US 20060078909A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2006/0078909 A1**

Srinivasan et al. (43) **Pub. Date: Apr. 13, 2006**

(54) **NOVEL SULFURYLASE-LUCIFERASE FUSION PROTEINS AND THERMOSTABLE SULFURYLASE**

(76) Inventors: **Maithreyan Srinivasan**, Branford, CT (US); **Michael Reifler**, Hamden, CT (US)

Correspondence Address:
MINTZ, LEVIN, COHN, et al
24th Floor
666 Third Avenue
New York, NY 10017 (US)

(21) Appl. No.: **11/147,763**

(22) Filed: **Jun. 7, 2005**

Related U.S. Application Data

(63) Continuation of application No. 10/154,515, filed on May 23, 2002, now Pat. No. 6,902,921, which is a continuation-in-part of application No. 10/122,706, filed on Apr. 11, 2002, now Pat. No. 6,956,114.

(60) Provisional application No. 60/335,949, filed on Oct. 30, 2001. Provisional application No. 60/349,076, filed on Jan. 16, 2002.

Publication Classification

(51) **Int. Cl.**
C12Q 1/68 (2006.01)
C12N 9/06 (2006.01)

(52) **U.S. Cl.** **435/6**; 435/191

(57) **ABSTRACT**

The present invention relates to the field of DNA recombinant technology. More specifically, this invention relates to fusion proteins comprising an ATP generating polypeptide joined to a polypeptide that converts ATP into a detectable entity. Accordingly, this invention focuses on sulfurylase-luciferase fusion proteins. This invention also relates to pharmaceutical compositions containing the fusion proteins and methods for using them.

FIGURE 1

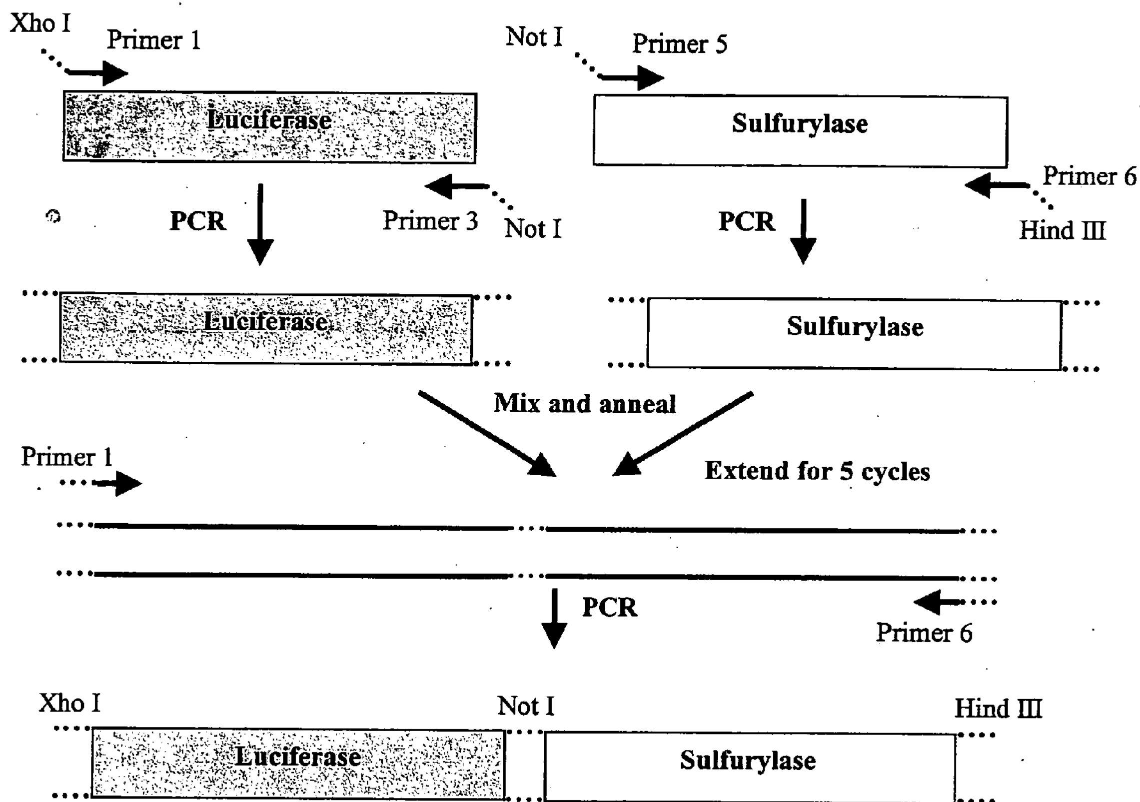
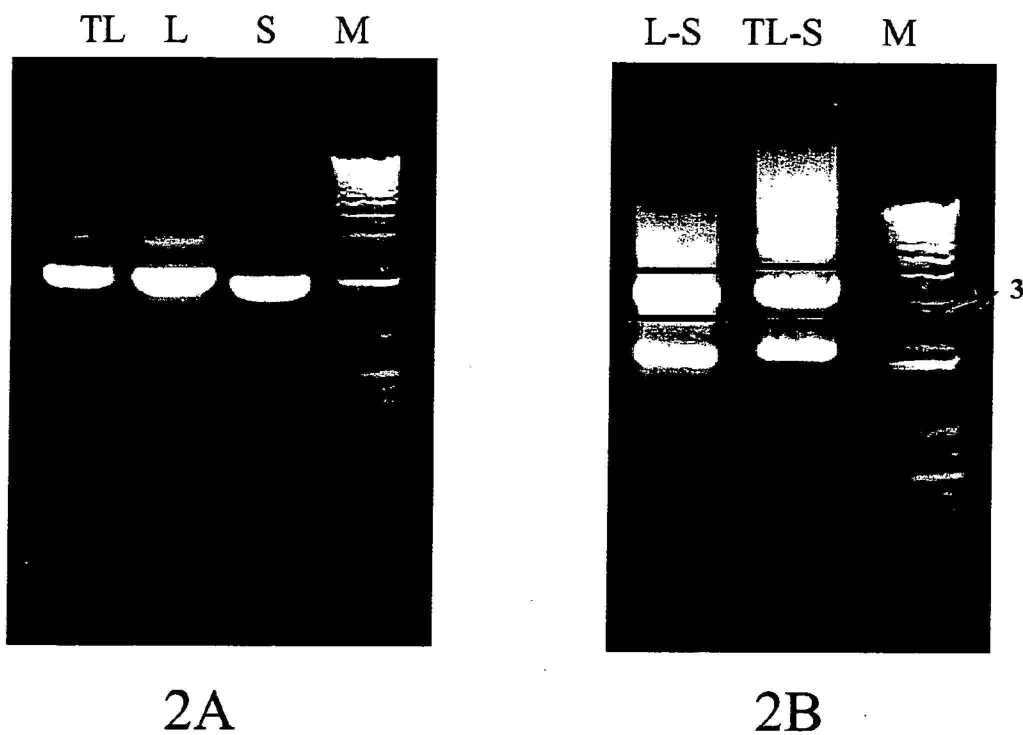
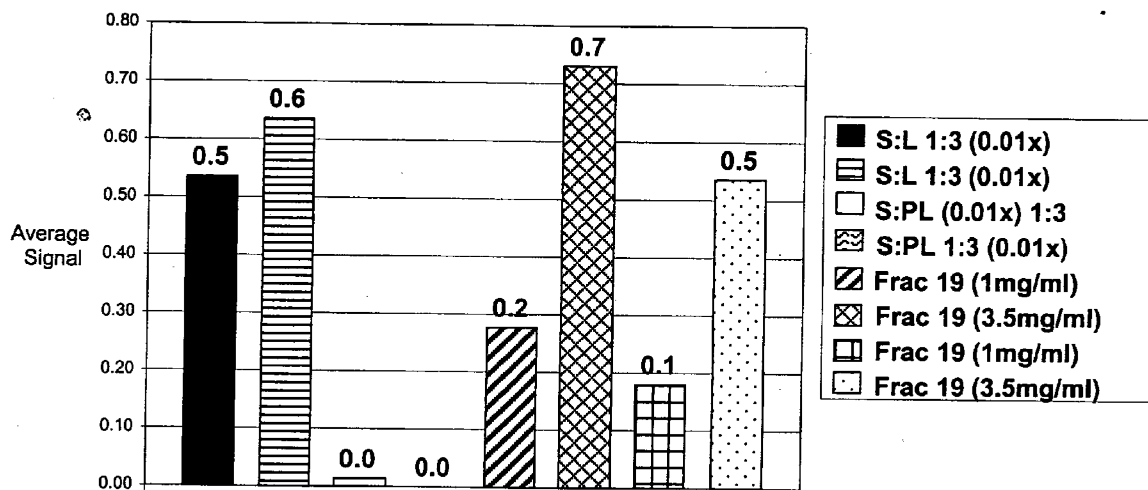


Figure 2



2A 2B

FIGURE 3



NOVEL SULFURYLASE-LUCIFERASE FUSION PROTEINS AND THERMOSTABLE SULFURYLASE

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application is a Continuation of U.S. patent application Ser. No. 10/154,515, filed May 23, 2002, which is a continuation in part of U.S. patent application Ser. No. 10/122,706 filed Apr. 11, 2002 which claims the benefit of priority to U.S. Patent Application 60/335,949 filed Oct. 30, 2001 and U.S. Patent Application 60/349,076 filed Jan. 16, 2002. All patents, patent applications and references cited in this specification is hereby incorporated by reference.

FIELD OF THE INVENTION

[0002] The invention relates generally to fusion proteins that are useful as reporter proteins, in particular to fusion proteins of ATP sulfurylase and luciferase which are utilized to achieve an efficient conversion of pyrophosphate (PPi) to light. This invention also relates to a novel thermostable sulfurylase which can be used in the detection of inorganic pyrophosphate, particularly in the sequencing of nucleic acid.

BACKGROUND OF THE INVENTION

[0003] ATP sulfurylase has been identified as being involved in sulfur metabolism. It catalyzes the initial reaction in the metabolism of inorganic sulfate (SO_4^{-2}); see e.g., Robbins and Lipmann, 1958. *J. Biol. Chem.* 233: 686-690; Hawes and Nicholas, 1973. *Biochem. J.* 133: 541-550). In this reaction SO_4^{-2} is activated to adenosine 5'-phosphosulfate (APS). ATP sulfurylase is also commonly used in pyrophosphate sequencing methods. In order to convert pyrophosphate (PPi) generated from the addition of dNMP to a growing DNA chain to light, PPi must first be converted to ATP by ATP sulfurylase.

[0004] ATP produced by an ATP sulfurylase can also be hydrolyzed using enzymatic reactions to generate light. Light-emitting chemical reactions (i.e., chemiluminescence) and biological reactions (i.e., bioluminescence) are widely used in analytical biochemistry for sensitive measurements of various metabolites. In bioluminescent reactions, the chemical reaction that leads to the emission of light is enzyme-catalyzed. For example, the luciferin-luciferase system allows for specific assay of ATP. Thus, both ATP generating enzymes, such as ATP sulfurylase, and light emitting enzymes, such as luciferase, could be useful in a number of different assays for the detection and/or concentration of specific substances in fluids and gases. Since high physical and chemical stability is sometimes required for enzymes involved in sequencing reactions, a thermostable enzyme is desirable.

[0005] Because the product of the sulfurylase reaction is consumed by luciferase, proximity between these two enzymes by covalently linking the two enzymes in the form of a fusion protein would provide for a more efficient use of the substrate. Substrate channeling is a phenomenon in which substrates are efficiently delivered from enzyme to enzyme without equilibration with other pools of the same substrates. In effect, this creates local pools of metabolites at high concentrations relative to those found in other areas of the cell. Therefore, a fusion of an ATP generating polypeptide and an ATP converting peptide could benefit from the phenomenon of substrate channeling and would reduce production costs and increase the number of enzymatic reactions that occur during a given time period.

[0006] All patents and publications cited throughout the specification are hereby incorporated by reference into this specification in their entirety in order to more fully describe the state of the art to which this invention pertains.

SUMMARY OF THE INVENTION

[0007] The invention provides a fusion protein comprising an ATP generating polypeptide bound to a polypeptide which converts ATP into an entity which is detectable. In one aspect, the invention provides a fusion protein comprising a sulfurylase polypeptide bound to a luciferase polypeptide. This invention provides a nucleic acid that comprises an open reading frame that encodes a novel thermostable sulfurylase polypeptide. In a further aspect, the invention provides for a fusion protein comprising a thermostable sulfurylase joined to at least one affinity tag.

[0008] In another aspect, the invention provides a recombinant polynucleotide that comprises a coding sequence for a fusion protein having a sulfurylase poylpeptide sequence joined to a luciferase polypeptide sequence. In a further aspect, the invention provides an expression vector for expressing a fusion protein. The expression vector comprises a coding sequence for a fusion protein having: (i) a regulatory sequence, (ii) a first polypeptide sequence of an ATP generating polypeptide and (iii) a second polypeptide sequence that converts ATP to an entity which is detectable. In an additional embodiment, the fusion protein comprises a sulfurylase polypeptide and a luciferase polypeptide. In another aspect, the invention provides a transformed host cell which comprises the expression vector. In an additional aspect, the invention provides a fusion protein bound to a mobile support. The invention also includes a kit comprising a sulfurylase-luciferase fusion protein expression vector.

[0009] The invention also includes a method for determining the nucleic acid sequence in a template nucleic acid polymer, comprising: (a) introducing the template nucleic acid polymer into a polymerization environment in which the nucleic acid polymer will act as a template polymer for the synthesis of a complementary nucleic acid polymer when nucleotides are added; (b) successively providing to the polymerization environment a series of feedstocks, each feedstock comprising a nucleotide selected from among the nucleotides from which the complementary nucleic acid polymer will be formed, such that if the nucleotide in the feedstock is complementary to the next nucleotide in the template polymer to be sequenced said nucleotide will be incorporated into the complementary polymer and inorganic pyrophosphate will be released; (c) separately recovering each of the feedstocks from the polymerization environment; and (d) measuring the amount of PPi with an ATP generating polypeptide-ATP converting polypeptide fusion protein in each of the recovered feedstocks to determine the identity of each nucleotide in the complementary polymer and thus the sequence of the template polymer. In one embodiment, the amount of inorganic pyrophosphate is measured by the steps of: (a) adding adenosine-5'-phosphosulfate to the feedstock; (b) combining the recovered feedstock containing adenosine-5'-phosphosulfate with an ATP

generating polypeptide-ATP converting polypeptide fusion protein such that any inorganic pyrophosphate in the recovered feedstock and the adenosine-5'-phosphosulfate will react to the form ATP and sulfate; (c) combining the ATP and sulfate-containing feedstock with luciferin in the presence of oxygen such that the ATP is consumed to produced AMP, inorganic pyrophosphate, carbon dioxide and light; and (d) measuring the amount of light produced.

[0010] In another aspect, the invention includes a method wherein each feedstock comprises adenosine-5'-phosphosulfate and luciferin in addition to the selected nucleotide base, and the amount of inorganic pyrophosphate is determined by reacting the inorganic pyrophosphate feedstock with an ATP generating polypeptide-ATP converting polypeptide fusion protein thereby producing light in an amount proportional to the amount of inorganic pyrophosphate, and measuring the amount of light produced.

[0011] In another aspect, the invention provides a method for sequencing a nucleic acid, the method comprising; (a) providing one or more nucleic acid anchor primers; (b) providing a plurality of single-stranded circular nucleic acid templates disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200 μm; (c) annealing an effective amount of the nucleic acid anchor primer to at least one of the single-stranded circular templates to yield a primed anchor primer-circular template complex; (d) combining the primed anchor primer-circular template complex with a polymerase to form an extended anchor primer covalently linked to multiple copies of a nucleic acid complementary to the circular nucleic acid template; (e) annealing an effective amount of a sequencing primer to one or more copies of said covalently linked complementary nucleic acid; (f) extending the sequencing primer with a polymerase and a predetermined nucleotide triphosphate to yield a sequencing product and, if the predetermined nucleotide triphosphate is incorporated onto the 3' end of said sequencing primer, a sequencing reaction byproduct; and (g) identifying the sequencing reaction byproduct with the use of a ATP generating polypeptide-ATP converting polypeptide fusion protein, thereby determining the sequence of the nucleic acid.

[0012] In one aspect, the invention provides a method for sequencing a nucleic acid, the method comprising: (a) providing at least one nucleic acid anchor primer; (b) providing a plurality of single-stranded circular nucleic acid templates in an array having at least 400,000 discrete reaction sites; (c) annealing a first amount of the nucleic acid anchor primer to at least one of the single-stranded circular templates to yield a primed anchor primer-circular template complex; (d) combining the primed anchor primer-circular template complex with a polymerase to form an extended anchor primer covalently linked to multiple copies of a nucleic acid complementary to the circular nucleic acid template; (e) annealing a second amount of a sequencing primer to one or more copies of the covalently linked complementary nucleic acid; (f) extending the sequencing primer with a polymerase and a predetermined nucleotide triphosphate to yield a sequencing product and, when the predetermined nucleotide triphosphate is incorporated onto the 3' end of the sequencing primer, to yield a sequencing reaction byproduct; and (g) identifying the sequencing reaction byproduct with the use of a ATP generating polypeptide-ATP converting polypeptide fusion protein, thereby determining the sequence of the nucleic acid at each reaction site that contains a nucleic acid template.

[0013] In another aspect, the invention includes a method of determining the base sequence of a plurality of nucleotides on an array, the method comprising the steps of: (a) providing a plurality of sample DNAs, each disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200 μm, (b) adding an activated nucleotide 5'-triphosphate precursor of one known nitrogenous base to a reaction mixture in each reaction chamber, each reaction mixture comprising a template-directed nucleotide polymerase and a single-stranded polynucleotide template hybridized to a complementary oligonucleotide primer strand at least one nucleotide residue shorter than the templates to form at least one unpaired nucleotide residue in each template at the 3'-end of the primer strand, under reaction conditions which allow incorporation of the activated nucleoside 5'-triphosphate precursor onto the 3'-end of the primer strands, provided the nitrogenous base of the activated nucleoside 5'-triphosphate precursor is complementary to the nitrogenous base of the unpaired nucleotide residue of the templates; (c) determining whether or not the nucleoside 5'-triphosphate precursor was incorporated into the primer strands through detection of a sequencing byproduct with a ATP generating polypeptide-ATP converting polypeptide fusion protein, thus indicating that the unpaired nucleotide residue of the template has a nitrogenous base composition that is complementary to that of the incorporated nucleoside 5'-triphosphate precursor; and (d) sequentially repeating steps (b) and (c), wherein each sequential repetition adds and, detects the incorporation of one type of activated nucleoside 5'-triphosphate precursor of known nitrogenous base composition; and

(e) determining the base sequence of the unpaired nucleotide residues of the template in each reaction chamber from the sequence of incorporation of said nucleoside precursors.

[0014] In one aspect, the invention includes a method for determining the nucleic acid sequence in a template nucleic acid polymer, comprising: (a) introducing a plurality of template nucleic acid polymers into a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200 μm, each reaction chamber having a polymerization environment in which the nucleic acid polymer will act as a template polymer for the synthesis of a complementary nucleic acid polymer when nucleotides are added; (b) successively providing to the polymerization environment a series of feedstocks, each feedstock comprising a nucleotide selected from among the nucleotides from which the complementary nucleic acid polymer will be formed, such that if the nucleotide in the feedstock is complementary to the next nucleotide in the template polymer to be sequenced said nucleotide will be incorporated into the complementary polymer and inorganic pyrophosphate will be released; (c) detecting the formation of inorganic pyrophosphate with an ATP generating polypeptide-ATP converting polypeptide fusion protein to determine the identify of each nucleotide in the complementary polymer and thus the sequence of the template polymer.

[0015] In one aspect, the invention provides a method of identifying the base in a target position in a DNA sequence of sample DNA including the steps comprising: (a) disposing sample DNA within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200 μm, said DNA being rendered single stranded either before or after being disposed in the reaction chambers, (b) providing an extension primer which hybridizes to said immobilized single-stranded DNA at a position immediately adjacent to said target position; (c) subjecting said immobilized single-stranded DNA to a polymerase reaction in the presence of a predetermined nucleotide triphosphate, wherein if the predetermined nucleotide triphosphate is incorporated onto the 3' end of said sequencing primer then a sequencing reaction byproduct is formed; and

(d) identifying the sequencing reaction byproduct with a ATP generating polypeptide-ATP converting polypeptide fusion protein, thereby determining the nucleotide complementary to the base at said target position.

[0016] The invention also includes a method of identifying a base at a target position in a sample DNA sequence comprising: (a) providing sample DNA disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200 μm, said DNA being rendered single stranded either before or after being disposed in the reaction chambers; (b) providing an extension primer which hybridizes to the sample DNA immediately adjacent to the target position; (c) subjecting the sample DNA sequence and the extension primer to a polymerase reaction in the presence of a nucleotide triphosphate whereby the nucleotide triphosphate will only become incorporated and release pyrophosphate (PPi) if it is complementary to the base in the target position, said nucleotide triphosphate being added either to separate aliquots of sample-primer mixture or successively to the same sample-primer mixture; (d) detecting the release of PPi with an ATP generating polypeptide-ATP converting polypeptide fusion protein to indicate which nucleotide is incorporated.

