PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/52, 9/02, 1/21, C12P 11/00, C12S 1/02, C10G 32/00

A1

(11) International Publication Number: **WO 98/45446**

(43) International Publication Date: 15 October 1998 (15.10.98)

(21) International Application Number: PCT/US98/06684

(22) International Filing Date: 3 April 1998 (03.04.98)

(30) Priority Data:

08/835,292	7 April 1997 (07.04.97)	US
08/851,089	5 May 1997 (05.05.97)	US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications

US	08/835,292 (CIP)
Filed on	7 April 1997 (07.04.97)
US	08/851,089 (CIP)
Filed on	5 May 1997 (05.05.97)

(71) Applicant *(for all designated States except US)*: ENERGY BIOSYSTEMS CORPORATION [US/US]; 4200 Research Forest Drive, The Woodlands, TX 77381 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: DARZINS, Aldis [US/US]; 23 Firethorn Place, The Woodlands, TX 77381 (US). MRACHKO, Gregory, T. [US/US]; Apartment 3802, 26001 Budde Road, Spring, TX 77380 (US).

(74) Agents: ELMORE, Carolyn, S. et al.; Hamilton, Brook, Smith & Reynolds, P.C., Two Militia Drive, Lexington, MA 02173 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: A *SPHINGOMONAS* BIODESULFURIZATION CATALYST

(57) Abstract

The invention relates to a novel microorganism, designated *Sphingomonas sp.* strain AD109, which is capable of selectively desulfurizing dibenzothiophene. The invention also includes isolated proteins and nucleic acid sequences obtained from this microorganism. In another embodiment, the invention provides a method of using this microorganism or enzyme preparations derived therefrom in the biocatalytic desulfurization of a fossil fuel containing organic sulfur compounds.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

A *SPHINGOMONAS* BIODESULFURIZATION CATALYST

RELATED APPLICATIONS

This is a continuation-in-part application of Serial No. 08/851,089, filed May 5, 1997 which is a continuation-in-part application of Serial No. 08/835,292, filed April 7, 1997, now abandoned, the contents of which are incorporated herein by reference in its entirety.

BACKGROUND OF THE INVENTION

The microbial desulfurization of fossil fuels has been an area of active investigation for over fifty years. The object of these investigations has been to develop biotechnology based methods for the pre-combustion removal of sulfur from fossil fuels, such as coal, crude oil and petroleum distillates. The driving forces for the development of desulfurization methods are the increasing levels of sulfur in fossil fuel and the increasingly stringent regulation of sulfur emissions. Monticello *et al.*, "Practical Considerations in Biodesulfurization of Petroleum," IGT's 3d Intl. Symp. on Gas, Oil, Coal and Env. Biotech., (Dec. 3-5, 1990) New Orleans, LA.

Many biocatalysts and processes have been developed to desulfurize fossil fuels, including those described in U.S. Patent Nos. 5,356,801, 5,358,870, 5,358,813, 5,198,341, 5,132,219, 5,344,778, 5,104,801 and 5,002,888, incorporated herein by reference. Economic

analyses indicate that one limitation in the commercialization of the technology is improving the reaction rates and specific activities of the biocatalysts, such as the bacteria and enzymes that are involved in the desulfurization reactions. The reaction rates and specific activities (sulfur removed/hour/gram of biocatalyst) that have been reported in the literature are much lower than those necessary for optimal commercial technology. Therefore, improvements in the longevity and specific activity of the biocatalyst are desirable.

SUMMARY OF THE INVENTION

The invention relates to a novel microorganism, designated *Sphingomonas* sp. strain AD109, as well as isolated proteins and nucleic acid sequences obtained from this microorganism. This microorganism was obtained using a soil enrichment process using 2-(2-hydroxyphenyl)benzenesulfinate as the sole sulfur source. A biologically pure sample of this microorganism has been isolated and characterized.

The invention also relates to a collection of desulfurization enzymes isolated from *Sphingomonas* sp. strain AD109 which, together, catalyze the oxidative desulfurization of dibenzothiophene (DBT).

In another embodiment, the invention includes an isolated nucleic acid molecule, such as a DNA or RNA nucleotide sequence or molecule, which encodes one or more of the *Sphingomonas* desulfurization enzymes, or a homologue or active fragment thereof. The invention also includes a recombinant microorganism containing one or more heterologous nucleic acid molecules which encode

one or more of the *Sphingomonas* desulfurization enzymes or homologues or active fragments thereof.

In a further embodiment, the invention provides a method of using the *Sphingomonas* microorganism or an enzyme preparation derived therefrom as a biocatalyst in the biocatalytic desulfurization of a fossil fuel containing organosulfur compounds. The method comprises the steps of (1) contacting the fossil fuel with an aqueous phase containing *a Sphingomonas* biocatalyst which is capable of biocatalytic desulfurization and, optionally, a flavoprotein, thereby forming a fossil fuel and aqueous phase mixture; (2) maintaining the mixture under conditions sufficient for sulfur oxidation and/or cleavage of the carbon-sulfur bonds of the organosulfur molecules by the biocatalyst, and (3) separating the fossil fuel having a reduced organic sulfur content from the resulting aqueous phase.

The invention also provides a method of oxidizing an organic compound. The method comprises the steps of: (1) contacting the organic compound with an aqueous phase containing a *Sphingomonas* biocatalyst comprising at least one enzyme capable of catalyzing at least one step in the oxidative cleavage of carbon-sulfur bonds, thereby forming an organic compound and aqueous phase mixture; (2) maintaining the mixture of step (1) under conditions sufficient for oxidation of the organic compound by the biocatalyst, thereby resulting in an oxidized organic compound, and, optionally, separating the oxidized organic compound from the aqueous phase.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A, 1B, 1C and 1D together set forth the DNA sequence and the corresponding amino acid sequence of open reading frame 1 (ORF-1, *dszA*) of the nucleotide sequence required for desulfurization activity in *Sphingomonas* sp. strain AD109.

Figures 2A, 2B and 2C together set forth the DNA sequence and the corresponding amino acid sequence of open reading frame 2 (ORF-2, *dszB*) of the nucleotide sequence required for desulfurization activity in *Sphingomonas* sp. strain AD109.

Figures 3A, 3B and 3C together set forth the DNA sequence and the corresponding amino acid sequence of open reading frame 3 (ORF-3, *dszC*) of the nucleotide sequence required for desulfurization activity in *Sphingomonas* sp. strain AD109.

Figure 4 is a graph showing the disappearance of 2-(2-phenyl)benzenesulfinate (HPBS) and the appearance of 2-hydroxybiphenyl (2-HBP) in the presence of *Sphingomonas* AD109 cell-free lysates.

Figure 5 shows a physical map of the *Sphingomonas dsz* gene cluster.

Figures 6A, 6B, 6C, 6D, 6E, 6F and 6G together set forth the nucleotide sequence of the *Sphingomonas dsz* gene cluster.

Figure 7 is a physical map of the plasmid pDA296.

Figure 8 presents the results of a GAP analysis of the DszA proteins from *Sphingomonas* sp. strain AD109 and *Rhodococcus* IGTS8.

Figure 9 presents the results of a GAP analysis of the DszB proteins from *Sphingomonas* sp. strain AD109 and *Rhodococcus* IGTS8.

Figure 10 presents the results of a GAP analysis of the sequences of the DszC proteins from *Sphingomonas* sp. strain AD109 and *Rhodococcus* IGTS8.

Figure 11 is a physical map of the plasmid pEBCtac.

Figure 12 is a graph of substrate concentration versus time for the desulfurization of DBT, 2,8-dimethyl-DBT and 4,6-dimethyl-DBT by a cell-free *Sphingomonas* AD109 lysate.

Figure 13 is a graph of substrate a dprodcut concentrations versus time for the desulfurization of DBT by a cell-free *Sphingomonas* AD109 lysate.

Figure 14 is a graph of product concentration versus time for the desulfurization of 2,8-dimethyl-DBT and 4,6-dimethyl-DBT by a cell-free *Sphingomonas* AD109 lysate.

Figure 15 is a graph of substrate concentration versus time for the desulfurization of DBT, 2,8-dimethyl-DBT and 4,6-dimethyl-DBT by a cell-free *Rhodococcus* lysate.

DETAILED DESCRIPTION OF THE INVENTION

The present invention is based on the discovery and isolation of a novel microorganism which is capable of selectively desulfurizing dibenzothiophene ("DBT"). As described in Example 1, this microorganism was obtained from soil samples obtained at sites contaminated with petroleum and petroleum by-products by a soil enrichment procedure using 2-(2-hydroxyphenyl)benzenesulfinate as the sole sulfur source. A biologically pure sample of the novel microorganism has been isolated and characterized. The microorganism is a motile, gram-negative rod. Based on a fatty acid analysis, as

described in Example 2, this microorganism has been identified as a *Sphingomonas* species, and designated strain AD-109. This microorganism has been deposited at the American Type Culture Collection (ATCC), 12301 Park Lawn Drive, Rockville, Maryland, U.S.A. 20852 under the terms of the Budapest Treaty and has been designated as ATCC Deposit No. 55954 on April 21, 1997.

The novel microorganism of the invention can be grown by fermentation under aerobic conditions in the presence of a sulfur-free mineral salts medium (e.g., 4 g/L K_2HPO_4, 4 g/L Na_2HPO_4, 2 g/L NH_4Cl, 0.2 g/L $MgCl_2.6H_2O$, 0.001 g/L $CaCl_2.2H_2O$, and 0.001 g/L $FeCl_3.6H_2O$), containing a sulfur-free source of assimilable carbon such as glucose. The sole source of sulfur provided can be a heterocyclic organosulfur compound, such as dibenzothiophene or a derivative thereof.

Sphingomonas sp. strain AD109 expresses a collection of enzymes which together catalyze the conversion of DBT to 2-hydroxybiphenyl (also referred to as "2-HBP") and inorganic sulfur. An enzyme which catalyzes one or more steps in this overall process is referred to herein as a "desulfurization enzyme". The nucleic acid sequence required for this overall process has been identified and cloned using the general method described in U.S. Patent No. 5,356,801, the contents of which are incorporated herein by reference, and is set forth in Figure 6 (SEQ ID NO.: 12). This nucleic acid sequence (also referred to as the "*Sphingomonas* dsz sequence") comprises three open reading frames, designated ORF-1 (base pairs 442-1800, also set forth in Figures 1A-1D and SEQ ID NO.: 1), ORF-2 (base pairs 1800-2909, also set forth in Figures 2A-2C and SEQ ID

NO.: 3) and ORF-3 (base pairs 2906-4141, sequence also set forth in Figures 3A-3C and SEQ ID NO.: 5). The predicted amino acid sequences encoded by these open reading frames are set forth in Figures 1A-1D (ORF-1, SEQ ID NO: 2), Figures 2A-2C (ORF-2, SEQ ID NO.: 4) and Figures 3A-3C (ORF-3, SEQ ID NO.: 6). Each of these open reading frames is homologous to the corresponding open reading frame of *Rhodococcus* sp. IGTS8; the sequences of the *Rhodococcus* open reading frames are disclosed in U.S. Patent No. 5,356,801.

In one embodiment, the present invention provides an isolated nucleic acid molecule comprising one or more nucleotide sequences which encode one or more of the biodesulfurization enzymes of *Sphingomonas* sp. strain AD109. The isolated nucleic acid molecule can be, for example, a nucleotide sequence, such as a deoxyribonucleic acid (DNA) sequence or a ribonucleic acid (RNA) sequence. Such a nucleic acid molecule comprises one or more nucleotide sequences which encode one or more of the amino acid sequences set forth in SEQ ID NO.: 2, SEQ ID NO.: 4, and SEQ ID NO.: 6. For example, the isolated nucleic acid molecule can comprise one or more of the nucleotide sequences of SEQ ID NO.: 1, SEQ ID NO.: 3, and SEQ ID NO.: 5, or a complement of any of these sequences. The isolated nucleic acid molecule can also comprise a nucleotide sequence which results from a silent mutation of one or more of the sequences set forth in SEQ ID NO.: 1, SEQ ID NO.: 3, and SEQ ID NO.: 5. Such a nucleotide sequence can result, for example, from a mutation of the native sequence in which one or more codons have been replaced with a degenerate codon, i.e., a codon which encodes the same

amino acid. Such mutant nucleotide sequences can be constructed using methods which are well known in the art, for example the methods discussed by Ausubel *et al.*, *Current Protocols in Molecular Biology*, Wiley-Interscience, New York (1997) (hereinafter "Ausubel *et al.*") and by Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, third edition, Cold Spring Harbor Laboratory Press (1992) (hereinafter "Sambrook *et al.*"), each of which are incorporated herein by reference.

In another embodiment, the invention includes an isolated nucleic acid molecule comprising a nucleotide sequence which is homologous to one or more of the sequences of SEQ ID NO.: 1, SEQ ID NO.: 3, and SEQ ID NO.: 5, or complements thereof. Such a nucleotide sequence exhibits at least about 80% homology, or sequence identity, with one of these *Sphingomonas* nucleotide sequences, preferably at least about 90% homology or sequence identity. Particularly preferred sequences have at least about 95% homology or have essentially the same sequence. Preparation of mutant nucleotide sequences can be accomplished by methods known in the art as are described in Old, *et al.*, *Principles of Gene Manipulation*, Fourth Edition, Blackwell Scientific Publications (1989), in Sambrook *et al.*, and in Ausubel *et al.*

The invention further includes nucleic acid molecules which are useful as hybridization probes, for example, for the isolation of the *Sphingomonas* genes encoding desulfurization enzymes or identical or homologous genes from other organisms. Such molecules comprise nucleotide sequences which hybridize to all or a portion of the nucleotide sequence of SEQ ID NO.: 1,

SEQ ID NO.: 3 or SEQ ID NO.: 5 or to non-coding regions immediately (within about 1000 nucleotides) 5' or 3' of each open reading frame. The invention also includes an isolated nucleic acid molecule which comprises a fragment of one or more of the nucleotide sequences set forth in SEQ ID NO.: 1, SEQ ID NO.: 3 or SEQ ID NO.: 5 or complements of any of these sequences. Such a fragment will generally comprise at least about 20 or at least about 40 contiguous nucleotides and, preferably, at least about 50 contiguous nucleotides of one of the disclosed sequences. Preferably, the hybridization probe of the invention hybridizes to one of these sequences under stringent conditions, such as those set forth by Sambrook *et al.* and Ausubel *et al.* For example, under conditions of high stringency, such as high temperatures and low salt concentrations, only DNA molecules which are essentially exact matches, or complements, will hybridize, particularly if the probe is relatively short. Hybridization under conditions of lower stringency, such as low temperatures, low formamide concentrations and high salt concentrations, allows greater mismatch between the probe and the target DNA molecule. It is particularly preferred that the nucleic acid molecule hybridizes selectively to the disclosed sequence(s).

The nucleic acid molecules can be synthesized chemically from the disclosed sequences. Alternatively, the nucleic acid molecules can be isolated from a suitable nucleic acid library (such as a DNA library) obtained from a microorganism which is believed to possess the nucleic acid molecule (such as, *Sphingomonas* sp. strain AD109), employing hybridizing primers and/or

probes designed from the disclosed sequences. Such a method can result in isolating the disclosed molecules (or spontaneous mutants thereof) for use in preparing recombinant enzymes, confirming the disclosed sequences, or for use in mutagenizing the native sequences.

In yet another embodiment, the nucleic acid molecule of the present invention can be a nucleic acid molecule, such as a recombinant DNA molecule, resulting from the insertion into its chain by chemical or biological means, of one or more of the nucleotide sequences described above. Recombinant DNA includes any DNA synthesized by procedures using restriction nucleases, nucleic acid hybridization, DNA cloning, DNA synthesis or any combination of the preceding. Methods of construction can be found in Sambrook *et al.* and Ausubel *et al.*, and additional methods are known by those skilled in the art.

The isolated nucleic acid molecule of the invention can further comprise a nucleotide sequence which encodes an oxidoreductase, such as a flavoprotein, such as a flavin reductase. For example, the nucleic acid molecule can encode an oxidoreductase which is native to *Sphingomonas* sp. strain AD109. The nucleic acid molecule can also encode the oxidoreductase denoted DszD described in copending U.S. Patent Application Serial No. 08/583,118; the flavin reductase from *Vibrio harveyii* described in copending U.S. Patent Application Serial No. 08/351,754; or the flavin reductase from *Rhodococcus* sp. IGTS8, described in copending U.S. Patent Application Serial No. 08/735,963. The contents of each of these applications are incorporated herein by reference.

The invention also includes a plasmid or vector comprising a recombinant DNA sequence or molecule which comprises one or more of the nucleic acid molecules, e.g. nucleotide sequences, of the invention, as described above. The terms "plasmid" and "vector" are intended to encompass any replication competent or replication incompetent plasmid or vector capable of having foreign or exogenous DNA inserted into it by chemical or biological means and subsequently, when transformed into an appropriate non-human host organism, of expressing the product of the foreign or exogenous DNA insert (e.g., of expressing the biocatalyst and flavoprotein of the present invention). In addition, the plasmid or vector is receptive to the insertion of a DNA molecule or fragment thereof containing the gene or genes of the present invention, said gene or genes encoding a biocatalyst as described herein. Procedures for the construction of DNA plasmid vectors include those described in Sambrook *et al.* and Ausubel *et al.* and others known by those skilled in the art.

The plasmids of the present invention include any DNA fragment containing a nucleotide sequence as described above. The DNA fragment should be transmittable, for example, to a host microorganism by transformation or conjugation. Procedures for the construction or extraction of DNA plasmids include those described in Sambrook *et al.* and Ausubel *et al.*, and others known by those skilled in the art. In one embodiment, the plasmid comprises a nucleotide sequence of the invention operatively linked to a competent or functional regulatory sequence. Examples of suitable regulatory sequences include promoters, enhancers,

transcription binding sites, ribosomal binding sites, transcription termination sequences, etc.

In one preferred embodiment, the regulatory or promoter sequences are those native to the *Sphingomonas* operon containing the genes disclosed herein. In yet another embodiment, one or more regulatory sequences (e.g. the promoter) is native to the selected host cell for expression. The promoter can be selected so that the gene or genes are inducible or constitutively expressed. Furthermore, the sequences can be regulated individually or together, as an operon. Examples of suitable promoters include the *E. coli lac* and *tac* promoters and the Pseudomonas P_G promoter (Yen, *J. Bacteriol.* **173** : 5328-5335 (1991)). An example of such a plasmid and its construction are described in Example 8.

In another embodiment, the invention relates to a recombinant or transformed non-human host organism which contains a heterologous DNA molecule of the invention as described above. The recombinant non-human host organisms of the present invention can be created by various methods by those skilled in the art. Any method for introducing a recombinant plasmid, such as a plasmid of the invention described above, into the organism of choice can be used, and a variety of such methods are described by Sambrook *et al.* and Ausubel *et al.* For example, the recombinant plasmid can be introduced via a suitable vector by transformation, conjugation, transduction or electroporation. By the term "non-human host organism" is intended any non-human organism capable of the uptake and expression of foreign, exogenous or recombinant DNA.

The recombinant microorganism can be derived from a host organism which does not express a native desulfurization biocatalyst. Such microorganisms include bacteria and yeasts, e.g., *E. coli, Bacillus,* and non-desulfurizing pseudomonads (as described in U. S. Patent Application Serial Number 08/851,088). In another embodiment, the recombinant microorganism is derived from a host organism which expresses a native biodesulfurization catalyst. Preferred microorganisms of this type are *Rhodococcus* sp. IGTS8 (ATCC 53968), recombinant microorganisms comprising one or more of the IGTS8 desulfurizing genes and *Sphingomonas* sp. strain AD109. Other desulfurizing microorganisms which are suitable host organisms include *Corynebacterium* sp. strain SY1, as disclosed by Omori *et al., Appl. Env. Microbiol.*, **58** : 911-915 (1992); *Rhodococcus erythropolis* D-1, as disclosed by Izumi *et al., Appl. Env. Microbiol.*, **60** :223-226 (1994); the *Arthrobacter* strain described by Lee *et al., Appl. Environ. Microbiol.* **61** : 4362-4366 (1995); the *Agrobacterium* strain disclosed by Constanti *et al., Enzyme Microb. Tech.* **19** : 214-219 (1996) and the *Rhodococcus* strains (ATCC 55309 and ATCC 55310) disclosed by Grossman *et al.*, U.S. Patent No. 5,607,857, each of which is incorporated herein by reference in its entirety. Each of these microorganisms produces one or more enzymes (protein biocatalysts) that catalyze one or more reactions in the desulfurization of DBT.

The invention also relates to desulfurization enzymes which can be isolated from *Sphingomonas* sp. strain AD109. These include desulfurization enzymes which catalyze one or more steps in the oxidative

desulfurization of DBT. The enzyme encoded by ORF-2 has been partially purified and exhibits 2-(2-hydroxyphenyl)benzenesulfinate (HPBS) desulfinase activity and has an apparent molecular weight by denaturing gel electrophoresis of about 40,000 daltons.

In one embodiment, the invention includes an isolated desulfurization enzyme from *Sphingomonas* sp. strain AD109 using methods and assays which are known the art, for example, the methods used by Gray *et al.* to isolate and characterize desulfurization enzymes from *Rhodococcus* IGTS8 (Gray *et al.*, *Nature Biotech.* **14** : 1705-1709 (1996)). These enzymes can be isolated or purified from the cell by lysing the cell and subjecting the cell lysate to known protein purification methods, and testing the fractions obtained thereby for the desired enzymatic activity. Examples of suitable protein purification methods include ammonium sulfate precipitation, ultrafiltration, diafiltration, immunoabsorption, anion exchange chromatography, gel filtration chromatography and hydrophobic interaction chromatography. The enzymes of the invention can also be recombinant proteins produced by heterologous expression of a nucleic acid molecule comprising a nucleotide sequence of SEQ ID NO.: 1, SEQ ID NO.: 3 or SEQ ID NO.: 5; or a mutation or fragment thereof, as discussed above. When the recombinant organism is derived from a non-*Sphingomonas* host, the recombinant proteins can be prepared in a form which is substantially free of other *Sphingomonas* proteins.

