

POLYPEPTIDES AND BIOSYNTHETIC PATHWAYS

CROSS-REFERENCE TO RELATED APPLICATION

This application claims priority to U.S. Patent Application 60/374,831 filed April 23, 2002.

FIELD

This disclosure provides polypeptides and biosynthetic pathways that are useful in the production of indole-3-pyruvate, 2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid (MP) and/or monatin.

BACKGROUND

Indole pyruvate.

Indole-3-pyruvate is a strong antioxidant that is believed to counter act oxidative stress in tissues with high oxygen concentrations (Politi *et al.* "Recent advances in Tryptophan Research", edited by G. A. Filippini *et al.* Plenum Press, New York, 1996, pp 291-8). Indole pyruvate also is an intermediate in a pathway to produce indole-acetic acid (IAA), the primary plant growth hormone auxin (diffusible growth promoting factor). IAA is active in submicrogram amounts in a range of physiological processes including apical dominance, tropisms, shoot elongation, induction of cambial cell division, and root initiation. Synthetic auxins are used in horticulture to induce rooting and to promote the set and development of fruit. At high concentrations the synthetic auxins are effective herbicides against broad-leaved plants. Natural auxins produced by fermentation may be considered more environmentally friendly than chemically produced herbicides. Growth regulators had world sales in 1999 of 0.4 billion pounds (1.4 billion U.S. dollars).

Some examples of patents on indole acetic acid and derivatives thereof include: US Patent No. 5,843,782 Micropropagation of rose plants, auxin used in culture medium and US 5,952,231 Patent No. Micropropagation of rose plants.

In addition to plant related utilities, indole acetic acid is useful in pharmaceutical applications. For example, US Patent No. 5,173,497 "Method of preparing alpha-oxopyrrolo[2,3-B]indole acetic acids and derivatives" proposes the use of these compounds in the treatment of memory impairment such as that associated with Alzheimer's disease and senile dementia. The mechanism proposed in US Patent No. 5,173,497 is that these compounds inhibit the polypeptide acetylcholinesterase and increase acetylcholine levels in the brain.

Indole-3-carbinol is produced from indole-3-acetic acid by peroxidase-catalyzed oxidation, and can easily be converted into diindolylmethane. Both compounds are reported to eliminate toxins and promote the production of hormones beneficial to women's health.

Tryptophan Derivatives

Chlorinated D-tryptophan has been identified as a nonnutritive sweetener, and there is increasing interest in pursuing other derivatives as well. Monatin is a natural sweetener that is similar

in composition to the amino acid tryptophan. It can be extracted from the bark of the roots of the South African shrub, *Sclerochiton ilicifolius*, and has promise in the food and beverage industry as a high-intensity sweetener. Some examples of patents on monatin include: US Patent No. 5994559 Synthesis of monatin-A high intensity natural sweetener, US Patent No. 4975298 3-(1-amino-1,3-dicarboxy-3-hydroxy-but-4-yl)-indole compounds, US Patent No. 5128164 Composition for human consumption containing 3-(1-amino-1,3-dicarboxy-3-hydroxy-but-4-yl)-indole compounds; and US Patent No. 5128482 Process for the production of 3-1(1-amino-1,3-dicarboxy-3-hydroxy-but-4-yl) indole.

Some of the precursors of monatin described here can also be useful as synthetic sweeteners or as intermediates in the synthesis of monatin derivatives.

SUMMARY

The disclosure provides several biosynthetic routes for making monatin from glucose, tryptophan, indole-3-lactic acid, and/or through monatin precursors such as indole-3-pyruvate and 2-hydroxy 2-(indole-3-ylmethyl)-4-keto glutaric acid. Polypeptides and nucleic acid sequences that can be used to make monatin, indole-3-pyruvate, and 2-hydroxy 2-(indole-3-ylmethyl)-4-keto glutaric acid are disclosed.

Monatin can be produced through indole-3-pyruvate, 2-hydroxy 2-(indole-3-ylmethyl)-4-keto glutaric acid (monatin precursor, MP, the alpha-keto form of monatin), indole-3-lactic acid, tryptophan, and/or glucose (FIG. 1). Methods of producing or making monatin or its intermediates shown in FIGS. 1-3 and 11-13 that involve converting a substrate to a first product, and then converting the first product to a second product, and so on, until the desired end product is created, are disclosed.

FIGS. 1-3 and 11-13 show potential intermediate products and end products in boxes. For example, a conversion from one product to another, such as glucose to tryptophan, tryptophan to indole-3-pyruvate, indole-3-pyruvate to MP, MP to monatin, or indole-3-lactic acid (indole-lactate) to indole-3-pyruvate, can be performed by using these methods. These conversions can be facilitated chemically or biologically. The term "convert" refers to the use of either chemical means or polypeptides in a reaction which changes a first intermediate to a second intermediate. The term "chemical conversion" refers to reactions that are not actively facilitated by polypeptides. The term "biological conversion" refers to reactions that are actively facilitated by polypeptides. Conversions can take place *in vivo* or *in vitro*. When biological conversions are used the polypeptides and/or cells can be immobilized on supports such as by chemical attachment on polymer supports. The conversion can be accomplished using any reactor known to one of ordinary skill in the art, for example in a batch or a continuous reactor.

Methods are also provided that include contacting a first polypeptide with a substrate and making a first product, and then contacting the first product created with a second polypeptide and creating a second product, and then contacting the second product created with a third polypeptide

and creating a third product, for example monatin. The polypeptides used and the products produced are shown in FIGS. 1-3 and 11-13.

Polypeptides, and their coding sequences, that can be used to perform the conversions shown in FIGS. 1-3 and 11-13 are disclosed. In some examples, polypeptides having one or more point mutations that allow the substrate specificity and/or activity of the polypeptides to be modified, are used to make monatin.

Isolated and recombinant cells that produce monatin are disclosed. These cells can be any cell, such as a plant, animal, bacterial, yeast, algal, archaeal, or fungal cell.

In a particular example, the disclosed cells include one or more of the following activities, for example two or more or three or more of the following activities: tryptophan aminotransferase (EC 2.6.1.27), tyrosine (aromatic) aminotransferase (EC 2.6.1.5), multiple substrate aminotransferase (EC 2.6.1.-), aspartate aminotransferase (EC 2.6.1.1), tryptophan dehydrogenase (EC 1.4.1.19), tryptophan-phenylpyruvate transaminase (EC 2.6.1.28), L-amino acid oxidase (EC 1.4.3.2), tryptophan oxidase (no EC number, Hadar *et al.*, *J. Bacteriol* 125:1096-1104, 1976 and Furuya *et al.*, *Biosci Biotechnol Biochem* 64:1486-93, 2000), D-amino acid dehydrogenase (EC 1.4.99.1), D-amino acid oxidase (EC 1.4.3.3), D-alanine aminotransferase (EC 2.6.1.21), synthase/lyase (EC 4.1.3.-), such as 4-hydroxy-4-methyl-2-oxoglutarate aldolase (EC 4.1.3.17) or 4-hydroxy-2-oxoglutarate aldolase (EC 4.1.3.16), synthase/lyase (4.1.2.-), D-tryptophan aminotransferase (Kohiba and Mito, Proceedings of the 8th International Symposium on Vitamin B₆ and Carbonyl Catalysis, Osaka, Japan 1990), phenylalanine dehydrogenase (EC 1.4.1.20) and/or glutamate dehydrogenase (EC 1.4.1.2, 1.4.1.3, 1.4.1.4).

In another example, cells include one or more, for example two or more, or three or more, of the following activities: indolelactate dehydrogenase (EC 1.1.1.110), R-4-hydroxyphenyllactate dehydrogenase (EC 1.1.1.222), 3-(4)-hydroxyphenylpyruvate reductase (EC 1.1.1.237), lactate dehydrogenase (EC 1.1.1.27, 1.1.1.28, 1.1.2.3), (3-imidazol-5-yl) lactate dehydrogenase (EC 1.1.1.111), lactate oxidase (EC 1.1.3.-), synthase/lyase (4.1.3.-) such as 4-hydroxy-4-methyl-2-oxoglutarate aldolase (EC 4.1.3.17) or 4-hydroxy-2-oxoglutarate aldolase (EC 4.1.3.16), synthase/lyase (4.1.2.-), tryptophan dehydrogenase (EC 1.4.1.19), tryptophan-phenylpyruvate transaminase (EC 2.6.1.28), tryptophan aminotransferase (EC 2.6.1.27), tyrosine (aromatic) aminotransferase (EC 2.6.1.5), multiple substrate aminotransferase (EC 2.6.1.-), aspartate aminotransferase (EC 2.6.1.1), phenylalanine dehydrogenase (EC 1.4.1.20), glutamate dehydrogenase (EC 1.4.1.2, 1.4.1.3, 1.4.1.4), D-amino acid dehydrogenase (EC 1.4.99.1), D-tryptophan aminotransferase, and/or D-alanine aminotransferase (EC 2.6.1.21).

In addition, the disclosed cells can include one or more of the following activities, for example two or more or three or more of the following activities: tryptophan aminotransferase (EC 2.6.1.27), tyrosine (aromatic) aminotransferase (EC 2.6.1.5), multiple substrate aminotransferase (EC 2.6.1.-), aspartate aminotransferase (EC 2.6.1.1), tryptophan dehydrogenase (EC 1.4.1.19), tryptophan-phenylpyruvate transaminase (EC 2.6.1.28), L-amino acid oxidase (EC 1.4.3.2), tryptophan oxidase (no EC number), D-amino acid dehydrogenase (EC 1.4.99.1), D-amino acid

oxidase (EC 1.4.3.3), D-alanine aminotransferase (EC 2.6.1.21), indolelactate dehydrogenase (EC 1.1.1.110), R-4-hydroxyphenyllactate dehydrogenase (EC 1.1.1.222), 3-(4)-hydroxyphenylpyruvate reductase (EC 1.1.1.237), lactate dehydrogenase (EC 1.1.1.27, 1.1.1.28, 1.1.2.3), (3-imidazol-5-yl) lactate dehydrogenase (EC 1.1.1.111), lactate oxidase (EC 1.1.3.-), synthase/lyase (4.1.3.-) such as 4-
5 hydroxy-4-methyl-2-oxoglutarate aldolase (EC 4.1.3.17) or 4-hydroxy-2-oxoglutarate aldolase (EC 4.1.3.16), synthase/lyase (4.1.2.-), glutamate dehydrogenase (EC 1.4.1.2, 1.4.1.3, 1.4.1.4), phenylalanine dehydrogenase (EC 1.4.1.20), and/or D-tryptophan aminotransferase.

Monatin can be produced by a method that includes contacting tryptophan and/or indole-3-lactic acid with a first polypeptide, wherein the first polypeptide converts tryptophan and/or indole-3-lactic acid to indole-3-pyruvate (either the D or the L form of tryptophan or indole-3-lactic acid can
10 be used as the substrate that is converted to indole-3-pyruvate; one of skill in the art will appreciate that the polypeptides chosen for this step ideally exhibit the appropriate specificity), contacting the resulting indole-3-pyruvate with a second polypeptide, wherein the second polypeptide converts the indole-3-pyruvate to 2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid (MP), and contacting the
15 MP with a third polypeptide, wherein the third polypeptide converts MP to monatin. Exemplary polypeptides that can be used for these conversions are shown in FIGS. 2 and 3.

Another aspect of the invention provides compositions such as MP, cells that contain at least two polypeptides, or sometimes at least three or at least four polypeptides, that are encoded on at least one exogenous nucleic acid sequence.

20 These and other aspects of the disclosure are apparent from the following detailed description and illustrative examples.

BRIEF DESCRIPTION OF THE FIGURES

25 **FIG. 1** shows biosynthetic pathways used to produce monatin and/or indole-3-pyruvate. One pathway produces indole-3-pyruvate via tryptophan, while another produces indole-3-pyruvate via indole-3-lactic acid. Monatin is subsequently produced via a 2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid (MP) intermediate.

30 Compounds shown in boxes are substrates and products produced in the biosynthetic pathways.

Compositions adjacent to the arrows are cofactors, or reactants that can be used during the conversion of a substrate to a product. The cofactor or reactant used will depend upon the polypeptide used for the particular step of the biosynthetic pathway. The cofactor PLP (pyridoxal 5'-phosphate) can catalyze reactions independent of a polypeptide, and therefore, merely providing
35 PLP can allow for the progression from substrate to product.

FIG. 2 is a more detailed diagram of the biosynthetic pathway that utilizes the MP intermediate. The substrates for each step in the pathways are shown in boxes. The polypeptides allowing for the conversion between substrates are listed adjacent to the arrows between the substrates. Each polypeptide is described by its common name and an enzymatic class (EC) number.

FIG. 3 shows a more detailed diagram of the biosynthetic pathway of the conversion of indole-3-lactic acid to indole-3-pyruvate. The substrates are shown in boxes, and the polypeptides allowing for the conversion between the substrates are listed adjacent to the arrow between the substrates. Each polypeptide is described by its common name and an enzymatic class (EC) number.

5 FIG. 4 shows one possible reaction for making MP via chemical means.

FIGS. 5A and 5B are chromatograms showing the LC/MS identification of monatin produced enzymatically.

FIG. 6 is an electrospray mass spectrum of enzymatically synthesized monatin.

10 FIGS. 7A and 7B show chromatograms of the LC/MS/MS daughter ion analyses of monatin produced in an enzymatic mixture.

FIG. 8 is a chromatogram showing the high resolution mass measurement of monatin produced enzymatically.

FIGS. 9A-9C are chromatograms showing the chiral separation of (A) R-tryptophan, (B) S-tryptophan, and (C) monatin produced enzymatically.

15 FIG. 10 is a bar graph showing the relative amount of monatin produced in bacterial cells following IPTG induction. The (-) indicates a lack of substrate addition (no tryptophan or pyruvate was added).

FIGS. 11-12 are schematic diagrams showing pathways used to increase the yield of monatin produced from tryptophan or indole-3-pyruvate.

20 FIG. 13 is a schematic diagram showing a pathway which can be used to increase the yield of monatin produced from tryptophan or indole-3-pyruvate.

SEQUENCE LISTING

25 The nucleic and amino acid sequences listed in the accompanying sequence listing are shown using standard letter abbreviations for nucleotide bases, and three-letter code for amino acids. Only one strand of each nucleic acid sequence is shown, but the complementary strand is understood to be included by any reference to the displayed strand.

30 SEQ ID NOS: 1 and 2 show the nucleic acid and amino acid sequences of an aminotransferase from *Sinorhizobium meliloti*, respectively (*tatA* gene, called a tyrosine or aromatic aminotransferase in literature).

SEQ ID NOS: 3 and 4 show the nucleic acid and amino acid sequences of a tyrosine aminotransferase from *Rhodobacter sphaeroides* (2.4.1), respectively (by homology with *tatA* (SEQ ID NOS: 1 and 2) predicted to be an "aspartate aminotransferase" by genomics software).

35 SEQ ID NOS: 5 and 6 show the nucleic acid and amino acid sequences of an aminotransferase from *Rhodobacter sphaeroides* (35053), respectively (novel, cloned based on 2.4.1 sequence SEQ ID NOS 3 and 4).

SEQ ID NOS: 7 and 8 show the nucleic acid and amino acid sequences of a broad substrate aminotransferase (*bsat*) from *Leishmania major*, respectively.

SEQ ID NOS: 9 and 10 show the nucleic acid and amino acid sequences of an aromatic aminotransferase (*araT*) from *Bacillus subtilis*, respectively.

5 SEQ ID NOS: 11 and 12 show novel nucleic acid and amino acid sequences of an aromatic aminotransferase (*araT*) from *Lactobacillus amylovorus*, respectively (by homology identified as an aromatic aminotransferase).

SEQ ID NOS: 13 and 14 show the nucleic acid and amino acid sequences of a multiple substrate aminotransferase (*msa*) from *R. sphaeroides* (35053), respectively (identified as a multiple substrate aminotransferase by homology to Accession No. AAAE01000093.1, bp 14743-16155 and Accession No. ZP00005082.1).

10 SEQ ID NOS: 15 and 16 show primers used to clone the *B. subtilis* D-alanine aminotransferase (*dat*) sequence.

SEQ ID NOS: 17 and 18 show primers used to clone the *S. meliloti tataA* sequence.

SEQ ID NOS: 19 and 20 show primers used to clone the *B. subtilis araT* aminotransferase sequence.

15 SEQ ID NOS: 21 and 22 show primers used to clone the *Rhodobacter sphaeroides* (2.4.1 and 35053) multiple substrate aminotransferase sequences.

SEQ ID NOS: 23 and 24 show primers used to clone the *Leishmania major bsat* sequence.

SEQ ID NOS: 25 and 26 show primers used to clone the *Lactobacillus amylovorus araT* sequence.

20 SEQ ID NOS: 27 and 28 show primers used to clone the *R. sphaeroides tataA* sequences (both 2.4.1 and 35053).

SEQ ID NOS: 29 and 30 show primers used to clone the *E. coli aspC* sequence (gene sequence Genbank Accession No.: AE000195.1, protein sequence Genbank Accession No.: AAC74014.1).

25 SEQ ID NOS: 31 and 32 show the nucleic acid and amino acid sequences of aromatic aminotransferase (*tyrB*) from *E. coli*, respectively.

SEQ ID NOS: 33 and 34 show primers used to clone the *E. coli tyrB* sequence.

SEQ ID NOS: 35-40 show primers used to clone polypeptides with 4-hydroxy-2-oxoglutarate aldolase (KHG) (EC 4.1.3.16) activity.

30 SEQ ID NOS: 41 and 42 show the nucleic acid sequences of tryptophanase (*tna*) from *E. coli* and tyrosine phenol-lyase (*tpI*) from *Citrobacter freundii*, coding for proteins P00913 (GI:401195) and P31013 (GI:401201), respectively.

SEQ ID NOS: 43 - 46 show primers used to clone tryptophanase polypeptides and β -tyrosinase (tyrosine phenol-lyase) polypeptides.

35 SEQ ID NOS: 47 - 54 show primers used to mutate tryptophanase polypeptides and β -tyrosinase polypeptides.

SEQ ID NOS: 55-64 show primers used to clone polypeptides with 4-hydroxy-4-methyl-2-oxoglutarate aldolase (EC 4.1.3.17) activity.

SEQ ID NOS: 65 and 66 show the nucleic acid and amino acid sequences of 4-hydroxy-4-methyl-2-oxoglutarate aldolase (*proA*) from *C. testosteroni*, respectively.

SEQ ID NOS: 67-68 show primers used to clone *C. testosteroni* 4-hydroxy-4-methyl-2-oxoglutarate aldolase (*proA*) in an operon with *E. coli aspC* in pET30 Xa/LIC.

5 SEQ ID NOS: 69-72 show primers used to clone *E. coli aspC* and *C. testosteroni proA* in pESC-his.

SEQ ID NOS: 73-74 show sequences added to the 5' end of primers used to clone the genes disclosed herein.

10 DETAILED DESCRIPTION OF SEVERAL EMBODIMENTS

Abbreviations and Terms

The following explanations of terms and methods are provided to better describe the present disclosure and to guide those of ordinary skill in the art in the practice of the present disclosure. As used herein, "comprising" means "including" and the singular forms "a" or "an" or "the" include
15 plural references unless the context clearly dictates otherwise. For example, reference to "comprising a protein" includes one or a plurality of such proteins, and reference to "comprising the cell" includes reference to one or more cells and equivalents thereof known to those skilled in the art, and so forth.

Unless explained otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this disclosure belongs.
20 Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present disclosure, suitable methods and materials are described below. The materials, methods, and examples are illustrative only and not intended to be limiting. Other features and advantages of the disclosure are apparent from the following detailed description and the claims.

cDNA (complementary DNA): A piece of DNA lacking internal, non-coding segments
25 (introns) and regulatory sequences which determine transcription. cDNA can be synthesized in the laboratory by reverse transcription from messenger RNA extracted from cells.

Conservative substitution: One or more amino acid substitutions (for example 2, 5, or 10 residues) for amino acid residues having similar biochemical properties. Typically, conservative substitutions have little to no impact on the activity of a resulting polypeptide. For example, ideally,
30 a tryptophan aminotransferase polypeptide including one or more conservative substitutions retains tryptophan aminotransferase activity. A polypeptide can be produced to contain one or more conservative substitutions by manipulating the nucleotide sequence that encodes that polypeptide using, for example, standard procedures such as site-directed mutagenesis or PCR.

Substitutional variants are those in which at least one residue in the amino acid sequence has
35 been removed and a different residue inserted in its place. Examples of amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative substitutions include: Ala substituted with ser or thr; arg substituted with gln, his, or lys; asn substituted with glu, gln, lys, his, asp; asp substituted with asn, glu, or gln; cys substituted with ser or ala; gln substituted with asn, glu, lys, his, asp, or arg; glu substituted with asn, gln, lys, or asp; gly

substituted with pro; his substituted with asn, lys, gln, arg, tyr; ile substituted with leu, met, val, phe; leu substituted with ile, met, val, phe; lys substituted with asn, glu, gln, his, arg; met substituted with ile, leu, val, phe; phe substituted with trp, tyr, met, ile, or leu; ser substituted with thr, ala; thr substituted with ser or ala; trp substituted with phe, tyr; tyr substituted with his, phe, or trp; and val substituted with met, ile, leu.

Further information about conservative substitutions can be found in, among other locations, Ben-Bassat *et al.*, (*J. Bacteriol.* 169:751-7, 1987), O'Regan *et al.*, (*Gene* 77:237-51, 1989), Sahin-Toth *et al.*, (*Protein Sci.* 3:240-7, 1994), Hochuli *et al.*, (*Bio/Technology* 6:1321-5, 1988), WO 00/67796 (Curd *et al.*) and in standard textbooks of genetics and molecular biology.

Exogenous: The term "exogenous" as used herein with reference to nucleic acid and a particular cell refers to any nucleic acid that does not originate from that particular cell as found in nature. Thus, non-naturally-occurring nucleic acid is considered to be exogenous to a cell once introduced into the cell. Nucleic acid that is naturally-occurring also can be exogenous to a particular cell. For example, an entire chromosome isolated from a cell of person X is an exogenous nucleic acid with respect to a cell of person Y once that chromosome is introduced into Y's cell.

Functionally Equivalent: Having an equivalent function. In the context of an enzyme, functionally equivalent molecules include different molecules that retain the function of the enzyme. For example, functional equivalents can be provided by sequence alterations in an enzyme sequence, wherein the peptide with one or more sequence alterations retains a function of the unaltered peptide, such that it retains its enzymatic activity. In a particular example, a tryptophan aminotransferase functional equivalent retains the ability to convert tryptophan to indole-3-pyruvate.

Examples of sequence alterations include, but are not limited to, conservative substitutions, deletions, mutations, frameshifts, and insertions. In one example, a given polypeptide binds an antibody, and a functional equivalent is a polypeptide that binds the same antibody. Thus a functional equivalent includes peptides that have the same binding specificity as a polypeptide, and that can be used as a reagent in place of the polypeptide. In one example a functional equivalent includes a polypeptide wherein the binding sequence is discontinuous, wherein the antibody binds a linear epitope. Thus, if the peptide sequence is MPELANDLGL (amino acids 1-10 of SEQ ID NO: 12) a functional equivalent includes discontinuous epitopes, that can appear as follows (**=any number of intervening amino acids): NH₂-**-M**P**E**L**A**N**D**L**G**L-COOH. In this example, the polypeptide is functionally equivalent to amino acids 1-10 of SEQ ID NO: 12 if the three dimensional structure of the polypeptide is such that it can bind a monoclonal antibody that binds amino acids 1-10 of SEQ ID NO: 12.

Hybridization: The term "hybridization" as used herein refers to a method of testing for complementarity in the nucleotide sequence of two nucleic acid molecules, based on the ability of complementary single-stranded DNA and/or RNA to form a duplex molecule. Nucleic acid hybridization techniques can be used to obtain an isolated nucleic acid within the scope of the disclosure. Briefly, any nucleic acid having some homology to a sequence set forth in SEQ ID NO: 11 can be used as a probe to identify a similar nucleic acid by hybridization under conditions of

moderate to high stringency. Once identified, the nucleic acid then can be purified, sequenced, and analyzed to determine whether it is within the scope of the present disclosure.

Hybridization can be done by Southern or Northern analysis to identify a DNA or RNA sequence, respectively, that hybridizes to a probe. The probe can be labeled with a biotin, digoxigenin, a polypeptide, or a radioisotope such as ^{32}P . The DNA or RNA to be analyzed can be electrophoretically separated on an agarose or polyacrylamide gel, transferred to nitrocellulose, nylon, or other suitable membrane, and hybridized with the probe using standard techniques well known in the art such as those described in sections 7.39-7.52 of Sambrook *et al.*, (1989) Molecular Cloning, second edition, Cold Spring Harbor Laboratory, Plainview, NY. Typically, a probe is at least about 20 nucleotides in length. For example, a probe corresponding to a contiguous 20 nucleotide sequence set forth in SEQ ID NO: 11 can be used to identify an identical or similar nucleic acid. In addition, probes longer or shorter than 20 nucleotides can be used.

The disclosure also provides isolated nucleic acid sequences that are at least about 12 bases in length (e.g., at least about 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, 60, 100, 250, 500, 750, 1000, 1500, 2000, 3000, 4000, or 5000 bases in length) and hybridize, under hybridization conditions, to the sense or antisense strand of a nucleic acid having the sequence set forth in SEQ ID NO: 11. The hybridization conditions can be moderately or highly stringent hybridization conditions.

For the purpose of this disclosure, moderately stringent hybridization conditions mean the hybridization is performed at about 42°C in a hybridization solution containing 25 mM KPO_4 (pH 7.4), 5X SSC, 5X Denhart's solution, 50 µg/mL denatured, sonicated salmon sperm DNA, 50% formamide, 10% Dextran sulfate, and 1-15 ng/mL probe (about 5×10^7 cpm/µg), while the washes are performed at about 50°C with a wash solution containing 2X SSC and 0.1% sodium dodecyl sulfate.

Highly stringent hybridization conditions mean the hybridization is performed at about 42°C in a hybridization solution containing 25 mM KPO_4 (pH 7.4), 5X SSC, 5X Denhart's solution, 50 µg/mL denatured, sonicated salmon sperm DNA, 50% formamide, 10% Dextran sulfate, and 1-15 ng/mL probe (about 5×10^7 cpm/µg), while the washes are performed at about 65°C with a wash solution containing 0.2X SSC and 0.1% sodium dodecyl sulfate.

Isolated: The term "isolated" as used herein with reference to nucleic acid refers to a naturally-occurring nucleic acid that is not immediately contiguous with both of the sequences with which it is immediately contiguous (one on the 5' end and one on the 3' end) in the naturally-occurring genome of the organism from which it is derived. For example, an isolated nucleic acid can be, without limitation, a recombinant DNA molecule of any length, provided one of the nucleic acid sequences normally found immediately flanking that recombinant DNA molecule in a naturally-occurring genome is removed or absent. Thus, an isolated nucleic acid includes, without limitation, a recombinant DNA that exists as a separate molecule (e.g., a cDNA or a genomic DNA fragment produced by PCR or restriction endonuclease treatment) independent of other sequences as well as recombinant DNA that is incorporated into a vector, an autonomously replicating plasmid, a virus (e.g., a retrovirus, adenovirus, or herpes virus), or into the genomic DNA of a prokaryote or eukaryote. In addition, an isolated nucleic acid can include a recombinant DNA molecule that is part

of a hybrid or fusion nucleic acid sequence.

The term "isolated" as used herein with reference to nucleic acid also includes any non-naturally-occurring nucleic acid since non-naturally-occurring nucleic acid sequences are not found in nature and do not have immediately contiguous sequences in a naturally-occurring genome. For example, non-naturally-occurring nucleic acid such as an engineered nucleic acid is considered to be isolated nucleic acid. Engineered nucleic acid can be made using common molecular cloning or chemical nucleic acid synthesis techniques. Isolated non-naturally-occurring nucleic acid can be independent of other sequences, or incorporated into a vector, an autonomously replicating plasmid, a virus (e.g., a retrovirus, adenovirus, or herpes virus), or the genomic DNA of a prokaryote or eukaryote. In addition, a non-naturally-occurring nucleic acid can include a nucleic acid molecule that is part of a hybrid or fusion nucleic acid sequence.

It will be apparent to those of skill in the art that a nucleic acid existing among hundreds to millions of other nucleic acid molecules within, for example, cDNA or genomic libraries, or gel slices containing a genomic DNA restriction digest is not to be considered an isolated nucleic acid.

Nucleic acid: The term "nucleic acid" as used herein encompasses both RNA and DNA including, without limitation, cDNA, genomic DNA, and synthetic (e.g., chemically synthesized) DNA. The nucleic acid can be double-stranded or single-stranded. Where single-stranded, the nucleic acid can be the sense strand or the antisense strand. In addition, nucleic acid can be circular or linear.

Operably linked: A first nucleic acid sequence is "operably linked" with a second nucleic acid sequence whenever the first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription of the coding sequence. Generally, operably linked DNA sequences are contiguous and, where necessary to join two polypeptide-coding regions, in the same reading frame.

Peptide Modifications: The present disclosure includes enzymes, as well as synthetic embodiments thereof. In addition, analogues (non-peptide organic molecules), derivatives (chemically functionalized peptide molecules obtained starting with the disclosed peptide sequences) and variants (homologs) having the desired enzymatic activity can be utilized in the methods described herein. The peptides disclosed herein include a sequence of amino acids, that can be either L- and/or D- amino acids, naturally occurring and otherwise.

Peptides can be modified by a variety of chemical techniques to produce derivatives having essentially the same activity as the unmodified peptides, and optionally having other desirable properties. For example, carboxylic acid groups of the protein, whether carboxyl-terminal or side chain, may be provided in the form of a salt of a pharmaceutically-acceptable cation or esterified to form a C1-C16 ester, or converted to an amide of formula NR₁R₂ wherein R₁ and R₂ are each independently H or C1-C16 alkyl, or combined to form a heterocyclic ring, such as a 5- or 6-membered ring. Amino groups of the peptide, whether amino-terminal or side chain, may be in the form of a pharmaceutically-acceptable acid addition salt such as the HCl, HBr, acetic, benzoic,

toluene sulfonic, maleic, tartaric and other organic salts, or may be modified to C1-C16 alkyl or dialkyl amino or further converted to an amide.

Hydroxyl groups of the peptide side chains may be converted to C1-C16 alkoxy or to a C1-C16 ester using well-recognized techniques. Phenyl and phenolic rings of the peptide side chains may be substituted with one or more halogen atoms, such as F, Cl, Br or I, or with C1-C16 alkyl, C1-C16 alkoxy, carboxylic acids and esters thereof, or amides of such carboxylic acids. Methylene groups of the peptide side chains can be extended to homologous C2-C4 alkylenes. Thiols can be protected with any one of a number of well-recognized protecting groups, such as acetamide groups. Those skilled in the art will also recognize methods for introducing cyclic structures into the peptides of this disclosure to select and provide conformational constraints to the structure that result in enhanced stability. For example, a C- or N-terminal cysteine can be added to the peptide, so that when oxidized the peptide will contain a disulfide bond, generating a cyclic peptide. Other peptide cyclizing methods include the formation of thioethers and carboxyl- and amino-terminal amides and esters.

Peptidomimetic and organomimetic embodiments are also within the scope of the present disclosure, whereby the three-dimensional arrangement of the chemical constituents of such peptido- and organomimetics mimic the three-dimensional arrangement of the peptide backbone and component amino acid side chains, resulting in such peptido- and organomimetics of the proteins of this disclosure having detectable enzyme activity. For computer modeling applications, a pharmacophore is an idealized, three-dimensional definition of the structural requirements for biological activity. Peptido- and organomimetics can be designed to fit each pharmacophore with current computer modeling software (using computer assisted drug design or CADD). See Walters, "Computer-Assisted Modeling of Drugs", in Klegerman & Groves (eds.), *Pharmaceutical Biotechnology*, 1993, Interpharm Press: Buffalo Grove, IL, pp. 165-74 and Ch. 102 in Munson (ed.), *Principles of Pharmacology*, 1995, Chapman & Hall, for descriptions of techniques used in CADD. Also included within the scope of the disclosure are mimetics prepared using such techniques. In one example, a mimetic mimics the enzyme activity generated by an enzyme or a variant, fragment, or fusion thereof.

Probes and primers: Nucleic acid probes and primers can be prepared readily based on the amino acid sequences and nucleic acid sequences provided herein. A "probe" includes an isolated nucleic acid containing a detectable label or reporter molecule. Exemplary labels include, but are not limited to, radioactive isotopes, ligands, chemiluminescent agents, and polypeptides. Methods for labeling and guidance in the choice of labels appropriate for various purposes are discussed in, for example, Sambrook *et al.* (ed.), *Molecular Cloning: A Laboratory Manual* 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989, and Ausubel *et al.* (ed.) *Current Protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience, New York (with periodic updates), 1987.