[0017] In one aspect, the invention provides a method of identifying a base at a target position in a single-stranded sample DNA sequence, the method comprising: (a) providing an extension primer which hybridizes to sample DNA immediately adjacent to the target position, said sample DNA disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200 um, said DNA being rendered single stranded either before or after being disposed in the reaction chambers; (b) subjecting the sample DNA and extension primer to a polymerase reaction in the presence of a predetermined deoxynucleotide or dideoxynucleotide whereby the deoxynucleotide or dideoxynucleotide will only become incorporated and release pyrophosphate (PPi) if it is complementary to the base in the target position, said predetermined deoxynucleotides or dideoxynucleotides being added either to separate aliquots of sample-primer mixture or successively to the same sample-primer mixture, (c) detecting any release of PPi with an ATP generating polypeptide-ATP converting polypeptide fusion protein to indicate which deoxynucleotide or dideoxynucleotide is incorporated;characterized in that, the PPi-detection enzyme(s) are included in the polymerase reaction step and in that in place of deoxy- or dideoxy adenosine triphosphate (ATP) a dATP or ddATP analogue is used which is capable of acting as a substrate for a polymerase but incapable of acting as a substrate for a said PPi-detection enzyme.

[0018] In another aspect, the invention includes a method of determining the base sequence of a plurality of nucleotides on an array, the method comprising: (a) providing a plurality of sample DNAs, each disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200 μm, (b) converting PPi into light with an ATP generating polypeptide-ATP converting polypeptide fusion protein; (c) detecting the light level emitted from a plurality of reaction sites on respective portions of an optically sensitive device; (d) converting the light impinging upon each of said portions of said optically sensitive device into an electrical signal which is distinguishable from the signals from all of said other regions; (e) determining a light intensity for each of said discrete regions from the corresponding electrical signal; (f) recording the variations of said electrical signals with time.

[0019] In one aspect, the invention provides a method for sequencing a nucleic acid, the method comprising:(a) providing one or more nucleic acid anchor primers; (b) providing a plurality of single-stranded circular nucleic acid templates disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200 μm;(c) converting PPi into a detectable entity with the use of an ATP generating polypeptide-ATP converting polypeptide fusion protein; (d) detecting the light level emitted from a plurality of reaction sites on respective portions of an optically sensitive device; (e) converting the light impinging upon each of said portions of said optically sensitive device into an electrical signal which is distinguishable from the signals from all of said other regions; (f) determining a light intensity for each of said discrete regions from the corresponding electrical signal; (g) recording the variations of said electrical signals with time.

[0020] In another aspect, the invention includes a method for sequencing a nucleic acid, the method comprising: (a) providing at least one nucleic acid anchor primer; (b) providing a plurality of single-stranded circular nucleic acid templates in an array having at least 400,000 discrete reaction sites; (c) converting PPi into a detectable entity with an ATP generating polypeptide-ATP converting polypeptide fusion protein; (d) detecting the light level emitted from a plurality of reaction sites on respective portions of an optically sensitive device; (e) converting the light impinging upon each of said portions of said optically sensitive device into an electrical signal which is distinguishable from the signals from all of said other regions; (f) determining a light intensity for each of said discrete regions from the corresponding electrical signal; (g) recording the variations of said electrical signals with time.

[0021] In another aspect, the invention includes an isolated polypeptide comprising an amino acid sequence selected from the group consisting of: (a) a mature form of an amino acid sequence of SEQ ID NO: 2; (b) a variant of a mature form of an amino acid sequence of SEQ ID NO: 2;

an amino acid sequence of SEQ ID NO: 2; (c) a variant of an amino acid sequence of SEQ ID NO: 2, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 5% of amino acid residues from said amino acid sequence; (d) and at least one conservative amino acid substitution to the amino acid sequences in (a), (b), (c) or (d). The invention also includes an antibody that binds immunospecifically to the polypeptide of (a), (b), (c) or (d).

[0022] In another aspect, the invention includes an isolated nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of: (a) a mature form of an amino acid sequence of SEQ ID NO: 2; (b) a variant of a mature form of an amino acid sequence of SEQ ID NO: 2, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 5% of the amino acid residues from the amino acid sequence of said mature form; (c) an amino acid sequence of SEQ ID NO: 2; (d) a variant of an amino acid sequence of SEQ ID NO: 2, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 15% of amino acid residues from said amino acid sequence; a nucleic acid fragment encoding at least a portion of a polypeptide comprising an amino acid sequence of SEQ ID NO: 2, or a variant of said polypeptide, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 5% of amino acid residues from said amino acid sequence; (e) and a nucleic acid molecule comprising the complement of (a), (b), (c), (d) or (e).

[0023] In a further aspect, the invention provides a nucleic acid molecule wherein the nucleic acid molecule comprises nucleotide sequence selected from the group consisting of: (a) a first nucleotide sequence comprising a coding sequence differing by one or more nucleotide sequences from a coding sequence encoding said amino acid sequence, provided that no more than 20% of the nucleotides in the coding sequence in said first nucleotide sequence differ from said coding sequence; an isolated second polynucleotide that is a complement of the first polynucleotide; (b) and a nucleic acid fragment of (a) or (b). The invention also includes a vector comprising the nucleic acid molecule of (a) or (b). In another aspect, the invention includes a cell comprising the vector.

[0024] In a further aspect, the invention includes a method for determining the nucleic acid sequence in a template nucleic acid polymer, comprising: (a) introducing the template nucleic acid polymer into a polymerization environment in which the nucleic acid polymer will act as a template polymer for the synthesis of a complementary nucleic acid polymer when nucleotides are added; (b) successively providing to the polymerization environment a series of feedstocks, each feedstock comprising a nucleotide selected from among the nucleotides from which the complementary nucleic acid polymer will be formed, such that if the nucleotide in the feedstock is complementary to the next nucleotide in the template polymer to be sequenced said nucleotide will be incorporated into the complementary polymer and inorganic pyrophosphate will be released; (c) separately recovering each of the feedstocks from the polymerization environment; and (d) measuring the amount of PPi with an ATP sulfurylase and a luciferase in each of the recovered feedstocks to determine the identity of each nucleotide in the complementary polymer and thus the sequence of the template polymer.

[0025] In another aspect, the invention provides a method for sequencing a nucleic acid, the method comprising: (a) providing one or more nucleic acid anchor primers; (b) providing a plurality of single-stranded circular nucleic acid templates disposed within a plurality of cavities in an array on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200 μm and at least 400,000 discrete sites; (c) annealing an effective amount of the nucleic acid anchor primer to at least one of the single-stranded circular templates to yield a primed anchor primer-circular template complex; (d) combining the primed anchor primer-circular template complex with a polymerase to form an extended anchor primer covalently linked to multiple copies of a nucleic acid complementary to the circular nucleic acid template; (e) annealing an effective amount of a sequencing primer to one or more copies of said covalently linked complementary nucleic acid; (f) extending the sequencing primer with a polymerase and a predetermined nucleotide triphosphate to yield a sequencing product and, if the predetermined nucleotide triphosphate is incorporated onto the 3' end of said sequencing primer, a sequencing reaction byproduct; and (g) identifying the sequencing reaction byproduct with the use of an ATP sulfurylase and a luciferase, thereby determining the sequence of the nucleic acid.

[0026] In another aspect, the invention provides a method for sequencing a nucleic acid, the method comprising: (a) providing at least one nucleic acid anchor primer; (b) providing a plurality of single-stranded circular nucleic acid templates in an array having at least 400,000 discrete reaction sites; (c) annealing a first amount of the nucleic acid anchor primer to at least one of the single-stranded circular templates to yield a primed anchor primer-circular template complex; (d) combining the primed anchor primer-circular template complex with a polymerase to form an extended anchor primer covalently linked to multiple copies of a nucleic acid complementary to the circular nucleic acid template; (e) annealing a second amount of a sequencing primer to one or more copies of the covalently linked complementary nucleic acid; (f) extending the sequencing primer with a polymerase and a predetermined nucleotide triphosphate to yield a sequencing product and, when the predetermined nucleotide triphosphate is incorporated onto the 3' end of the sequencing primer, to yield a sequencing reaction byproduct; and (g) identifying the sequencing reaction byproduct with the use of a thermostable sulfurylase and a luciferase, thereby determining the sequence of the nucleic acid at each reaction site that contains a nucleic acid template.

[0027] In a further aspect, the invention includes a method of determining the base sequence of a plurality of nucleotides on an array, the method comprising: (a) providing a plurality of sample DNAs, each disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 5 to 200 μm, (b)

adding an activated nucleotide 5'-triphosphate precursor of one known nitrogenous base to a reaction mixture in each reaction chamber, each reaction mixture comprising a template-directed nucleotide polymerase and a single-stranded polynucleotide template hybridized to a complementary oligonucleotide primer strand at least one nucleotide residue shorter than the templates to form at least one unpaired nucleotide residue in each template at the 3'-end of the primer strand, under reaction conditions which allow incorporation of the activated nucleoside 5'-triphosphate precursor onto the 3'-end of the primer strands, provided the nitrogenous base of the activated nucleoside 5'-triphosphate precursor is complementary to the nitrogenous base of the unpaired nucleotide residue of the templates; (c) detecting whether or not the nucleoside 5'-triphosphate precursor was incorporated into the primer strands through detection of a sequencing byproduct with a thermostable sulfurylase and luciferase, thus indicating that the unpaired nucleotide residue of the template has a nitrogenous base composition that is complementary to that of the incorporated nucleoside 5'-triphosphate precursor; and (d) sequentially repeating steps (b) and (c), wherein each sequential repetition adds and, detects the incorporation of one type of activated nucleoside 5'-triphosphate precursor of known nitrogenous base composition; and (e) determining the base sequence of the unpaired nucleotide residues of the template in each reaction chamber from the sequence of incorporation of said nucleoside precursors.

BRIEF DESCRIPTION OF THE DRAWINGS

[0028] **FIG. 1** is one embodiment for a cloning strategy for obtaining the luciferase-sulfurylase sequence.

[0029] **FIG. 2A and 2B** show the preparative agarose gel of luciferase and sulfurylase as well as sulfurylase-luciferase fusion genes.

[0030] **FIG. 3** shows the results of experiments to determine the activity of the luciferase-sulfurylase fusion protein on NTA-agarose and MPG-SA solid supports.

DETAILED DESCRIPTION OF THE INVENTION

[0031] This invention provides a fusion protein containing an ATP generating polypeptide bound to a polypeptide which converts ATP into an entity which is detectable. As used herein, the term "fusion protein" refers to a chimeric protein containing an exogenous protein fragment joined to another exogenous protein fragment. The fusion protein could include an affinity tag to allow attachment of the protein to a solid support or to allow for purification of the recombinant fusion protein from the host cell or culture supernatant, or both.

[0032] In a preferred embodiment, the ATP generating polypeptide and ATP converting polypeptide are from a eukaryote or a prokaryote. The eukaryote could be an animal, plant, fungus or yeast. In some embodiments, the animal is a mammal, rodent, insect, worm, mollusk, reptile, bird and amphibian. Plant sources of the polypeptides include but are not limited to *Arabidopsis thaliana, Brassica napus, Allium sativum, Amaranthus caudatus, Hevea brasiliensis, Hordeum vulgare, Lycopersicon esculentum, Nicotiana tabacum, Oryza sativum, Pisum sativum, Populus trichocarpa, Solanum tuberosum, Secale cereale, Sambucus nigra, Ulmus americana* or *Triticum aestivum*. Examples of fungi include but are not limited to *Penicillum chrysogenum, Stachybotrys chartarum, Aspergillus fumigatus, Podospora anserina and Trichoderma reesei*. Examples of sources of yeast include but are not limited to *Saccharomyces cerevisiae, Candida tropicalis, Candida lypolitica, Candida utilis, Kluyveromyces lactis, Schizosaccharomyces pombe, Yarrowia lipolytica, Candida* spp., *Pichia* spp. and *Hansenula* spp.

[0033] The prokaryote source could be bacteria or archaea. In some embodiments, the bacteria is *E. coli, B. subtilis, Streptococcus gordonii*, flavobacteria or green sulfur bacteria. In other embodiments, the archaea is *Sulfolobus, Thermococcus, Methanobacterium, Halococcus, Halobacterium or Methanococcus jannaschii.*

[0034] The ATP generating polypeptide can be a ATP sulfurylase, hydrolase or an ATP synthase. In a preferred embodiment, the ATP generating polypeptide is ATP sulfurylase. In one embodiment, the ATP sulfurylase is a thermostable sulfurylase cloned from *Bacillus stearothermophilus* (Bst) and comprising the nucleotide sequence of SEQ ID NO:1. This putative gene was cloned using genomic DNA acquired from ATCC (Cat. No. 12980D). The gene is shown to code for a functional ATP sulfurylase that can be expressed as a fusion protein with an affinity tag. The disclosed Bst sulfurylase nucleic acid (SEQ ID NO:1) includes the 1247 nucleotide sequence. An open reading frame (ORF) for the mature protein was identified beginning with an ATG codon at nucleotides 1-3 and ending with a TAA codon at nucleotides 1159-1161. The start and stop codons of the open reading frame are highlighted in bold type. The putative untranslated regions are underlined and found upstream of the initiation codon and downstream from the termination codon.

```
Bst Thermostable Sulfurylase Nucleotide Sequence
(SEQ ID NO: 1)
GTTATGAACATGAGTTTGAGCATTCCGCATGGCGGCACATTGATCAACCGTTGGAATCGG   60

GATTACCCAATGGATGAAGCAACGAAAACGATGGAGGTGTCCAAAGCCGAAGTAAGCGAC  120

CTTGAGCTGATCGGCACAGGCGCCTACAGCCCGCTCACCGGGTTTTTAAGGAAAGCCGAT  180

TACGATGCGGTCGTAGAAACGATGCGCCTCGCTGATGGCACTGTCTGGAGCATTCCGATC  240

ACGCTGGCGGTGACGGAAGAAAAAGCGAGTGAACTCACTGTCGGCGACAAAGCGAAACTC  300

GTTTATGGCGGCGACGTCTAGGGCGTCATTGAAATCGCCGATATTTACCGCCCGGATAAA  360
```

-continued

```
ACGAAAGAAGCCAAGCTCGTCTATAAAACCGATGAACTCGCTCACCCGGGCGTGGGCAAG     420

CTGTTTGAAAAACCAGATGTGTAGGTCGGCGGAGCGGTTAGGCTCGTCAAACGGAGCGAC     480

AAAGGCCAGTTTGCTCCGTTTTATTTCGATCCGGCCGAAACGCGGAAACGATTTGCCGAA     540

CTCGGCTGGAATACCGTCGTGGGCTTCCAAACACGCAACCCGGTTCACCGGGCCCATGAA     600

TACATTCAAAAATGCGCGCTTGAAATCGTGGACGGCTTGTTTTTAAACCCGCTCGTCGGC     660

GAAACGAAAGCGGACGATATTCCGGCCGACATCCGGATGGAAAGCTATCAAGTGCTGCTG     720

GAAAACTATTATCCGAAAGACCGCGTTTTCTTGGGCGTCTTCCAAGCTGCGATGCGGTAT     780

GCCGGTCCGCGCGAAGCGATTTTCCATGCCATGGTGCGGAAAAACTTCGGCTGCACGCAC     840

TTCATCGTCGGCCGGGACCATGCGGGCGTCGGCAACTATTACGGCACGTATGATGCGCAA     900

AAAATCTTCTCGAACTTTACAGCCGAAGAGCTTGGCATTACACCGCTCTTTTTCGAACAC     960

AGCTTTTATTGCAGGAAATGCGAAGGGATGGCATCGAGGAAAACATGCCCGCACGACGCA    1020

CAATATCACGTTGTCCTTTCTGGCACGAAAGTCCGTGAAATGTTGCGTAACGGCCAAGTG    1080

CCGCCGAGCACATTCAGCCGTCCGGAAGTGGCCGGCGTTTTGATCAAAGGGCTGCAAGAA    1140

CGCGAAACGGTCACCCCGTCGACACGCTAAAGGAGGAGCGAGATGAGCACGAATATCGTT    1200

TGGCATCATACATCGGTGACAAAAGAAGATCGCCGCCAACGCAACGG                 1247
```

[**0035**] The Bst sulfurylase polypeptide (SEQ ID NO:2) is 386 amino acid residues in length and is presented using the three letter amino acid code.

```
Bst Sulfurylase Amino Acid Sequence (SEQ ID NO: 2)
Met Ser Leu Ser Ile Pro His Gly Gly Thr Leu Ile
  1               5                  10

Asn Arg Trp Asn Pro Asp Tyr Pro Ile Asp Glu Ala
         15                  20

Thr Lys Thr Ile Glu Leu Ser Lys Ala Glu Leu Ser
     25                  30                  35

Asp Leu Glu Leu Ile Gly Thr Gly Ala Tyr Ser Pro
                 40                  45

Leu Thr Gly Phe Leu Thr Lys Ala Asp Tyr Asp Ala
         50                  55

Val Val Glu Thr Met Arg Leu Ala Asp Gly Thr Val
 60                  65                  70

Trp Ser Ile Pro Ile Thr Leu Ala Val Thr Glu Glu
             75                  80

Lys Ala Ser Glu Leu Thr Val Gly Asp Lys Ala Lys
     85                  90                  95

Leu Val Tyr Gly Gly Asp Val Tyr Gly Val Ile Glu
                100                 105

Ile Ala Asp Ile Tyr Arg Pro Asp Lys Thr Lys Glu
        110                 115

Ala Lys Leu Val Tyr Lys Thr Asp Glu Leu Ala His
120                 125                 130

Pro Gly Val Arg Lys Leu Phe Glu Lys Pro Asp Val
            135                 140

Tyr Val Gly Gly Ala Val Thr Leu Val Lys Arg Thr
```

-continued

```
    145                 150                 155

Asp Lys Gly Gln Phe Ala Pro Phe Tyr Phe Asp Pro
                160                 165

Ala Glu Thr Arg Lys Arg Phe Ala Glu Leu Gly Trp
        170                 175

Asn Thr Val Val Gly Phe Gln Thr Arg Asn Pro Val
180                 185                 190

His Arg Ala His Glu Tyr Ile Gln Lys Cys Ala Leu
            195                 200

Glu Ile Val Asp Gly Leu Phe Leu Asn Pro Leu Val
    205                 210                 215

Gly Glu Thr Lys Ala Asp Asp Ile Pro Ala Asp Ile
                220                 225

Arg Met Glu Ser Tyr Gln Val Leu Leu Glu Asn Tyr
        230                 235

Tyr Pro Lys Asp Arg Val Phe Leu Gly Val Phe Gln
240                 245                 250

Ala Ala Met Arg Tyr Ala Gly Pro Arg Glu Ala Ile
            255                 260

Phe His Ala Met Val Arg Lys Asn Phe Gly Cys Thr
    265                 270                 275

His Phe Ile Val Gly Arg Asp His Ala Gly Val Gly
                280                 285

Asn Tyr Tyr Gly Thr Tyr Asp Ala Gln Lys Ile Phe
        290                 295

Ser Asn Phe Thr Ala Glu Glu Leu Gly Ile Thr Pro
300                 305                 310

Leu Phe Phe Glu His Ser Phe Tyr Cys Thr Lys Cys
```

-continued

```
            315                     320

Glu Gly Met Ala Ser Thr Lys Thr Cys Pro His Asp
    325                     330                     335

Ala Gln Tyr His Val Val Leu Ser Gly Thr Lys Val
                340                     345

Arg Glu Met Leu Arg Asn Gly Gln Val Pro Pro Ser
        350                     355

Thr Phe Ser Arg Pro Glu Val Ala Ala Val Leu Ile
360                     365                     370

Lys Gly Leu Gln Glu Arg Glu Thr Val Thr Pro Ser
            375                     380

Thr Arg
    385
```

[0036] In one embodiment, the thermostable sulfurylase is active at temperatures above ambient to at least 50° C. This property is beneficial so that the sulfurylase will not be denatured at higher temperatures commonly utilized in polymerase chain reaction (PCR) reactions or sequencing reactions. In one embodiment, the ATP sulfurylase is from a thermophile. The thermostable sulfurylase can come from thermophilic bacteria, including but not limited to, *Bacillus stearothermophilus, Thermus thermophilus, Bacillus caldolyticus, Bacillus subtilis, Bacillus thermoleovorans, Pyrococcus furiosus, Sulfolobus acidocaldarius, Rhodothermus obamensis, Aquifex aeolicus, Archaeoglobus fulgidus, Aeropyrum pernix, Pyrobaculum aerophilum, Pyrococcus abyssi, Penicillium chrysogenum, Sulfolobus solfataricus* and *Thermomonospora fusca.*

[0037] The homology of twelve ATP sulfurylases can be shown graphically in the ClustalW analysis in Table 1. The alignment is of ATP sulfurylases from the following species: *Bacillus stearothermophilus* (Bst), University of Oklahoma—Strain 10 (Univ of OK), *Aquifex aeolicus* (Aae), *Pyrococcus furiosus* (Pfu), *Sulfolobus solfataricus* (Sso), *Pyrobaculum aerophilum* (Pae), *Archaeoglobus fulgidus* (Afu), *Penicillium chrysogenum* (Pch), *Aeropyrum pernix* (Ape), *Saccharomyces cerevisiae* (Sce), and *Thermomonospora fusca* (Tfu).

