAGTGCATCCAGGACATCTTCGGCGGCCTCAAGAACCCGTCGGTCAAGGAC
CTCGCCACCTCGCTCAAGCAGATCCCGAACGCCGCCAAGCTCTCGCAGCCGTACATCAAGGAGC
CGGACCAGTACGCCTACGGCCGCAACGCCATCTACCGCAACAACGAGCTCGAGATCATCGTCAT
CAACATCCCGCCGAACAAGGAGACGACCGTCCACGACCACGGCCAGTCGATCGGCTGCGCCATG
GTCCTCGAGGGCAAGCTCCTCAACTCGATCTACCGCTCGACCGGCGAGCACGCCGAGCTCTCGA
ACTCGTACTTCGTCCACGAGGGCGAGTGCCTCATCTCGACCAAGGGCCTCATCCACAAGATGTCG
AACCCGACCTCGGAGCGCATGGTGTCGCTCCACGTCTACTCGCCGCCGCTCGAGGACATGACCGT
GTTCGAGGAGCAGAAGGAGGTCCTCGAGAACTCGTGA

SEQ ID NO:59
Description: Mammalian CDO
Alias: P21816, M35266.1, CDO_Rat
Length: 603
Type: DNA
Codon Optimization: *M. extorquens*
>ATGGAGCGCACCGAGCTCCTCAAGCCGCGCACCCTCGCCGACCTCATCCGCATCCTCCACGAGC
TCTTCGCCGGCGACGAGGTCAACGTCGAGGAGGTCCAGGCCGTCCTCGAGGCCTACGAGTCGAA
CCCGGCCGAGTGGGCCCTCTACGCCAAGTTCGACCAGTACCGCTACACCCGCAACCTCGTCGACC
AGGGCAACGGCAAGTTCAACCTCATGATCCTCTGCTGGGGCGAGGGCCACGGCTCGTCGATCCA
CGACCACACCGACTCGCACTGCTTCCTCAAGCTCCTCCAGGGCAACCTCAAGGAGACGCTCTTCG
ACTGGCCGGACAAGAAGTCGAACGAGATGATCAAGAAGTCGGAGCGCACCCTCCGCGAGAACC
AGTGCGCCTACATCAACGACTCGATCGGCCTCCACCGCGTCGAGAACGTCTCGCACACCGAGCC
GGCCGTCTCGCTCCACCTCTACTCGCCGCCGTTCGACACGTGCCACGCCTTCGACCAGCGCACCG
GCCACAAGAACAAGGTCACCATGACCTTCCACTCGAAGTTCGGCATCCGCACCCCGTTCACCACC
TCGGGCTCGCTCGAGAACAACTAA

SEQ ID NO:60
Description: p3MDO
Alias: Q9I0N5, PA2602
Length: 606
Type: DNA
Codon Optimization: *M. extorquens*
>ATGTCGTCGATCCTCCGCCTTGACCGTCTCCGCCAGTTCATCGGCGAGCTCGCCACCCTCCTCGA
TTCGCGCCCGGATGAATCGACCCTCCTCGCCCAGGCCCATCCGCTCCTCGCCGAACTTGTCCATC
AGGATGACTGGCTCCCGGAGGATTGCGCCCGCCCGGACCCGCAGCGCTATCAGCAGTACCTCCT
CCACGTCGACTCGCGTCAGCGCTTCTCGGTCGTCTCGTTCGTCTGGGGCCCGGGTCAGATCACCC

CGGTCCACGATCACCGCGTCTGGGGCCTCATCGGCATGCTTCGTGGCGCCGAGTACTCGCAGCCG
TATGCCTTCGATGCCGGTGGCAGACCGCATCCGTCGGGTGCCAGACGTCGCCTTGAGCCGGGCG
AAGTCGAGGCTCTCTCGCCTCGCATCGGCGATGTCCACCAGGTGTCGAACGCCTTCTCGGACCGC
ACCTCGATCTCGATCCACGTCTACGGCGCTAACATCGGCGCCGTCCGCCGCGCCGTGTTCTCGGC
CGAGGGTGAGGAGAAGCCGTTCATCTCGGGCTACTCGAACTCGCGCCTCCCGAACATCTGGGAC
CTCTCGAAGGAGAACCCGGCCTGA

SEQ ID NO:61
Description: PAPS-AS
Alias: OT_ostta05g01260
Length: 696
Type: DNA
Codon Optimization: *M. extorquens*
>ATGCCGCGCGGCTGGACCAAGACCCGCGCCTATGACTCGCATCATTTCGATGCCGACGCCTGGT
CGGTCGTCACCCCGCGCGCCGGTGATGTCATTATCGCCACCGCCTACAAGTCGGGCACCACCTGG
ATGCAGCAGATCGTCTCGCAGCTCGTTTTCGAGGGCGCCGCCCCGGCTGCCCTCGGCGAACTTAG
TCCTTGGGTCGATCTCCGTGTTCCTCCTCGCGAAGTCAAGCGCGGTATGATTGAGGGCCTCCCGT
CGCCGCGCATTCTCAAGACCCATCTCCCGACCACCGGCCTCGAGTATGACGAGAACGCCAAGTA
CATCTACGTCGCCCGCGACGGCCGCGACGCCTTCATGTCGCTCATGAACCACTACAAGAACGGC
AACGAGGCCTTCTATGGCGCCCTCAACGGCCCGGGCCTCAAGGGTGCTCCGCTCCCGACCTGGG
AAGAAGCTTGCGAGGGCGAGGGCGATGAAAAGCTCAGAGCCCTCTTCGACAAGTGGCTCAACAC
CCCGTGGGGCCAGCACCCGTGGGAGGAGGACGGCTGGCCGTTCTGGTCGCTCTTCTACAACATG
AAAACCTGGTGGGACGCCCGCGAGTCGAAGAACATCATCTTCGTCCACTTCTCGGACCTCAAGA
AGGACCTCAAGGGCCAGATGCGCCGCATCGCCAAGTTCCTCAACGCCCCGATCGACGAGTCGAA
GTTTGACGCCCAGGTCACCGCCTGCACCTTCGAATCGATGAAGGGTAATGCCGCTTCGGTCGCCC
CTCTCGGCGGCGCCCTCTGGAAGGGCGGTGCCGAAACCTTCATCAACAAAGGCACTAACGGCCG
CTGGCGCAACGTCCTCACCAAGGAGCAGGTCAAGCAGTACGAGCAGGTCGCCGAGAAGCGCCTC
GGCAAGGATTGCGCCAAGTGGCTCGCCAACGGCGGCGATATGAACGGCCGCGGCTGCGTCATCA
TGTGA

SEQ ID NO:62
Description: Adenylyl-Sulfate Kinase
Alias: cysC, NC_000913.3
Length: 606
Type: DNA
Codon Optimization: *M. extorquens*
>ATGGCCCTCCACGACGAGAACGTCGTCTGGCACTCGCACCCGGTCACCGTCCAGCAGCGCGAA
CTCCATCATGGCCATCGCGGCGTCGTCCTCTGGTTCACCGGCCTCTCGGGTTCGGGTAAATCGAC
CGTCGCCGGCGCCCTCGAAGAGGCCCTCCACAAGCTCGGTGTCTCGACCTACCTCCTCGATGGCG
ATAACGTCCGCCACGGTCTGTGCTCGGATCTCGGCTTCTCGGACGCCGACCGCAAGGAGAACAT
CCGCCGCGTCGGCGAGGTCGCCAACCTCATGGTCGAAGCCGGTCTGGTCGTCCTCACCGCCTTCA
TCTCGCCGCATCGCGCTGAACGCCAAATGGTCCGTGAGCGCGTCGGCGAGGGCCGCTTCATCGA
GGTGTTCGTCGATACCCCGCTCGCCATCTGCGAAGCCCGTGATCCGAAGGGCCTCTACAAGAAG
GCCCGCGCCGGCGAGCTCCGCAACTTCACCGGTATCGACTCGGTCTACGAAGCCCCGGAGTCGG
CCGAGATCCATCTCAACGGCGAGCAGCTCGTCACCAACCTCGTCCAGCAGCTCCTCGACCTCCTC
CGCCAGAACGACATCATCCGCTCGTGA

SEQ ID NO:63
Description: PAPSS1 - Bifunctional 3'-phosphoadenosine 5'-phosphosulfate synthase 1
Alias: E1C8P2
Length: 1875
Type: DNA
Codon Optimization: *M. extorquens*
>ATGGAGCTCCCGGAGTCGCAGTGCAAGAAGGCCAAGCTCTCGAACCGCGTCCCGAACTGGGGC
ATGCAGCGCGCTACCAACGTCACCTACCAGGCCCACCATGTTTCGCGCAACAAGCGTGGCCAGG
TCGTCGGTACTCGCAGTGGTTTCCGCGGTTGCACCGTTTGGCTTACCGGCCTTTCGGGCGCTGGC
AAGACCACCGTCAGTATGGCCCTCGAGGAGTATCTCGTCTGCCACGGCATCCCGTGCTATACCCT
CGACGGCGACAACATCCGCCAGGGACTCAACAAGAATCTCGGCTTCACCCCGGAGGACCGCGAG