"Primers" are typically nucleic acid molecules having ten or more nucleotides (e.g., nucleic acid molecules having between about 10 nucleotides and about 100 nucleotides). A primer can be

annealed to a complementary target nucleic acid strand by nucleic acid hybridization to form a hybrid between the primer and the target nucleic acid strand, and then extended along the target nucleic acid strand by, for example, a DNA polymerase polypeptide. Primer pairs can be used for amplification of a nucleic acid sequence, for example, by the polymerase chain reaction (PCR) or other nucleic-acid amplification methods known in the art.

Methods for preparing and using probes and primers are described, for example, in references such as Sambrook *et al.* (ed.), *Molecular Cloning: A Laboratory Manual*, 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989; Ausubel *et al.* (ed.), *Current Protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience, New York (with periodic updates), 1987; and Innis *et al.* (eds.), *PCR Protocols: A Guide to Methods and Applications*, Academic Press: San Diego, 1990. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, © 1991, Whitehead Institute for Biomedical Research, Cambridge, Mass.). One of skill in the art will appreciate that the specificity of a particular probe or primer increases with the length, but that a probe or primer can range in size from a full-length sequence to sequences as short as five consecutive nucleotides. Thus, for example, a primer of 20 consecutive nucleotides can anneal to a target with a higher specificity than a corresponding primer of only 15 nucleotides. Thus, in order to obtain greater specificity, probes and primers can be selected that comprise, for example, 10, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, 2000, 2050, 2100, 2150, 2200, 2250, 2300, 2350, 2400, 2450, 2500, 2550, 2600, 2650, 2700, 2750, 2800, 2850, 2900, 3000, 3050, 3100, 3150, 3200, 3250, 3300, 3350, 3400, 3450, 3500, 3550, 3600, 3650, 3700, 3750, 3800, 3850, 3900, 4000, 4050, 4100, 4150, 4200, 4250, 4300, 4350, 4400, 4450, 4500, 4550, 4600, 4650, 4700, 4750, 4800, 4850, 4900, 5000, 5050, 5100, 5150, 5200, 5250, 5300, 5350, 5400, 5450, or more consecutive nucleotides.

Promoter: An array of nucleic acid control sequences which direct transcription of a nucleic acid. A promoter includes necessary nucleic acid sequences near the start site of transcription, such as, in the case of a polymerase II type promoter, a TATA element. A promoter can include distal enhancer or repressor elements which can be located as much as several thousand base pairs from the start site of transcription.

Purified: The term "purified" as used herein does not require absolute purity; rather, it is intended as a relative term. Thus, for example, a purified polypeptide or nucleic acid preparation can be one in which the subject polypeptide or nucleic acid, respectively, is at a higher concentration than the polypeptide or nucleic acid would be in its natural environment within an organism. For example, a polypeptide preparation can be considered purified if the polypeptide content in the preparation represents at least 50%, 60%, 70%, 80%, 85%, 90%, 92%, 95%, 98%, or 99% of the total protein content of the preparation.

Recombinant: A "recombinant" nucleic acid is one having (1) a sequence that is not naturally occurring in the organism in which it is expressed or (2) a sequence made by an artificial

combination of two otherwise-separated, shorter sequences. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques. "Recombinant" is also used to describe nucleic acid molecules that have been artificially manipulated, but contain the same
5 regulatory sequences and coding regions that are found in the organism from which the nucleic acid was isolated.

Sequence identity: The similarity between amino acid sequences is expressed in terms of the similarity between the sequences, otherwise referred to as sequence identity. Sequence identity is frequently measured in terms of percentage identity (or similarity or homology); the higher the percent
10 the more similar the two sequences are. Homologs or variants of a peptide, such as SEQ ID NO: 12, possess a relatively high degree of sequence identity when aligned using standard methods.

Methods of alignment of sequences for comparison are well known in the art. Various programs and alignment algorithms are described in: Smith and Waterman, *Adv. Appl. Math.* 2:482, 1981; Needleman and Wunsch, *J. Mol. Biol.* 48:443-53, 1970; Pearson and Lipman, *Proc. Natl. Acad. Sci. U.*
15 85:2444-8, 1988; Higgins and Sharp, *Gene* 73:237-44, 1988; Higgins and Sharp, *CABIOS* 5:151-3, 1989; Corpet et al., *Nucleic Acids Research* 16:10881-90, 1988; and Altschul et al., *Nature Genet.* 6:119-29, 1994.

The NCBI Basic Local Alignment Search Tool (BLAST™) (Altschul et al., *J. Mol. Biol.* 215: 10, 1990) is available from several sources, including the National Center for Biotechnology Information (NCBI, Bethesda, MD) and on the Internet, for use in connection with the sequence analysis programs
20 blastp, blastn, blastx, tblastn and tblastx.

Variants of a peptide are typically characterized by possession of at least 50% sequence identity counted over the full length alignment with the amino acid sequence using the NCBI Blast 2.0, gapped blastp set to default parameters. For comparisons of amino acid sequences of greater
25 than about 30 amino acids, the Blast 2 sequences function is employed using the default BLOSUM62 matrix set to default parameters, (gap existence cost of 11, and a per residue gap cost of 1). When aligning short peptides (fewer than around 30 amino acids), the alignment is performed using the Blast 2 sequences function, employing the PAM30 matrix set to default parameters (open gap 9, extension gap 1 penalties). Proteins with even greater similarity to the reference sequences will show
30 increasing percentage identities when assessed by this method, such as at least 80%, at least 90%, at least 95%, at least 98%, or even at least 99% sequence identity. When less than the entire sequence is being compared for sequence identity, homologs and variants will typically possess at least 80% sequence identity over short windows of 10-20 amino acids, and may possess sequence identities of
35 at least 85%, at least 90%, at least 95%, or 98% depending on their similarity to the reference sequence. Methods for determining sequence identity over such short windows are described at the website that is maintained by the National Center for Biotechnology Information in Bethesda, Maryland. One of skill in the art will appreciate that these sequence identity ranges are provided for guidance only; it is entirely possible that strongly significant homologs could be obtained that fall outside of the ranges provided.

Similar methods can be used to determine the percent sequence identity of a nucleic acid sequence. In a particular example, a homologous sequence is aligned to a native sequence, and the number of matches is determined by counting the number of positions where an identical nucleotide or amino acid residue is presented in both sequences. The percent sequence identity is determined by dividing the number of matches either by the length of the sequence set forth in the identified sequence (e.g., SEQ ID NO: 11), or by an articulated length (e.g., 100 consecutive nucleotides or amino acid residues from a sequence set forth in an identified sequence), followed by multiplying the resulting value by 100. For example, a nucleic acid sequence that has 1166 matches when aligned with the sequence set forth in SEQ ID NO: 11 is 75.0 percent identical to the sequence set forth in SEQ ID NO: 11 (i.e., $1166 \div 1554 \times 100 = 75.0$). It is noted that the percent sequence identity value is rounded to the nearest tenth. For example, 75.11, 75.12, 75.13, and 75.14 is rounded down to 75.1, while 75.15, 75.16, 75.17, 75.18, and 75.19 is rounded up to 75.2. It is also noted that the length value will always be an integer. In another example, a target sequence containing a 20-nucleotide region that aligns with 20 consecutive nucleotides from an identified sequence as follows contains a region that shares 75 percent sequence identity to that identified sequence (i.e., $15 \div 20 \times 100 = 75$).

	1	20
Target Sequence:	AGGTCGTGTACTGTCAGTCA	
Identified Sequence:	ACGTGGTGAAGTCCAGTGA	

Specific binding agent: An agent that is capable of specifically binding to any of the polypeptide described herein. Examples include, but are not limited to, polyclonal antibodies, monoclonal antibodies (including humanized monoclonal antibodies), and fragments of monoclonal antibodies such as Fab, F(ab')₂, and Fv fragments as well as any other agent capable of specifically binding to an epitope of such polypeptides.

Antibodies to the polypeptides provided herein (or fragments, variants, or fusions thereof) can be used to purify or identify such polypeptides. The amino acid and nucleic acid sequences provided herein allow for the production of specific antibody-based binding agents that recognize the polypeptides described herein.

Monoclonal or polyclonal antibodies can be produced to the polypeptides, portions of the polypeptides, or variants thereof. Optimally, antibodies raised against one or more epitopes on a polypeptide antigen will specifically detect that polypeptide. That is, antibodies raised against one particular polypeptide would recognize and bind that particular polypeptide, and would not substantially recognize or bind to other polypeptides. The determination that an antibody specifically binds to a particular polypeptide is made by any one of a number of standard immunoassay methods; for instance, Western blotting (See, e.g., Sambrook *et al.* (ed.), *Molecular Cloning: A Laboratory Manual*, 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989).

To determine that a given antibody preparation (such as a preparation produced in a mouse against a polypeptide having the amino acid sequence set forth in SEQ ID NO: 12) specifically detects the appropriate polypeptide (e.g., a polypeptide having the amino acid sequence set forth in

SEQ ID NO: 12) by Western blotting, total cellular protein can be extracted from cells and separated by SDS-polyacrylamide gel electrophoresis.

The separated total cellular protein can then be transferred to a membrane (e.g., nitrocellulose), and the antibody preparation incubated with the membrane. After washing the
5 membrane to remove non-specifically bound antibodies, the presence of specifically bound antibodies can be detected using an appropriate secondary antibody (e.g., an anti-mouse antibody) conjugated to a polypeptide such as alkaline phosphatase since application of 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium results in the production of a densely blue-colored compound by immuno-localized alkaline phosphatase.

10 Substantially pure polypeptides suitable for use as an immunogen can be obtained from transfected cells, transformed cells, or wild-type cells. Polypeptide concentrations in the final preparation can be adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms per milliliter. In addition, polypeptides ranging in size from full-length polypeptides to polypeptides having as few as nine amino acid residues can be utilized as
15 immunogens. Such polypeptides can be produced in cell culture, can be chemically synthesized using standard methods, or can be obtained by cleaving large polypeptides into smaller polypeptides that can be purified. Polypeptides having as few as nine amino acid residues in length can be immunogenic when presented to an immune system in the context of a Major Histocompatibility Complex (MHC) molecule such as an MHC class I or MHC class II molecule. Accordingly,
20 polypeptides having at least 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 900, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, or more consecutive amino acid residues of any amino acid sequence disclosed herein can be used as immunogens for producing antibodies.

Monoclonal antibodies to any of the polypeptides disclosed herein can be prepared from
25 murine hybridomas according to the classic method of Kohler & Milstein (Nature 256:495-7, 1975) or a derivative method thereof.

Polyclonal antiserum containing antibodies to the heterogeneous epitopes of any polypeptide disclosed herein can be prepared by immunizing suitable animals with the polypeptide (or fragment thereof), which can be unmodified or modified to enhance immunogenicity. An effective
30 immunization protocol for rabbits can be found in Vaitukaitis *et al.* (*J. Clin. Endocrinol. Metab.* 33:988-91, 1971).

Antibody fragments can be used in place of whole antibodies and can be readily expressed in prokaryotic host cells. Methods of making and using immunologically effective portions of monoclonal antibodies, also referred to as "antibody fragments," are well known and include those
35 described in Better & Horowitz (Methods Enzymol. 178:476-96, 1989), Glockshuber *et al.* (*Biochemistry* 29:1362-7, 1990), U.S. Pat. No. 5,648,237 ("Expression of Functional Antibody Fragments"), U.S. Pat. No. 4,946,778 ("Single Polypeptide Chain Binding Molecules"), U.S. Pat. No. 5,455,030 ("Immunotherapy Using Single Chain Polypeptide Binding Molecules"), and references cited therein.

Transformed: A "transformed" cell is a cell into which a nucleic acid molecule has been introduced by, for example, molecular biology techniques. Transformation encompasses all techniques by which a nucleic acid molecule can be introduced into such a cell including, without limitation, transfection with a viral vector, conjugation, transformation with a plasmid vector, and
5 introduction of naked DNA by electroporation, lipofection, and particle gun acceleration.

Variants, fragments or fusion proteins: The disclosed proteins, include variants, fragments, and fusions thereof. DNA sequences (for example SEQ ID NO: 11) which encode for a protein (for example SEQ ID NO: 12), fusion protein, or a fragment or variant of a protein, can be engineered to allow the protein to be expressed in eukaryotic cells, bacteria, insects, and/or plants.
10 To obtain expression, the DNA sequence can be altered and operably linked to other regulatory sequences. The final product, which contains the regulatory sequences and the protein, is referred to as a vector. This vector can be introduced into eukaryotic, bacteria, insect, and/or plant cells. Once inside the cell the vector allows the protein to be produced.

A fusion protein including a protein, such as a tryptophan aminotransferase (or variant, polymorphism, mutant, or fragment thereof), for example SEQ ID NO: 12, linked to other amino acid
15 sequences that do not inhibit the desired activity of the protein, for example the ability to convert tryptophan to indole-3-pyruvate. In one example, the other amino acid sequences are no more than about 10, 12, 15, 20, 25, 30, or 50 amino acids in length.

One of ordinary skill in the art will appreciate that a DNA sequence can be altered in
20 numerous ways without affecting the biological activity of the encoded protein. For example, PCR can be used to produce variations in the DNA sequence which encodes an protein. Such variants can be variants optimized for codon preference in a host cell used to express the protein, or other sequence changes that facilitate expression.

Vector: A nucleic acid molecule as introduced into a cell, thereby producing a transformed
25 cell. A vector may include nucleic acid sequences that permit it to replicate in the cell, such as an origin of replication. A vector may also include one or more selectable marker genes and other genetic elements known in the art.

Overview of Biosynthetic Pathways

30 As shown in FIGS. 1-3 and 11-13, many biosynthetic pathways can be used to produce monatin or its intermediates such as indole-3-pyruvate or MP. For the conversion of each substrate (glucose, tryptophan, indole-3-lactic acid, indole-3-pyruvate, and MP) to each product (tryptophan, indole-3-pyruvate, MP and monatin) several different polypeptides can be used. Moreover, these reactions can be carried out *in vivo*, *in vitro*, or through a combination of *in vivo* reactions and *in vitro*
35 reactions, such as *in vitro* reactions that include non-enzymatic chemical reactions. Therefore, FIGS. 1-3 and 11-13 are exemplary, and show multiple different pathways that can be used to obtain desired products.

Glucose to Tryptophan

Many organisms can synthesize tryptophan from glucose. The construct(s) containing the gene(s) necessary to produce monatin, MP, and/or indole-3-pyruvate from glucose and/or tryptophan can be cloned into such organisms. It is shown herein that tryptophan can be converted into monatin.

5 In other examples, an organism is engineered using known polypeptides to produce tryptophan, or overproduce tryptophan. For example, U.S. Patent No. 4,371,614 (herein incorporated by reference) describes an *E. coli* strain transformed with a plasmid containing a wild type tryptophan operon.

Maximum titers of tryptophan disclosed in U.S. Patent No. 4,371,614 are about 230 ppm. Similarly, WO 8701130 (herein incorporated by reference) describes an *E. coli* strain that has been genetically engineered to produce tryptophan and discusses increasing fermentative production of L-tryptophan. Those skilled in the art will recognize that organisms capable of producing tryptophan from glucose are also capable of utilizing other carbon and energy sources that can be converted to glucose or fructose-6-phosphate, with similar results. Exemplary carbon and energy sources include, but are not limited to, sucrose, fructose, starch, cellulose, or glycerol.

Tryptophan to Indole-3-pyruvate

Several polypeptides can be used to convert tryptophan to indole-3-pyruvate. Exemplary polypeptides include members of the enzyme classes (EC) 2.6.1.27, 1.4.1.19, 1.4.99.1, 2.6.1.28, 1.4.3.2, 1.4.3.3, 2.6.1.5, 2.6.1.-, 2.6.1.1, and 2.6.1.21. These classes include polypeptides termed tryptophan aminotransferase (also termed L-phenylalanine-2-oxoglutarate aminotransferase, tryptophan transaminase, 5-hydroxytryptophan-ketoglutaric transaminase, hydroxytryptophan aminotransferase, L-tryptophan aminotransferase, L-tryptophan transaminase, and L-tryptophan:2-oxoglutarate aminotransferase) which converts L-tryptophan and 2-oxoglutarate to indole-3-pyruvate and L-glutamate; D-tryptophan aminotransferase which converts D-tryptophan and a 2-oxo acid to indole-3-pyruvate and an amino acid; tryptophan dehydrogenase (also termed NAD(P)-L-tryptophan dehydrogenase, L-tryptophan dehydrogenase, L-Trip-dehydrogenase, TDH and L-tryptophan:NAD(P) oxidoreductase (deaminating)) which converts L-tryptophan and NAD(P) to indole-3-pyruvate and NH₃ and NAD(P)H; D-amino acid dehydrogenase, which converts D-amino acids and FAD to indole-3-pyruvate and NH₃ and FADH₂; tryptophan-phenylpyruvate transaminase (also termed L-tryptophan- α -ketoisocaproate aminotransferase and L-tryptophan:phenylpyruvate aminotransferase) which converts L-tryptophan and phenylpyruvate to indole-3-pyruvate and L-phenylalanine; L-amino acid oxidase (also termed ophio-amino-acid oxidase and L-amino-acid:oxygen oxidoreductase (deaminating)) which converts an L-amino acid and H₂O and O₂ to a 2-oxo acid and NH₃ and H₂O₂; D-amino acid oxidase (also termed ophio-amino-acid oxidase and D-amino-acid:oxygen oxidoreductase (deaminating)) which converts a D-amino acid and H₂O and O₂ to a 2-oxo acid and NH₃ and H₂O₂; and tryptophan oxidase which converts L-tryptophan and H₂O and O₂ to indole-3-pyruvate and NH₃ and H₂O₂. These classes also contain tyrosine (aromatic) aminotransferase, aspartate aminotransferase, D-amino acid (or D-alanine) aminotransferase, and broad (multiple

substrate) aminotransferase which have multiple aminotransferase activities, some of which can convert tryptophan and a 2-oxo acid to indole-3-pyruvate and an amino acid.

Eleven members of the aminotransferase class that have such activity are described below in Example 1, including a novel aminotransferase shown in SEQ ID NOS: 11 and 12. Therefore, this disclosure provides isolated nucleic acid and protein sequences having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or even at least 99% sequence identity to SEQ ID NOS: 11 and 12. Also encompassed by this disclosure are fragments and fusions of SEQ ID NOS: 11 and 12 that retain aminotransferase activity or encode a protein having aminotransferase activity. Exemplary fragments include, but are not limited to at least 10, 12, 15, 20, 25, 50, 100, 200, 500, or 1000 contiguous nucleotides of SEQ ID NO: 11 or at least 6, 10, 15, 20, 25, 50, 75, 100, 200, 300 or 350 contiguous amino acids of SEQ ID NO: 12. The disclosed sequences (and variants, fragments, and fusions thereof) can be part of a vector. The vector can be used to transform host cells, thereby producing recombinant cells which can produce indole-3-pyruvate from tryptophan, and in some examples can further produce MP and/or monatin.

L-amino acid oxidases (1.4.3.2) are known, and sequences can be isolated from several different sources, such as *Vipera lebetine* (sp P81375), *Ophiophagus hannah* (sp P81383), *Agkistrodon rhodostoma* (sp P81382), *Crotalus atrox* (sp P56742), *Burkholderia cepacia*, *Arabidopsis thaliana*, *Caulobacter crescentus*, *Chlamydomonas reinhardtii*, *Mus musculus*, *Pseudomonas syringae*, and *Rhodococcus* str. In addition, tryptophan oxidases are described in the literature and can be isolated, for example, from *Coprinus* sp. SF-1, Chinese cabbage with club root disease, *Arabidopsis thaliana*, and mammalian liver. One member of the L-amino acid oxidase class that can convert tryptophan to indole-3-pyruvate is discussed below in Example 3, as well as alternative sources for molecular cloning. Many D-amino acid oxidase genes are available in databases for molecular cloning.

Tryptophan dehydrogenases are known, and can be isolated, for example, from spinach, *Pisum sativum*, *Prosopis juliflora*, pea, mesquite, wheat, maize, tomato, tobacco, *Chromobacterium violaceum*, and *Lactobacilli*. Many D-amino acid dehydrogenase gene sequences are known.

As shown in FIGS. 11-13, if an amino acid oxidase, such as tryptophan oxidase, is used to convert tryptophan to indole-3-pyruvate, catalase can be added to reduce or even eliminate the presence of hydrogen peroxide.

Indole-3-lactate to Indole-3-pyruvate

The reaction that converts indole-3-lactate to indole-3-pyruvate can be catalyzed by a variety of polypeptides, such as members of the 1.1.1.110, 1.1.1.27, 1.1.1.28, 1.1.2.3, 1.1.1.222, 1.1.1.237, 1.1.3.-, or 1.1.1.111 classes of polypeptides. The 1.1.1.110 class of polypeptides includes indolelactate dehydrogenases (also termed indolelactic acid: NAD⁺ oxidoreductase). The 1.1.1.27, 1.1.1.28, and 1.1.2.3 classes include lactate dehydrogenases (also termed lactic acid dehydrogenases, lactate: NAD⁺ oxidoreductase). The 1.1.1.222 class contains (R)-4-hydroxyphenyllactate dehydrogenase (also termed D-aromatic lactate dehydrogenase, R-aromatic lactate dehydrogenase).

and R-3-(4-hydroxyphenyl)lactate:NAD(P)⁺ 2-oxidoreductase) and the 1.1.1.237 class contains 3-(4-hydroxyphenyl)pyruvate reductase (also termed hydroxyphenylpyruvate reductase and 4-hydroxyphenyllactate:NAD⁺ oxidoreductase). The 1.1.3.- class contains lactate oxidases, and the 1.1.1.111 class contains (3-imidazol-5-yl) lactate dehydrogenases (also termed (S)-3-(imidazol-5-yl)lactate:NAD(P)⁺ oxidoreductase). It is likely that several of the polypeptides in these classes allow for the production of indole-3-pyruvate from indole-3-lactic acid. Examples of this conversion are provided in Example 2.

Chemical reactions can also be used to convert indole-3-lactic acid to indole-3-pyruvate. Such chemical reactions include an oxidation step that can be accomplished using several methods, for example: air oxidation using a B2 catalyst (China Chemical Reporter, v 13, n 28, p 18 (1), 2002), dilute permanganate and perchlorate, or hydrogen peroxide in the presence of metal catalysts.

Indole-3-pyruvate to 2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid (MP)

Several known polypeptides can be used to convert indole-3-pyruvate to MP. Exemplary polypeptide classes include 4.1.3.-, 4.1.3.16, 4.1.3.17, and 4.1.2.-. These classes include carbon-carbon synthases/lyases, such as aldolases that catalyze the condensation of two carboxylic acid substrates. Peptide class EC 4.1.3.- are synthases/lyases that form carbon-carbon bonds utilizing oxo-acid substrates (such as indole-3-pyruvate) as the electrophile, while EC 4.1.2.- are synthases/lyases that form carbon-carbon bonds utilizing aldehyde substrates (such as benzaldehyde) as the electrophile.

For example, the polypeptide described in EP 1045-029 (EC 4.1.3.16, 4-hydroxy-2-oxoglutarate glyoxylate-lyase also termed 4-hydroxy-2-oxoglutarate aldolase, 2-oxo-4-hydroxyglutarate aldolase or KHG aldolase) converts glyoxylic acid and pyruvate to 4-hydroxy-2-ketoglutaric acid, and the polypeptide 4-hydroxy-4-methyl-2-oxoglutarate aldolase (EC 4.1.3.17, also termed 4-hydroxy-4-methyl-2-oxoglutarate pyruvate-lyase or ProA aldolase), condenses two keto-acids such as two pyruvates to 4-hydroxy-4-methyl-2-oxoglutarate. Reactions utilizing these lyases are described herein.

FIGS. 1-2 and 11-13 show schematic diagrams of these reactions in which a 3-carbon (C3) molecule is combined with indole-3-pyruvate. Many members of EC 4.1.2.- and 4.1.3.-, particularly PLP-utilizing polypeptides, can utilize C3 molecules that are amino acids such as serine, cysteine, and alanine, or derivatives thereof. Aldol condensations catalyzed by representatives of EC 4.1.2.- and 4.1.3.- require the three carbon molecule of this pathway to be pyruvate or a derivative of pyruvate. However, other compounds can serve as a C3 carbon source and be converted to pyruvate. Alanine can be transaminated by many PLP-utilizing transaminases, including many of those mentioned above, to yield pyruvate. Pyruvate and ammonia can be obtained by beta-elimination reactions (such as those catalyzed by tryptophanase or β -tyrosinase) of L-serine, L-cysteine, and derivatives of serine and cysteine with sufficient leaving groups, such as O-methyl-L-serine, O-benzyl-L-serine, S-methylcysteine, S-benzylcysteine, S-alkyl-L-cysteine, O-acyl-L-serine, and 3-chloro-L-alanine. Aspartate can serve as a source of pyruvate in PLP-mediated beta-lyase reactions

such as those catalyzed by tryptophanase (EC 4.1.99.1) and/or β -tyrosinase (EC 4.1.99.2, also termed tyrosine-phenol lyase). The rate of beta-lyase reactions can be increased by performing site-directed mutagenesis on the (4.1.99.1-2) polypeptides as described by Mouratou *et al.* (*J. Biol. Chem* 274:1320-5, 1999) and in Example 8. These modifications allow the polypeptides to accept

5 dicarboxylic amino acid substrates. Lactate can also serve as a source of pyruvate, and is oxidized to pyruvate by the addition of lactate dehydrogenase and an oxidized cofactor or lactate oxidase and oxygen. Examples of these reactions are described below. For example, as shown in FIG. 2 and FIGS. 11-13, ProA aldolase can be contacted with indole-3-pyruvate when pyruvate is used as the C3 molecule.

10 The MP can also be generated using chemical reactions, such as the aldol condensations provided in Example 5.

MP to Monatin

Conversion of MP to monatin can be catalyzed by one or more of: tryptophan

15 aminotransferases (2.6.1.27), tryptophan dehydrogenases (1.4.1.19), D-amino acid dehydrogenases (1.4.99.1), glutamate dehydrogenases (1.4.1.2-4), phenylalanine dehydrogenase (EC 1.4.1.20), tryptophan-phenylpyruvate transaminases (2.6.1.28), or more generally members of the aminotransferase family (2.6.1.-) such as aspartate aminotransferase (EC 2.6.1.1), tyrosine (aromatic) aminotransferase (2.6.1.5), D-tryptophan aminotransferase, or D-alanine (2.6.1.21) aminotransferase

20 (FIG. 2). Eleven members of the aminotransferase class are described below (Example 1), including a novel member of the class shown in SEQ ID NOS: 11 and 12, and reactions demonstrating the activity of aminotransferase and dehydrogenase enzymes are provided in Example 7.

This reaction can also be performed using chemical reactions. Amination of the keto acid (MP) is performed by reductive amination using ammonia and sodium cyanoborohydride.

25 FIGS. 11-13 show additional polypeptides that can be used to convert MP to monatin, as well as providing increased yields of monatin from indole-3-pyruvate or tryptophan. For example, if aspartate is used as the amino donor, aspartate aminotransferase can be used to convert the aspartate to oxaloacetate (FIG. 11). The oxaloacetate is converted to pyruvate and carbon dioxide by a decarboxylase, such as oxaloacetate decarboxylase (FIG. 11). In addition, if lysine is used as the

30 amino donor, lysine epsilon aminotransferase can be used to convert the lysine to allysine (FIG. 12). The allysine is spontaneously converted to 1-piperidine 6-carboxylate (FIG. 12). If a polypeptide capable of catalyzing reductive amination reactions (e.g., glutamate dehydrogenase) is used to convert MP to monatin, a polypeptide that can recycle NAD(P)H and/or produce a volatile product (FIG. 13) can be used, such as formate dehydrogenase.

Additional Considerations in the Design of the Biosynthetic Pathways

Depending on which polypeptides are used to generate indole-3-pyruvate, MP and/or monatin, cofactors, substrates, and/or additional polypeptides can be provided to the production cell to enhance product formation.

Removal of Hydrogen Peroxide

Hydrogen peroxide (H_2O_2) is a product that, if generated, can be toxic to production cells and can damage the polypeptides or intermediates produced. The L-amino acid oxidase described above generates H_2O_2 as a product. Therefore, if L-amino acid oxidase is used, the resulting H_2O_2 can be removed or its levels decreased to decrease potential injury to the cell or product.

Catalases can be used to reduce the level of H_2O_2 in the cell (FIGS. 11-13). The production cell can express a gene or cDNA sequence that encodes a catalase (EC 1.11.1.6), which catalyzes the decomposition of hydrogen peroxide into water and oxygen gas. For example, a catalase can be expressed from a vector transfected into the production cell. Examples of catalases that can be used to include, but are not limited to: tr|Q9EV50 (*Staphylococcus xylosus*), tr|Q9KBE8 (*Bacillus halodurans*), tr|Q9URJ7 (*Candida albicans*), tr|P77948 (*Streptomyces coelicolor*), tr|Q9RBJ5 (*Xanthomonas campestris*) (SwissProt Accession Nos.). Biocatalytic reactors utilizing L-amino acid oxidase, D-amino acid oxidase, or tryptophan oxidase can also contain a catalase polypeptide.

Modulation of PLP (pyridoxal-5'-phosphate) Availability

As shown in FIG. 1, PLP can be utilized in one or more of the biosynthetic steps described herein. The concentration of PLP can be supplemented so that PLP does not become a limitation on the overall efficiency of the reaction.

The biosynthetic pathway for vitamin B₆ (the precursor of PLP) has been thoroughly studied in *E. coli* and some of the proteins have been crystallized (Laber *et al.*, *FEBS Letters*, 449:45-8, 1999). Two of the genes (*epd* or *gapB* and *serC*) are required in other metabolic pathways, while three genes (*pdxA*, *pdxB*, and *pdxJ*) are unique to pyridoxal phosphate biosynthesis. One of the starting materials in the *E. coli* pathway is 1-deoxy-D-xylulose-5-phosphate (DXP). Synthesis of this precursor from common 2 and 3 carbon central metabolites is catalyzed by the polypeptide 1-deoxy-D-xylulose-5-phosphate synthase (DSX). The other precursor is a threonine derivative formed from the 4-carbon sugar, D-erythrose 4-phosphate. The genes required for the conversion to phospho-4-hydroxyl-L threonine (HTP) are *epd*, *pdxB*, and *serC*. The last reaction for the formation of PLP is a complex intramolecular condensation and ring-closure reaction between DXP and HTP, catalyzed by the gene products of *pdxA* and *pdxJ*.

If PLP becomes a limiting nutrient during the fermentation to produce monatin, increased expression of one or more of the pathway genes in a production host cell can be used to increase the yield of monatin. A host organism can contain multiple copies of its native pathway genes or copies of non-native pathway genes can be incorporated into the organism's genome. Additionally, multiple copies of the salvage pathway genes can be cloned into the host organism.

One salvage pathway that is conserved in all organisms recycles the various derivatives of vitamin B₆ to the active PLP form. The polypeptides involved in this pathway are *pdxK* kinase, *pdxH* oxidase, and *pdxY* kinase. Over-expression of one or more of these genes can increase PLP availability.

-22-

Vitamin B₆ levels can be elevated by elimination or repression of the metabolic regulation of the native biosynthetic pathway genes in the host organism. PLP represses polypeptides involved in the biosynthesis of the precursor threonine derivative in the bacterium *Flavobacterium sp.* strain 238-7. This bacterial strain, freed of metabolic control, overproduces pyridoxal derivatives and can excrete up to 20 mg/L of PLP. Genetic manipulation of the host organism producing monatin in a similar fashion will allow the increased production PLP without over-expression of the biosynthetic pathway genes.

Ammonium Utilization

Tryptophanase reactions can be driven toward the synthetic direction (production of tryptophan from indole) by making ammonia more available or by removal of water. Reductive amination reactions, such as those catalyzed by glutamate dehydrogenase, can also be driven forward by an excess of ammonium.

Ammonia can be made available as an ammonium carbonate or ammonium phosphate salt in a carbonate or phosphate buffered system. Ammonia can also be provided as ammonium pyruvate or ammonium formate. Alternatively, ammonia can be supplied if the reaction is coupled with a reaction that generates ammonia, such as glutamate dehydrogenase or tryptophan dehydrogenase. Ammonia can be generated by addition of the natural substrates of EC 4.1.99.- (tyrosine or tryptophan), which will be hydrolyzed to phenol or indole, pyruvate and NH₃. This also allows for an increased yield of synthetic product over the normal equilibrium amount by allowing the enzyme to hydrolyze its preferred substrate.

Removal of products and byproducts

The conversion of tryptophan to indole-3-pyruvate via a tryptophan aminotransferase may adversely affect the production rate of indole-3-pyruvate because the reaction produces glutamate and requires the co-substrate 2-oxoglutarate (α -ketoglutarate). Glutamate may cause inhibition of the aminotransferase, and the reaction will consume large amounts of the co-substrate. Moreover, high glutamate concentrations are detrimental to downstream separation processes.

The polypeptide glutamate dehydrogenase (GLDH) converts glutamate to 2-oxoglutarate, thereby recycling the co-substrate in the reaction catalyzed by tryptophan aminotransferase. GLDH also generates reducing equivalents (NADH or NADPH) that can be used to generate energy for the cell (ATP) under aerobic conditions. The utilization of glutamate by GLDH also reduces byproduct formation. Additionally, the reaction generates ammonia, which can serve as a nitrogen source for the cell or as a substrate in a reductive amination for the final step shown in FIG. 1. Therefore, a production cell that over-expresses a GLDH polypeptide can be used to increase the yield and reduce the cost of media and/or separation processes.