The invention also includes an isolated enzyme having an amino acid sequence which is homologous to the amino acid sequence of SEQ ID NO.: 2, SEQ ID NO.: 4 or

SEQ ID NO.: 6, or fragments thereof. The term "homologous" or "homologue", as used herein, describes a protein (which is not obtained from *Rhodococcus* or *Rhodococcus* sp IGTS8) having at least about 80% sequence identity or homology with the reference protein, and preferably about 90% sequence homology, in an amino acid alignment. Most preferably, the protein exhibits at least about 95% homology or essentially the same sequence as the disclosed sequence. An amino acid alignment of two or more proteins can be produced by methods known in the art, for example, using a suitable computer program, such as BLAST (Altschul *et al.*, *J. Mol. Biol.* **215** : 403-410 (1990)). A homologous protein can also have one or more additional amino acids appended at the carboxyl terminus or amino terminus, such as a fusion protein.

The homologous enzymes described herein can be native to an organism, such as a desulfurizing microorganism, including *Sphingomonas* sp. strain AD109 and mutants thereof. Such enzymes can be isolated from such sources using standard techniques and assays, as are described in the Exemplification and others known in the art. For example, the *Sphingomonas* desulfurization enzymes can be used to induce the formation of antibodies, such as monoclonal antibodies, according to known methods. The antibodies can then be used to purify the desulfurization enzymes from a desulfurizing organism via affinity chromatography, as is well known in the art.

The homologous enzymes of the invention can also be non-naturally occurring. For example, a homologous enzyme can be a mutant desulfurization enzyme which has

a modified amino acid sequence resulting from the deletion, insertion or substitution of one or more amino acid residues in the amino acid sequence of a *Sphingomonas* desulfurization enzyme. Such amino acid sequence variants can be prepared by methods known in the art. For example, the desired polypeptide can be synthesized *in vitro* using known methods of peptide synthesis. The amino acid sequence variants are preferably made by introducing appropriate nucleotide changes into a DNA molecule encoding the native enzyme, followed by expression of the mutant enzyme in an appropriate vector, such as *E. coli*. These methods include site-directed mutagenesis or random mutagenesis, for example.

Particularly preferred mutants include those having amino acid sequences which include the amino acid residues which are encoded by both SEQ ID NO.: 1, SEQ ID NO.: 3 or SEQ ID NO.: 5 and the corresponding open reading frame of *Rhodococcus* sp. IGTS8, as disclosed in U.S. Patent No. 5,356,801. That is, these mutants include the amino acid residues which are conserved in these two organisms in an amino acid alignment. Mutants which result from conservative substitution of one or more of these conserved residues, as well as non-conserved residues, are also included. Conservative and non-conservative substitutions (including deletions and insertions) can be made in non-conserved regions of the amino acid sequence and mutants resulting from both conservative and non-conservative substitutions of these residues are included herein.

Conservative substitutions are those in which a first amino acid residue is substituted by a second

residue having similar side chain properties. An example of such a conservative substitution is replacement of one hydrophobic residue, such as valine, with another hydrophobic residue, such as leucine. A non-conservative substitution involves replacing a first residue with a second residue having different side chain properties. An example of this type of substitution is the replacement of a hydrophobic residue, such as valine, with an acidic residue, such as glutamic acid.

The two primary variables in the construction of amino acid sequence variants are (1) the location of the mutation site and (2) the nature of the mutation. These variables can be manipulated to identify amino acid residues at the active site of the enzyme. For example, an amino acid substitution which yields a mutant enzyme having significantly different activity than the native enzyme suggests that the substituted amino acid residue is at the active site. Such mutants can have the same or similar, increased or decreased activity relative to that of the native enzyme.

Amino acids can be modified, for example, by substituting first with a conservative choice, followed by non-conservative choices depending upon the results achieved, by deleting the target residue(s) or by inserting residues adjacent to a particular site. Variants can also be constructed using a combination of these approaches.

The proteins of the present invention can be produced using techniques to overexpress the gene, as are described by Sambrook *et al.* and Ausubel *et al.* Improved expression, activity or overexpression of the

Sphingomonas desulfurization enzymes (in *Sphingomonas* sp AD 109 or in recombinant host cells harboring the disclosed nucleic acid molecules) can also be accomplished by mutagenesis. Suitable mutagens include radiation, e.g., ultraviolet radiation, and chemical mutagens, such as N-methyl-N'-nitroso-guanidine, hydroxylamine, ethylmethanesulfonate and nitrous acid. Furthermore, spontaneous mutants can be selected where the microorganism is subjected to an enrichment culture, as exemplified herein. The mutagenesis and subsequent screening for mutants harboring increased enzymatic activity can be conducted according to methods generally known in the art.

The present invention also provides a method of desulfurizing a carbonaceous material containing organosulfur molecules. The carbonaceous material can be, for example, a DBT-containing material or a fossil fuel, such as petroleum, a petroleum distillate fraction or coal. The method comprises the steps of (1) contacting the carbonaceous material with an aqueous phase containing a *Sphingomonas*-derived biocatalyst comprising at least one enzyme capable of catalyzing at least one step in the oxidative cleavage of carbon-sulfur bonds, thereby forming a carbonaceous material and aqueous phase mixture; (2) maintaining the mixture of step (1) under conditions sufficient for biocatalysis; and (3) separating the carbonaceous material having a reduced organic sulfur content from the resulting aqueous phase.

The term "*Sphingomonas*-derived biocatalyst", as used herein, is a biocatalyst which includes one or more desulfurization enzymes encoded by SEQ ID NO.: 1, SEQ ID

NO.: 3 and SEQ ID NO.: 5; or a mutant or homologue thereof. In one embodiment, the biocatalyst is a microorganism, such as *Sphingomonas sp.* strain AD109. The biocatalyst can also be a recombinant organism which contains one or more heterologous nucleotide sequences or nucleic acid molecules as described above.

Although living microorganisms (e.g., a culture) can be used as the biocatalyst herein, this is not required. Biocatalytic enzyme preparations that are useful in the present invention include microbial lysates, extracts, fractions, subfractions, or purified products obtained by conventional means and capable of carrying out the desired biocatalytic function. Generally, such enzyme preparations are substantially free of intact microbial cells. In a particularly preferred embodiment, the biocatalyst is overexpressed in the recombinant host cell (such as a cell which contains more than one copy of the gene or genes). Enzyme biocatalyst preparations suitable for use herein can optionally be affixed to a solid support, e.g., a membrane, filter, polymeric resin, glass particles or beads, or ceramic particles or beads. The use of immobilized enzyme preparations facilitates the separation of the biocatalyst from the treated fossil fuel which has been depleted of refractory organosulfur compounds.

A fossil fuel that is suitable for desulfurization treatment according to the present invention is one that contains organic sulfur. Such a fossil fuel is referred to as a "substrate fossil fuel". Substrate fossil fuels that are rich in thiophenic sulfur are particularly suitable for desulfurization according to the method

described herein. Examples of such substrate fossil fuels include Cerro Negro or Orinoco heavy crude oils; Athabascan tar and other types of bitumen; petroleum refining fractions such as gasoline, kerosene, diesel, fuel oil, residual oils and miscellaneous refinery by-products; shale oil and shale oil fractions; and coal-derived liquids manufactured from sources such as Pocahontas #3, Lewis-Stock, Australian Glencoe or Wyodak coal.

In the petroleum extraction and refining arts, the term "organic sulfur" is generally understood as referring to organic molecules having a hydrocarbon framework to which one or more sulfur atoms are covalently joined. These sulfur atoms can be directly bonded to the hydrocarbon framework, e.g., by one or more carbon-sulfur bonds, or can be present in a substituent bonded to the hydrocarbon framework of the molecule, e.g., a sulfate group. The general class of organic molecules having one or more sulfur heteroatoms are sometimes referred to as "organosulfur compounds". The hydrocarbon portion of these compounds can be aliphatic and/or aromatic.

Sulfur-bearing heterocycles, such as substituted and unsubstituted thiophene, benzothiophene, and dibenzothiophene, are known to be stable to conventional desulfurization treatments, such as hydrodesulfurization (HDS). Sulfur-bearing heterocycles can have relatively simple or relatively complex chemical structures. In complex heterocycles, multiple condensed aromatic rings, one or more of which can be heterocyclic, are present. The difficulty of desulfurization generally increases with the structural complexity of the molecule. That

is, refractory behavior is particularly accentuated in complex sulfur-bearing heterocycles, such as dibenzothiophene (DBT, $C_{12}H_8S$).

Much of the residual post-HDS organic sulfur in fossil fuel refining intermediates and combustible products is thiophenic sulfur. The majority of this residual thiophenic sulfur is present in DBT and derivatives thereof having one or more alkyl or aryl groups attached to one or more carbon atoms present in one or both flanking benzo rings. DBT itself is accepted as a model compound illustrative of the behavior of the class of compounds encompassing DBT and derivatives thereof in reactions involving thiophenic sulfur (Monticello and Finnerty, *Ann. Rev. Microbiol.*, **39** : 371-389 (1985)). DBT and derivatives thereof can account for a significant percentage of the total sulfur content of particular crude oils, coals and bitumen. For example, these sulfur-bearing heterocycles have been reported to account for as much as 70 wt% of the total sulfur content of West Texas crude oil, and up to 40 wt% of the total sulfur content of some Middle East crude oils. Thus, DBT is considered to be particularly relevant as a model compound for the forms of thiophenic sulfur found in fossil fuels, such as crude oils, coals or bitumen of particular geographic origin, and various refining intermediates and fuel products manufactured therefrom (Monticello and Finnerty (1985), *supra*). Another characteristic of DBT and derivatives thereof is that, following a release of fossil fuel into the environment, these sulfur-bearing heterocycles persist for long periods of time without significant biodegradation. Gundlach *et al.*, *Science* **221** : 122-129

(1983). Thus, most prevalent naturally occurring microorganisms do not effectively metabolize and break down sulfur-bearing heterocycles.

Biocatalytic desulfurization (biocatalysis or BDS) is the excision (liberation or removal) of sulfur from organosulfur compounds, including refractory organosulfur compounds such as sulfur-bearing heterocycles, as a result of the oxidative, preferably selective, cleavage of carbon-sulfur bonds in said compounds by a biocatalyst. BDS treatment yields the desulfurized combustible hydrocarbon framework of the former refractory organosulfur compound, along with inorganic sulfur substances which can be readily separated from each other by known techniques such as fractional distillation or water extraction. For example, DBT is converted into 2-hydroxybiphenyl when subjected to BDS treatment. A suitable biocatalyst for BDS comprises *Sphingomonas* sp. strain AD109 or an enzyme preparation derived therefrom, optionally, in combination with one or more additional non-human desulfurizing organisms (e.g., microorganisms); or an enzyme preparation derived from such an organism. Suitable additional desulfurizing organisms include those described above.

The specific activity of a given biocatalyst is a measure of its biocatalytic activity per unit mass. Thus, the specific activity of a particular biocatalyst depends on the nature or identity of the microorganism used or used as a source of biocatalytic enzymes, as well as the procedures used for preparing and/or storing the biocatalyst preparation. The concentration of a particular biocatalyst can be adjusted as desired for

use in particular circumstances. For example, where a culture of living microorganisms, such as *Sphingomonas* sp. strain AD109, is used as the biocatalyst preparation, a suitable culture medium lacking a sulfur source other than sulfur-bearing heterocycles can be inoculated with suitable microorganisms and grown until a desired culture density is reached. The resulting culture can be diluted with additional medium or another suitable buffer, or microbial cells present in the culture can be retrieved e.g., by centrifugation, and resuspended at a greater concentration than that of the original culture. The concentrations of microorganism and enzyme biocatalyst can be adjusted similarly. In this manner, appropriate volumes of biocatalyst preparations having predetermined specific activities and/or concentrations can be obtained.

In the biocatalytic desulfurization stage, the liquid fossil fuel containing sulfur-bearing heterocycles is combined with the biocatalyst. The relative amounts of biocatalyst and liquid fossil fuel can be adjusted to suit particular conditions, or to produce a particular level of residual sulfur in the treated, deeply desulfurized fossil fuel. The amount of biocatalyst preparation to be combined with a given quantity of liquid fossil fuel will reflect the nature, concentration and specific activity of the particular biocatalyst used, as well as the nature and relative abundance of inorganic and organic sulfur compounds present in the substrate fossil fuel and the degree of deep desulfurization sought or considered acceptable.

The method of desulfurizing a fossil fuel of the present invention involves two aspects. First, a host

organism or biocatalytic preparation obtained therefrom is contacted with a fossil fuel to be desulfurized. This can be done in any appropriate container, optionally fitted with an agitation or mixing device. The mixture is combined thoroughly and maintained or allowed to incubate for a sufficient time to allow for biocatalysis. In one embodiment, an aqueous emulsion or microemulsion is produced with an aqueous culture of the organism or enzyme fraction and the fossil fuel, allowing the organism to propagate in the emulsion while the expressed biocatalyst cleaves carbon-sulfur bonds.

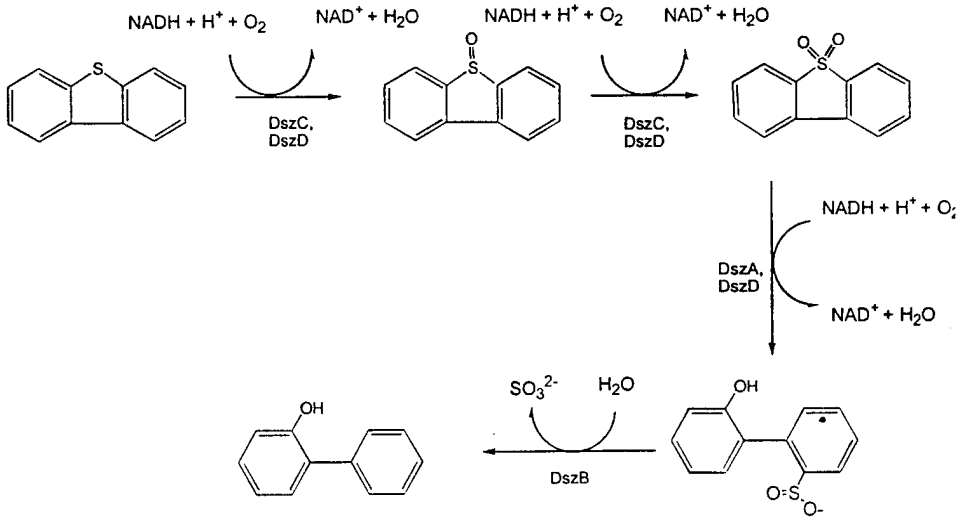
Variables such as temperature, pH, oxidation levels, mixing rate and rate of desulfurization will vary according to the nature of the biocatalyst used. Optimal parameters can generally be determined through no more than routine experimentation.

When the fossil fuel is a liquid hydrocarbon, such as petroleum, the desulfurized fossil fuel and the aqueous phase can form an emulsion. The components of such emulsions can be separated by a variety of methods, such as those described in U.S. Patent No. 5,358,870 and U.S. Patent Application Serial No. 08/640,129, which are incorporated herein by reference. For example, some emulsions reverse spontaneously when maintained under stationary conditions for a suitable period of time. Other emulsions can be reversed by adding an additional amount of an aqueous phase. Still other emulsions can be separated by the addition of a suitable chemical agent, such as a demulsifying agent or by employing suitable physical conditions, such as a particular temperature range.

The biocatalyst can be recovered from the aqueous phase, for example, by centrifugation, filtration or lyophilization. When the biocatalyst is a microorganism, the biocatalyst can be resuspended in fresh sulfur-free nutrient medium and/or any fresh microorganism culture as necessary to reconstitute or replenish to the desired level of biocatalytic activity. The biocatalyst can then be reintroduced into the reaction system.

Several suitable techniques for monitoring the rate and extent of desulfurization are well-known and readily available to those skilled in the art. Baseline and time course samples can be collected from the incubation mixture, and prepared for a determination of the residual organic sulfur in the fossil fuel. The disappearance of sulfur from organosulfur compounds, such as DBT, in the sample being subjected to biocatalytic treatment can be monitored using, e.g., X-ray fluorescence (XRF) or atomic emission spectrometry (flame spectrometry). Preferably, the molecular components of the sample are first separated, e.g., by gas chromatography.

Without being limited to any particular mechanism or theory, it is believed that the pathway of the desulfurization reaction in *Sphingomonas* sp. strain AD109 and other desulfurizing organisms, such as *Rhodococcus* sp. IGTS8, is set forth below:



Here the flavin reductase provides an electron transport chain which delivers, via $FMNH_2$, the reducing equivalents from NADH (or other electron donor) to the enzymes DszC and/or DszA. The enzyme DszC is responsible for the biocatalysis of the oxidation reaction of DBT to $DBTO_2$. The enzyme DszA is responsible for the reaction of $DBTO_2$ to 2-(2-hydroxyphenyl)benzenesulfinate (HPBS). The enzyme DszB catalyzes the conversion of HPBS to 2-hydroxybiphenyl and inorganic sulfur.

Another method of use of the *Sphingomonas* desulfurization enzymes, or mutants, homologues or active fragments thereof, is as a biocatalyst for the oxidation of organic compounds, such as substituted or unsubstituted dibenzothiophenes. The method comprises the steps of (1) contacting the organic compound with an aqueous phase containing a *Sphingomonas*-derived biocatalyst comprising at least one enzyme capable of

catalyzing at least one step in the oxidative cleavage of carbon-sulfur bonds, thereby forming an organic compound and aqueous phase mixture; (2) maintaining the mixture of step (1) under conditions sufficient for oxidation of the organic compound by the biocatalyst, thereby resulting in an oxidized organic compound, and, optionally, separating the oxidized organic compound from the aqueous phase. In one embodiment, the organic compound is a heteroorganic compound, such as an organonitrogen compound or an organosulfur compound. In one embodiment, the organic compound is an organosulfur compound which is a component of a fossil fuel, such as petroleum or a petroleum distillate fraction. In a second embodiment, the organic compound is a substituted or unsubstituted indole, as described in U.S. Provisional Patent Application Serial Number 60/020563, filed July 2, 1996, which is incorporated herein by reference.

The enzyme encoded by the nucleotide sequence of ORF-3 catalyzes the oxidation of dibenzothiophene to dibenzothiophene-5,5-dioxide (dibenzothiophene sulfone), and the enzyme encoded by the nucleotide sequence of ORF-1 catalyzes the oxidation of dibenzothiophene-5,5-dioxide to 2-(2-hydroxyphenyl)benzenesulfinate (also referred to as "HPBS"). In one embodiment the biocatalyst comprises the enzyme encoded by ORF-3, or a mutant, homologue or active fragment thereof; the organosulfur compound is substituted or unsubstituted dibenzothiophene; and the oxidized organosulfur is a substituted or unsubstituted dibenzothiophene-5,5-dioxide or dibenzothiophene-5-oxide (dibenzothiophene sulfoxide). In another embodiment the biocatalyst

comprises the enzymes encoded by ORF-1 and ORF-3, or a mutant, homologue or active fragment thereof; the organosulfur compound is a substituted or unsubstituted dibenzothiophene; and the oxidized organosulfur compound is a substituted or unsubstituted 2-(2-hydroxyphenyl)benzenesulfinate. In yet another embodiment, the biocatalyst comprises the enzyme encoded by ORF-1 or a mutant, homologue or active fragment thereof; the organosulfur compound is a substituted or unsubstituted dibenzothiophene-5,5-dioxide; and the oxidized organosulfur compound is a substituted or unsubstituted 2-(2-hydroxyphenyl)benzenesulfinate.

The oxidized organosulfur compound can, optionally, be further processed, for example, via a non-biological process or an enzyme-catalyzed reaction. In one embodiment, the oxidized organosulfur compound is desulfurized in a process employing suitable desulfurization enzymes from an organism other than a *Sphingomonas*.

The biocatalyst can be an organism, such as *Sphingomonas* sp. strain AD109, a desulfurizing mutant thereof, or a recombinant organism or enzyme preparation, as discussed above. When the organosulfur compound is a component of a fossil fuel, suitable reaction conditions and fossil fuel sources can be determined as described above.

The invention will now be further illustrated by the way of the following examples.

EXAMPLES

General Methods and Materials

Bacterial strains and plasmids

E. coli DH10ß (F⁻ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC)phi80d*lac*ZΔM15 Δ*lac*X74 *deo*R *rec*A1 *end*A1 *ara*Δ139 Δ(*ara, leu*)7697 *gal*U *gal*K lambda⁻ *rps*L *nup*G; Gibco-BRL, Gaithersburg, MD) was used as the cloning host. Plasmids pUC18 (Ap^R ; Vieria and Messing, *Gene* **19** : 259-268, (1982)), pOK12 (Km^R; Vieria and Messing, *Gene* **100** : 189-194 (1991)) and pSL1180 (Ap^R; Brosius, *DNA* **8** : 759, (1989)) were used as cloning vectors. Plasmid pEBCtac (Ap^R Tc^R *lacI*^{q} *tac*, shown in Figure 11, was used to overexpress the *Sphingomonas dszB* in *E. coli*.