GAAAACGTCCGCCGCATCGCCGAGGTCGCTAAGCTCTTCGCCGATGCTGGCCTCGTCTGCATCAC
CAGTTTCATCTCGCCGTACGCTCAGGACCGCAACAATGCCCGCCGCATCCACGAAGGTGCCTCGC
TCCCGTTCTTCGAGGTGTTCGTCGATGCCCCGCTCCATGTCTGCGAACAGCGCGATGTCAAAGGC
CTCTACAAGAAGGCCCGCGCCGGCGAGATCAAGGGTTTCACCGGCATCGACTCGGAGTACGAGA
AGCCTGAGGCCCCGGAGCTCGTCCTTAAGACTGACTCGTGCGACGTCAACGACTGCGTCCAGCA
GGTCGTCGAGCTCCTCCAGGAGCGCGACATTGTCCCGGTCGACGCCTCGTACGAGGTCAAGGAG
CTCTACGTCCCGGAGAACAAGCTCAAGCTCGCCAAGACCGATGCCGAGTCGCTCCTTACCCTCGA
GATCAACAAGGTCGATATGCAGTGGGTCCAGGTCCTCGCCGAGGGCTGGGCCACCCCGCTCTCG
GGTTTCATGCGCGAGCGCGAATACCTCCAGTGCCTTCATTTCGATTGCCTTCTCGATGGCGGCGT
CATCAACCTCTCGGTTCCGATTGTCCTCACCGCTACCCAGGAGGACAAAGAACGTCTCGACGGCT
GCACCGCCATCGCCCTCGTCTACGAGGGCCGTCGTGTCGCCATTCTTCGCAACCCGGAGTTCTAC
GAACACCGTAAGGAAGAGCGCTGCGCCCGTCAGTGGGGCACCACGTGCAAGGATCACCCGTACA
TCAAGATGGTCATGGAGCAGGGCAACTGGCTCGTCGGCGGTGACCTCCAGGTTCTCGATCGCAT
CTACTGGAACGATGGCCTCGACCAGTATCGCCTCACCCCGGCCGAACTCCGCCAGAAGTTCAAG
GAGATGAACGCCGACGCCGTCTTTGCCTTCCAGCTCCGCAACCCGGTCCACAACGGTCATGCCCT
CCTCATGCAAGACACCCACAAGCAGCTCCTCGAGCGCGGTTACCGTCGCCCTGTCCTCCTCCTCC
ATCCTCTCGGCGGCTGGACCAAAGAGGATGACGTCCCGCTTATGTGGCGCATGAAACAGCACGC
CGCCGTCCTCGAGGAAGGCGTCCTCAACCCGGAGACGACCGTCGTTGCCATCTTCCCGTCGCCTA
TGATGTATGCCGGTCCGACCGAGGTTCAGTGGCATTGCCGTTCGCGCATGGTCGCTGGCGCCAAC
TTCTATATCGTCGGCCGTGATCCTGCCGGTATGCCGCATCCGGGCACCGGCAAAGACCTTTACGA
ACCGACTCATGGCGCCAAGGTTCTTACCATGGCCCCGGGCCTCCGTGCCCTCGAGATCGTCCCTT
TCCGCGTCGCCGCCTACAACAAGAAGAAGAAGTCGATGGACTACTACGACTCGGAGCACCATGA
GGACTTCGAGTTCATCTCGGGCACCCATATGCGCAAGCTCGCCCGCGAAGGCCAGAACCCGCCG
GAGGGCTTCATGGCTCCGAAGGCTTGGACCGTCCTCACCGAATACTACAAGTCGCTCGAGAAGG
CCTAG

SEQ ID NO:64
Description: MA_3297
Length: 1251
Type: DNA
Type: DNA
Codon Optimization: *M. extorquens*
>ATGGGCCGCTTCATCCTCAAGTGCCTCAAGTGCGGCCGCGAGTACTCGCAGGAGTACCGCCTGA
CCTGCGAGAACGACGACTCGTTCCTCCGCGCCGAGTACCTCGAGAAGAAGCTCGAGCTCCGCAA
GCAGCCGGGCATCGGCCGCTTCCACTCGTGGCTCCCGGTCCAGGAGGAGCTCACCACCGAGGCC
GGCCCGATCACCTACAAGTCGGAGGCCCTCGCCCGCGAGCTCGGCCTCTCGAACCTCTACATCGG
CTTCTCGGGCTACTGGCCGGAGAAGGGCGCCTTCATCAAGACCTGCTCGTTCAAGGAGCTCGAG
GCCCACCCGACCATGCAGCTCCTCAAGGAGTCGGGCGGCAAGGCCATCGTCCTCGCCTCGGCCG
GCAACACCGGCCGCGCCTTCGCCCACGTCTCGGCCCTCACCGGCACCGACGTCTACATCGTCGTC
CCGGACTCGGGCATCCCGAAGCTCTGGCTCCCGGAGGAGCCGACCGACTCGATCCACCTCATCTC
GATGACCCCGGGCAACGACTACACCGACGCCATCAACCTCGCCGGCCGCATCGCCAAGCTCCCG
GGCATGGTCCCGGAGGGCGGCGCCCGCAACGTCGCCCGCCGCGAGGGCATGGGCACCGTCATGC
TCGACGCCGCCGTCACCATCGGCAAGATGCCGGACCACTACTTCCAGGCCGTCGGCTCGGGCAC
CGGCGGCATCTCGGCCTGGGAGGCCTCGCTCCGCCTCCGCGAGGACGGCCGCTTCGGCTCGAAG
CTCCCGAAGCTCCAGCTCACCCAGAACCTCCCGTTCGTCCCGATGTACAACGCCTGGCAGGAGG
GCCGCCGCGACATCATCCCGGAGATCGACATGAAGGACGCCAAGAAGCGCATCGAGGAAACCT
ACGCCACCGTCCTCACCAACCGCGCCCCGCCGTACTCGGTCACCGGCGGCCTCTACGACGCCCTC
GTCGACACCGACGGCATCATGTACGCCGTCTCGAAGGAGGAGGCCCTCGACGCCAAGGCCCTCT
TCGAGTCGCTCGAGGGCATCGACATCCTCCCGCCGTCGGCCGTCGCCGCCGCCTCGCTCCTCAAG
GCCGTCGAGGCCGGCAACGTCGGCAAGGACGACACCATCCTCCTCAACATCGCCGGCGGCGGCT
TCAAGCGCCTCAAGGAGGACTTCACCCTCTTCCAGATCGAGCCGGAGATCACCGTCTCGAACCC
GGACGTCCCGCTCGAGGAGCTCAAGCTCTGA

SEQ ID NO:65
Description: Cystathionine gamma-lyase (CGL)
Alias: mccB, BSU27250

Length: 1140
Type: DNA
Codon Optimization: *M. extorquens*
>ATGAAGAAGAAAACCCTCATGATCCACGGCGGCATCACCGGCGACGAAAAGACCGGCGCCGTC
TCGGTCCCGATCTATCAGGTGTCGACCTACAAGCAGCCGAAGGCCGGCCAGCATACTGGCTACG
AGTATTCGCGCACCGCCAACCCGACCAGAACCGCCTTAGAGGCCCTCGTCACCGAGCTCGAAAG
TGGCGAAGCCGGCTACGCCTTCTCGTCGGGTATGGCTGCCATCACCGCCGTCATGATGCTCTTCA
ACTCGGGCGACCACGTCGTCCTCACCGACGACGTCTACGGCGGCACCTACCGCGTCATGACCAA
GGTCCTCAACCGCCTCGGCATCGAGTCGACCTTCGTCGACACCTCGTCGCGCGAGGAGGTCGAG
AAGGCCATCCGCCCGAACACCAAGGCCATCTACATCGAGACGCCGACCAACCCGCTCCTCAAGA
TCACCGACCTCACCCTCATGGCCGACATCGCCAAGAAGGCCGGCGTCCTCCTCATCGTCGACAAC
ACCTTCAACACCCCGTACTTCCAGCAGCCGCTTACTCTCGGCGCCGACATCGTCCTCCATTCGGC
CACCAAGTACCTCGGTGGCCATTCGGATGTCGTCGGCGGCCTCGTTGTCACCGCCTCGAAGGAGC
TCGGTGAGGAACTCCACTTCGTCCAGAACTCGACCGGTGGCGTCCTCGGTCCGCAGGATAGTTGG
CTCCTCATGCGCGGCATCAAGACCCTCGGCCTCCGCATGGAGGCCATCGATCAGAACGCCCGTA
AGATCGCCTCGTTCCTCGAGAACCATCCGGCCGTCCAGACCCTCTATTACCCGGGCTCGTCGAAC
CATCCGGGTCATGAACTCGCCAAGACCCAGGGCGCTGGCTTCGGCGGCATGATCTCGTTCGATAT
CGGCTCGGAGGAGCGCGTCGACGCCTTCCTCGGCAACCTCAAGCTCTTCACCATCGCCGAATCGC
TTGGCGCCGTCGAGTCGCTTATCTCGGTTCCGGCCCGCATGACCCACGCCAGTATCCCGCGTGAG
CGTCGCCTTGAACTCGGCATCACCGATGGCCTCATCCGCATCTCGGTCGGCATCGAAGATGCCGA
GGACCTCCTCGAGGACATCGGCCAGGCCCTCGAGAACATCTAA

SEQ ID NO:66
Description: cuyA
Length: 1020
Type: DNA
Codon Optimization: *M. extorquens*
>ATGCATCTCGCCCGTTACCCGCGCCGCTTCATCGCCCATCTTCCGACTCCGCTCGAGAGACTCG
ACCGTCTCACCGCCGAACTCGGTGGCCCGGAAATCTGGATCAAGCGCGACGATTGCACTGGCCT
CTCGACCGGCGGCAACAAGACCCGCAAGCTCGAGTTCCTCATGGCCGAGGCCGAGCTCCAAGGC
GCCGATATGGTCATGACCCAGGGTGCTACCCAGTCGAATCATGCTCGTCAGACCGCCGCCTTCGC
CGCCAAGCTCGGTATGGACTGCCACATCCTCCTCGAGGACCGCACCGGCTCGAACAACGCCAAC
TACAACAACAACGGCAACGTCCTCCTCGACCATCTCCACGGCGCCACCACCGAAAAGCGCCCGG
GCTCGGGCCTCGATATGAACGCCGAAATGGAGAAGGTCGCCGAGAAGTTCCGCGCCGATGGTCG
CAAGGTCTACACCATCCCTGGCGGTGGTTCGAACCCGACCGGCGCCCTCGGTTACGTCAACTGCG
CCTTCGAGATGCTCAACCAGTTCAACGAGCGCGGCCTCAAGGTCGACCACATCGTCCATGCCACC
GGTAGTGCCGGCACCCAAGCCGGCCTCATCACCGGCCTCCAGGCTATGAATGCCCAGATTCCGCT
TCTTGGCATCGGTGTCCGTGCCCCGAAGCCGAAGCAGGAAGAGAACGTCTATAATCTCGCCTGC
GCCACCGCCGAGAAACTTGGCTGCCCGGGCGTCGTCGCTCGCGAGGACGTCGTCGCCAATACCG
ACTATGTCGGTGAGGGCTATGGCATTCCTACCGAGTCGGGCCTCGAAGCCATCCGCATGTTCGCC
GAGCTCGAAGCCATCCTCCTCGACCCGGTCTATTCGGCCAAGGGTGCCGCCGGCTTCATCGACCT
TATCCGCAAGGGCCATTTTAAGAAGGGCGAGCGCGTCGTCTTTCTCCACACCGGCGGCGCCGTCG
CCCTCTTCGGCTACGACAACGCCTTCGACTACTCGGGCCGCTGGGTCGCCTAA