In the tryptophan to monatin pathway, the amino donor of step three (e.g., glutamate or aspartate) can be converted back to the amino acceptor required for step 1 (e.g., 2-oxo-glutarate or oxaloacetate) if an aminotransferase from the appropriate enzyme classes is used. Utilization of two

separate transaminases for this pathway, in which the substrate of one transaminase does not competitively inhibit the activity of the other transaminase, can increase the efficiency of this pathway.

Many of the reactions in the described pathways are reversible and will, therefore, reach an equilibrium between substrates and products. The yield of the pathway can be increased by continuous removal of the products from the polypeptides. For example, secretion of monatin into the fermentation broth using a permease or other transport protein, or selective crystallization of monatin from a biocatalytic reactor stream with concomitant recycle of substrates will increase the reaction yield.

The removal of byproducts by additional enzymatic reactions or by substitution of amino donor groups is another way to increase the reaction yield. Several examples are discussed in Example 13 and shown in FIGS. 11-13. Ideally a byproduct is produced that is unavailable to react in the reverse direction, either by phase change (evaporation) or by spontaneous conversion to an unreactive endproduct, such as carbon dioxide.

Modulation of the Substrate Pools

The indole pool can be modulated by increasing production of tryptophan precursors and/or altering catabolic pathways involving indole-3-pyruvate and/or tryptophan. For example, the production of indole-3-acetic acid from indole-3-pyruvate can be reduced or eliminated by functionally deleting the gene coding for EC 4.1.1.74 in the host cell. Production of indole from tryptophan can be reduced or eliminated by functionally deleting the gene coding for EC 4.1.99.1 in the host cell. Alternatively, an excess of indole can be utilized as a substrate in an *in vitro* or *in vivo* process in combination with increased amounts of the gene coding for EC 4.1.99.1 (Kawasaki *et al.*, *J. Ferm. and Bioeng.*, 82:604-6, 1996). Genetic modifications can be made to increase the level of intermediates such as D-erythrose-4-phosphate and chorismate.

Tryptophan production is regulated in most organisms. One mechanism is via feedback inhibition of certain enzymes in the pathway; as tryptophan levels increase, the production rate of tryptophan decreases. Thus, when using a host cell engineered to produce monatin via a tryptophan intermediate, an organism can be used that is not sensitive to tryptophan concentrations. For example, a strain of *Catharanthus roseus* that is resistant to growth inhibition by various tryptophan analogs was selected by repeated exposure to high concentrations of 5-methyltryptophan (Schallenberg and Berlin, *Z Naturforsch* 34:541-5, 1979). The resulting tryptophan synthase activity of the strain was less effected by product inhibition, likely due to mutations in the gene. Similarly, a host cell used for monatin production can be optimized.

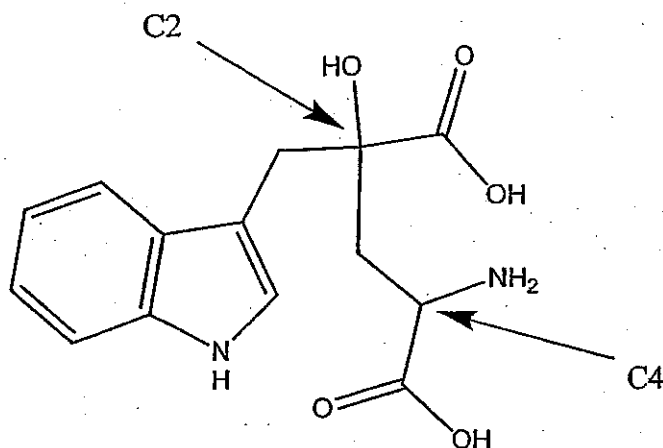
Tryptophan production can be optimized through the use of directed evolution to evolve polypeptides that are less sensitive to product inhibition. For example, screening can be performed on plates containing no tryptophan in the medium, but with high levels of non-metabolizable tryptophan analogs. U.S. Patent Nos. 5,756,345; 4,742,007; and 4,371,614 describe methods used to increase tryptophan productivity in a fermentation organism. The last step of tryptophan biosynthesis

is the addition of serine to indole; therefore the availability of serine can be increased to increase tryptophan production.

The amount of monatin produced by a fermentation organism can be increased by increasing the amount of pyruvate produced by the host organism. Certain yeasts, such as *Trichosporon cutaneum* (Wang *et al.*, *Lett. Appl. Microbiol.* 35:338-42, 2002) and *Torulopsis glabrata* (Li *et al.*, *Appl Microbiol. Biotechnol.* 57:451-9, 2001) overproduce pyruvate and can be used to practice the methods disclosed herein. In addition, genetic modifications can be made to organisms to promote pyruvic acid production, such as those in *E. coli* strain W1485lip2 (Kawasaki *et al.*, *J. Ferm. and Bioeng.* 82:604-6, 1996).

Controlling Chirality

The taste profile of monatin can be altered by controlling the stereochemistry (chirality) of the molecule. For example, different monatin isomers may be desired in different blends of concentrations for different food systems. Chirality can be controlled via a combination of pH and polypeptides.



Racemization at the C-4 position of monatin (see numbered molecule above) can occur by deprotonation and reprotonation of the alpha carbon, which can occur by a shift in pH or by reaction with the cofactor PLP. In a microorganism, the pH is unlikely to shift enough to cause the racemization, but PLP is abundant. Methods to control the chirality with polypeptides depend upon the biosynthetic route utilized for monatin production.

When monatin is formed using the pathway shown in FIG. 2, the following can be considered. In a biocatalytic reaction, the chirality of carbon-2 is determined by the enzyme that converts indole-3-pyruvate to MP. Multiple enzymes (e.g. from EC 4.1.2.-, 4.1.3.-) can convert indole-3-pyruvate to MP, thus, one can choose the enzyme that forms the desired isomer. Alternatively, the enantiospecificity of the enzyme that converts indole-3-pyruvate to MP can be modified through the use of directed evolution or catalytic antibodies can be engineered to catalyze the desired reaction. Once MP is produced (either enzymatically or by chemical condensation), the amino group can be added stereospecifically using a transaminase, such as those described herein.

Either the R or S configuration of carbon-4 can be generated depending on whether a D- or L-aromatic acid aminotransferase is used. Most aminotransferases are specific for the L-isomer, however D-tryptophan aminotransferases exist in certain plants (Kohiba and Mito, Proceedings of the 8th International Symposium on Vitamin B₆ and Carbonyl Catalysis, Osaka, Japan 1990). Moreover, D-alanine aminotransferases (2.6.1.21), D-methionine-pyruvate aminotransferases (2.6.1.41) and both (R)-3-amino-2-methylpropanoate aminotransferase (2.6.1.61) and (S)-3-amino-2-methylpropanoate aminotransferase (2.6.1.22) have been identified. Certain aminotransferases may only accept the substrate for this reaction with a particular configuration at the C2 carbon. Therefore, even if the conversion to MP is not stereospecific, the stereochemistry of the final product can be controlled through the appropriate selection of a transaminase. Since the reactions are reversible, the unreacted MP (undesired isomer) can be recycled back to its constituents and a racemic mixture of MP can be reformed.

Activation of substrates

Phosphorylated substrates, such as phosphoenolpyruvate (PEP), can be used in the reactions disclosed herein. Phosphorylated substrates can be more energetically favorable and, therefore, can be used to increase the reaction rates and/or yields. In aldol condensations, the addition of a phosphate group stabilizes the enol tautomer of the nucleophilic substrate, making it more reactive. In other reactions, a phosphorylated substrate often provides a better leaving group. Similarly, substrates can be activated by conversion to CoA derivatives or pyrophosphate derivatives.

Example 1

Cloning and Expression of Tryptophan Aminotransferases

This example describes methods that were used to clone tryptophan aminotransferases, which can be used to convert tryptophan to indole-3-pyruvate.

Experimental Overview

Eleven genes encoding aminotransferases were cloned into *E. coli*. These genes were *Bacillus subtilis* D-alanine aminotransferase (*dat*, Genbank Accession No. Y14082.1 bp 28622-29470 and Genbank Accession No. NP_388848.1, nucleic acid sequence and amino acid sequence, respectively), *Sinorhizobium meliloti* (also termed *Rhizobium meliloti*) tyrosine aminotransferase (*tatA*, SEQ ID NOS: 1 and 2, nucleic acid sequence and amino acid sequence, respectively), *Rhodobacter sphaeroides* strain 2.4.1 tyrosine aminotransferase (*tatA* asserted by homology, SEQ ID NOS: 3 and 4, nucleic acid sequence and amino acid sequence, respectively), *R. sphaeroides* 35053 tyrosine aminotransferase (asserted by homology, SEQ ID NOS: 5 and 6, nucleic acid sequence and amino acid sequence, respectively), *Leishmania major* broad substrate aminotransferase (*bsat*, asserted by homology to peptide fragments from *L. mexicana*, SEQ ID NOS: 7 and 8, nucleic acid sequence and amino acid sequence, respectively), *Bacillus subtilis* aromatic aminotransferase (*araT*, asserted by homology, SEQ ID NOS: 9 and 10, nucleic acid sequence and amino acid sequence,

respectively), *Lactobacillus amylovorus* aromatic aminotransferase (*araT* asserted by homology, SEQ ID NOS: 11 and 12, nucleic acid sequence and amino acid sequence, respectively), *R. sphaeroides* 35053 multiple substrate aminotransferase (asserted by homology, SEQ ID NOS: 13 and 14, nucleic acid sequence and amino acid sequence, respectively), *Rhodobacter sphaeroides* strain 2.4.1 multiple
5 substrate aminotransferase (*msa* asserted by homology, Genbank Accession No. AAAE01000093.1, bp 14743-16155 and Genbank Accession No. ZP00005082.1, nucleic acid sequence and amino acid sequence, respectively), *Escherichia coli* aspartate aminotransferase (*aspC*, Genbank Accession No. AE000195.1 bp 2755-1565 and Genbank Accession No. AAC74014.1, nucleic acid sequence and amino acid sequence, respectively), and *E. coli* tyrosine aminotransferase (*tyrB*, SEQ ID NOS: 31
10 and 32, nucleic acid sequence and amino acid sequence, respectively).

The genes were cloned, expressed, and tested for activity in conversion of tryptophan to indole-3-pyruvate, along with commercially available enzymes. All eleven clones had activity.

Identification of Bacterial Strains that May Contain Polypeptides with the Desired Activity

15 No genes in the NCBI (National Center for Biotechnology Information) database were designated as tryptophan aminotransferases. However, organisms having this enzymatic activity have been identified. L-tryptophan aminotransferase (TAT) activity has been measured in cell extracts or from purified protein from the following sources: Rhizobacterial isolate from *Festuca octoflora*, pea mitochondria and cytosol, sunflower crown gall cells, *Rhizobium leguminosarum*
20 biovar *trifoli*, *Erwinia herbicola* pv *gypsophylae*, *Pseudomonas syringae* pv. *savastanoi*, *Agrobacterium tumefaciens*, *Azospirillum lipoferum* & *brasilense*, *Enterobacter cloacae*, *Enterobacter agglomerans*, *Bradyrhizobium elkanii*, *Candida maltosa*, *Azotobacter vinelandii*, rat brain, rat liver, *Sinorhizobium meliloti*, *Pseudomonas fluorescens* CHA0, *Lactococcus lactis*, *Lactobacillus casei*, *Lactobacillus helveticus*, wheat seedlings, barley, *Phaseolus aureus* (mung bean), *Saccharomyces*
25 *uvarum* (carlsbergensis), *Leishmania* sp., maize, tomato shoots, pea plants, tobacco, pig, *Clostridium sporogenes*, and *Streptomyces griseus*.

Isolation of Genomic DNA for Cloning

S. meliloti (ATCC number 9930) was grown in TY media at 25°C, pH 7.2. Cells were
30 grown to an optical density at 600 nm (OD₆₀₀) of 1.85 and a 2% inoculum was used for genomic DNA preparations. The Qiagen genomic tip 20/G kit (Valencia, CA) was used for genomic DNA isolation.

Bacillus subtilis 6051 (ATCC) was grown at 30°C in Bereto Nutrient Broth (Difco; Detroit, MI). The Qiagen genomic tip 20/G protocol was used to isolate the genomic DNA with the
35 following changes: the concentrations of proteinase K and lysozyme were doubled and incubation times were increased 2-3 fold.

Leishmania major ATCC 50122 genomic DNA was supplied by IDI, Inc. (Quebec, Canada) in TE buffer pH 8.0, 17 ng/μL.

Rhodobacter sphaeroides 2.4.1 (provided by Professor Sam Kaplan, University of Texas, Houston), *R. sphaeroides* 35053 (ATCC number), and *L. amylovorus* genomic DNA was prepared by standard phenol extraction. Cells were harvested in late log phase, resuspended in TEN buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 100 mM NaCl), and lysed by the addition of 0.024 mL sodium lauryl sarcosine per mL cell suspension. After extracting at least three times with an equal volume of phenol saturated with TE buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA), the DNA solution was extracted once with 9:1 chloroform:octanol and three times with chloroform. The DNA was precipitated by the addition of 0.1 volume of 3 M sodium acetate, pH 6.8 and 2 volumes ethanol. The precipitate was collected by centrifugation and washed once with 70% ethanol. Finally the DNA was dissolved in 0.10 mL distilled water.

Escherichia coli genomic DNA was isolated from strain DH10B (Invitrogen) and prepared using the Qiagen Genomic-tip™ (500/G) kit. From 30 mL of this strain grown in LB to an OD₆₅₀ of 1.87, 0.3 mg of purified DNA was obtained. The purified DNA was dissolved in Qiagen elution buffer (EB) at a concentration of 0.37 µg/µL.

Polymerase Chain Reaction Protocol

Primers were designed with compatible overhangs for the pET 30 Xa/LIC vector (Novagen, Madison, WI). The pET vector has a 12 base single stranded overhang on the 5' side of the Xa/LIC site and a 15-base single stranded overhang on the 3' side of the Xa/LIC site. The plasmid is designed for ligation independent cloning, with N-terminal His and S-tags and an optional C-terminal His-tag. The Xa protease recognition site (IEGR) sits directly in front of the start codon of the gene of interest, such that the fusion protein tags can be removed.

The following sequences were added to the 5' ends of the organism specific sequences when designing primers: forward primer, 5' GGTATTGAGGGTCGC (SEQ ID NO: 73); reverse primer: 5' AGAGGAGAGTTAGAGCC (SEQ ID NO: 74).

Bacillus subtilis *dat* primers: N term: 5'-

GGTATTGAGGGTCGCATGAAGGTTTGTAGTCAATGG-3' and C term: 5'-

AGAGGAGAGTTAGAGCCTTATGAAATGCTAGCAGCCT-3' (SEQ ID NOS: 15 and 16).

Sinorhizobium meliloti *tatA* primers: N term: 5'-

GGTATTGAGGGTCGCATGTTTCGACGCCCTCGCCCG and C term: 5'-

AGAGGAGAGTTAGAGCCTCAGAGACTGGTGAACCTGTC (SEQ ID NOS: 17 and 18).

Bacillus subtilis *araT* primers: N term: 5'-

GGTATTGAGGGTCGCATGGAACATTGCTGAATCC and C term: 5'-

AGAGGAGAGTTAGAGCCTTAAACGCCGTTGTTTATCG (SEQ ID NOS: 19 and 20).

Rhodobacter sphaeroides *msa* (both 2.4.1 and 35053): N term: 5'-

GGTATTGAGGGTCGCATGCGCGAGCCTCTTGCCCT and C term: 5'-

AGAGGAGAGTTAGAGCCTCAGCCGGGGAAGCTCCGGG (SEQ ID NOS: 21 and 22).

Leishmania major bsat: N term: 5'-

GGTATTGAGGGTCGCATGTCCACGCAGGCGGCCAT and C term: 5'-
AGAGGAGAGTTAGAGCCTCACTCACGATTACATTGC (SEQ ID NOS: 23 and 24).

Lactobacillus amylovorus araT: N term: 5'-

5 GGTATTGAGGGTCGCATGCCAGAATTAGCTAATGA and C term: 5'-
AGAGGAGAGTTAGAGCCTTATTCGTCCTCTTGTAATAA (SEQ ID NOS: 25 and 26).

Rhodobacter sphaeroides tata (both 2.4.1 and 35053 strains): N term: 5'-

GGTATTGAGGGTCGCATGCGCTCTACGACGGCTCC and C term: 5'-
AGAGGAGAGTTAGAGCCTCAGCCGCGCAGCACCTTGG (SEQ ID NOS: 27 and 28).

10 *Escherichia coli aspC*: N term: 5'-

GGTATTGAGGGTCGCATGTTTGAGAACATTACCGC-3' and C term: 5'-
AGAGGAGAGTTAGAGCCTTACAGCACTGCCACAATCG-3' (SEQ ID NOS: 29 and 30).

Escherichia coli tyrB: N term: 5'-

GGTATTGAGGGTCGCGTGTTCAAAAAGTTGACGC and C term: 5'-
15 AGAGGAGAGTTAGAGCCTTACATCACCGCAGCAAACG-3' (SEQ ID NOS: 33 and 34).

The gene derived from *S. meliloti* (*tata*) was amplified using the following PCR protocol.

In a 50 µL reaction 0.1-0.5 µg template, 1.5 µM of each primer, 0.4 mM each dNTP, 3.5 U Expand High Fidelity Polymerase (Roche, Indianapolis, IN), and 1X ExpandTM buffer with Mg were used.

The thermocycler program used included a hot start at 96°C for 5 minutes, followed by 29 repetitions
20 of the following steps: 94°C for 30 seconds, 55°C for 2 minutes, and 72°C for 2.5 minutes. After the
29 repetitions the sample was maintained at 72°C for 10 minutes and then stored at 4°C. This PCR
protocol produced a product of 1199 bp.

The sequences of the genes derived from *R. sphaeroides* (*msa* and *tata*), *L. amylovorus*
araT, and *Bacillus araT* were amplified using the following PCR protocol. In a 50 µL reaction, 0.1-

25 0.5 µg template, 1.5 µM of each primer, 0.4 mM each dNTP, 3.5 U Expand High FidelityTM
Polymerase, and 1X ExpandTM buffer with Mg were added. The thermocycler program used included
a hot start at 96°C for 5 minutes, followed by 29 repetitions of the following steps: 94°C for 30
seconds, 40-60°C for 1 minute, 45 seconds (gradient thermocycler) and 72°C for 2 minutes, 15
seconds. After the 29 repetitions the sample was maintained at 72°C for 10 minutes and then stored
30 at 4°C.

For each *R. sphaeroides msa* gene, the 42°C and 48°C annealing temperatures produced
multiple products, but a distinct band at approximately 1464 bp. For *L. amylovorus araT*, the 42°C,
48°C, and 56°C annealing temperatures yielded single products with intense bands at 1173 bp. For *B.*
subtilis araT, the 40°C, 45°C, 50°C, 55°C annealing temperatures generated single intense products
35 (1173 bp), from both genomic DNA and colonies. For *L. major bsat*, the 55°C annealing temperature
gave the cleanest product (1239 bp). For *Rhodobacter tata* genes, the 50-55°C annealing
temperatures gave clean products at the correct size (1260 bp). For both *E. coli* genes and the *B.*
subtilis dat gene, an annealing temperature of 55-60°C was used, and the annealing time was

shortened to 45 seconds. Clean products of the correct sizes were obtained (approximately 1.3 kb for the *E. coli* genes, 850 bp for the *dat* gene).

Cloning

5 The PCR products were gel purified from 0.8 or 1% TAE-agarose gels using the Qiagen gel extraction kit (Valencia, CA). The PCR products were quantified by comparison to standards on an agarose gel, and then treated with T4 DNA polymerase following the manufacturer's recommended protocols for Ligation Independent Cloning (Novagen, Madison, WI).

10 Briefly, approximately 0.2 pmol of purified PCR product was treated with 1 U T4 DNA polymerase in the presence of dGTP for 30 minutes at 22°C. The polymerase removes successive bases from the 3' ends of the PCR product. When the polymerase encounters a guanine residue, the 5' to 3' polymerase activity of the enzyme counteracts the exonuclease activity to effectively prevent further excision. This creates single stranded overhangs that are compatible with the pET Xa/LIC vector. The polymerase is inactivated by incubating at 75°C for 20 minutes.

15 The vector and treated insert were annealed as recommended by Novagen. Approximately 0.02 pmol of treated insert and 0.01 pmol vector were incubated for 5 minutes at 22°C, 6.25 mM EDTA (final concentration) was added, and the incubation at 22°C was repeated. The annealing reaction (1 µL) was added to NovaBlue™ singles competent cells (Novagen, Madison, WI), and incubated on ice for 5 minutes. After mixing, the cells were transformed by heat shock for 30
20 seconds at 42°C. The cells were placed on ice for 2 minutes, and allowed to recover in 250 µL of room temperature SOC for 30 minutes at 37°C with shaking at 225 rpm. Cells were plated on LB plates containing kanamycin (25-50 µg/mL).

Plasmid DNA was purified using the Qiagen spin miniprep kit and screened for the correct inserts by restriction digest with *Xho*I and *Xba*I. The sequences of plasmids that appeared to have the
25 correct insert were verified by dideoxy chain termination DNA sequencing.

SEQ ID NOS: 1-14 and 31-32 show nucleotide and corresponding amino acid sequences of the recombinant aminotransferases, any changes from the Genbank sequences were either silent or generated conservative substitutions in the protein sequence. SEQ ID NOS: 11 and 12 are novel
30 sequences.

Gene Expression and Assays

Plasmid DNA, verified by sequence analysis, was subcloned into *E. coli* expression hosts BLR(DE3) or BL21(DE3) (Novagen, Madison, WI). The cultures were grown and the plasmids were isolated using Qiagen miniprep kit, and analyzed by restriction digest to confirm identity.

35 Induction was initially performed with *L. amylovorus araT*, *B. subtilis araT*, and *S. meliloti tata* in both BLR(DE3) and BL21(DE3) cells. A time course study was performed with cultures grown in LB containing kanamycin (30 mg/L) to an OD₆₀₀ of 0.5-0.8 and induced with 1 mM IPTG (isopropyl thiogalactoside) and sampled at 0, 1, 2, and 4 hours post induction. Cells from 2.0 mL were resuspended in 0.10 mL 120 mM Tris-HCl, pH 6.8 containing 10% sodium dodecyl sulfate,

10% 2-mercaptoethanol, and 20% glycerol, heated at 95°C for 10 min, and cooled, and diluted with 0.10 mL H₂O. Aliquots of these total cellular protein samples were analyzed by SDS-PAGE using a 4-15% gradient gel. There were no significant differences in the amount of protein expressed between the 2 hour and 4 hour induction, nor between the BLR(DE3) and BL21(DE3) cells.

5 Cell extracts were also prepared from the 4 hour samples by suspending cell pellets from 2 mL of culture in 0.25 mL Novagen BugBuster™ reagent containing 0.25 µL benzonase nuclease, incubating at room temperature for 20 minutes with gentle shaking, and centrifuging at 16,000 x g to remove cell debris. The supernatants (cell extracts) were loaded onto 4-15% gradient gels for analysis of the cellular soluble proteins.

10 The three clones, (*L. amylovorus araT* (SEQ ID NOS: 11 and 12), *B. subtilis araT* (SEQ ID NOS: 9 and 10), and *S. meliloti tatA* (SEQ ID NOS: 1 and 2) showed soluble protein that corresponded to the correct size (approximately 45 kDa). The *B. subtilis araT* gene product was over-expressed at the highest level and/or was more soluble than the other two gene products.

15 In subsequent expression methods, plasmid DNA from positive clones was subcloned into BL21(DE3) due to the better growth characteristics of this host. Induction was repeated using 1 mM IPTG with cultures grown in LB containing kanamycin at 50 mg/L, inducing when the OD₆₀₀ reached approximately 0.8. Cells were harvested after 4 hours of growth at 37°C, centrifuged at 3000 rpm for 10 minutes (4°C), washed with TEGGP buffer (50 mM Tris-HCl (pH 7.0), 0.5 mM EDTA, 100 mg/L glutathione, 5% glycerol, with Roche complete protease inhibitor cocktail), and flash frozen in -80°C ethanol.

20 Samples were resuspended in 5 mL/g wet cell weight of BugBuster™ (Novagen) reagent containing 5 µL/mL protease inhibitor cocktail set #3 (Calbiochem-Novabiochem Corp., San Diego, CA) and 1 µL/mL benzonase nuclease. Samples were incubated at room temperature for 20 minutes on an orbital shaker. Insoluble cell debris was removed by centrifugation at 16,000 X g for 20 minutes at 4°C.

25 Cell extracts were analyzed by SDS-PAGE, and assayed for tryptophan aminotransferase activity by following production of indole-pyruvic acid using the following protocol. One mL reactions were carried out in 50 mM sodium tetraborate (pH 8.5), 0.5 mM EDTA, 0.5 mM sodium arsenate, 50 µM pyridoxal phosphate, 5 mM α-ketoglutarate, and 5 mM L-tryptophan. The reactions were initiated by the addition of cell free extracts or purified enzyme and were incubated 30 minutes at 30°C. 20% TCA (200 µL) was added to stop the reaction, and the precipitated protein was removed by centrifugation. The absorbance at 327 nm was measured and compared to a standard curve of freshly prepared indole-3-pyruvate in the assay buffer. Control reactions without the substrate tryptophan or using cell-free extracts from clones transformed with pET30a alone were also performed.

35 Due to background from the native *E. coli* aminotransferases in cell extracts, recombinant proteins were purified using His-Bind cartridges following manufacturer's protocols (Novagen, Madison, WI). The eluent fractions were desalted on PD-10 (Amersham Biosciences, Piscataway,

NJ) columns and eluted in 50 mM Tris, pH 7.0. Purified proteins were analyzed by SDS-PAGE and assayed for aminotransferase activity.

Results from the 37°C induction with 1 mM IPTG (4 hours) demonstrate that *L. major bsat*, *S. meliloti tata*, *E. coli aspC*, and both *R. sphaeroides tata* clones have significant levels of tryptophan aminotransferase activity. The *araT* protein from *B. subtilis* was over-expressed and soluble, but showed little enzymatic activity. The *L. amylovorus araT* gene product appeared to be soluble in the cell extract, but purification using a His-Bind cartridge resulted in only small amounts of protein with the correct molecular weight. The *msa* gene products were insoluble and further expression experiments were done at 24°C to minimize inclusion body formation. Several concentrations of IPTG between 10 µM and 1 mM were used to maximize the amount of soluble protein.

Table 1 lists the specific activities measured in micrograms of indole-3-pyruvate (I3P) formed per milligram protein per minute. In some cases, very small amounts of recombinant protein showed high levels of activity above the effective linear range of the assay. In these cases a '>' precedes the specific activity number.

Table 1: Specific Activities of Clones in Cell Extracts (CE) and Purified (P) and Commercial Enzymes

Enzyme	Specific Activity (µg I3P/mg protein/min)	Note
<i>L. major bsat</i> CE	>49.3	
<i>L. major bsat</i> P	>4280	
<i>S. meliloti tata</i> CE	>28.6	
<i>S. meliloti tata</i> P	>931	
2.4.1 <i>tata</i> CE	>41.2	
2.4.1 <i>tata</i> P	1086	
35053 <i>tata</i> CE	>62.3	
35053 <i>tata</i> P	>486	
<i>L. amylovorus araT</i> CE	1.26	
<i>L. amylovorus araT</i> P	0	little protein after His-Bind cartridge
<i>B. subtilis araT</i> CE	0	undetectable
<i>B. subtilis araT</i> P	1.5-4.5	
2.4.1 <i>msa</i> CE	2.05	very little soluble protein
2.4.1 <i>msa</i> P	0	no protein after His-Bind cartridge
35053 <i>msa</i> CE	3.97	very little soluble protein
35053 <i>msa</i> P	0	no protein after His-Bind cartridge
<i>E. coli aspC</i> (P)	800	
<i>E. coli tyrB</i> (P)	1	not very soluble
<i>B. subtilis</i> D-aminotransf.(P)	2.7	using D-tryptophan as substrate
broad range transaminase	22	Sigma cat # T 7684
Porcine type II-A	1.5	Sigma G7005
Porcine type I	1	Sigma G2751

An alignment comparing all of the recombinant proteins cloned illustrates that there are not many highly conserved areas between the *araT*, *tata*, *bsat*, and *msa* sequences. An alignment of

highest activity recombinant proteins: *Rhodobacter tata* gene product homologs, *L. major* broad substrate aminotransferase, and the *Sinorhizobium meliloti* tyrosine aminotransferase showed several conserved regions, however they are only approximately 30-43% identical at the protein level. The availability of the broad range, D-specific (D-alanine) aminotransferase can be useful in the production of other stereoisomers of monatin.

Example 2

Conversion of Indole-3-lactate to Indole-3-pyruvate

As shown in FIGS. 1 and 3, indole-3-lactic acid can be used to produce indole-3-pyruvate. Conversion between lactic acid and pyruvate is a reversible reaction, as is conversion between indole-3-pyruvate and indole-3-lactate. The oxidation of indole-lactate was typically followed due to the high amount of background at 340 nm from indole-3-pyruvate.

The standard assay mixture contained 100 mM potassium phosphate, pH 8.0, 0.3 mM NAD^+ , 7 units of lactate dehydrogenase (LDH) (Sigma-L2395, St. Louis, MO), and 2 mM substrate in 0.1 mL. The assay was performed in duplicate in a UV-transparent microtiter plate, using a Molecular Devices SpectraMax Plus platereader. Polypeptide and buffer were mixed and pipetted into wells containing the indole-3-lactic acid and NAD^+ and the absorbance at 340 nm of each well was read at intervals of 9 seconds after brief mixing. The reaction was held at 25°C for 5 minutes. The increase in absorbance at 340 nm follows the production of NADH from NAD^+ . Separate negative controls were performed without NAD^+ and without substrate. D-LDH from *Leuconostoc mesenteroides* (Sigma catalog number L2395) appeared to exhibit more activity with the indole-derivative substrates than did L-LDH from *Bacillus stearothermophilus* (Sigma catalog number L5275).

Similar methods were utilized with D-lactic acid and NAD^+ or NADH and pyruvate, the natural substrates of D-LDH polypeptides. The $V_{\max}$ for the reduction of pyruvate was 100-1000 fold higher than the $V_{\max}$ for the oxidation of lactate. The $V_{\max}$ for the oxidation reaction of indole-3-lactic with D-LDH was approximately one-fifth of that with lactic acid. The presence of indole-3-pyruvate was also measured by following the change in absorbance at 327 (the enol-borate derivative) using 50 mM sodium borate buffer containing 0.5 mM EDTA and 0.5 mM sodium arsenate. Small, but repeatable, absorbance changes were observed, as compared to the negative controls for both L and D-LDH polypeptides.

Additionally, broad specificity lactate dehydrogenases (enzymes with activity associated with EC 1.1.1.27, EC 1.1.1.28, and/or EC 1.1.2.3), can be cloned and used to make indole-3-pyruvate from indole-3-lactic acid. Sources of broad specificity dehydrogenases include *E. coli*, *Neisseria gonorrhoeae*, and *Lactobacillus plantarum*.

Alternatively, indole-3-pyruvate can be produced by contacting indole-3-lactate with cellular extracts from *Clostridium sporogenes* which contain an indolelactate dehydrogenase (EC 1.1.1.110); or *Trypanosoma cruzi* epimastigotes cellular extracts which contain *p*-hydroxyphenyllactate dehydrogenase (EC 1.1.1.222) known to have activity on indole-3-pyruvate; or *Pseudomonas*

acidovorans or *E. coli* cellular extracts, which contain an imidazol-5-yl lactate dehydrogenase (EC 1.1.1.111); or *Coleus blumei*, which contains a hydroxyphenylpyruvate reductase (EC 1.1.1.237); or *Candida maltosa* which contains a D-aromatic lactate dehydrogenase (EC 1.1.1.222). References describing such activities include, Nowicki *et al.* (*FEMS Microbiol Lett* 71:119-24, 1992), Jean and DeMoss (*Canadian J. Microbiol.* 14 1968, Coote and Hassall (*Biochem. J.* 111: 237-9, 1969), Cortese *et al.* (*C.R. Seances Soc. Biol. Fil.* 162 390-5, 1968), Petersen and Alfermann (*Z. Naturforsch. C: Biosci.* 43 501-4, 1988), and Bhatnagar *et al.* (*J. Gen Microbiol* 135:353-60, 1989). In addition, a lactate oxidase such as the one from *Pseudomonas sp.* (Gu *et al.* *J. Mol. Catalysis B: Enzymatic*: 18:299-305, 2002), can be utilized for oxidation of indole-3-lactic to indole-3-pyruvate.