Media and Reagents

Luria broth (LB) medium was routinely used to propagate *E. coli*. LB medium is 1% tryptone (Difco), 0.5% yeast extract (Difco) and 0.5% NaCl. Rich medium (RM) was used to propagate *Sphingomonas* strain AD109. RM medium is 0.8% nutrient broth, 0.05% yeast extract and 1% glucose. 2YT medium, used in gene expression studies, is 1.6% tryptone, 1% yeast extract and 0.5% NaCl. Basal salts medium (BSM-glucose) contained the following (per liter): phosphate buffer 100 mmol (pH 7.2); glucose, 20 g: NH_4Cl, 2 g; $MgCl_2 \cdot 6H_2O$, 644 mg; $MnCl_2 \cdot 4H_2O$, 1 mg; nitriloacetic acid, 0.1 g; $FeCl_2 \cdot 4H_2O$, 2.6 mg; $Na_2B_4O_7 \cdot 10H_2O$, 0.1 mg; $CuCl_2 \cdot 2H_2O$, 0.15 mg; $Co(NO_3)_2 \cdot 6H_2O$, 0.125 mg; $ZnCl_2$, 2.6 mg; $CaCl_2 \cdot 2H_2O$, 33 mg; $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$, 0.09 mg; and EDTA, 1.25 mg. When required the sulfur source was either 2 mM $MgSO_4$, 300 μM Dibenzothiophene (DBT), 300 μM Dibenzothiophene sulfone ($DBTO_2$) or 300 μM 2-(2-hydroxyphenyl) benzenesulfinate

(HPBS). For solid media, agar or agarose was added at a concentration of 1.5% (wt/wt). The antibiotic concentrations for *E. coli* were as follows: ampicillin, 100 μg/ml; kanamycin, 30 μg/ml; tetracycline, 10 μg/ml.

DNA Methods

Restriction enzymes and T4 DNA ligase were purchased from New England Biolabs, Inc. (Beverly, MA) and used as recommended by the supplier. Chromosomal DNA was isolated by the method described by Woo *et al.*, *BioTechniques* **13**: 696-698 (1992). Small scale plasmid preparations from *E. coli* were carried out as described by Birboim and Doly, *Nuc. Acids Res.* **7** : 1513-1523 (1979). Larger scale DNA preparations were carried out with Midi-prep columns from Qiagen (Chatsworth, CA). DNA fragments were purified from agarose gels after electrophoretic separation by the method of Vogelstein and Gillespie (*Proc. Natl. Acad. Sci. USA* **76**: 615-619 (1979). DNA fragments were cloned into vectors by using techniques described by Sambrook *et al.*

Degenerate oligonucleotide probes were end-labeled using standard digoxygenin protocols according to the Boehringer Mannheim DIG Oligonucleotide 3'-End Labeling Kit (Cat. No. 1362372). Hybridization was performed in 5X SSC with blocking solution containing 50% ultrapure deionized formamide at 42°C overnight (16 hr). Detection of hybrids was by enzyme immunoassay according to the Boehringer Mannheim Nonradioactive DIG DNA Labeling and Detection Kit (Cat. No. 1093657).

DNA samples were sequenced by SeqWright (Houston, TX) using a dye-terminator cycling sequencing kit from Perkin Elmer and the 373A and 377 ABI automatic DNA

sequencer. The sequence was extended by synthesizing overlapping oligonucleotides to previously read sequence. The synthesized oligonucleotides were used as primers for continuing sequence reactions. Sequencing reads were assembled and edited to 99.99% accuracy using Genecode's *Sequencher*, version 3.0 computer software.

DNA and protein sequence analysis was performed with the MacVector software program (Oxford Molecular Group, Campbell, CA). Nucleotide and amino acid sequences were compared to sequences in the available databases using BLAST. The Wisconsin Genetics Computer Group (GCG) software (Devereux *et al.*, *Nucl. Acids Res.* **12** : 387-395 (1984)) program GAP was used to generate comparisons of the protein sequences.

Transformation of *E. coli*

Plasmid DNA was introduced into *E. coli* DH10ß by electroporation. Competent ElectroMAX DH10ß (Gibco-BRL, Gaithersburg, MD) were used according to the manufacturer's suggestions.

Preparation of cell-free extracts

Cells grown in the appropriate medium were concentrated to an optical density at 600 nm of 50 by centrifugation and resuspended in 10 mM phosphate buffer (pH 7.0). Cells were disrupted in a French press and debris was removed by centrifugation at 32,000 x g for 20 min. Cell lysates were stored on ice at 4°C.

Desulfurization assays and analytical analysis

HPBS desulfinase activity was assayed by the ability of cell-free lysates to convert HPBS (substrate)

to 2-HBP (product) in a one hour assay at 30°C. The amounts of product made and substrate consumed during the reaction were quantitated by high-pressure liquid chromatography (HPLC) analysis. HPBS desulfinase activity was also measured by fluorescence spectroscopy. In a typical enzyme assay, enzyme activity is determined by the change in fluorescence at an excitation wavelength of 288 nm and an emission wavelength of 414 nm as HPBS is converted to 2-HBP. The assay is initiated by the addition of 20 - 100 μg total protein to a 3 mL solution of 200 μM HPBS in 50 mM phosphate buffer pH 7.5 containing 0.1 M NaCl.

Expression studies

E. coli DH10ß harboring the *Sphingomonas dszB* overexpression plasmid pDA296 was inoculated into 100ml of 2YT medium containing ampicillin and allowed to grow with shaking at 30°C. At an OD_{600} of approximately 0.3, the culture was divided into two parts. One half of the culture was induced by the addition of isopropylthio-β-galactoside (IPTG) (final conc. 1 mM) and the remaining culture was used as an uninduced control (no IPTG was added). Following incubation for an additional 3 hr, both cultures were harvested and cell-free lysates were prepared.

Protein purification and N-terminal sequencing

Sphingomonas AD109 cell paste was resuspended in an approximately equal weight of 25 mM phosphate buffer pH 7.5 containing 0.1 mM EDTA, 0.5 mM dithiothreitol (DTT), 10 μg/mL DNAse and 1 mM phenylmethylsulfonyl fluoride and passed through a French press mini-cell at about

20,000 psi. Cell debris was removed by centrifugation and the cell lysate was fractionated over an Econo-Pac High Q cartridge manufactured by Bio-Rad. A linear 0-0.5 M NaCl gradient was used to elute the bound protein into fractions. The active fractions were identified by a 2-HBP fluorescence enzyme assay (excitation/emission wavelengths set at 288/414 nm). The active fractions were pooled and desalted over a Bio-Rad P6 gel filtration cartridge, diluted to 1.7 M ammonium sulfate and fractionated over a Phenyl Superose HR 5/5 column manufactured by Pharmacia. A linear 1.7-0.0 M ammonium sulfate gradient was used to elute protein into fractions. Active fractions were identified and pooled as described above. Identity and purity of the AD109 HPBS desulfinase protein was also determined by SDS-PAGE and Western blots using antibodies generated against the DszB protein from *Rhodococcus erythroplis* strain IGTS8. N-terminal microsequencing of the HPBS desulfinase was carried out by Edman degradation after transfer of the purified protein to a polyvinylidene difluoride (PVDF) membrane.

SDS-PAGE and Western Blot Analysis

Protein separations were done with Novex (San Diego, CA) precast 10% polyacrylamide gels with Tris-Glycine- sodium dodecyl sulfate (SDS) (Laemmli) running buffer. Western blot analysis was carried out by first transferring the proteins electrophoretically to nitrocellulose membranes as recommended by Biorad (Hercules, CA). Blots were treated with antisera raised against the purified IGTS8 DszB protein (primary antibody) and then with goat anti-rabbit antisera

conjugated to horseradish peroxidase as the secondary antibody. Finally, the proteins were detected with a horseradish peroxidase catalyzed chemiluminescent reaction.

Example 1 Soil enrichments and isolation of a microorganism that can use HPBS as a sole sulfur source

Three independent soil samples from oil-contaminated sites were used to perform soil enrichments for microorganisms able to use HPBS as a sole sulfur source. Approximately 5 grams of each soil sample was placed into a sterile 250 ml flask along with 50 ml of BSM Glucose medium containing HPBS (300 μM) as the sole source of sulfur. Following incubation for 96 hrs at 30°C, a 3 ml sample of each enrichment was transferred to fresh BSM Glucose medium containing HPBS. After 72 hrs, one of the three flasks (flask #3) showed visible turbidity, while the two remaining flasks showed no visible increase in turbidity (even after more than a week of incubation). Microscopic analysis of the contents of flask #3 revealed the presence of a mixed population of bacterial cells (i.e., sessile and motile rods of varying shapes; large and small coccoid shaped bacteria). After repeated liquid subculture enrichments with HPBS as the sole sulfur source, the contents of the flask was plated onto several RM and LB agar plates. Following incubation at 30°C, a variety of microorganisms with different colony morphologies was present. Analysis of individual colonies from these plates identified a pure isolate that efficiently used

HPBS as a sole sulfur source. This strain, designated AD109, was selected for further analysis.

Example 2 Characterization and identification of strain AD109

The HPBS utilizing strain AD109 is a Gram-negative, motile rod that forms distinctive yellow colonies on agar plates. It grows somewhat poorly on LB agar, but grows rather well on RM agar plates. Like *Rhodococcus* IGTS8, strain AD109 also has the ability to produce clearing zones on a BSM Glucose DBT-sulfone plate. The optimal growth temperature of AD109 was found to be between 30 and 37°C.

Based on fatty acid analysis (Acculab, Inc., Newark, DE), this strain was identified as a *Sphingomonas* species. Strain AD109 was a "good" match to *S. paucimobilis* (formerly *Pseudomonas paucimobilis*) based on its "similarity index". The similarity index is a mathematical expression of the extent to which the fatty acid profile of a given unknown matches the mean profile for an organism in the TSBA database. Strain AD109 had an index value of 0.426 which indicates that it is from a strain of a species that differs significantly from those represented in the database. A similarity index of 0.5 or above is considered to be an "excellent" match (a value of 1.0 being the highest possible). On the other hand, an index below 0.3 indicates that the sample is from a species that is not likely to be in the database. Based on 16S rRNA sequence analysis and the presence of sphingoglycolipids, Yabuuchi *et al.* (*Microbiol. Immunol.*

34 : 99-119 (1990)) proposed that *P. paucimobilis* be reclassified and placed into the genus *Sphingomonas*.

Example 3 Growth characteristics of *Sphingomonas* species strain AD109

Evidence for the existence of an HPBS desulfinase activity was demonstrated by monitoring the supernatant of a AD109 culture growing in BSM Glucose HPBS (300 μM). By the time the culture was well into stationary phase all of the HPBS had been converted with no apparent accumulation of identifiable intermediates. There was, however, a transient production of a small amount of 2-HBP, as determined by HPLC analysis, which also disappeared with time. This preliminary result suggested that AD109 may also be capable of metabolizing 2-HBP. *Sphingomonas* strain AD109 was also capable of utilizing DBT-sulfone ($DBTO_2$) as a sole sulfur source. The ability to utilize DBT-sulfone as a sole sulfur source suggests that strain AD109 may also contain a gene that encodes DBT-sulfone monooxygenase activity.

During the course of growth studies it was discovered that strain AD109 could utilize DBT as a sole sulfur source. While growing with DBT, however, the culture supernatant takes on a very characteristic orange/brown color with an absorption maximum of approximately 470 nm. Orange-colored oxidation products have been previously identified in a number of *Pseudomonas* species that are capable of degrading DBT (Monticello *et al.*, *Appl. Environ. Microbiol.* **49** : 756-760 (1985)); Foght and Westlake, *Can. J. Microbiol.* **36** : 718-724 (1990)). No such color development was detected

in cultures growing with either HPBS or DBT-sulfone as sulfur sources.

Example 4 Demonstration of HPBS desulfinase activity in AD109 cell-free lysates

A cell-free lysate prepared from a culture of *Sphingomonas* strain AD109 (grown in BSM Glucose medium containing HPBS) was used in a time course study to examine the rate at which HPBS is converted to 2-HBP. As presented in Figure 4, at a protein concentration of 4 mg/ml there was a linear increase in 2-HBP production and a concomitant disappearance of HPBS.

The product of the *in vitro* reaction was confirmed to be 2-HBP by a spectral comparison to authentic 2-HBP. The ultraviolet absorption spectrum of the suspected 2-HBP peak produced by the action of the AD109 lysate is virtually identical with that of the 2-HBP standard. Furthermore, the molecular weight of the unknown compound was exactly that of authentic 2-HBP as determined by GC-MS analysis.

Example 5 Purification of the HPBS desulfinase from *Sphingomonas* AD109

HPBS desulfinase was purified from AD109 by a series of chromatographic steps using a Bio-Rad low pressure column chromatography Econo system and a Pharmacia FPLC (Gray *et al.*, *Nature Biotech.* **14** : 1705-1709 (1996)). The steps included fractionation over an anion exchange resin followed by a hydrophobic interaction column chromatography step. These protein

purification steps are described above. A 15-20 fold purification was achieved in these two steps which is comparable to protein preparations from a *Rhodococcus* IGTS8 lysate.

The molecular weight of this protein by SDS-PAGE was estimated to be 40,000 daltons, which is approximately the same size as DszB purified from IGTS8. Western analysis demonstrated that the purified protein shows some cross-reactivity with anti-DszB antisera.

Nonlinear regression analysis of an enzyme progress curve was performed according to the general method described by Duggleby, *Methods Enzymol.* **249** : 61-90 (1995). The analysis involves fitting the integrated Michealis-Menton rate equation $V_m^*t = y-K_m^*\ln(1-y/[A]_0)$ to concentration vs. time data from the enzyme catalyzed reaction of 2-(2-phenyl)benzenesulfinate to 2-hydroxybiphenyl monitored to completion by fluorescence.

The semi-pure protein sample was generated by fractionation of a crude lysate over Q Sepharose Fast Flow resin (Pharmacia) by a linear 0-0.5 M NaCl gradient, as discussed in more detail above. The purity of the active fraction was determined by SDS-PAGE. Pure enzyme is not necessary for the application of enzyme progress curve analysis, however, the calculation of k_{cat} ($V_m = [E]_t^*k_{cat}$) was limited to a value range as only a crude estimate of the enzyme concentration was available. The reaction conditions were as follows. A 3 mL reaction solution containing 1 μM HPBS and 0.1 M NaCl in 50 mM phosphate at pH 7.5 and 30°C was initiated by the addition of 0.023 mg total protein and was monitored for 30 min by fluorescence at an excitation wavelength of 288 nm and an emission wavelength of 414

nm. The data were fit to the equation using the Kaleidagraph data analysis/graphics application (Abelleck Software).

Based on the kinetic parameters calculated from the enzyme progress assay (K_m=0.3 μM and V_m=0.1 μM/min), the minimum k_{cat} = 0.5 min^{-1}. However, a more realistic value would be on the order of 2 min^{-1} in view of the fact that the preparation is estimated to be about 25% pure. Therefore, the HPBS desulfinase from *Sphingomonas* AD109 appears to be comparable to that from *Rhodococcus* IGTS8 with the possibility of a higher catalytic efficiency (k_{cat}/K_m).

The N-terminal amino acid sequence of the purified *Sphingomonas* HPBS desulfinase was also determined. Protein microsequencing using standard methods of analysis resulted in the following amino acid sequence:

1 10 20

TTDIHPASAA SSPAARATIT YS (SEQ ID NO.: 7)

A comparison of the putative AD109 HPBS desulfinase N-terminal sequence with that of the N-terminus of the IGTS8 DszB protein revealed that 9 out the 22 amino acid residues were identical (41%). In order to determine whether the purified protein is, in fact, the *Sphingomonas* desulfinase protein, a degenerate (192 permutations) 17-mer oligonucleotide probe with the following sequence: 5' ACN GAY ATH CAY CCN GC 3' (SEQ ID NO.: 8), was designed based on the determined N-terminal sequence. Following labeling with a

non-isotopic label this probe was used in hybridization studies using the cloned *Sphingomonas* AD109 HPBS desulfinase gene (see below) and the *dszB* gene from IGTS8 (Denome *et al.*, *J. Bacteriol.* **176** : 6707-6716 (1994); Piddington *et al.*, *App. Environ. Microbiol.* **61** : 468-475 (1995). The labeled oligonucleotide probe hybridized to the cloned *Sphingomonas* HPBS desulfinase gene which indicated that the correct protein had been purified. However, no signal was detected in the lane containing a fragment harboring the *Rhodococcus dsz* B gene.

Example 6 Cloning of the *Sphingomonas* AD109 HPBS desulfinase gene

Strain AD109 has been shown to be capable of using HPBS as a sole sulfur source and clearing a $DBTO_2$ plate. On the assumption that the gene(s) responsible for $DBTO_2$ clearing and HPBS desulfinase activity are genetically closely linked, as they are in *Rhodococcus* IGTS8, a cloning scheme was devised to isolate the HPBS desulfinase gene from *Sphingomonas* strain AD109. Total genomic DNA from strain AD109 was digested with either *Eco*RI, *Bam*HI, and *Hin*dIII and the resulting fragments were ligated into pUC18 or pSL1180. Following transformation of *E. coli* DH10ß, approximately 1000-2000 Lac-negative, ampicillin-resistant colonies of each library were screened for the ability to clear a $DBTO_2$ plate. No clearing colonies were detected amongst transformants derived from either the *Eco*RI or *Bam*HI libraries. However, two clearing colonies were detected utilizing the *Hin*dIII library and one clearing colony

was detected with the *Not*I library. Based on restriction endonuclease profiles, both colonies from the *Hind*III library contained the same large fragment (~20 kb). Furthermore, there was measurable HPBS desulfinase activity in cell-free lysates of these strains.

The single clearing colony from the *Not*I library contained a 6.5 kb fragment which, according to restriction endonuclease mapping, overlapped the 20 kb *Hind*III fragment. This clone also contained measurable HPBS desulfinase activity.

Subcloning analysis localized the genes responsible for $DBTO_2$ clearing and HPBS desulfinase activity to a 6 kb *Hind*III-*Not*I fragment. A smaller 2.7 kb *Hind*III-*Sma*I fragment was subsequently found to retain HPBS desulfinase activity, but lost the ability to clear a $DBTO_2$ plate. It is likely, therefore, that the gene that confers the ability to produce clearing zones on a DBT-sulfone plate spans the *Sma*I site.

Example 7 DNA sequence analysis of the *Sphingomonas* sp. strain AD109 desulfurization gene cluster

The nucleotide sequence of a 4144 bp region which encompasses the AD109 HPBS desulfinase gene was determined from both DNA strands and is present in Figure 6 (SEQ ID NO.: 12). The overall G+C content of the first 3837 base pairs of the AD109 sequence is 64.5%, a value which is consistent with the range of G+C values (61.7 - 67.2%) reported for various *Sphingomonas* species (Yabuuchi *et al.* (1990)). A

comparison of the AD109 nucleotide sequence with the IGTS8 *dsz* sequence by DNA matrix analysis revealed that a considerable amount of homology exists between the two sequences as evidenced by the presence of a near continuous diagonal line.

Open reading frame analysis of the AD109 sequence revealed the presence of a number of ORFs on both DNA strands, but of these, only three contained the codon-choice pattern characteristic of microorganisms with G-C rich genomes (West *et al.*, *Nucl. Acids Res.* **16:** 9323-9334 (1988)). All three identified ORFs were in the same transcriptional orientation. A strong preference for codons with either G or C occurred in positions 1 and 3. The first codon position of all three ORFs ranged from 67 to 72%, while the third codon position of all three ORFs ranged from 79-81%. In addition, the predicted translation initiation sites of all three ORFs are preceded by sequences that resemble a consensus ribosome binding site.

The entire nucleotide sequence of the AD109 region was used to conduct a BLAST search of the available DNA databases. The *Rhodococcus* IGTS8 *dsz* genes were the highest scoring sequences that demonstrated homology to the *Sphingomonas* sequences. The only other nucleotide sequence that demonstrated any significant homology to the *Sphingomonas* DNA, was the *Streptomyces pristinaespiralis snaA* gene which encodes the large subunit of the PII_A synthase (Blanc *et al.*, *J. Bacteriol.* **177** : 5206-5214 (1995)). The *Sphingomonas dszA* and *S. pristinaespiralis snaA* genes

demonstrate about 60% identity over a 800 bp region proximal to the 5′ end of each gene.

The first ORF (bp 442-1800; Figures 1A-1D) is 71% identical (at the nucleotide level) to the *Rhodococcus dszA* gene. The primary translation product of ORF-1 would encode a protein (*Sphingomonas* DszA or Dsz(S)) that contains 453 amino acids with a predicted molecular weight of 50,200. More importantly, this protein demonstrates considerable homology to the amino acid sequence of *Rhodococcus* DszA (Dsz(R), SEQ ID NO.: 9) over the entire length of the polypeptide (76% identity and 87% similarity; Figure 8). The protein databases were also searched with the *Sphingomonas* DszA protein sequences. Aside from the DszA protein of *Rhodococcus* IGTS8, several other proteins demonstrated significant homology to the *Sphingomonas* DszA protein. These include a hypothetical 49.3 kD protein in the IDH-DEOR intergenic region of *Bacillus subtilis* which showed 45% identity over 382 residues, the PII_A synthase SnaA subunit of *S. pristineaspiralis* (Blanc *et al.*, *J. Bacteriol.* **177** : 5206-5214 (1995)) which was 49% identical over 358 residues and the nitrilotriacetate monooxygenase of *Chelatobacter heintzii* (Xu *et al.*, Abstracts of the 95th General Meeting of the American Society for Microbiology, Q-281) which was 50% identical over the 335 residues examined.

The stop site of the *Sphingomonas* ORF-1 shows a 4-bp overlap with the translation start site of the second ORF (bp 1800-2906; Figures 2A-2C), which shows a high degree of homology to the *Rhodococcus* IGTS8 *dszB* gene (67% identity). It was determined that the

primary translation product of ORF-2 would encode a 369-amino acid polypeptide with a predicted molecular weight of 40,000 (*Sphingomonas* DszB or Dsz(S)). At the amino acid level this putative protein is 66% identical (75% similarity) to the *Rhodococcus* HPBS desulfinase protein DszB (DszB(R), SEQ ID NO: 10), as shown in Figure 9. Except for the IGTS8 DszB protein, a BLAST search with the *Sphingomonas* DszB sequence did not identify any other significant homologous sequences in the available databases. The predicted N-terminus of the *Sphingomonas* DszB protein matches identically the N-terminus of the HPBS desulfinase purified from AD109 cell lysates, except that the amino-terminal methionine was absent. Removal of the methionine residue has been shown to occur when the second amino acid is Ala, Ser, Gly, Pro, Thr or Val (Hirel *et al.*, *Proc. Nat. Acad. Sci. USA* 86 : 8247-8251 (1989)).