SEQ ID NO:67
Description: ComA
Alias: phosphosulfolactate synthase
Length: 759Type: DNA
Codon Optimization: *M. extorquens*
>ATGAACGCCTTCAAGTTCCTCGACGAGATCGGCCCGGTCAACACCAACACCATGGTCCTCGACA
AGGCCCTCGGCTACAAGACCGTCGAGGACATGCTCACCATCTCGGGCAACTACTTCAACCTCCTC
AAGTACGGCTGGGGCACCTCGATCCTCTACGACGAGGAGATCATCAAGGACAAGAACGAGCTCT
ACCACTCGTACAACATCCGCACCTACACCGGCGGCACCCTCTTCGAGCTCGCCAACAAGCAGAA
CAAGATCGACGAGTACTTCAACGAGATCGATCGCCTCGGCTTCAACGCCGTCGAGATCTCGGAT
GGCTCGACCACCATCGACTCGGACCGCCGCGCCCAGCTCATCAACAAGTCGAAGGAGCTCGGCT
TCTACACCCTCTCGGAGATCGGCAAGAAGAACCCGCAGAAGGACTCGGAGTACACCACCCAGCA
GCGCATCGACCTCATCAACACCGACATCGAGGCCGGCTCGGACATGGTCATCATCGAGGGCCGC
GAGTCGGGCAAGAACATCGGCATCTACGACGACAAGGGCAACGTCAAGAAGGACGACCTCACC

TCGATCTACGAGAACACCCCGAAGGAGAAGGTCCTCTGGGAGGCCCCGCAGAAGAACCAGCAG
GTCGAGCTCATCCTCACCCTCTCGAACGACGTCAACCTCGGCAACATCAACTCGAACGAAATCGT
CTCGCTCGAAACCCTCCGCCGCGGCCTCCGCGGCGACACCCTCGGCAAGCTCTAA

SEQ ID NO:68
Description: ComB1
Length: 342
Type: DNA
Codon Optimization: *M. extorquens*
>ATGAAGATCAACGTCTCGCTCTACAACTCGCGCACCAACGACCTCGCCATCGTCATCGACCTCC
TCCGCGCCTCGACCACCATCTCGGTCGCCCTCAACACCTTCAAGCGCATCGTCCCGATCAACGAC
ATCGACGAGGCCATCAAGCTCAAGGAGAAGCACAACGCCATCCTCGCCGGCGAGATCAAGTCGT
CGGACTTCGACGTCTCGAACTCGCCGGTCCAGATCTCGAACTACGCCGGCGACACCCTCATCCTC
AAGACCACCAACGGCACCAAGGTCCTCGAGAACATCAAGCAGCGCAACTCGGAGGTCAACATCC
TCGTCGGCGCCTCGATCAACGCCAAGACCGTCGCCCAGAAGGCCCTCGATATCGCCGATAACGA
AATCGAACTCGTCATGGCCGGCCGCCATCAGCGCTTCACCATCGAGGACTGCATCGGCGCCGGC
ATCATCATCAACGAGATCGTCAACATCGCCAAGGAGAAGAACATCTACCTCGAGCTCTCGGAGT
CGGCCAAGGCCTCGAAGATCATCTCGAACAACTCGAACATCATCAAGCAGCTCATCAACACCTC
GCACTCGGCCGACAAGCTCCGCTACCTCGGCTTCGGCGAGGACATCGAGATCTGCTCGCTCATCA
ACAAGATCGACACCGTCCCGATCTACAAGAACAACTACATCGTCTCGCTCGACTAA

SEQ ID NO:69
Description: ComC
Alias: Sulfolactate dehydrogenase
Type: DNA
Codon Optimization: *M. extorquens*
>ATGAACATCACCCCGGAGCAGGAGCTCTCGCTCATCATCGACATCCTCACCAAGTTCGACGTCC
CGGAGGACCAGGCCTCGATCATCGCCGAGGTCACCCTCGATGGCGATCTCAAGGGCTTCTCGTC
GCACGGCATCGGCCGTTTCCCGCAGTACATCAAGGGCCTCGAATGCGGCCACATTAAGCCGCAC
ACCGAGATCGTCGTCGAGAAGGAGACGGCCGCCACCGCCCTCATCAACGGCAACCACGGCTTCG
GCCACGTCGTCACCTACCAGGCCATGAAGATGGCCATCGAGAAGGCCAAGGAGGTCGGCATCGG
CCTCGTCGGCATCCACAACTCGAACCACTTCGGCGTCGCCGGCTACTACTCGGACATGGCCCTCA
TGGAGGACATCATTGGTATCGTCACCGCCAACACCGAACCGGCCGTCGCCCCGATTGGCGGCAA
AGAACCGATCCTTGGCACCAACCCGCTCGCCATCGGTATCCCGTCGGGCAGTCATTACCTCTCGG
TCGATATGGCCACCTCGGCCTCGGCCCGCGGTAAGCTCATGGAAGCCAAGCGCCTTGGCGAGCC
GATCCCGGAAAATGTCGCCCTCGATTCGGATGGCAACCCTACCACCGATCCGGCTGAGGCCCTTA
AGGGCTCGATCCTCCCGTTCGGCGCCCACAAGGGCTATGCCCTCTCGCTCATGATCGAAGTCATC
GCCGGTCCGCTTGTCCGCGCCTCGTATGGCAAGGGTGTCACCGGTACGGCCGACCCGGAGGTTCC
GTGCACTAAGGGCGATCTTATCGCTGCCATCGACCCGTCGAAGTTCGTCGACATCGACCAGTTCA
AGGAGGAGGTCGACGACCTCATCTCGGAGCTCAAGTCGACCCCGAACGTCATGATCCCGGGCGA
CTTCGAGGTCCTCAACGTCAAGCGCCACCAGAAGGAGGGCATCGCCCTCGACGAGACGCTCGTC
CAGCAGCTCCGCGAAATCGCCTCGAACGTCGACGTCGACGTCTCGGATATCCTCGGCGACTAA

SEQ ID NO:70
Description: Cystathionine-β-lyase / L-cysteine desulfhydrase
Alias: MetC
Length: 395
Type: Protein
Organism: *E. coli*
>MADKKLDTQLVNAGRSKKYTLGAVNSVIQRASSLVFDSVEAKKHATRNRANGELFYGRRG
TLTHFSLQQAMCELEGGAGCVLFPCGAAAVANSILAFIEQGDHVLMTNTAYEPSQDFCSK
ILSKLGVTTSWFDPLIGADIVKHLQPNTKIVFLESPGSITMEVHDVPAIVAAVRSVVPDA
IIMIDNTWAAGVLFKALDFGIDVSIQAATKYLVGHSDAMIGTAVCNARCWEQLRENAYLM
GQMVDADTAYITSRGLRTLGVRLRQHHESSLKVAEWLAEHPQVARVNHPALPGSKGHEFW
KRDFTGSSGLFSFVLKKKLNNEELANYLDNFSLFSMAYSWGGYESLILANQPEHIAAIRP
QGEIDFSGTLIRLHIGLEDVDDLIADLDAGFARIV

SEQ ID NO:71

Description: Cystathionine-β-lyase / L-cysteine desulfhydrase
Alias: MetC
Length: 1188
Type: DNA
Organism: *E. coli*
>atgGCGGACAAAAAGCTTGATACTCAACTGGTGAATGCAGGACGCAGCAAAAAATACACT
CTCGGCGCGGTAAATAGCGTGATTCAGCGCGCTTCTTCGCTGGTCTTTGACAGTGTAGAA
GCCAAAAAACACGCGACACGTAATCGCGCCAATGGAGAGTTGTTCTATGGACGGCGCGGA
ACGTTAACCCATTTCTCCTTACAACAAGCGATGTGTGAACTGGAAGGTGGCGCAGGCTGC
GTGCTATTTCCCTGCGGGGCGGCAGCGGTTGCTAATTCCATTCTTGCTTTTATCGAACAG
GGCGATCATGTGTTGATGACCAACACCGCCTATGAACCGAGTCAGGATTTCTGTAGCAAA
ATCCTCAGCAAACTGGGCGTAACGACATCATGGTTTGATCCGCTGATTGGTGCCGATATC
GTTAAGCATCTGCAGCCAAACACTAAAATCGTGTTTCTGGAATCGCCAGGCTCCATCACC
ATGGAAGTCCACGACGTTCCGGCGATTGTTGCCGCCGTACGCAGTGTGGTGCCGGATGCC
ATCATTATGATCGACAACACCTGGGCAGCCGGTGTGCTGTTTAAGGCGCTGGATTTTGGC
ATCGATGTTTCTATTCAAGCCGCCACCAAATATCTGGTTGGGCATTCAGATGCGATGATT
GGCACTGCCGTGTGCAATGCCCGTTGCTGGGAGCAGCTACGGGAAAATGCCTATCTGATG
GGCCAGATGGTCGATGCCGATACCGCCTATATAACCAGCCGTGGCCTGCGCACATTAGGT
GTGCGTTTGCGTCAACATCATGAAAGCAGTCTGAAAGTGGCTGAATGGCTGGCAGAACAT
CCGCAAGTTGCGCGAGTTAACCACCCTGCTCTGCCTGGCAGTAAAGGTCACGAATTCTGG
AAACGAGACTTTACAGGCAGCAGCGGGCTATTTTCCTTTGTGCTTAAGAAAAAACTCAAT
AATGAAGAGCTGGCGAACTATCTGGATAACTTCAGTTTATTCAGCATGGCCTACTCGTGG
GGCGGGTATGAATCGTTGATCCTGGCAAATCAACCAGAACATATCGCCGCCATTCGCCCA
CAAGGCGAGATCGATTTTAGCGGGACCTTGATTCGCCTGCATATTGGTCTGGAAGATGTC
GACGATCTGATTGCCGATCTGGACGCCGGTTTTGCGCGAATTGTAtaa

SEQ ID NO:72
Description: Cysteine synthase B with L-cysteine desulfhydrase activity
Alias: CysM
Length: 303
Type: Protein
Organism: *E. coli*
>MSTLEQTIGNTPLVKLQRMGPDNGSEVWLKLEGNNPAGSVKDRAALSMIVEAEKRGEIKP
GDVLIEATSGNTGIALAMIAALKGYRMKLLMPDNMSQERRAAMRAYGAELILVTKEQGME
GARDLALEMANRGEGKLLDQFNNPDNPYAHYTTTGPEIWQQTGGRITHFVSSMGTTGTIT
GVSRFMREQSKPVTIVGLQPEEGSSIPGIRRWPTEYLPGIFNASLVDEVLDIHQRDAENT
MRELAVREGIFCGVSSGGAVAGALRVAKANPDAVVVAIICDRGDRYLSTGVFGEEHFSQG
AGI