Example 3

Conversion of L-tryptophan to Indole-3-pyruvate utilizing L-amino acid oxidase

This example describes methods used to convert tryptophan to indole-3-pyruvate via an oxidase (EC 1.4.3.2), as an alternative to using a tryptophan aminotransferase as described in Example 1. L-amino acid oxidase was purified from *Crotalus durissus* (Sigma, St. Louis, MO, catalog number A-2805). The accession numbers of L-amino acid oxidases for molecular cloning include: CAD21325.1, AAL14831, NP_490275, BAB78253, A38314, CAB71136, JE0266, T08202, S48644, CAC00499, P56742, P81383, O93364, P81382, P81375, S62692, P23623, AAD45200, AAC32267, CAA88452, AP003600, and Z48565.

Reactions were performed in microcentrifuge tubes in a total volume of 1 mL, incubated for 10 minutes while shaking at 37°C. The reaction mix contained 5 mM L-tryptophan, 100 mM sodium phosphate buffer pH 6.6, 0.5 mM sodium arsenate, 0.5 mM EDTA, 25 mM sodium tetraborate, 0.016 mg catalase (83 U, Sigma C-3515), 0.008 mg FAD (Sigma), and 0.005-0.125 Units of L-amino acid oxidase. Negative controls contained all components except tryptophan, and blanks contained all components except the oxidase. Catalase was used to remove the hydrogen peroxide formed during the oxidative deamination. The sodium tetraborate and arsenate were used to stabilize the enol-borate form of indole-3-pyruvate, which shows a maximum absorbance at 327 nm. Indole-3-pyruvate standards were prepared at concentrations of 0.1-1 mM in the reaction mix.

The purchased L-amino acid oxidase had a specific activity of 540 μ g indole-3-pyruvate formed per minute per mg protein. This is the same order of magnitude as the specific activity of tryptophan aminotransferase enzymes.

Example 4

Converting Indole-3-pyruvate to 2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid with an Aldolase

This example describes methods that can be used to convert indole-3-pyruvate to the 2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid monatin precursor (MP) using an aldolase (lyase) (FIG. 2). Aldol condensations are reactions that form carbon-carbon bonds between the β -carbon of an aldehyde or ketone and the carbonyl carbon of another aldehyde or ketone. A carbanion is formed

on the carbon adjacent to the carbonyl group of one substrate, and serves as a nucleophile attacking the carbonyl carbon of the second substrate (the electrophilic carbon). Most commonly, the electrophilic substrate is an aldehyde, so most aldolases fall into EC 4.1.2.- category. Quite often, the nucleophilic substrate is pyruvate. It is less common for aldolases to catalyze the condensation between two keto-acids or two aldehydes.

However, aldolases that catalyze the condensation of two carboxylic acids have been identified. For example, EP 1045-029 describes the production of L-4-hydroxy-2-ketoglutaric acid from glyoxylic acid and pyruvate using a *Pseudomonas* culture (EC 4.1.3.16). In addition, 4-hydroxy-4-methyl-2-oxoglutarate aldolase (4-hydroxy-4-methyl-2-oxoglutarate pyruvate lyase, EC 4.1.3.17) can catalyze the condensation of two keto acids. Therefore, similar aldolase polypeptides were used to catalyze the condensation of indole-3-pyruvate with pyruvate.

Cloning

4-Hydroxy-4-methyl-2-oxoglutarate pyruvate lyases (ProA aldolase, EC 4.1.3.17) and 4-hydroxy-2-oxoglutarate glyoxylate-lyase (KHG aldolase, EC 4.1.3.16) catalyze reactions very similar to the aldolase reaction of FIG. 2. Primers were designed with compatible overhangs for the pET30 Xa/LIC vector (Novagen, Madison, WI). The design of these primers is described above in Example 1.

The following primers were designed for pET30 Xa/LIC cloning:

1. *Pseudomonas straminea proA* gene (Genbank Accession No.: 12964663 Version: 12964663) and *Comamonas testosteroni proA* gene (SEQ ID NOS: 65-66, nucleic acid sequence and amino acid sequence, respectively) forward 5'-GGTATTGAGGGTCGCATGTACGAACTGGGAGTTGT-3' and reverse 5'-AGAGGAGAGTTAGAGCCTTAGTCAATATATTTTCAGGC-3' (SEQ ID NOS: 55 and 56).

2. *Sinorhizobium meliloti* 1021 SMc00502 gene (homologous to *proA*, Genbank Accession Nos.: 15074579 and CAC46344, nucleic acid sequence and amino acid sequence, respectively) forward 5'-GGTATTGAGGGTCGCATGAGCGTGGTTCACCGGAA-3' and reverse 5'-AGAGGAGAGTTAGAGCCTCAATCGATATATTTTCAGTC-3' (SEQ ID NOS: 61 and 62).

3. *Sphingomonas sp.* LB126 fldZ gene (Genbank Accession No.: 7573247 Version: 7573247, codes for a putative acyl transferase) forward 5'-GGTATTGAGGGTCGCATGTCCGGCATCGTTGTCCA-3' and reverse 5'-AGAGGAGAGTTAGAGCCTCAGACATATTTTCAGTCCCA-3' (SEQ ID NOS: 57 and 58).

4. *Arthrobacter keyseri* pcmE gene (Genbank Accession No.: AF331043 Version: AF331043.1, codes for an oxalocitramalate aldolase) forward 5'-GGTATTGAGGGTCGCATGCGACTGAACAACCTCGG-3' and reverse 5'-AGAGGAGAGTTAGAGCCTCAGTTCTCCACGTATTTCCA-3' (SEQ ID NOS: 59 and 60).

5. *Yersinia pestis* strain CO92 YPO0082 gene (Genbank Accession No.: 15978115 Version: 15978115, codes for a possible transferase) forward 5'-

GGTATTGAGGGTCGCATGAGCCTGGTTAATATGAA-3' and reverse 5'-
AGAGGAGAGTTAGAGCCTTATGACTTTAACGCGTTGA-3' (SEQ ID NOS: 63 and 64).

6. *Bacillus subtilis khg* gene (Genbank Accession Nos. Z99115.1 GI:2634478, 126711-
127301 and CAB14127.1, nucleic acid sequence and amino acid sequence, respectively) forward 5'-
5 GGTATTGAGGGTCGCATGGAGTCCAAAGTCGTTGA-3' and reverse 5'-
AGAGGAGAGTTAGAGCCTTACACTTGGAACAGCCT-3' (SEQ ID NOS: 35 and 36).

7. *E. coli khg* gene (Genbank Accession Nos. AE000279.1 1331-1972 and AAC74920.1,
nucleic acid and amino acid sequence, respectively) forward 5'-
GGTATTGAGGGTCGCATGAAAACTGGAAAAACAAG-3' and reverse 5'-
10 AGAGGAGAGTTAGAGCCTTACAGCTTAGCGCCTTCTA-3' (SEQ ID NOS: 37 and 38).

8. *S. meliloti khg* gene (Genbank Accession Nos. AL591792.1 GI:15075850, 65353-64673
and CAC47463.1, nucleic acid and amino acid sequence, respectively) forward 5'-
GGTATTGAGGGTCGCATGCGAGGGGCATTATTCAA-3' and reverse 5'-
AGAGGAGAGTTAGAGCCTCAGCCCTTGAGCGCGAAG-3' (SEQ ID NOS: 39 and 40).

15 Genomic DNA from the organisms described in 1-2 and 6-8, above, was purified using the
Qiagen genomic-tip protocol. Using similar techniques the genomic DNA from organisms described
in 3-5 can be purified.

Pseudomonas straminea (ATCC 33636) was grown at 30°C in Nutrient Broth and
hydroxybenzoate medium. *Comamonas testosteroni* (ATCC 49249) was grown at 26°C in Nutrient
20 Broth and hydroxybenzoate medium. *Sphingomonas* sp. LB126 (Flemish Institute for Technological
Research, VITO, B-2400 Mol, Belgium) is grown according to the method described by Wattiau *et al.*
(*Research in Microbiol.* 152:861-72, 2001). *Arthrobacter keyseri* (Gulf Ecology Division, National
Health and Environmental Effects Research Laboratory, U.S. Environmental Protection Agency, Gulf
Breeze, FL 32561, USA) is grown according to the protocol described by Eaton (*J. Bacteriol.*
25 183:3689-3703, 2001). *Sinorhizobium meliloti* 1021 (ATCC 51124) was grown at 26°C in ATCC
TY medium and hydroxybenzoate medium. *Yersinia pestis* strain CO92 (ATCC) is grown at 26°C in
ATCC medium 739 Horse blood agar. *Bacillus subtilis* 6051 (ATCC) was grown at 30°C in Bereto
Nutrient Broth (Difco; Detroit, MI). *E. coli* genomic DNA was isolated from strain DH10B
(Invitrogen) as described in Example 1.

30 The PCR, cloning, and screening protocols described in Example 1 were used to clone the *C.*
testosteroni and the *S. meliloti proA* sequences, as well as the *E. coli*, *B. subtilis*, and *S. meliloti khg*
sequences. The same methods can be used to clone the other sequences described above.

Positive clones were sequenced using dideoxy chain termination sequencing (Seqwright,
Houston, TX) with S-tag and T7 terminator primers (Novagen), and internal primers from Integrated
35 DNA Technologies, Inc. (Coralville, IA).

Expression and Activity Assays

Plasmid DNA (verified by sequence analysis) was subcloned into expression host
BL21(DE3) (Novagen). The cultures were grown in LB medium with 50 mg/L kanamycin, the

plasmids isolated using a Qiagen spin plasmid miniprep kit and subsequently analyzed by restriction digest to confirm identity. Induction experiments were done with the BL21(DE3) constructs grown in LB medium containing 50 mg/L kanamycin at 37°C. Protein expression was induced using 0.1 mM IPTG after the OD₆₀₀ reached approximately 0.6. The cells were grown for 4 hours at 30°C and harvested by centrifugation. The cells were then lysed using Bugbuster™ reagent (Novagen) and the His-tag recombinant proteins were purified using His-Bind cartridges as described above (Example 1). Purified proteins were desalted on PD-10 disposable columns and eluted in 50 mM Tris-HCl buffer, pH 7.3 with 2 mM MgCl₂.

The proteins were analyzed by SDS-PAGE on 4-15% gradient gels to detect soluble protein levels at the predicted MW of the recombinant fusion protein.

The proteins were assayed for activity using indole-3-pyruvate and sodium pyruvate as substrates. The assay mixture contained 100 mM Tris-HCl (pH 7-pH 8.9), 0-8 mM MgCl₂, 3 mM potassium phosphate (pH 8), and 6 mM of each substrate in 1 mL. The reaction was started by adding varying amounts of polypeptide (for example from 10 to 100 µg), and was incubated at 25°C-37°C for 30 minutes, filtered, and then frozen at -80 °C.

Activity Results with *proA* gene products

Both the *C. testosteroni proA* and *S. meliloti SMc00502* gene constructs had high levels of expression when induced with IPTG. The recombinant proteins were highly soluble, as determined by SDS-PAGE analysis of total protein and cellular extract samples. The *C. testosteroni* gene product was purified to > 95% purity. Because the yield of the *S. meliloti* gene product was very low after affinity purification using a His-Bind cartridge, cellular extract was used for the enzymatic assays.

Both recombinant aldolases catalyzed the formation of MP from indole-3-pyruvate and pyruvate. The presence of both divalent magnesium and potassium phosphate were required for enzymatic activity. No product was apparent when indole-3-pyruvate, pyruvate, or potassium phosphate was absent. A small amount of the product was also formed in the absence of enzyme (typically one order of magnitude less than when enzyme was present).

The product peak eluted from the reverse phase C18 column slightly later than the indole-3-pyruvate standard, the mass spectrum of this peak showed a collisionally-induced parent ion ([M + H]⁺) of 292.1, the parent ion expected for the product MP. The major daughter fragments present in the mass spectrum included those with m/z = 158 (1H-indole-3-carbaldehyde carbonium ion), 168 (3-buta-1,3-dienyl-1H-indole carbonium ion), 274 (292 - H₂O), 256 (292 - 2 H₂O), 238 (292 - 3 H₂O), 228 (292 - CH₄O₃), and 204 (loss of pyruvate). The product also exhibited a UV spectrum characteristic of other indole-containing compounds such as tryptophan, with the λ_{max} of 279-280 and a small shoulder at approximately 290 nm.

The amount of MP produced by the *C. testosteroni* aldolase increased with an increase in reaction temperature from room temperature to 37°C, amount of substrate, and amount of magnesium. The synthetic activity of the enzyme decreased with increasing pH, the maximum

product observed was at pH 7. Based on tryptophan standards, the amount of MP produced under a standard assay using 20 µg of purified protein was approximately 10-40 µg per one mL reaction.

Due to the high degree of homology of the *S. meliloti* and *C. testosteroni* ProA aldolase coding sequences with the other genes described above, it is expected that all of the recombinant gene products can catalyze this reaction. Moreover, it is expected that aldolases that have threonine (T) at positions 59 and 87, arginine (R) at 119, aspartate (D) at 120, and histidine (H) at 31 and 71, (based on the numbering system of *C. testosteroni*) will have similar activity.

Activity Results with *khg* gene products

Both the *B. subtilis* and *E. coli khg* gene constructs had high levels of expression of protein when induced with IPTG, while the *S. meliloti khg* had a lower level of expression. The recombinant proteins were highly soluble, as judged by SDS-PAGE analysis of total proteins and cellular extracts. The *B. subtilis* and *E. coli khg* gene products were purified to > 95% purity; the yield of the *S. meliloti* gene product was not as high after affinity purification using a His-Bind cartridge.

There is no evidence that magnesium and phosphate are required for activity for this enzyme. However, the literature reports performing the assays in sodium phosphate buffer, and the enzyme reportedly is bifunctional and has activity on phosphorylated substrates such as 2-keto-3-deoxy-6-phosphogluconate (KDPG). The enzymatic assays were performed as described above, and in some instances the phosphate was omitted. The results indicate that the recombinant KHG aldolases produced MP, but were not as active as the ProA aldolases. In some cases the level of MP produced by KHG was almost identical to the amount produced by magnesium and phosphate alone. Phosphate did not appear to increase the KHG activities. The *Bacillus* enzyme had the highest activity, approximately 20-25% higher activity than the magnesium and phosphate alone, as determined by SRM (see Example 10). The *Sinorhizobium* enzyme had the least amount of activity, which may be associated with folding and solubility problems noted in the expression. All three enzymes have the active site glutamate (position 43 in *B. subtilis* numbering system) as well as the lysine required for Schiff base formation with pyruvate (position 130); however, the *B. subtilis* enzyme contains a threonine in position 47, an active site residue, rather than arginine. The *B. subtilis* KHG is smaller and appears to be in a cluster distinct from the *S. meliloti* and *E. coli* enzymes, with other enzymes having the active site threonine. The differences in the active site may be the reason for the increased activity of the *B. subtilis* enzyme.

Improvement of Aldolase Activity

Catalytic antibodies can be as efficient as natural aldolases, accept a broad range of substrates, and can be used to catalyze the reaction shown in FIG. 2.

Aldolases can also be improved by directed evolution, for example as previously described for a KDPG aldolase (highly homologous to KHG described above) evolved by DNA shuffling and error-prone PCR to remove the requirement for phosphate and to invert the enantioselectivity. The KDPG aldolase polypeptides are useful in biochemical reactions since they are highly specific for the donor substrate (herein pyruvate), but are relatively flexible with respect to the acceptor substrate

(i.e. indole-3-pyruvate) (Koeller & Wong, *Nature* 409:232-9, 2001). KHG aldolase has activity for condensation of pyruvate with a number of carboxylic acids. Mammalian versions of the KHG aldolase are thought to have broader specificity than bacterial versions, including higher activity on 4-hydroxy 4-methyl 2-oxoglutarate and acceptance of both stereoisomers of 4-hydroxy-2-ketoglutarate. Bacterial sources appear to have a 10-fold preference for the R isomer. There are nearly 100 KHG homologs available in genomic databases, and activity has been demonstrated in *Pseudomonas*, *Paracoccus*, *Providencia*, *Sinorhizobium*, *Morganella*, *E. coli*, and mammalian tissues. These enzymes can be used as a starting point for tailoring the enantiospecificity that is desired for monatin production.

Aldolases that utilize pyruvate and another substrate that is either a keto acid and/or has a bulky hydrophobic group like indole can be "evolved" to tailor the polypeptide's specificity, speed, and selectivity. In addition to KHG and ProA aldolases demonstrated herein, examples of these enzymes include, but are not limited to: KDPG aldolase and related polypeptides (KDPH); transcarboxybenzalpyruvate hydratase-aldolase from *Nocardioides* st; 4-(2-carboxyphenyl)-2-oxobut-3-enoate aldolase (2'-carboxybenzalpyruvate aldolase) which condenses pyruvate and 2-carboxybenzaldehyde (an aromatic ring-containing substrate); trans-O-hydroxybenzylidenepyruvate hydratase-aldolase from *Pseudomonas putida* and *Sphingomonas aromaticivorans*, which also utilizes pyruvate and an aromatic-containing aldehyde as substrates; 3-hydroxyaspartate aldolase (erythro-3-hydroxy-L-aspartate glyoxylate lyase), which uses 2-oxo acids as the substrates and is thought to be in the organism *Micrococcus denitrificans*; benzoin aldolase (benzaldehyde lyase), which utilizes substrates containing benzyl groups; dihydroneopterin aldolase; L-threo-3-phenylserine benzaldehyde-lyase (phenylserine aldolase) which condenses glycine with benzaldehyde; 4-hydroxy-2-oxovalerate aldolase; 1,2-dihydroxybenzylpyruvate aldolase; and 2-hydroxybenzalpyruvate aldolase.

A polypeptide having the desired activity can be selected by screening clones of interest using the following methods. Tryptophan auxotrophs are transformed with vectors carrying the clones of interest on an expression cassette and are grown on a medium containing small amounts of monatin or MP. Since aminotransferases and aldolase reactions are reversible, the cells are able to produce tryptophan from a racemic mixture of monatin. Similarly, organisms (both recombinant and wildtype) can be screened by ability to utilize MP or monatin as a carbon and energy source. One source of target aldolases is expression libraries of various *Pseudomonas* and rhizobacterial strains. *Pseudomonads* have many unusual catabolic pathways for degradation of aromatic molecules and they also contain many aldolases; whereas the rhizobacteria contain aldolases, are known to grow in the plant rhizosphere, and have many of the genes described for construction of a biosynthetic pathway for monatin.

Example 5

Chemical Synthesis of the Monatin Precursor

Example 4 described a method of using an aldolase to convert indole-3-pyruvate to the 2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid monatin precursor (MP). This example describes an alternative method of chemically synthesizing MP.

The MP is formed by using a typical aldol-type condensation (FIG. 4). Briefly, a typical aldol-type reaction involves the generation of a carbanion of the pyruvate ester using a strong base, such as LDA (lithium diisopropylamide), lithium hexamethyldisilazane or butyl lithium. The carbanion that is generated reacts with the indole-pyruvate to form the coupled product.

Protecting groups that can be used for protecting the indole nitrogen include, but are not limited to: t-butyloxycarbonyl (Boc), and benzyloxycarbonyl (Cbz). Blocking groups for carboxylic acids include, but are not limited to, alkyl esters (for example, methyl, ethyl, benzyl esters). When such protecting groups are used, it is not possible to control the stereochemistry of the product that is formed. However, if R₂ and/or R₃ are chiral protecting groups (FIG. 4), such as (S)-2-butanol, menthol, or a chiral amine, this can favor the formation of one MP enantiomer over the other.

Example 6

Conversion of Tryptophan or Indole-3-Pyruvate to Monatin

An *in vitro* process utilizing two enzymes, an aminotransferase and an aldolase, produced monatin from tryptophan and pyruvate. In the first step alpha-ketoglutarate was the acceptor of the amino group from tryptophan in a transamination reaction generating indole-3-pyruvate and glutamate. An aldolase catalyzed the second reaction in which pyruvate was reacted with indole-3-pyruvate, in the presence of Mg²⁺ and phosphate, generating the alpha-keto derivative of monatin (MP), 2-hydroxy-2-(indol-3-ylmethyl)-4-ketoglutaric acid. Transfer of the amino group from the glutamate formed in the first reaction produced the desired product, monatin. Purification and characterization of the product established that the isomer formed was S,S-monatin. Alternative substrates, enzymes, and conditions are described as well as improvements that were made to this process.

Enzymes

The aldolase, 4-hydroxy-4-methyl-2-oxoglutarate pyruvate lyase (ProA aldolase, *proA* gene) (EC 4.1.3.17) from *Comamonas testosteroni* was cloned, expressed and purified as described in Example 4. The 4-hydroxy-2-oxoglutarate glyoxylate lyases (KHG aldolases) (EC 4.1.3.16) from *B. subtilis*, *E. coli*, and *S. meliloti* were cloned, expressed and purified as described in Example 4.

The aminotransferases used in conjunction with the aldolases to produce monatin were L-aspartate aminotransferase encoded by the *E. coli aspC* gene, the tyrosine aminotransferase encoded by the *E. coli tyrB* gene, the *S. meliloti* TatA enzyme, the broad substrate aminotransferase encoded by the *L. major bsat* gene, or the glutamic-oxaloacetic transaminase from pig heart (Type IIa). The

cloning, expression and purification of the non-mammalian proteins are described in Example 1. Glutamic-oxaloacetic transaminase from pig heart (type IIa) was obtained from Sigma (# G7005).

Method using ProA aldolase and L-aspartate aminotransferase

5 The reaction mixture contained 50 mM ammonium acetate, pH 8.0, 4 mM MgCl₂, 3 mM potassium phosphate, 0.05 mM pyridoxal phosphate, 100 mM ammonium pyruvate, 50 mM tryptophan, 10 mM alpha-ketoglutarate, 160 mg of recombinant *C. testosteroni* ProA aldolase (unpurified cell extract, ~30% aldolase), 233 mg of recombinant *E. coli* L-aspartate aminotransferase (unpurified cell extract, ~40% aminotransferase) in one liter. All components except the enzymes
10 were mixed together and incubated at 30°C until the tryptophan dissolved. The enzymes were then added and the reaction solution was incubated at 30°C with gentle shaking (100 rpm) for 3.5 hours. At 0.5 and 1 hour after the addition of the enzymes aliquots of solid tryptophan (50 mmoles each) were added to the reaction. All of the added tryptophan did not dissolve, but the concentration was maintained at 50 mM or higher. After 3.5 hours, the solid tryptophan was filtered off. Analysis of
15 the reaction mixture by LC/MS using a defined amount of tryptophan as a standard showed that the concentration of tryptophan in the solution was 60.5 mM and the concentration of monatin was 5.81 mM (1.05 g).

 The following methods were used to purify the final product. Ninety percent of the clear solution was applied to a column of BioRad AG50W-X8 resin (225 mL; binding capacity of 1.7
20 meq/mL). The column was washed with water, collecting 300 mL fractions, until the absorbance at 280 nm was <5% of the first flow through fraction. The column was then eluted with 1 M ammonium acetate, pH 8.4, collecting 4 300-mL fractions. All 4 fractions contained monatin and were evaporated to 105 mL using a roto-evaporator with a tepid water bath. A precipitate formed as the volume reduced and was filtered off over the course of the evaporation process.

25 Analysis of the column fractions by LC/MS showed that 99% of the tryptophan and monatin bound to the column. The precipitate that formed during the evaporation process contained >97% tryptophan and <2% of monatin. The ratio of tryptophan to product in the supernatant was approximately 2:1.

 The supernatant (7 ml) was applied to a 100 mL Fast Flow DEAE Sepharose (Amersham
30 Biosciences) column previously converted to the acetate form by washing with 0.5 L 1 M NaOH, 0.2 L water, 1.0 L of 1.0 M ammonium acetate, pH 8.4, and 0.5 L water. The supernatant was loaded at <2 mL/min and the column was washed with water at 3-4 mL/min until the absorbance at 280 nm was ~0. Monatin was eluted with 100 mM ammonium acetate, pH 8.4, collecting 4 100-mL fractions.

35 Analysis of the fractions showed that the ratio of tryptophan to monatin in the flow through fractions was 85:15 and the ratio in the eluent fractions was 7:93. Assuming the extinction coefficient at 280 nm of monatin is the same as tryptophan, the eluent fractions contained 0.146

mmole of product. Extrapolation to the total 1 L reaction would produce ~2.4 mmoles (~710 mg) of monatin, for a recovery of 68%.

The eluent fractions from the DEAE Sepharose column were evaporated to <20 mL. An aliquot of the product was further purified by application to a C₈ preparative reversed-phase column using the same chromatographic conditions as those described in Example 10 for the analytical-scale monatin characterization. Waters Fractionlynx™ software was employed to trigger automated fraction collection of monatin based on detection of the $m/z = 293$ ion. The fraction from the C₈ column with the corresponding protonated molecular ion for monatin was collected, evaporated to dryness, and then dissolved in a small volume of water. This fraction was used for characterization of the product.

The resulting product was characterized using the following methods.

UV/Visible Spectroscopy. UV/visible spectroscopic measurements of monatin produced enzymatically were carried out using a Cary 100 Bio UV/visible spectrophotometer. The purified product, dissolved in water, showed an absorption maximum of 280 nm with a shoulder at 288 nm, characteristics typical of indole containing compounds.

LC/MS Analysis. Analyses of mixtures for monatin derived from the *in vitro* biochemical reactions were carried out as described in Example 10. A typical LC/MS analysis of monatin in an *in vitro* enzymatic synthetic mixture is illustrated in FIG. 5. The lower panel of FIG. 5 illustrates a selected ion chromatogram for the protonated molecular ion of monatin at $m/z = 293$. This identification of monatin in the mixture was corroborated by the mass spectrum illustrated in FIG. 6. Analysis of the purified product by LC/MS showed a single peak with a molecular ion of 293 and absorbance at 280 nm. The mass spectrum was identical to that shown in FIG. 6.

MS/MS Analysis. LC/MS/MS daughter ion experiments, as described in Example 10, were also performed on monatin. A daughter ion mass spectrum of monatin is illustrated in FIG. 7. Tentative structural assignments of all fragment ions labeled in FIG. 7 were made. These include fragment ions of $m/z = 275$ ($293 - \text{H}_2\text{O}$), 257 ($293 - (2 \times \text{H}_2\text{O})$), 230 ($275 - \text{COOH}$), 212 ($257 - \text{COOH}$), 168 (3-buta-1,3-dienyl-1H-indole carbonium ion), 158 (1H-indole-3-carbaldehyde carbonium ion), 144 (3-ethyl-1H-indole carbonium ion), 130 (3-methylene-1H-indole carbonium ion), and 118 (indole carbonium ion). Many of these are the same as those obtained for MP (Example 4), as expected if derived from the indole portion of the molecule. Some are 1 mass unit higher than those seen for MP, due to the presence of an amino group instead of a ketone.

High Resolution MS analysis. FIG. 8 illustrates the mass spectrum obtained for purified monatin employing an Applied Biosystems-Perkin Elmer Q-Star hybrid quadrupole/time-of-flight mass spectrometer. The measured mass for protonated monatin using tryptophan as an internal mass calibration standard was 293.1144. The calculated mass of protonated monatin, based on the elemental composition C₁₄H₁₇N₂O₅ is 293.1137. This is a mass measurement error of less than 2 parts per million (ppm), providing conclusive evidence of the elemental composition of monatin produced enzymatically.

NMR Spectroscopy. The NMR experiments were performed on a Varian Inova 500 MHz instrument. The sample of monatin (~3 mg) was dissolved in 0.5 ml of D₂O. Initially, the solvent (D₂O) was used as the internal reference at 4.78 ppm. Since the peak for water was large, the ¹H-NMR was run with suppression of the peak for water. Subsequently, due to the broadness of the water peak, the C-2 proton of monatin was used as the reference peak, and set at the published value of 7.192 ppm.

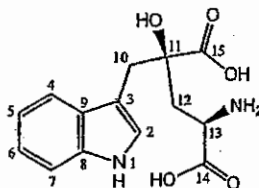
For ¹³C-NMR, an initial run of several hundred scans indicated that the sample was too dilute to obtain an adequate ¹³C spectrum in the allotted time. Therefore, a heteronuclear multiple quantum coherence (HMQC) experiment was performed, which enabled the correlation of the hydrogens and the carbons to which they were attached, and also providing information on the chemical shifts of the carbons.

A summary of the ¹H and HMQC data is shown in Tables 2 and 3. By comparison to published values, the NMR data indicated that the enzymatically produced monatin was either (S,S), (R,R), or a mixture of both.

Chiral LC/MS Analysis. To establish that the monatin produced *in vitro* was one isomer, and not a mixture of the (R,R) and (S,S) enantiomers, chiral LC/MS analyses were carried out using the instrumentation described in Example 10.

Chiral LC separations were made using an Chirobiotic T (Advanced Separations Technology) chiral chromatography column at room temperature. Separation and detection, based on published protocols from the vendor, were optimized for the R- (D) and S- (L) isomers of tryptophan. The LC mobile phase consisted of A) water containing 0.05% (v/v) trifluoroacetic acid; B) Methanol containing 0.05% (v/v) trifluoroacetic acid. The elution was isocratic at 70% A and 30% B. The flow rate was 1.0 mL/min, and PDA absorbance was monitored from 200 nm to 400 nm. The instrumental parameters used for chiral LC/MS analysis of tryptophan and monatin are identical to those described in Example 10 for LC/MS analysis. Collection of mass spectra for the region *m/z* 150–400 was utilized. Selected ion chromatograms for protonated molecular ions ([M + H]⁺ = 205 for both R- and S-tryptophan and [M + H]⁺ = 293 for monatin) allowed direct identification of these analytes in the mixtures.

The chromatograms of R- and S-tryptophan and monatin, separated by chiral chromatography and monitored by MS, are shown in FIG. 9. The single peak in the chromatogram of monatin indicates that the compound is one isomer, with a retention time almost identical to S-tryptophan.

Table 2: ^1H NMR data

Atom	Cargill		Vleggaar et al. ¹		Takeshi et al. ²	
	δ_{H}	J(HH) Hz	δ_{H}	J(HH) Hz	δ_{H}	J(HH) Hz
2	7.192 (1H, s)		7.192 (s)		7.18 (s)	
4	7.671 (d)	7.99	7.686 (d)	7.9	7.67 (d)	8.0
5	7.104 (dd)	7.99	7.102 (dd)	8.0, 8.0	7.11 (dd)	7.5, 7.5
6	7.178 (dd)	*	7.176 (dd)	8.0, 8.0	7.17 (dd)	7.5, 7.5
7	7.439 (d)	7.99	7.439 (d)	8.1	7.43 (d)	8.0
10a	3.242 (d)	14.5	3.243 (d)	14.3	3.24 (d)	14.5
10b	3.033 (d)	14.5	3.051 (d)	14.3	3.05 (d)	14.5
12	2.626 (dd)	15.5, 1.5	2.651 (dd)	15.3, 1.7	2.62 (dd)	15.5, 1.8
	2.015 (dd)	15.0, 12.0	2.006 (dd)	15.3, 11.7	2.01 (dd)	15.5, 12.0
13	3.571 (dd)	10.75*, 1.5	3.168 (dd)	11.6, 1.8	3.57 (dd)	12.0, 1.8

¹ Vleggaar et al. (*J.C.S. Perkin Trans. 1*:3095-8, 1992).5 ² Takeshi and Shusuke (JP2002060382, 2002-02-26).Table 3: ^{13}C NMR data (from HMQC spectrum)

Atom	Cargill		Vleggaar et al. ¹	
	δ_{C}		δ_{C}	
2	126.1		126.03	
3	*		110.31	
4	120.4		120.46	
5	120.2		120.25	
6	122.8		122.74	
7	112.8		112.79	
8	*		137.06	
9	*		129.23	
10a	36.4		36.53	
12	39.5		39.31	
13	54.9		54.89	
14	*		175.30	
15	*		181.18	

¹ Vleggaar et al. (*J.C.S. Perkin Trans. 1*:3095-8, 1992).

10

Polarimetry. The optical rotation was measured on a Rudolph Autopol III polarimeter.

The monatin was prepared as a 14.6 mg/mL solution in water. The expected specific rotation ($[\alpha]_{\text{D}}^{20}$) for S,S monatin (salt form) is -49.6 for a 1 g/mL solution in water (Vleggaar et al). The observed $[\alpha]_{\text{D}}^{20}$ was -28.1 for the purified, enzymatically produced monatin indicating that it was the S, S isomer.

Improvements

The reaction conditions, including reagent and enzyme concentrations, were optimized and yields of 10 mg/mL were produced using the following reagent mix:

50 mM ammonium acetate pH 8.3, 2 mM MgCl₂, 200 mM pyruvate (sodium or ammonium salt), 5
5 mM alpha-ketoglutarate (sodium salt), 0.05 mM pyridoxal phosphate,
deaerated water to achieve a final volume of 1 mL after the addition of the enzymes,
3 mM potassium phosphate, 50 µg/mL of recombinant ProA aldolase (cell extract; total protein
concentration of 167 µg/mL), 1000 µg/mL of L-aspartate aminotransferase encoded by the *E. coli*
aspC gene (cell extract; total protein concentration of 2500 µg/mL), and solid tryptophan to afford a
10 concentration of > 60 mM (saturated; some undissolved throughout the reaction). The mixture was
incubated at 30°C for 4 hours with gentle stirring or mixing.