Table 1: ClustalW Analysis of ATP Sulfurylase Amino Acid Sequence

```
                       10        20        30        40        50
               ....|....|....|....|....|....|....|....|....|....|
Bst-Sulf       -----MSLSIPHGGTLINRWNPDYP-----IDEATK---------TIELS 31
Univ of OK     -----MSVSIPHGGTLINRWNPDYP-----LDEATK---------TIELS 31
Aae-Sulf       --------------------------MEKIKYLKS----------IQIS 13
Pfu-Sulf       ------MVSKPHGGKLIRRIAAPRTRERILSEQHEYP--------RVQID 36
Sso-Sulf       ------MNLIGHG----KVEIVERIKTISDFKELHRIEVK---------- 30
Pae-Sulf       --MPMPAPLEPHGGRLVYNVIEDRDKAAAMIQGLPSIEIEPTLGPDGSPI 48
Afu-Sulf       --MPLIKTPPPHGGKLVERVVKKRDIAEKMIAGCPTYELKPTTLPDGTPI 48
Pch-Sulf       -----MAN-APHGGVLKDLLARDAPRQAELAAEAES--LP-----AVTLT 37
Ape-Sulf       MGCSVGLVSRPHGGRLVRRVLSGRRREIFESQYREMP--------RLEVP 42
Sce-Sulf       -----MP--APHGGILQDLIARDALKKNELLSEAQSSDIL-----VWNLT 38
Tfu-Sulf       ----------------MSQVSDAVGR----YQLSQLDF----------LE 20

                       60        70        80        90       100
               ....|....|....|....|....|....|....|....|....|....|
Bst-Sulf       KAELSDLELIGTGAYSPLTGFLTKADYDAVVETMRLADGTVWSIPITLAV 81
Univ of OK     KAELSDLELIGTGAYSPLTGFLTKTDYDAVVETMRLSDGTVWSIPVTLAV 81
Aae-Sulf       QRSVLDLKLLAVGAFTPLDRFMGEEDYRNVVESMRLKSGTLFPIPITLPM 63
Pfu-Sulf       HGRAIDLENIAHGVYSPLKGFLTREDFESVLDYMRLSDDTPWTIPIVLDV 86
Sso-Sulf       RQLAHEIVSIAYGFLSPLKGFMNYEEVDGVVENMRLPNGVLWPIPLVFDY 80
Pae-Sulf       RNPYREIMSIAYGFFSPVEGFMTRNEVESILKERRLLNGWLFPFPLIYDV 98
Afu-Sulf       RHVYREIMSVCYGFFSPVEGSMVQNELERVLNERRLLSEWIFPYPILFDI 98
Pch-Sulf       ERQLCDLELIMNGGFSPLEGFMNQADYDRVCEDNRLADGNVFSMPITLDA 87
Ape-Sulf       LERAIDAEDLARGVFSPLEGFMVEDDYLSVLSRMRLSNDLPWTIPIVLDA 92
Sce-Sulf       PRQLCDIELILNGGFSPLTGFLNENDYSSVVTDSRLADGTLWTIPITLDV 88
Tfu-Sulf       AEAIFIMREVAAEFERPVLLFSGGKD-S--VVMLRIAEKAFWPAPIPFPV 67

                      110       120       130       140       150
               ....|....|....|....|....|....|....|....|....|....|
Bst-Sulf       TEEKAS--ELTVGDKAKLV-YGGDVY--GVIEIADIYRP-DKTKEAKLVY 125
Univ of OK     TEEKAK--ELAVGDKAKLV-YRGDVY--GVIEIADIYRP-DKTKEAKLVY 125
Aae-Sulf       EKEIAK--DLKEGEWIVLR--DPKNVPLAIMRVEEVYKW-NLEYEAKNVL 108
Pfu-Sulf       G-----EPTFEGGDAILLY---YENPPIARMHVEDIYTY-DKKEFAVKVF 127
Sso-Sulf       SQN--EK--VKEGDTIGIT---YLGKPLAIMKVKEIFKY-DKLKIAEKVY 122
Pae-Sulf       DEE--KIKGIKEGDSVLLK---LKGKPLAVLNVEEIWRLPDRKELADAVF 143
Afu-Sulf       SEEDYKALDVKEGDRLLLM---LKGQPFATLDIEEVYKI-DPVDVATRTF 144
Pch-Sulf       SQEVIDEKKLQAASRITLRDFRDDRN-LAILTIDDIYRP-DKTKEAKLVF 135
Ape-Sulf       NREWVLNEGVSAGDDIILT---YHGLPIAVLTLEDIYSW-DKGLHAEKVF 138
Sce-Sulf       DEAFAN--QIKPDTRIAL--FQDDEIPIAILTVQDVYKP-NKTIEAEKVF 133
Tfu-Sulf       MH-------VDTGHNFPEV-IEFRDKRVAELGVRLIVASVQDLIDAGKVV 109

                      160       170       180       190       200
               ....|....|....|....|....|....|....|....|....|....|
Bst-Sulf       KTD-----------ELAHPGVRKLF-EKPDVYVGGAVTLVKRTDKG-QFA 162
Univ of OK     KTD-----------ELAHPGVRKLF-EKPDVYVGGEITLVKRTDKG-QFA 162
Aae-Sulf       GTT-----------DPRHPLVAEMH-TWGEYYISGELKVIQLPKYY-DFP 145
Pfu-Sulf       KTD-----------DPNHLGVARVY-SMGKYLVGGGIELLNELPN--PFA 163
Sso-Sulf       KT--K---------DIKHPGVKRTL-SYADAFLAGDVWLVREPQFNKPYS 160
Pae-Sulf       GTPERNKEVVKKRFDEKHPGWLIYR-SMRPMALAGKITVVNPPRFKEPYS 192
Afu-Sulf       GTPEKNPEVVREPFDDKHPGYVIYK-MHNPIILAGKYTIVNEPKFKEPYD 193
Pch-Sulf       GG------------DPEHPAIVYLNNTVKEFYIGGKIEAVNKLNHY-DYV 172
Ape-Sulf       KTR-----------DPNHPGVEATY-KRGDILLGGRLELIQGPPN--PLE 174
Sce-Sulf       RG------------DPEHPAISYLFNVAGDYYVGGSLEAIQLPQHY-DYP 170
Tfu-Sulf       EP-----------------KGR-WASRNRLQTAALLEAIEKYGFDAAFG 140
```

```
                    210       220       230       240       250
               ....|....|....|....|....|....|....|....|....|....|
Bst-Sulf       PEYFDPAETRKRFA-ELGWNTVVGFQTRNPVHRAHEYIQKCALEIVDG--  209
Univ of OK     AFYFDPAETRKKFA-EFGWNTVVGFQTRNPVHRAHEYIQKCALEIVDG--  209
Aae-Sulf       EYRKTPKQVREEIK-SLGLDKDIVAFQTRNPMHRVHEELTKRAMEKVGGG-  193
Pfu-Sulf       KYTLRPVETRILFK-ERGWKTIVAFQTRNVPHLGHEYVQKAALTFVDG--  210
Sso-Sulf       EFWLTPRMHRTVFE-KKGWKRVVAFQTRNVPHTGHEYLMKFAWFAANENQ  209
Pae-Sulf       RFWMPPRVSREYVE-KKGWRIVVAHQTRNVPHIGHEMLMKRAMFVAGG--  239
Afu-Sulf       RFWFPPSKCREVIKNEKKWRTVIAHQTRNVPHVGHEMLMKCAAYTG-D--  240
Pch-Sulf       ALRYTPAELRVHFD-KLGWSRVVAFQTRNPMHRAHRELTVRAARSRQAN-  220
Ape-Sulf       RYTLWPVETRVLFK-EKGWRTVAAFQTRNVPHLGHEYVQKAALTFVDG--  221
Sce-Sulf       GLRKTPAQLRLEFQ-SRQWDRVVAFQTRNPMHRAHRELTVRAAREANAK-  218
Tfu-Sulf       GARRDEEKARAKER---------VFSFRDEFGQWDPKNQRPEEWNLYN--  179

                    260       270       280       290       300
               ....|....|....|....|....|....|....|....|....|....|
Bst-Sulf       --------LFLNPLVGETKADDIPADIRMESYQVLLEN-YYPKDRVFLGV  250
Univ of OK     --------LFLNPLVGETKSDDIPADIRMESYQVLLEN-YYPKDRVFLGV  250
Aae-Sulf       --------LLLHPVVGLTKPGDVDVYTRMRIYKVLYEK-YYDKKKTILAF  234
Pfu-Sulf       --------LFINPVLGRKKKGDYKDEVIIKAYY-LIMK-YCSN-------  243
Sso-Sulf       KVDEPRTGILVNVVIGEKRVGDYIDEAILLTHDALSKYGYISPKVHLLSF  259
Pae-Sulf       --ERPGDAVLVNAIIGAKRPGDYVDEAILEGHEALNKAGYFHPDRHVVTM  287
Afu-Sulf       --IEPCHGILVNAIIGAKRRGDYPDEAILEGHEAVNKYGYIKPERHMVTF  288
Pch-Sulf       --------VLIHPVVGLTKPGDIDHFTRVRAYQALLPR-YPNG-MAVLGL  260
Ape-Sulf       --------LLVHPLAGWKKRGDYRDEVIIRAYEALITH-YYPRGVVVLSV  262
Sce-Sulf       --------VLIHPVVGLTKPGDIDHHTRVRVYQEIIKR-YPNG-IAFLSL  258
Tfu-Sulf       ---------TR------VHRGENIRVFPLSNWTELDVWHYIRRE------  208

                    310       320       330       340       350
               ....|....|....|....|....|....|....|....|....|....|
Bst-Sulf       FQAAMRYAGPREAIFHAMVRKNFGCTHFIVGRDHAGVG---N---YYGTY  294
Univ of OK     FQAAMRYAGPREAIFHAMVRKNFGCTHFIVGRDHAGVG---N---YYGTY  294
Aae-Sulf       LPLAMRMAGPREALWHGIIRRNYGATHFIVGRDHASPGKDSKGKPFYDPY  284
Pfu-Sulf       ------------TTHHAIMRKTSTSSQT----------------------  259
Sso-Sulf       TLWDMRYAGPREALLHAIIRSNLGCTHHVFGRDHAGVG------NYYSPY  303
Pae-Sulf       TLWDMRYGNPIESLLHGIIRQNMGATHHMFGRDHAATG------DYYDPY  331
Afu-Sulf       TLWDMRYGNPIESILHGVIRQNMGCTHHMFGRDHAAVG------EYYDMY  332
Pch-Sulf       LGLAMRMGGPREAIWHAIIRKNHGATHFIVGRDHAGPGSNSKGEDFYGPY  310
Ape-Sulf       LRMNMNYAGPREAVHHAIVRKNFGATHFIVGRDHAGVG------SYYGPY  306
Sce-Sulf       LPLAMRMSGDREAVWHAIIRKNYGASHFIVGRDHAGPGKNSKGVDFYGPY  308
Tfu-Sulf       ---GLRLP----SIYFAHRRRVFERDGILLP---DS--------PYVTRD  240

                    360       370       380       390       400
               ....|....|....|....|....|....|....|....|....|....|
Bst-Sulf       DAQKIFSNFT------------AEELGITPLFFEHSFYCTKCEGMASTK-  331
Univ of OK     DAQKIFLNFT------------AEELGITPLFFEHSFYCTKCEGMASTK-  331
Aae-Sulf       EAQELFKKY-------------EDEIGIKMVPFEELVYVPELDQYVEIN-  320
Pfu-Sulf       --------------------------------------------------  259
Sso-Sulf       EAHEIFDS----------INEED--LLIKPIFLRENYYCPRCGSIENE--  339
Pae-Sulf       ATQYLWTRGLPSYGLNEPPHMTDKGLRIKPVNLGEFAYCPKCGEYTYLGM  381
Afu-Sulf       ATQILWSQGIPSFGFEAPPNEVDYGLKIIPQNMAEFWYCPICQEIAYS--  380
Pch-Sulf       DAQHAVEKY-------------KDELGIEVVEFQMVTYLPDTDEYRPVD-  346
Ape-Sulf       EAWEIFREF--------------PDLGITPLFVREAYYCRRCGGMVNEK-  341
Sce-Sulf       DAQELVESY-------------KHELDIEVVPFRMVTYLPDEDRYAPID-  344
Tfu-Sulf       EDEEVFEAS--------------------VRYRTVGDMTCTGAVLST--  267

                    410       420       430       440       450
               ....|....|....|....|....|....|....|....|....|....|
Bst-Sulf       --------TCPHDAQYHVVLSGTKVRE-MLRNGQVPPSTFSRPEVAAVLI  372
Univ of OK     --------TCPHDAKYHVVLSGTKVRE-MLRNGQVPPSTFSRPEVAAVLI  372
```

```
Aae-Sulf      --------EAKKRNLKYINIISGTEIIRENFILKQGRKLPEWFTRPEVAEILA 362
Pfu-Sulf      ----------------------------------------------------- 259
Sso-Sulf      --------ILCDHKDEKQEFSCSLIRS-IILDEVKPTKMVMRPEVYDVLM 380
Pae-Sulf      SYEGYKEVALCGHT--PERISCSLLRG-IIIEGLRPPKVVMRPEVYDVIV 428
Afu-Sulf      --------ENCGHTDAKQKFSCSFLRG-MVAEGVFPPRVVMRPEVYKQIV 421
Pch-Sulf      --------QVPAG-VKTLNIISGTELRR-RLRSGAHIPEWFSYPEVVKILR 386
Ape-Sulf      --------VCPHGDEYRVRISGTRLRE-MLGRGERPPEYMMRPEVADAII 382
Sce-Sulf      --------QIDTTKTRTLNISGTELRR-RLRVGGEIPEWFSYPEVVKILR 385
Tfu-Sulf      --------AT-TLDEVIAEIAATRITE---RGQTRADDRGSEAAMEERKR 305
```

```
                   460       470       480       490       500
              ....|....|....|....|....|....|....|....|....|....|
Bst-Sulf      KGLQER-----------------ETVT----------------PSTR--- 386
Univ of OK    KGLQER-----------------ETVA----------------PSAR--- 386
Aae-Sulf      ETYVPKHKQGFCVWLTGLPCAGKSTIA-EILATMLQARGRKVTLLDGDVV 411
Pfu-Sulf      -------------------------------------------------- 259
Sso-Sulf      KAAEQYGFGS--------PFVTEEYLE--------------KRQSILG-- 406
Pae-Sulf      KWWRVYGY----------PYVTDKYLR--------------IKEQELEVE 454
Afu-Sulf      KWWKVYNY----------PFVNRKYLE--------------LKNKELEID 447
Pch-Sulf      ESNPPRATQGFTIFLTGYMNSGKDAIARALQVTLNQQGGRSVSLLLGDTV 436
Ape-Sulf      SHPDPFI------------------------------------------- 389
Sce-Sulf      ESNPPRPKQGFSIVLGNSLTVSREQLSIALLSTFLQFGGGRYYKIFEHNN 435
Tfu-Sulf      EGYF---------------------------------------------- 309
```

```
                   510       520       530       540       550
              ....|....|....|....|....|....|....|....|....|....|
Bst-Sulf      -------------------------------------------------- 386
Univ of OK    -------------------------------------------------- 386
Aae-Sulf      RTHLSRGLGFSKEDRITNILRVGFVASEIVKHNGVVICALVSPYRSARN- 460
Pfu-Sulf      -------------------------------------------------- 259
Sso-Sulf      -------------------------------------------------- 406
Pae-Sulf      L------------------------------------------------- 455
Afu-Sulf      LPAMEVPKA----------------------------------------- 456
Pch-Sulf      RHELSSELGFTREDRHTNIQRIAFVATELTRAGAAVIAAPIAPYEESRKF 486
Ape-Sulf      -------------------------------------------------- 389
Sce-Sulf      KTELLSLI------------------QDFIGSGSGLIIP--NQWEDD--- 462
Tfu-Sulf      -------------------------------------------------- 309
```

```
                   560       570       580       590       600
              ....|....|....|....|....|....|....|....|....|....|
Bst-Sulf      -------------------------------------------------- 386
Univ of OK    -------------------------------------------------- 386
Aae-Sulf      QVRNMMEEGKFIEVFVDAPVEVCEERDVKGLYKKAKEGLIKGFTGVDDPY 510
Pfu-Sulf      -------------------------------------------------- 259
Sso-Sulf      -------------------------------------------------- 406
Pae-Sulf      -------------------------------------------------- 455
Afu-Sulf      -------------------------------------------------- 456
Pch-Sulf      ARDAVSQAGSFFLVHVATPLEHCEQSDKRGIYAAARRGEIKGFTGVDDPY 536
Ape-Sulf      -------------------------------------------------- 389
Sce-Sulf      -KDSVVGKQNVYLLDTSS-------------------------------- 479
Tfu-Sulf      -------------------------------------------------- 309
```

```
                   610       620       630
              ....|....|....|....|....|....|....|..
Bst-Sulf      ------------------------------------- 386 (SEQ ID NO:2)
Univ of OK    ------------------------------------- 386 (SEQ ID NO:22)
Aae-Sulf      EPPVAPEVRVDTTKLTPEESALKILEFLKKEGFIKD- 546 (SEQ ID NO:23)
Pfu-Sulf      ------------------------------------- 259 (SEQ ID NO:24)
Sso-Sulf      ------------------------------------- 406 (SEQ ID NO:25)
Pae-Sulf      ------------------------------------- 455 (SEQ ID NO:26)
```

```
Afu-Sulf     ------------------------------------- 456 (SEQ ID NO:27)
Pch-Sulf     ETPEKADLVVDFSKQSVRSIVHEIILVLESQGFLERQ 573 (SEQ ID NO:28)
Ape-Sulf     ------------------------------------- 389 (SEQ ID NO:29)
Sce-Sulf     ----SADIQLESADEPISHIVQKVVLFLEDNGFFVF- 511 (SEQ ID NO:30)
Tfu-Sulf     ------------------------------------- 309 (SEQ ID NO:31)
```

[0038] A thermostable sulfurylase polypeptide is encoded by the open reading frame ("ORF") of a thermostable sulfurylase nucleic acid. An ORF corresponds to a nucleotide sequence that could potentially be translated into a polypeptide. A stretch of nucleic acids comprising an ORF is uninterrupted by a stop codon. An ORF that represents the coding sequence for a full protein begins with an ATG "start" codon and terminates with one of the three "stop" codons, namely, TAA, TAG, or TGA. For the purposes of this invention, an ORF may be any part of a coding sequence, with or without a start codon, a stop codon, or both. For an ORF to be considered as a good candidate for coding for a bona fide cellular protein, a minimum size requirement is often set, e.g., a stretch of DNA that would encode a protein of 50 amino acids or more.

[0039] The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences shown in SEQ ID NO:1 due to degeneracy of the genetic code and thus encode the same thermostable sulfurylase proteins as that encoded by the nucleotide sequences shown in SEQ ID NO:1. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NO:2. In addition to the thermostable sulfurylase nucleotide sequence shown in SEQ ID NO:1 it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the thermostable sulfurylase polypeptides may exist within a population (e.g., the bacterial population). Such genetic polymorphism in the thermostable sulfurylase genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a thermostable sulfurylase protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the thermostable sulfurylase genes. Any and all such nucleotide variations and resulting amino acid polymorphisms in the thermostable sulfurylase polypeptides, which are the result of natural allelic variation and that do not alter the functional activity of the thermostable sulfurylase polypeptides, are intended to be within the scope of the invention.

[0040] Moreover, nucleic acid molecules encoding thermostable sulfurylase proteins from other species, and thus that have a nucleotide sequence that differs from the sequence SEQ ID NO:1 are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the thermostable sulfurylase cDNAs of the invention can be isolated based on their homology to the thermostable sulfurylase nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. The invention further includes the nucleic acid sequence of SEQ ID NO:1 and mature and variant forms thereof, wherein a first nucleotide sequence comprising a coding sequence differing by one or more nucleotide sequences from a coding sequence encoding said amino acid sequence, provided that no more than 11% of the nucleotides in the coding sequence differ from the coding sequence.

[0041] Another aspect of the invention pertains to nucleic acid molecules encoding a thermostable sulfurylase protein that contains changes in amino acid residues that are not essential for activity. Such thermostable sulfurylase proteins differ in amino acid sequence from SEQ ID NO:2 yet retain biological activity. In separate embodiments, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 96%, 97%, 98% or 99% homologous to the amino acid sequence of SEQ ID NO:2. An isolated nucleic acid molecule encoding a thermostable sulfurylase protein homologous to the protein of SEQ ID NO: 2 can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:1 such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

[0042] Mutations can be introduced into SEQ ID NO:2 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted, non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined within the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g. threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted non-essential amino acid residue in the thermostable sulfurylase protein is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a thermostable sulfurylase coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for thermostable sulfurylase biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:1, the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

[0043] The relatedness of amino acid families may also be determined based on side chain interactions. Substituted amino acids may be fully conserved "strong" residues or fully conserved "weak" residues. The "strong" group of conserved amino acid residues may be any one of the following groups: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW, wherein the single letter amino acid codes are grouped by those amino acids that may be substituted for each other. Likewise, the "weak" group of conserved residues may be any one of the following: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, VLIM, HFY, wherein the letters within each group represent the single letter amino acid code.