SEQ ID NO:73
Description: Cysteine synthase B with L-cysteine desulfhydrase activity
Alias: CysM
Length: 912
Type: DNA
Organism: *E. coli*
>gtgAGTACATTAGAACAAACAATAGGCAATACGCCTCTGGTGAAGTTGCAGCGAATGGGG
CCGGATAACGGCAGTGAAGTGTGGTTAAAACTGGAAGGCAATAACCCGGCAGGTTCGGTG
AAAGATCGTGCGGCACTTTCGATGATCGTCGAGGCGGAAAAGCGCGGGGAAATTAAACCG
GGTGATGTCTTAATCGAAGCCACCAGTGGTAACACCGGCATTGCGCTGGCAATGATTGCC
GCGCTGAAAGGCTATCGCATGAAATTGCTGATGCCCGACAACATGAGCCAGGAACGCCGT
GCGGCGATGCGTGCTTATGGTGCGGAACTGATTCTTGTCACCAAAGAGCAGGGCATGGAA
GGTGCGCGCGATCTGGCGCTGGAGATGGCGAATCGTGGCGAAGGAAAGCTGCTCGATCAG
TTCAATAATCCCGATAACCCTTATGCGCATTACACCACCACTGGGCCGGAAATCTGGCAG
CAAACCGGCGGGCGCATCACTCATTTTGTCTCCAGCATGGGGACGACCGGCACTATCACC
GGCGTCTCACGCTTTATGCGCGAACAATCCAAACCGGTGACCATTGTCGGCCTGCAACCG
GAAGAGGGCAGCAGCATTCCCGGCATTCGCCGCTGGCCTACGGAATATCTGCCGGGGATT
TTCAACGCTTCTCTGGTGGATGAGGTGCTGGATATTCATCAGCGCGATGCGGAAAACACC
ATGCGCGAACTGGCGGTGCGGGAAGGAATATTCTGTGGCGTCAGCTCCGGCGGCGCGGTT

GCCGGAGCACTGCGGGTGGCAAAAGCTAACCCTGACGCGGTGGTGGTGGCGATCATCTGC
GATCGTGGCGATCGCTACCTTTCTACCGGGGTGTTTGGGGAAGAGCATTTTAGCCAGGGG
GCGGGGATTtaa

[256] Although the foregoing invention has been described in some detail by way of illustration and examples for purposes of clarity of understanding, it will be apparent to those skilled in the art that certain changes and modifications may be practiced without departing from the spirit and scope of the invention, which is delineated in the appended claims. Therefore, the description should not be construed as limiting the scope of the invention.

[257] All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entireties for all purposes and to the same extent as if each individual publication, patent, or patent application were specifically and individually indicated to be so incorporated by reference.

CLAIMS

We claim:

1. A non-naturally occurring microorganism that produces taurine and/or at least one taurine precursor selected from cysteate, sulfoacetaldehyde, and hypotaurine, wherein said microorganism comprises one or more exogenous polynucleotide(s), and wherein said microorganism expresses one or more enzyme(s) for production of taurine and/or taurine precursor(s) in the microorganism, selected from:

(a) cysteamine (2-aminoethanethiol) dioxygenase (ADO);

(b) cysteine dioxygenase (CDO), and cysteine sulfinic acid decarboxylase (CSAD) or glutamate decarboxylase (GAD);

(c) 3-mercaptopropionate dioxygenase (MDO; p3MDO), and CSAD or GAD;

(d) L-serine dehydratase; adenylyl-sulfate kinase (APSK) and sulfate adenyltransferase, or 3'-phosphoadenosine 5'-phosphosulfate synthase (PAPSS1); 3'-phosphoadenylyl sulfate:2-aminoacrylate C-sulfotransferase (PAPS-AS); and CSAD or GAD;

(e) cysteate synthase; optionally, L-serine dehydratase; and CSAD or GAD;

(f) L-cysteine dehydrase (CD); optionally, cystathionine gamma-lyase (CGL); sulfate adenyltransferase and APSK, or PAPSS1; PAPS-AS; and CSAD or GAD;

(g) CD; optionally CGL; cysteate synthase, and CSAD or GAD;

(h) cysteate sulfo-lyase (CuyA), and CSAD or GAD;

(i) phosphosulfolactate synthase (ComA), 2-phospho-3-sulfolactate phosphohydrolase (ComB), sulfolactate dehydrogenase (ComC), aspartate aminotransferase (AspAT), and CSAD or GAD;

(j) sulfoacetaldehyde acetyltransferase (Xsc) and taurine-pyruvate aminotransferase (Tpa);

(k) ComA, ComB, ComC, sulfopyruvate decarboxylase (ComDE), and Tpa; and

(l) aspartate aminotransferase, ComDE, and Tpa,

wherein at least one of said enzyme activities is encoded by an exogenous polynucleotide that is expressed in the microorganism.

2. A non-naturally occurring microorganism according to claim 1, wherein the microorganism is from genera selected from *Methylobacterium*, *Methylomonas*, *Methylobacter*. *Methylococcus*, *Methylosinus*, *Methylocyctis*, *Methylomicrobium*, *Methylomonas*, *Methylpophilus*, *Methylobacillus*, *Methylobacterium*, *Hyphomicrobium*, *Xanthobacter*, *Bacillus*, *Paracoccus*, *Nocardia*, *Arthrobacter*, *Rhodopseudomonas*, *Pseudomonas*, *Candida*, *Hansenula, Pichia*, *Torulopsis*, *Rhodotorula*, *Escherichia*, and *Saccharomyces*.

3. A non-naturally occurring microorganism according to claim 2, wherein the microorganism is selected from *Methylobacterium*, *Escherichia*, *Saccharomyces*, and *Bacillus*.

4. A non-naturally occurring microorganism according to claim 1, wherein the microorganism is a methylotrophic bacterium.

5. A non-naturally occurring microorganism according to claim 1, wherein said one or more exogenous polynucleotide(s) is codon optimized for expression in the microorganism.

6. A non-naturally occurring microorganism according to claim 1, wherein said one or more exogenous polynucleotide(s) is operably linked to promoter(s) for expression in the microorganism.

7. A non-naturally occurring microorganism according to claim 1, comprising deletion of one or more genes that encode enzyme(s) that degrade taurine, cysteate, or sulfoacetaldehyde or modification of one or more genes that encode enzyme(s) that degrade taurine, cysteate, or sulfoacetaldehyde such that activity of the one or more enzyme(s) is lower than in the microorganism parent strain from which the non-naturally microorganism is derived.

8. A non-naturally occurring microorganism according to claim 7, wherein said one or more enzyme(s) that degrade taurine, cysteate, or sulfoacetaldehyde comprises taurine dehydrogenase, Tpa, taurine dioxygenase, CuyA, sulfoacetaldehyde acetyltransferase, and/or gamma-glutamyltransferase.

9. A non-naturally occurring microorganism according to claim 1, wherein said microorganism is genetically modified or artificially pre-selected to produce elevated levels of a carotenoid compound relative to the corresponding unmodified or unselected microorganism.

10. A non-naturally occurring microorganism according to claim 9, wherein said carotenoid compound is selected from β-carotene, lycopene, rhodopsin, zeaxanthin, lutein, canthaxanthin, astaxanthin, and sprilloxanthin.

11. A method for producing biomass that comprises taurine, comprising culturing the microorganism according to claim 1 in a culture medium under conditions suitable for growth of the microorganism and expression of said enzyme(s) for production of taurine and/or taurine precursor(s), wherein biomass comprising said taurine and/or taurine precursor(s) is produced in the culture.

12. A feed or nutritional supplement composition comprising biomass produced according to claim 11.

13. A method for producing biomass that comprises taurine and a carotenoid compound, comprising culturing the microorganism according to claim 9 in a culture medium under conditions suitable for growth of the microorganism and expression of said enzyme(s) for production of taurine and/or taurine precursor(s), wherein biomass comprising said taurine or taurine precursor(s) and a carotenoid compound is produced in the culture.

14. A feed or nutritional supplement composition comprising biomass produced according to claim 13.

15. A non-naturally occurring microorganism according to claim 1, wherein said microorganism accumulates intracellular taurine and/or hypotaurine, wherein said taurine and/or hypotaurine aids in the folding of one or more native or heterologous protein(s), thereby increasing enzymatic activity and/or protein yield of said one or more native or heterologous protein(s) in comparison to the parent microorganism from which the non-naturally occurring microorganism is derived.

Figure 1

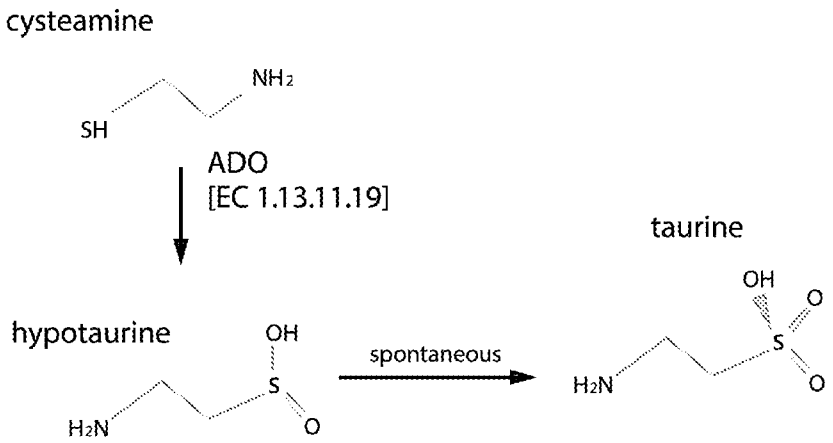

cysteamine
NH2
SH
ADO
[EC 1.13.11.19]
taurine
hypotaurine
OH
S
O
H2N
spontaneous
OH
O
S
O
H2N