The stop site of the *Sphingomonas dszB* gene also shows a 4-bp overlap with the translation start site of the third ORF. This ORF (bp 2906-4141; Figures 3A-3C), shows significant homology to the *Rhodococcus* IGTS8 *dszC* gene. For example, over the first 931 bp, this ORF is 69% identical to the IGTS8 *dszC* gene and the N-terminus of the protein predicted by this sequence (*Sphingomonas* DszC, DszC(S)) is 67% identical to the N-terminus of *Rhodococcus* DszC (DszC(R), SEQ ID NO: 11), as shown in Figure 10. A BLAST search of the protein databases with the available *Sphingomonas* DszC sequence identified a number of proteins in addition to the IGTS8 DszC protein. The *Sphingomonas* DszC protein is 32% identical (over 199 residues) to

Isobutylamine *N*-Hydroxylase (IBAH) of *Streptomyces viridifaciens*. It has previously been shown that IBAH exhibits the greatest similarity to the IGTS8 DszC protein (Parry *et al.*, *J. Bacteriol.*, **179**: 409-416 (1997)). In addition, the AD109 DszC protein showed variable homology to a number of acyl coenzyme A dehydrogenases. For example, the N-terminal 300 residues of the *Sphingomonas* DszC protein is 29% identical to the acyl CoA dehydrogenase of *B. subtilis*.

The sequences (400 bp) directly upstream of the *dszA* start site contain regulatory elements (i.e., promoter elements) that control transcription of the AD109 *dsz* gene cluster. A comparison of this potential promoter region with the IGTS8 *dsz* promoter region failed to reveal any significant homology. It has been shown that the IGTS8 *dsz* promoter region encompasses a region of potential diad symmetry that may contain an operator (Li *et al.*, *J. Bacteriol.* **178** : 6409-6418 (1996)). An examination of the AD109 sequences directly upstream of *dszA* revealed no such palindromic sequence.

Example 8 Expression of the *Sphingomonas dszB* gene in *E. coli*

The AD109 *dszB* gene was subcloned into the *tac* promoter expression vector, pEBCtac, in two steps. The first step involved cloning a 1.2 kb *Pst*I-*Bgl*II fragment that contained the entire coding region of the AD109 *dszB* gene (Figures 2A-2C) into the polylinker plasmid pOK12. The resulting plasmid,

designated pDA295, contained a unique *Xba*I site upstream of the *dszB* gene. In the second step, a 1.2 kb *Xba*I-*Bgl*II fragment from pDA295 that contained the entire *dszB* gene was cloned into the *Xba*I and *Bgl*II sites of pEBCtac, thus placing the AD109 *dszB* gene under the transcriptional control of the *tac* promoter. This plasmid, designated pDA296 and presented in Figure 7, was introduced into *E. coli* DH10ß for expression studies.

HPBS desulfinase assays (2 mg/ml protein) using cell-free lysates prepared from induced and uninduced cultures of DH10ß/pDA296 were performed. In the absence of IPTG the cell-free lysate contained very little HPBS desulfinase activity. Only 22 nmoles of 2-HBP were produced during the 60 min. incubation period which corresponds to a specific activity of 0.2 (nmoles 2-HBP formed/min/mg protein). The lysate prepared from the IPTG-induced culture, however, had approximately 20 times more HPBS desulfinase activity (4.2 nmoles 2-HBP formed/min/mg protein) than the lysate prepared from the uninduced culture.

Example 9 Desulfurization of DBT and alkylated derivatives by AD109 cell-free lysates

To a cell free *Sphingomonas* AD109 lysate having a total protein concentration of 10 mg/mL were added NADH (4 mM) and FMN (10 μM). The lysate was then treated with either DBT, 2,8-diMeDBT or 4,6-diMeDBT at a concentration of approximately 90 μM and maintained at 37°C. Aliquots were removed from the reaction mixture at approximately 10 minute time intervals, and

the substrate and product concentrations of each aliquot were determined using high performance liquid chromatography. A similar set of experiments was conducted using a cell-free lysate of a *Rhodococcus* strain which expresses the *Rhodococcus* ATCC 53968 DszA, DszB and DszC enzymes.

The results of substrate consumption by the *Sphingomonas* AD109 lysate are presented in Figures 12-14. The time dependence of substrate concentration for each of the three substrates is illustrated in Figure 12, which shows that 4,6-diMeDBT is more rapidly consumed than the other two substrates, which disappear at similar rates. Figure 13 indicates that the concentration of 2-HBP, the expected product of DBT desulfurization, increases as the DBT concentration decreases. Figure 14 shows the time dependence of product formation for both 4,6-diMeDBT (product: 2-(2-hydroxy-3-methylphenyl)-6-methylbenzenesulfinate (4,6-dimethyl HBP)) and 2,8-diMeDBT (product: 2-(2-hydroxy-6-methylphenyl)-3-methylbenzenesulfinate (2,8-dimethyl HBP)). The desulfurization product of 4,6-diMeDBT is formed more rapidly than the product resulting from 2,8-diMeDBT.

Figure 15 shows the time dependence of substrate disappearance in similar experiments with the *Rhodococcus* cell-free lysate. In this case, DBT and 2,8-diMeDBT are consumed at similar rates, while 4,6-diMeDBT is consumed at a much slower rate.

The results indicate that the *Sphingomonas* and *Rhodococcus* desulfurization enzymes have different substrate preferences. In particular, the *Sphingomonas* AD109 lysate desulfurizes 4,6-diMeDBT, in

which the sulfur atom is sterically hindered by the adjacent methyl groups, more rapidly than the unhindered 2,8-diMeDBT and DBT. *Rhodococcus* shows the opposite preference, desulfurizing the unhindered substrates significantly more rapidly than 4,6-diMeDBT.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

(A) NAME: Energy BioSystems Corporation

(B) STREET: 4200 Research Forest Drive

(C) CITY: The Woodlands

(D) STATE/PROVINCE: Texas

(E) COUNTRY: USA

(F) POSTAL CODE/ZIP: 77381

(G) TELEPHONE: (281) 364-6100

(i) TELEFAX: (281) 364-6112

(ii) TITLE OF INVENTION: A Sphingomonas Biodesulfurization Catalyst

(iii) NUMBER OF SEQUENCES: 13

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

(B) STREET: Two Militia Drive

(C) CITY: Lexington

(D) STATE: Massachusetts

(E) COUNTRY: USA

(F) ZIP: 02173

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: PC-DOS/MS-DOS

(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 08/851,089

(B) FILING DATE: 05-MAY-1997

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 08/835,292

(B) FILING DATE: 07-APR-1997

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Elmore, Carolyn S.

(B) REGISTRATION NUMBER: 37,567

(C) REFERENCE/DOCKET NUMBER: EBC97-06A2

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (781) 861-6240

(B) TELEFAX: (781) 861-9540

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1362 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..1359

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATG ACC GAT CCA CGT CAG CTG CAC CTG GCC GGA TTC TTC TGT GCC GGC 48
Met Thr Asp Pro Arg Gln Leu His Leu Ala Gly Phe Phe Cys Ala Gly
1 5 10 15

```
AAC GTC ACG CAC GCC CAC GGA GCG TGG CGC CAC GCC GAC GAC TCC AAC     96
Asn Val Thr His Ala His Gly Ala Trp Arg His Ala Asp Asp Ser Asn
             20                  25                  30

GGC TTC CTC ACC AAG GAG TAC TAC CAG CAG ATT GCC CGC ACG CTC GAG    144
Gly Phe Leu Thr Lys Glu Tyr Tyr Gln Gln Ile Ala Arg Thr Leu Glu
         35                  40                  45

CGC GGC AAG TTC GAC CTG CTG TTC CTT CCC GAC GCG CTC GCC GTG TGG    192
Arg Gly Lys Phe Asp Leu Leu Phe Leu Pro Asp Ala Leu Ala Val Trp
     50                  55                  60

GAC AGC TAC GGC GAC AAT CTG GAG ACC GGT CTG CGG TAT GGC GGG CAA    240
Asp Ser Tyr Gly Asp Asn Leu Glu Thr Gly Leu Arg Tyr Gly Gly Gln
 65                  70                  75                  80

GGC GCG GTG ATG CTG GAG CCC GGC GTA GTT ATC GCC GCG ATG GCC TCG    288
Gly Ala Val Met Leu Glu Pro Gly Val Val Ile Ala Ala Met Ala Ser
                 85                  90                  95

GTG ACC GAA CAT CTG GGG CTG GGC GCC ACC ATT TCC ACC ACC TAC TAC    336
Val Thr Glu His Leu Gly Leu Gly Ala Thr Ile Ser Thr Thr Tyr Tyr
            100                 105                 110

CCG CCC TAC CAT GTA GCC CGG GTC GTC GCT TCG CTG GAC CAG CTG TCC    384
Pro Pro Tyr His Val Ala Arg Val Val Ala Ser Leu Asp Gln Leu Ser
        115                 120                 125

TCC GGG CGA GTG TCG TGG AAC GTG GTC ACC TCG CTC AGC AAT GCA GAG    432
Ser Gly Arg Val Ser Trp Asn Val Val Thr Ser Leu Ser Asn Ala Glu
    130                 135                 140

GCG CGC AAC TTC GGC TTC GAT GAA CAT CTC GAC CAC GAT GCC CGC TAC    480
Ala Arg Asn Phe Gly Phe Asp Glu His Leu Asp His Asp Ala Arg Tyr
145                 150                 155                 160
```

GAT CGC GCC GAT GAA TTC CTC GAG GTC GTG CGC AAG CTC TGG AAC AGC 528
Asp Arg Ala Asp Glu Phe Leu Glu Val Val Arg Lys Leu Trp Asn Ser
165 170 175

TGG GAT CGC GAT GCG CTG ACA CTC GAC AAG GCA ACC GGC CAG TTC GCC 576
Trp Asp Arg Asp Ala Leu Thr Leu Asp Lys Ala Thr Gly Gln Phe Ala
180 185 190

GAT CCG GCT AAG GTG CGC TAC ATC GAC CAC CGC GGC GAA TGG CTC AAC 624
Asp Pro Ala Lys Val Arg Tyr Ile Asp His Arg Gly Glu Trp Leu Asn
195 200 205

GTA CGC GGG CCG CTT CAG GTG CCG CGC TCC CCC CAG GGC GAG CCT GTC 672
Val Arg Gly Pro Leu Gln Val Pro Arg Ser Pro Gln Gly Glu Pro Val
210 215 220

ATT CTG CAG GCC GGG CTT TCG GCG CGG GGC AAG CGC TTC GCC GGG CGC 720
Ile Leu Gln Ala Gly Leu Ser Ala Arg Gly Lys Arg Phe Ala Gly Arg
225 230 235 240

TGG GCG GAC GCG GTG TTC ACG ATT TCG CCC AAT CTG GAC ATC ATG CAG 768
Trp Ala Asp Ala Val Phe Thr Ile Ser Pro Asn Leu Asp Ile Met Gln
245 250 255

GCC ACG TAC CGC GAC ATA AAG GCG CAG GTC GAG GCC GCC GGA CGC GAT 816
Ala Thr Tyr Arg Asp Ile Lys Ala Gln Val Glu Ala Ala Gly Arg Asp
260 265 270

CCC GAG CAG GTC AAG GTG TTT GCC GCG GTG ATG CCG ATC CTC GGC GAG 864
Pro Glu Gln Val Lys Val Phe Ala Ala Val Met Pro Ile Leu Gly Glu
275 280 285

ACC GAG GCG ATC GCC AGG CAG CGT CTC GAA TAC ATA AAT TCG CTG GTG 912
Thr Glu Ala Ile Ala Arg Gln Arg Leu Glu Tyr Ile Asn Ser Leu Val
290 295 300

```
CAT CCC GAA GTC GGG CTT TCT ACG TTG TCC AGC CAT GTC GGG GTC AAC    960
His Pro Glu Val Gly Leu Ser Thr Leu Ser Ser His Val Gly Val Asn
305                 310                 315                 320

CTT GCC GAC TAT TCG CTC GAT ACC CCG CTG ACC GAG GTC CTG GGC GAT   1008
Leu Ala Asp Tyr Ser Leu Asp Thr Pro Leu Thr Glu Val Leu Gly Asp
                325                 330                 335

CTC GCC CAG CGC AAC GTG CCC ACC CAA CTG GGC ATG TTC GCC AGG ATG   1056
Leu Ala Gln Arg Asn Val Pro Thr Gln Leu Gly Met Phe Ala Arg Met
            340                 345                 350

TTG CAG GCC GAG ACG CTG ACC GTG GGA GAA ATG GGC CGG CGT TAT GGC   1104
Leu Gln Ala Glu Thr Leu Thr Val Gly Glu Met Gly Arg Arg Tyr Gly
        355                 360                 365

GCC AAC GTG GGC TTC GTC CCG CAG TGG GCG GGA ACC CGC GAG CAG ATC   1152
Ala Asn Val Gly Phe Val Pro Gln Trp Ala Gly Thr Arg Glu Gln Ile
    370                 375                 380

GCG GAC CTG ATC GAG ATC CAT TTC AAG GCC GGC GGC GCC GAT GGC TTC   1200
Ala Asp Leu Ile Glu Ile His Phe Lys Ala Gly Gly Ala Asp Gly Phe
385                 390                 395                 400

ATC ATC TCG CCG GCG TTC CTG CCC GGA TCT TAC GAG GAA TTC GTC GAT   1248
Ile Ile Ser Pro Ala Phe Leu Pro Gly Ser Tyr Glu Glu Phe Val Asp
                405                 410                 415

CAG GTG GTG CCC ATC CTG CAG CAC CGC GGA CTG TTC CGC ACT GAT TAC   1296
Gln Val Val Pro Ile Leu Gln His Arg Gly Leu Phe Arg Thr Asp Tyr
            420                 425                 430

GAA GGC CGC ACC CTG CGC AGC CAT CTG GGA CTG CGT GAA CCC GCA TAC   1344
Glu Gly Arg Thr Leu Arg Ser His Leu Gly Leu Arg Glu Pro Ala Tyr
        435                 440                 445
```

CTG GGA GAG TAC GCA TGA 1362
Leu Gly Glu Tyr Ala
450

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 453 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Thr Asp Pro Arg Gln Leu His Leu Ala Gly Phe Phe Cys Ala Gly
1 5 10 15

Asn Val Thr His Ala His Gly Ala Trp Arg His Ala Asp Asp Ser Asn
20 25 30

Gly Phe Leu Thr Lys Glu Tyr Tyr Gln Gln Ile Ala Arg Thr Leu Glu
35 40 45

Arg Gly Lys Phe Asp Leu Leu Phe Leu Pro Asp Ala Leu Ala Val Trp
50 55 60

Asp Ser Tyr Gly Asp Asn Leu Glu Thr Gly Leu Arg Tyr Gly Gly Gln
65 70 75 80

Gly Ala Val Met Leu Glu Pro Gly Val Val Ile Ala Ala Met Ala Ser
85 90 95

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

Pro Pro Tyr His Val Ala Arg Val Val Ala Ser Leu Asp Gln Leu Ser
115 120 125

Ser Gly Arg Val Ser Trp Asn Val Val Thr Ser Leu Ser Asn Ala Glu
130 135 140

Ala Arg Asn Phe Gly Phe Asp Glu His Leu Asp His Asp Ala Arg Tyr
145 150 155 160

Asp Arg Ala Asp Glu Phe Leu Glu Val Val Arg Lys Leu Trp Asn Ser
165 170 175

Trp Asp Arg Asp Ala Leu Thr Leu Asp Lys Ala Thr Gly Gln Phe Ala
180 185 190

Asp Pro Ala Lys Val Arg Tyr Ile Asp His Arg Gly Glu Trp Leu Asn
195 200 205

Val Arg Gly Pro Leu Gln Val Pro Arg Ser Pro Gln Gly Glu Pro Val
210 215 220

Ile Leu Gln Ala Gly Leu Ser Ala Arg Gly Lys Arg Phe Ala Gly Arg
225 230 235 240

Trp Ala Asp Ala Val Phe Thr Ile Ser Pro Asn Leu Asp Ile Met Gln
245 250 255

Ala Thr Tyr Arg Asp Ile Lys Ala Gln Val Glu Ala Ala Gly Arg Asp
260 265 270

Pro Glu Gln Val Lys Val Phe Ala Ala Val Met Pro Ile Leu Gly Glu
275 280 285

Thr Glu Ala Ile Ala Arg Gln Arg Leu Glu Tyr Ile Asn Ser Leu Val
290 295 300

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

Leu Ala Asp Tyr Ser Leu Asp Thr Pro Leu Thr Glu Val Leu Gly Asp
325 330 335

Leu Ala Gln Arg Asn Val Pro Thr Gln Leu Gly Met Phe Ala Arg Met
340 345 350

Leu Gln Ala Glu Thr Leu Thr Val Gly Glu Met Gly Arg Arg Tyr Gly
355 360 365

Ala Asn Val Gly Phe Val Pro Gln Trp Ala Gly Thr Arg Glu Gln Ile
370 375 380

Ala Asp Leu Ile Glu Ile His Phe Lys Ala Gly Gly Ala Asp Gly Phe
385 390 395 400

Ile Ile Ser Pro Ala Phe Leu Pro Gly Ser Tyr Glu Glu Phe Val Asp
405 410 415

Gln Val Val Pro Ile Leu Gln His Arg Gly Leu Phe Arg Thr Asp Tyr
420 425 430

Glu Gly Arg Thr Leu Arg Ser His Leu Gly Leu Arg Glu Pro Ala Tyr
435 440 445

Leu Gly Glu Tyr Ala
450

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1110 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..1107

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
ATG ACG ACA GAC ATC CAC CCG GCG AGC GCC GCA TCG TCG CCG GCG GCG     48
Met Thr Thr Asp Ile His Pro Ala Ser Ala Ala Ser Ser Pro Ala Ala
 1               5                  10                  15

CGC GCG ACG ATC ACC TAC AGC AAC TGC CCC GTG CCT AAT GCC CTG CTC     96
Arg Ala Thr Ile Thr Tyr Ser Asn Cys Pro Val Pro Asn Ala Leu Leu
            20                  25                  30

GCC GCG CTC GGC TCA GGT ATT CTG GAC AGT GCC GGG ATC ACA CTT GCC    144
Ala Ala Leu Gly Ser Gly Ile Leu Asp Ser Ala Gly Ile Thr Leu Ala
        35                  40                  45

CTG CTG ACC GGA AAG CAG GGC GAG GTG CAC TTC ACC TAC GAC CGA GAT    192
Leu Leu Thr Gly Lys Gln Gly Glu Val His Phe Thr Tyr Asp Arg Asp
    50                  55                  60

GAC TAC ACC CGC TTC GGC GGC GAG ATT CCG CCG CTG GTC AGC GAG GGA    240
Asp Tyr Thr Arg Phe Gly Gly Glu Ile Pro Pro Leu Val Ser Glu Gly
65                  70                  75                  80

CTG CGT GCG CCG GGG CGG ACC CGC CTG CTG GGA CTG ACG CCG GTG CTG    288
Leu Arg Ala Pro Gly Arg Thr Arg Leu Leu Gly Leu Thr Pro Val Leu
                85                  90                  95

GGC CGC TGG GGC TAC TTC GTC CGG GGC GAC AGC GCG ATC CGC ACC CCG    336
Gly Arg Trp Gly Tyr Phe Val Arg Gly Asp Ser Ala Ile Arg Thr Pro
            100                 105                 110

GCC GAT CTT GCC GGC CGC CGC GTC GGA GTA TCC GAT TCG GCC AGG AGG    384
Ala Asp Leu Ala Gly Arg Arg Val Gly Val Ser Asp Ser Ala Arg Arg
        115                 120                 125
```

```
ATA TTG ACC GGA AGG CTG GGC GAC TAC CGC GAA CTT GAT CCC TGG CGG    432
Ile Leu Thr Gly Arg Leu Gly Asp Tyr Arg Glu Leu Asp Pro Trp Arg
    130                 135                 140

CAG ACC CTG GTC GCG CTG GGG ACA TGG GAG GCG CGT GCC TTG CTG AGC    480
Gln Thr Leu Val Ala Leu Gly Thr Trp Glu Ala Arg Ala Leu Leu Ser
145                 150                 155                 160

ACG CTC GAG ACG GCG GGG CTT GGC GTC GGC GAC GTC GAG CTG ACG CGC    528
Thr Leu Glu Thr Ala Gly Leu Gly Val Gly Asp Val Glu Leu Thr Arg
                165                 170                 175

ATC GAG AAC CCG TTC GTC GAC GTG CCG ACC GAA CGA CTG CAT GCC GCC    576
Ile Glu Asn Pro Phe Val Asp Val Pro Thr Glu Arg Leu His Ala Ala
            180                 185                 190

GGC TCG CTC AAA GGA ACC GAC CTG TTC CCC GAC GTG ACC AGC CAG CAG    624
Gly Ser Leu Lys Gly Thr Asp Leu Phe Pro Asp Val Thr Ser Gln Gln
        195                 200                 205

GCC GCA GTC CTT GAG GAT GAG CGC GCC GAC GCC CTG TTC GCG TGG CTT    672
Ala Ala Val Leu Glu Asp Glu Arg Ala Asp Ala Leu Phe Ala Trp Leu
    210                 215                 220

CCC TGG GCG GCC GAG CTC GAG ACC CGC ATC GGT GCA CGG CCG GTC CTA    720
Pro Trp Ala Ala Glu Leu Glu Thr Arg Ile Gly Ala Arg Pro Val Leu
225                 230                 235                 240