[0044] The thermostable sulfurylase nucleic acid of the invention includes the nucleic acid whose sequence is provided herein, or fragments thereof. The invention also includes mutant or variant nucleic acids any of whose bases may be changed from the corresponding base shown herein while still encoding a protein that maintains its sulfurylase-like activities and physiological functions, or a fragment of

such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

[0045] A thermostable sulfurylase nucleic acid can encode a mature thermostable sulfurylase polypeptide. As used herein, a "mature" form of a polypeptide or protein disclosed in the present invention is the product of a naturally occurring polypeptide or precursor form or proprotein. The naturally occurring polypeptide, precursor or proprotein includes, by way of nonlimiting example, the full-length gene product, encoded by the corresponding gene. Alternatively, it may be defined as the polypeptide, precursor or proprotein encoded by an ORF described herein. The product "mature" form arises, again by way of nonlimiting example, as a result of one or more naturally occurring processing steps as they may take place within the cell, or host cell, in which the gene product arises. Examples of such processing steps leading to a "mature" form of a polypeptide or protein include the cleavage of the N-terminal methionine residue encoded by the initiation codon of an ORF, or the proteolytic cleavage of a signal peptide or leader sequence. Thus a mature form arising from a precursor polypeptide or protein that has residues 1 to N, where residue 1 is the N-terminal methionine, would have residues 2 through N remaining after removal of the N-terminal methionine. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an N-terminal signal sequence from residue 1 to residue M is cleaved, would have the residues from residue M+1 to residue N remaining. Further as used herein, a "mature" form of a polypeptide or protein may arise from a step of post-translational modification other than a proteolytic cleavage event. Such additional processes include, by way of non-limiting example, glycosylation, myristoylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or a combination of any of them.

[0046] The term "isolated" nucleic acid molecule, as utilized herein, is one, which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5'- and 3'-termini of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated thermostable sulfurylase nucleic acid molecules can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell/tissue from which the nucleic acid is derived (e.g., brain, heart, liver, spleen, etc.). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.

[0047] A nucleic acid molecule of the invention, e.g., a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1 or a complement of this aforementioned nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NO:1 as a hybridization probe, thermostable sulfurylase molecules can be isolated using standard hybridization and cloning techniques (e.g., as described in Sambrook, et al., (eds.), MOLECULAR CLONING: A LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989; and Ausubel, et al., (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, N.Y., 1993.)

[0048] A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to thermostable sulfurylase nucleotide sequences can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

[0049] As used herein, the term "complementary" refers to Watson-Crick or Hoogsteen base pairing between nucleotides units of a nucleic acid molecule, and the term "binding" means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, van der Waals, hydrophobic interactions, and the like. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

[0050] Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, respectively, and are at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice. Derivatives are nucleic acid sequences or amino acid sequences formed from the native compounds either directly or by modification or partial substitution. Analogs are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound but differs from it in respect to certain components or side chains. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type. Homologs are nucleic acid sequences or amino acid sequences of a particular gene that are derived from different species.

[0051] Derivatives and analogs may be full length or other than full length, if the derivative or analog contains a modified nucleic acid or amino acid, as described below.

Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 89% identity over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the aforementioned proteins under stringent, moderately stringent, or low stringent conditions. See e.g. Ausubel, et al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, N.Y., 1993, and below.

[0052] A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences encode those sequences coding for isoforms of thermostable sulfurylase polypeptides. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the invention, homologous nucleotide sequences include nucleotide sequences encoding for a thermostable sulfurylase polypeptide of species other than humans, including, but not limited to: vertebrates, and thus can include, e.g., frog, mouse, rat, rabbit, dog, cat cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions in SEQ ID NO:1, as well as a polypeptide possessing thermostable sulfurylase biological activity. Various biological activities of the thermostable sulfurylase proteins are described below.

[0053] The thermostable sulfurylase proteins of the invention include the sulfurylase protein whose sequence is provided herein. The invention also includes mutant or variant proteins any of whose residues may be changed from the corresponding residue shown herein while still encoding a protein that maintains its sulfurylase-like activities and physiological functions, or a functional fragment thereof. The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})2$, that bind immunospecifically to any of the proteins of the invention. This invention also includes a variant or a mature form of the amino acid sequence of SEQ ID NO:2, wherein one or more amino acid residues in the variant differs in no more than 4% of the amino acic residues from the amino acid sequence of the mature form.

[0054] Several assays have been developed for detection of the forward ATP sulfurylase reaction. The colorimetric molybdolysis assay is based on phosphate detection (see e.g., Wilson and Bandurski, 1958. *J. Biol. Chem.* 233: 975-981), whereas the continuous spectrophotometric molybdolysis assay is based upon the detection of NADH oxidation (see e.g., Seubert, et al., 1983. *Arch. Biochem. Biophys.* 225: 679-691; Seubert, et al., 1985. *Arch. Biochem. Biophys.* 240: 509-523). The later assay requires the presence of several detection enzymes.

[0055] Suitable enzymes for converting ATP into light include luciferases, e.g., insect luciferases. Luciferases produce light as an end-product of catalysis. The best known light-emitting enzyme is that of the firefly, *Photinus pyralis* (Coleoptera). The corresponding gene has been cloned and expressed in bacteria (see e.g., de Wet, et al., 1985. *Proc. Natl. Acad. Sci. USA* 80: 7870-7873) and plants (see e.g., Ow, et al., 1986. *Science* 234: 856-859), as well as in insect (see e.g., Jha, et al., 1990. *FEBS Lett.* 274: 24-26) and mammalian cells (see e.g., de Wet, et al., 1987. *Mol. Cell. Biol.* 7: 725-7373; Keller, et al., 1987. *Proc. Natl. Acad. Sci. USA* 82: 3264-3268). In addition, a number of luciferase genes from the Jamaican click beetle, *Pyroplorus plagiophihalamus* (Coleoptera), have recently been cloned and partially characterized (see e.g., Wood, et al., 1989. *J. Biolumin. Chemilumin.* 4: 289-301; Wood, et al., 1989. *Science* 244: 700-702). Distinct luciferases can sometimes produce light of different wavelengths, which may enable simultaneous monitoring of light emissions at different wavelengths. Accordingly, these aforementioned characteristics are unique, and add new dimensions with respect to the utilization of current reporter systems.

[0056] Firefly luciferase catalyzes bioluminescence in the presence of luciferin, adenosine 5'-triphosphate (ATP), magnesium ions, and oxygen, resulting in a quantum yield of 0.88 (see e.g., McElroy and Selinger, 1960. *Arch. Biochem. Biophys.* 88: 136-145). The firefly luciferase bioluminescent reaction can be utilized as an assay for the detection of ATP with a detection limit of approximately 1×10^{-13} M (see e.g., Leach, 1981. *J. Appl. Biochem.* 3: 473-517). In addition, the overall degree of sensitivity and convenience of the luciferase-mediated detection systems have created considerable interest in the development of firefly luciferase-based biosensors (see e.g., Green and Kricka, 1984. *Talanta* 31: 173-176; Blum, et al., 1989. *J. Biolumin. Chemilumin.* 4: 543-550).

[0057] The development of new reagents have made it possible to obtain stable light emission proportional to the concentrations of ATP (see e.g., Lundin, 1982. Applications of firefly luciferase In; *Luminescent Assays* (Raven Press, New York). With such stable light emission reagents, it is possible to make endpoint assays and to calibrate each individual assay by addition of a known amount of ATP. In addition, a stable light-emitting system also allows continuous monitoring of ATP-converting systems.

[0058] In a preferred embodiment, the ATP generating-ATP converting fusion protein is attached to an affinity tag. The term "affinity tag" is used herein to denote a peptide segment that can be attached to a polypeptide to provide for purification or detection of the polypeptide or provide sites for attachment of the polypeptide to a substrate. In principal, any peptide or protein for which an antibody or other specific binding agent is available can be used as an affinity tag. Affinity tags include a poly-histidine tract or a biotin carboxyl carrier protein (BCCP) domain, protein A (Nilsson et al., EMBO J. 4:1075, 1985; Nilsson et al., Methods Enzymol. 198:3, 1991), glutathione S transferase (Smith and Johnson, Gene 67:31, 1988), substance P, Flag.™. peptide (Hopp et al., Biotechnology 6:1204-1210, 1988; available from Eastman Kodak Co., New Haven, Conn.), streptavidin binding peptide, or other antigenic epitope or binding domain. See, in general Ford et al., Protein Expression and

Purification 2: 95-107, 1991. DNAs encoding affinity tags are available from commercial suppliers (e.g., Pharmacia Biotech, Piscataway, N.J.).

[0059] As used herein, the term "poly-histidine tag," when used in reference to a fusion protein refers to the presence of two to ten histidine residues at either the amino- or carboxy-terminus of a protein of interest. A poly-histidine tract of six to ten residues is preferred. The poly-histidine tract is also defined functionally as being a number of consecutive histidine residues added to the protein of interest which allows the affinity purification of the resulting fusion protein on a nickel-chelate or IDA column.

[0060] In some embodiments, the fusion protein has an orientation such that the sulfurylase polypeptide is N-terminal to the luciferase polypeptide. In other embodiments, the luciferase polypeptide is N-terminal to the sulfurylase polypeptide. As used herein, the term sulfurylase-luciferase fusion protein refers to either of these orientations. The terms "amino-terminal" (N-terminal) and "carboxyl-terminal" (C-terminal) are used herein to denote positions within polypeptides and proteins. Where the context allows, these terms are used with reference to a particular sequence or portion of a polypeptide or protein to denote proximity or relative position. For example, a certain sequence positioned carboxyl-terminal to a reference sequence within a protein is located proximal to the carboxyl terminus of the reference sequence, but is not necessarily at the carboxyl terminus of the complete protein.

[0061] The fusion protein of this invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, e.g. by employing blunt-ended or "sticky"-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, Ausubel et al. (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). The two polypeptides of the fusion protein can also be joined by a linker, such as a unique restriction site, which is engineered with specific primers during the cloning procedure. In one embodiment, the sulfurylase and luciferase polypeptides are joined by a linker, for example an ala-ala-ala linker which is encoded by a Notl restriction site.

[0062] In one embodiment, the invention includes a recombinant polynucleotide that comprises a coding sequence for a fusion protein having an ATP generating polypeptide sequence and an ATP converting polypeptide sequence. In a preferred embodiment, the recombinant polynucleotide encodes a sulfurylase-luciferase fusion protein. The term "recombinant DNA molecule" or "recombinant polynucleotide" as used herein refers to a DNA molecule which is comprised of segments of DNA joined together by means of molecular biological techniques. The term "recombinant protein" or "recombinant polypeptide" as used herein refers to a protein molecule which is expressed from a recombinant DNA molecule.

[0063] In one aspect, this invention discloses a sulfurylase-luciferase fusion protein with an N-terminal hexahistidine tag and a BCCP tag. The nucleic acid sequence of the disclosed N-terminal hexahistidine-BCCP luciferase-sulfurylase gene (His6-BCCP L-S) gene is shown below:

```
His6-BCCP L-S Nucleotide Sequence (SEQ ID NO: 3):
ATGCGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGGAAGCGCCAGCAGCA   60

GCGGAAATCAGTGGTCACATCGTACGTTCCCCGATGGTTGGTAGTTTCTACCGCACCCCA  120

AGCCCGGACGCAAAAGCGTTCATCGAAGTGGGTCAGAAAGTCAACGTGGGCGATACCCTG  180

TGCATCGTTGAAGCCATGAAAATGATGAACCAGATCGAAGCGGACAAATCCGGTACCGTG  240

AAAGCAATTCTGGTCGAAAGTGGACAACCGGTAGAATTTGACGAGCCGCTGGTCGTCATC  300

GAGGGATCCGAGCTCGAGATCCAAATGGAAGACGCCAAAAACATAAAGAAAGGCCCGGCG  360

CCATTCTATCCTCTAGAGGATGGAACCGCTGGAGAGCAACTGCATAAGGCTATGAAGAGA  420

TACGCCCTGGTTCCTGGAACAATTGCTTTTACAGATGCACATATCGAGGTGAACATCACG  480

TACGCGGAATACTTCGAAATGTCCGTTCGGTTGGCAGAAGCTATGAAACGATATGGGCTG  540

AATACAAATCACAGAATCGTCGTATGCAGTGAAAACTCTCTTCAATTCTTTATGCCGGTG  600

TTGGGCGCGTTATTTATCGGAGTTGCAGTTGCGCCCGCGAACGACATTTATAATGAACGT  660

GAATTGCTCAACAGTATGAACATTTCGCAGCCTACCGTAGTGTTTGTTTCCAAAAAGGGG  720

TTGCAAAAAATTTTGAACGTGCAAAAAAAATTACCAATAATCCAGAAAATTATTATCATG  780

GATTCTAAAACGGATTACCAGGGATTTCAGTCGATGTACACGTTCGTCACATCTGATCTA  840

CCTCCCGGTTTTAATGAATACGATTTTGTACCAGAGTCCTTTGATCGTGACAAAACAATT  900
```

-continued

```
GCACTGATAATGAATTCCTCTGGATCTACTGGGTTACCTAAGGGTGTGGCCCTTCCGCAT    960
AGAACTGCCTGCGTCAGATTCTCGCATGCCAGAGATCCTATTTTTGGCAATCAAATCATT   1020
CCGGATACTGCGATTTTAAGTGTTGTTCCATTCCATCACGGTTTTGGAATGTTTACTACA   1080
CTCGGATATTTGATATGTGGATTTCGAGTCGTCTTAATGTATAGATTTGAAGAAGAGCTG   1140
TTTTTACGATCCCTTCAGGATTACAAAATTCAAAGTGCGTTGCTAGTACCAACCCTATTT   1200
TCATTCTTCGCCAAAAGCACTCTGATTGACAAATACGATTTATCTAATTTACACGAAATT   1260
GCTTCTGGGGGCGCACCTCTTTCGAAAGAAGTCGGGGAAGCGGTTGCAAAACGCTTCCAT   1320
CTTCCAGGGATACGACAAGGATATGGGCTCACTGAGACTACATCAGCTATTCTGATTACA   1380
CCCGAGGGGGATGATAAACCGGGCGCGGTCGGTAAAGTTGTTCCATTTTTTGAAGCGAAG   1440
GTTGTGGATCTGGATACCGGGAAAACGCTGGGCGTTAATCAGAGAGGCGAATTATGTGTC   1500
AGAGGACCTATGATTATGTCCGGTTATGTAAACAATCCGGAAGCGACCAACGCCTTGATT   1560
GACAAGGATGGATGGCTACATTCTGGAGACATAGCTTACTGGGACGAAGACGAACACTTC   1620
TTCATAGTTGACCGCTTGAAGTCTTTAATTAAATACAAAGGATATCAGGTGGCCCCCGCT   1680
GAATTGGAATCGATATTGTTACAACACCCCAACATCTTCGACGCGGGCGTGGCAGGTCTT   1740
CCCGACGATGACGCCGGTGAACTTCCCGGCGCCGTTGTTGTTTTGGAGCACGGAAAGACG   1800
ATGACGGAAAAAGAGATCGTGGATTACGTCGCCAGTCAAGTAACAACCGCGAAAAAGTTG   1860
CGCGGAGGAGTTGTGTTTGTGGACGAAGTACCGAAAGGTCTTACCGGAAAACTCGACGCA   1920
AGAAAAATCAGAGAGATCCTCATAAAGGCCAAGAAGGGCGGAAAGTCCAAATTGGCGGCC   1980
GCTATGCCTGCTCCTCACGGTGGTATTCTACAAGACTTGATTGCTAGAGATGCGTTAAAG   2040
AAGAATGAATTGTTATCTGAAGCGCAATCTTCGGACATTTTAGTATGGAACTTGACTCCT   2100
AGACAACTATGTGATATTGAATTGATTCTAAATGGTGGGTTTTCTCCTCTGACTGGGTTT   2160
TTGAACGAAAACGATTACTCCTCTGTTGTTACAGATTCGAGATTAGCAGACGGCACATTG   2220
TGGACCATCCCTATTACATTAGATGTTGATGAAGCATTTGCTAACCAAATTAAACCAGAC   2280
ACAAGAATTGCCCTTTTCCAAGATGATGAAATTCCTATTGCTATACTTACTGTCCAGGAT   2340
GTTTACAAGCCAAACAAAACTATCGAAGCCGAAAAAGTCTTCAGAGGTGACCCAGAACAT   2400
CCAGCCATTAGCTATTTATTTAACGTTGCCGGTGATTATTACGTCGGCGGTTCTTTAGAA   2460
GCGATTCAATTACCTCAACATTATGACTATCCAGGTTTGCGTAAGACACCTGCCCAACTA   2520
AGACTTGAATTCCAATCAAGACAATGGGACCGTGTCGTAGCTTTCCAAACTCGTAATCCA   2580
ATGCATAGAGCCCACAGGGAGTTGACTGTGAGAGCCGCCAGAGAAGCTAATGCTAAGGTG   2640
CTGATCCATCCAGTTGTTGGACTAACCAAACCAGGTGATATAGACCATCACACTCGTGTT   2700
CGTGTCTACCAGGAAATTATTAAGCGTTATCCTAATGGTATTGCTTTCTTATCCCTGTTG   2760
CCATTAGCAATGAGAATGAGTGGTGATAGAGAAGCCGTATGGCATGCTATTATTAGAAAG   2820
AATTATGGTGCCTCCCACTTCATTGTTGGTAGAGACCATGCGGGCCCAGGTAAGAACTCC   2880
AAGGGTGTTGATTTCTACGGTCCATACGATGCTCAAGAATTGGTCGAATCCTACAAGCAT   2940
GAACTGGACATTGAAGTTGTTGCATTCAGAATGGTCACTTATTTGCCAGACGAAGACCGT   3000
TATGCTCCAATTGATCAAATTGACACCACAAAGACGAGAACCTTGAACATTTCAGGTACA   3060
GAGTTGAGACGCCGTTTAAGAGTTGGTGGTGAGATTCCTGAATGGTTCTCATATCCTGAA   3120
GTGGTTAAAATCCTAAGAGAATCCAACCCACCAAGACCAAAACAAGGTTTTTCAATTGTT   3180
```

-continued

```
TTAGGTAATTCATTAACCGTTTCTCGTGAGCAATTATCCATTGCTTTGTTGTCAACATTC 3240

TTGCAATTCGGTGGTGGCAGGTATTACAAGATCTTTGAACACAATAATAAGACAGAGTTA 3300

CTATCTTTGATTCAAGATTTCATTGGTTCTGGTAGTGGACTAATTATTCCAAATCAATGG 3360

GAAGATGACAAGGACTCTGTTGTTGGCAAGCAAAACGTTTACTTATTAGATACCTCAAGC 3420

TCAGCCGATATTCAGCTAGAGTCAGCGGATGAACCTATTTCACATATTGTACAAAAAGTT 3480

GTCCTATTCTTGGAAGACAATGGCTTTTTTGTATTTTAA                      3519
```

[0064] The amino acid sequence of the disclosed His6-BCCP L-S polypeptide is presented using the three letter amino acid code (SEQ ID NO:4).

```
His6-BCCP L-S Amino Acid Sequence (SEQ ID NO: 4)
Met Arg Gly Ser His His His His His His Gly Met
  1               5                  10

Ala Ser Met Glu Ala Pro Ala Ala Ala Glu Ile Ser
         15                  20

Gly His Ile Val Arg Ser Pro Met Val Gly Thr Phe
 25                  30                  35

Tyr Arg Thr Pro Ser Pro Asp Ala Lys Ala Phe Ile
         40                  45

Glu Val Gly Gln Lys Val Asn Val Gly Asp Thr Leu
     50                  55                  60

Cys Ile Val Glu Ala Met Lys Met Met Asn Gln Ile
                 65                  70

Glu Ala Asp Lys Ser Gly Thr Val Lys Ala Ile Leu
         75                  80

Val Glu Ser Gly Gln Pro Val Glu Phe Asp Glu Pro
 85                  90                  95

Leu Val Val Ile Glu Gly Ser Glu Leu Glu Ile Gln
             100                 105

Met Glu Asp Ala Lys Asn Ile Lys Lys Gly Pro Ala
    110                 115                 120

Pro Phe Tyr Pro Leu Glu Asp Gly Thr Ala Gly Glu
                125                 130

Gln Leu His Lys Ala Met Lys Arg Tyr Ala Leu Val
        135                 140

Pro Gly Thr Ile Ala Phe Thr Asp Ala His Ile Glu
145                 150                 155

Val Asn Ile Thr Tyr Ala Glu Tyr Phe Glu Met Ser
            160                 165

Val Arg Leu Ala Glu Ala Met Lys Arg Tyr Gly Leu
    170                 175                 180

Asn Thr Asn His Arg Ile Val Val Cys Ser Glu Asn
                185                 190

Ser Leu Gln Phe Phe Met Pro Val Leu Gly Ala Leu
        195                 200

Phe Ile Gly Val Ala Val Ala Pro Ala Asn Asp Ile
205                 210                 215

Tyr Asn Glu Arg Glu Leu Leu Asn Ser Met Asn Ile
            220                 225
```

-continued

```
Ser Gln Pro Thr Val Val Phe Val Ser Lys Lys Gly
    230                 235                 240

Leu Gln Lys Ile Leu Asn Val Gln Lys Lys Leu Pro
                245                 250

Ile Ile Gln Lys Ile Ile Ile Met Asp Ser Lys Thr
        255                 260

Asp Tyr Gln Gly Phe Gln Ser Met Tyr Thr Phe Val
265                 270                 275

Thr Ser His Leu Pro Pro Gly Phe Asn Glu Tyr Asp
            280                 285

Phe Val Pro Glu Ser Phe Asp Arg Asp Lys Thr Ile
    290                 295                 300

Ala Leu Ile Met Asn Ser Ser Gly Ser Thr Gly Leu
                305                 310

Pro Lys Gly Val Ala Leu Pro His Arg Thr Ala Cys
        315                 320

Val Arg Phe Ser His Ala Arg Asp Pro Ile Phe Gly
325                 330                 335

Asn Gln Ile Ile Pro Asp Thr Ala Ile Leu Ser Val
            340                 345

Val Pro Phe His His Gly Phe Gly Met Phe Thr Thr
    350                 355                 360

Leu Gly Tyr Leu Ile Cys Gly Phe Arg Val Val Leu
                365                 370

Met Tyr Arg Phe Glu Glu Glu Leu Phe Leu Arg Ser
        375                 380

Leu Gln Asp Tyr Lys Ile Gln Ser Ala Leu Leu Val
385                 390                 395

Pro Thr Leu Phe Ser Phe Phe Ala Lys Ser Thr Leu
            400                 405

Ile Asp Lys Tyr Asp Leu Ser Asn Leu His Glu Ile
    410                 415                 420

Ala Ser Gly Gly Ala Pro Leu Ser Lys Glu Val Gly
                425                 430

Glu Ala Val Ala Lys Arg Phe His Leu Pro Gly Ile
        435                 440

Arg Gln Gly Tyr Gly Leu Thr Glu Thr Thr Ser Ala
445                 450                 455

Ile Leu Ile Thr Pro Glu Gly Asp Asp Lys Pro Gly
            460                 465
```