Figure 2

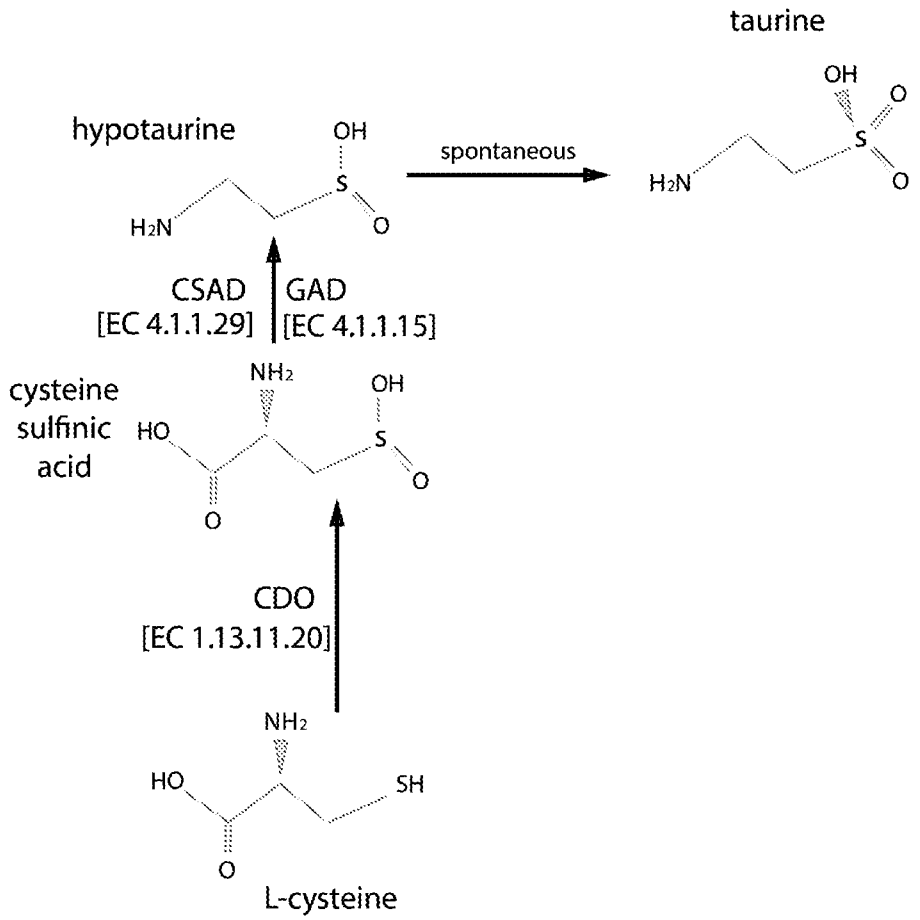

taurine
hypotaurine
OH
O
S
O
H2N
spontaneous
OH
O
S
O
H2N
CSAD
[EC 4.1.1.29]
GAD
[EC 4.1.1.15]
cysteine
sulfinic
acid
NH2
OH
HO
S
O
O
CDO
[EC 1.13.11.20]
NH2
HO
SH
O
L-cysteine

Figure 3

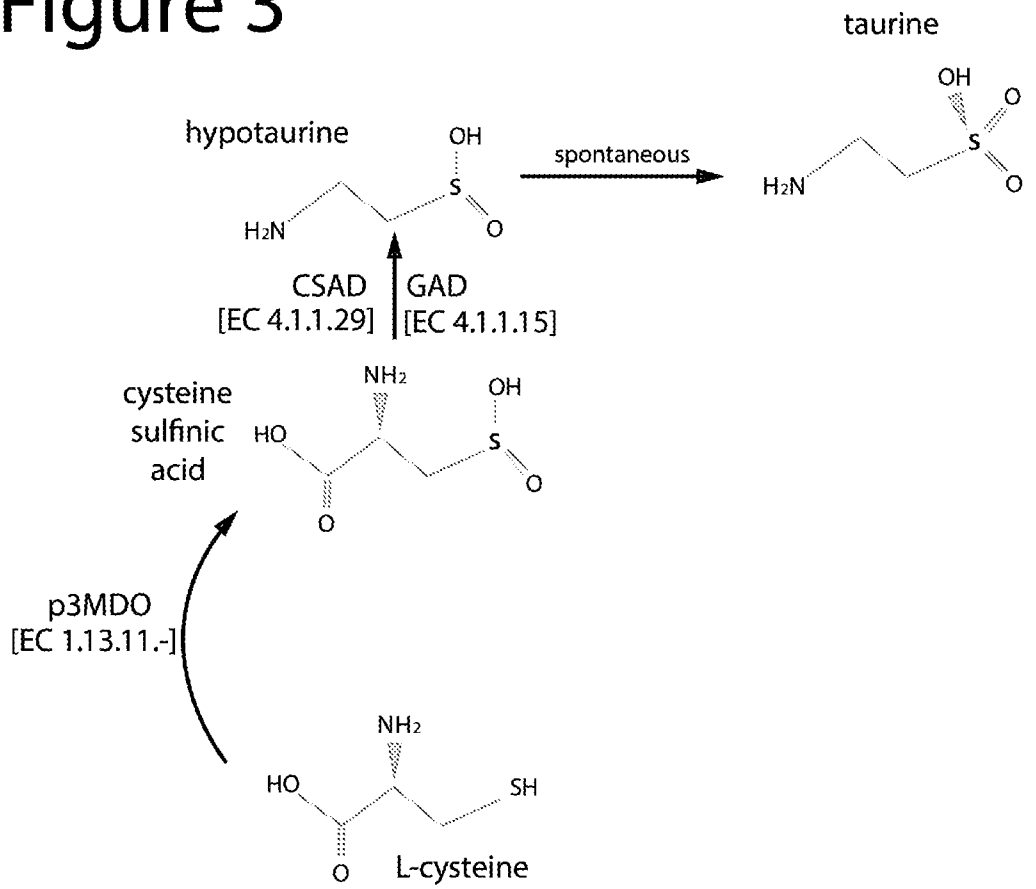

taurine
hypotaurine
OH
O
S
O
H2N
spontaneous
CSAD
[EC 4.1.1.29]
GAD
[EC 4.1.1.15]
cysteine
sulfinic
acid
NH2
OH
HO
S
O
O
p3MDO
[EC 1.13.11.-]
NH2
HO
SH
O
L-cysteine

Figure 4

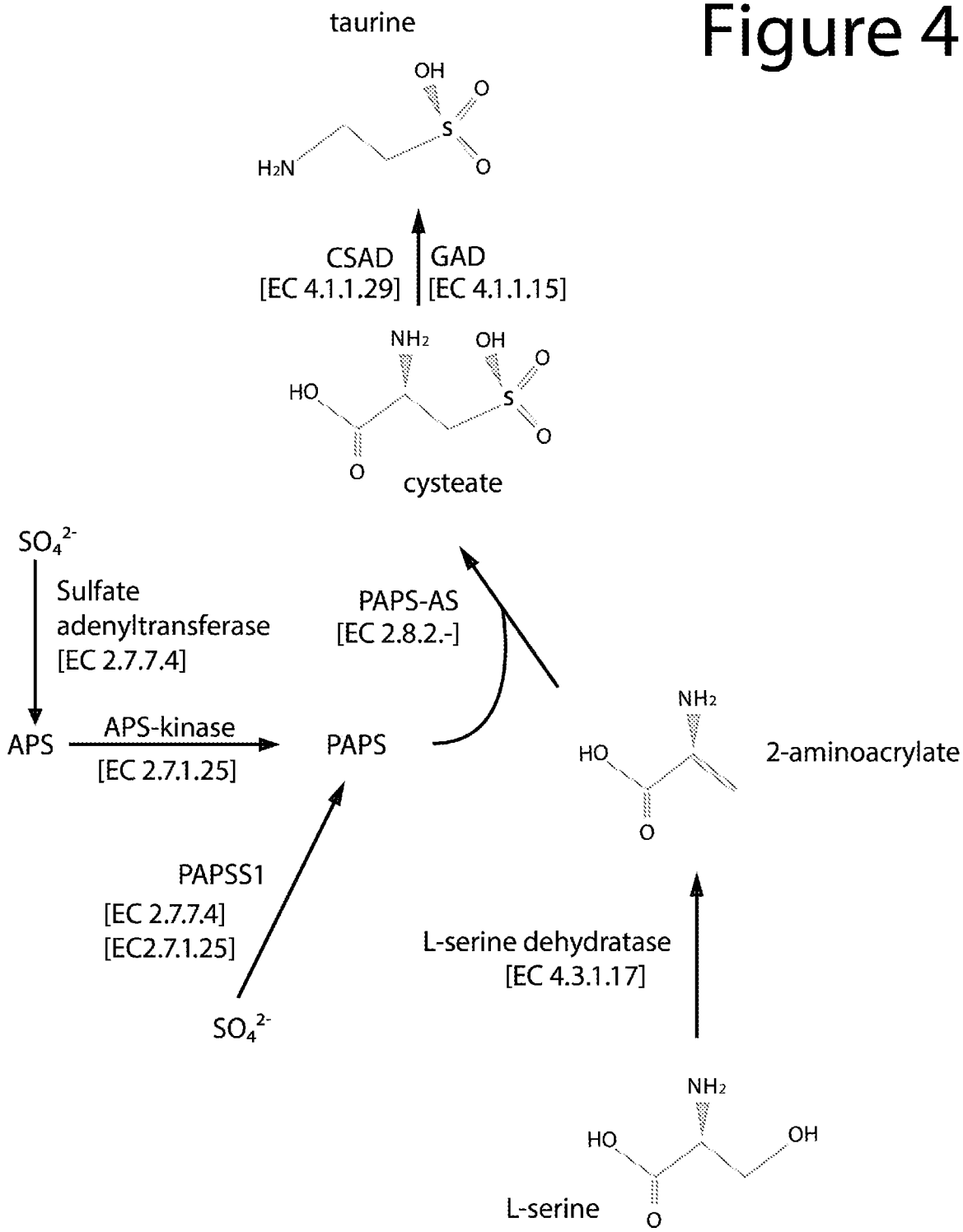

taurine
OH
O
S
O
H2N
CSAD
[EC 4.1.1.29]
GAD
[EC 4.1.1.15]
NH2
OH
HO
S
O
O
O
cysteate
SO4 2-
Sulfate
adenyltransferase
[EC 2.7.7.4]
PAPS-AS
[EC 2.8.2.-]
APS
APS-kinase
[EC 2.7.1.25]
PAPS
NH2
HO
O
2-aminoacrylate
PAPSS1
[EC 2.7.7.4]
[EC2.7.1.25]
L-serine dehydratase
[EC 4.3.1.17]
SO4 2-
NH2
HO
OH
O
L-serine