GAC CTC AGC GCA GAC GAC CGC AAT GCC TAT GCG AGC ACC TGG ACG GTG    768
Asp Leu Ser Ala Asp Asp Arg Asn Ala Tyr Ala Ser Thr Trp Thr Val
                245                 250                 255

AGC GCC GAG CTG GTG GAC CGG CAG CCC GAA CTG GTG CAG CGG CTC GTC    816
Ser Ala Glu Leu Val Asp Arg Gln Pro Glu Leu Val Gln Arg Leu Val
            260                 265                 270
```

```
GAT GCC GTG GTG GAT GCA GGG CGG TGG GCC GAG GCC AAT GGC GAT GTC    864
Asp Ala Val Val Asp Ala Gly Arg Trp Ala Glu Ala Asn Gly Asp Val
        275                 280                 285

GTC TCC CGC CTG CAC GCC GAT AAC CTC GGT GTC AGT CCC GAA AGC GTC    912
Val Ser Arg Leu His Ala Asp Asn Leu Gly Val Ser Pro Glu Ser Val
    290                 295                 300

CGC CAG GGA TTC GGA GCC GAT TTT CAC CGC CGC CTG ACG CCG CGG CTC    960
Arg Gln Gly Phe Gly Ala Asp Phe His Arg Arg Leu Thr Pro Arg Leu
305                 310                 315                 320

GAC AGC GAT GCT ATC GCC ATC CTG GAG CGT ACT CAG CGG TTC CTG AAG   1008
Asp Ser Asp Ala Ile Ala Ile Leu Glu Arg Thr Gln Arg Phe Leu Lys
                325                 330                 335

GAT GCG AAC CTG ATC GAT CGG TCG TTG GCG CTC GAT CGG TGG GCT GCA   1056
Asp Ala Asn Leu Ile Asp Arg Ser Leu Ala Leu Asp Arg Trp Ala Ala
            340                 345                 350

CCT GAA TTC CTC GAA CAA AGT CTC TCA CGC CAG GTC GAA GGG CAG ATA   1104
Pro Glu Phe Leu Glu Gln Ser Leu Ser Arg Gln Val Glu Gly Gln Ile
        355                 360                 365

GCA TGA                                                          1110
Ala
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 369 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Thr Thr Asp Ile His Pro Ala Ser Ala Ala Ser Ser Pro Ala Ala
1 5 10 15

Arg Ala Thr Ile Thr Tyr Ser Asn Cys Pro Val Pro Asn Ala Leu Leu
20 25 30

Ala Ala Leu Gly Ser Gly Ile Leu Asp Ser Ala Gly Ile Thr Leu Ala
35 40 45

Leu Leu Thr Gly Lys Gln Gly Glu Val His Phe Thr Tyr Asp Arg Asp
50 55 60

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

Leu Arg Ala Pro Gly Arg Thr Arg Leu Leu Gly Leu Thr Pro Val Leu
85 90 95

Gly Arg Trp Gly Tyr Phe Val Arg Gly Asp Ser Ala Ile Arg Thr Pro
100 105 110

Ala Asp Leu Ala Gly Arg Arg Val Gly Val Ser Asp Ser Ala Arg Arg
115 120 125

Ile Leu Thr Gly Arg Leu Gly Asp Tyr Arg Glu Leu Asp Pro Trp Arg
130 135 140

Gln Thr Leu Val Ala Leu Gly Thr Trp Glu Ala Arg Ala Leu Leu Ser
145 150 155 160

Thr Leu Glu Thr Ala Gly Leu Gly Val Gly Asp Val Glu Leu Thr Arg
165 170 175

Ile Glu Asn Pro Phe Val Asp Val Pro Thr Glu Arg Leu His Ala Ala
180 185 190

Gly Ser Leu Lys Gly Thr Asp Leu Phe Pro Asp Val Thr Ser Gln Gln
195 200 205

Ala Ala Val Leu Glu Asp Glu Arg Ala Asp Ala Leu Phe Ala Trp Leu
210 215 220

Pro Trp Ala Ala Glu Leu Glu Thr Arg Ile Gly Ala Arg Pro Val Leu
225 230 235 240

Asp Leu Ser Ala Asp Asp Arg Asn Ala Tyr Ala Ser Thr Trp Thr Val
245 250 255

Ser Ala Glu Leu Val Asp Arg Gln Pro Glu Leu Val Gln Arg Leu Val
260 265 270

Asp Ala Val Val Asp Ala Gly Arg Trp Ala Glu Ala Asn Gly Asp Val
275 280 285

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

Arg Gln Gly Phe Gly Ala Asp Phe His Arg Arg Leu Thr Pro Arg Leu
305 310 315 320

Asp Ser Asp Ala Ile Ala Ile Leu Glu Arg Thr Gln Arg Phe Leu Lys
325 330 335

Asp Ala Asn Leu Ile Asp Arg Ser Leu Ala Leu Asp Arg Trp Ala Ala
340 345 350

Pro Glu Phe Leu Glu Gln Ser Leu Ser Arg Gln Val Glu Gly Gln Ile
355 360 365

Ala

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1236 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1236

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
ATG AAC GAA CTC GTC AAA GAT CTC GGC CTC AAT CGA TCC GAT CCG ATC     48
Met Asn Glu Leu Val Lys Asp Leu Gly Leu Asn Arg Ser Asp Pro Ile
 1               5                  10                  15

GGC GCT GTG CGG CGA CTG GCC GCG CAG TGG GGG GCC ACC GCT GTT GAT     96
Gly Ala Val Arg Arg Leu Ala Ala Gln Trp Gly Ala Thr Ala Val Asp
            20                  25                  30

CGG GAC CGG GCC GGC GGA TCG GCA ACC GCC GAA CTC GAT CAA CTG CGC    144
Arg Asp Arg Ala Gly Gly Ser Ala Thr Ala Glu Leu Asp Gln Leu Arg
        35                  40                  45

GGC AGC GGC CTG CTC TCG CTG TCC ATT CCC GCC GCA TAT GGC GGC TGG    192
Gly Ser Gly Leu Leu Ser Leu Ser Ile Pro Ala Ala Tyr Gly Gly Trp
    50                  55                  60

GGC GCC GAC TGG CCA ACG ACT CTG GAA GTT ATC CGC GAA GTC GCA ACG    240
Gly Ala Asp Trp Pro Thr Thr Leu Glu Val Ile Arg Glu Val Ala Thr
65                  70                  75                  80
```

```
GTG GAC GGA TCG CTG GCG CAT CTA TTC GGC TAC CAC CTC GGC TGC GTA    288
Val Asp Gly Ser Leu Ala His Leu Phe Gly Tyr His Leu Gly Cys Val
                85                  90                  95

CCG ATG ATC GAG CTG TTC GGC TCG GCG CCA CAA AAG GAA CGG CTG TAC    336
Pro Met Ile Glu Leu Phe Gly Ser Ala Pro Gln Lys Glu Arg Leu Tyr
            100                 105                 110

CGC CAG ATC GCA AGC CAT GAT TGG CGG GTC GGG AAT GCG TCG AGC GAA    384
Arg Gln Ile Ala Ser His Asp Trp Arg Val Gly Asn Ala Ser Ser Glu
        115                 120                 125

AAC AAC AGC CAC GTG CTC GAG TGG AAG CTT GCC GCC ACC GCC GTC GAT    432
Asn Asn Ser His Val Leu Glu Trp Lys Leu Ala Ala Thr Ala Val Asp
    130                 135                 140

GAT GGC GGG TTC GTC CTC AAC GGC GCG AAG CAC TTC TGC AGC GGC GCC    480
Asp Gly Gly Phe Val Leu Asn Gly Ala Lys His Phe Cys Ser Gly Ala
145                 150                 155                 160

AAA AGC TCC GAC CTG CTC ATC GTG TTC GGC GTG ATC CAG GAC GAA TCC    528
Lys Ser Ser Asp Leu Leu Ile Val Phe Gly Val Ile Gln Asp Glu Ser
                165                 170                 175

CCC CTG CGC GGC GCG ATC ATC ACC GCG GTC ATT CCC ACC GAC CGG GCC    576
Pro Leu Arg Gly Ala Ile Ile Thr Ala Val Ile Pro Thr Asp Arg Ala
            180                 185                 190

GGT GTT CAG ATC AAT GAC GAC TGG CGC GCA ATC GGG ATG CGC CAG ACC    624
Gly Val Gln Ile Asn Asp Asp Trp Arg Ala Ile Gly Met Arg Gln Thr
        195                 200                 205

GAC AGC GGC AGC GCC GAA TTT CGC GAC GTC CGA GTC TAC CCA GAC GAG    672
Asp Ser Gly Ser Ala Glu Phe Arg Asp Val Arg Val Tyr Pro Asp Glu
    210                 215                 220
```

```
ATC TTG GGG GCA CCA AAC TCA GTC GTT GAG GCG TTC GTG ACA AGC AAC    720
Ile Leu Gly Ala Pro Asn Ser Val Val Glu Ala Phe Val Thr Ser Asn
225                 230                 235                 240

CGC GGC AGC CTG TGG ACG CCG GCG ATT CAG TCG ATC TTC TCG AAC GTT    768
Arg Gly Ser Leu Trp Thr Pro Ala Ile Gln Ser Ile Phe Ser Asn Val
                245                 250                 255

TAT CTG GGG CTC GCG CGT GGC GCG CTC GAG GCG GCA GCG GAT TAC ACC    816
Tyr Leu Gly Leu Ala Arg Gly Ala Leu Glu Ala Ala Ala Asp Tyr Thr
            260                 265                 270

CGG ACC CAG AGC CGC CCC TGG ACA CCC GCC GGC GTG GCG AAG GCG ACA    864
Arg Thr Gln Ser Arg Pro Trp Thr Pro Ala Gly Val Ala Lys Ala Thr
        275                 280                 285

GAG GAT CCC CAC ATC ATC GCC ACC TAC GGT GAA CTG GCG ATC GCG CTC    912
Glu Asp Pro His Ile Ile Ala Thr Tyr Gly Glu Leu Ala Ile Ala Leu
    290                 295                 300

CAG GGC GCC GAG GCG GCC GCG CGC GAG GTC GCG GCC CTG TTG CAA CAG    960
Gln Gly Ala Glu Ala Ala Ala Arg Glu Val Ala Ala Leu Leu Gln Gln
305                 310                 315                 320

GCG TGG GAC AAG GGC GAT GCG GTG ACG CCC GAA GAG CGC GGC CAG CTG   1008
Ala Trp Asp Lys Gly Asp Ala Val Thr Pro Glu Glu Arg Gly Gln Leu
                325                 330                 335

ATG GTG AAG GTT TCG GGT GTG AAG GCC CTC TCG ACG AAG GCC GCC CTC   1056
Met Val Lys Val Ser Gly Val Lys Ala Leu Ser Thr Lys Ala Ala Leu
            340                 345                 350

GAC ATC ACC AGC CGT ATT TTC GAG ACA ACG GGC TCG CGA TCG ACG CAT   1104
Asp Ile Thr Ser Arg Ile Phe Glu Thr Thr Gly Ser Arg Ser Thr His
        355                 360                 365
```

```
CCC AGA TAC GGA TTC GAT CGG TTC TGG CGT AAC ATC CGG ACT CAT ACG   1152
Pro Arg Tyr Gly Phe Asp Arg Phe Trp Arg Asn Ile Arg Thr His Thr
    370                 375                 380

CTG CAC GAT CCG GTA TCG TAT AAA ATC GTC GAT GTG GGG AAC TAC ACG   1200
Leu His Asp Pro Val Ser Tyr Lys Ile Val Asp Val Gly Asn Tyr Thr
385                 390                 395                 400

CTC AAC GGG ACA TTC CCG GTT CCC GGA TTT ACG TCA                   1236
Leu Asn Gly Thr Phe Pro Val Pro Gly Phe Thr Ser
                405                 410
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 412 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Met Asn Glu Leu Val Lys Asp Leu Gly Leu Asn Arg Ser Asp Pro Ile
 1               5                  10                  15

Gly Ala Val Arg Arg Leu Ala Ala Gln Trp Gly Ala Thr Ala Val Asp
            20                  25                  30

Arg Asp Arg Ala Gly Gly Ser Ala Thr Ala Glu Leu Asp Gln Leu Arg
        35                  40                  45

Gly Ser Gly Leu Leu Ser Leu Ser Ile Pro Ala Ala Tyr Gly Gly Trp
    50                  55                  60