-continued

```
Ala Val Gly Lys Val Val Pro Phe Phe Glu Ala Lys
    470                 475                 480

Val Val Asp Leu Asp Thr Gly Lys Thr Leu Gly Val
                485                 490

Asn Gln Arg Gly Glu Leu Cys Val Arg Gly Pro Met
        495                 500

Ile Met Ser Gly Tyr Val Asn Asn Pro Glu Ala Thr
505                 510                 515

Asn Ala Leu Ile Asp Lys Asp Gly Trp Leu His Ser
            520                 525

Gly Asp Ile Ala Tyr Trp Asp Glu Asp Glu His Phe
    530                 535                 540

Phe Ile Val Asp Arg Leu Lys Ser Leu Ile Lys Tyr
                545                 550

Lys Gly Tyr Gln Val Ala Pro Ala Glu Leu Glu Ser
        555                 560

Ile Leu Leu Gln His Pro Asn Ile Phe Asp Ala Gly
565                 570                 575

Val Ala Gly Leu Pro Asp Asp Asp Ala Gly Glu Leu
            580                 585

Pro Ala Ala Val Val Val Leu Glu His Gly Lys Thr
    590                 595                 600

Met Thr Glu Lys Glu Ile Val Asp Tyr Val Ala Ser
                605                 610

Gln Val Thr Thr Ala Lys Lys Leu Arg Gly Gly Val
        615                 620

Val Phe Val Asp Glu Val Pro Lys Gly Leu Thr Gly
625                 630                 635

Lys Leu Asp Ala Arg Lys Ile Arg Glu Ile Leu Ile
            640                 645

Lys Ala Lys Lys Gly Gly Lys Ser Lys Leu Ala Ala
    650                 655                 660

Ala Met Pro Ala Pro His Gly Gly Ile Leu Gln Asp
                665                 670

Leu Ile Ala Arg Asp Ala Leu Lys Lys Asn Glu Leu
            675                 680

Leu Ser Glu Ala Gln Ser Ser Asp Ile Leu Val Trp
    685                 690                 695

Asn Leu Thr Pro Arg Gln Leu Cys Asp Ile Glu Leu
                700                 705

Ile Leu Asn Gly Gly Phe Ser Pro Leu Thr Gly Phe
        710                 715

Leu Asn Glu Asn Asp Tyr Ser Ser Val Val Thr Asp
720                 725                 730

Ser Arg Leu Ala Asp Gly Thr Leu Trp Thr Ile Pro
            735                 740

Ile Thr Leu Asp Val Asp Glu Ala Phe Ala Asn Gln
    745                 750                 755

Ile Lys Pro Asp Thr Arg Ile Ala Leu Phe Gln Asp
                760                 765

Asp Glu Ile Pro Ile Ala Ile Leu Thr Val Gln Asp
        770                 775
```

-continued

```
Val Tyr Lys Pro Asn Lys Thr Ile Glu Ala Glu Lys
780                 785                 790

Val Phe Arg Gly Asp Pro Glu His Pro Ala Ile Ser
            795                 800

Tyr Leu Phe Asn Val Ala Gly Asp Tyr Tyr Val Gly
    805                 810                 815

Gly Ser Leu Glu Ala Ile Gln Leu Pro Gln His Tyr
                820                 825

Asp Tyr Pro Gly Leu Arg Lys Thr Pro Ala Gln Leu
        830                 835

Arg Leu Glu Phe Gln Ser Arg Gln Trp Asp Arg Val
840                 845                 850

Val Ala Phe Gln Thr Arg Asn Pro Met His Arg Ala
            855                 860

His Arg Glu Leu Thr Val Arg Ala Ala Arg Glu Ala
    865                 870                 875

Asn Ala Lys Val Leu Ile His Pro Val Val Gly Leu
                880                 885

Thr Lys Pro Gly Asp Ile Asp His His Thr Arg Val
        890                 895

Arg Val Tyr Gln Glu Ile Ile Lys Arg Tyr Pro Asn
900                 905                 910

Gly Ile Ala Phe Leu Ser Leu Leu Pro Leu Ala Met
            915                 920

Arg Met Ser Gly Asp Arg Glu Ala Val Trp His Ala
    925                 930                 935

Ile Ile Arg Lys Asn Tyr Gly Ala Ser His Phe Ile
                940                 945

Val Gly Arg Asp His Ala Gly Pro Gly Lys Asn Ser
        950                 955

Lys Gly Val Asp Phe Tyr Gly Pro Tyr Asp Ala Gln
960                 965                 970

Glu Leu Val Glu Ser Tyr Lys His Glu Leu Asp Ile
            975                 980

Glu Val Val Pro Phe Arg Met Val Thr Tyr Leu Pro
    985                 990                 995

Asp Glu Asp Arg Tyr Ala Pro Ile Asp Gln Ile Asp
                1000                1005

Thr Thr Lys Thr Arg Thr Leu Asn Ile Ser Gly Thr
        1010                1015

Glu Leu Arg Arg Arg Leu Arg Val Gly Gly Glu Ile
1020                1025                1030

Pro Glu Trp Phe Ser Tyr Pro Glu Val Val Lys Ile
            1035                1040

Leu Arg Glu Ser Asn Pro Pro Arg Pro Lys Gln Gly
    1045                1050                1055

Phe Ser Ile Val Leu Gly Asn Ser Leu Thr Val Ser
                1060                1065

Arg Glu Gln Leu Ser Ile Ala Leu Leu Ser Thr Phe
        1070                1075

Leu Gln Phe Gly Gly Gly Arg Tyr Tyr Lys Ile Phe
```

-continued

```
1080                1085                1090

Glu His Asn Asn Lys Thr Glu Leu Leu Ser Leu Ile
            1095                1100

Gln Asp Phe Ile Gly Ser Gly Ser Gly Leu Ile Ile
    1105                1110                1115

Pro Asn Gln Trp Glu Asp Asp Lys Asp Ser Val Val
        1120                1125

Gly Lys Gln Asn Val Tyr Leu Leu Asp Thr Ser Ser
        1130                1135

Ser Ala Asp Ile Gln Leu Glu Ser Ala Asp Glu Pro
1140                1145                1150
```

-continued

```
Ile Ser His Ile Val Gln Lys Val Val Leu Phe Leu
            1155                1160

Glu Asp Asn Gly Phe Phe Val Phe
    1165                1170
```

[0065] Accordingly, in one aspect, the invention provides for a fusion protein comprising a thermostable sulfurylase joined to at least one affinity tag. The nucleic acid sequence of the disclosed N-terminal hexahistidine-BCCP Bst ATP Sulfurylase (His6-BCCP Bst Sulfurylase) gene is shown below:

```
His6-BCCP Bst Sulfurylase Nucleotide Sequence (SEQ ID NO: 5)
ATGCGGGGTTCTCATGATCATCATCATCATGGTATGGCTAGCATGGAAGGGCCAGCAGCA     60

GCGGAAATCAGTGGTCACATCGTACGTTCCCCGATGGTTGGTACTTTCTACCGCACCCCA    120

AGCCCGGACGCAAAAGCGTTCATCGAAGTGGGTCAGAAAGTCAACGTGGGCGATACCCTG    180

TGCATCGTTGAAGCCATGAAAATGATGAACCAGATCGAAGCGGACAAATCCGGTACCGTG    240

AAAGCAATTCTGGTCGAAAGTGGACAACCGGTAGAATTTGACGAGCCGCTGGTCGTCATC    300

GAGGGATCCGAGCTCGAGATCTGCAGCATGAGCGTAAGCATCCCGCATGGCGGCACATTG    360

ATCAACCGTTGGAATCCGGATTACCCAATCGATGAAGCAACGAAAACGATCGAGCTGTCC    420

AAAGCCGAACTAAGCGACCTTGAGCTGATCGGCACAGGCGCCTACAGCCCGCTCACCGGG    480

TTTTTAACGAAAGCCGATTACGATGCGGTCGTAGAAACGATGCGCCTCGCTGATGGCACT    540

GTCTGGAGCATTCCGATCACGCTGGCGGTGACGGAAGAAAAAGCGAGTGAACTCACTGTC    600

GGCGACAAAGCGAAACTCGTTTATGGCGGCGACGTCTACGGCGTCATTGAAATCGCCGAT    660

ATTTACCGCCCGGATAAAACGAAAGAAGCCAAGCTCGTCTATAAAACCGATGAACTCGCT    720

CACCCGGGCGTGCGCAAGCTGTTTGAAAAACCAGATGTGTACGTCGGCGGAGCGGTTACG    780

CTCGTCAAACGGACCGACAAAGGCCAGTTTGCTCCGTTTTATTTCGATCCGGCCGAAACG    840

CGGAAACGATTTGCCGAACTCGGCTGGAATACCGTCGTCGGCTTCCAAACACGCAACCCG    900

GTTCACCGCGCCCATGAATACATTCAAAAATGCGCGCTTGAAATCGTGGACGGCTTGTTT    960

TTAAACCCGCTCGTCGGCGAAACGAAAGCGGACGATATTCCGGCCGACATCCGGATGGAA   1020

AGCTATCAAGTGCTGCTGGAAAACTATTATCCGAAAGACCGCGTTTTCTTGGGCGTCTTC   1080

CAAGCTGCGATGCGCTATGCCGGTCCGCGCGAAGCGATTTTCCATGCCATGGTGCGGAAA   1140

AACTTCGGCTGCACGCACTTCATCGTCGGCCGCGACCATGCGGGCGTCGGCAACTATTAC   1200

GGCACGTATGATGCGCAAAAAATCTTCTCGAACTTTACAGCCGAAGAGCTTGGCATTACA   1260

CCGCTCTTTTTCGAACACAGCTTTTATTGCACGAAATGCGAAGGCATGGCATCGACGAAA   1320

ACATGCCCGCACGACGCACAATATCACGTTGTCCTTTCTGGCACGAAAGTCCGTGAAATG   1380

TTGCGTAACGGCCAAGTGCCGCCGAGCACATTCAGCCGTCCGGAAGTGGCCGCCGTTTTG   1440

ATCAAAGGGCTGCAAGAACGCGAAACGGTCGCCCCGTCAGCGGGCTAA                1488
```

[0066] The amino acid sequence of the His6-BCCP Bst Sulfurylase polypeptide is presented using the three letter amino acid code in Table 6 (SEQ ID NO:6).

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 31

<210> SEQ ID NO 1
<211> LENGTH: 1247
<212> TYPE: DNA
<213> ORGANISM: Bacillus stearothermophilus

<400> SEQUENCE: 1

gttatgaaca tgagtttgag cattccgcat ggcggcacat tgatcaaccg ttggaatccg 60

gattacccaa tcgatgaagc aacgaaaacg atcgagctgt ccaaagccga actaagcgac 120

cttgagctga tcggcacagg cgcctacagc ccgctcaccg ggtttttaac gaaagccgat 180

tacgatgcgg tcgtagaaac gatgcgcctc gctgatggca ctgtctggag cattccgatc 240

acgctggcgg tgacggaaga aaaagcgagt gaactcactg tcggcgacaa agcgaaactc 300

gtttatggcg gcgacgtcta cggcgtcatt gaaatcgccg atatttaccg cccggataaa 360

acgaaagaag ccaagctcgt ctataaaacc gatgaactcg ctcacccggg cgtgcgcaag 420

ctgtttgaaa aaccagatgt gtacgtcggc ggagcggtta cgctcgtcaa acggaccgac 480

aaaggccagt ttgctccgtt ttatttcgat ccggccgaaa cgcggaaacg atttgccgaa 540

ctcggctgga ataccgtcgt cggcttccaa acacgcaacc cggttcaccg cgcccatgaa 600

tacattcaaa aatgcgcgct tgaaatcgtg gacggcttgt ttttaaaccc gctcgtcggc 660

gaaacgaaag cggacgatat tccggccgac atccggatgg aaagctatca agtgctgctg 720

gaaaactatt atccgaaaga ccgcgttttc ttgggcgtct tccaagctgc gatgcgctat 780

gccggtccgc gcgaagcgat tttccatgcc atggtgcgga aaaacttcgg ctgcacgcac 840

ttcatcgtcg gccgcgacca tgcgggcgtc ggcaactatt acggcacgta tgatgcgcaa 900

aaaatcttct cgaactttac agccgaagag cttggcatta caccgctctt tttcgaacac 960

agcttttatt gcacgaaatg cgaaggcatg gcatcgacga aaacatgccc gcacgacgca 1020

caatatcacg ttgtcctttc tggcacgaaa gtccgtgaaa tgttgcgtaa cggccaagtg 1080

ccgccgagca cattcagccg tccggaagtg gccgccgttt tgatcaaagg gctgcaagaa 1140

cgcgaaacgg tcaccccgtc gacacgctaa aggaggagcg agatgagcac gaatatcgtt 1200

tggcatcata catcggtgac aaaagaagat cgccgccaac gcaacgg 1247

<210> SEQ ID NO 2
<211> LENGTH: 386
<212> TYPE: PRT
<213> ORGANISM: Bacillus stearothermophilus

<400> SEQUENCE: 2

Met Ser Leu Ser Ile Pro His Gly Gly Thr Leu Ile Asn Arg Trp Asn
1 5 10 15

Pro Asp Tyr Pro Ile Asp Glu Ala Thr Lys Thr Ile Glu Leu Ser Lys
20 25 30

Ala Glu Leu Ser Asp Leu Glu Leu Ile Gly Thr Gly Ala Tyr Ser Pro
35 40 45

-continued

```
Leu Thr Gly Phe Leu Thr Lys Ala Asp Tyr Asp Ala Val Val Glu Thr
     50                  55                  60

Met Arg Leu Ala Asp Gly Thr Val Trp Ser Ile Pro Ile Thr Leu Ala
 65                  70                  75                  80

Val Thr Glu Glu Lys Ala Ser Glu Leu Thr Val Gly Asp Lys Ala Lys
                 85                  90                  95

Leu Val Tyr Gly Gly Asp Val Tyr Gly Val Ile Glu Ile Ala Asp Ile
            100                 105                 110

Tyr Arg Pro Asp Lys Thr Lys Glu Ala Lys Leu Val Tyr Lys Thr Asp
        115                 120                 125

Glu Leu Ala His Pro Gly Val Arg Lys Leu Phe Glu Lys Pro Asp Val
    130                 135                 140

Tyr Val Gly Gly Ala Val Thr Leu Val Lys Arg Thr Asp Lys Gly Gln
145                 150                 155                 160

Phe Ala Pro Phe Tyr Phe Asp Pro Ala Glu Thr Arg Lys Arg Phe Ala
                165                 170                 175

Glu Leu Gly Trp Asn Thr Val Val Gly Phe Gln Thr Arg Asn Pro Val
            180                 185                 190

His Arg Ala His Glu Tyr Ile Gln Lys Cys Ala Leu Glu Ile Val Asp
        195                 200                 205

Gly Leu Phe Leu Asn Pro Leu Val Gly Glu Thr Lys Ala Asp Asp Ile
    210                 215                 220

Pro Ala Asp Ile Arg Met Glu Ser Tyr Gln Val Leu Leu Glu Asn Tyr
225                 230                 235                 240

Tyr Pro Lys Asp Arg Val Phe Leu Gly Val Phe Gln Ala Ala Met Arg
                245                 250                 255

Tyr Ala Gly Pro Arg Glu Ala Ile Phe His Ala Met Val Arg Lys Asn
            260                 265                 270

Phe Gly Cys Thr His Phe Ile Val Gly Arg Asp His Ala Gly Val Gly
        275                 280                 285

Asn Tyr Tyr Gly Thr Tyr Asp Ala Gln Lys Ile Phe Ser Asn Phe Thr
    290                 295                 300

Ala Glu Glu Leu Gly Ile Thr Pro Leu Phe Phe Glu His Ser Phe Tyr
305                 310                 315                 320

Cys Thr Lys Cys Glu Gly Met Ala Ser Thr Lys Thr Cys Pro His Asp
                325                 330                 335

Ala Gln Tyr His Val Val Leu Ser Gly Thr Lys Val Arg Glu Met Leu
            340                 345                 350

Arg Asn Gly Gln Val Pro Pro Ser Thr Phe Ser Arg Pro Glu Val Ala
        355                 360                 365

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

Thr Arg
385
```

<210> SEQ ID NO 3
<211> LENGTH: 3519
<212> TYPE: DNA
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 3

```
atgcggggtt ctcatcatca tcatcatcat ggtatggcta gcatggaagc gccagcagca     60
```

-continued

```
gcggaaatca gtggtcacat cgtacgttcc ccgatggttg gtactttcta ccgcacccca    120
agcccggacg caaaagcgtt catcgaagtg ggtcagaaag tcaacgtggg cgataccctg    180
tgcatcgttg aagccatgaa aatgatgaac cagatcgaag cggacaaatc cggtaccgtg    240
aaagcaattc tggtcgaaag tggacaaccg gtagaatttg acgagccgct ggtcgtcatc    300
gagggatccg agctcgagat ccaaatggaa gacgccaaaa acataaagaa aggcccggcg    360
ccattctatc ctctagagga tggaaccgct ggagagcaac tgcataaggc tatgaagaga    420
tacgccctgg ttcctggaac aattgctttt acagatgcac atatcgaggt gaacatcacg    480
tacgcggaat acttcgaaat gtccgttcgg ttggcagaag ctatgaaacg atatgggctg    540
aatacaaatc acagaatcgt cgtatgcagt gaaaactctc ttcaattctt tatgccggtg    600
ttgggcgcgt tatttatcgg agttgcagtt gcgcccgcga acgacattta taatgaacgt    660
gaattgctca acagtatgaa catttcgcag cctaccgtag tgtttgtttc caaaaagggg    720
ttgcaaaaaa ttttgaacgt gcaaaaaaaa ttaccaataa tccagaaaat tattatcatg    780
gattctaaaa cggattacca gggatttcag tcgatgtaca cgttcgtcac atctcatcta    840
cctcccggtt ttaatgaata cgattttgta ccagagtcct ttgatcgtga caaaacaatt    900
gcactgataa tgaattcctc tggatctact gggttaccta agggtgtggc ccttccgcat    960
agaactgcct gcgtcagatt ctcgcatgcc agagatccta tttttggcaa tcaaatcatt   1020
ccggatactg cgattttaag tgttgttcca ttccatcacg gttttggaat gtttactaca   1080
ctcggatatt tgatatgtgg atttcgagtc gtcttaatgt atagatttga agaagagctg   1140
tttttacgat cccttcagga ttacaaaatt caaagtgcgt tgctagtacc aaccctattt   1200
tcattcttcg ccaaaagcac tctgattgac aaatacgatt tatctaattt acacgaaatt   1260
gcttctgggg gcgcacctct ttcgaaagaa gtcggggaag cggttgcaaa acgcttccat   1320
cttccaggga tacgacaagg atatgggctc actgagacta catcagctat tctgattaca   1380
cccgaggggg atgataaacc gggcgcggtc ggtaaagttg ttccattttt tgaagcgaag   1440
gttgtggatc tggataccgg gaaaacgctg ggcgttaatc agagaggcga attatgtgtc   1500
agaggaccta tgattatgtc cggttatgta aacaatccgg aagcgaccaa cgccttgatt   1560
gacaaggatg gatggctaca ttctggagac atagcttact gggacgaaga cgaacacttc   1620
ttcatagttg accgcttgaa gtctttaatt aaatacaaag gatatcaggt ggcccccgct   1680
gaattggaat cgatattgtt acaacacccc aacatcttcg acgcgggcgt ggcaggtctt   1740
cccgacgatg acgccggtga acttcccgcc gccgttgttg ttttggagca cggaaagacg   1800
atgacggaaa aagagatcgt ggattacgtc gccagtcaag taacaaccgc gaaaaagttg   1860
cgcggaggag ttgtgtttgt ggacgaagta ccgaaaggtc ttaccggaaa actcgacgca   1920
agaaaaatca gagagatcct cataaaggcc aagaagggcg gaaagtccaa attggcggcc   1980
gctatgcctg ctcctcacgg tggtattcta caagacttga ttgctagaga tgcgttaaag   2040
aagaatgaat tgttatctga agcgcaatct tcggacattt tagtatggaa cttgactcct   2100
agacaactat gtgatattga attgattcta aatggtgggt tttctcctct gactgggttt   2160
ttgaacgaaa acgattactc ctctgttgtt acagattcga gattagcaga cggcacattg   2220
tggaccatcc ctattacatt agatgttgat gaagcatttg ctaaccaaat taaaccagac   2280
acaagaattg cccttttcca agatgatgaa attcctattg ctatacttac tgtccaggat   2340
```