Figure 5

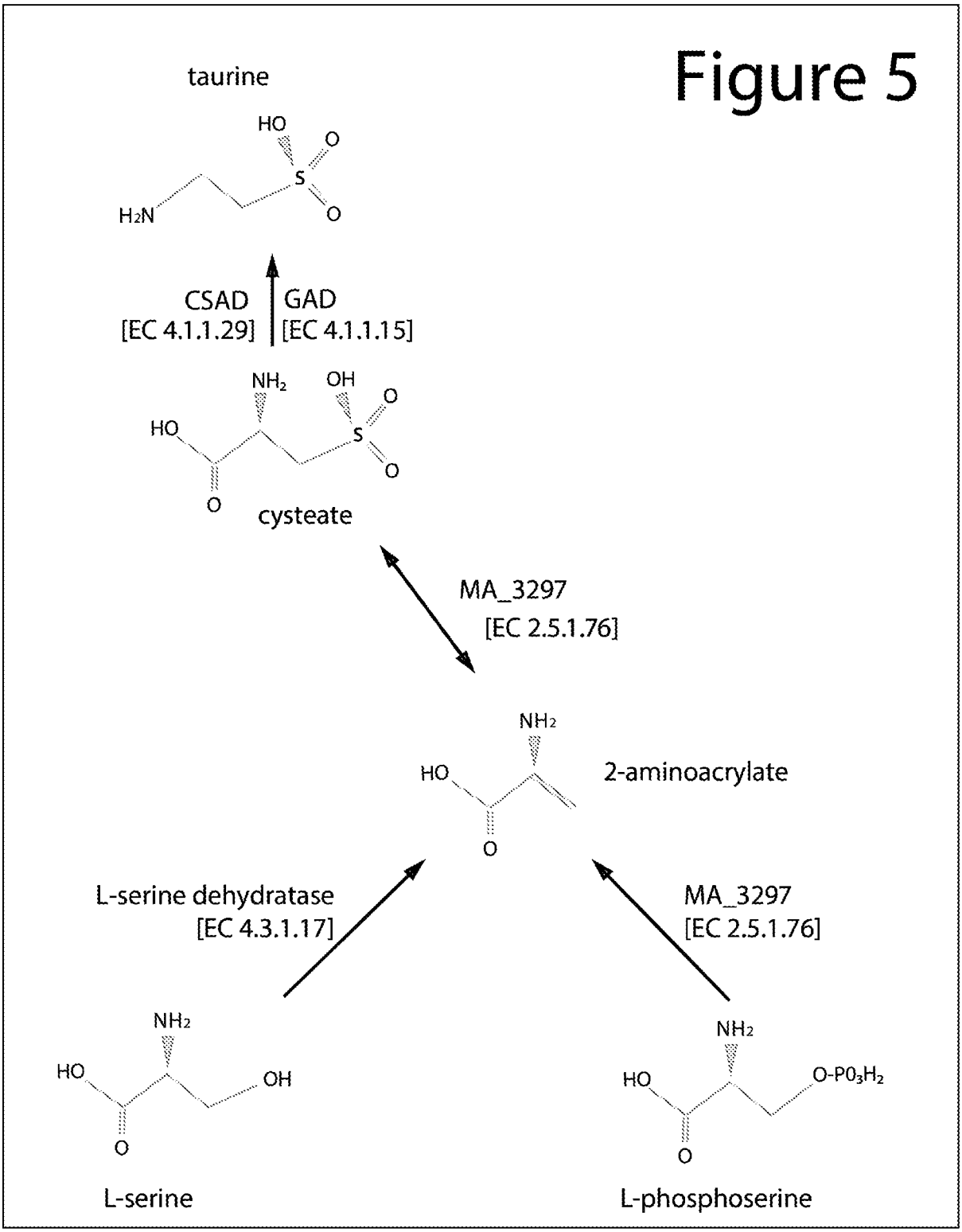

taurine
CSAD
[EC 4.1.1.29]
GAD
[EC 4.1.1.15]
cysteate
MA_3297
[EC 2.5.1.76]
2-aminoacrylate
L-serine dehydratase
[EC 4.3.1.17]
MA_3297
[EC 2.5.1.76]
L-serine
L-phosphoserine

Figure 6

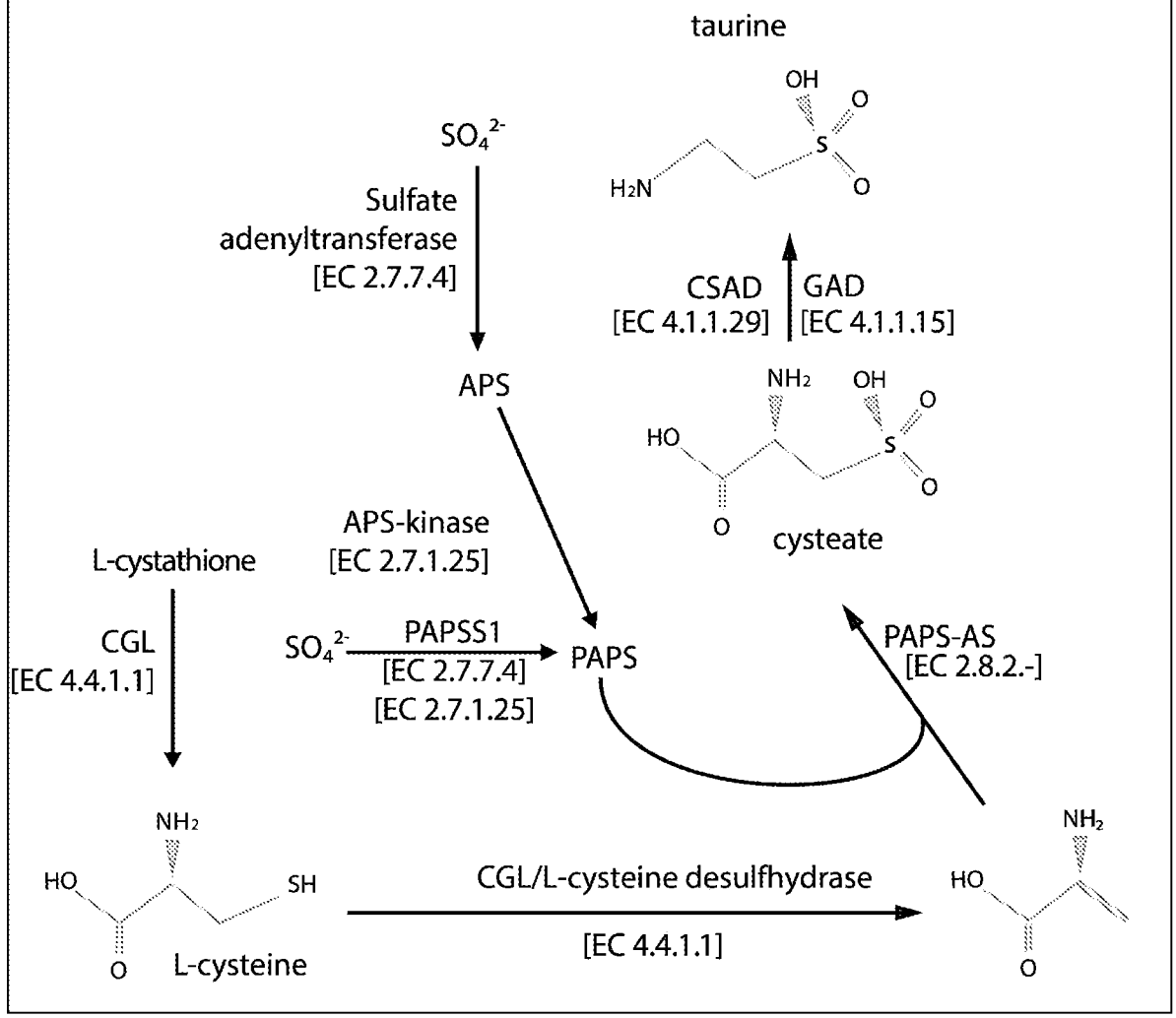

taurine
SO_4^{2-}
Sulfate adenyltransferase [EC 2.7.7.4]
CSAD [EC 4.1.1.29]
GAD [EC 4.1.1.15]
APS
cysteate
APS-kinase [EC 2.7.1.25]
L-cystathione
CGL [EC 4.4.1.1]
SO_4^{2-}
PAPSS1 [EC 2.7.7.4] [EC 2.7.1.25]
PAPS
PAPS-AS [EC 2.8.2.-]
L-cysteine
CGL/L-cysteine desulfhydrase
[EC 4.4.1.1]

Figure 7

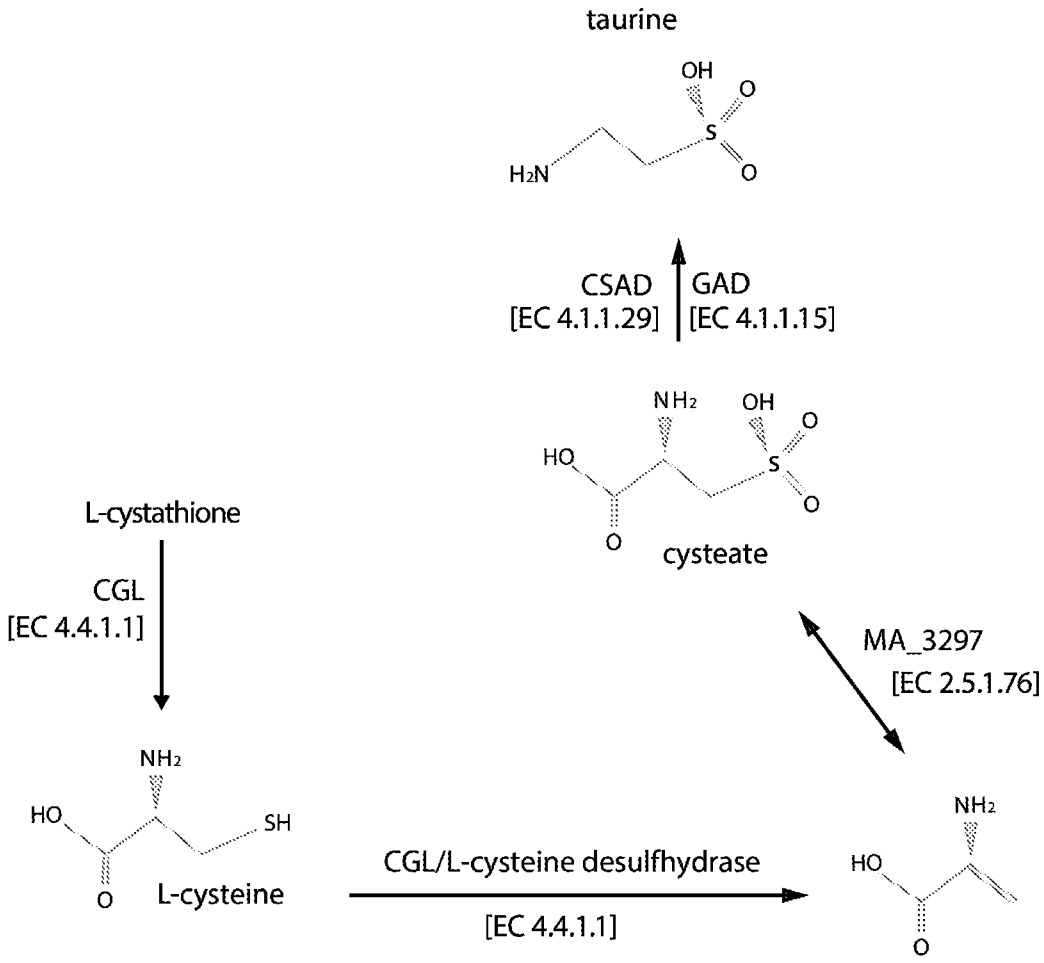

taurine
OH
O
S
O
H2N
CSAD
[EC 4.1.1.29]
GAD
[EC 4.1.1.15]
NH2
OH
O
HO
S
O
O
cysteate
L-cystathione
CGL
[EC 4.4.1.1]
MA_3297
[EC 2.5.1.76]
NH2
HO
SH
O
L-cysteine
CGL/L-cysteine desulfhydrase
[EC 4.4.1.1]
NH2
HO
O