Substitutions

The concentration of alpha-ketoglutarate can be reduced to 1 mM and supplemented with 9
mM aspartate with an equivalent yield of monatin. Alternative amino acid acceptors can be utilized
15 in the first step, such as oxaloacetate.

When recombinant *L. major* broad substrate aminotransferase was used in place of the *E.*
coli L-aspartate aminotransferase, similar yields of monatin were achieved. However, a second
unidentified product (3-10% of the major product) with a molecular mass of 292 was also detected by
LC-MS analysis. Monatin concentrations of 0.1-0.5 mg/mL were produced when the *E. coli tyrB*
20 encoded enzyme, the *S. meliloti tat A* encoded enzyme or the glutamic-oxaloacetic transaminase from
pig heart (type IIa) was added as the aminotransferase. When starting the reaction from indole-3-
pyruvate, a reductive amination can be done for the last step with glutamate dehydrogenase and
NADH (as in Example 7).

The KHG aldolases from *B. subtilis*, *E. coli*, and *S. meliloti* were also used with the *E. coli* L-
25 aspartate aminotransferase to produce monatin enzymatically. The following reaction conditions
were used: 50 mM NH₄-OAc pH 8.3, 2 mM MgCl₂, 200 mM pyruvate, 5 mM glutamate, 0.05 mM
pyridoxal phosphate, deaerated water to achieve a final volume of 0.5 mL after the addition of the
enzymes, 3 mM potassium phosphate, 20 µg/mL of recombinant *B. subtilis* KHG aldolase (purified),
ca. 400 µg/mL of *E. coli* L-aspartate aminotransferase (AspC) unpurified from cell extract, and 12
30 mM indole-3-pyruvate. The reactions were incubated at 30°C for 30 minutes with shaking. The
amount of monatin produced using the *B. subtilis* enzyme was 80 ng/mL, and increased with
increasing amounts of aldolase. If indole-3-pyruvate and glutamate were replaced by saturating
amounts of tryptophan and 5 mM alpha-ketoglutarate, the production of monatin was increased to
360 ng/mL. Reactions were repeated with 30 µg/mL of each of the three KHG enzymes in 50 mM
35 Tris pH 8.3, with saturating amounts of tryptophan, and were allowed to proceed for an hour in order
to increase detection. The *Bacillus* enzyme had the highest activity as in Example 4, producing
approximately 4000 ng/mL monatin. The *E. coli* KHG produced 3000 ng/mL monatin, and the *S.*
meliloti enzyme produced 2300 ng/mL.

Example 7

Interconversion between MP and Monatin

The amination of MP to form monatin can be catalyzed by aminotransferases such as those identified in Examples 1 and 6, or by dehydrogenases that require a reducing cofactor such as NADH or NADPH. These reactions are reversible and can be measured in either direction. The directionality, when using a dehydrogenase enzyme, can be largely controlled by the concentration of ammonium salts.

Dehydrogenase activity. The oxidative deamination of monatin was monitored by following the increase in absorbance at 340 nm as NAD(P)⁺ was converted to the more chromophoric NAD(P)H. Monatin was enzymatically produced and purified as described in Example 6.

A typical assay mixture contained 50 mM Tris-HCl, pH 8.0 to 8.9, 0.33 mM NAD⁺ or NADP⁺, 2 to 22 units of glutamate dehydrogenase (Sigma), and 10-15 mM substrate in 0.2 mL. The assay was performed in duplicate in a UV-transparent microtiter plate, on a Molecular Devices SpectraMax Plus platereader. A mix of the enzyme, buffer, and NAD(P)⁺ were pipetted into wells containing the substrate and the increase in absorbance at 340 nm was monitored at 10 second intervals after brief mixing. The reaction was incubated at 25°C for 10 minutes. Negative controls were carried out without the addition of substrate, and glutamate was utilized as a positive control. The type III glutamate dehydrogenase from bovine liver (Sigma # G-7882) catalyzed the conversion of the monatin to the monatin precursor at a rate of conversion approximately one-hundredth the rate of the conversion of glutamate to alpha-ketoglutarate.

Transamination activity. Monatin aminotransferase assays were conducted with the aspartate aminotransferase (AspC) from *E. coli*, the tyrosine aminotransferase (TyrB) from *E. coli*, the broad substrate aminotransferase (BSAT) from *L. major*, and the two commercially available porcine glutamate-oxaloacetate aminotransferases described in Example 1. Both oxaloacetate and alpha-ketoglutarate were tested as the amino acceptor. The assay mixture contained (in 0.5 mL) 50 mM Tris-HCl, pH 8.0, 0.05 mM PLP, 5 mM amino acceptor, 5 mM monatin, and 25 µg of aminotransferase. The assays were incubated at 30°C for 30 minutes, and the reactions were stopped by addition of 0.5 mL isopropyl alcohol. The loss of monatin was monitored by LC/MS (Example 10). The highest amount of activity was noted with *L. major* BSAT with oxaloacetate as the amino acceptor, followed by the same enzyme with alpha-ketoglutarate as the amino acceptor. The relative activity with oxaloacetate was: BSAT > AspC > porcine type IIa > porcine type I = TyrB. The relative activity with alpha-ketoglutarate was: BSAT > AspC > porcine type I > porcine type IIa > TyrB.

Example 8

Production of Monatin from Tryptophan and C3 Sources Other than Pyruvate

As described above in Example 6, indole-3-pyruvate or tryptophan can be converted to monatin using pyruvate as the C3 molecule. However, in some circumstances, pyruvate may not be a desirable raw material. For example, pyruvate may be more expensive than other C3 carbon sources, or may have adverse effects on fermentations if added to the medium. Alanine can be transaminated by many PLP-enzymes to produce pyruvate.

Tryptophanase-like enzymes perform beta-elimination reactions at faster rates than other PLP enzymes such as aminotransferases. Enzymes from this class (4.1.99.-) can produce ammonia and pyruvate from amino acids such as L-serine, L-cysteine, and derivatives of serine and cysteine with good leaving groups such as O-methyl-L-serine, O-benzyl-L-serine, S-methylcysteine, S-benzylcysteine, S-alkyl-L-cysteine, O-acyl-L-serine, 3-chloro-L-alanine.

Processes to produce monatin using EC 4.1.99.- polypeptides can be improved by mutating the β -tyrosinase (TPL) or tryptophanase according to the method of Mouratou *et al.* (*J. Biol. Chem* 274:1320-5, 1999). Mouratou *et al.* describe the ability to convert the β -tyrosinase into a dicarboxylic amino acid β -lyase, which has not been reported to occur in nature. The change in specificity was accomplished by converting valine (V) 283 to arginine (R) and arginine (R) 100 to threonine (T). These amino acid changes allow for the lyase to accept a dicarboxylic amino acid for the hydrolytic deamination reaction (such as aspartate). Aspartate, therefore, can also be used as a source of pyruvate for subsequent aldol condensation reactions.

Additionally, cells or enzymatic reactors can be supplied with lactate and an enzyme that converts lactate to pyruvate. Examples of enzymes capable of catalyzing this reaction include lactate dehydrogenase and lactate oxidase.

Isolation of Genomic DNA

Tryptophanase polypeptides have previously been reported in, for example, Mouratou *et al.* (*JBC* 274:1320-5, 1999). To isolate genes that encode tryptophanase polypeptides, genomic DNA from *E. coli* DH10B was used as a template for PCR as described in Example 1.

The gene for tyrosine-phenol lyase was isolated from *C. freundii* (ATCC catalog number 8090, Designation ATCC 13316; NCTC 9750) and grown on Nutrient agar (Difco 0001) and nutrient broth (Difco 0003) at 37°C to an OD of 2.0. The genomic DNA was purified using a Qiagen Genomic-tip™ 100/G kit.

PCR Amplification of Coding Sequences

Primers were designed with compatible overhangs for the pET 30 Xa/LIC vector (Novagen, Madison, WI) as described above in Example 1.

E. coli tna (SEQ ID NO: 41). N-terminal primer for pET30 Xa/LIC cloning: 5'-GGT ATT GAG GGT CGC ATG GAA AAC TTT AAA CAT CT-3' (SEQ ID NO: 43). C-terminal primer for

pET30 Xa/LIC cloning: 5'-AGA GGA GAG TTA GAG CCT TAA ACT TCT TTA AGT TTT G-3' (SEQ ID NO: 44).

C. freundii *tpl* (SEQ ID NO: 42). N-terminal primer for pET30 Xa/LIC cloning: 5'-GGT ATT GAG GGT CGC ATGAATTATCCGGCAGAACC-3' (SEQ ID NO: 45). C-terminal primer for pET 30 Xa/LIC cloning: 5'-AGA GGA GAG TTA GAG CCTTAGATGTAATCAAAGCGTG-3' (SEQ ID NO: 46).

The Eppendorf Mastercycler™ Gradient 5331 Thermal Cycler was used for all PCR reactions. In 50 µL was added 0.5 µg template (genomic DNA), 1.0 µM of each primer, 0.4 mM each dNTP, 3.5 U Expand High Fidelity Polymerase (Roche), 1X Expand buffer with Mg, and 5% DMSO (final concentration). The thermocycler PCR program used was as follows: 96°C hot start (5 minutes), 94°C - 30 seconds, 40-60°C - 1 minute 45 seconds, 72°C - 2 minutes 15 seconds; 30 repetitions. The final polymerization step was for 7 minutes, and the samples were then stored at 4°C.

15 Cloning

Cloning and positive clone identification procedures detailed above in Example 1 were used to identify the appropriate clones.

Gene Expression and Activity Assays

20 Plasmid DNA (verified by sequence analysis) was subcloned into the expression host BL21(DE3) (Novagen). The cultures were grown in LB medium with 30 mg/L kanamycin, the plasmids were isolated using a Qiagen miniprep kit, and analyzed by restriction digest to confirm identity.

25 Induction experiments were done with the BL21(DE3) expression host, the constructs were grown in LB medium containing 50 mg/L kanamycin at 37°C. Protein expression was induced using 0.1 mM IPTG after the OD₆₀₀ of the culture reached approximately 0.6. The cells were grown for 4 hours at 30°C and harvested by centrifugation. The cells were then lysed in 5 mL/g wet cell weight BugBuster™ (Novagen) reagent containing 5 µL/mL protease inhibitor cocktail set #III (Calbiochem) and 1 µL/mL benzonase nuclease (Novagen), and the His-tagged recombinant proteins 30 were purified using the His-Bind cartridges as described above in Example 1. Purified proteins were desalted on a PD-10 (G25 Sephadex, Amersham Biosciences) column and eluted in 100 mM Tris-Cl buffer, pH 8.0. The proteins were analyzed by SDS-PAGE on 4-15% gradient gels to check for soluble protein levels at the predicted MW of the recombinant fusion protein.

35 Mutagenesis

Some members of polypeptide class 4.1.99.- (tryptophanase and β-tyrosinase) will perform the beta-lyase reaction with aspartate or similar amino acids without any modification. However, some members of the class may need to be mutagenized to allow for the use of the substrates and/or

the creation of the product. Moreover, in some cases polypeptides that can perform the conversion may be further optimized by mutagenesis.

Site directed mutagenesis was performed based on 3D structure analysis of PLP-binding polypeptides. Two examples for changing the substrate specificity of the polypeptides are shown below.

Mutagenesis of Tryptophanase Example 1

The mutagenesis protocol provided below introduced two point mutations in the amino acid sequence. The first point mutation changed arginine (R) at position 103 to threonine (T) and the second point mutation changed valine (V) at position 299 to arginine (R) (numbering system for *E. coli* mature protein). Mutagenesis experiments were performed by ATG Laboratories (Eden Prairie, MN). Mutations were introduced sequentially by PCR of gene fragments and reassembly of the fragments was accomplished by PCR as well. Primers for converting arginine (R)103 to threonine (T): 5'-CCAGGGCACCGGCGCAGAGCAAATCTATATT-3' (SEQ ID NO: 47) and 5'-TGCGCCGGTGCCCTGGTGAGTCGGAATGGT-3' (SEQ ID NO: 48).

Primers for converting valine (V)299 to arginine (R): 5'-TCCTGCACGCGGCAAAGGGTTCTGCACTCGGT-3' (SEQ ID NO: 49) and 5'-CTTTGCCGCGTGCAGGAAGGCTTCCCGACA-3' (SEQ ID NO: 50).

Mutants were screened by restriction digest with Xba I/HindIII and SphI, and verified by sequencing.

Mutagenesis of Tyrosine Phenol Lyase (β -tyrosinase) Example 2

Two point mutations were made to the tyrosine phenol lyase amino acid sequence. These mutations converted arginine (R) at position 100 to threonine (T) and valine (V) at position 283 to arginine (R) (in *C. freundii* mature protein sequence).

Primers for the R100T conversion were: 5'-AGGGGACCGGCGCAGAAAACCTGTTATCG-3' (SEQ ID NO: 51) and 5'-AGGGGACCGGCGCAGAAAACCTGTTATCG-3' (SEQ ID NO: 52). Primers for the V283R conversion were: 5'-GTTAGTCCGCGTCTACGAAGGGATGCCAT-3' (SEQ ID NO: 53) and 5'-GTAGACGCGGACTAACTCTTTGGCAGAAG-3' (SEQ ID NO: 54).

The methods described above were used, and the clones were screened by KpnI/SacI digestion, and BstX I digestion. The sequences were verified by dideoxy chain termination sequencing. Recombinant protein was produced as described above for the wildtype enzymes.

The reaction mixture consisted of 50 mM Tris-Cl pH 8.3, 2 mM MgCl₂, 200 mM C3 carbon source, 5 mM alpha-ketoglutarate, sodium salt, 0.05 mM pyridoxal phosphate, deaerated water to achieve a final volume of 0.5 mL after the addition of the enzymes, 3 mM potassium phosphate pH 7.5, 25 μ g of crude recombinant *C. testosteroni* ProA aldolase as prepared as in Example 4, 500 μ g of crude L-aspartate aminotransferase (AspC) as prepared in Example 1, and solid tryptophan to afford

concentration of > 60 mM (saturated; some undissolved throughout the reaction). The reaction mix was incubated at 30°C for 30 minutes with mixing. Serine, alanine, and aspartate were supplied as 3-carbon sources. Assays were performed with and without secondary PLP enzymes (purified) capable of performing beta-elimination and beta-lyase reactions (tryptophanase (TNA), double mutant tryptophanase, β -tyrosinase (TPL)). The results are shown in Table 4:

Table 4: Production of monatin utilizing alternative C3-carbon sources

C3-carbon source	Additional PLP enzyme	Relative Activity
none	none	0%
pyruvate	none	100%
serine	none	3%
serine	11 μ g wildtype TNA (1 U)	5.1%
serine	80 μ g double mutant TNA	4.6%
alanine	none	32%
alanine	11 μ g wildtype TNA	41.7%
alanine	80 μ g mutant TNA	43.9%
aspartate	110 μ g wildtype TNA (10 U)	7.7%
aspartate	5 U wildtype TPL (crude)	5.1%
aspartate	80 μ g mutant TNA	3.3%

The monatin produced from alanine and serine as 3-carbon sources was verified by LC/MS/MS daughter scan analysis, and was identical to the characterized monatin produced in Example 6. Alanine was the best alternative tested, and was transaminated by the AspC enzyme. The amount of monatin produced was increased by addition of the tryptophanase, which is capable of transamination as a secondary activity. The amount of monatin produced with serine as a carbon source nearly doubled with the addition of the tryptophanase enzymes, even though only one-fifth of the amount of tryptophanase was added in comparison to the aminotransferase. AspC is capable of some amount of beta-elimination activity alone. The results with aspartate indicate that the tryptophanase activity on aspartate does not increase with the same site-directed mutations as previously suggested for β -tyrosinase. It is expected that the mutant β -tyrosinase will have higher activity for production of monatin.

Example 9

Chemical Synthesis of Monatin

The addition of alanine to indole-3-pyruvic acid produces monatin, and this reaction can be performed synthetically with a Grignard or organolithium reagent.

For example, to 3-chloro- or 3-bromo-alanine which has been appropriately blocked at the carboxyl and amino groups, is added magnesium under anhydrous conditions. Indole-3-pyruvate (appropriately blocked) is then added to form the coupled product followed by removal of the protecting groups to form monatin. Protecting groups that are particularly useful include THP (tetrahydropyranyl ether) which is easily attached and removed.

Example 10

Detection of Monatin and MP

This example describes methods used to detect the presence of monatin, or its precursor 2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid.

5 LC/MS Analysis

Analyses of mixtures for the alpha-keto acid form of monatin (monatin precursor, MP) and monatin derived from *in vitro* or *in vivo* biochemical reactions were performed using a Waters/Micromass liquid chromatography-tandem mass spectrometry (LC/MS/MS) instrument including a Waters 2690 liquid chromatograph with a Waters 996 Photo-Diode Array (PDA) absorbance monitor placed in series between the chromatograph and a Micromass Quattro Ultima triple quadrupole mass spectrometer. LC separations were made using a Supelco Discovery C₁₈ reversed-phase chromatography column, 2.1mm x 150 mm, or an Xterra MS C₈ reversed-phase chromatography column, 2.1mm x 250 mm, at room temperature. The LC mobile phase consisted of A) water containing 0.05% (v/v) trifluoroacetic acid and B) methanol containing 0.05% (v/v) trifluoroacetic acid.

The gradient elution was linear from 5% B to 35% B, 0-9 min, linear from 35% B to 90% B, 9-16 min, isocratic at 90% B, 16-20 min, linear from 90% B to 5% B, 20-22 min, with a 10 min re-equilibration period between runs. The flow rate was 0.25 mL/min, and PDA absorbance was monitored from 200 nm to 400 nm. All parameters of the ESI-MS were optimized and selected based on generation of protonated molecular ions ($[M + H]^+$) of the analytes of interest, and production of characteristic fragment ions.

The following instrumental parameters were used for LC/MS analysis of monatin: Capillary: 3.5 kV; Cone: 40 V; Hex 1: 20 V; Aperture: 0 V; Hex 2: 0 V; Source temperature: 100°C; Desolvation temperature: 350°C; Desolvation gas: 500 L/h; Cone gas: 50 L/h; Low mass resolution (Q1): 15.0; High mass resolution (Q1): 15.0; Ion energy: 0.2; Entrance: 50V; Collision Energy: 2; Exit: 50V; Low mass resolution (Q2): 15; High mass resolution (Q2): 15; Ion energy (Q2): 3.5; Multiplier: 650. Uncertainties for reported mass/charge ratios (m/z) and molecular masses are $\pm$ 0.01%. Initial detection of the alpha-keto acid form of monatin (MP) and monatin in the mixtures was accomplished by LC/MS monitoring with collection of mass spectra for the region m/z 150-400. Selected ion chromatograms for protonated molecular ions ($[M + H]^+ = 292$ for MP, $[M + H]^+ = 293$ for monatin) allowed direct identification of these analytes in the mixtures.

MS/MS Analysis

LC/MS/MS daughter ion experiments were performed on monatin as follows. A daughter ion analysis involves transmission of the parent ion (e.g., $m/z = 293$ for monatin) of interest from the first mass analyzer (Q1) into the collision cell of the mass spectrometer, where argon is introduced and chemically dissociates the parent into fragment (daughter) ions. These fragment ions are then

detected with the second mass analyzer (Q2), and can be used to corroborate the structural assignment of the parent.

The following instrumental parameters were used for LC/MS/MS analysis of monatin:

Capillary: 3.5 kV; Cone: 40 V; Hex 1: 20 V; Aperture: 0 V; Hex 2: 0 V; Source temperature: 100 °C;
Desolvation temperature: 350 °C; Desolvation gas: 500 L/h; Cone gas: 50 L/h; Low mass resolution (Q1): 13.0; High mass resolution (Q1): 13.0; Ion energy: 0.2; Entrance: -5 V; Collision Energy: 14; Exit: 1V; Low mass resolution (Q2): 15; High mass resolution (Q2): 15; Ion energy (Q2): 3.5; Multiplier: 650.

10 **High-Throughput Determination of Monatin**

High-throughput analyses (< 5 min/sample) of mixtures for monatin derived from *in vitro* or *in vivo* reactions were carried out using instrumentation described above, and the same parameters as described for LC/MS/MS. LC separations were made using Waters Xterra MS C₈ (2.1mm x 50 mm) chromatography at room temperature with isocratic elution in 15% aqueous MeOH, 0.25% acetic acid at a flow rate of 0.3 mL/min. Detection of monatin in the mixtures was accomplished using selected reaction monitoring (SRM)-tandem mass spectrometry. This involved monitoring specific collisionally-induced parent ion ($[M + H]^+ = 293.1$) to daughter ion (e.g., the fragment ion at $m/z = 168.1$, tentatively assigned as a 3-buta-1,3-dienyl-1*H*-indole carbonium ion) transitions to maximize sensitivity, selectivity, and throughput for the detection of monatin. PDA absorbance data were collected in parallel for further verification of monatin identity.

Example 11

Production of Monatin in Bacteria

This example describes methods used to produce monatin in *E. coli* cells. One skilled in the art will understand that similar methods can be used to produce monatin in other bacterial cells. In addition, vectors containing other genes in the monatin synthesis pathway (FIG. 2) can be used.

Trp-1 + glucose medium, a minimal medium that has been used for increased production of tryptophan in *E. coli* cells (Zeman *et al. Folia Microbiol.* 35:200-4, 1990), was prepared as follows. To 700 mL nanopure water the following reagents were added: 2 g (NH₄)₂SO₄, 13.6 g KH₂PO₄, 0.2 g MgSO₄·7H₂O, 0.01 g CaCl₂·2H₂O, and 0.5 mg FeSO₄·7H₂O. The pH was adjusted to 7.0, the volume was increased to 850 mL, and the medium was autoclaved. A 50% glucose solution was prepared separately, and sterile-filtered. Forty mL was added to the base medium (850 mL) for a 1 L final volume.

A 10 g/L L-tryptophan solution was prepared in 0.1 M sodium phosphate pH 7, and sterile-filtered. One-tenth volume was typically added to cultures as specified below. A 10% sodium pyruvate solution was also prepared and sterile-filtered. A 10 mL aliquot was typically used per liter of culture. Stocks of ampicillin (100 mg/mL), kanamycin (25 mg/mL) and IPTG (840 mM) were prepared, sterile-filtered, and stored at -20°C before use. Tween 20 (polyoxyethylene 20-Sorbitan

monolaurate) was utilized at a 0.2% (vol/vol) final concentration. Ampicillin was used at non-lethal concentrations, typically 1-10 µg/mL final concentration.

Fresh plates of *E. coli* BL21(DE3)::*C. testosteroni proA/pET 30 Xa/LIC* (described in Example 4) were prepared on LB medium containing 50 µg/mL kanamycin. Overnight cultures (5 mL) were inoculated from a single colony and grown at 30°C in LB medium with kanamycin. Typically a 1 to 50 inoculum was used for induction in trp-1 + glucose medium. Fresh antibiotic was added to a final concentration of 50 mg/L. Shake flasks were grown at 37°C prior to induction.

Cells were sampled every hour until an OD₆₀₀ of 0.35-0.8 was obtained. Cells were then induced with 0.1 mM IPTG, and the temperature reduced to 34 °C. Samples (1 mL) were collected prior to induction (zero time point) and centrifuged at 5000 x g. The supernatant was frozen at -20°C for LC/MS analysis. Four hours post-induction, another 1 mL sample was collected, and centrifuged to separate the broth from the cell pellet. Tryptophan, sodium pyruvate, ampicillin, and Tween were added as described above.

The cells were grown for 48 hours post-induction, and another 1 mL sample was taken and prepared as above. At 48 hours, another aliquot of tryptophan and pyruvate were added. The entire culture volume was centrifuged after approximately 70 hours of growth (post-induction), for 20 minutes at 4°C and 3500 rpm. The supernatant was decanted and both the broth and the cells were frozen at -80°C. The broth fractions were filtered and analyzed by LC/MS. The heights and areas of the $[M+H]^+ = 293$ peaks were monitored as described in Example 10. The background level of the medium was subtracted. The data was also normalized for cell growth by plotting the height of the $[M+H]^+ = 293$ peak divided by the optical density of the culture at 600 nm.

Higher levels of monatin were produced when pyruvate, ampicillin, and Tween were added 4 hours post induction rather than at induction. Other additives such as PLP, additional phosphate, or additional MgCl₂ did not increase the production of monatin. Higher titers of monatin were obtained when tryptophan was utilized instead of indole-3-pyruvate, and when the tryptophan was added post-induction rather than at inoculation, or at induction. Prior to induction, and 4 hours post-induction (at time of substrate addition), there was typically no detectable level of monatin in the fermentation broth or cellular extracts. Negative controls were done utilizing cells with pET30a vector only, as well as cultures where tryptophan and pyruvate were not added. A parent MS scan demonstrated that the compound with $(m+1)/z = 293$ was not derived from larger molecules, and daughter scans (performed as in Example 10) were similar to monatin made *in vitro*.

The effect of Tween was studied by utilizing 0, 0.2% (vol/vol), and 0.6% final concentrations of Tween-20. The highest amount of monatin produced by shake flasks was at 0.2% Tween. The ampicillin concentration was varied between 0 and 10 µg/mL. The amount of monatin in the cellular broth increased rapidly (2.5 X) between 0 and 1 µg/mL, and increased 1.3 X when the ampicillin concentration was increased from 1 to 10 µg/mL.

A time course experiment showing typical results is shown in FIG. 10. The amount of monatin secreted into the cell broth increased, even when the values are normalized for cell growth.

By using the molar extinction coefficient of tryptophan, the amount of monatin in the broth was estimated to be less than 10 µg/mL. The same experiment was repeated with the cells containing vector without *proA* insert. Many of the numbers were negative, indicating the peak height at $m/z=293$ was less in these cultures than in the medium alone (FIG. 10). The numbers were
 5 consistently lower when tryptophan and pyruvate were absent, demonstrating that monatin production is a result of an enzymatic reaction catalyzed by the aldolase enzyme.

The *in vivo* production of monatin in bacterial cells was repeated in 800 mL shake flask experiments and in fermentors. A 250 mL sample of monatin (in cell-free broth) was purified by anion exchange chromatography and preparative reverse-phase liquid chromatography. This sample
 10 was evaporated, and submitted for high resolution mass analysis (described in Example 6). The high resolution MS indicated that the metabolite being produced is monatin.

In vitro assays indicate that aminotransferase needs to be present at higher levels than aldolase (see Example 6), therefore the aspartate aminotransferase from *E. coli* was overexpressed in combination with the aldolase gene to increase the amount of monatin produced. Primers were
 15 designed to introduce *C. testosteroni proA* into an operon with *aspC*/pET30 Xa/LIC, as follows: 5' primer: ACTCGGATCCGAAGGAGATATACATATGTACGAACTGGGACT (SEQ ID NO: 67) and 3' primer: CGGCTGTCGACCGTTAGTCAATATATTTTCAGGC (SEQ ID NO: 68). The 5' primer contains a BamHI site, the 3' primer contains a SalI site for cloning. PCR was performed as described in Example 4, and gel purified. The *aspC*/pET30 Xa/LIC construct was digested with
 20 BamHI and SalI, as was the PCR product. The digests were purified using a Qiagen spin column. The *proA* PCR product was ligated to the vector using the Roche Rapid DNA Ligation kit (Indianapolis, IN) according to manufacturer's instructions. Chemical transformations were done using Novablues Singles (Novagen) as described in Example 1. Colonies were grown up in LB medium containing 50 mg/L kanamycin and plasmid DNA was purified using the Qiagen spin
 25 miniprep kit. Clones were screened by restriction digest analysis and sequence was confirmed by Seqwright (Houston, TX). Constructs were subcloned into BLR(DE3), BLR(DE3)pLysS, BL21(DE3) and BL21(DE3)pLysS (Novagen). The *proA*/pET30 Xa/LIC construct was also transformed into BL21(DE3)pLysS.

Initial comparisons of BLR(DE3) shake flask samples under the standard conditions
 30 described above demonstrated that the addition of the second gene (*aspC*) improved the amount of monatin produced by seven-fold. To hasten growth, BL21(DE3)-derived host strains were used. The *proA* clones and the two gene operon clones were induced in Trp-1 medium as above, the pLysS hosts had chloramphenicol (34 mg/L) added to the medium as well. Shake flask experiments were performed with and without the addition of 0.2% Tween-20 and 1 mg/L ampicillin. The amount of
 35 monatin in the broth was calculated using *in vitro* produced purified monatin as a standard. SRM analyses were performed as described in Example 10. Cells were sampled at zero, 4 hours, 24 hours, 48 hours, 72 hours, and 96 hours of growth.

The results are shown in Table 5 for the maximum amounts produced in the culture broths. In most instances, the two gene construct gave higher values than the *proA* construct alone. The

pLysS strains, which should have leakier cell envelopes, had higher levels of monatin secreted, even though these strains typically grow at a slower rate. The additions of Tween and ampicillin were beneficial.

5 **Table 5: Amount of Monatin Produced by *E. coli* Bacteria**

Construct	Host	Tween + Amp	µg/mL monatin	time
<i>proA</i>	BL21(DE3)	-	0.41	72 hr
<i>proA</i>	BL21(DE3)	+	1.58	48 hr
<i>proA</i>	BL21(DE3)pLysS	-	1.04	48 hr
<i>proA</i>	BL21(DE3)pLysS	+	1.60	48 hr
<i>aspC:proA</i>	BL21(DE3)	-	0.09	48 hr
<i>aspC:proA</i>	BL21(DE3)	+	0.58	48 hr
<i>aspC:proA</i>	BL21(DE3)pLysS	-	1.39	48 hr
<i>aspC:proA</i>	BL21(DE3)pLysS	+	6.68	48 hr

Example 12

Production of Monatin in Yeast

10 This example describes methods used to produce monatin in eukaryotic cells. One skilled in the art will understand that similar methods can be used to produce monatin in any cell of interest. In addition, other genes can be used (e.g., those listed in FIG. 2) in addition to, or alternatively to those described in this example.

The pESC Yeast Epitope Tagging Vector System (Stratagene, La Jolla, CA) was used to clone and express the *E. coli aspC* and *C. testosteroni proA* genes into *Saccharomyces cerevisiae*.
 15 The pESC vectors contain both the GAL1 and the GAL10 promoters on opposite strands, with two distinct multiple cloning sites, allowing for expression of two genes at the same time. The pESC-His vector also contains the *His3* gene for complementation of histidine auxotrophy in the host (YPH500). The GAL1 and GAL10 promoters are repressed by glucose and induced by galactose; a Kozak sequence is utilized for optimal expression in yeast. The pESC plasmids are shuttle vectors,
 20 allowing the initial construct to be made in *E. coli* (with the *bla* gene for selection); however, no bacterial ribosome binding sites are present in the multiple cloning sites.

The following primers were designed for cloning into pESC-His (restriction sites are underlined, Kozak sequence is in bold): *aspC* (*Bam*HI/*Sal*I), GAL1: 5'-
 CGCGGATCCATAATGGTTGAGAACATTACCG-3' (SEQ ID NO: 69) and 5'-
 25 ACGCGTCGACTTACAGCACTGCCACAATCG-3' (SEQ ID NO: 70). *proA* (*Eco*RI/*Nof*I), GAL10: 5'-CCGGAATTCATAATGGTCAACTGGGAGTTGT-3' (SEQ ID NO: 71) and 5'-
 GAATGCGGCCGCTTAGTCAATATATTTTCAGGCC-3' (SEQ ID NO: 72).

The second codon for both mature proteins was changed from an aromatic amino acid to valine due to the introduction of the Kozak sequence. The genes of interest were amplified using
 30 pET30 Xa/LIC miniprep DNA from the clones described in Examples 1 and Example 4 as template. PCR was performed using an Eppendorf Master cycler gradient thermocycler and the following protocol for a 50 µL reaction: 1.0 µL template, 1.0 µM of each primer, 0.4 mM each dNTP, 3.5 U

Expand High Fidelity Polymerase (Roche, Indianapolis, IN), and 1X ExpandTM buffer with Mg. The thermocycler program used consisted of a hot start at 94°C for 5 minutes, followed by 29 repetitions of the following steps: 94°C for 30 seconds, 50°C for 1 minute 45 seconds, and 72°C for 2 minutes 15 seconds. After the 29 repetitions the sample was maintained at 72°C for 10 minutes and then
5 stored at 4°C. The PCR products were purified by separation on a 1% TAE-agarose gel followed by recovery using a QIAquick Gel Extraction Kit (Qiagen, Valencia, CA).

The pESC-His vector DNA (2.7 µg) was digested with *Bam*HI/*Sa*II and gel-purified as above. The *aspC* PCR product was digested with *Bam*HI/*Sa*II and purified with a QIAquick PCR Purification Column. Ligations were performed with the Roche Rapid DNA Ligation Kit following
10 the manufacturer's protocols. Desalted ligations were electroporated into 40 µl Electromax DH10B competent cells (Invitrogen) in a 0.2 cm Biorad disposable cuvette using a Biorad Gene Pulser II with pulse controller plus, according to the manufacturer's instructions. After 1 hour of recovery in 1 mL of SOC medium, the transformants were plated on LB medium containing 100 µg/mL ampicillin. Plasmid DNA preparations for clones were done using QIAprep Spin Miniprep Kits. Plasmid DNA
15 was screened by restriction digest, and sequenced (Seqwright) for verification using primers designed for the vector.