Gly Ala Asp Trp Pro Thr Thr Leu Glu Val Ile Arg Glu Val Ala Thr
65                  70                  75                  80
```

Val Asp Gly Ser Leu Ala His Leu Phe Gly Tyr His Leu Gly Cys Val
85 90 95

Pro Met Ile Glu Leu Phe Gly Ser Ala Pro Gln Lys Glu Arg Leu Tyr
100 105 110

Arg Gln Ile Ala Ser His Asp Trp Arg Val Gly Asn Ala Ser Ser Glu
115 120 125

Asn Asn Ser His Val Leu Glu Trp Lys Leu Ala Ala Thr Ala Val Asp
130 135 140

Asp Gly Gly Phe Val Leu Asn Gly Ala Lys His Phe Cys Ser Gly Ala
145 150 155 160

Lys Ser Ser Asp Leu Leu Ile Val Phe Gly Val Ile Gln Asp Glu Ser
165 170 175

Pro Leu Arg Gly Ala Ile Ile Thr Ala Val Ile Pro Thr Asp Arg Ala
180 185 190

Gly Val Gln Ile Asn Asp Asp Trp Arg Ala Ile Gly Met Arg Gln Thr
195 200 205

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

Ile Leu Gly Ala Pro Asn Ser Val Val Glu Ala Phe Val Thr Ser Asn
225 230 235 240

Arg Gly Ser Leu Trp Thr Pro Ala Ile Gln Ser Ile Phe Ser Asn Val
245 250 255

Tyr Leu Gly Leu Ala Arg Gly Ala Leu Glu Ala Ala Ala Asp Tyr Thr
260 265 270

Arg Thr Gln Ser Arg Pro Trp Thr Pro Ala Gly Val Ala Lys Ala Thr
275 280 285

Glu Asp Pro His Ile Ile Ala Thr Tyr Gly Glu Leu Ala Ile Ala Leu
290 295 300

Gln Gly Ala Glu Ala Ala Ala Arg Glu Val Ala Ala Leu Leu Gln Gln
305 310 315 320

Ala Trp Asp Lys Gly Asp Ala Val Thr Pro Glu Glu Arg Gly Gln Leu
325 330 335

Met Val Lys Val Ser Gly Val Lys Ala Leu Ser Thr Lys Ala Ala Leu
340 345 350

Asp Ile Thr Ser Arg Ile Phe Glu Thr Thr Gly Ser Arg Ser Thr His
355 360 365

Pro Arg Tyr Gly Phe Asp Arg Phe Trp Arg Asn Ile Arg Thr His Thr
370 375 380

Leu His Asp Pro Val Ser Tyr Lys Ile Val Asp Val Gly Asn Tyr Thr
385 390 395 400

Leu Asn Gly Thr Phe Pro Val Pro Gly Phe Thr Ser
405 410

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Thr Thr Asp Ile His Pro Ala Ser Ala Ala Ser Ser Pro Ala Ala Arg
1 5 10 15

Ala Thr Ile Thr Tyr Ser
20

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

ACNGAYATHC AYCCNGC 17

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 453 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met Thr Gln Gln Arg Gln Met His Leu Ala Gly Phe Phe Ser Ala Gly
1 5 10 15

Asn Val Thr His Ala His Gly Ala Trp Arg His Thr Asp Ala Ser Asn
20 25 30

Asp Phe Leu Ser Gly Lys Tyr Tyr Gln His Ile Ala Arg Thr Leu Glu
35 40 45

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

Asp Ser Tyr Gly Asp Asn Leu Asp Thr Gly Val Gly Leu Gly Gly Gln
65 70 75 80

Gly Ala Val Ala Leu Glu Pro Ala Ser Val Val Ala Thr Met Ala Ala
85 90 95

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

Pro Pro Tyr His Val Ala Arg Val Phe Ala Thr Leu Asp Gln Leu Ser
115 120 125

Gly Gly Arg Val Ser Trp Asn Val Val Thr Ser Leu Asn Asp Ala Glu
130 135 140

Ala Arg Asn Phe Gly Ile Asn Gln His Leu Glu His Asp Ala Arg Tyr
145 150 155 160

Asp Arg Ala Asp Glu Phe Leu Glu Ala Val Lys Lys Leu Trp Asn Ser
165 170 175

Trp Asp Glu Asp Ala Leu Val Leu Asp Lys Ala Ala Gly Val Phe Ala
180 185 190

Asp Pro Ala Lys Val His Tyr Val Asp His His Gly Glu Trp Leu Asn
195 200 205

Val Arg Gly Pro Leu Gln Val Pro Arg Ser Pro Gln Gly Glu Pro Val
210 215 220

Ile Leu Gln Ala Gly Leu Ser Pro Arg Gly Arg Arg Phe Ala Gly Lys
225 230 235 240

Trp Ala Glu Ala Val Phe Ser Leu Ala Pro Asn Leu Glu Val Met Gln
245 250 255

Ala Thr Tyr Gln Gly Ile Lys Ala Glu Val Asp Ala Ala Gly Arg Asp
260 265 270

Pro Asp Gln Thr Lys Ile Phe Thr Ala Val Met Pro Val Leu Gly Glu
275 280 285

Ser Gln Ala Val Ala Gln Glu Arg Leu Glu Tyr Leu Asn Ser Leu Val
290 295 300

His Pro Glu Val Gly Leu Ser Thr Leu Ser Ser His Thr Gly Ile Asn
305 310 315 320

Leu Ala Ala Tyr Pro Leu Asp Thr Pro Ile Lys Asp Ile Leu Arg Asp
325 330 335

Leu Gln Asp Arg Asn Val Pro Thr Gln Leu His Met Phe Ala Ala Ala
340 345 350

Thr His Ser Glu Glu Leu Thr Leu Ala Glu Met Gly Arg Arg Tyr Gly
355 360 365

Thr Asn Val Gly Phe Val Pro Gln Trp Ala Gly Thr Gly Glu Gln Ile
370 375 380

Ala Asp Glu Leu Ile Arg His Phe Glu Gly Gly Ala Ala Asp Gly Phe
385 390 395 400

Ile Ile Ser Pro Ala Phe Leu Pro Gly Ser Tyr Asp Glu Phe Val Asp
405 410 415

Gln Val Val Pro Val Leu Gln Asp Arg Gly Tyr Phe Arg Thr Glu Tyr
420 425 430

Gln Gly Asn Thr Leu Arg Asp His Leu Gly Leu Arg Val Pro Gln Leu
435 440 445

Gln Gly Gln Pro Ser
450

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 365 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Met Thr Ser Arg Val Asp Pro Ala Asn Pro Gly Ser Glu Leu Asp Ser
1 5 10 15

Ala Ile Arg Asp Thr Leu Thr Tyr Ser Asn Cys Pro Val Pro Asn Ala
20 25 30

Leu Leu Thr Ala Ser Glu Ser Gly Phe Leu Asp Ala Ala Gly Ile Glu
35 40 45

Leu Asp Val Leu Ser Gly Gln Gln Gly Thr Val His Phe Thr Tyr Asp
50 55 60

Gln Pro Ala Tyr Thr Arg Phe Gly Gly Glu Ile Pro Pro Leu Leu Ser
65 70 75 80

Glu Gly Leu Arg Ala Pro Gly Arg Thr Arg Leu Leu Gly Ile Thr Pro
85 90 95

Leu Leu Gly Arg Gln Gly Phe Phe Val Arg Asp Asp Ser Pro Ile Thr
100 105 110

Ala Ala Ala Asp Leu Ala Gly Arg Arg Ile Gly Val Ser Ala Ser Ala
115 120 125

Ile Arg Ile Leu Arg Gly Gln Leu Gly Asp Tyr Leu Glu Leu Asp Pro
130 135 140

Trp Arg Gln Thr Leu Val Ala Leu Gly Ser Trp Glu Ala Arg Ala Leu
145 150 155 160

Leu His Thr Leu Glu His Gly Glu Leu Gly Val Asp Asp Val Glu Leu
165 170 175

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

Glu Ser Ala Thr Val Lys Gly Ala Asp Leu Phe Pro Asp Val Ala Arg
195 200 205

Gly Gln Ala Ala Val Leu Ala Ser Gly Asp Val Asp Ala Leu Tyr Ser
210 215 220

Trp Leu Pro Trp Ala Gly Glu Leu Gln Ala Thr Gly Ala Arg Pro Val
225 230 235 240

Val Asp Leu Gly Leu Asp Glu Arg Asn Ala Tyr Ala Ser Val Trp Thr
245 250 255

Val Ser Ser Gly Leu Val Arg Gln Arg Pro Gly Leu Val Gln Arg Leu
260 265 270

Val Asp Ala Ala Val Asp Ala Gly Leu Trp Ala Arg Asp His Ser Asp
275 280 285

Ala Val Thr Ser Leu His Ala Ala Asn Leu Gly Val Ser Thr Gly Ala
290 295 300

Val Gly Gln Gly Phe Gly Ala Asp Phe Gln Gln Arg Leu Val Pro Arg
305 310 315 320

Leu Asp His Asp Ala Leu Ala Leu Leu Glu Arg Thr Gln Gln Phe Leu
325 330 335

Leu Thr Asn Asn Leu Leu Gln Glu Pro Val Ala Leu Asp Gln Trp Ala
340 345 350

Ala Pro Glu Phe Leu Asn Asn Ser Leu Asn Arg His Arg
355 360 365

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 417 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Met Thr Leu Ser Pro Glu Lys Gln His Val Arg Pro Arg Asp Ala Ala
1 5 10 15

Asp Asn Asp Pro Val Ala Val Ala Arg Gly Leu Ala Glu Lys Trp Arg
20 25 30

Ala Thr Ala Val Glu Arg Asp Arg Ala Gly Gly Ser Ala Thr Ala Glu
35 40 45

Arg Glu Asp Leu Arg Ala Ser Ala Leu Leu Ser Leu Leu Val Pro Arg
50 55 60

Glu Tyr Gly Gly Trp Gly Ala Asp Trp Pro Thr Ala Ile Glu Val Val
65 70 75 80

Arg Glu Ile Ala Ala Ala Asp Gly Ser Leu Gly His Leu Phe Gly Tyr
85 90 95

His Leu Thr Asn Ala Pro Met Ile Glu Leu Ile Gly Ser Gln Glu Gln
100 105 110

Glu Glu His Leu Tyr Thr Gln Ile Ala Gln Asn Asn Trp Trp Thr Gly
115 120 125

Asn Ala Ser Ser Glu Asn Asn Ser His Val Leu Asp Trp Lys Val Ser
130 135 140

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

Phe Cys Ser Gly Ala Lys Gly Ser Asp Leu Leu Phe Val Phe Gly Val
165 170 175

Val Gln Asp Asp Ser Pro Gln Gln Gly Ala Ile Ile Ala Ala Ala Ile
180 185 190

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

Gly Met Arg Gln Thr Asp Ser Gly Ser Thr Asp Phe His Asn Val Lys
210 215 220

Val Glu Pro Asp Glu Val Leu Gly Ala Pro Asn Ala Phe Val Leu Ala
225 230 235 240

Phe Ile Gln Ser Glu Arg Gly Ser Leu Phe Ala Pro Ile Ala Gln Leu
245 250 255

Ile Phe Ala Asn Val Tyr Leu Gly Ile Ala His Gly Ala Leu Asp Ala
260 265 270

Ala Arg Glu Tyr Thr Arg Thr Gln Ala Arg Pro Trp Thr Pro Ala Gly
275 280 285

Ile Gln Gln Ala Thr Glu Asp Pro Tyr Thr Ile Arg Ser Tyr Gly Glu
290 295 300

Phe Thr Ile Ala Leu Gln Gly Ala Asp Ala Ala Ala Arg Glu Ala Ala
305 310 315 320

His Leu Leu Gln Thr Val Trp Asp Lys Gly Asp Ala Leu Thr Pro Glu
325 330 335

Asp Arg Gly Glu Leu Met Val Lys Val Ser Gly Val Lys Ala Leu Ala
340 345 350

Thr Asn Ala Ala Leu Asn Ile Ser Ser Gly Val Phe Glu Val Ile Gly
355 360 365

Ala Arg Gly Thr His Pro Arg Tyr Gly Phe Asp Arg Phe Trp Arg Asn
370 375 380

Val Arg Thr His Ser Leu His Asp Pro Val Ser Tyr Lys Ile Ala Asp
385 390 395 400

Val Gly Lys His Thr Leu Asn Gly Gln Tyr Pro Ile Pro Gly Phe Thr
405 410 415

Ser

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4144 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GGTTCGAGAT CGATCTGACC GTCGAACCCG GCGCGGTTCA AACCATCCTC TGGGGCCTCT 60

TCTTGCACTT GACATAGGAA TCTCTACTAA ATAAATAGAT ATTTATTCGA CACTAAGTTC 120

GGTGATCAGG CCGACCGTGT GTCTCAAGTG CTCGCTCCGG GTTGCCACGA GCTAAAGCGC 180

GCGATGCTGG GGCGACAGCG CTAGGCATTG CGTTCCCTCA CACCAATGAT GAGATGATAC 240

GATGCGCATG ACCACTATCC GCACCTAGCA CGAAAGATCC GTGCATTTCG CGAATGCCAA 300

TGAAGAGGAC CGACGTACGG CAGCTTCCTA CGCTTTCGCG CCATCGTTCA TAGCCAAGGT 360

CTTTTCGACG CCGGTTCGCG TGGGCGACTG ACGGCGGTAG CGCCGCGACT ATTCGTTTCA 420

AACTCACGAG GATAAGAGCC TATGACCGAT CCACGTCAGC TGCACCTGGC CGGATTCTTC 480

TGTGCCGGCA ACGTCACGCA CGCCCACGGA GCGTGGCGCC ACGCCGACGA CTCCAACGGC 540

TTCCTCACCA AGGAGTACTA CCAGCAGATT GCCCGCACGC TCGAGCGCGG CAAGTTCGAC 600

CTGCTGTTCC TTCCCGACGC GCTCGCCGTG TGGGACAGCT ACGGCGACAA TCTGGAGACC 660

GGTCTGCGGT ATGGCGGGCA AGGCGCGGTG ATGCTGGAGC CCGGCGTAGT TATCGCCGCG 720

ATGGCCTCGG TGACCGAACA TCTGGGGCTG GGCGCCACCA TTTCCACCAC CTACTACCCG 780

CCCTACCATG TAGCCCGGGT CGTCGCTTCG CTGGACCAGC TGTCCTCCGG GCGAGTGTCG 840

TGGAACGTGG TCACCTCGCT CAGCAATGCA GAGGCGCGCA ACTTCGGCTT CGATGAACAT 900

CTCGACCACG ATGCCCGCTA CGATCGCGCC GATGAATTCC TCGAGGTCGT GCGCAAGCTC 960

TGGAACAGCT GGGATCGCGA TGCGCTGACA CTCGACAAGG CAACCGGCCA GTTCGCCGAT 1020

CCGGCTAAGG TGCGCTACAT CGACCACCGC GGCGAATGGC TCAACGTACG CGGGCCGCTT 1080

CAGGTGCCGC GCTCCCCCCA GGGCGAGCCT GTCATTCTGC AGGCCGGGCT TTCGGCGCGG 1140

GGCAAGCGCT TCGCCGGGCG CTGGGCGGAC GCGGTGTTCA CGATTTCGCC CAATCTGGAC 1200

ATCATGCAGG CCACGTACCG CGACATAAAG GCGCAGGTCG AGGCCGCCGG ACGCGATCCC 1260

GAGCAGGTCA AGGTGTTTGC CGCGGTGATG CCGATCCTCG GCGAGACCGA GGCGATCGCC 1320

AGGCAGCGTC TCGAATACAT AAATTCGCTG GTGCATCCCG AAGTCGGGCT TTCTACGTTG 1380

TCCAGCCATG TCGGGGTCAA CCTTGCCGAC TATTCGCTCG ATACCCCGCT GACCGAGGTC 1440

CTGGGCGATC TCGCCCAGCG CAACGTGCCC ACCCAACTGG GCATGTTCGC CAGGATGTTG 1500

CAGGCCGAGA CGCTGACCGT GGGAGAAATG GGCCGGCGTT ATGGCGCCAA CGTGGGCTTC 1560

GTCCCGCAGT GGGCGGGAAC CCGCGAGCAG ATCGCGGACC TGATCGAGAT CCATTTCAAG 1620

GCCGGCGGCG CCGATGGCTT CATCATCTCG CCGGCGTTCC TGCCCGGATC TTACGAGGAA 1680

TTCGTCGATC AGGTGGTGCC CATCCTGCAG CACCGCGGAC TGTTCCGCAC TGATTACGAA 1740

GGCCGCACCC TGCGCAGCCA TCTGGGACTG CGTGAACCCG CATACCTGGG AGAGTACGCA 1800

TGACGACAGA CATCCACCCG GCGAGCGCCG CATCGTCGCC GGCGGCGCGC GCGACGATCA 1860

CCTACAGCAA CTGCCCCGTG CCTAATGCCC TGCTCGCCGC GCTCGGCTCA GGTATTCTGG 1920

ACAGTGCCGG GATCACACTT GCCCTGCTGA CCGGAAAGCA GGGCGAGGTG CACTTCACCT 1980

ACGACCGAGA TGACTACACC CGCTTCGGCG GCGAGATTCC GCCGCTGGTC AGCGAGGGAC 2040

TGCGTGCGCC GGGGCGGACC CGCCTGCTGG GACTGACGCC GGTGCTGGGC CGCTGGGGCT 2100

ACTTCGTCCG GGGCGACAGC GCGATCCGCA CCCCGGCCGA TCTTGCCGGC CGCCGCGTCG 2160

GAGTATCCGA TTCGGCCAGG AGGATATTGA CCGGAAGGCT GGGCGACTAC CGCGAACTTG 2220

ATCCCTGGCG GCAGACCCTG GTCGCGCTGG GGACATGGGA GGCGCGTGCC TTGCTGAGCA 2280

CGCTCGAGAC GGCGGGGCTT GGCGTCGGCG ACGTCGAGCT GACGCGCATC GAGAACCCGT 2340

TCGTCGACGT GCCGACCGAA CGACTGCATG CCGCCGGCTC GCTCAAAGGA ACCGACCTGT 2400

TCCCCGACGT GACCAGCCAG CAGGCCGCAG TCCTTGAGGA TGAGCGCGCC GACGCCCTGT 2460

TCGCGTGGCT TCCCTGGGCG GCCGAGCTCG AGACCCGCAT CGGTGCACGG CCGGTCCTAG 2520

ACCTCAGCGC AGACGACCGC AATGCCTATG CGAGCACCTG GACGGTGAGC GCCGAGCTGG 2580

TGGACCGGCA GCCCGAACTG GTGCAGCGGC TCGTCGATGC CGTGGTGGAT GCAGGGCGGT 2640

GGGCCGAGGC CAATGGCGAT GTCGTCTCCC GCCTGCACGC CGATAACCTC GGTGTCAGTC 2700

CCGAAAGCGT CCGCCAGGGA TTCGGAGCCG ATTTTCACCG CCGCCTGACG CCGCGGCTCG 2760

ACAGCGATGC TATCGCCATC CTGGAGCGTA CTCAGCGGTT CCTGAAGGAT GCGAACCTGA 2820

TCGATCGGTC GTTGGCGCTC GATCGGTGGG CTGCACCTGA ATTCCTCGAA CAAAGTCTCT 2880

CACGCCAGGT CGAAGGGCAG ATAGCATGAA CGAACTCGTC AAAGATCTCG GCCTCAATCG 2940

ATCCGATCCG ATCGGCGCTG TGCGGCGACT GGCCGCGCAG TGGGGGGCCA CCGCTGTTGA 3000

TCGGGACCGG GCCGGCGGAT CGGCAACCGC CGAACTCGAT CAACTGCGCG GCAGCGGCCT 3060

GCTCTCGCTG TCCATTCCCG CCGCATATGG CGGCTGGGGC GCCGACTGGC CAACGACTCT 3120

GGAAGTTATC CGCGAAGTCG CAACGGTGGA CGGATCGCTG GCGCATCTAT TCGGCTACCA 3180

CCTCGGCTGC GTACCGATGA TCGAGCTGTT CGGCTCGGCG CCACAAAAGG AACGGCTGTA 3240

CCGCCAGATC GCAAGCCATG ATTGGCGGGT CGGGAATGCG TCGAGCGAAA ACAACAGCCA 3300

CGTGCTCGAG TGGAAGCTTG CCGCCACCGC CGTCGATGAT GGCGGGTTCG TCCTCAACGG 3360

CGCGAAGCAC TTCTGCAGCG GCGCCAAAAG CTCCGACCTG CTCATCGTGT TCGGCGTGAT 3420

CCAGGACGAA TCCCCCCTGC GCGGCGCGAT CATCACCGCG GTCATTCCCA CCGACCGGGC 3480

CGGTGTTCAG ATCAATGACG ACTGGCGCGC AATCGGGATG CGCCAGACCG ACAGCGGCAG 3540

CGCCGAATTT CGCGACGTCC GAGTCTACCC AGACGAGATC TTGGGGGCAC CAAACTCAGT 3600

CGTTGAGGCG TTCGTGACAA GCAACCGCGG CAGCCTGTGG ACGCCGGCGA TTCAGTCGAT 3660

CTTCTCGAAC GTTTATCTGG GGCTCGCGCG TGGCGCGCTC GAGGCGGCAG CGGATTACAC 3720

CCGGACCCAG AGCCGCCCCT GGACACCCGC CGGCGTGGCG AAGGCGACAG AGGATCCCCA 3780

CATCATCGCC ACCTACGGTG AACTGGCGAT CGCGCTCCAG GGCGCCGAGG CGGCCGCGCG 3840

CGAGGTCGCG GCCCTGTTGC AACAGGCGTG GGACAAGGGC GATGCGGTGA CGCCCGAAGA 3900

GCGCGGCCAG CTGATGGTGA AGGTTTCGGG TGTGAAGGCC CTCTCGACGA AGGCCGCCCT 3960

CGACATCACC AGCCGTATTT TCGAGACAAC GGGCTCGCGA TCGACGCATC CCAGATACGG 4020

ATTCGATCGG TTCTGGCGTA ACATCCGGAC TCATACGCTG CACGATCCGG TATCGTATAA 4080

AATCGTCGAT GTGGGGAACT ACACGCTCAA CGGGACATTC CCGGTTCCCG GATTTACGTC 4140

ATGA 4144

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4144 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TCATGACGTA AATCCGGGAA CCGGGAATGT CCCGTTGAGC GTGTAGTTCC CCACATCGAC 60

GATTTTATAC GATACCGGAT CGTGCAGCGT ATGAGTCCGG ATGTTACGCC AGAACCGATC 120

GAATCCGTAT CTGGGATGCG TCGATCGCGA GCCCGTTGTC TCGAAAATAC GGCTGGTGAT 180

GTCGAGGGCG GCCTTCGTCG AGAGGGCCTT CACACCCGAA ACCTTCACCA TCAGCTGGCC 240

GCGCTCTTCG GGCGTCACCG CATCGCCCTT GTCCCACGCC TGTTGCAACA GGGCCGCGAC 300

CTCGCGCGCG GCCGCCTCGG CGCCCTGGAG CGCGATCGCC AGTTCACCGT AGGTGGCGAT 360

GATGTGGGGA TCCTCTGTCG CCTTCGCCAC GCCGGCGGGT GTCCAGGGGC GGCTCTGGGT 420

CCGGGTGTAA TCCGCTGCCG CCTCGAGCGC GCCACGCGCG AGCCCCAGAT AAACGTTCGA 480

GAAGATCGAC TGAATCGCCG GCGTCCACAG GCTGCCGCGG TTGCTTGTCA CGAACGCCTC 540

AACGACTGAG TTTGGTGCCC CCAAGATCTC GTCTGGGTAG ACTCGGACGT CGCGAAATTC 600

GGCGCTGCCG CTGTCGGTCT GGCGCATCCC GATTGCGCGC CAGTCGTCAT TGATCTGAAC 660

ACCGGCCCGG TCGGTGGGAA TGACCGCGGT GATGATCGCG CCGCGCAGGG GGGATTCGTC 720

CTGGATCACG CCGAACACGA TGAGCAGGTC GGAGCTTTTG GCGCCGCTGC AGAAGTGCTT 780

CGCGCCGTTG AGGACGAACC CGCCATCATC GACGGCGGTG GCGGCAAGCT TCCACTCGAG 840

CACGTGGCTG TTGTTTTCGC TCGACGCATT CCCGACCCGC CAATCATGGC TTGCGATCTG 900

GCGGTACAGC CGTTCCTTTT GTGGCGCCGA GCCGAACAGC TCGATCATCG GTACGCAGCC 960

GAGGTGGTAG CCGAATAGAT GCGCCAGCGA TCCGTCCACC GTTGCGACTT CGCGGATAAC 1020

TTCCAGAGTC GTTGGCCAGT CGGCGCCCCA GCCGCCATAT GCGGCGGGAA TGGACAGCGA 1080

GAGCAGGCCG CTGCCGCGCA GTTGATCGAG TTCGGCGGTT GCCGATCCGC CGGCCCGGTC 1140

CCGATCAACA GCGGTGGCCC CCCACTGCGC GGCCAGTCGC CGCACAGCGC CGATCGGATC 1200

GGATCGATTG AGGCCGAGAT CTTTGACGAG TTCGTTCATG CTATCTGCCC TTCGACCTGG 1260

CGTGAGAGAC TTTGTTCGAG GAATTCAGGT GCAGCCCACC GATCGAGCGC CAACGACCGA 1320

TCGATCAGGT TCGCATCCTT CAGGAACCGC TGAGTACGCT CCAGGATGGC GATAGCATCG 1380

CTGTCGAGCC GCGGCGTCAG GCGGCGGTGA AAATCGGCTC CGAATCCCTG GCGGACGCTT 1440

TCGGGACTGA CACCGAGGTT ATCGGCGTGC AGGCGGGAGA CGACATCGCC ATTGGCCTCG 1500

GCCCACCGCC CTGCATCCAC CACGGCATCG ACGAGCCGCT GCACCAGTTC GGGCTGCCGG 1560

TCCACCAGCT CGGCGCTCAC CGTCCAGGTG CTCGCATAGG CATTGCGGTC GTCTGCGCTG 1620

AGGTCTAGGA CCGGCCGTGC ACCGATGCGG GTCTCGAGCT CGGCCGCCCA GGGAAGCCAC 1680

GCGAACAGGG CGTCGGCGCG CTCATCCTCA AGGACTGCGG CCTGCTGGCT GGTCACGTCG 1740

GGGAACAGGT CGGTTCCTTT GAGCGAGCCG GCGGCATGCA GTCGTTCGGT CGGCACGTCG 1800

ACGAACGGGT TCTCGATGCG CGTCAGCTCG ACGTCGCCGA CGCCAAGCCC CGCCGTCTCG 1860

AGCGTGCTCA GCAAGGCACG CGCCTCCCAT GTCCCCAGCG CGACCAGGGT CTGCCGCCAG 1920

GGATCAAGTT CGCGGTAGTC GCCCAGCCTT CCGGTCAATA TCCTCCTGGC CGAATCGGAT 1980

ACTCCGACGC GGCGGCCGGC AAGATCGGCC GGGGTGCGGA TCGCGCTGTC GCCCCGGACG 2040

AAGTAGCCCC AGCGGCCCAG CACCGGCGTC AGTCCCAGCA GGCGGGTCCG CCCCGGCGCA 2100

CGCAGTCCCT CGCTGACCAG CGGCGGAATC TCGCCGCCGA AGCGGGTGTA GTCATCTCGG 2160

TCGTAGGTGA AGTGCACCTC GCCCTGCTTT CCGGTCAGCA GGGCAAGTGT GATCCCGGCA 2220

CTGTCCAGAA TACCTGAGCC GAGCGCGGCG AGCAGGGCAT TAGGCACGGG GCAGTTGCTG 2280

TAGGTGATCG TCGCGCGCGC CGCCGGCGAC GATGCGGCGC TCGCCGGGTG GATGTCTGTC 2340

GTCATGCGTA CTCTCCCAGG TATGCGGGTT CACGCAGTCC CAGATGGCTG CGCAGGGTGC 2400

GGCCTTCGTA ATCAGTGCGG AACAGTCCGC GGTGCTGCAG GATGGGCACC ACCTGATCGA 2460

CGAATTCCTC GTAAGATCCG GGCAGGAACG CCGGCGAGAT GATGAAGCCA TCGGCGCCGC 2520

CGGCCTTGAA ATGGATCTCG ATCAGGTCCG CGATCTGCTC GCGGGTTCCC GCCCACTGCG 2580

GGACGAAGCC CACGTTGGCG CCATAACGCC GGCCCATTTC TCCCACGGTC AGCGTCTCGG 2640

CCTGCAACAT CCTGGCGAAC ATGCCCAGTT GGGTGGGCAC GTTGCGCTGG GCGAGATCGC 2700

CCAGGACCTC GGTCAGCGGG GTATCGAGCG AATAGTCGGC AAGGTTGACC CCGACATGGC 2760

TGGACAACGT AGAAAGCCCG ACTTCGGGAT GCACCAGCGA ATTTATGTAT TCGAGACGCT 2820

GCCTGGCGAT CGCCTCGGTC TCGCCGAGGA TCGGCATCAC CGCGGCAAAC ACCTTGACCT 2880

GCTCGGGATC GCGTCCGGCG GCCTCGACCT GCGCCTTTAT GTCGCGGTAC GTGGCCTGCA 2940

TGATGTCCAG ATTGGGCGAA ATCGTGAACA CCGCGTCCGC CCAGCGCCCG GCGAAGCGCT 3000

TGCCCCGCGC CGAAAGCCCG GCCTGCAGAA TGACAGGCTC GCCCTGGGGG GAGCGCGGCA 3060

CCTGAAGCGG CCCGCGTACG TTGAGCCATT CGCCGCGGTG GTCGATGTAG CGCACCTTAG 3120

CCGGATCGGC GAACTGGCCG GTTGCCTTGT CGAGTGTCAG CGCATCGCGA TCCCAGCTGT 3180

TCCAGAGCTT GCGCACGACC TCGAGGAATT CATCGGCGCG ATCGTAGCGG GCATCGTGGT 3240

CGAGATGTTC ATCGAAGCCG AAGTTGCGCG CCTCTGCATT GCTGAGCGAG GTGACCACGT 3300

TCCACGACAC TCGCCCGGAG GACAGCTGGT CCAGCGAAGC GACGACCCGG GCTACATGGT 3360

AGGGCGGGTA GTAGGTGGTG GAAATGGTGG CGCCCAGCCC CAGATGTTCG GTCACCGAGG 3420

CCATCGCGGC GATAACTACG CCGGGCTCCA GCATCACCGC GCCTTGCCCG CCATACCGCA 3480

GACCGGTCTC CAGATTGTCG CCGTAGCTGT CCCACACGGC GAGCGCGTCG GGAAGGAACA 3540

GCAGGTCGAA CTTGCCGCGC TCGAGCGTGC GGGCAATCTG CTGGTAGTAC TCCTTGGTGA 3600

GGAAGCCGTT GGAGTCGTCG GCGTGGCGCC ACGCTCCGTG GGCGTGCGTG ACGTTGCCGG 3660

CACAGAAGAA TCCGGCCAGG TGCAGCTGAC GTGGATCGGT CATAGGCTCT TATCCTCGTG 3720

AGTTTGAAAC GAATAGTCGC GGCGCTACCG CCGTCAGTCG CCCACGCGAA CCGGCGTCGA 3780

AAAGACCTTG GCTATGAACG ATGGCGCGAA AGCGTAGGAA GCTGCCGTAC GTCGGTCCTC 3840

TTCATTGGCA TTCGCGAAAT GCACGGATCT TTCGTGCTAG GTGCGGATAG TGGTCATGCG 3900

CATCGTATCA TCTCATCATT GGTGTGAGGG AACGCAATGC CTAGCGCTGT CGCCCCAGCA 3960

TCGCGCGCTT TAGCTCGTGG CAACCCGGAG CGAGCACTTG AGACACACGG TCGGCCTGAT 4020

CACCGAACTT AGTGTCGAAT AAATATCTAT TTATTTAGTA GAGATTCCTA TGTCAAGTGC 4080

AAGAAGAGGC CCCAGAGGAT GGTTTGAACC GCGCCGGGTT CGACGGTCAG ATCGATCTCG 4140

AACC 4144

CLAIMS

We claim:

1. A culture of *Sphingomonas* sp. strain AD109 or a mutant thereof.

2. A nucleotide molecule encoding an enzyme having an amino acid sequence set forth in SEQ ID NO.: 2; or a mutant, fragment or homologue thereof.

3. The nucleotide molecule of Claim 2 having substantially the same sequence as the sequence set forth in SEQ ID NO.: 1.

4. A nucleotide molecule encoding an enzyme having an amino acid sequence set forth in SEQ ID NO.: 4; or a mutant, fragment or homologue thereof.

5. The nucleotide molecule of Claim 4 having substantially the same sequence set forth in SEQ ID NO.: 3.

6. A nucleotide molecule encoding an enzyme having an amino acid sequence set forth in SEQ ID NO.: 6; or a mutant, fragment or homologue thereof.

7. The nucleotide molecule of Claim 6 having substantially the same sequence as the sequence set forth in SEQ ID NO.: 5.

8. A nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO.: 1, SEQ ID NO.: 3, or SEQ ID NO.: 5; a mutant or fragment thereof; or a combination thereof.

9. A nucleotide sequence comprising at least about 20 contiguous nucleotides from the sequence of SEQ ID NO.: 1, or the complement thereof.

10. The nucleotide sequence of Claim 9 comprising at least about 40 contiguous nucleotides from the sequence of SEQ ID NO.: 1 or the complement thereof.

11. The nucleotide sequence of Claim 9 comprising at least about 50 contiguous nucleotides from the sequence of SEQ ID NO.: 1 or the complement thereof.

12. A nucleotide sequence comprising at least about 20 contiguous nucleotides from the sequence of SEQ ID NO.: 3, or the complement thereof.

13. The nucleotide sequence of Claim 12 comprising at least about 40 contiguous nucleotides from the sequence of SEQ ID NO.: 3 or the complement thereof.

14. The nucleotide sequence of Claim 12 comprising at least about 50 contiguous nucleotides from the sequence of SEQ ID NO.: 3 or the complement thereof.

15. A nucleotide sequence comprising at least about 20 contiguous nucleotides from the sequence of SEQ ID NO.: 5, or the complement thereof.

16. The nucleotide sequence of Claim 15 comprising at least about 40 contiguous nucleotides.

17. The nucleotide sequence of Claim 15 comprising at least about 50 contiguous nucleotides.

18. A nucleotide sequence which specifically hybridizes to a polynucleotide molecule comprising the nucleotide sequence set forth in SEQ ID NO.: 1.

19. The nucleotide sequence of Claim 18, wherein the isolated nucleotide sequence hybridizes to the polynucleotide molecule under conditions of high stringency.

20. A nucleotide sequence which hybridizes to a polynucleotide molecule comprising the nucleotide sequence set forth in SEQ ID NO.: 3.

21. The nucleotide sequence of Claim 20, wherein the isolated nucleotide sequence hybridizes to the polynucleotide molecule under conditions of high stringency.

22. A nucleotide sequence which hybridizes to a polynucleotide molecule comprising the nucleotide sequence set forth in SEQ ID NO.: 5.

23. The nucleotide sequence of Claim 22, wherein the isolated nucleotide sequence hybridizes to the polynucleotide molecule under conditions of high stringency.

24. An enzyme having substantially the amino acid sequence set forth in SEQ ID NO.: 2, or an enzymatically active fragment thereof.

25. The enzyme of Claim 24, wherein said enzyme is isolated from a microorganism.

26. The enzyme of Claim 25 wherein the microorganism is a *Sphingomonas*.

27. The enzyme of Claim 26 wherein the microorganism is *Sphingomonas* sp. strain AD109.

28. The enzyme of Claim 27 having substantially the amino acid sequence set forth in SEQ ID NO.: 2, or fragment thereof, said enzyme being substantially free of other *Sphingomonas* proteins.

29. An enzyme having substantially the amino acid sequence set forth in SEQ ID NO.: 4, or an enzymatically active fragment thereof.

30. The enzyme of Claim 29, wherein said enzyme is isolated from a microorganism.

31. The enzyme of Claim 30 wherein the microorganism is a *Sphingomonas*.

32. The enzyme of Claim 31 wherein the microorganism is *Sphingomonas* sp. strain AD109.

33. The enzyme of Claim 32 having a molecular weight of about 40,000 daltons.

34. An enzyme having substantially the amino acid sequence set forth in SEQ ID NO.: 4, or fragment thereof, said enzyme being substantially free of other *Sphingomonas* proteins.

35. An enzyme having substantially the amino acid sequence set forth in SEQ ID NO.: 6, or an enzymatically active fragment thereof.

36. The enzyme of Claim 35, wherein said enzyme is isolated from a microorganism.

37. The enzyme of Claim 36 wherein the microorganism is a *Sphingomonas*.

38. The enzyme of Claim 37 wherein the microorganism is *Sphingomonas* sp. strain AD109.

39. An enzyme having substantially the amino acid sequence set forth in SEQ ID NO.: 6, or fragment thereof, said enzyme being substantially free of other *Sphingomonas* proteins.

40. An enzyme comprising the amino acid sequence set forth in SEQ ID NO.: 2.

41. An enzyme comprising the amino acid sequence set forth in SEQ ID NO.: 4.

42. An enzyme comprising the amino acid sequence set forth in SEQ ID NO.: 6.

43. A *Sphingomonas* enzyme catalyzing the conversion of dibenzothiophene to dibenzothiophene-5,5-dioxide.

44. A *Sphingomonas* enzyme catalyzing the conversion of dibenzothiophene-5,5-dioxide to 2-(2-hydroxyphenyl)benzenesulfinate.

45. A *Sphingomonas* enzyme catalyzing the conversion of 2-(2-hydroxyphenyl)benzenesulfinate to 2-hydroxybiphenyl and inorganic sulfur.

46. A plasmid comprising a nucleic acid molecule of Claim 2 operatively linked to a promoter.

47. A plasmid comprising a nucleic acid molecule of Claim 4 operatively linked to a promoter.

48. A plasmid comprising a nucleic acid molecule of Claim 6 operatively linked to a promoter.

49. A plasmid comprising a nucleotide molecule of Claim 8 operatively linked to a promoter.

50. A transformed microorganism containing a recombinant DNA plasmid comprising a DNA molecule encoding an enzyme set forth in SEQ ID NO.: 2, a mutant or a homologue thereof.

51. A transformed microorganism containing a recombinant DNA plasmid comprising a DNA molecule encoding an enzyme having the sequence set forth in SEQ ID NO.: 4, or a mutant, fragment or homologue thereof.

52. A transformed microorganism containing a recombinant DNA plasmid comprising a DNA molecule encoding an enzyme having the amino acid sequence set forth in SEQ ID NO.: 6; or a mutant or homologue thereof.

53. A transformed microorganism containing a recombinant DNA plasmid comprising a DNA sequence encoding an enzyme having the sequence set forth in SEQ ID NO.: 2, or a mutant, fragment or homologue thereof; a DNA sequence encoding an enzyme having the sequence set forth in SEQ ID NO.: 4, or a mutant, fragment or homologue thereof; and a DNA sequence encoding an enzyme having the amino acid sequence set forth in SEQ ID NO.: 6, or a mutant, fragment or homologue thereof.

54. A method of desulfurizing a fossil fuel containing organosulfur molecules, comprising the steps of:

(a) contacting the fossil fuel with an aqueous phase containing a *Sphingomonas*-derived desulfurization biocatalyst thereby forming a fossil fuel and aqueous phase mixture;

(b) maintaining the mixture under conditions sufficient for desulfurization, thereby resulting in a fossil fuel having a reduced organic sulfur content; and

(c) separating the fossil fuel having a reduced organic sulfur content from the resulting aqueous phase.

55. The method of Claim 54 wherein the biocatalyst comprises an enzyme preparation or a microorganism comprising an enzyme having the sequence set forth in SEQ ID NO.: 2, or a mutant, active fragment or homologue thereof; an enzyme having the sequence set forth in SEQ ID NO.: 4, or a mutant, active fragment or homologue thereof; and an enzyme having the amino acid sequence set forth in SEQ ID NO.: 6, or a mutant, active fragment or homologue thereof.

56. The method of Claim 55 further comprising the steps of adding a flavoprotein, flavin, NADH or a combination thereof.

57. The method of Claim 55 wherein the fossil fuel is a liquid hydrocarbon.

58. The method of Claim 57 wherein the liquid hydrocarbon is a petroleum.

59. The method of Claim 55 wherein the biocatalyst is a microorganism.

60. The method of Claim 59 wherein the microorganism is *Sphingomonas* sp. strain AD109.

61. The method of Claim 59 wherein the microorganism contains a heterologous DNA molecule which encodes the biocatalyst.

62. The method of Claim 61 wherein the heterologous DNA is derived from *Sphingomonas* sp. strain AD109.

63. The method of Claim 55 wherein the biocatalyst is a cell-free fraction.

64. The method of Claim 63 wherein the biocatalyst is a cell-free fraction of *Sphingomonas* sp. strain AD109.

65. A method of oxidizing organic molecules, comprising the steps of:

(a) contacting the organic molecules with an aqueous phase containing a *Sphingomonas*-derived biocatalyst capable of oxidizing organosulfur compounds, thereby forming an organic compound and aqueous phase mixture; and

(b) maintaining the mixture under conditions sufficient for oxidation of the organic molecules by the biocatalyst, thereby forming an oxidized organic compound.

66. The method of Claim 65 wherein the biocatalyst comprises an enzyme having the sequence set forth in SEQ ID NO.: 2, or a mutant, active fragment or homologue thereof; an enzyme having the sequence set forth in SEQ ID NO.: 4, or a mutant, active fragment or homologue thereof; an enzyme having the amino acid sequence set forth in SEQ ID NO.: 6, or a mutant, active fragment or homologue thereof; or a combination thereof.

67. The method of Claim 66 wherein the organic is an organosulfur compound which is a component of a fossil fuel.

68. The method of Claim 67 wherein the organosulfur compound is a substituted or unsubstituted dibenzothiophene and the oxidized organosulfur compound is a substituted or unsubstituted dibenzothiophene-5-5-dioxide.

69. The method of Claim 67 wherein the organosulfur compound is a substituted or unsubstituted dibenzothiophene-5-5-dioxide and the oxidized organosulfur compound is a substituted or unsubstituted 2-(2-hydroxyphenyl)benzenesulfinate.

70. The method of Claim 66 wherein the biocatalyst capable of oxidizing organosulfur molecules is a microorganism.

71. The method of Claim 70 wherein the microorganism is *Sphingomonas* sp. strain AD109.

72. The method of Claim 70 wherein the microorganism contains a heterologous DNA molecule which encodes the biocatalyst.

73. The method of Claim 72 wherein the heterologous DNA is derived from *Sphingomonas* sp. strain AD109.

74. The method of Claim 66 wherein the biocatalyst capable of oxidizing organosulfur compounds is a cell-free fraction.

75. The method of Claim 74 wherein the biocatalyst is a cell-free fraction of *Sphingomonas* sp. strain AD109.

Sphingamonas ORF1