Figure 8

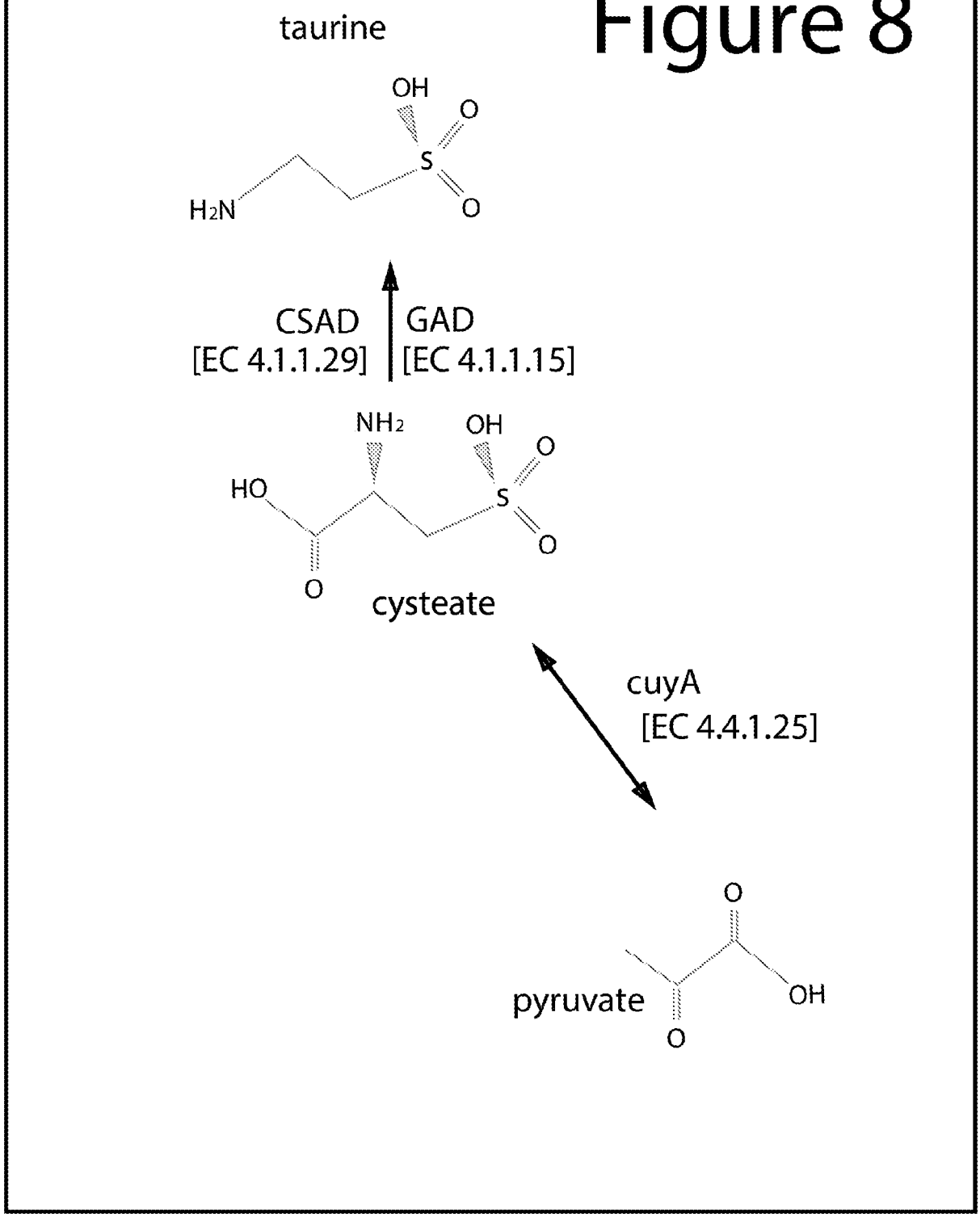

taurine
OH
O
S
O
H2N
CSAD
[EC 4.1.1.29]
GAD
[EC 4.1.1.15]
NH2
OH
O
HO
S
O
O
cysteate
cuyA
[EC 4.4.1.25]
O
OH
pyruvate
O

Figure 9

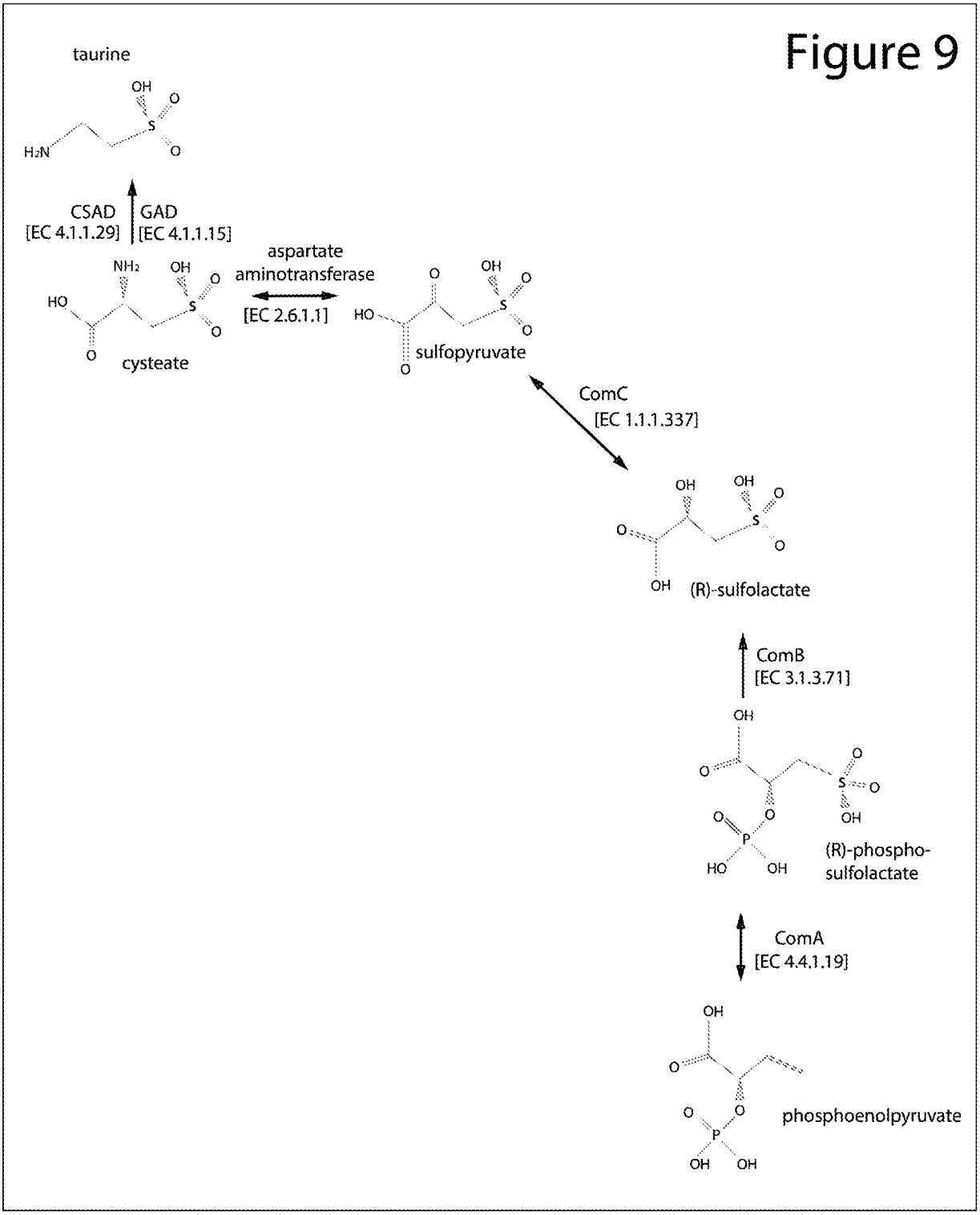
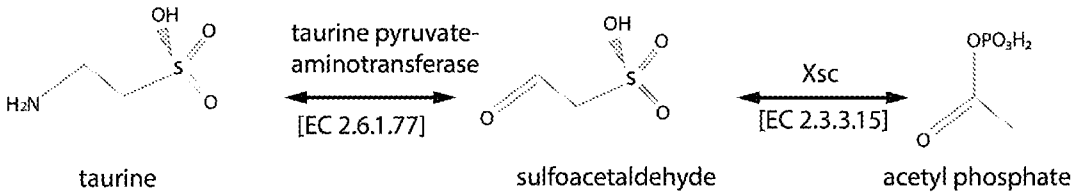

taurine
CSAD
[EC 4.1.1.29]
GAD
[EC 4.1.1.15]
cysteate
aspartate
aminotransferase
[EC 2.6.1.1]
sulfopyruvate
ComC
[EC 1.1.1.337]
(R)-sulfolactate
ComB
[EC 3.1.3.71]
(R)-phospho-
sulfolactate
ComA
[EC 4.4.1.19]
phosphoenolpyruvate