The *aspC*/pESC-His clone was digested with *Eco*RI and *Not*I, as was the *proA* PCR product. DNA was purified as above, and ligated as above. The two gene construct was transformed into DH10B cells and screened by restriction digest and DNA sequencing.

20 The construct was transformed into *S. cerevisiae* strain YPH500 using the S.c. EasyCompTM Transformation Kit (Invitrogen). Transformation reactions were plated on SC-His minimal medium (Invitrogen pYES2 manual) containing 2% glucose. Individual yeast colonies were screened for the presence of the *proA* and *aspC* genes by colony PCR using the PCR primers above. Pelleted cells (2 µl) were suspended in 20 µL of Y-Lysis Buffer (Zymo Research) containing 1 µl of zymolase and
25 heated at 37°C for 10 minutes. Four µL of this suspension was then used in a 50 µL PCR reaction using the PCR reaction mixture and program described above.

Five mL cultures were grown overnight on SC-His + glucose at 30°C and 225 rpm. The cells were gradually adjusted to growth on raffinose in order to minimize the lag period prior to induction with galactose. After approximately 12 hours of growth, absorbance measurements at 600
30 nm were taken, and an appropriate volume of cells was spun down and resuspended to give an OD of 0.4 in the fresh SC-His medium. The following carbon sources were used sequentially: 1% raffinose + 1 % glucose, 0.5% glucose + 1.5% raffinose, 2% raffinose, and finally 1% raffinose + 2% galactose for induction.

After approximately 16 hours of growth in induction medium, the 50 mL cultures were
35 divided into duplicate 25 mL cultures, and the following were added to only one of the duplicates: (final concentrations) 1 g/L L-tryptophan, 5 mM sodium phosphate pH 7.1, 1 g/L sodium pyruvate, 1 mM MgCl₂. Samples of broths and cell pellets from the non-induction medium, and from the 16 hour cultures prior to addition of substrates for the monatin pathway, were saved as negative controls. In

addition, constructs containing only a functional *aspC* gene (and a truncated *proA* gene) were utilized as another negative control. The cells were allowed to grow for a total of 69 hours post-induction. Occasionally the yeast cells were induced at a lower OD, and only grown for 4 hours prior to addition of tryptophan and pyruvate. However, these monatin substrates appear to inhibit growth and the addition at higher OD was more effective.

The cell pellets from the cultures were lysed with 5 mL of YeastBuster™ + 50 µl THP (Novagen) per gram (wet weight) of cells following manufacturer's protocols, with the addition of protease inhibitors and benzonase nuclease as described in previous examples. The culture broth and cell extracts were filtered and analyzed by SRM as described in Example 10. Using this method, no monatin was detected in the broth samples, indicating that the cells could not secrete monatin under these conditions. The proton motive force may be insufficient under these conditions or the general amino acid transporters may be saturated with tryptophan. Protein expression was not at a level that allowed for detection of changes using SDS-PAGE.

Monatin was detectable (approximately 60 µg/mL) transiently in cell extracts of the culture with two functional genes, when tryptophan and pyruvate were added to the medium. Monatin was not detected in any of the negative control cell extracts. *In vitro* assays for monatin were performed in duplicate with 4.4 mg/mL of total protein (about double what is typically used for *E. coli* cell extracts) using the optimized assay described in Example 6. Other assays were performed with the addition of either 32 µg/mL *C. testosteroni* ProA aldolase or 400 µg/mL AspC aminotransferase, to determine which enzyme was limiting in the cell extract. Negative controls were performed with no addition of enzyme, or the addition of only AspC aminotransferase (the aldol condensation can occur to some extent without enzyme). Positive controls were performed with partially pure enzymes (30-40%), using 16 µg/mL aldolase and 400 µg/mL aminotransferase.

In vitro results were analyzed by SRM. The analysis of cell extracts showed that tryptophan was effectively transported into the cells when it was added to the medium post-induction, resulting in tryptophan levels two orders of magnitude higher than those in which no additional tryptophan was added. The results for *in vitro* monatin analysis are shown in Table 6 (numbers indicate ng/mL).

Table 6: Monatin production with yeast cell extracts.

	<i>aspC</i> construct	+ aldolase	+ AspC	two-gene construct	+ aldolase	+ AspC
repressed (glucose medium)	0	888.3	173.5	0	465.2	829
24 hr induced	0	2832.8	642.4	0	1375.6	9146.6
69 hr induced	0	4937.3	340.3	71.9	1652.8	23693.5
69 hr + subs.	0	556.9	659.1	21.9	755.6	16688.2
+ control (purified enzymes)	21853			21853		
-control (no enzymes)	0		254.3	0		254.3

Positive results were obtained with the full two-gene construct cell extracts with and without substrate added to the growth medium. These results, in comparison to the positive controls, indicate that the enzymes were expressed at levels of close to 1% of the total protein in yeast. The amount of monatin produced when the cell extract of the *aspC* construct (with truncated *proA*) was assayed with aldolase was significantly greater than when cell extracts were assayed alone, and indicates that the recombinant AspC aminotransferase comprises approximately 1-2% of the yeast total protein. The cell extracts of uninduced cultures had a small amount of activity when assayed with aldolase due to the presence of native aminotransferases in the cells. When assayed with AspC aminotransferase, the activity of the extracts from uninduced cells increased to the amount of monatin produced by the negative control with AspC (ca. 200 ng/ml). In contrast, the activity observed when assaying the two gene construct cell extract increases more when aminotransferase is supplemented than when aldolase is added. Since both genes should be expressed at the same level, this indicates that the amount of monatin produced is maximized when the level of aminotransferase is higher than that of aldolase, in agreement with results shown in Example 6.

The addition of pyruvate and tryptophan not only inhibits cellular growth, but apparently inhibits protein expression as well. The addition of the pESC-Trp plasmid can be used to correct for tryptophan auxotrophy of the YPH500 host cells, to provide a means of supplying tryptophan with fewer effects on growth, expression, and secretion.

Example 13

Improvement of Enzymatic Processes using Coupled Reactions

In theory, if no side reactions or degradation of substrates or intermediates occurs, the maximum amount of product formed from the enzymatic reaction illustrated in FIG. 1 is directly proportional to the equilibrium constants of each reaction, and the concentrations of tryptophan and pyruvate. Tryptophan is not a highly soluble substrate, and concentrations of pyruvate greater than 200 mM appear to have a negative effect on the yield (see Example 6).

Ideally, the concentration of monatin is maximized with respect to substrates, in order to decrease the cost of separation. Physical separations can be performed such that the monatin is removed from the reaction mixture, preventing the reverse reactions from occurring. The raw materials and catalysts can then be regenerated. Due to the similarity of monatin in size, charge, and hydrophobicity to several of the reagents and intermediates, physical separations will be difficult unless there is a high amount of affinity for monatin (such as an affinity chromatography technique). However, the monatin reactions can be coupled to other reactions such that the equilibrium of the system is shifted toward monatin production. The following are examples of processes for improving the yield of monatin obtained from tryptophan or indole-3-pyruvate.

Coupled reactions using oxaloacetate decarboxylase (EC 4.1.1.3)

FIG. 11 is an illustration of the reaction. Tryptophan oxidase and catalase are utilized to drive the reaction in the direction of indole-3-pyruvate production. Catalase is used in excess such

that hydrogen peroxide is not available to react in the reverse direction or to damage the enzymes or intermediates. Oxygen is regenerated during the catalase reaction. Alternatively, indole-3-pyruvate can be used as the substrate.

Aspartate is used as the amino donor for the amination of MP, and an aspartate
5 aminotransferase is utilized. Ideally, an aminotransferase that has a low specificity for the tryptophan/indole-3-pyruvate reaction in comparison to the MP to monatin reaction is used so that the aspartate is not utilized to reaminate the indole-3-pyruvate. Oxaloacetate decarboxylase (from *Pseudomonas sp.*) can be added to convert the oxaloacetate to pyruvate and carbon dioxide. Since CO₂ is volatile, it is not available for reaction with the enzymes, decreasing or even preventing the
10 reverse reactions. The pyruvate produced in this step can also be utilized in the aldol condensation reaction. Other decarboxylase enzymes can be used, and homologs are known to exist in *Actinobacillus actinomycetemcomitans*, *Aquifex aeolicus*, *Archaeoglobus fulgidus*, *Azotobacter vinelandii*, *Bacteroides fragilis*, several *Bordetella* species, *Campylobacter jejuni*, *Chlorobium tepidum*, *Chloroflexus aurantiacus*, *Enterococcus faecalis*, *Fusobacterium nucleatum*, *Klebsiella*
15 *pneumoniae*, *Legionella pneumophila*, *Magnetococcus MC-1*, *Mannheimia haemolytica*, *Methylobacillus flagellatus KT*, *Pasteurella multocida Pm70*, *Petrotoga miotherma*, *Porphyromonas gingivalis*, several *Pseudomonas* species, several *Pyrococcus* species, *Rhodococcus*, several *Salmonella* species, several *Streptococcus* species, *Thermochromatium tepidum*, *Thermotoga maritima*, *Treponema pallidum*, and several *Vibrio* species.

20 Tryptophan aminotransferase assays were performed with the aspartate aminotransferase (AspC) from *E. coli*, the tyrosine aminotransferase (TyrB) from *E. coli*, the broad substrate aminotransferase (BSAT) from *L. major*, and the two commercially available porcine glutamate-oxaloacetate aminotransferases as described in Example 1. Both oxaloacetate and alpha-ketoglutarate were tested as the amino acceptor. The ratio of activity using monatin (Example 7) versus activity
25 using tryptophan was compared, to determine which enzyme had the highest specificity for the monatin aminotransferase reaction. These results indicated that the enzyme with the highest specificity for the monatin reaction versus the tryptophan reaction is the Porcine type II-A glutamate-oxaloacetate aminotransferase, GOAT (Sigma G7005). This specificity was independent of which amino acceptor was utilized. Therefore, this enzyme was used in the coupled reactions with
30 oxaloacetate decarboxylase.

A typical reaction starting from indole-3-pyruvate included (final concentrations) 50 mM Tris-Cl pH 7.3, 6 mM indole-3-pyruvate, 6 mM sodium pyruvate, 6 mM aspartate, 0.05 mM PLP, 3 mM potassium phosphate, 3 mM MgCl₂, 25 µg/mL aminotransferase, 50 µg/mL *C. testosteronei* ProA aldolase, and 3 Units/mL of decarboxylase (Sigma O4878). The reactions were allowed to proceed
35 for 1 hour at 26°C. In some cases, the decarboxylase was omitted or the aspartate was substituted with alpha-ketoglutarate (as negative controls). The aminotransferase enzymes described above were also tested in place of the GOAT to confirm earlier specificity experiments. Samples were filtered and analyzed by LC/MS as described in Example 10. The results demonstrate that the GOAT
enzyme produced the highest amount of monatin per mg of protein with the least amount of

tryptophan produced as a byproduct. In addition, there was a 2-3 fold benefit from having the decarboxylase enzyme added. The *E. coli* AspC enzyme also produced large amounts of monatin in comparison to the other aminotransferases.

Monatin production was increased by: 1) periodically adding 2 mM additions of indole-
5 pyruvate, pyruvate, and aspartate (every half hour to hour), 2) performing the reactions in an anaerobic environment or with degassed buffers, 3) allowing the reactions to proceed overnight, and 4) using freshly prepared decarboxylase that has not been freeze-thawed multiple times. The decarboxylase was inhibited by concentrations of pyruvate greater than 12 mM. At concentrations of indole-3-pyruvate higher than 4 mM, side reactions with indole-3-pyruvate were hastened. The
10 amount of indole-3-pyruvate used in the reaction could be increased if the amount of aldolase was also increased. High levels of phosphate (50 mM) and aspartate (50 mM) were found to be inhibitory to the decarboxylase enzyme. The amount of decarboxylase enzyme added could be reduced to 0.5 U/mL with no decrease in monatin production in a one hour reaction. The amount of monatin produced increased when the temperature was increased from 26°C to 30°C and from 30°C to 37°C;
15 however, at 37°C the side reactions of indole-3-pyruvate were also hastened. The amount of monatin produced increased with increasing pH from 7 to 7.3, and was relatively stable from pH 7.3-8.3.

A typical reaction starting with tryptophan included (final concentrations) 50 mM Tris-Cl pH 7.3, 20 mM tryptophan, 6 mM aspartate, 6 mM sodium pyruvate, 0.05 mM PLP, 3 mM potassium phosphate, 3 mM MgCl₂, 25 µg/mL aminotransferase, 50 µg/mL *C. testosteroni* ProA aldolase, 4
20 Units/mL of decarboxylase, 5-200 mU/mL L-amino acid oxidase (Sigma A-2805), 168 U/mL catalase (Sigma C-3515), and 0.008 mg FAD. Reactions were carried out for 30 minutes at 30°C. Improvement was observed with the addition of decarboxylase. The greatest amount of monatin was produced when 50 mU/mL of oxidase was used. Improvements were similar to those observed when indole-3-pyruvate was used as the substrate. In addition, the amount of monatin produced increased
25 when 1) the tryptophan level was low (i.e., below the K_m of the aminotransferase enzyme and therefore unable to compete with MP in the active site), and 2) the ratio of oxidase to aldolase and aminotransferase was maintained at a level such that indole-3-pyruvate could not accumulate.

Whether starting with either indole-3-pyruvate or tryptophan, the amount of monatin produced in assays with incubation times of 1-2 hours increased when 2-4 times the amounts of all
30 the enzymes were used while maintaining the same enzyme ratio. Using either substrate, concentrations of approximately 1 mg/mL of monatin were achieved. The amount of tryptophan produced if starting from indole-pyruvate was typically less than 20% of the amount of product, which shows the benefit of utilizing coupled reactions. With further optimization and control of the concentrations of intermediates and side reactions, the productivity and yield can be improved
35 greatly.

Coupled reactions using lysine epsilon aminotransferase (EC 2.6.1.36)

Lysine epsilon aminotransferase (L-Lysine 6-transaminase) is found in several organisms, including *Rhodococcus*, *Mycobacterium*, *Streptomyces*, *Nocardia*, *Flavobacterium*, *Candida utilis*,

and *Streptomyces*. It is utilized by organisms as the first step in the production of some beta-lactam antibiotics (Rius and Demain, *J. Microbiol. Biotech.*, 7:95-100, 1997). This enzyme converts lysine to L-2-aminoadipate 6-semialdehyde (allysine), by a PLP-mediated transamination of the C-6 of lysine, utilizing alpha-ketoglutarate as the amino acceptor. Allysine is unstable and spontaneously

- 5 undergoes an intramolecular dehydration to form 1-piperidine 6-carboxylate, a cyclic molecule. This effectively inhibits any reverse reaction from occurring. The reaction scheme is depicted in FIG. 12. An alternative enzyme, lysine-pyruvate 6-transaminase (EC 2.6.1.71), can also be used.

A typical reaction contained in 1 mL: 50 mM Tris-HCl pH 7.3, 20 mM indole-3-pyruvate, 0.05 mM PLP, 6 mM potassium phosphate pH 8, 2-50 mM sodium pyruvate, 1.5 mM $MgCl_2$, 50 mM
10 lysine, 100 μ g aminotransferase (lysine epsilon aminotransferase LAT-101, BioCatalytics Pasadena, CA), and 200 μ g *C. testosteroni* ProA aldolase. The amount of monatin produced increased with increasing concentrations of pyruvate. The maximum amount using these reaction conditions (at 50 mM pyruvate) was 10-fold less than what was observed with coupled reactions using oxaloacetate decarboxylase (approximately 0.1 mg/mL).

- 15 A peak with $[M+H]^+ = 293$ eluted at the expected time for monatin and the mass spectrum contained several of the same fragments observed with other enzymatic processes. A second peak with the correct mass to charge ratio (293) eluted slightly earlier than what is typically observed for the S,S monatin produced in Example 6, and may indicate the presence of another isomer of monatin. Very little tryptophan was produced by this enzyme. However, there is likely some activity on
20 pyruvate (producing alanine as a byproduct). Also, the enzyme is known to be unstable. Improvements can be made by performing directed evolution experiments to increase stability, reduce the activity with pyruvate, and increase the activity with MP. These reactions can also be coupled to L-amino acid oxidase/catalase as described above.

25 *Other coupled reactions*

- Another coupling reaction that can improve monatin yield from tryptophan or indole-pyruvate is shown in FIG. 13. Formate dehydrogenase (EC 1.2.1.2 or 1.2.1.43) is a common enzyme. Some formate dehydrogenases require NADH while others can utilize NADPH. Glutamate dehydrogenase catalyzed the interconversion between the monatin precursor and monatin in previous
30 examples, using ammonium based buffers. The presence of ammonium formate and formate dehydrogenase is an efficient system for regeneration of cofactors, and the production of carbon dioxide is an efficient way to decrease the rate of the reverse reactions (Bommarius *et al.*, *Biocatalysis* 10:37, 1994 and Galkin *et al.* *Appl. Environ. Microbiol.* 63:4651-6, 1997). In addition, large amounts of ammonium formate can be dissolved in the reaction buffer. The yield of monatin
35 produced by glutamate dehydrogenase reactions (or similar reductive aminations) can be improved by the addition of formate dehydrogenase and ammonium formate.

Other processes can be used to drive the equilibrium toward monatin production. For instance, if aminopropane is utilized as the amino acid donor in the conversion of MP to monatin with

an omega-amino acid aminotransferase (EC 2.6.1.18) such as those described by in US patents 5,360,724 and 5,300,437, one of the resulting products would be acetone, a more volatile product than the substrate, aminopropane. The temperature can be raised periodically for short periods to flash off the acetone, thereby alleviating equilibrium. Acetone has a boiling point of 47°C, a temperature not likely to degrade the intermediates if used for short periods of time. Most aminotransferases that have activity on alpha-ketoglutarate also have activity on the monatin precursor. Similarly, if a glyoxylate/aromatic acid aminotransferase (EC 2.6.1.60) is used with glycine as the amino donor, glyoxylate is produced which is relatively unstable and has a highly reduced boiling point in comparison to glycine.

EXAMPLE 14

Recombinant Expression

With publicly available enzyme cDNA and amino acid sequences, and the enzymes and sequences disclosed herein, such as SEQ ID NOS: 11 and 12, as well as variants, polymorphisms, mutants, fragments and fusions thereof, the expression and purification of any protein, such as an enzyme, by standard laboratory techniques is enabled. One skilled in the art will understand that enzymes and fragments thereof can be produced recombinantly in any cell or organism of interest, and purified prior to use, for example prior to production of SEQ ID NO: 12 and derivatives thereof.

Methods for producing recombinant proteins are well known in the art. Therefore, the scope of this disclosure includes recombinant expression of any protein or fragment thereof, such as an enzyme. For example, see U.S. Patent No: 5,342,764 to Johnson *et al.*; U.S. Patent No: 5,846,819 to Pausch *et al.*; U.S. Patent No: 5,876,969 to Fleer *et al.* and Sambrook *et al.* (*Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, New York, 1989, Ch. 17).

Briefly, partial, full-length, or variant cDNA sequences, which encode for a protein or peptide, can be ligated into an expression vector, such as a bacterial or eukaryotic expression vector. Proteins and/or peptides can be produced by placing a promoter upstream of the cDNA sequence. Examples of promoters include, but are not limited to *lac*, *trp*, *tac*, *trc*, major operator and promoter regions of phage lambda, the control region of fd coat protein, the early and late promoters of SV40, promoters derived from polyoma, adenovirus, retrovirus, baculovirus and simian virus, the promoter for 3-phosphoglycerate kinase, the promoters of yeast acid phosphatase, the promoter of the yeast alpha-mating factors and combinations thereof.

Vectors suitable for the production of intact native proteins include pKC30 (Shimatake and Rosenberg, 1981, *Nature* 292:128), pKK177-3 (Amann and Brosius, 1985, *Gene* 40:183) and pET-3 (Studier and Moffatt, 1986, *J. Mol. Biol.* 189:113). A DNA sequence can be transferred to other cloning vehicles, such as other plasmids, bacteriophages, cosmids, animal viruses and yeast artificial chromosomes (YACs) (Burke *et al.*, 1987, *Science* 236:806-12). These vectors can be introduced into a variety of hosts including somatic cells, and simple or complex organisms, such as bacteria, fungi (Timberlake and Marshall, 1989, *Science* 244:1313-7), invertebrates, plants (Gasser and Fraley,

1989, *Science* 244:1293), and mammals (Pursel *et al.*, 1989, *Science* 244:1281-8), which are rendered transgenic by the introduction of the heterologous cDNA.

For expression in mammalian cells, a cDNA sequence can be ligated to heterologous promoters, such as the simian virus SV40, promoter in the pSV2 vector (Mulligan and Berg, 1981, *Proc. Natl. Acad. Sci. USA* 78:2072-6), and introduced into cells, such as monkey COS-1 cells (Gluzman, 1981, *Cell* 23:175-82), to achieve transient or long-term expression. The stable integration of the chimeric gene construct may be maintained in mammalian cells by biochemical selection, such as neomycin (Southern and Berg, 1982, *J. Mol. Appl. Genet.* 1:327-41) and mycophenolic acid (Mulligan and Berg, 1981, *Proc. Natl. Acad. Sci. USA* 78:2072-6).

The transfer of DNA into eukaryotic, such as human or other mammalian cells, is a conventional technique. The vectors are introduced into the recipient cells as pure DNA (transfection) by, for example, precipitation with calcium phosphate (Graham and vander Eb, 1973, *Virology* 52:466) strontium phosphate (Brash *et al.*, 1987, *Mol. Cell Biol.* 7:2013), electroporation (Neumann *et al.*, 1982, *EMBO J.* 1:841), lipofection (Felgner *et al.*, 1987, *Proc. Natl. Acad. Sci. USA* 84:7413), DEAE dextran (McCuthan *et al.*, 1968, *J. Natl. Cancer Inst.* 41:351), microinjection (Mueller *et al.*, 1978, *Cell* 15:579), protoplast fusion (Schafner, 1980, *Proc. Natl. Acad. Sci. USA* 77:2163-7), or pellet guns (Klein *et al.*, 1987, *Nature* 327:70). Alternatively, the cDNA can be introduced by infection with virus vectors, for example retroviruses (Bernstein *et al.*, 1985, *Gen. Engrg.* 7:235) such as adenoviruses (Ahmad *et al.*, 1986, *J. Virol.* 57:267) or Herpes (Spaete *et al.*, 1982, *Cell* 30:295).

In view of the many possible embodiments to which the principles of our disclosure may be applied, it should be recognized that the illustrated embodiments are only particular examples of the disclosure and should not be taken as a limitation on the scope of the disclosure. Rather, the scope of the disclosure is in accord with the following claims. We therefore claim as our invention all that comes within the scope and spirit of these claims.

We claim:

1. A method of producing monatin, comprising converting tryptophan and/or indole-3-lactic acid to monatin, wherein the method comprises: contacting tryptophan or indole-3-lactic acid with a first polypeptide, wherein the first polypeptide converts the tryptophan or indole-3-lactic acid to indole-3-pyruvate; contacting the indole-3-pyruvate with a second polypeptide and a C3 carbon source, wherein the second polypeptide converts the indole-3-pyruvate and the C3 carbon source to 2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid; and contacting the 2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid with a third polypeptide, wherein the third polypeptide converts the 2-hydroxy 2-(indol-3-ylmethyl)-4-keto glutaric acid to monatin.
2. The method of claim 1, wherein the first polypeptide is selected from the group consisting of tryptophan aminotransferase (EC 2.6. 1.27), tryptophan dehydrogenase (EC 1. 4.1. 19), tyrosine (aromatic) aminotransferase (EC 2.6. 1.5), tryptophan-phenylpyruvate transaminase (EC 2.6. 1.28), multiple substrate aminotransferase (EC 2.6. 1.-), aspartate aminotransferase (EC 2.6. 1.1), tryptophan oxidase, L-amino acid oxidase (EC 1. 4. 3. 2), D-amino acid dehydrogenase (EC 1.4. 99.1), D-amino acid oxidase (EC 1. 4. 3. 3), D-tryptophan aminotransferase, D-alanine aminotransferase (EC 2.6. 1.21), glutamate dehydrogenase (EC 1.4. 1.2-4), phenylalanine dehydrogenase (EC 1.4. 1.20), and a combination thereof, and wherein the first polypeptide converts tryptophan to indole-3-pyruvate.
3. The method of claim 1, wherein the first polypeptide is selected from the group consisting of indolelactate dehydrogenase (EC 1. 1. 1. 110), R-4-hydroxyphenyllactate dehydrogenase (EC 1.1. 1.222), 3- (4)-hydroxyphenylpyruvate reductase (EC 1.1. 1.237), lactate dehydrogenase (EC 1.1. 1.27, 1.1. 1.28, 1.1. 2.3), (3-imidazol-5-yl) lactate dehydrogenase (EC 1. 1. 1. 111), and lactate oxidase (EC 1. 1.3.-), or a combination thereof, and wherein the first polypeptide converts indole-3-lactic acid to indole-3-pyruvate.
4. The method of claim 1 or 2, wherein the first and/or third polypeptide comprises:

an amino acid sequence shown in SEQ ED NO: 2,4, 6,8, 10,12, 14,32, Genbank Accession No: Nu-388848. 1, Genbank Accession No: ZP00005082. 1, or Genbank Accession No : AAC74014.1 ;

an amino acid sequence comprising at least 90% sequence identity with the sequence shown in SEQ ID NO : 2,4, 6,8, 10,12, 14,32, Genbank Accession No: Nu-388848. 1, Genbank Accession No: ZP00005082.1, or Genbank Accession No : AAC74014.1 ; or

amino acid sequences that differ from SEQ ID NO: 2,4, 6,8, 10,12, 14,32, Genbank Accession No: NP388848. 1, Genbank Accession No: ZP00005082.1, or Genbank Accession No: AAC74014.1 by less than 50 conservative amino acid substitutions, wherein the amino acid sequence converts tryptophan to indole-3-pyruvate.

5. The method of claim 1, wherein the second polypeptide is an EC 4.1. 2.- or EC 4. 1. 3.- enzyme.

6. The method of claim 5, wherein the EC 4.1. 3.- enzyme is 4-hydroxy-2-oxoglutarate glyoxylate- lyase (EC 4. 1. 3. 16), 4-hydroxy-4-methyl-2-oxoglutarate pyruvate-lyase (EC 4.1. 3.17) or a combination thereof.

7. The method of claim 1, wherein the third polypeptide is selected from the group consisting of tyrosine (aromatic) aminotransferase (EC 2.6. 1.5), tryptophan dehydrogenase (EC 1.4. 1.19), tryptophan-phenylpyruvate transaminase (EC 2.6. 1.28), tryptophan aminotransferase (EC 2.6. 1.27), aspartate aminotransferase (2.6. 1.1), D-alanine aminotransferase (EC 2.6. 1.21), D-tryptophan aminotransferase, glutamate dehydrogenase (EC 1.4. 1.2-4), phenylalanine dehydrogenase (EC 1.4. 1.20), multiple substrate aminotransferase (EC 2.6. 1. -), D-amino acid dehydrogenase (EC 1.4. 99.1), and a combination thereof.

8. The method of claim 1, wherein the first polypeptide comprises an amino acid oxidase (EC 1.4. 3. -) and a catalase (EC 1.11. 1.6), and wherein the first polypeptide converts tryptophan to indole-3- pyruvate.

9. The method of claim 8, wherein the amino acid oxidase comprises a tryptophan oxidase.
10. The method of claim 1, wherein the C3 carbon source is selected from the group consisting of pyruvate, phosphoenolpyruvate, alanine, serine, cysteine, or a combination thereof.
11. The method of claim 10, wherein the pyruvate is produced by alanine and a polypeptide capable of transamination of alanine; serine and a polypeptide capable of beta-elimination of serine; cysteine and a polypeptide capable of beta-elimination of cysteine; aspartate and a polypeptide capable of beta-lyase reactions with a dicarboxylic amino acid; lactate and lactate oxidase (EC 1. 1. 3.-) or lactate dehydrogenase (EC 1.1. 1.27, 1.1. 1.28, 1.1. 2.3) ; or a combination thereof.
12. The method of claim 1, wherein the method further comprises reducing an amount of hydrogen peroxide produced when tryptophan is converted to indole-3-pyruvate, by contacting the hydrogen peroxide with a catalase.
13. The method of claim 1, wherein the second polypeptide comprises an amino acid sequence selected from the group consisting of : SEQ ID NO: 66, Genbank Accession No: CAC46344, Genbank Accession No: CAB 14127. 1, Genbank Accession No: AAC74920.1, or Genbank Accession No: CAC47463.1 ; an amino acid sequence comprising at least 90% sequence identity to SEQ ID NO: 66, Genbank Accession No : CAC46344, Genbank Accession No : CAB14127. 1, Genbank Accession No : AAC74920.1, or Genbank Accession No : CAC47463.1 ; or amino acid sequences that differ from SEQ ID NO: 66, Genbank Accession No: CAC46344, Genbank Accession No : CAB 14127. 1, Genbank Accession No: AAC74920.1, or Genbank Accession No: CAC47463.1 by less than 50 conservative amino acid substitutions, wherein the amino acid sequence converts indole-3-pyruvate and pyruvate to MP.
14. The method of claim 7, wherein the third polypeptide comprises an aspartate aminotransferase, wherein the 2-hydroxy 2- (indol-3-ylmethyl)-4-keto glutaric acid is contacted with aspartate to produce oxaloacetate and monatin.
15. The method of claim 14, further comprising contacting oxaloacetate with a decarboxylase.

16. The method of claim 1, wherein the third polypeptide comprises a lysine epsilon aminotransferase, and wherein the 2-hydroxy 2- (indol-3-ylmethyl)-4-keto glutaric acid is further contacted with lysine.
17. The method of claim 1, wherein the third polypeptide comprises an enzyme having reductive amination activity and a polypeptide that is capable of recycling NAD (P) H.
18. The method of claim 17, wherein the enzyme having reductive amination activity is selected from the group consisting of glutamate dehydrogenase, phenylalanine dehydrogenase, tryptophan dehydrogenase, or a combination thereof.
19. The method of claim 17, wherein the polypeptide capable of recycling NAD (P) H further produces a volatile product.
20. The method of claim 19, wherein the polypeptide that is capable of recycling NAD (P) H and producing a volatile product comprises formate dehydrogenase, and wherein the volatile product produced is carbon dioxide.
21. The method of claim 1, wherein the method comprises converting indole-3-lactic acid to monatin.
22. The method of claim 1, where at least 10 g/L of monatin is produced.
23. A method of producing monatin comprising converting indole-3-pyruvate to monatin.
24. A method of producing monatin comprising converting tryptophan to monatin.
25. A polypeptide comprising an amino acid sequence selected from the group consisting of Genbank Accession No : P00913), Genbank Accession No : NP290344, Genbank Accession No: CAA34096, Genbank Accession No : NP246359, GenbankAccessionNo : AAB96579, Genbank Accession No: NP232561, GenbankAccessionNo : Q59342, Genbank Accession No : P28796Genbank Accession No: lAX4 A, Genbank Accession No : BAA34638, Genbank Accession No: P31014, Genbank Accession No : P31015, Genbank Accession No : NP280989, Genbank Accession No : AAC33284, Genbank Accession No : NP_147838, Genbank

Accession No : AAF63441, or Genbank Accession No: AAA24679, where the amino acid residues at position 103 and 299 with respect to Genbank Accession No: Nu-418164 have been changed to broaden a specificity of the polypeptide to accept dicarboxylic amino acid substrates.

26. The polypeptide of claim 25, where the change in the amino acid sequence is selected from the group consisting of arginine 103 changed to threonine, arginine 103 changed to threonine and valine 299 to arginine; and valine 299 to arginine.

27. A polypeptide comprising an amino acid sequence selected from the group consisting of Genbank Accession No : 2TPL A, Genbank Accession No : P31012, Genbank Accession No : P31013, A49493, Genbank Accession No : P31011, Genbank Accession No: AAB24234, Genbank Accession No: 1TPL A,, Genbank Accession No : 1TPL B, Genbank Accession No : NP_245748, Genbank Accession No : AAB41499, Genbank Accession No : Q08897, GenbankAccessionNo : T45297, and Genbank Accession No: AAA71928, where the amino acid residues at position 100 and 283 with respect to Genbank Accession No: X66978 have been changed to broaden a specificity of the polypeptide to accept dicarboxylic acid amino acid substrates.

28. The polypeptide of claim 27, where the change in the amino acid sequence is selected from the group consisting of : arginine 100 changed to threonine, arginine 100 changed to threonine and valine 283 to arginine; and valine 283 to arginine.

29. The polypeptide of claim 25 or 27, wherein the substrate is aspartate.

30. A method of making monatin comprising reacting at least one alanine derivative blocked at carboxyl and amino groups, and at least one indole-3-pyruvate derivative blocked at a carboxyl group.

31. The method of making monatin according to claim 30, wherein the alanine derivative is 3- chloro-alanine or 3-bromo-alanine, and wherein the indole-3-pyruvate is reacted with the alanine derivative in a Grignard or organolithium reaction.