```
                                     30                                      60
          *           *           *           *           *           *
ATG ACC GAT CCA CGT CAG CTG CAC CTG GCC GGA TTC TTC TGT GCC GGC AAC GTC ACG CAC
 M   T   D   P   R   Q   L   H   L   A   G   F   F   C   A   G   N   V   T  H>

                                     90                                     120
          *           *           *           *           *           *
GCC CAC GGA GCG TGG CGC CAC GCC GAC GAC TCC AAC GGC TTC CTC ACC AAG GAG TAC TAC
 A   H   G   A   W   R   H   A   D   D   S   N   G   F   L   T   K   E   Y  Y>

                                    150                                     180
          *           *           *           *           *           *
CAG CAG ATT GCC CGC ACG CTC GAG CGC GGC AAG TTC GAC CTG CTG TTC CTT CCC GAC GCG
 Q   Q   I   A   R   T   L   E   R   G   K   F   D   L   L   F   L   P   D  A>

                                    210                                     240
          *           *           *           *           *           *
CTC GCC GTG TGG GAC AGC TAC GGC GAC AAT CTG GAG ACC GGT CTG CGG TAT GGC GGG CAA
 L   A   V   W   D   S   Y   G   D   N   L   E   T   G   L   R   Y   G   G  Q>

                                    270                                     300
          *           *           *           *           *           *
GGC GCG GTG ATG CTG GAG CCC GGC GTA GTT ATC GCC GCG ATG GCC TCG GTG ACC GAA CAT
 G   A   V   M   L   E   P   G   V   V   I   A   A   M   A   S   V   T   E  H>

                                    330                                     360
          *           *           *           *           *           *
CTG GGG CTG GGC GCC ACC ATT TCC ACC ACC TAC TAC CCG CCC TAC CAT GTA GCC CGG GTC
 L   G   L   G   A   T   I   S   T   T   Y   Y   P   P   Y   H   V   A   R  V>

                                    390                                     420
          *           *           *           *           *           *
GTC GCT TCG CTG GAC CAG CTG TCC TCC GGG CGA GTG TCG TGG AAC GTG GTC ACC TCG CTC
 V   A   S   L   D   Q   L   S   S   G   R   V   S   W   N   V   V   T   S  L>
```

FIGURE 1A

Sphingomonas ORF1 (cont)

```
                                      450                                    480
           *           *           *           *           *           *
AGC AAT GCA GAG GCG CGC AAC TTC GGC TTC GAT GAA CAT CTC GAC CAC GAT GCC CGC TAC
 S   N   A   E   A   R   N   F   G   F   D   E   H   L   D   H   D   A   R   Y>

                                      510                                    540
           *           *           *           *           *           *
GAT CGC GCC GAT GAA TTC CTC GAG GTC GTG CGC AAG CTC TGG AAC AGC TGG GAT CGC GAT
 D   R   A   D   E   F   L   E   V   V   R   K   L   W   N   S   W   D   R   D>

                                      570                                    600
           *           *           *           *           *           *
GCG CTG ACA CTC GAC AAG GCA ACC GGC CAG TTC GCC GAT CCG GCT AAG GTG CGC TAC ATC
 A   L   T   L   D   K   A   T   G   Q   F   A   D   P   A   K   V   R   Y   I>

                                      630                                    660
           *           *           *           *           *           *
GAC CAC CGC GGC GAA TGG CTC AAC GTA CGC GGG CCG CTT CAG GTG CCG CGC TCC CCC CAG
 D   H   R   G   E   W   L   N   V   R   G   P   L   Q   V   P   R   S   P   Q>

                                      690                                    720
           *           *           *           *           *           *
GGC GAG CCT GTC ATT CTG CAG GCC GGG CTT TCG GCG CGG GGC AAG CGC TTC GCC GGG CGC
 G   E   P   V   I   L   Q   A   G   L   S   A   R   G   K   R   F   A   G   R>

                                      750                                    780
           *           *           *           *           *           *
TGG GCG GAC GCG GTG TTC ACG ATT TCG CCC AAT CTG GAC ATC ATG CAG GCC ACG TAC CGC
 W   A   D   A   V   F   T   I   S   P   N   L   D   I   M   Q   A   T   Y   R>

                                      810                                    840
           *           *           *           *           *           *
GAC ATA AAG GCG CAG GTC GAG GCC GCC GGA CGC GAT CCC GAG CAG GTC AAG GTG TTT GCC
 D   I   K   A   Q   V   E   A   A   G   R   D   P   E   Q   V   K   V   F   A>
```

FIGURE 1B

Sphingomonas ORF1 (cont)

```
                                        870                                     900
                    *                   *                   *                   *
GCG GTG ATG CCG ATC CTC GGC GAG ACC GAG GCG ATC GCC AGG CAG CGT CTC GAA TAC ATA
 A   V   M   P   I   L   G   E   T   E   A   I   A   R   Q   R   L   E   Y   I>

                                        930                                     960
                    *                   *                   *                   *
AAT TCG CTG GTG CAT CCC GAA GTC GGG CTT TCT ACG TTG TCC AGC CAT GTC GGG GTC AAC
 N   S   L   V   H   P   E   V   G   L   S   T   L   S   S   H   V   G   V   N>

                                        990                                    1020
                    *                   *                   *                   *
CTT GCC GAC TAT TCG CTC GAT ACC CCG CTG ACC GAG GTC CTG GGC GAT CTC GCC CAG CGC
 L   A   D   Y   S   L   D   T   P   L   T   E   V   L   G   D   L   A   Q   R>

                                       1050                                    1080
                    *                   *                   *                   *
AAC GTG CCC ACC CAA CTG GGC ATG TTC GCC AGG ATG TTG CAG GCC GAG ACG CTG ACC GTG
 N   V   P   T   Q   L   G   M   F   A   R   M   L   Q   A   E   T   L   T   V>

                                       1110                                    1140
                    *                   *                   *                   *
GGA GAA ATG GGC CGG CGT TAT GGC GCC AAC GTG GGC TTC GTC CCG CAG TGG GCG GGA ACC
 G   E   M   G   R   R   Y   G   A   N   V   G   F   V   P   Q   W   A   G   T>

                                       1170                                    1200
                    *                   *                   *                   *
CGC GAG CAG ATC GCG GAC CTG ATC GAG ATC CAT TTC AAG GCC GGC GGC GCC GAT GGC TTC
 R   E   Q   I   A   D   L   I   E   I   H   F   K   A   G   G   A   D   G   F>

                                       1230                                    1260
                    *                   *                   *                   *
ATC ATC TCG CCG GCG TTC CTG CCC GGA TCT TAC GAG GAA TTC GTC GAT CAG GTG GTG CCC
 I   I   S   P   A   F   L   P   G   S   Y   E   E   F   V   D   Q   V   V   P>
```

FIGURE 1C

Sphingomonas ORF1 (cont)

```
                                        1290                                  1320
                  *           *           *           *           *           *
ATC CTG CAG CAC CGC GGA CTG TTC CGC ACT GAT TAC GAA GGC CGC ACC CTG CGC AGC CAT
 I   L   Q   H   R   G   L   F   R   T   D   Y   E   G   R   T   L   R   S   H>

                                        1350
                  *           *           *           *
CTG GGA CTG CGT GAA CCC GCA TAC CTG GGA GAG TAC GCA TGA
 L   G   L   R   E   P   A   Y   L   G   E   Y   A   *>
```

FIGURE 1D

Sphingomonas ORF2

```
                                   30                                  60
                    *                   *                   *           *
ATG ACG ACA GAC ATC CAC CCG GCG AGC GCC GCA TCG TCG CCG GCG GCG CGC GCG ACG ATC
 M   T   T   D   I   H   P   A   S   A   A   S   S   P   A   A   R   A   T   I>

                                   90                                 120
                *                   *                   *           *
ACC TAC AGC AAC TGC CCC GTG CCT AAT GCC CTG CTC GCC GCG CTC GGC TCA GGT ATT CTG
 T   Y   S   N   C   P   V   P   N   A   L   L   A   A   L   G   S   G   I   L>

                                  150                                 180
                *                   *                   *           *
GAC AGT GCC GGG ATC ACA CTT GCC CTG CTG ACC GGA AAG CAG GGC GAG GTG CAC TTC ACC
 D   S   A   G   I   T   L   A   L   L   T   G   K   Q   G   E   V   H   F   T>

                                  210                                 240
                *                   *                   *           *
TAC GAC CGA GAT GAC TAC ACC CGC TTC GGC GGC GAG ATT CCG CCG CTG GTC AGC GAG GGA
 Y   D   R   D   D   Y   T   R   F   G   G   E   I   P   P   L   V   S   E   G>

                                  270                                 300
                *                   *                   *           *
CTG CGT GCG CCG GGG CGG ACC CGC CTG CTG GGA CTG ACG CCG GTG CTG GGC CGC TGG GGC
 L   R   A   P   G   R   T   R   L   L   G   L   T   P   V   L   G   R   W   G>

                                  330                                 360
                *                   *                   *           *
TAC TTC GTC CGG GGC GAC AGC GCG ATC CGC ACC CCG GCC GAT CTT GCC GGC CGC CGC GTC
 Y   F   V   R   G   D   S   A   I   R   T   P   A   D   L   A   G   R   R   V>

                                  390                                 420
                *                   *                   *           *
GGA GTA TCC GAT TCG GCC AGG AGG ATA TTG ACC GGA AGG CTG GGC GAC TAC CGC GAA CTT
 G   V   S   D   S   A   R   R   I   L   T   G   R   L   G   D   Y   R   E   L>
```

FIGURE 2A

Sphingamonas ORF2 (cont)

```
                                                  450                                                     480
                      *                 *                   *                 *                 *                   *
GAT CCC TGG CGG CAG ACC CTG GTC GCG CTG GGG ACA TGG GAG GCG CGT GCC TTG CTG AGC
 D   P   W   R   Q   T   L   V   A   L   G   T   W   E   A   R   A   L   L   S>