-continued

```
gtttacaagc caaacaaaac tatcgaagcc gaaaaagtct tcagaggtga cccagaacat   2400

ccagccatta gctatttatt taacgttgcc ggtgattatt acgtcggcgg ttctttagaa   2460

gcgattcaat tacctcaaca ttatgactat ccaggtttgc gtaagacacc tgcccaacta   2520

agacttgaat tccaatcaag acaatgggac cgtgtcgtag ctttccaaac tcgtaatcca   2580

atgcatagag cccacaggga gttgactgtg agagccgcca gagaagctaa tgctaaggtg   2640

ctgatccatc cagttgttgg actaaccaaa ccaggtgata tagaccatca cactcgtgtt   2700

cgtgtctacc aggaaattat taagcgttat cctaatggta ttgctttctt atccctgttg   2760

ccattagcaa tgagaatgag tggtgataga gaagccgtat ggcatgctat tattagaaag   2820

aattatggtg cctcccactt cattgttggt agagaccatg cgggcccagg taagaactcc   2880

aagggtgttg atttctacgg tccatacgat gctcaagaat tggtcgaatc ctacaagcat   2940

gaactggaca ttgaagttgt tccattcaga atggtcactt atttgccaga cgaagaccgt   3000

tatgctccaa ttgatcaaat tgacaccaca aagacgagaa ccttgaacat ttcaggtaca   3060

gagttgagac gccgtttaag agttggtggt gagattcctg aatggttctc atatcctgaa   3120

gtggttaaaa tcctaagaga atccaaccca ccaagaccaa aacaaggttt ttcaattgtt   3180

ttaggtaatt cattaaccgt ttctcgtgag caattatcca ttgctttgtt gtcaacattc   3240

ttgcaattcg gtggtggcag gtattacaag atctttgaac acaataataa gacagagtta   3300

ctatctttga ttcaagattt cattggttct ggtagtggac taattattcc aaatcaatgg   3360

gaagatgaca aggactctgt tgttggcaag caaaacgttt acttattaga tacctcaagc   3420

tcagccgata ttcagctaga gtcagcggat gaacctattt cacatattgt acaaaaagtt   3480

gtcctattct tggaagacaa tggctttttt gtattttaa                          3519
```

<210> SEQ ID NO 4
<211> LENGTH: 1172
<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 4

```
Met Arg Gly Ser His His His His His His Gly Met Ala Ser Met Glu
  1               5                  10                  15

Ala Pro Ala Ala Ala Glu Ile Ser Gly His Ile Val Arg Ser Pro Met
             20                  25                  30

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

Glu Val Gly Gln Lys Val Asn Val Gly Asp Thr Leu Cys Ile Val Glu
     50                  55                  60

Ala Met Lys Met Met Asn Gln Ile Glu Ala Asp Lys Ser Gly Thr Val
 65                  70                  75                  80

Lys Ala Ile Leu Val Glu Ser Gly Gln Pro Val Glu Phe Asp Glu Pro
                 85                  90                  95

Leu Val Val Ile Glu Gly Ser Glu Leu Glu Ile Gln Met Glu Asp Ala
            100                 105                 110

Lys Asn Ile Lys Lys Gly Pro Ala Pro Phe Tyr Pro Leu Glu Asp Gly
        115                 120                 125

Thr Ala Gly Glu Gln Leu His Lys Ala Met Lys Arg Tyr Ala Leu Val
    130                 135                 140

Pro Gly Thr Ile Ala Phe Thr Asp Ala His Ile Glu Val Asn Ile Thr
```

-continued

```
145                 150                 155                 160

Tyr Ala Glu Tyr Phe Glu Met Ser Val Arg Leu Ala Glu Ala Met Lys
                165                 170                 175

Arg Tyr Gly Leu Asn Thr Asn His Arg Ile Val Val Cys Ser Glu Asn
            180                 185                 190

Ser Leu Gln Phe Phe Met Pro Val Leu Gly Ala Leu Phe Ile Gly Val
        195                 200                 205

Ala Val Ala Pro Ala Asn Asp Ile Tyr Asn Glu Arg Glu Leu Leu Asn
    210                 215                 220

Ser Met Asn Ile Ser Gln Pro Thr Val Val Phe Val Ser Lys Lys Gly
225                 230                 235                 240

Leu Gln Lys Ile Leu Asn Val Gln Lys Lys Leu Pro Ile Ile Gln Lys
                245                 250                 255

Ile Ile Ile Met Asp Ser Lys Thr Asp Tyr Gln Gly Phe Gln Ser Met
            260                 265                 270

Tyr Thr Phe Val Thr Ser His Leu Pro Pro Gly Phe Asn Glu Tyr Asp
        275                 280                 285

Phe Val Pro Glu Ser Phe Asp Arg Asp Lys Thr Ile Ala Leu Ile Met
    290                 295                 300

Asn Ser Ser Gly Ser Thr Gly Leu Pro Lys Gly Val Ala Leu Pro His
305                 310                 315                 320

Arg Thr Ala Cys Val Arg Phe Ser His Ala Arg Asp Pro Ile Phe Gly
                325                 330                 335

Asn Gln Ile Ile Pro Asp Thr Ala Ile Leu Ser Val Val Pro Phe His
            340                 345                 350

His Gly Phe Gly Met Phe Thr Thr Leu Gly Tyr Leu Ile Cys Gly Phe
        355                 360                 365

Arg Val Val Leu Met Tyr Arg Phe Glu Glu Glu Leu Phe Leu Arg Ser
    370                 375                 380

Leu Gln Asp Tyr Lys Ile Gln Ser Ala Leu Leu Val Pro Thr Leu Phe
385                 390                 395                 400

Ser Phe Phe Ala Lys Ser Thr Leu Ile Asp Lys Tyr Asp Leu Ser Asn
                405                 410                 415

Leu His Glu Ile Ala Ser Gly Gly Ala Pro Leu Ser Lys Glu Val Gly
            420                 425                 430

Glu Ala Val Ala Lys Arg Phe His Leu Pro Gly Ile Arg Gln Gly Tyr
        435                 440                 445

Gly Leu Thr Glu Thr Thr Ser Ala Ile Leu Ile Thr Pro Glu Gly Asp
    450                 455                 460

Asp Lys Pro Gly Ala Val Gly Lys Val Val Pro Phe Phe Glu Ala Lys
465                 470                 475                 480

Val Val Asp Leu Asp Thr Gly Lys Thr Leu Gly Val Asn Gln Arg Gly
                485                 490                 495

Glu Leu Cys Val Arg Gly Pro Met Ile Met Ser Gly Tyr Val Asn Asn
            500                 505                 510

Pro Glu Ala Thr Asn Ala Leu Ile Asp Lys Asp Gly Trp Leu His Ser
        515                 520                 525

Gly Asp Ile Ala Tyr Trp Asp Glu Asp Glu His Phe Phe Ile Val Asp
    530                 535                 540

Arg Leu Lys Ser Leu Ile Lys Tyr Lys Gly Tyr Gln Val Ala Pro Ala
545                 550                 555                 560
```

-continued

```
Glu Leu Glu Ser Ile Leu Leu Gln His Pro Asn Ile Phe Asp Ala Gly
                565                 570                 575

Val Ala Gly Leu Pro Asp Asp Asp Ala Gly Glu Leu Pro Ala Ala Val
            580                 585                 590

Val Val Leu Glu His Gly Lys Thr Met Thr Glu Lys Glu Ile Val Asp
        595                 600                 605

Tyr Val Ala Ser Gln Val Thr Thr Ala Lys Lys Leu Arg Gly Gly Val
    610                 615                 620

Val Phe Val Asp Glu Val Pro Lys Gly Leu Thr Gly Lys Leu Asp Ala
625                 630                 635                 640

Arg Lys Ile Arg Glu Ile Leu Ile Lys Ala Lys Lys Gly Gly Lys Ser
                645                 650                 655

Lys Leu Ala Ala Ala Met Pro Ala Pro His Gly Gly Ile Leu Gln Asp
            660                 665                 670

Leu Ile Ala Arg Asp Ala Leu Lys Lys Asn Glu Leu Leu Ser Glu Ala
        675                 680                 685

Gln Ser Ser Asp Ile Leu Val Trp Asn Leu Thr Pro Arg Gln Leu Cys
    690                 695                 700

Asp Ile Glu Leu Ile Leu Asn Gly Gly Phe Ser Pro Leu Thr Gly Phe
705                 710                 715                 720

Leu Asn Glu Asn Asp Tyr Ser Ser Val Val Thr Asp Ser Arg Leu Ala
                725                 730                 735

Asp Gly Thr Leu Trp Thr Ile Pro Ile Thr Leu Asp Val Asp Glu Ala
            740                 745                 750

Phe Ala Asn Gln Ile Lys Pro Asp Thr Arg Ile Ala Leu Phe Gln Asp
        755                 760                 765

Asp Glu Ile Pro Ile Ala Ile Leu Thr Val Gln Asp Val Tyr Lys Pro
    770                 775                 780

Asn Lys Thr Ile Glu Ala Glu Lys Val Phe Arg Gly Asp Pro Glu His
785                 790                 795                 800

Pro Ala Ile Ser Tyr Leu Phe Asn Val Ala Gly Asp Tyr Tyr Val Gly
                805                 810                 815

Gly Ser Leu Glu Ala Ile Gln Leu Pro Gln His Tyr Asp Tyr Pro Gly
            820                 825                 830

Leu Arg Lys Thr Pro Ala Gln Leu Arg Leu Glu Phe Gln Ser Arg Gln
        835                 840                 845

Trp Asp Arg Val Val Ala Phe Gln Thr Arg Asn Pro Met His Arg Ala
    850                 855                 860

His Arg Glu Leu Thr Val Arg Ala Ala Arg Glu Ala Asn Ala Lys Val
865                 870                 875                 880

Leu Ile His Pro Val Val Gly Leu Thr Lys Pro Gly Asp Ile Asp His
                885                 890                 895

His Thr Arg Val Arg Val Tyr Gln Glu Ile Ile Lys Arg Tyr Pro Asn
            900                 905                 910

Gly Ile Ala Phe Leu Ser Leu Leu Pro Leu Ala Met Arg Met Ser Gly
        915                 920                 925

Asp Arg Glu Ala Val Trp His Ala Ile Ile Arg Lys Asn Tyr Gly Ala
    930                 935                 940

Ser His Phe Ile Val Gly Arg Asp His Ala Gly Pro Gly Lys Asn Ser
945                 950                 955                 960
```

-continued

```
Lys Gly Val Asp Phe Tyr Gly Pro Tyr Asp Ala Gln Glu Leu Val Glu
                965                 970                 975

Ser Tyr Lys His Glu Leu Asp Ile Glu Val Val Pro Phe Arg Met Val
            980                 985                 990

Thr Tyr Leu Pro Asp Glu Asp Arg Tyr Ala Pro Ile Asp Gln Ile Asp
        995                 1000                1005

Thr Thr Lys Thr Arg Thr Leu Asn Ile Ser Gly Thr Glu Leu Arg Arg
    1010                1015                1020

Arg Leu Arg Val Gly Gly Glu Ile Pro Glu Trp Phe Ser Tyr Pro Glu
1025                1030                1035                1040

Val Val Lys Ile Leu Arg Glu Ser Asn Pro Pro Arg Pro Lys Gln Gly
                1045                1050                1055

Phe Ser Ile Val Leu Gly Asn Ser Leu Thr Val Ser Arg Glu Gln Leu
            1060                1065                1070

Ser Ile Ala Leu Leu Ser Thr Phe Leu Gln Phe Gly Gly Gly Arg Tyr
        1075                1080                1085

Tyr Lys Ile Phe Glu His Asn Asn Lys Thr Glu Leu Leu Ser Leu Ile
    1090                1095                1100

Gln Asp Phe Ile Gly Ser Gly Ser Gly Leu Ile Ile Pro Asn Gln Trp
1105                1110                1115                1120

Glu Asp Asp Lys Asp Ser Val Val Gly Lys Gln Asn Val Tyr Leu Leu
                1125                1130                1135

Asp Thr Ser Ser Ser Ala Asp Ile Gln Leu Glu Ser Ala Asp Glu Pro
            1140                1145                1150

Ile Ser His Ile Val Gln Lys Val Val Leu Phe Leu Glu Asp Asn Gly
        1155                1160                1165

Phe Phe Val Phe
    1170
```

<210> SEQ ID NO 5
<211> LENGTH: 1488
<212> TYPE: DNA
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 5

```
atgcggggtt ctcatcatca tcatcatcat ggtatggcta gcatggaagc gccagcagca      60

gcggaaatca gtggtcacat cgtacgttcc ccgatggttg gtactttcta ccgcacccca     120

agcccggacg caaaagcgtt catcgaagtg ggtcagaaag tcaacgtggg cgataccctg     180

tgcatcgttg aagccatgaa aatgatgaac cagatcgaag cggacaaatc cggtaccgtg     240

aaagcaattc tggtcgaaag tggacaaccg gtagaatttg acgagccgct ggtcgtcatc     300

gagggatccg agctcgagat ctgcagcatg agcgtaagca tcccgcatgg cggcacattg     360

atcaaccgtt ggaatccgga ttacccaatc gatgaagcaa cgaaaacgat cgagctgtcc     420

aaagccgaac taagcgacct tgagctgatc ggcacaggcg cctacagccc gctcaccggg     480

tttttaacga aagccgatta cgatgcggtc gtagaaacga tgcgcctcgc tgatggcact     540

gtctggagca ttccgatcac gctggcggtg acggaagaaa aagcgagtga actcactgtc     600

ggcgacaaag cgaaactcgt ttatggcggc gacgtctacg gcgtcattga aatcgccgat     660

atttaccgcc cggataaaac gaaagaagcc aagctcgtct ataaaaccga tgaactcgct     720

cacccgggcg tgcgcaagct gtttgaaaaa ccagatgtgt acgtcggcgg agcggttacg     780
```

-continued

```
ctcgtcaaac ggaccgacaa aggccagttt gctccgtttt atttcgatcc ggccgaaacg     840

cggaaacgat ttgccgaact cggctggaat accgtcgtcg gcttccaaac acgcaacccg     900

gttcaccgcg cccatgaata cattcaaaaa tgcgcgcttg aaatcgtgga cggcttgttt     960

ttaaacccgc tcgtcggcga aacgaaagcg gacgatattc cggccgacat ccggatggaa    1020

agctatcaag tgctgctgga aaactattat ccgaaagacc gcgttttctt gggcgtcttc    1080

caagctgcga tgcgctatgc cggtccgcgc gaagcgattt tccatgccat ggtgcggaaa    1140

aacttcggct gcacgcactt catcgtcggc cgcgaccatg cgggcgtcgg caactattac    1200

ggcacgtatg atgcgcaaaa aatcttctcg aactttacag ccgaagagct tggcattaca    1260

ccgctctttt tcgaacacag cttttattgc acgaaatgcg aaggcatggc atcgacgaaa    1320

acatgcccgc acgacgcaca atatcacgtt gtcctttctg gcacgaaagt ccgtgaaatg    1380

ttgcgtaacg gccaagtgcc gccgagcaca ttcagccgtc cggaagtggc cgccgttttg    1440

atcaaagggc tgcaagaacg cgaaacggtc gccccgtcag cgcgctaa                 1488
```

```
<210> SEQ ID NO 6
<211> LENGTH: 495
<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 6

Met Arg Gly Ser His His His His His His Gly Met Ala Ser Met Glu
  1               5                  10                  15

Ala Pro Ala Ala Ala Glu Ile Ser Gly His Ile Val Arg Ser Pro Met
             20                  25                  30

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

Glu Val Gly Gln Lys Val Asn Val Gly Asp Thr Leu Cys Ile Val Glu
     50                  55                  60

Ala Met Lys Met Met Asn Gln Ile Glu Ala Asp Lys Ser Gly Thr Val
 65                  70                  75                  80

Lys Ala Ile Leu Val Glu Ser Gly Gln Pro Val Glu Phe Asp Glu Pro
                 85                  90                  95

Leu Val Val Ile Glu Gly Ser Glu Leu Glu Ile Cys Ser Met Ser Val
            100                 105                 110

Ser Ile Pro His Gly Gly Thr Leu Ile Asn Arg Trp Asn Pro Asp Tyr
        115                 120                 125

Pro Ile Asp Glu Ala Thr Lys Thr Ile Glu Leu Ser Lys Ala Glu Leu
    130                 135                 140

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

Phe Leu Thr Lys Ala Asp Tyr Asp Ala Val Val Glu Thr Met Arg Leu
                165                 170                 175

Ala Asp Gly Thr Val Trp Ser Ile Pro Ile Thr Leu Ala Val Thr Glu
            180                 185                 190

Glu Lys Ala Ser Glu Leu Thr Val Gly Asp Lys Ala Lys Leu Val Tyr
        195                 200                 205

Gly Gly Asp Val Tyr Gly Val Ile Glu Ile Ala Asp Ile Tyr Arg Pro
    210                 215                 220

Asp Lys Thr Lys Glu Ala Lys Leu Val Tyr Lys Thr Asp Glu Leu Ala
225                 230                 235                 240
```

-continued

```
His Pro Gly Val Arg Lys Leu Phe Glu Lys Pro Asp Val Tyr Val Gly
                245                 250                 255

Gly Ala Val Thr Leu Val Lys Arg Thr Asp Lys Gly Gln Phe Ala Pro
            260                 265                 270

Phe Tyr Phe Asp Pro Ala Glu Thr Arg Lys Arg Phe Ala Glu Leu Gly
        275                 280                 285

Trp Asn Thr Val Val Gly Phe Gln Thr Arg Asn Pro Val His Arg Ala
    290                 295                 300

His Glu Tyr Ile Gln Lys Cys Ala Leu Glu Ile Val Asp Gly Leu Phe
305                 310                 315                 320

Leu Asn Pro Leu Val Gly Glu Thr Lys Ala Asp Asp Ile Pro Ala Asp
                325                 330                 335

Ile Arg Met Glu Ser Tyr Gln Val Leu Leu Glu Asn Tyr Tyr Pro Lys
            340                 345                 350

Asp Arg Val Phe Leu Gly Val Phe Gln Ala Ala Met Arg Tyr Ala Gly
        355                 360                 365

Pro Arg Glu Ala Ile Phe His Ala Met Val Arg Lys Asn Phe Gly Cys
    370                 375                 380

Thr His Phe Ile Val Gly Arg Asp His Ala Gly Val Gly Asn Tyr Tyr
385                 390                 395                 400

Gly Thr Tyr Asp Ala Gln Lys Ile Phe Ser Asn Phe Thr Ala Glu Glu
                405                 410                 415

Leu Gly Ile Thr Pro Leu Phe Phe Glu His Ser Phe Tyr Cys Thr Lys
            420                 425                 430

Cys Glu Gly Met Ala Ser Thr Lys Thr Cys Pro His Asp Ala Gln Tyr
        435                 440                 445

His Val Val Leu Ser Gly Thr Lys Val Arg Glu Met Leu Arg Asn Gly
    450                 455                 460

Gln Val Pro Pro Ser Thr Phe Ser Arg Pro Glu Val Ala Ala Val Leu
465                 470                 475                 480

Ile Lys Gly Leu Gln Glu Arg Glu Thr Val Ala Pro Ser Ala Arg
                485                 490                 495
```

<210> SEQ ID NO 7
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: primer

<400> SEQUENCE: 7

ccttctgca gcatgagcgt aagcatcccg catggcggca cattg 45

<210> SEQ ID NO 8
<211> LENGTH: 47
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: primer

<400> SEQUENCE: 8

cccgtaagct tttagcgcgc tgacggggcg accgtttcgc gttcttg 47

<210> SEQ ID NO 9
<211> LENGTH: 48

-continued

<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: primer

<400> SEQUENCE: 9

ccccctcgag atccaaatgg aagacgccaa aaacataaag aaaggccc 48

<210> SEQ ID NO 10
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: primer

<400> SEQUENCE: 10

ccccctcgag atccaaatgg ctgacaaaaa catcctgtat ggccc 45

<210> SEQ ID NO 11
<211> LENGTH: 67
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: primer

<400> SEQUENCE: 11

ttgtagaata ccaccgtgag gagcaggcat agcggccgcc aatttggact ttccgccctt 60

cttggcc 67

<210> SEQ ID NO 12
<211> LENGTH: 64
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: primer

<400> SEQUENCE: 12

ttgtagaata ccaccgtgag gagcaggcat agcggccgca ccgttggtgt gtttctcgaa 60

catc 64

<210> SEQ ID NO 13
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: primer

<400> SEQUENCE: 13

gcggccgcta tgcctgctcc tcacggtggt attctac 37

<210> SEQ ID NO 14
<211> LENGTH: 49
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: primer

<400> SEQUENCE: 14

ccccaagctt ttaaaataca aaaaagccat tgtcttccaa gaataggac 49

<210> SEQ ID NO 15
<211> LENGTH: 49

-continued

<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: primer

<400> SEQUENCE: 15

ccccggatcc atccaaatgc ctgctcctca cggtggtatt ctacaagac 49

<210> SEQ ID NO 16
<211> LENGTH: 62
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: primer

<400> SEQUENCE: 16

gggcctttct ttatgttttt ggcgtcttcc atagcggccg caaatacaaa aaagccattg 60

tc 62

<210> SEQ ID NO 17
<211> LENGTH: 41
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: primer

<400> SEQUENCE: 17

gcggccgcta tggaagacgc caaaaacata aagaaaggcc c 41

<210> SEQ ID NO 18
<211> LENGTH: 41
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: primer