32. A method of making monatin comprising converting a substrate selected from the group consisting of tryptophan, indole-3-lactic acid, indole-3-pyruvate, and 2-

hydroxy 2- (indol-3- ylmethyl) -4-keto glutaric acid, to one or more intermediates, wherein monatin is made.

33. The method of claim 32, wherein at least one of the intermediate products is converted via a chemical reaction.

34. The method of claim 32, wherein at least one of the intermediate products is converted via a polypeptide.

35. An isolated nucleic acid comprising at least 90% sequence identity to SEQ ID NO: 11.

36. The isolated nucleic acid of claim 35, wherein the nucleic acid comprises SEQ ID NO: 11.

37. The isolated nucleic acid of claim 35, wherein the nucleic acid comprises at least 20 consecutive nucleotides of SEQ ID NO: 11.

38. A vector comprising the isolated nucleic acid of claim 35.

39. A recombinant cell comprising the vector of claim 38.

40. An isolated protein encoded by the isolated nucleic acid of claim 35.

41. The isolated protein of claim 40, wherein the protein comprises a sequence that has at least 90% sequence identity to SEQ ID NO: 12.

42. The isolated protein of claim 41, wherein the protein comprises SEQ ID NO: 12.

43. A method for producing monatin comprising converting 2-hydroxy 2- (indol-3- ylmethyl)-4-keto glutaric acid to monatin.

44. The method of claim 43, wherein the method further comprises detection of the 2-hydroxy 2- (indol-3-ylmethyl) -4-keto glutaric acid by mass spectrometry.

45. A composition comprising 2-hydroxy 2- (indol-3-ylmethyl)-4-keto glutaric acid.

46. The method of claim 1, wherein the tryptophan is produced from indole using a polypeptide from EC 4.1. 99.1.

47. A method for producing monatin comprising reacting tryptophan with one or more polypeptides.

48. The method of claim 47, wherein the one or more polypeptides is selected from the group consisting of: tryptophan aminotransferase (EC 2.6. 1.27), tryptophan dehydrogenase (EC 1.4. 1.19), tyrosine (aromatic) aminotransferase (EC 2.6. 1.5), tryptophan-phenylpyruvate transaminase (EC 2.6. 1.28), multiple substrate aminotransferase (EC 2.6. 1.-), aspartate aminotransferase (EC 2.6. 1.1), tryptophan oxidase, L-amino acid oxidase (EC 1.4. 3.2), D-amino acid dehydrogenase (EC 1. 4.99. 1), D-amino acid oxidase (EC 1.4. 3. 3), D-alanine aminotransferase (EC 2.6. 1.21), D-tryptophan aminotransferase, synthase/lyase (EC 4.1. 3. -), synthase/lyase (4.1. 2. -), phenylalanine dehydrogenase (EC 1.4. 1.20), glutamate dehydrogenase (EC 1.4. 1.2, 1.4. 1. 3, 1.4. 1.4) and a combination thereof.

SEQUENCE LISTING

<110> Cargill, Incorporated

<120> Polypeptides and Biosynthetic Pathways

<130> 6682-64349

<150> 60/374,831

<151> 2002-04-23

<160> 74

<170> PatentIn version 3.1

<210> 1

<211> 1170

<212> DNA

<213> Sinorhizobium meliloti

<400> 1

```

atgttcgacg ccctcgcccg ccaagccgac gatcccttgc ttttcctgat cggcctgttc      60
aggaaggatg agcgccccgg aaaggctgat ctggcgtag gagtctatcg cgaagagacc      120
ggacgcacgc cgatcttcgg ggcggtcaag gcggcggaag agcggttctt cgaaacacag      180
gacagcaagg cctatatcgg ccccggaagg gacctcgtct ttctcgatcg gctctgggaa      240
ctcgtcggcg gcgacacgat cgagcggagc catgttgccg gcgtccagac gcccgggcgc      300
tccggcgcgc tccgtttggc ggcggacctc atcgcccgca tgggcggccg aggcattctgg      360
ctcgggctgc cgagctggcc gaaccacgcg ccgatcttca aggcggcccg gctcgatata      420
gccacctacg acttcttcga cattccgtcg cagtcgggtca tcttcgataa tctggtgagc      480
gcgctggaag gcgcccgcac cggcgatgcg gtgctgctgc atgcaagctg ccacaaccgg      540
accggcgggc tcttgagcga agcacaatgg atggagatcg ccgctgtggt ggccgagcgc      600
ggcctgctgc cgctcgtcga tctcgcttat caggggttcc gcccgggcct cgaccaggat      660
gtcgcggggc tccggcatct tctcggcgtg gtcccgaag cgctcgtcgc ggtttctctgc      720
tcgaagtcct tcgggcttta tcgcgagcgc gcgggcgcga tcttcgcgcg gaccagctcg      780
actgcctcgg cggacagggg gcgctcaaac ctgcggggcc tcgacgcac cagctattcc      840
atgccgcccg atcacggcgc agccgtcgtg cggacgatcc ttgacgaccc ggaactcagg      900
cgcgactgga cggaggagct cgagacgatg cggctcagga tgacgggcct ccggcggtcg      960
cttgccgagg gactccgcac ccgctggcag agcctcggcg cagtcgccga tcaggagggc     1020
atgtttctca tgctgccgct ttccgaagcg gaggttatgc ggctcaggac cgagcacggc     1080
atctatatgc cggcatccgg ccgcatcaac atcgccgggc tgaagacggc ggaagccgcc     1140
gagattgccg gcaagttcac cagtctctga                                     1170

```

<210> 2
 <211> 389
 <212> PRT
 <213> *Sinorhizobium meliloti*

<400> 2

Met Phe Asp Ala Leu Ala Arg Gln Ala Asp Asp Pro Leu Leu Phe Leu
 1 5 10 15

Ile Gly Leu Phe Arg Lys Asp Glu Arg Pro Gly Lys Val Asp Leu Gly
 20 25 30

Val Gly Val Tyr Arg Asp Glu Thr Gly Arg Thr Pro Ile Phe Arg Ala
 35 40 45

Val Lys Ala Ala Glu Lys Arg Leu Leu Glu Thr Gln Asp Ser Lys Ala
 50 55 60

Tyr Ile Gly Pro Glu Gly Asp Leu Val Phe Leu Asp Arg Leu Trp Glu
 65 70 75 80

Leu Val Gly Gly Asp Thr Ile Glu Arg Ser His Val Ala Gly Val Gln
 85 90 95

Thr Pro Gly Gly Ser Gly Ala Leu Arg Leu Ala Ala Asp Leu Ile Ala
 100 105 110

Arg Met Gly Gly Arg Gly Ile Trp Leu Gly Leu Pro Ser Trp Pro Asn
 115 120 125

His Ala Pro Ile Phe Lys Ala Ala Gly Leu Asp Ile Ala Thr Tyr Asp
 130 135 140

Phe Phe Asp Ile Pro Ser Gln Ser Val Ile Phe Asp Asn Leu Val Ser
 145 150 155 160

Ala Leu Glu Gly Ala Ala Ser Gly Asp Ala Val Leu Leu His Ala Ser
 165 170 175

Cys His Asn Pro Thr Gly Gly Val Leu Ser Glu Ala Gln Trp Met Glu
 180 185 190

Ile Ala Ala Leu Val Ala Glu Arg Gly Leu Leu Pro Leu Val Asp Leu
 195 200 205

Ala Tyr Gln Gly Phe Gly Arg Gly Leu Asp Gln Asp Val Ala Gly Leu

210 215 220
 Arg His Leu Leu Gly Val Val Pro Glu Ala Leu Val Ala Val Ser Cys
 225 230 235 240
 Ser Lys Ser Phe Gly Leu Tyr Arg Glu Arg Ala Gly Ala Ile Phe Ala
 245 250 255
 Arg Thr Ser Ser Thr Ala Ser Ala Asp Arg Val Arg Ser Asn Leu Ala
 260 265 270
 Gly Leu Ala Arg Thr Ser Tyr Ser Met Pro Pro Asp His Gly Ala Ala
 275 280 285
 Val Val Arg Thr Ile Leu Asp Asp Pro Glu Leu Arg Arg Asp Trp Thr
 290 295 300
 Glu Glu Leu Glu Thr Met Arg Leu Arg Met Thr Gly Leu Arg Arg Ser
 305 310 315 320
 Leu Ala Glu Gly Leu Arg Thr Arg Trp Gln Ser Leu Gly Ala Val Ala
 325 330 335
 Asp Gln Glu Gly Met Phe Ser Met Leu Pro Leu Ser Glu Ala Glu Val
 340 345 350
 Met Arg Leu Arg Thr Glu His Gly Ile Tyr Met Pro Ala Ser Gly Arg
 355 360 365
 Ile Asn Ile Ala Gly Leu Lys Thr Ala Glu Ala Ala Glu Ile Ala Gly
 370 375 380
 Lys Phe Thr Ser Leu
 385

<210> 3
 <211> 1260
 <212> DNA
 <213> Rhodobacter sphaeroides

<400> 3
 atgcgctcta cgacggctcc tgggtccgagt ggggcatgta tgacgatctc aaggctcgca 60
 aaggatgacg aaggaatgct gaccgcctg aagccgcagc ccgaggacaa gatcctgcaa 120
 ctgatccaga tgttccgcga ggatgcgcgc gccgacaaga tcgatctggg cgtgggcgtc 180
 tacaaggacc cgaccgggt caccgccgtc atgcgggccc tgaaggcggc cgagaagcgg 240

ctctgggagg tcgagaccac caagacctac accggccttg ccgacgagcc ggcctacaat 300
 gccgcgatgg cgaagctgat cctcgcgggc gcgggtcccg ccgaccgggt ggcctcggtc 360
 gccacccccg gcggcacggg cgcgggtgct caggcgctcg agctgatccg catggcctcg 420
 cccgaggcca ccgtctggat ctogaacccg acctggccga accatctgtc gatcgtgaaa 480
 tatctcggca tcccgatgcg ggaataccgc tatttcgacg ccgagaccgg cgccgtcgat 540
 gccgagggca tgatggagga tetggcccag gtgaaggcgg gcgacgtggt gctgctgcac 600
 ggctgctgcc acaacccgac cggcgccaac ccgaacccgg tgcagtggct ggccatctgc 660
 gagagcctgg cccggacagg cgcgggtgcc ctgatcgacc tcgcctatca gggcttcggc 720
 gacgggctcg agatggatgc ggcggcgacg cggcttcttg ccaccagact gcccgagggtg 780
 ctgatcgcg cctcctgctc gaagaacttc ggcattctacc gcgagcgcac gggcatcctg 840
 atcgccatcg gcgaggcggc gggccggggc acggtgcagg ccaacctcaa cttcctgaac 900
 cggcagaact actccttccc gccggaccat ggcgcgcggc tcgtgaccat gatcctcgag 960
 gacgagacgc tgagcgccga ctggaaggcg gaactcgagg aggtgcggct caacatgctg 1020
 aactgcgcc gccagcttgc cgatgcgctg caggccgaga ccggctcgaa ccgcttcggc 1080
 ttcgtggccg agcatcgcg catgttctcg cgcctcggga tcacgcccgc cgaggtggag 1140
 cggctgcgga ccgagcacgg ggtctacatg gtgggcgatt cgcggctgaa catcgcgggg 1200
 ctgaaccgga cgaccgtgcc ggtgctggcg cgcgcggtgg ccaagggtgt gcgcggctga 1260

<210> 4
 <211> 419
 <212> PRT
 <213> Rhodobacter sphaeroides
 <400> 4

Met Arg Ser Thr Thr Ala Pro Gly Pro Ser Gly Ala Cys Met Thr Ile
1 5 10 15

Ser Arg Ser Arg Lys Asp Asp Glu Gly Met Leu Thr Ala Leu Lys Pro
20 25 30

Gln Pro Ala Asp Lys Ile Leu Gln Leu Ile Gln Met Phe Arg Glu Asp
35 40 45

Ala Arg Ala Asp Lys Ile Asp Leu Gly Val Gly Val Tyr Lys Asp Pro
50 55 60

Thr Gly Leu Thr Pro Val Met Arg Ala Val Lys Ala Ala Glu Lys Arg
65 70 75 80

Leu Trp Glu Val Glu Thr Thr Lys Thr Tyr Thr Gly Leu Ala Asp Glu
85 90 95

Pro Ala Tyr Asn Ala Ala Met Ala Lys Leu Ile Leu Ala Gly Ala Val
100 105 110

Pro Ala Asp Arg Val Ala Ser Val Ala Thr Pro Gly Gly Thr Gly Ala
115 120 125

Val Arg Gln Ala Leu Glu Leu Ile Arg Met Ala Ser Pro Glu Ala Thr
130 135 140

Val Trp Ile Ser Asn Pro Thr Trp Pro Asn His Leu Ser Ile Val Lys
145 150 155 160

Tyr Leu Gly Ile Pro Met Arg Glu Tyr Arg Tyr Phe Asp Ala Glu Thr
165 170 175

Gly Ala Val Asp Ala Glu Gly Met Met Glu Asp Leu Ala Gln Val Lys
180 185 190

Ala Gly Asp Val Val Leu Leu His Gly Cys Cys His Asn Pro Thr Gly
195 200 205

Ala Asn Pro Asn Pro Val Gln Trp Leu Ala Ile Cys Glu Ser Leu Ala
210 215 220

Arg Thr Gly Ala Val Pro Leu Ile Asp Leu Ala Tyr Gln Gly Phe Gly
225 230 235 240

Asp Gly Leu Glu Met Asp Ala Ala Ala Thr Arg Leu Leu Ala Thr Arg
245 250 255

Leu Pro Glu Val Leu Ile Ala Ala Ser Cys Ser Lys Asn Phe Gly Ile
260 265 270

Tyr Arg Glu Arg Thr Gly Ile Leu Ile Ala Ile Gly Glu Ala Ala Gly
275 280 285

Arg Gly Thr Val Gln Ala Asn Leu Asn Phe Leu Asn Arg Gln Asn Tyr
290 295 300

Ser Phe Pro Pro Asp His Gly Ala Arg Leu Val Thr Met Ile Leu Glu
305 310 315 320

Asp Glu Thr Leu Ser Ala Asp Trp Lys Ala Glu Leu Glu Glu Val Arg
 325 330 335

Leu Asn Met Leu Thr Leu Arg Arg Gln Leu Ala Asp Ala Leu Gln Ala
 340 345 350

Glu Thr Gly Ser Asn Arg Phe Gly Phe Val Ala Glu His Arg Gly Met
 355 360 365

Phe Ser Arg Leu Gly Ile Thr Pro Ala Glu Val Glu Arg Leu Arg Thr
 370 375 380

Glu His Gly Val Tyr Met Val Gly Asp Ser Arg Leu Asn Ile Ala Gly
 385 390 395 400

Leu Asn Arg Thr Thr Val Pro Val Leu Ala Arg Ala Val Ala Lys Val
 405 410 415

Leu Arg Gly

<210> 5
 <211> 1260
 <212> DNA
 <213> Rhodobacter sphaeroides

<400> 5
 atgcgctcta cgacggctcc tgggtccgagt ggggcatgta tgacgatctc aaggtcgcga 60
 aaggatgacg aaggaatgct gaccgccctg aagccgcagc ccgcggacaa gatcctgcaa 120
 ctgatccaga tgttccgcga ggatgcgcgc gcggacaaga tcgatctggg cgtgggcgtc 180
 tacaaggacc cgaccgggct caccocggctc atgcgggccc tgaaggccgc cgagaagcgg 240
 ctctgggagg tcgagaccac caagacctac accggccttg ccggcgagcc cgcctacaat 300
 gccgcgatgg cgaagctgat cctcgcaggc ggggtcccgg ccgaccgggt ggccctcggtc 360
 gccacccccg gcggcacggg cgcggtgcgt caggcgcctc agctgatccg catggcctcg 420
 cccgaggcca ctgtctggat ctgaaccgc acctggccga accatctgtc gatcgtgaaa 480
 tatctcggca tcccgatgcg ggaataccgc tatttcgacg ccgagaccgg cgcctcgat 540
 gccgagggtc tgatggagga tctggcccag gtgaaggcgg gcgacgtggt gctgctgcac 600
 ggctgctgcc acaaccgcac cggcgccaac ccgaaccogg tgcagtggct ggccgtctgc 660
 gagagcctgg cccggacagg cgcggtgccg ctgatcgacc tcgctatca gggcttcggc 720
 gacgggctcg agatggatgc ggcggcgacg cggtttctgg ccaccagact gcccgagggtg 780
 ctgatcgagg cctcctgctc gaagaacttc ggcattctacc gcgagcgaac gggcattcctg 840

atcgccatcg gcgaggcggc gggccggggc acggtgcagg ccaacctcaa cttcctgaac 900
 cggcagaact actccttccc gccggaccat ggcgcgcggc tcgtgaccat gatcctcgag 960
 gacgagacgc tgagcgccga ctggaaggcg gaactcgagg aggtgctggct caacatgctg 1020
 acgctgcgcc gccagcttgc cgatgcgctg caggccgaga ccggctcgaa ccgcttcggc 1080
 ttctgtggccg agcatcgcg catgttctcg cgcttcggga tcacgcccgc cgaggtggag 1140
 cggctgcgga ccgagcacgg ggtctacatg gtgggcgatt cgcggctgaa catcgcgggg 1200
 ctgaaccgga cgaccgtgcc ggtgctggcg cgcgcggtgg ccaaggtgct gcgcggctga 1260

<210> 6
 <211> 419
 <212> PRT
 <213> Rhodobacter sphaeroides

<400> 6

Met Arg Ser Thr Thr Ala Pro Gly Pro Ser Gly Ala Cys Met Thr Ile
 1 5 10 15

Ser Arg Ser Arg Lys Asp Asp Glu Gly Met Leu Thr Ala Leu Lys Pro
 20 25 30

Gln Pro Ala Asp Lys Ile Leu Gln Leu Ile Gln Met Phe Arg Glu Asp
 35 40 45

Ala Arg Ala Asp Lys Ile Asp Leu Gly Val Gly Val Tyr Lys Asp Pro
 50 55 60

Thr Gly Leu Thr Pro Val Met Arg Ala Val Lys Ala Ala Glu Lys Arg
 65 70 75 80

Leu Trp Glu Val Glu Thr Thr Lys Thr Tyr Thr Gly Leu Ala Gly Glu
 85 90 95

Pro Ala Tyr Asn Ala Ala Met Ala Lys Leu Ile Leu Ala Gly Ala Val
 100 105 110

Pro Ala Asp Arg Val Ala Ser Val Ala Thr Pro Gly Gly Thr Gly Ala
 115 120 125

Val Arg Gln Ala Leu Glu Leu Ile Arg Met Ala Ser Pro Glu Ala Thr
 130 135 140

Val Trp Ile Ser Asn Pro Thr Trp Pro Asn His Leu Ser Ile Val Lys
 145 150 155 160

Tyr Leu Gly Ile Pro Met Arg Glu Tyr Arg Tyr Phe Asp Ala Glu Thr
 165 170 175

Gly Ala Val Asp Ala Glu Gly Leu Met Glu Asp Leu Ala Gln Val Lys
 180 185 190

Ala Gly Asp Val Val Leu Leu His Gly Cys Cys His Asn Pro Thr Gly
 195 200 205

Ala Asn Pro Asn Pro Val Gln Trp Leu Ala Val Cys Glu Ser Leu Ala
 210 215 220

Arg Thr Gly Ala Val Pro Leu Ile Asp Leu Ala Tyr Gln Gly Phe Gly
 225 230 235 240

Asp Gly Leu Glu Met Asp Ala Ala Ala Thr Arg Leu Leu Ala Thr Arg
 245 250 255

Leu Pro Glu Val Leu Ile Ala Ala Ser Cys Ser Lys Asn Phe Gly Ile
 260 265 270

Tyr Arg Glu Arg Thr Gly Ile Leu Ile Ala Ile Gly Glu Ala Ala Gly
 275 280 285

Arg Gly Thr Val Gln Ala Asn Leu Asn Phe Leu Asn Arg Gln Asn Tyr
 290 295 300

Ser Phe Pro Pro Asp His Gly Ala Arg Leu Val Thr Met Ile Leu Glu
 305 310 315 320

Asp Glu Thr Leu Ser Ala Asp Trp Lys Ala Glu Leu Glu Glu Val Arg
 325 330 335

Leu Asn Met Leu Thr Leu Arg Arg Gln Leu Ala Asp Ala Leu Gln Ala
 340 345 350

Glu Thr Gly Ser Asn Arg Phe Gly Phe Val Ala Glu His Arg Gly Met
 355 360 365

Phe Ser Arg Leu Gly Ile Thr Pro Ala Glu Val Glu Arg Leu Arg Thr
 370 375 380

Glu His Gly Val Tyr Met Val Gly Asp Ser Arg Leu Asn Ile Ala Gly
 385 390 395 400

Leu Asn Arg Thr Thr Val Pro Val Leu Ala Arg Ala Val Ala Lys Val
 405 410 415

Leu Arg Gly

<210> 7
 <211> 1239
 <212> DNA
 <213> Leishmania major

<400> 7
 atgtccatgc aggcggccat gaccacggcg gacgctggc agaagattca ggcacaagct 60
 cccgatgtca tcttcgatct cgcaaaacgc gccgcccgtg ccaagggccc caaggccaac 120
 ctgctcattg gtgcctaccg cgacgagcag ggccgtccct atccgctacg cgtgggtccg 180
 aaggctgagc agcttctctt ggacatgaat ctgcactacg agtacctacc catctcgggc 240
 taccagccct tcatcgatga ggccggtaaag attatctacg gcaataccgt cgagctggag 300
 aacctgggtg cgggtgcagac gctgagcggg accggtgctg tctctctcgg ggcgaagctg 360
 ctgactcgcg tcttcgacgc tgagacgacg cccatctacc ttcccgaccc cactgtggcc 420
 aaccactacg gcgtcgtgaa ggctgctggc tggaagaaca tctgcacgta cgcctactac 480
 gacccaaga cggtcagcct gaatttcgag ggcatgaaga aagacattct ggccggcgcg 540
 gacggctccg tgttcattct gcaccagtgc gcgcacaacc ccaccggcgt ggaccgcgtc 600
 caggagcagt ggaacgagat cgcgtcactg atgctggcca agcaccatca ggtgttcttc 660
 gactccgcct accaaggcta tgcgagcggc agcctcgaca cggacgcgta tctgtcccgc 720
 ctgtttgccc gccgcggcat cgaggtactg ctggcgcagt cgttctccaa gaacatgggc 780
 ttgtacagcg agcgtgcagg cactgtgtcg ctgctcctca aggacaagac gaagcgcgcg 840
 gatgtaaaga gcgtgatgga ttcgctgac cgtgaggagt acacgtgccc ccagccac 900
 ggtgcccgt tagccacct aatcctgagc aacaacgaac tgcgaaagga gtgggaggca 960
 gagctatcag ccatggcaga gcgcacccgt acgatgcgcc gcaccgtgta cgacgagctg 1020
 ctgcgcctgc agacgcccgg gagctgggaa catgtcatta accagattgg catgttttcc 1080
 ttctcgggc tgtcaaaggc gcagtgcgaa tactgcaaaa accacaacat cttcatcaca 1140
 gtgtcgggcc gcgctaacat ggcaggtctg acgcatgaga cggcgcgtgat gctagcacag 1200
 acgatcaacg atgctgtgcg caatgtgaat cgtgagtga 1239

<210> 8
 <211> 412
 <212> PRT

<213> Leishmania major

<400> 8

Met Ser Met Gln Ala Ala Met Thr Thr Ala Glu Arg Trp Gln Lys Ile
1 5 10 15

Gln Ala Gln Ala Pro Asp Val Ile Phe Asp Leu Ala Lys Arg Ala Ala
20 25 30

Ala Ala Lys Gly Pro Lys Ala Asn Leu Val Ile Gly Ala Tyr Arg Asp
35 40 45

Glu Gln Gly Arg Pro Tyr Pro Leu Arg Val Val Arg Lys Ala Glu Gln
50 55 60

Leu Leu Leu Asp Met Asn Leu Asp Tyr Glu Tyr Leu Pro Ile Ser Gly
65 70 75 80

Tyr Gln Pro Phe Ile Asp Glu Ala Val Lys Ile Ile Tyr Gly Asn Thr
85 90 95

Val Glu Leu Glu Asn Leu Val Ala Val Gln Thr Leu Ser Gly Thr Gly
100 105 110

Ala Val Ser Leu Gly Ala Lys Leu Leu Thr Arg Val Phe Asp Ala Glu
115 120 125

Thr Thr Pro Ile Tyr Leu Ser Asp Pro Thr Trp Pro Asn His Tyr Gly
130 135 140

Val Val Lys Ala Ala Gly Trp Lys Asn Ile Cys Thr Tyr Ala Tyr Tyr
145 150 155 160

Asp Pro Lys Thr Val Ser Leu Asn Phe Glu Gly Met Lys Lys Asp Ile
165 170 175

Leu Ala Ala Pro Asp Gly Ser Val Phe Ile Leu His Gln Cys Ala His
180 185 190

Asn Pro Thr Gly Val Asp Pro Ser Gln Glu Gln Trp Asn Glu Ile Ala
195 200 205

Ser Leu Met Leu Ala Lys His His Gln Val Phe Phe Asp Ser Ala Tyr
210 215 220

Gln Gly Tyr Ala Ser Gly Ser Leu Asp Thr Asp Ala Tyr Ala Ala Arg

225					230					235					240
Leu	Phe	Ala	Arg	Arg	Gly	Ile	Glu	Val	Leu	Leu	Ala	Gln	Ser	Phe	Ser
			245						250					255	
Lys	Asn	Met	Gly	Leu	Tyr	Ser	Glu	Arg	Ala	Gly	Thr	Leu	Ser	Leu	Leu
			260					265					270		
Leu	Lys	Asp	Lys	Thr	Lys	Arg	Ala	Asp	Val	Lys	Ser	Val	Met	Asp	Ser
		275					280						285		
Leu	Ile	Arg	Glu	Glu	Tyr	Thr	Cys	Pro	Pro	Ala	His	Gly	Ala	Arg	Leu
	290					295					300				
Ala	His	Leu	Ile	Leu	Ser	Asn	Asn	Glu	Leu	Arg	Lys	Glu	Trp	Glu	Ala
305					310					315					320
Glu	Leu	Ser	Ala	Met	Ala	Glu	Arg	Ile	Arg	Thr	Met	Arg	Arg	Thr	Val
				325					330					335	
Tyr	Asp	Glu	Leu	Leu	Arg	Leu	Gln	Thr	Pro	Gly	Ser	Trp	Glu	His	Val
			340					345					350		
Ile	Asn	Gln	Ile	Gly	Met	Phe	Ser	Phe	Leu	Gly	Leu	Ser	Lys	Ala	Gln
		355					360						365		
Cys	Glu	Tyr	Cys	Gln	Asn	His	Asn	Ile	Phe	Ile	Thr	Val	Ser	Gly	Arg
	370					375					380				
Ala	Asn	Met	Ala	Gly	Leu	Thr	His	Glu	Thr	Ala	Leu	Met	Leu	Ala	Gln
385					390					395					400
Thr	Ile	Asn	Asp	Ala	Val	Arg	Asn	Val	Asn	Arg	Glu				
				405					410						

```
<210> 9
<211> 1182
<212> DNA
<213> Bacillus subtilis
```

```

<400> 9
atggaacatt tgctgaatcc gaaagcaaga gagatcgaa tttcaggaat acgcaaattc 60
tcgaatctt cagcccaaca cgaagacgtc atttcactta caatcggcca gcctgatttt 120
ttcacaccgc atcatgtgaa agctgccgca aaaaaagcca ttgatgaaaa cgtgacgtca 180
tatactccga atgccggcta cctggagctg agacaagctg tgcagcttta tatqaagaaa 240

```

aaagcggatt tcaactatga tgctgaatct gaaattatca tcacaacagg cgcaagccaa 300
 gccattgatg ctgcattccg gacgatttta tctcccgtg atgaagtcac tatgccaggg 360
 cctattttatc cgggctatga acctattatc aatttgtgcg gggccaagcc tgtcattggt 420
 gatactacgt cacacggctt taagcttacc gcccggtga ttgaagatgc tctgacaccc 480
 aacaccaagt gtgtcgtgct tccttatccg tcaaacccta ccggcgtgac tttatctgaa 540
 yaagaactga aaagcatcgc agctctctta aaaggcagaa atgtcttcgt attgtctgat 600
 gaaatataca gtgaattaac atatgacaga ccgcattact ccatcgcaac ctatttgcgg 660
 gatcaaacga ttgtcattaa cgggttgta aaatcacaca gcatgaccgg ttggagaatt 720
 ggatttttat ttgcaccgaa agacattgca aagcacattt taaaggttca tcaatacaat 780
 gtgtcgtgcg cctcatccat ttctcaaaaa gccgcgcttg aagctgtcac aaacggcttt 840
 gacgatgcat tgattatgag agaacaatac aaaaaacgtc tggactatgt ttatgaccgt 900
 cttgtttcca tgggacttga cgtagttaaa ccgtccggtg cgttttatat cttcccttct 960
 attaaatcat ttggaatgac ttcatittgat tttagtatgg ctcttttgga agacgctggc 1020
 gtggcactcg tgccgggcag ctggttctca acatattggtg aaggatatgt aaggctgtct 1080
 tttgcatgct caatggacac gctgagagaa ggcctagacc gtttagaatt atttgtatta 1140
 aaaaaacgtg aagcaatgca gacgataaac aacggcgctt aa 1182

<210> 10
 <211> 393
 <212> PRT
 <213> *Bacillus subtilis*

<400> 10

Met Glu His Leu Leu Asn Pro Lys Ala Arg Glu Ile Glu Ile Ser Gly
 1 5 10 15

Ile Arg Lys Phe Ser Asn Leu Val Ala Gln His Glu Asp Val Ile Ser
 20 25 30

Leu Thr Ile Gly Gln Pro Asp Phe Phe Thr Pro His His Val Lys Ala
 35 40 45

Ala Ala Lys Lys Ala Ile Asp Glu Asn Val Thr Ser Tyr Thr Pro Asn
 50 55 60

Ala Gly Tyr Leu Glu Leu Arg Gln Ala Val Gln Leu Tyr Met Lys Lys
 65 70 75 80

Lys Ala Asp Phe Asn Tyr Asp Ala Glu Ser Glu Ile Ile Ile Thr Thr

85	90	95
Gly Ala Ser Gln Ala Ile Asp Ala Ala Phe Arg Thr Ile Leu Ser Pro 100 105 110		
Gly Asp Glu Val Ile Met Pro Gly Pro Ile Tyr Pro Gly Tyr Glu Pro 115 120 125		
Ile Ile Asn Leu Cys Gly Ala Lys Pro Val Ile Val Asp Thr Thr Ser 130 135 140		
His Gly Phe Lys Leu Thr Ala Arg Leu Ile Glu Asp Ala Leu Thr Pro 145 150 155 160		
Asn Thr Lys Cys Val Val Leu Pro Tyr Pro Ser Asn Pro Thr Gly Val 165 170 175		
Thr Leu Ser Glu Glu Glu Leu Lys Ser Ile Ala Ala Leu Leu Lys Gly 180 185 190		
Arg Asn Val Phe Val Leu Ser Asp Glu Ile Tyr Ser Glu Leu Thr Tyr 195 200 205		
Asp Arg Pro His Tyr Ser Ile Ala Thr Tyr Leu Arg Asp Gln Thr Ile 210 215 220		
Val Ile Asn Gly Leu Ser Lys Ser His Ser Met Thr Gly Trp Arg Ile 225 230 235 240		
Gly Phe Leu Phe Ala Pro Lys Asp Ile Ala Lys His Ile Leu Lys Val 245 250 255		
His Gln Tyr Asn Val Ser Cys Ala Ser Ser Ile Ser Gln Lys Ala Ala 260 265 270		
Leu Glu Ala Val Thr Asn Gly Phe Asp Asp Ala Leu Ile Met Arg Glu 275 280 285		
Gln Tyr Lys Lys Arg Leu Asp Tyr Val Tyr Asp Arg Leu Val Ser Met 290 295 300		
Gly Leu Asp Val Val Lys Pro Ser Gly Ala Phe Tyr Ile Phe Pro Ser 305 310 315 320		
Ile Lys Ser Phe Gly Met Thr Ser Phe Asp Phe Ser Met Ala Leu Leu 325 330 335		