Figure 10



taurine sulfoacetaldehyde acetyl phosphate

Figure 11

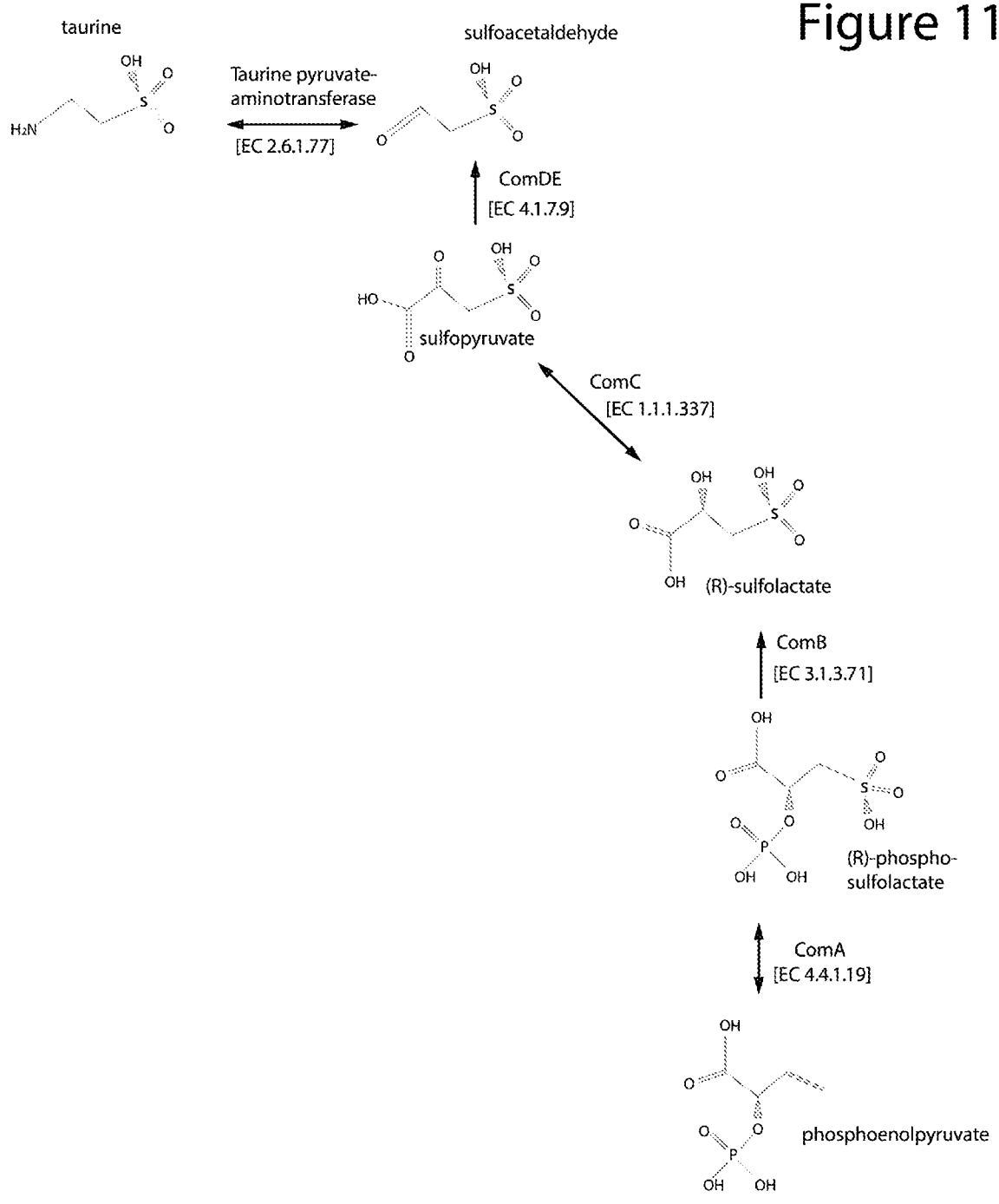

taurine
sulfoacetaldehyde
Taurine pyruvate-aminotransferase
[EC 2.6.1.77]
ComDE
[EC 4.1.7.9]
sulfopyruvate
ComC
[EC 1.1.1.337]
(R)-sulfolactate
ComB
[EC 3.1.3.71]
(R)-phospho-sulfolactate
ComA
[EC 4.4.1.19]
phosphoenolpyruvate

Figure 12

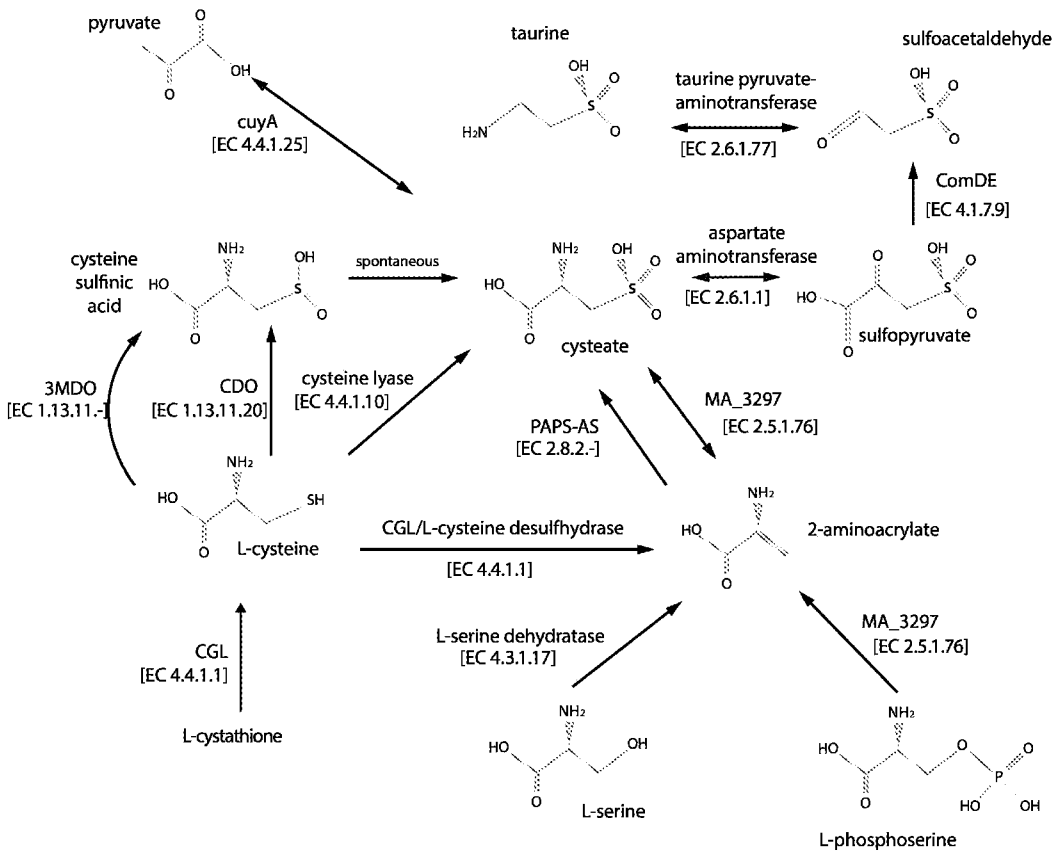
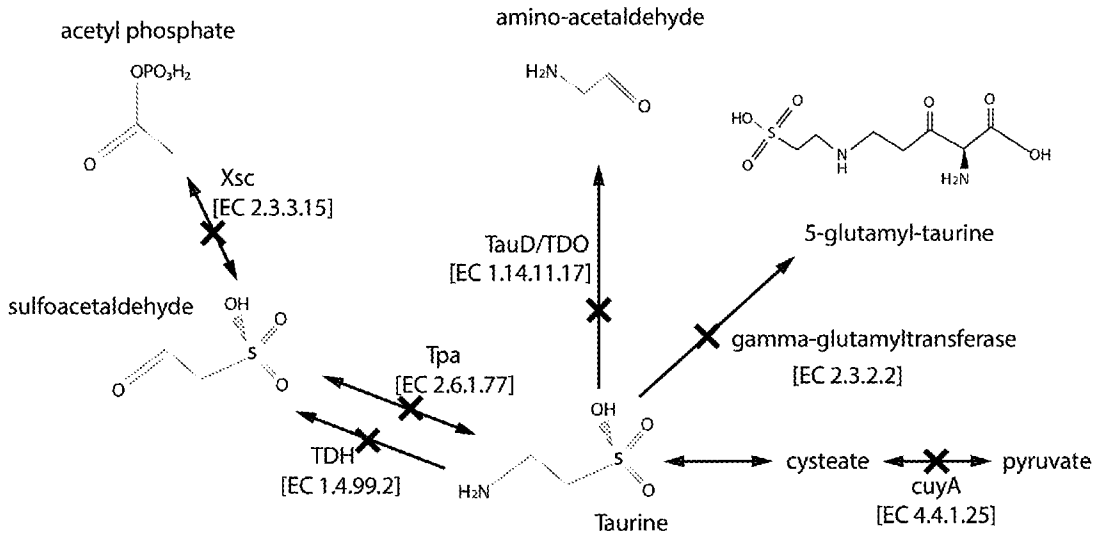

pyruvate
cuyA
[EC 4.4.1.25]
taurine
taurine pyruvate-
aminotransferase
[EC 2.6.1.77]
sulfoacetaldehyde
ComDE
[EC 4.1.7.9]
cysteine
sulfinic
acid
spontaneous
cysteate
aspartate
aminotransferase
[EC 2.6.1.1]
sulfopyruvate
3MDO
[EC 1.13.11.-]
CDO
[EC 1.13.11.20]
cysteine lyase
[EC 4.4.1.10]
PAPS-AS
[EC 2.8.2.-]
MA_3297
[EC 2.5.1.76]
L-cysteine
CGL/L-cysteine desulfhydrase
[EC 4.4.1.1]
2-aminoacrylate
CGL
[EC 4.4.1.1]
L-cystathione
L-serine dehydratase
[EC 4.3.1.17]
MA_3297
[EC 2.5.1.76]
L-serine
L-phosphoserine

Figure 13


acetyl phosphate
amino-acetaldehyde
Xsc
[EC 2.3.3.15]
TauD/TDO
[EC 1.14.11.17]
5-glutamyl-taurine
sulfoacetaldehyde
Tpa
[EC 2.6.1.77]
gamma-glutamyltransferase
[EC 2.3.2.2]
TDH
[EC 1.4.99.2]
cysteate
pyruvate
cuyA
[EC 4.4.1.25]
Taurine

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US16/61081

A. CLASSIFICATION OF SUBJECT MATTER

IPC - C12N 1/20, 1/21,15/75; A23K 10/10; A01N 41/08; C07C 403/24; A23L 33/175, 5/44 (2017.01)
CPC - C12N 1/20, 15/75; A23K 10/10; A01N 41/08; C07C 403/24; A23L 33/175, 5/44

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 2012/0222148 A1 (TURANO, FJ et al.) 30 August, 2012; paragraphs [0013]-[0014], [0025], [0045], [0231], [0272]	1-3, 5-6 --- 4, 7-14
Y	US 2015/0044327 A1 (KNIPBIO) 12 February, 2015; [0018], [0021], [0030], [0053], [0059], [0065]	4, 9-14
Y	US 2012/0107360 A1 (LE BUTT, H et al.) 03 May, 2012; paragraph [0013]	7-8
A	(AGNELLO, G et al.) Discovery of a Substrate Selectivity Motif in Amino Acid Decarboxylases Unveils a Taurine Biosynthesis Pathway in Prokaryotes. ACS Chemical Biology. 23 August, 2013; Vol. 8, No. 10; pages 1-17; abstract; page 2, paragraph 3 – page 3, paragraph 1; DOI: 10.1021/cb400335k.	15
A	(ABE, Y et al.) Role of the Osmolyte Taurine on the Folding of a Model protein, Hen Egg White Lysozyme, Under a Crowding Condition. Amino Acids. 21 January, 2015; Vol. 47, No. 5; pages 1-7; abstract; page 6, paragraph 1; DOI: 10.1007/s00726-015-1918-0.	15

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
02 February 2017 (02.02.2017)	21 FEB 2017
Name and mailing address of the ISA/ Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Shane Thomas PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)