<400> SEQUENCE: 18

ccccccatgg ttacaatttg gactttccgc ccttcttggc c 41

<210> SEQ ID NO 19
<211> LENGTH: 59
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: primer

<400> SEQUENCE: 19

gggccataca ggatgttttt gtcagccata gcggccgcaa atacaaaaaa gccattgtc 59

<210> SEQ ID NO 20
<211> LENGTH: 38
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: primer

<400> SEQUENCE: 20

gcggccgcta tggctgacaa aaacatcctg tatggccc 38

<210> SEQ ID NO 21
<211> LENGTH: 44
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: primer

<400> SEQUENCE: 21

ccccaagctt ctaaccgttg gtgtgtttct cgaacatctg acgc 44

<210> SEQ ID NO 22
<211> LENGTH: 386
<212> TYPE: PRT
<213> ORGANISM: Bacillus stearothermophilus

<400> SEQUENCE: 22

Met Ser Val Ser Ile Pro His Gly Gly Thr Leu Ile Asn Arg Trp Asn
1 5 10 15

Pro Asp Tyr Pro Leu Asp Glu Ala Thr Lys Thr Ile Glu Leu Ser Lys
20 25 30

Ala Glu Leu Ser Asp Leu Glu Leu Ile Gly Thr Gly Ala Tyr Ser Pro
35 40 45

Leu Thr Gly Phe Leu Thr Lys Thr Asp Tyr Asp Ala Val Val Glu Thr
50 55 60

Met Arg Leu Ser Asp Gly Thr Val Trp Ser Ile Pro Val Thr Leu Ala
65 70 75 80

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

Leu Val Tyr Arg Gly Asp Val Tyr Gly Val Ile Glu Ile Ala Asp Ile
100 105 110

Tyr Arg Pro Asp Lys Thr Lys Glu Ala Lys Leu Val Tyr Lys Thr Asp
115 120 125

Glu Leu Ala His Pro Gly Val Arg Lys Leu Phe Glu Lys Pro Asp Val
130 135 140

Tyr Val Gly Gly Glu Ile Thr Leu Val Lys Arg Thr Asp Lys Gly Gln
145 150 155 160

Phe Ala Ala Phe Tyr Phe Asp Pro Ala Glu Thr Arg Lys Lys Phe Ala
165 170 175

Glu Phe Gly Trp Asn Thr Val Val Gly Phe Gln Thr Arg Asn Pro Val
180 185 190

His Arg Ala His Glu Tyr Ile Gln Lys Cys Ala Leu Glu Ile Val Asp
195 200 205

Gly Leu Phe Leu Asn Pro Leu Val Gly Glu Thr Lys Ser Asp Asp Ile
210 215 220

Pro Ala Asp Ile Arg Met Glu Ser Tyr Gln Val Leu Leu Glu Asn Tyr
225 230 235 240

Tyr Pro Lys Asp Arg Val Phe Leu Gly Val Phe Gln Ala Ala Met Arg
245 250 255

Tyr Ala Gly Pro Arg Glu Ala Ile Phe His Ala Met Val Arg Lys Asn
260 265 270

Phe Gly Cys Thr His Phe Ile Val Gly Arg Asp His Ala Gly Val Gly
275 280 285

Asn Tyr Tyr Gly Thr Tyr Asp Ala Gln Lys Ile Phe Leu Asn Phe Thr
290 295 300

Ala Glu Glu Leu Gly Ile Thr Pro Leu Phe Phe Glu His Ser Phe Tyr
305 310 315 320

Cys Thr Lys Cys Glu Gly Met Ala Ser Thr Lys Thr Cys Pro His Asp

-continued

```
                325                 330                 335

Ala Lys Tyr His Val Val Leu Ser Gly Thr Lys Val Arg Glu Met Leu
            340                 345                 350

Arg Asn Gly Gln Val Pro Pro Ser Thr Phe Ser Arg Pro Glu Val Ala
        355                 360                 365

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

Ala Arg
385


<210> SEQ ID NO 23
<211> LENGTH: 546
<212> TYPE: PRT
<213> ORGANISM: Aquifex aeolicus

<400> SEQUENCE: 23

Met Glu Lys Ile Lys Tyr Leu Lys Ser Ile Gln Ile Ser Gln Arg Ser
  1               5                  10                  15

Val Leu Asp Leu Lys Leu Leu Ala Val Gly Ala Phe Thr Pro Leu Asp
             20                  25                  30

Arg Phe Met Gly Glu Glu Asp Tyr Arg Asn Val Val Glu Ser Met Arg
         35                  40                  45

Leu Lys Ser Gly Thr Leu Phe Pro Ile Pro Ile Thr Leu Pro Met Glu
     50                  55                  60

Lys Glu Ile Ala Lys Asp Leu Lys Glu Gly Glu Trp Ile Val Leu Arg
 65                  70                  75                  80

Asp Pro Lys Asn Val Pro Leu Ala Ile Met Arg Val Glu Glu Val Tyr
                 85                  90                  95

Lys Trp Asn Leu Glu Tyr Glu Ala Lys Asn Val Leu Gly Thr Thr Asp
            100                 105                 110

Pro Arg His Pro Leu Val Ala Glu Met His Thr Trp Gly Glu Tyr Tyr
        115                 120                 125

Ile Ser Gly Glu Leu Lys Val Ile Gln Leu Pro Lys Tyr Tyr Asp Phe
    130                 135                 140

Pro Glu Tyr Arg Lys Thr Pro Lys Gln Val Arg Glu Glu Ile Lys Ser
145                 150                 155                 160

Leu Gly Leu Asp Lys Ile Val Ala Phe Gln Thr Arg Asn Pro Met His
                165                 170                 175

Arg Val His Glu Glu Leu Thr Lys Arg Ala Met Glu Lys Val Gly Gly
            180                 185                 190

Gly Leu Leu Leu His Pro Val Val Gly Leu Thr Lys Pro Gly Asp Val
        195                 200                 205

Asp Val Tyr Thr Arg Met Arg Ile Tyr Lys Val Leu Tyr Glu Lys Tyr
    210                 215                 220

Tyr Asp Lys Lys Lys Thr Ile Leu Ala Phe Leu Pro Leu Ala Met Arg
225                 230                 235                 240

Met Ala Gly Pro Arg Glu Ala Leu Trp His Gly Ile Ile Arg Arg Asn
                245                 250                 255

Tyr Gly Ala Thr His Phe Ile Val Gly Arg Asp His Ala Ser Pro Gly
            260                 265                 270

Lys Asp Ser Lys Gly Lys Pro Phe Tyr Asp Pro Tyr Glu Ala Gln Glu
        275                 280                 285
```

-continued

```
Leu Phe Lys Lys Tyr Glu Asp Glu Ile Gly Ile Lys Met Val Pro Phe
    290                 295                 300

Glu Glu Leu Val Tyr Val Pro Glu Leu Asp Gln Tyr Val Glu Ile Asn
305                 310                 315                 320

Glu Ala Lys Lys Arg Asn Leu Lys Tyr Ile Asn Ile Ser Gly Thr Glu
                325                 330                 335

Ile Arg Glu Asn Phe Leu Lys Gln Gly Arg Lys Leu Pro Glu Trp Phe
            340                 345                 350

Thr Arg Pro Glu Val Ala Glu Ile Leu Ala Glu Thr Tyr Val Pro Lys
        355                 360                 365

His Lys Gln Gly Phe Cys Val Trp Leu Thr Gly Leu Pro Cys Ala Gly
    370                 375                 380

Lys Ser Thr Ile Ala Glu Ile Leu Ala Thr Met Leu Gln Ala Arg Gly
385                 390                 395                 400

Arg Lys Val Thr Leu Leu Asp Gly Asp Val Val Arg Thr His Leu Ser
                405                 410                 415

Arg Gly Leu Gly Phe Ser Lys Glu Asp Arg Ile Thr Asn Ile Leu Arg
            420                 425                 430

Val Gly Phe Val Ala Ser Glu Ile Val Lys His Asn Gly Val Val Ile
        435                 440                 445

Cys Ala Leu Val Ser Pro Tyr Arg Ser Ala Arg Asn Gln Val Arg Asn
    450                 455                 460

Met Met Glu Glu Gly Lys Phe Ile Glu Val Phe Val Asp Ala Pro Val
465                 470                 475                 480

Glu Val Cys Glu Glu Arg Asp Val Lys Gly Leu Tyr Lys Lys Ala Lys
                485                 490                 495

Glu Gly Leu Ile Lys Gly Phe Thr Gly Val Asp Asp Pro Tyr Glu Pro
            500                 505                 510

Pro Val Ala Pro Glu Val Arg Val Asp Thr Thr Lys Leu Thr Pro Glu
        515                 520                 525

Glu Ser Ala Leu Lys Ile Leu Glu Phe Leu Lys Lys Glu Gly Phe Ile
    530                 535                 540

Lys Asp
545


<210> SEQ ID NO 24
<211> LENGTH: 259
<212> TYPE: PRT
<213> ORGANISM: Pyrococcus furiosus

<400> SEQUENCE: 24

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

Pro Arg Thr Arg Glu Arg Ile Leu Ser Glu Gln His Glu Tyr Pro Arg
            20                  25                  30

Val Gln Ile Asp His Gly Arg Ala Ile Asp Leu Glu Asn Ile Ala His
        35                  40                  45

Gly Val Tyr Ser Pro Leu Lys Gly Phe Leu Thr Arg Glu Asp Phe Glu
     50                  55                  60

Ser Val Leu Asp Tyr Met Arg Leu Ser Asp Asp Thr Pro Trp Thr Ile
 65                  70                  75                  80

Pro Ile Val Leu Asp Val Gly Glu Pro Thr Phe Glu Gly Gly Asp Ala
                 85                  90                  95
```

-continued

```
Ile Leu Leu Tyr Tyr Glu Asn Pro Pro Ile Ala Arg Met His Val Glu
            100                 105                 110

Asp Ile Tyr Thr Tyr Asp Lys Lys Glu Phe Ala Val Lys Val Phe Lys
        115                 120                 125

Thr Asp Asp Pro Asn His Leu Gly Val Ala Arg Val Tyr Ser Met Gly
    130                 135                 140

Lys Tyr Leu Val Gly Gly Gly Ile Glu Leu Leu Asn Glu Leu Pro Asn
145                 150                 155                 160

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

Lys Glu Arg Gly Trp Lys Thr Ile Val Ala Phe Gln Thr Arg Asn Val
            180                 185                 190

Pro His Leu Gly His Glu Tyr Val Gln Lys Ala Ala Leu Thr Phe Val
        195                 200                 205

Asp Gly Leu Phe Ile Asn Pro Val Leu Gly Arg Lys Lys Lys Gly Asp
    210                 215                 220

Tyr Lys Asp Glu Val Ile Ile Lys Ala Tyr Tyr Leu Ile Met Lys Tyr
225                 230                 235                 240

Cys Ser Asn Thr Thr His His Ala Ile Met Arg Lys Thr Ser Thr Ser
                245                 250                 255

Ser Gln Thr


<210> SEQ ID NO 25
<211> LENGTH: 406
<212> TYPE: PRT
<213> ORGANISM: Sulfolobus solfataricus

<400> SEQUENCE: 25

Met Asn Leu Ile Gly His Gly Lys Val Glu Ile Val Glu Arg Ile Lys
  1               5                  10                  15

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

Leu Ala His Glu Ile Val Ser Ile Ala Tyr Gly Phe Leu Ser Pro Leu
         35                  40                  45

Lys Gly Phe Met Asn Tyr Glu Glu Val Asp Gly Val Val Glu Asn Met
     50                  55                  60

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

Ser Gln Asn Glu Lys Val Lys Glu Gly Asp Thr Ile Gly Ile Thr Tyr
                 85                  90                  95

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

Asp Lys Leu Lys Ile Ala Glu Lys Val Tyr Lys Thr Lys Asp Ile Lys
        115                 120                 125

His Pro Gly Val Lys Arg Thr Leu Ser Tyr Ala Asp Ala Phe Leu Ala
    130                 135                 140

Gly Asp Val Trp Leu Val Arg Glu Pro Gln Phe Asn Lys Pro Tyr Ser
145                 150                 155                 160

Glu Phe Trp Leu Thr Pro Arg Met His Arg Thr Val Phe Glu Lys Lys
                165                 170                 175

Gly Trp Lys Arg Val Val Ala Phe Gln Thr Arg Asn Val Pro His Thr
            180                 185                 190
```

-continued

```
Gly His Glu Tyr Leu Met Lys Phe Ala Trp Phe Ala Ala Asn Glu Asn
        195                 200                 205

Gln Lys Val Asp Glu Pro Arg Thr Gly Ile Leu Val Asn Val Val Ile
    210                 215                 220

Gly Glu Lys Arg Val Gly Asp Tyr Ile Asp Glu Ala Ile Leu Leu Thr
225                 230                 235                 240

His Asp Ala Leu Ser Lys Tyr Gly Tyr Ile Ser Pro Lys Val His Leu
                245                 250                 255

Leu Ser Phe Thr Leu Trp Asp Met Arg Tyr Ala Gly Pro Arg Glu Ala
            260                 265                 270

Leu Leu His Ala Ile Ile Arg Ser Asn Leu Gly Cys Thr His His Val
        275                 280                 285

Phe Gly Arg Asp His Ala Gly Val Gly Asn Tyr Tyr Ser Pro Tyr Glu
    290                 295                 300

Ala His Glu Ile Phe Asp Ser Ile Asn Glu Glu Asp Leu Leu Ile Lys
305                 310                 315                 320

Pro Ile Phe Leu Arg Glu Asn Tyr Tyr Cys Pro Arg Cys Gly Ser Ile
                325                 330                 335

Glu Asn Glu Ile Leu Cys Asp His Lys Asp Glu Lys Gln Glu Phe Ser
            340                 345                 350

Gly Ser Leu Ile Arg Ser Ile Ile Leu Asp Glu Val Lys Pro Thr Lys
        355                 360                 365

Met Val Met Arg Pro Glu Val Tyr Asp Val Leu Met Lys Ala Ala Glu
    370                 375                 380

Gln Tyr Gly Phe Gly Ser Pro Phe Val Thr Glu Glu Tyr Leu Glu Lys
385                 390                 395                 400

Arg Gln Ser Ile Leu Gly
                405


<210> SEQ ID NO 26
<211> LENGTH: 455
<212> TYPE: PRT
<213> ORGANISM: Pyrobaculum aerophilum

<400> SEQUENCE: 26

Met Pro Met Pro Ala Pro Leu Glu Pro His Gly Gly Arg Leu Val Tyr
  1               5                  10                  15

Asn Val Ile Glu Asp Arg Asp Lys Ala Ala Ala Met Ile Gln Gly Leu
            20                  25                  30

Pro Ser Ile Glu Ile Glu Pro Thr Leu Gly Pro Asp Gly Ser Pro Ile
        35                  40                  45

Arg Asn Pro Tyr Arg Glu Ile Met Ser Ile Ala Tyr Gly Phe Phe Ser
     50                  55                  60

Pro Val Glu Gly Phe Met Thr Arg Asn Glu Val Glu Ser Ile Leu Lys
 65                  70                  75                  80

Glu Arg Arg Leu Leu Asn Gly Trp Leu Phe Pro Phe Pro Leu Ile Tyr
                85                  90                  95

Asp Val Asp Glu Glu Lys Ile Lys Gly Ile Lys Glu Gly Asp Ser Val
            100                 105                 110

Leu Leu Lys Leu Lys Gly Lys Pro Leu Ala Val Leu Asn Val Glu Glu
        115                 120                 125

Ile Trp Arg Leu Pro Asp Arg Lys Glu Leu Ala Asp Ala Val Phe Gly
```

-continued

```
    130                 135                 140

Thr Pro Glu Arg Asn Lys Glu Val Val Lys Lys Arg Phe Asp Glu Lys
145                 150                 155                 160

His Pro Gly Trp Leu Ile Tyr Arg Ser Met Arg Pro Met Ala Leu Ala
                165                 170                 175

Gly Lys Ile Thr Val Val Asn Pro Pro Arg Phe Lys Glu Pro Tyr Ser
            180                 185                 190

Arg Phe Trp Met Pro Pro Arg Val Ser Arg Glu Tyr Val Glu Lys Lys
        195                 200                 205

Gly Trp Arg Ile Val Val Ala His Gln Thr Arg Asn Val Pro His Ile
    210                 215                 220

Gly His Glu Met Leu Met Lys Arg Ala Met Phe Val Ala Gly Gly Glu
225                 230                 235                 240

Arg Pro Gly Asp Ala Val Leu Val Asn Ala Ile Ile Gly Ala Lys Arg
                245                 250                 255

Pro Gly Asp Tyr Val Asp Glu Ala Ile Leu Glu Gly His Glu Ala Leu
            260                 265                 270

Asn Lys Ala Gly Tyr Phe His Pro Asp Arg His Val Val Thr Met Thr
        275                 280                 285

Leu Trp Asp Met Arg Tyr Gly Asn Pro Leu Glu Ser Leu Leu His Gly
    290                 295                 300

Ile Ile Arg Gln Asn Met Gly Ala Thr His His Met Phe Gly Arg Asp
305                 310                 315                 320

His Ala Ala Thr Gly Asp Tyr Tyr Asp Pro Tyr Ala Thr Gln Tyr Leu
                325                 330                 335

Trp Thr Arg Gly Leu Pro Ser Tyr Gly Leu Asn Glu Pro Pro His Met
            340                 345                 350

Thr Asp Lys Gly Leu Arg Ile Lys Pro Val Asn Leu Gly Glu Phe Ala
        355                 360                 365

Tyr Cys Pro Lys Cys Gly Glu Tyr Thr Tyr Leu Gly Met Ser Tyr Glu
    370                 375                 380

Gly Tyr Lys Glu Val Ala Leu Cys Gly His Thr Pro Glu Arg Ile Ser
385                 390                 395                 400

Gly Ser Leu Leu Arg Gly Ile Ile Ile Glu Gly Leu Arg Pro Pro Lys
                405                 410                 415

Val Val Met Arg Pro Glu Val Tyr Asp Val Ile Val Lys Trp Trp Arg
            420                 425                 430

Val Tyr Gly Tyr Pro Tyr Val Thr Asp Lys Tyr Leu Arg Ile Lys Glu
        435                 440                 445

Gln Glu Leu Glu Val Glu Leu
    450                 455


<210> SEQ ID NO 27
<211> LENGTH: 456
<212> TYPE: PRT
<213> ORGANISM: Archaeoglobus fulgidus

<400> SEQUENCE: 27

Met Pro Leu Ile Lys Thr Pro Pro Pro His Gly Gly Lys Leu Val Glu
  1               5                  10                  15

Arg Val Val Lys Lys Arg Asp Ile Ala Glu Lys Met Ile Ala Gly Cys
            20                  25                  30
```

-continued

```
Pro Thr Tyr Glu Leu Lys Pro Thr Thr Leu Pro Asp Gly Thr Pro Ile
        35                  40                  45

Arg His Val Tyr Arg Glu Ile Met Ser Val Cys Tyr Gly Phe Phe Ser
    50                  55                  60

Pro Val Glu Gly Ser Met Val Gln Asn Glu Leu Glu Arg Val Leu Asn
65                  70                  75                  80

Glu Arg Arg Leu Leu Ser Glu Trp Ile Phe Pro Tyr Pro Ile Leu Phe
                85                  90                  95

Asp Ile Ser Glu Glu Asp Tyr Lys Ala Leu Asp Val Lys Glu Gly Asp
            100                 105                 110

Arg Leu Leu Leu Met Leu Lys Gly Gln Pro Phe Ala Thr Leu Asp Ile
        115                 120                 125

Glu Glu Val Tyr Lys Ile Asp Pro Val Asp Val Ala Thr Arg Thr Phe
    130                 135                 140

Gly Thr Pro Glu Lys Asn Pro Glu Val Val Arg Glu Pro Phe Asp Asp
145                 150                 155                 160

Lys His Pro Gly Tyr Val Ile Tyr Lys Met His Asn Pro Ile Ile Leu
                165                 170                 175

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

Asp Arg Phe Trp Phe Pro Pro Ser Lys Cys Arg Glu Val Ile Lys Asn
        195                 200                 205

Glu Lys Lys Trp Arg Thr Val Ile Ala His Gln Thr Arg Asn Val Pro
    210                 215                 220

His Val Gly His Glu Met Leu Met Lys Cys Ala Ala Tyr Thr Gly Asp
225                 230                 235                 240

Ile Glu Pro Cys His Gly Ile Leu Val Asn Ala Ile Ile Gly Ala Lys
                245                 250                 255

Arg Arg Gly Asp Tyr Pro Asp Glu Ala Ile Leu Glu Gly His Glu Ala
            260                 265                 270

Val Asn Lys Tyr Gly Tyr Ile Lys Pro Glu Arg His Met Val Thr Phe
        275                 280                 285

Thr Leu Trp Asp Met Arg Tyr Gly Asn Pro Ile Glu Ser Leu Leu His
    290                 295                 300

Gly Val Ile Arg Gln Asn Met Gly Cys Thr His His Met Phe Gly Arg
305                 310                 315                 320

Asp His Ala Ala Val Gly Glu Tyr Tyr Asp Met Tyr Ala Thr Gln Ile
                325                 330                 335

Leu Trp Ser Gln Gly Ile Pro Ser Phe Gly Phe Glu Ala Pro Pro Asn
            340                 345                 350

Glu Val Asp Tyr Gly Leu Lys Ile Ile Pro Gln Asn Met Ala Glu Phe
        355                 360                 365

Trp Tyr Cys Pro Ile Cys Gln Glu Ile Ala Tyr Ser Glu Asn Cys Gly
    370                 375                 380

His Thr Asp Ala Lys Gln Lys Phe Ser Gly Ser Phe Leu Arg Gly Met
385                 390                 395                 400

Val Ala Glu Gly Val Phe Pro Pro Arg Val Val Met Arg Pro Glu Val
                405                 410                 415

Tyr Lys Gln Ile Val Lys Trp Trp Lys Val Tyr Asn Tyr Pro Phe Val
            420                 425                 430

Asn Arg Lys Tyr Leu Glu Leu Lys Asn Lys Glu Leu Glu Ile Asp Leu
```