Glu Asp Ala Gly Val Ala Leu Val Pro Gly Ser Ser Phe Ser Thr Tyr
 340 345 350

Gly Glu Gly Tyr Val Arg Leu Ser Phe Ala Cys Ser Met Asp Thr Leu
 355 360 365

Arg Glu Gly Leu Asp Arg Leu Glu Leu Phe Val Leu Lys Lys Arg Glu
 370 375 380

Ala Met Gln Thr Ile Asn Asn Gly Val
 385 390

<210> 11
 <211> 1176
 <212> DNA
 <213> Lactobacillus amylovorus

<400> 11
 atgccagaat tagctaata tttaggatta agcaaaaaga tcactgatgt aaaagcttca 60
 ggaattagaa tctttgataa caaagtttca gctattcctg gcattatcaa attgactttg 120
 ggtgaaccag atatgaatac tcctgagcat gttaagcaag cggctattaa gaatattgca 180
 gataatgatt cacactatgc tccacaaaag ggaaagcttg aattaagaaa agctatcagt 240
 aaatatttga aaaagattac tggaattgaa tatgatccag aacagaaat cgtagtaaca 300
 gttggtgcaa ctgaagcaat taacgctacc ttgtttgcta ttactaatcc gggtgacaag 360
 gttgcaattc ctaogccagt cttttctcta tattggcccg tggctacact tgctgatgcc 420
 gattatgttt tgatgaatac tgcagaagat ggttttaagt taacacctaa gaagttagaa 480
 gaaactatca aagaaaatcc aacaattaaa gcagtaattt tgaattatcc aactaaccca 540
 actggtgttg aatatagcga agatgaaatt aaagctttgg ctaaggtaat taaagataat 600
 catctgtacg taattaccga tgaaatttac agtactttga cttacggtgt aaaacacttt 660
 tcaattgcca gcttaattcc agaaagagca atttatatct ctggtttatc taaatcacat 720
 gcgatgactg gttatcgttt aggctatggt gccggacctg caaaaattat ggcagaaatt 780
 ggtaaagttc atggccttat ggtgacgact acgacggatt catcacaagc tgccgcaatt 840
 gaagcacttg aacacggact tgatgacoot gagaaatata ~~gggaagttta~~ tgaaaagcgt 900
 cgtgactatg ~~tttttaagga~~ attagccgag atagagatgc aagcagttaa gccagaaggt 960
 gcattttata tctttgctaa aattccagct aagtatggca aagacgatat gaaatttgcc 1020
 ttggatttag cttttaaaga aaaagtgggt atcaactccag gtagtgcatt tggctctggt 1080
 ggtgaaggtc atattagatt atcttatgca tcaagtgatg aaaacttgca tgaggcaatg 1140

aagcgaatga agaaagtttt acaagaggac gaataa

1176

<210> 12

<211> 391

<212> PRT

<213> Lactobacillus amylovorus

<400> 12

Met Pro Glu Leu Ala Asn Asp Leu Gly Leu Ser Lys Lys Ile Thr Asp
1 5 10 15

Val Lys Ala Ser Gly Ile Arg Ile Phe Asp Asn Lys Val Ser Ala Ile
20 25 30

Pro Gly Ile Ile Lys Leu Thr Leu Gly Glu Pro Asp Met Asn Thr Pro
35 40 45

Glu His Val Lys Gln Ala Ala Ile Lys Asn Ile Ala Asp Asn Asp Ser
50 55 60

His Tyr Ala Pro Gln Lys Gly Lys Leu Glu Leu Arg Lys Ala Ile Ser
65 70 75 80

Lys Tyr Leu Lys Lys Ile Thr Gly Ile Glu Tyr Asp Pro Glu Thr Glu
85 90 95

Ile Val Val Thr Val Gly Ala Thr Glu Ala Ile Asn Ala Thr Leu Phe
100 105 110

Ala Ile Thr Asn Pro Gly Asp Lys Val Ala Ile Pro Thr Pro Val Phe
115 120 125

Ser Leu Tyr Trp Pro Val Ala Thr Leu Ala Asp Ala Asp Tyr Val Leu
130 135 140

Met Asn Thr Ala Glu Asp Gly Phe Lys Leu Thr Pro Lys Lys Leu Glu
145 150 155 160

Glu Thr Ile Lys Glu Asn Pro Thr Ile Lys Ala Val Ile Leu Asn Tyr
165 170 175

Pro Thr Asn Pro Thr Gly Val Glu Tyr Ser Glu Asp Glu Ile Lys Ala
180 185 190

Leu Ala Lys Val Ile Lys Asp Asn His Leu Tyr Val Ile Thr Asp Glu
195 200 205

Ile Tyr Ser Thr Leu Thr Tyr Gly Val Lys His Phe Ser Ile Ala Ser
210 215 220

Leu Ile Pro Glu Arg Ala Ile Tyr Ile Ser Gly Leu Ser Lys Ser His
225 230 235 240

Ala Met Thr Gly Tyr Arg Leu Gly Tyr Val Ala Gly Pro Ala Lys Ile
245 250 255

Met Ala Glu Ile Gly Lys Val His Gly Leu Met Val Thr Thr Thr Thr
260 265 270

Asp Ser Ser Gln Ala Ala Ala Ile Glu Ala Leu Glu His Gly Leu Asp
275 280 285

Asp Pro Glu Lys Tyr Arg Glu Val Tyr Glu Lys Arg Arg Asp Tyr Val
290 295 300

Leu Lys Glu Leu Ala Glu Ile Glu Met Gln Ala Val Lys Pro Glu Gly
305 310 315 320

Ala Phe Tyr Ile Phe Ala Lys Ile Pro Ala Lys Tyr Gly Lys Asp Asp
325 330 335

Met Lys Phe Ala Leu Asp Leu Ala Phe Lys Glu Lys Val Gly Ile Thr
340 345 350

Pro Gly Ser Ala Phe Gly Pro Gly Gly Glu Gly His Ile Arg Leu Ser
355 360 365

Tyr Ala Ser Ser Asp Glu Asn Leu His Glu Ala Met Lys Arg Met Lys
370 375 380

Lys Val Leu Gln Glu Asp Glu
385 390

<210> 13
<211> 1413
<212> DNA
<213> R. sphaeroides

<400> 13
atgcgcgagc ctcttgccct cgagatcgac ccgggccacg gcgcccgcgt gttcctcgcc 60
atcgccgagg cgatcaccct cgacatcacc cgcgggcggc tgaggcccgagg agcgagactg 120
cccggcacac gcgcgctggc gcgggcgctc ggcgtgcatc gcaacacggt ggatgccgcc 180

tatcaggagt tgctgacca gggctggctg caggccgagc ccgcgcgggg caccttcgtg 240
 gcgcaggatc tgccgcaggg gatgctgggt cacaggcccg cgcgcgcgcc ggtcgagccg 300
 gtcgcgatgc gcgcggggct cgccttctcc gatggcgcgc cggaccccca gctgggtccc 360
 gacaaggcgc tggcgcgggc ctttcgcggg gcgctcctgt cgcgcgcctt ccgcgcggga 420
 gcggattacg gcgatgcccg cggcacctcc tcgctgcggg aggcgctggc agcctatctc 480
 gcctcggacc ggggcgtggg cgcggatcct gcgcggctgc tgatcgcgcg gggcagccag 540
 atggcgctgt tcctggtagc ccgggcggcg ctggcgcccg gagaggcgat cgcggctgag 600
 gagccgggct atccgctggc ctgggaggcg ttccgcgcag cgggagcgga ggtgcgcggc 660
 gtgccgggtg atggcgggcg cctcaggatc gacgcgctcg aggcgcgctt ggcccgggat 720
 ccgcgaatcc gggcggtcta tgtcacgcc catcaccagt atccgacgac cgtcaccatg 780
 ggcgcggcgc ggcggttgca gcttctggaa ctggcagagc gccaccggct cgcgctgac 840
 gaggacgact acgaccacga ataccgcttc gagggccgtc cggctgctgc gctggctgcc 900
 cgcgcgcccg aaggctctgc gctgatctat gtgggctcgc tgcgaaact gctctcgc 960
 ggtatccggc tgggatacgc gctggcgccc gagcggctgc tgaccgcgat ggccgcggcg 1020
 cgcgccgcca tcgaccggca gggcgacgc cgcctcgagg cggcgctggc cgagctgac 1080
 cgcgacggcg atctgggccc tcatgcccgc aaggcgcgca gggctctacc ggcgcggcgg 1140
 gatctgctgg cggagcgtct cacggcgag ctggccgggc gcgcgcctt cgatctgcc 1200
 gccggggggc tcgcgctgtg gctgcgctgc gcgggcgtct cggccgagac ctgggccgaa 1260
 gccgcagggc aggcggggct cgccttgctg ccgggcacgc gcttcgcgct ggagagccc 1320
 gcgccgcagg ccttcgggct gggctatgcg gcgctggacg aggggcagat cgcgcggcg 1380
 gtggagatcc tcgcccggag cttccccggc tga 1413

<210> 14
 <211> 470
 <212> PRT
 <213> R. sphaeroides

<400> 14

Met Arg Glu Pro Leu Ala Leu Glu Ile Asp Pro Gly His Gly Gly Pro
1 5 10 15

Leu Phe Leu Ala Ile Ala Glu Ala Ile Thr Leu Asp Ile Thr Arg Gly
20 25 30

Arg Leu Arg Pro Gly Ala Arg Leu Pro Gly Thr Arg Ala Leu Ala Arg
35 40 45

Ala Leu Gly Val His Arg Asn Thr Val Asp Ala Ala Tyr Gln Glu Leu
50 55 60

Leu Thr Gln Gly Trp Leu Gln Ala Glu Pro Ala Arg Gly Thr Phe Val
65 70 75 80

Ala Gln Asp Leu Pro Gln Gly Met Leu Val His Arg Pro Ala Pro Ala
85 90 95

Pro Val Glu Pro Val Ala Met Arg Ala Gly Leu Ala Phe Ser Asp Gly
100 105 110

Ala Pro Asp Pro Glu Leu Val Pro Asp Lys Ala Leu Ala Arg Ala Phe
115 120 125

Arg Arg Ala Leu Leu Ser Pro Ala Phe Arg Ala Gly Ala Asp Tyr Gly
130 135 140

Asp Ala Arg Gly Thr Ser Ser Leu Arg Glu Ala Leu Ala Ala Tyr Leu
145 150 155 160

Ala Ser Asp Arg Gly Val Val Ala Asp Pro Ala Arg Leu Leu Ile Ala
165 170 175

Arg Gly Ser Gln Met Ala Leu Phe Leu Val Ala Arg Ala Ala Leu Ala
180 185 190

Pro Gly Glu Ala Ile Ala Val Glu Glu Pro Gly Tyr Pro Leu Ala Trp
195 200 205

Glu Ala Phe Arg Ala Ala Gly Ala Glu Val Arg Gly Val Pro Val Asp
210 215 220

Gly Gly Gly Leu Arg Ile Asp Ala Leu Glu Ala Ala Leu Ala Arg Asp
225 230 235 240

Pro Arg Ile Arg Ala Val Tyr Val Thr Pro His His Gln Tyr Pro Thr
245 250 255

Thr Val Thr Met Gly Ala Ala Arg Arg Leu Gln Leu Leu Glu Leu Ala
260 265 270

Glu Arg His Arg Leu Ala Leu Ile Glu Asp Asp Tyr Asp His Glu Tyr
275 280 285

Arg Phe Glu Gly Arg Pro Val Leu Pro Leu Ala Ala Arg Ala Pro Glu
290 295 300

Gly Leu Pro Leu Ile Tyr Val Gly Ser Leu Ser Lys Leu Leu Ser Pro
305 310 315 320

Gly Ile Arg Leu Gly Tyr Ala Leu Ala Pro Glu Arg Leu Leu Thr Arg
325 330 335

Met Ala Ala Ala Arg Ala Ala Ile Asp Arg Gln Gly Asp Ala Pro Leu
340 345 350

Glu Ala Ala Leu Ala Glu Leu Ile Arg Asp Gly Asp Leu Gly Arg His
355 360 365

Ala Arg Lys Ala Arg Arg Val Tyr Arg Ala Arg Arg Asp Leu Leu Ala
370 375 380

Glu Arg Leu Thr Ala Gln Leu Ala Gly Arg Ala Ala Phe Asp Leu Pro
385 390 395 400

Ala Gly Gly Leu Ala Leu Trp Leu Arg Cys Ala Gly Val Ser Ala Glu
405 410 415

Thr Trp Ala Glu Ala Ala Gly Gln Ala Gly Leu Ala Leu Leu Pro Gly
420 425 430

Thr Arg Phe Ala Leu Glu Ser Pro Ala Pro Gln Ala Phe Arg Leu Gly
435 440 445

Tyr Ala Ala Leu Asp Glu Gly Gln Ile Ala Arg Ala Val Glu Ile Leu
450 455 460

Ala Arg Ser Phe Pro Gly
465 470

<210> 15
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 15
ggatttgagg gtcgcatgaa ggttttagtc aatgg

<210> 16
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 16
agaggagagt tagagcctta tgaaatgcta gcagcct

37

<210> 17
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 17
ggtattgagg gtcgcatggt cgacgccctc gcccg

35

<210> 18
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 18
agaggagagt tagagcctca gagactgggtg aacttgc

37

<210> 19
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 19
ggtattgagg gtcgcatgga acatttgctg aatcc

35

<210> 20
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 20
agaggagagt tagagcctta aacgccgttg tttatcg

37

<210> 21

<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 21
ggtattgagg gtcgcatgcg cgagcctctt gccct

35

<210> 22
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 22
agaggagagt tagagcctca gccggggaag ctccggg

37

<210> 23
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 23
ggtattgagg gtcgcatgtc caccgaggcg gccat

35

<210> 24
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 24
agaggagagt tagagcctca ctcacgattc acattgc

37

<210> 25
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 25
ggtattgagg gtcgcatgcc agaattagct aatga

35

<210> 26
<211> 37

<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 26
agaggagagt tagagcctta ttcgtcctct tgtaaaa

37

<210> 27
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 27
ggtattgagg gtcgcatgcg ctctacgacg gctcc

35

<210> 28
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 28
agaggagagt tagagcctca gccgcgcagc accttgg

37

<210> 29
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 29
ggtattgagg gtcgcatggt tgagaacatt accgc

35

<210> 30
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 30
agaggagagt tagagcctta cagcactgcc acaatcg

37

<210> 31
<211> 1194
<212> DNA

<213> E. coli

<400> 31

gtgtttcaaa aagttgacgc ctacgctggc gacccgattc ttacgcttat ggagcgtttt 60
aaagaagacc ctgcgacgca caaagtgaat ttaagtatcg gtctgtacta caacgaagac 120
ggaattattc cacaactgca agccgtggcg gaggcggaag cgcgctgaa tgcgcagcct 180
catggcgctt cgctttattt accgatggaa gggcttaact gctatcgcca tgccattgcg 240
ccgctgctgt ttggtgcgga ccatccggta ctgaaacaac agcgcgtagc aaccattcaa 300
acccttggcg gctccggggc attgaaagtg ggcgcggatt tcctgaaacg ctacttcccg 360
gaatcaggcg tctgggtcag cgatccctacc tgggaaaacc gcgtagcaat attcgccggg 420
gctggattcg aagtgagtac ttaccctcgg tatgacgaag cgactaacgg cgtgcgcttt 480
aatgacctgt tggcgacgct gaaaacatta cctgcccgca gtattgtgtt gctgcatcca 540
tgttgccaca acccaacggg tgccgatctc actaatgatc agtgggatgc ggtgattgaa 600
attctcaaag cccgcgagct tattccattc ctcgatattg cctatcaagg atttggtgcc 660
ggtatggaag aggatgccta cgctattcgc gccattgcca gcgctggatt acccgctctg 720
gtgagcaatt cgttctcgaa aattttctcc ctttacggcg agcgcgctcg cggaactttct 780
gttatgtgtg aagatgccga agccgctggc cgcgctactg ggcaattgaa agcaacagtt 840
cgccgcaact actccagccc gccgaatttt ggtgcgcagg tgggtggctgc agtgctgaat 900
gacgaggcat tgaaagccag ctggctggcg gaagtagaag agatgcgtac tcgcattctg 960
gcaatgcgtc aggaattggg gaaggtatta agcacagaga tgccagaacg caatttcgat 1020
tatctgctta atcagcgcgg catgttcagt tataccgggt taagtgcgcg tcagggtgac 1080
cgactacgtg aagaatttgg tgtctatctc atcgccagcg gtgcgcatgt tgcgcgggg 1140
ttaaatacgg caaatgtaca acgtgtggca aaggcgtttg ctgcggtgat gtaa 1194

<210> 32

<211> 397

<212> PRT

<213> E. coli

<400> 32

Val Phe Gln Lys Val Asp Ala Tyr Ala Gly Asp Pro Ile Leu Thr Leu
1 5 10 15

Met Glu Arg Phe Lys Glu Asp Pro Arg Ser Asp Lys Val Asn Leu Ser
20 25 30

Ile Gly Leu Tyr Tyr Asn Glu Asp Gly Ile Ile Pro Gln Leu Gln Ala
35 40 45

Val Ala Glu Ala Glu Ala Arg Leu Asn Ala Gln Pro His Gly Ala Ser
50 55 60

Leu Tyr Leu Pro Met Glu Gly Leu Asn Cys Tyr Arg His Ala Ile Ala
65 70 75 80

Pro Leu Leu Phe Gly Ala Asp His Pro Val Leu Lys Gln Gln Arg Val
85 90 95

Ala Thr Ile Gln Thr Leu Gly Gly Ser Gly Ala Leu Lys Val Gly Ala
100 105 110

Asp Phe Leu Lys Arg Tyr Phe Pro Glu Ser Gly Val Trp Val Ser Asp
115 120 125

Pro Thr Trp Glu Asn Arg Val Ala Ile Phe Ala Gly Ala Gly Phe Glu
130 135 140

Val Ser Thr Tyr Pro Trp Tyr Asp Glu Ala Thr Asn Gly Val Arg Phe
145 150 155 160

Asn Asp Leu Leu Ala Thr Leu Lys Thr Leu Pro Ala Arg Ser Ile Val
165 170 175

Leu Leu His Pro Cys Cys His Asn Pro Thr Gly Ala Asp Leu Thr Asn
180 185 190

Asp Gln Trp Asp Ala Val Ile Glu Ile Leu Lys Ala Arg Glu Leu Ile
195 200 205

Pro Phe Leu Asp Ile Ala Tyr Gln Gly Phe Gly Ala Gly Met Glu Glu
210 215 220

Asp Ala Tyr Ala Ile Arg Ala Ile Ala Ser Ala Gly Leu Pro Ala Leu
225 230 235 240

Val Ser Asn Ser Phe Ser Lys Ile Phe Ser Leu Tyr Gly Glu Arg Val
245 250 255

Gly Gly Leu Ser Val Met Cys Glu Asp Ala Glu Ala Ala Gly Arg Val
260 265 270

Leu Gly Gln Leu Lys Ala Thr Val Arg Arg Asn Tyr Ser Ser Pro Pro
275 280 285

Asn Phe Gly Ala Gln Val Val Ala Ala Val Leu Asn Asp Glu Ala Leu
 290 295 300

Lys Ala Ser Trp Leu Ala Glu Val Glu Glu Met Arg Thr Arg Ile Leu
 305 310 315 320

Ala Met Arg Gln Glu Leu Val Lys Val Leu Ser Thr Glu Met Pro Glu
 325 330 335

Arg Asn Phe Asp Tyr Leu Leu Asn Gln Arg Gly Met Phe Ser Tyr Thr
 340 345 350

Gly Leu Ser Ala Ala Gln Val Asp Arg Leu Arg Glu Glu Phe Gly Val
 355 360 365

Tyr Leu Ile Ala Ser Gly Arg Met Cys Val Ala Gly Leu Asn Thr Ala
 370 375 380

Asn Val Gln Arg Val Ala Lys Ala Phe Ala Ala Val Met
 385 390 395

<210> 33
 <211> 35
 <212> DNA
 <213> Artificial sequence

<220>
 <223> primer

<400> 33
 ggtattgagg gtcgctgttt tcaaaaagtt gacgc

35

<210> 34
 <211> 37
 <212> DNA
 <213> Artificial sequence

<220>
 <223> primer

<400> 34
 agaggagagt tagagcotta catcacgcga gcaaaacg

37

<210> 35
 <211> 35
 <212> DNA
 <213> Artificial sequence

<220>
 <223> primer

<400> 35
 ggtattgagg gtgcgatgga gtccaaagtc gttga 35

<210> 36
 <211> 37
 <212> DNA
 <213> Artificial sequence

<220>
 <223> primer

<400> 36
 agaggagagt tagagcotta cacttggaac acagcct 37

<210> 37
 <211> 35
 <212> DNA
 <213> Artificial sequence

<220>
 <223> primer

<400> 37
 ggtattgagg gtgcgatgaa aaactggaaa acaag 35

<210> 38
 <211> 37
 <212> DNA
 <213> Artificial sequence

<220>
 <223> primer

<400> 38
 agaggagagt tagagcotta cagcttagcg ccttcta 37

<210> 39
 <211> 35
 <212> DNA
 <213> Artificial sequence

<220>
 <223> primer

<400> 39
 ggtattgagg gtgcgatgag aggggcatta ttcaa 35

<210> 40
 <211> 36
 <212> DNA
 <213> Artificial sequence

<220>
 <223> primer

<400> 40

agaggagagt tagagcctca gcccttgagc gcgaag

36

<210> 41

<211> 1416

<212> DNA

<213> E. coli

<400> 41

atggaaaact ttaaaccatct ccctgaaccg ttccgcattc gtgttattga gccagtaaaa 60
cgtaccacccc gcgcttatcg tgaagaggca attattaaat ccggtatgaa cccgttctctg 120
ctggatagcg aagatgtttt tatcgattta ctgaccgaca gcggcaccgg ggcggtgacg 180
cagagcatgc aggctgcgat gatgcgcggc gacgaagcct acagcggcag tcgtagctac 240
tatgcgttag ccgagtcagt gaaaaatatc tttggttatt aatacaccat tccgactcac 300
cagggccgtg gcgcagagca aatctatatt ccggtactga ttaaaaaacg cgagcaggaa 360
aaaggcctgg atcgcagcaa aatggtggcg ttctctaact atttctttga taccacgcag 420
ggccatagcc agatcaacgg ctgtaccgtg cgtaacgtct atatcaaaga agccttcgat 480
acgggctgtc gttacgactt taaaggcaac tttgaccttg agggattaga acgcggtatt 540
gaagaagttg gtccgaataa cgtgccgtat atcgttgcaa ccatcaccag taactctgca 600
ggtggtcagc cggtttctact ggcaaaactta aaagcgatgt acagcatcgc gaagaaatac 660
gatattccgg tggtaatgga ctccgcgcgc tttgctgaaa acgcctatct catcaagcag 720
cgtgaagcag aatacaaaga ctggaccatc gagcagatca cccgcgaaac ctacaaatat 780
gccgatatgc tggcgatgtc cgccaagaaa gatgcgatgg tgccgatggg cggcctgctg 840
tgcatgaaag acgacagctt ctttgatgtg tacaccgagt gcagaaccct ttgctgggtg 900
caggaaggct tcccgacata tggcggcctg gaaggcggcg cgatggagcg tctggcggta 960
ggtctgtatg acggcatgaa tctcgactgg ctggccttatt gtatcgcgca ggtacagtat 1020
ctggtcgatg gtctggaaga gattggcggt gtctgccagc aggcggggcg tcacgcggca 1080
ttcgttgatg ccggtaaact gttgccgat atcccggcag accagttccc ggcacaggcg 1140
ctggcctgcg agctgtataa agtcgccggt atccgtgcgg tagaaattgg ctctttcctg 1200
ttaggcccgcg atccgaaaac cggtaaacia etgccatgcc cggctgaact gctgcgttta 1260
accattccgc gcgcaacata tactcaaaca catatggact tcattattga agcctttaa 1320
catgtgaaag agaacgcggc gaatattaaa ggattaacct ttacgtacga accgaaagta 1380
ttgogtcact tcaccgcaaa acttaaagaa gtttaa 1416

<210> 42

<211> 1371

<212> DNA

<213> *Citrobacter freundii*

<400> 42

```
atgaattatc cggcagaacc cttccgtatt aaaagcgttg aaactgtatc tatgatcccg      60
cgtgatgaac gcctcaagaa aatgcaggaa gcggtttaca atactttcct gttaaattcg      120
aaagatattt atattgacct gctgacagac agtggcacta acgcaatgag cgacaagcag      180
tgggccggaa tgatgatggg tgatgaagcg tacgcgggca gcgaaaactt ctatcatctg      240
gaaagaaccg tgcaggaact gtttggcttt aaacatatfg ttccgactca ccaggggagc      300
ggcgcagaaa acctgttatc gcagctggct attaaacctg ggcaatatgt tgccgggaat      360
atgtatttca ctaccaccgg ttatcaccag gaaaaaatg gtgcggtgtt tgcgatatc      420
gttcgtgacg aagcgcacga tgccggtctg aatattgcgt ttaaagggtga tatcgatctt      480
aaaaaattac aaaagctgat tgatgaaaaa ggcgagaga atattgcgta tatctgcctg      540
gcggtgacgg ttaacctcgc gggcgggcaa ccggtctcga tggctaacat gcgtgcggtg      600
cgtgaactga cagaagcgca tggcattaaa gtgttctacg acgctaccgg ctgcgtagaa      660
aacgcctact ttatcaaaga gcaagagcag ggctttgaga acaagagcat cgccgagatc      720
gtgcatgaga tgttcagcta cgccgacggg tgtaccatga gtggtaaaaa agactgtctg      780
gtgaacatcg gcggcttcct gtgcatgaac gatgacgaaa tgttctcttc tgccaaagag      840
ttagtcgtgg tctacgaagg gatgccatct tacggcggcc tggccggacg tgatatggaa      900
gcgatggcga ttggcctgcg tgaagccatg cagtacgaat atattgagca ccgcgtgaag      960
caggttcgtc acctgggcga taagctgaaa gccgctggcg taccgattgt tgaaccggta     1020
ggcggtcacg cggatttcct cgatgcgcgt cgcttctgcg agcatctgac gcaagatgag     1080
ttcccgccac aaagtctggc tgccagcatc tatgtggaaa ccggcgtgcg cagtatggag     1140
cgcggaatta tctctgcggg ccgtaataac gtgaccggtg aacaccacag accgaaactg     1200
gaaaccgtgc gtctgactat tccacgtcgt gtttatacct acgcacatat ggatgtttgtg     1260
gctgacggta ttattaaact ttaccagcac aaagaagata ttcgcggggt gaagtttatt     1320
tacgagccga agcagttgcg tttctttact gcacgctttg attacatcta a              1371
```

<210> 43

<211> 35

<212> DNA

<213> Artificial sequence

<220>

<223> primer

<400> 43

ggtattgagg gtcgcatgga aaactttaaa catct

35

<210> 44
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 44
agaggagagt tagagcotta aacttcttta agttttg

37

<210> 45
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 45
ggtattgagg gtogcatgaa ttatccggca gaacc

35

<210> 46
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 46
agaggagagt tagagcotta gatgtaatca aagcgtg

37

<210> 47
<211> 31
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 47
ccagggcacc ggcgcagagc aaatctatat t

31

<210> 48
<211> 30
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 48
tgcgcgggtg ccctgggtgag tcggaatggt

30

<210> 49
<211> 32
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 49
tcctgcacgc ggcaaagggt tctgcactcg gt

32

<210> 50
<211> 30
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 50
ctttgccgcg tgcaggaagg cttcccgaca

30

<210> 51
<211> 29
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 51
aggggaccgg cgcagaaaac ctgttatcg

29

<210> 52
<211> 32
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 52
tctgcgcgg tccctggtg agtcggaaca at

32

<210> 53
<211> 29
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 53
gttagtcgc gtctacgaag ggatgccat

29

<210> 54

<211> 29
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 54
gtagacgcgg actaactctt tggcagaag

29

<210> 55
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 55
ggtattgagg gtcgcatgta cgaactggga gttgt

35

<210> 56
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 56
agaggagagt tagagcctta gtcaatatat ttcaggc

37

<210> 57
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 57
ggtattgagg gtcgcatgtc cggcatcggt gtcca

35

<210> 58
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 58
agaggagagt tagagcctca gacatatattc agtccca

37

<210> 59
<211> 35

<212> DNA
 <213> Artificial sequence

 <220>
 <223> primer

 <400> 59
 ggtattgagg gtcgcatgcg actgaacaac ctcgg 35

 <210> 60
 <211> 37
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> primer

 <400> 60
 agaggagagt tagagcctca gttctccacg tattcca 37

 <210> 61
 <211> 35
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> primer

 <400> 61
 ggtattgagg gtcgcatgag cgtgggtcac cggaa 35

 <210> 62
 <211> 37
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> primer

 <400> 62
 agaggagagt tagagcctca atcgatatat ttcagtc 37

 <210> 63
 <211> 35
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> primer

 <400> 63
 ggtattgagg gtcgcatgag cctgggtaat atgaa 35

 <210> 64
 <211> 37
 <212> DNA

<213> Artificial sequence

<220>

<223> primer

<400> 64

agaggagagt tagagcctta tgactttaac gcgttga

37

<210> 65

<211> 684

<212> DNA

<213> C. testosteroni

<400> 65

atgtacgaac tgggagttgt ctaccgcaat atccagcgcg ccgaccgcgc tgctgctgac 60

ggcctggccg ccttgggctc cgccaccgtg caccaggcca tgggcccgcgt cgggtctgctc 120

aagccctata tgcgccccat ctatgccggc aagcaggtct cgggcaccgc cgtcacgggtg 180

ctgctgcagc ccggcgacaa ctggatgatg catgtggctg ccgagcagat tcagcccggc 240

gacatcgtgg tcgcagccgt caccgcagag tgcaccgacg gctacttcgg cgatctgctg 300

gccaccagct tccaggcgcg cggcgccagt gcgctgatca tcgatgccgg cgtgcgcgac 360

gtgaagacgc tgcaggagat ggactttccg gtctggagca aggccatctc ttccaagggc 420

acgatcaagg ccaccctggg ctccgtcaac atccccatcg tctgcgccgg catgctggtc 480

acgcccgggtg acgtgatcgt ggccgacgac gacggcgtgg tctgcgtgcc cgccgcgcgt 540

gccgtggaag tgctggccgc cgcccagaag cgtgaaagct tcgaaggcga aaagcgcgcc 600

aagctggcct cgggcacccct cggcctggat atgtacaaga tgcgcgagcc cctggaaaag 660

gccggcctga aatatattga ctaa 684

<210> 66

<211> 227

<212> PRT

<213> C. testosteroni

<400> 66

Met Tyr Glu Leu Gly Val Val Tyr Arg Asn Ile Gln Arg Ala Asp Arg
1 5 10 15

Ala Ala Ala Asp Gly Leu Ala Ala Leu Gly Ser Ala Thr Val His Glu
20 25 30

Ala Met Gly Arg Val Gly Leu Leu Lys Pro Tyr Met Arg Pro Ile Tyr
35 40 45

Ala Gly Lys Gln Val Ser Gly Thr Ala Val Thr Val Leu Leu Gln Pro
50 55 60

Gly Asp Asn Trp Met Met His Val Ala Ala Glu Gln Ile Gln Pro Gly
65 70 75 80

Asp Ile Val Val Ala Ala Val Thr Ala Glu Cys Thr Asp Gly Tyr Phe
85 90 95

Gly Asp Leu Leu Ala Thr Ser Phe Gln Ala Arg Gly Ala Arg Ala Leu
100 105 110

Ile Ile Asp Ala Gly Val Arg Asp Val Lys Thr Leu Gln Glu Met Asp
115 120 125

Phe Pro Val Trp Ser Lys Ala Ile Ser Ser Lys Gly Thr Ile Lys Ala
130 135 140

Thr Leu Gly Ser Val Asn Ile Pro Ile Val Cys Ala Gly Met Leu Val
145 150 155 160

Thr Pro Gly Asp Val Ile Val Ala Asp Asp Asp Gly Val Val Cys Val
165 170 175

Pro Ala Ala Arg Ala Val Glu Val Leu Ala Ala Ala Gln Lys Arg Glu
180 185 190

Ser Phe Glu Gly Glu Lys Arg Ala Lys Leu Ala Ser Gly Ile Leu Gly
195 200 205

Leu Asp Met Tyr Lys Met Arg Glu Pro Leu Glu Lys Ala Gly Leu Lys
210 215 220

Tyr Ile Asp
225

<210> 67
<211> 42
<212> DNA
<213> Artificial sequence

<220>
<223> primer

<400> 67
actcggatcc gaaggagata tacatatgta cgaactggga ct

42

<210> 68
<211> 33
<212> DNA

<213> Artificial sequence

<220>

<223> primer

<400> 68

cggctgtcga cggtagtca atatatttca ggc

33

<210> 69

<211> 31

<212> DNA

<213> Artificial sequence

<220>

<223> primer

<400> 69

cgcggatcca taatggttga gaacattacc g

31

<210> 70

<211> 30

<212> DNA

<213> Artificial sequence

<220>

<223> primer

<400> 70

acgcgtcgac ttacagcact gccacaatcg

30

<210> 71

<211> 32

<212> DNA

<213> Artificial sequence

<220>

<223> primer

<400> 71

ccggaattca taatggtcga actgggagtt gt

32

<210> 72

<211> 33

<212> DNA

<213> Artificial sequence

<220>

<223> primer

<400> 72

gaatgcggcc gcttagtcaa tatatttcag gcc

33

<210> 73

<211> 15

<212> DNA

<213> Artificial sequence

<220>

<223> primer

<400> 73

ggtattgagg gtcgc

15

<210> 74

<211> 17

<212> DNA

<213> Artificial sequence

<220>

<223> primer

<400> 74

agaggagagt tagagcc

17