-continued

```
        435                 440                 445

Pro Ala Met Glu Val Pro Lys Ala
    450                 455


<210> SEQ ID NO 28
<211> LENGTH: 573
<212> TYPE: PRT
<213> ORGANISM: Penicillium chrysogenum

<400> SEQUENCE: 28

Met Ala Asn Ala Pro His Gly Gly Val Leu Lys Asp Leu Leu Ala Arg
  1               5                  10                  15

Asp Ala Pro Arg Gln Ala Glu Leu Ala Ala Glu Ala Glu Ser Leu Pro
             20                  25                  30

Ala Val Thr Leu Thr Glu Arg Gln Leu Cys Asp Leu Glu Leu Ile Met
         35                  40                  45

Asn Gly Gly Phe Ser Pro Leu Glu Gly Phe Met Asn Gln Ala Asp Tyr
     50                  55                  60

Asp Arg Val Cys Glu Asp Asn Arg Leu Ala Asp Gly Asn Val Phe Ser
 65                  70                  75                  80

Met Pro Ile Thr Leu Asp Ala Ser Gln Glu Val Ile Asp Glu Lys Lys
                 85                  90                  95

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

Asn Leu Ala Ile Leu Thr Ile Asp Asp Ile Tyr Arg Pro Asp Lys Thr
        115                 120                 125

Lys Glu Ala Lys Leu Val Phe Gly Gly Asp Pro Glu His Pro Ala Ile
    130                 135                 140

Val Tyr Leu Asn Asn Thr Val Lys Glu Phe Tyr Ile Gly Gly Lys Ile
145                 150                 155                 160

Glu Ala Val Asn Lys Leu Asn His Tyr Asp Tyr Val Ala Leu Arg Tyr
                165                 170                 175

Thr Pro Ala Glu Leu Arg Val His Phe Asp Lys Leu Gly Trp Ser Arg
            180                 185                 190

Val Val Ala Phe Gln Thr Arg Asn Pro Met His Arg Ala His Arg Glu
        195                 200                 205

Leu Thr Val Arg Ala Ala Arg Ser Arg Gln Ala Asn Val Leu Ile His
    210                 215                 220

Pro Val Val Gly Leu Thr Lys Pro Gly Asp Ile Asp His Phe Thr Arg
225                 230                 235                 240

Val Arg Ala Tyr Gln Ala Leu Leu Pro Arg Tyr Pro Asn Gly Met Ala
                245                 250                 255

Val Leu Gly Leu Leu Gly Leu Ala Met Arg Met Gly Gly Pro Arg Glu
            260                 265                 270

Ala Ile Trp His Ala Ile Ile Arg Lys Asn His Gly Ala Thr His Phe
        275                 280                 285

Ile Val Gly Arg Asp His Ala Gly Pro Gly Ser Asn Ser Lys Gly Glu
    290                 295                 300

Asp Phe Tyr Gly Pro Tyr Asp Ala Gln His Ala Val Glu Lys Tyr Lys
305                 310                 315                 320

Asp Glu Leu Gly Ile Glu Val Val Glu Phe Gln Met Val Thr Tyr Leu
                325                 330                 335
```

-continued

```
Pro Asp Thr Asp Glu Tyr Arg Pro Val Asp Gln Val Pro Ala Gly Val
            340                 345                 350

Lys Thr Leu Asn Ile Ser Gly Thr Glu Leu Arg Arg Arg Leu Arg Ser
        355                 360                 365

Gly Ala His Ile Pro Glu Trp Phe Ser Tyr Pro Glu Val Val Lys Ile
    370                 375                 380

Leu Arg Glu Ser Asn Pro Pro Arg Ala Thr Gln Gly Phe Thr Ile Phe
385                 390                 395                 400

Leu Thr Gly Tyr Met Asn Ser Gly Lys Asp Ala Ile Ala Arg Ala Leu
                405                 410                 415

Gln Val Thr Leu Asn Gln Gln Gly Gly Arg Ser Val Ser Leu Leu Leu
            420                 425                 430

Gly Asp Thr Val Arg His Glu Leu Ser Ser Glu Leu Gly Phe Thr Arg
        435                 440                 445

Glu Asp Arg His Thr Asn Ile Gln Arg Ile Ala Phe Val Ala Thr Glu
    450                 455                 460

Leu Thr Arg Ala Gly Ala Ala Val Ile Ala Ala Pro Ile Ala Pro Tyr
465                 470                 475                 480

Glu Glu Ser Arg Lys Phe Ala Arg Asp Ala Val Ser Gln Ala Gly Ser
                485                 490                 495

Phe Phe Leu Val His Val Ala Thr Pro Leu Glu His Cys Glu Gln Ser
            500                 505                 510

Asp Lys Arg Gly Ile Tyr Ala Ala Ala Arg Arg Gly Glu Ile Lys Gly
        515                 520                 525

Phe Thr Gly Val Asp Asp Pro Tyr Glu Thr Pro Glu Lys Ala Asp Leu
    530                 535                 540

Val Val Asp Phe Ser Lys Gln Ser Val Arg Ser Ile Val His Glu Ile
545                 550                 555                 560

Ile Leu Val Leu Glu Ser Gln Gly Phe Leu Glu Arg Gln
                565                 570


<210> SEQ ID NO 29
<211> LENGTH: 389
<212> TYPE: PRT
<213> ORGANISM: Aeropyrum pernix

<400> SEQUENCE: 29

Met Gly Cys Ser Val Gly Leu Val Ser Arg Pro His Gly Gly Arg Leu
  1               5                  10                  15

Val Arg Arg Val Leu Ser Gly Arg Arg Arg Glu Ile Phe Glu Ser Gln
            20                  25                  30

Tyr Arg Glu Met Pro Arg Leu Glu Val Pro Leu Glu Arg Ala Ile Asp
        35                  40                  45

Ala Glu Asp Leu Ala Arg Gly Val Phe Ser Pro Leu Glu Gly Phe Met
    50                  55                  60

Val Glu Asp Asp Tyr Leu Ser Val Leu Ser Arg Met Arg Leu Ser Asn
 65                  70                  75                  80

Asp Leu Pro Trp Thr Ile Pro Ile Val Leu Asp Ala Asn Arg Glu Trp
                85                  90                  95

Val Leu Asn Glu Gly Val Ser Ala Gly Asp Asp Ile Ile Leu Thr Tyr
            100                 105                 110

His Gly Leu Pro Ile Ala Val Leu Thr Leu Glu Asp Ile Tyr Ser Trp
        115                 120                 125
```

-continued

```
Asp Lys Gly Leu His Ala Glu Lys Val Phe Lys Thr Arg Asp Pro Asn
    130                 135                 140

His Pro Gly Val Glu Ala Thr Tyr Lys Arg Gly Asp Ile Leu Leu Gly
145                 150                 155                 160

Gly Arg Leu Glu Leu Ile Gln Gly Pro Pro Asn Pro Leu Glu Arg Tyr
                165                 170                 175

Thr Leu Trp Pro Val Glu Thr Arg Val Leu Phe Lys Glu Lys Gly Trp
            180                 185                 190

Arg Thr Val Ala Ala Phe Gln Thr Arg Asn Val Pro His Leu Gly His
        195                 200                 205

Glu Tyr Val Gln Lys Ala Ala Leu Thr Phe Val Asp Gly Leu Leu Val
    210                 215                 220

His Pro Leu Ala Gly Trp Lys Lys Arg Gly Asp Tyr Arg Asp Glu Val
225                 230                 235                 240

Ile Ile Arg Ala Tyr Glu Ala Leu Ile Thr His Tyr Tyr Pro Arg Gly
                245                 250                 255

Val Val Val Leu Ser Val Leu Arg Met Asn Met Asn Tyr Ala Gly Pro
            260                 265                 270

Arg Glu Ala Val His His Ala Ile Val Arg Lys Asn Phe Gly Ala Thr
        275                 280                 285

His Phe Ile Val Gly Arg Asp His Ala Gly Val Gly Ser Tyr Tyr Gly
    290                 295                 300

Pro Tyr Glu Ala Trp Glu Ile Phe Arg Glu Phe Pro Asp Leu Gly Ile
305                 310                 315                 320

Thr Pro Leu Phe Val Arg Glu Ala Tyr Tyr Cys Arg Arg Cys Gly Gly
                325                 330                 335

Met Val Asn Glu Lys Val Cys Pro His Gly Asp Glu Tyr Arg Val Arg
            340                 345                 350

Ile Ser Gly Thr Arg Leu Arg Glu Met Leu Gly Arg Gly Glu Arg Pro
        355                 360                 365

Pro Glu Tyr Met Met Arg Pro Glu Val Ala Asp Ala Ile Ile Ser His
    370                 375                 380

Pro Asp Pro Phe Ile
385


<210> SEQ ID NO 30
<211> LENGTH: 511
<212> TYPE: PRT
<213> ORGANISM: Saccharomyces cerevisiae

<400> SEQUENCE: 30

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

Ala Leu Lys Lys Asn Glu Leu Leu Ser Glu Ala Gln Ser Ser Asp Ile
            20                  25                  30

Leu Val Trp Asn Leu Thr Pro Arg Gln Leu Cys Asp Ile Glu Leu Ile
         35                  40                  45

Leu Asn Gly Gly Phe Ser Pro Leu Thr Gly Phe Leu Asn Glu Asn Asp
     50                  55                  60

Tyr Ser Ser Val Val Thr Asp Ser Arg Leu Ala Asp Gly Thr Leu Trp
 65                  70                  75                  80

Thr Ile Pro Ile Thr Leu Asp Val Asp Glu Ala Phe Ala Asn Gln Ile
```

-continued

```
                85                  90                  95

Lys Pro Asp Thr Arg Ile Ala Leu Phe Gln Asp Asp Glu Ile Pro Ile
            100                 105                 110

Ala Ile Leu Thr Val Gln Asp Val Tyr Lys Pro Asn Lys Thr Ile Glu
        115                 120                 125

Ala Glu Lys Val Phe Arg Gly Asp Pro Glu His Pro Ala Ile Ser Tyr
    130                 135                 140

Leu Phe Asn Val Ala Gly Asp Tyr Tyr Val Gly Gly Ser Leu Glu Ala
145                 150                 155                 160

Ile Gln Leu Pro Gln His Tyr Asp Tyr Pro Gly Leu Arg Lys Thr Pro
                165                 170                 175

Ala Gln Leu Arg Leu Glu Phe Gln Ser Arg Gln Trp Asp Arg Val Val
            180                 185                 190

Ala Phe Gln Thr Arg Asn Pro Met His Arg Ala His Arg Glu Leu Thr
        195                 200                 205

Val Arg Ala Ala Arg Glu Ala Asn Ala Lys Val Leu Ile His Pro Val
    210                 215                 220

Val Gly Leu Thr Lys Pro Gly Asp Ile Asp His His Thr Arg Val Arg
225                 230                 235                 240

Val Tyr Gln Glu Ile Ile Lys Arg Tyr Pro Asn Gly Ile Ala Phe Leu
                245                 250                 255

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

Trp His Ala Ile Ile Arg Lys Asn Tyr Gly Ala Ser His Phe Ile Val
        275                 280                 285

Gly Arg Asp His Ala Gly Pro Gly Lys Asn Ser Lys Gly Val Asp Phe
    290                 295                 300

Tyr Gly Pro Tyr Asp Ala Gln Glu Leu Val Glu Ser Tyr Lys His Glu
305                 310                 315                 320

Leu Asp Ile Glu Val Val Pro Phe Arg Met Val Thr Tyr Leu Pro Asp
                325                 330                 335

Glu Asp Arg Tyr Ala Pro Ile Asp Gln Ile Asp Thr Thr Lys Thr Arg
            340                 345                 350

Thr Leu Asn Ile Ser Gly Thr Glu Leu Arg Arg Arg Leu Arg Val Gly
        355                 360                 365

Gly Glu Ile Pro Glu Trp Phe Ser Tyr Pro Glu Val Val Lys Ile Leu
    370                 375                 380

Arg Glu Ser Asn Pro Pro Arg Pro Lys Gln Gly Phe Ser Ile Val Leu
385                 390                 395                 400

Gly Asn Ser Leu Thr Val Ser Arg Glu Gln Leu Ser Ile Ala Leu Leu
                405                 410                 415

Ser Thr Phe Leu Gln Phe Gly Gly Gly Arg Tyr Tyr Lys Ile Phe Glu
            420                 425                 430

His Asn Asn Lys Thr Glu Leu Leu Ser Leu Ile Gln Asp Phe Ile Gly
        435                 440                 445

Ser Gly Ser Gly Leu Ile Ile Pro Asn Gln Trp Glu Asp Asp Lys Asp
    450                 455                 460

Ser Val Val Gly Lys Gln Asn Val Tyr Leu Leu Asp Thr Ser Ser Ser
465                 470                 475                 480

Ala Asp Ile Gln Leu Glu Ser Ala Asp Glu Pro Ile Ser His Ile Val
                485                 490                 495
```

-continued

```
Gln Lys Val Val Leu Phe Leu Glu Asp Asn Gly Phe Phe Val Phe
            500                 505                 510


<210> SEQ ID NO 31
<211> LENGTH: 309
<212> TYPE: PRT
<213> ORGANISM: Thermomonospora fusca

<400> SEQUENCE: 31

Met Ser Gln Val Ser Asp Ala Val Gly Arg Tyr Gln Leu Ser Gln Leu
  1               5                  10                  15

Asp Phe Leu Glu Ala Glu Ala Ile Phe Ile Met Arg Glu Val Ala Ala
            20                  25                  30

Glu Phe Glu Arg Pro Val Leu Leu Phe Ser Gly Gly Lys Asp Ser Val
        35                  40                  45

Val Met Leu Arg Ile Ala Glu Lys Ala Phe Trp Pro Ala Pro Ile Pro
    50                  55                  60

Phe Pro Val Met His Val Asp Thr Gly His Asn Phe Pro Glu Val Ile
 65                  70                  75                  80

Glu Phe Arg Asp Lys Arg Val Ala Glu Leu Gly Val Arg Leu Ile Val
                 85                  90                  95

Ala Ser Val Gln Asp Leu Ile Asp Ala Gly Lys Val Val Glu Pro Lys
            100                 105                 110

Gly Arg Trp Ala Ser Arg Asn Arg Leu Gln Thr Ala Ala Leu Leu Glu
        115                 120                 125

Ala Ile Glu Lys Tyr Gly Phe Asp Ala Ala Phe Gly Gly Ala Arg Arg
    130                 135                 140

Asp Glu Glu Lys Ala Arg Ala Lys Glu Arg Val Phe Ser Phe Arg Asp
145                 150                 155                 160

Glu Phe Gly Gln Trp Asp Pro Lys Asn Gln Arg Pro Glu Leu Trp Asn
                165                 170                 175

Leu Tyr Asn Thr Arg Val His Arg Gly Glu Asn Ile Arg Val Phe Pro
            180                 185                 190

Leu Ser Asn Trp Thr Glu Leu Asp Val Trp His Tyr Ile Arg Arg Glu
        195                 200                 205

Gly Leu Arg Leu Pro Ser Ile Tyr Phe Ala His Arg Arg Arg Val Phe
    210                 215                 220

Glu Arg Asp Gly Ile Leu Leu Pro Asp Ser Pro Tyr Val Thr Arg Asp
225                 230                 235                 240

Glu Asp Glu Glu Val Phe Glu Ala Ser Val Arg Tyr Arg Thr Val Gly
                245                 250                 255

Asp Met Thr Cys Thr Gly Ala Val Leu Ser Thr Ala Thr Thr Leu Asp
            260                 265                 270

Glu Val Ile Ala Glu Ile Ala Ala Thr Arg Ile Thr Glu Arg Gly Gln
        275                 280                 285

Thr Arg Ala Asp Asp Arg Gly Ser Glu Ala Ala Met Glu Glu Arg Lys
    290                 295                 300

Arg Glu Gly Tyr Phe
305
```

1-221. (canceled)

222. A method of determining the base sequence of a plurality of single stranded template nucleotides on an array, the method comprising:

(a) providing a planar surface comprises at least 400,000 discrete cavities, wherein each cavity forms a reaction chamber containing single-stranded nucleic acid templates of a single species, wherein the reaction chambers have a center to center spacing of between 5 to 200 μm, wherein each reaction chamber contains a reaction mixture comprising a template-directed nucleotide polymerase and said one of said plurality of single-stranded template nucleotides hybridized to a complementary oligonucleotide primer strand at least one nucleotide residue shorter than the single-stranded template nucleotides to form at least one unpaired nucleotide residue in each template at the 3'-end of the primer strand;

(b) adding an activated nucleotide 5'-triphosphate precursor of one known nitrogenous base to the reaction chambers under conditions which allow incorporation of the activated nucleoside 5'-triphosphate precursor onto the 3'-end of the primer strand, provided the nitrogenous base of the activated nucleoside 5'-triphosphate precursor is complementary to the nitrogenous base of the unpaired nucleotide residue of the templates;

(c) detecting whether or not the nucleoside 5'-triphosphate precursor was incorporated into the primer strands in each reaction chamber by detecting a sequencing byproduct with an ATP generating polypeptide-ATP converting polypeptide fusion protein or an ATP generating protein and an ATP converting protein, thus indicating that the unpaired nucleotide residue of the template has a nitrogenous base composition that is complementary to that of the incorporated nucleoside 5'-triphosphate precursor in each reaction chamber;

(d) sequentially repeating steps (b) and (c), wherein each sequential repetition adds and, detects the incorporation of one type of activated nucleoside 5'-triphosphate precursor of known nitrogenous base composition; and

(e) determining the base sequence of the unpaired nucleotide residues of the template in each reaction chamber from the sequence of incorporation of said nucleoside precursors.

223. The method of claim 222 wherein said sequencing byproduct is pyrophosphate.

224. The method of claim 222 wherein the ATP generating polypeptide-ATP converting polypeptide fusion protein comprises an ATP generating polypeptide portion with an amino acid sequence which is at least 96% homologous to SEQ ID NO:2.

225. The method of claim 222 wherein the ATP generating polypeptide-ATP converting polypeptide fusion protein comprises an ATP generating polypeptide portion with an amino acid sequence which is SEQ ID NO:6.

226. The method of claim 222 wherein the ATP generating polypeptide-ATP converting polypeptide fusion protein comprises an amino acid sequence of SEQ ID NO:4.

227. The method of claim 222 wherein the ATP generating protein comprises an amino acid sequence which is at least 96% homologous to SEQ ID NO:2.

228. The method of claim 222 wherein the ATP generating protein comprises an amino acid sequence of SEQ ID NO:2 or SEQ ID NO:6.

229. The method of claim 222 wherein said ATP generating polypeptide-ATP converting polypeptide fusion protein comprise an amino acid sequence encoded by a polynucleotide with an open reading frame of SEQ ID NO:3.

230. The method of claim 222 wherein said ATP generating polypeptide comprise an amino acid sequence encoded by a polynucleotide with an open reading frame which is no more than 11% different from an open reading frame of SEQ ID NO:1.

231. The method of claim 222 wherein said ATP generating polypeptide comprises an amino acid sequence encoded by an open reading frame of SEQ ID NO:1 or SEQ ID NO:5.

232. The method of claim 222 wherein said ATP generating polypeptide-ATP converting polypeptide fusion protein or said ATP generating protein further comprises an affinity tag.

233. The method of claim 222 wherein said ATP generating polypeptide-ATP converting polypeptide fusion protein, said ATP generating protein, or said ATP converting polypeptide is bound to a bead.

234. A method of identifying a base at a target position in a sample nucleic acid sequence, comprising providing a sample nucleic acid and a primer which hybridizes to the sample nucleic acid immediately adjacent to the target position, subjecting the sample nucleic acid and primer to a polymerase reaction in the presence of a nucleotide whereby the nucleotide will only become incorporated if it is complementary to the base in the target position, and detecting said incorporation of the nucleotide by monitoring the release of inorganic pyrophosphate, whereby detection of incorporation of said nucleotide is indicative of identification of a base at a target position that is complementary to said nucleotide, and wherein the release of inorganic pyrophosphate is detected using a thermostable sulfurylase-luciferase fusion protein or a thermostable sulfurylase.

235. The method of claim 234 wherein the thermostable sulfurylase-luciferase fusion protein or the thermostable sulfurylase comprises an amino acid of at least 96% homology to SEQ ID NO:2, SEQ ID NO:4 or SEQ ID NO:6.

236. The method of claim 234 wherein the thermostable sulfurylase-luciferase fusion protein or the thermostable sulfurylase is encoded by an open reading frame of SEQ ID NO: 1, 3 or 5.

237. The method of claim 234 wherein the thermostable sulfurylase-luciferase fusion protein or the thermostable sulfurylase further comprises an affinity tag.

238. The method of claim 234 wherein said the thermostable sulfurylase-luciferase fusion protein or the thermostable sulfurylase is bound to a bead.

* * * * *