                                                  510                                                     540
                      *                 *                   *                 *                 *                   *
ACG CTC GAG ACG GCG GGG CTT GGC GTC GGC GAC GTC GAG CTG ACG CGC ATC GAG AAC CCG
 T   L   E   T   A   G   L   G   V   G   D   V   E   L   T   R   I   E   N   P>

                                                  570                                                     600
                      *                 *                   *                 *                 *                   *
TTC GTC GAC GTG CCG ACC GAA CGA CTG CAT GCC GCC GGC TCG CTC AAA GGA ACC GAC CTG
 F   V   D   V   P   T   E   R   L   H   A   A   G   S   L   K   G   T   D   L>

                                                  630                                                     660
                      *                 *                   *                 *                 *                   *
TTC CCC GAC GTG ACC AGC CAG CAG GCC GCA GTC CTT GAG GAT GAG CGC GCC GAC GCC CTG
 F   P   D   V   T   S   Q   Q   A   A   V   L   E   D   E   R   A   D   A   L>

                                                  690                                                     720
                      *                 *                   *                 *                 *                   *
TTC GCG TGG CTT CCC TGG GCG GCC GAG CTC GAG ACC CGC ATC GGT GCA CGG CCG GTC CTA
 F   A   W   L   P   W   A   A   E   L   E   T   R   I   G   A   R   P   V   L>

                                                  750                                                     780
                      *                 *                   *                 *                 *                   *
GAC CTC AGC GCA GAC GAC CGC AAT GCC TAT GCG AGC ACC TGG ACG GTG AGC GCC GAG CTG
 D   L   S   A   D   D   R   N   A   Y   A   S   T   W   T   V   S   A   E   L>

                                                  810                                                     840
                      *                 *                   *                 *                 *                   *
GTG GAC CGG CAG CCC GAA CTG GTG CAG CGG CTC GTC GAT GCC GTG GTG GAT GCA GGG CGG
 V   D   R   Q   P   E   L   V   Q   R   L   V   D   A   V   V   D   A   G   R>
```

FIGURE 2B

Sphingomonas ORF2 (cont)

```
                                        870                                 900
                    *           *           *           *           *           *
TGG GCC GAG GCC AAT GGC GAT GTC GTC TCC CGC CTG CAC GCC GAT AAC CTC GGT GTC AGT
 W   A   E   A   N   G   D   V   V   S   R   L   H   A   D   N   L   G   V   S>

                                        930                                 960
                    *           *           *           *           *           *
CCC GAA AGC GTC CGC CAG GGA TTC GGA GCC GAT TTT CAC CGC CGC CTG ACG CCG CGG CTC
 P   E   S   V   R   Q   G   F   G   A   D   F   H   R   R   L   T   P   R   L>

                                        990                                1020
                    *           *           *           *           *           *
GAC AGC GAT GCT ATC GCC ATC CTG GAG CGT ACT CAG CGG TTC CTG AAG GAT GCG AAC CTG
 D   S   D   A   I   A   I   L   E   R   T   Q   R   F   L   K   D   A   N   L>

                                       1050                                1080
                    *           *           *           *           *           *
ATC GAT CGG TCG TTG GCG CTC GAT CGG TGG GCT GCA CCT GAA TTC CTC GAA CAA AGT CTC
 I   D   R   S   L   A   L   D   R   W   A   A   P   E   F   L   E   Q   S   L>

                                       1110
                    *           *           *
TCA CGC CAG GTC GAA GGG CAG ATA GCA TGA
 S   R   Q   V   E   G   Q   I   A   *>
```

FIGURE 2C

Sphingomonas ORF3

```
          10         20         30         40         50         60
           *          *          *          *          *          *
ATG AAC GAA CTC GTC AAA GAT CTC GGC CTC AAT CGA TCC GAT CCG ATC GGC GCT GTG CGG
 M   N   E   L   V   K   D   L   G   L   N   R   S   D   P   I   G   A   V   R>

          70         80         90        100        110        120
           *          *          *          *          *          *
CGA CTG GCC GCG CAG TGG GGG GCC ACC GCT GTT GAT CGG GAC CGG GCC GGC GGA TCG GCA
 R   L   A   A   Q   W   G   A   T   A   V   D   R   D   R   A   G   G   S   A>

         130        140        150        160        170        180
           *          *          *          *          *          *
ACC GCC GAA CTC GAT CAA CTG CGC GGC AGC GGC CTG CTC TCG CTG TCC ATT CCC GCC GCA
 T   A   E   L   D   Q   L   R   G   S   G   L   L   S   L   S   I   P   A   A>

         190        200        210        220        230        240
           *          *          *          *          *          *
TAT GGC GGC TGG GGC GCC GAC TGG CCA ACG ACT CTG GAA GTT ATC CGC GAA GTC GCA ACG
 Y   G   G   W   G   A   D   W   P   T   T   L   E   V   I   R   E   V   A   T>

         250        260        270        280        290        300
           *          *          *          *          *          *
GTG GAC GGA TCG CTG GCG CAT CTA TTC GGC TAC CAC CTC GGC TGC GTA CCG ATG ATC GAG
 V   D   G   S   L   A   H   L   F   G   Y   H   L   G   C   V   P   M   I   E>

         310        320        330        340        350        360
           *          *          *          *          *          *
CTG TTC GGC TCG GCG CCA CAA AAG GAA CGG CTG TAC CGC CAG ATC GCA AGC CAT GAT TGG
 L   F   G   S   A   P   Q   K   E   R   L   Y   R   Q   I   A   S   H   D   W>

         370        380        390        400        410        420
           *          *          *          *          *          *
CGG GTC GGG AAT GCG TCG AGC GAA AAC AAC AGC CAC GTG CTC GAG TGG AAG CTT GCC GCC
 R   V   G   N   A   S   S   E   N   N   S   H   V   L   E   W   K   L   A   A>
```

FIGURE 3A

Sphingomonas ORF3 (cont)

```
          430         440         450         460         470         480
           *           *           *           *           *           *
ACC GCC GTC GAT GAT GGC GGG TTC GTC CTC AAC GGC GCG AAG CAC TTC TGC AGC GGC GCC
 T   A   V   D   D   G   G   F   V   L   N   G   A   K   H   F   C   S   G   A>

          490         500         510         520         530         540
           *           *           *           *           *           *
AAA AGC TCC GAC CTG CTC ATC GTG TTC GGC GTG ATC CAG GAC GAA TCC CCC CTG CGC GGC
 K   S   S   D   L   L   I   V   F   G   V   I   Q   D   E   S   P   L   R   G>

          550         560         570         580         590         600
           *           *           *           *           *           *
GCG ATC ATC ACC GCG GTC ATT CCC ACC GAC CGG GCC GGT GTT CAG ATC AAT GAC GAC TGG
 A   I   I   T   A   V   I   P   T   D   R   A   G   V   Q   I   N   D   D   W>

          610         620         630         640         650         660
           *           *           *           *           *           *
CGC GCA ATC GGG ATG CGC CAG ACC GAC AGC GGC AGC GCC GAA TTT CGC GAC GTC CGA GTC
 R   A   I   G   M   R   Q   T   D   S   G   S   A   E   F   R   D   V   R   V>

          670         680         690         700         710         720
           *           *           *           *           *           *
TAC CCA GAC GAG ATC TTG GGG GCA CCA AAC TCA GTC GTT GAG GCG TTC GTG ACA AGC AAC
 Y   P   D   E   I   L   G   A   P   N   S   V   V   E   A   F   V   T   S   N>

          730         740         750         760         770         780
           *           *           *           *           *           *
CGC GGC AGC CTG TGG ACG CCG GCG ATT CAG TCG ATC TTC TCG AAC GTT TAT CTG GGG CTC
 R   G   S   L   W   T   P   A   I   Q   S   I   F   S   N   V   Y   L   G   L>

          790         800         810         820         830         840
           *           *           *           *           *           *
GCG CGT GGC GCG CTC GAG GCG GCA GCG GAT TAC ACC CGG ACC CAG AGC CGC CCC TGG ACA
 A   R   G   A   L   E   A   A   A   D   Y   T   R   T   Q   S   R   P   W   T>
```

FIGURE 3B

Sphingomonas ORF3 (cont)

```
                850         860         870         880         890         900
                 *           *           *           *           *           *
CCC GCC GGC GTG GCG AAG GCG ACA GAG GAT CCC CAC ATC ATC GCC ACC TAC GGT GAA CTG
 P   A   G   V   A   K   A   T   E   D   P   H   I   I   A   T   Y   G   E   L>

                910         920         930         940         950         960
                 *           *           *           *           *           *
GCG ATC GCG CTC CAG GGC GCC GAG GCG GCC GCG CGC GAG GTC GCG GCC CTG TTG CAA CAG
 A   I   A   L   Q   G   A   E   A   A   A   R   E   V   A   A   L   L   Q   Q>

                970         980         990         1000        1010        1020
                 *           *           *           *           *           *
GCG TGG GAC AAG GGC GAT GCG GTG ACG CCC GAA GAG CGC GGC CAG CTG ATG GTG AAG GTT
 A   W   D   K   G   D   A   V   T   P   E   E   R   G   Q   L   M   V   K   V>

                1030        1040        1050        1060        1070        1080
                 *           *           *           *           *           *
TCG GGT GTG AAG GCC CTC TCG ACG AAG GCC GCC CTC GAC ATC ACC AGC CGT ATT TTC GAG
 S   G   V   K   A   L   S   T   K   A   A   L   D   I   T   S   R   I   F   E>

                1090        1100        1110        1120        1130        1140
                 *           *           *           *           *           *
ACA ACG GGC TCG CGA TCG ACG CAT CCC AGA TAC GGA TTC GAT CGG TTC TGG CGT AAC ATC
 T   T   G   S   R   S   T   H   P   R   Y   G   F   D   R   F   W   R   N   I>

                1150        1160        1170        1180        1190        1200
                 *           *           *           *           *           *
CGG ACT CAT ACG CTG CAC GAT CCG GTA TCG TAT AAA ATC GTC GAT GTG GGG AAC TAC ACG
 R   T   H   T   L   H   D   P   V   S   Y   K   I   V   D   V   G   N   Y   T>

                1210        1220        1230
                 *           *           *
CTC AAC GGG ACA TTC CCG GTT CCC GGA TTT ACG TCA
 L   N   G   T   F   P   V   P   G   F   T   S>
```

FIGURE 3C

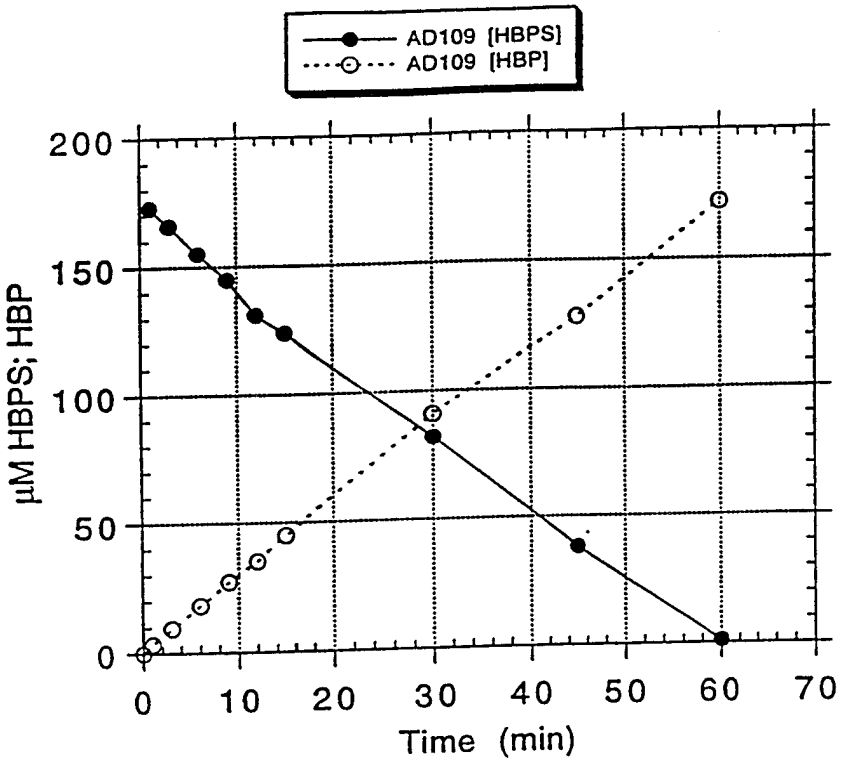

AD109 [HBPS]
AD109 [HBP]
200
150
100
50
0
µM HBPS; HBP
0
10
20
30
40
50
60
70
Time (min)

FIGURE 4

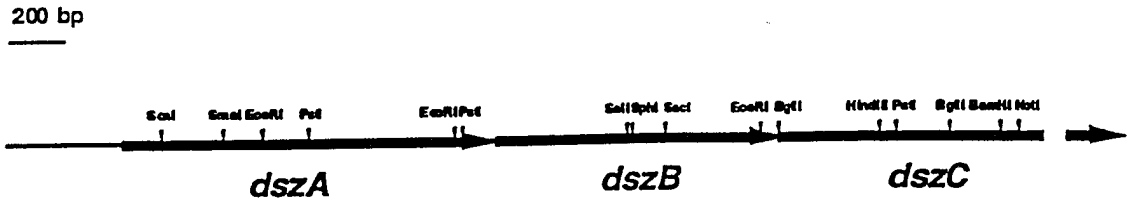

200 bp
ScaI
SmaI EcoRI
PstI
EcoRI PstI
SalI SphI SacI
EcoRI BglI
HindIII PstI
BglI BamHI NotI
dszA
dszB
dszC

FIGURE 5

Sphingomonas dsz sequence

```
        10          20          30          40          50          60
         *           *           *           *           *           *
GGTTCGAGAT CGATCTGACC GTCGAACCCG GCGCGGTTCA AACCATCCTC TGGGGCCTCT
CCAAGCTCTA GCTAGACTGG CAGCTTGGGC CGCGCCAAGT TTGGTAGGAG ACCCCGGAGA

        70          80          90         100         110         120
         *           *           *           *           *           *
TCTTGCACTT GACATAGGAA TCTCTACTAA ATAAATAGAT ATTTATTCGA CACTAAGTTC
AGAACGTGAA CTGTATCCTT AGAGATGATT TATTTATCTA TAAATAAGCT GTGATTCAAG

       130         140         150         160         170         180
         *           *           *           *           *           *
GGTGATCAGG CCGACCGTGT GTCTCAAGTG CTCGCTCCGG GTTGCCACGA GCTAAAGCGC
CCACTAGTCC GGCTGGCACA CAGAGTTCAC GAGCGAGGCC CAACGGTGCT CGATTTCGCG

       190         200         210         220         230         240
         *           *           *           *           *           *
GCGATGCTGG GGCGACAGCG CTAGGCATTG CGTTCCCTCA CACCAATGAT GAGATGATAC
CGCTACGACC CCGCTGTCGC GATCCGTAAC GCAAGGGAGT GTGGTTACTA CTCTACTATG

       250         260         270         280         290         300
         *           *           *           *           *           *
GATGCGCATG ACCACTATCC GCACCTAGCA CGAAAGATCC GTGCATTTCG CGAATGCCAA
CTACGCGTAC TGGTGATAGG CGTGGATCGT GCTTTCTAGG CACGTAAAGC GCTTACGGTT

       310         320         330         340         350         360
         *           *           *           *           *           *
TGAAGAGGAC CGACGTACGG CAGCTTCCTA CGCTTTCGCG CCATCGTTCA TAGCCAAGGT
ACTTCTCCTG GCTGCATGCC GTCGAAGGAT GCGAAAGCGC GGTAGCAAGT ATCGGTTCCA

       370         380         390         400         410         420
         *           *           *           *           *           *
CTTTTCGACG CCGGTTCGCG TGGGCGACTG ACGGCGGTAG CGCCGCGACT ATTCGTTTCA
GAAAAGCTGC GGCCAAGCGC ACCCGCTGAC TGCCGCCATC GCGGCGCTGA TAAGCAAAGT

       430         440         450         460         470         480
         *           *           *           *           *           *
AACTCACGAG GATAAGAGCC TATGACCGAT CCACGTCAGC TGCACCTGGC CGGATTCTTC
TTGAGTGCTC CTATTCTCGG ATACTGGCTA GGTGCAGTCG ACGTGGACCG GCCTAAGAAG

       490         500         510         520         530         540
         *           *           *           *           *           *
TGTGCCGGCA ACGTCACGCA CGCCCACGGA GCGTGGCGCC ACGCCGACGA CTCCAACGGC
ACACGGCCGT TGCAGTGCGT GCGGGTGCCT CGCACCGCGG TGCGGCTGCT GAGGTTGCCG

       550         560         570         580         590         600
         *           *           *           *           *           *
TTCCTCACCA AGGAGTACTA CCAGCAGATT GCCCGCACGC TCGAGCGCGG CAAGTTCGAC
AAGGAGTGGT TCCTCATGAT GGTCGTCTAA CGGGCGTGCG AGCTCGCGCC GTTCAAGCTG
```

FIGURE 6A

Sphingomonas dsz sequence (page 2)

```
        610         620         630         640         650         660
          *           *           *           *           *           *
CTGCTGTTCC TTCCCGACGC GCTCGCCGTG TGGGACAGCT ACGGCGACAA TCTGGAGACC
GACGACAAGG AAGGGCTGCG CGAGCGGCAC ACCCTGTCGA TGCCGCTGTT AGACCTCTGG

        670         680         690         700         710         720
          *           *           *           *           *           *
GGTCTGCGGT ATGGCGGGCA AGGCGCGGTG ATGCTGGAGC CCGGCGTAGT TATCGCCGCG
CCAGACGCCA TACCGCCCGT TCCGCGCCAC TACGACCTCG GGCCGCATCA ATAGCGGCGC

        730         740         750         760         770         780
          *           *           *           *           *           *
ATGGCCTCGG TGACCGAACA TCTGGGGCTG GGCGCCACCA TTTCCACCAC CTACTACCCG
TACCGGAGCC ACTGGCTTGT AGACCCCGAC CCGCGGTGGT AAAGGTGGTG GATGATGGGC

        790         800         810         820         830         840
          *           *           *           *           *           *
CCCTACCATG TAGCCCGGGT CGTCGCTTCG CTGGACCAGC TGTCCTCCGG GCGAGTGTCG
GGGATGGTAC ATCGGGCCCA GCAGCGAAGC GACCTGGTCG ACAGGAGGCC CGCTCACAGC

        850         860         870         880         890         900
          *           *           *           *           *           *
TGGAACGTGG TCACCTCGCT CAGCAATGCA GAGGCGCGCA ACTTCGGCTT CGATGAACAT
ACCTTGCACC AGTGGAGCGA GTCGTTACGT CTCCGCGCGT TGAAGCCGAA GCTACTTGTA

        910         920         930         940         950         960
          *           *           *           *           *           *
CTCGACCACG ATGCCCGCTA CGATCGCGCC GATGAATTCC TCGAGGTCGT GCGCAAGCTC
GAGCTGGTGC TACGGGCGAT GCTAGCGCGG CTACTTAAGG AGCTCCAGCA CGCGTTCGAG

        970         980         990        1000        1010        1020
          *           *           *           *           *           *
TGGAACAGCT GGGATCGCGA TGCGCTGACA CTCGACAAGG CAACCGGCCA GTTCGCCGAT
ACCTTGTCGA CCCTAGCGCT ACGCGACTGT GAGCTGTTCC GTTGGCCGGT CAAGCGGCTA

       1030        1040        1050        1060        1070        1080
          *           *           *           *           *           *
CCGGCTAAGG TGCGCTACAT CGACCACCGC GGCGAATGGC TCAACGTACG CGGGCCGCTT
GGCCGATTCC ACGCGATGTA GCTGGTGGCG CCGCTTACCG AGTTGCATGC GCCCGGCGAA

       1090        1100        1110        1120        1130        1140
          *           *           *           *           *           *
CAGGTGCCGC GCTCCCCCCA GGGCGAGCCT GTCATTCTGC AGGCCGGGCT TTCGGCGCGG
GTCCACGGCG CGAGGGGGGT CCCGCTCGGA CAGTAAGACG TCCGGCCCGA AAGCCGCGCC

       1150        1160        1170        1180        1190        1200
          *           *           *           *           *           *
GGCAAGCGCT TCGCCGGGCG CTGGGCGGAC GCGGTGTTCA CGATTTCGCC CAATCTGGAC
CCGTTCGCGA AGCGGCCCGC GACCCGCCTG CGCCACAAGT GCTAAAGCGG GTTAGACCTG
```

FIGURE 6B

Sphingomonas dsz sequence (page 3)

```
      1210        1220        1230        1240        1250        1260
        *           *           *           *           *           *
ATCATGCAGG CCACGTACCG CGACATAAAG GCGCAGGTCG AGGCCGCCGG ACGCGATCCC
TAGTACGTCC GGTGCATGGC GCTGTATTTC CGCGTCCAGC TCCGGCGGCC TGCGCTAGGG

      1270        1280        1290        1300        1310        1320
        *           *           *           *           *           *
GAGCAGGTCA AGGTGTTTGC CGCGGTGATG CCGATCCTCG GCGAGACCGA GGCGATCGCC
CTCGTCCAGT TCCACAAACG GCGCCACTAC GGCTAGGAGC CGCTCTGGCT CCGCTAGCGG

      1330        1340        1350        1360        1370        1380
        *           *           *           *           *           *
AGGCAGCGTC TCGAATACAT AAATTCGCTG GTGCATCCCG AAGTCGGGCT TTCTACGTTG
TCCGTCGCAG AGCTTATGTA TTTAAGCGAC CACGTAGGGC TTCAGCCCGA AAGATGCAAC

      1390        1400        1410        1420        1430        1440
        *           *           *           *           *           *
TCCAGCCATG TCGGGGTCAA CCTTGCCGAC TATTCGCTCG ATACCCCGCT GACCGAGGTC
AGGTCGGTAC AGCCCCAGTT GGAACGGCTG ATAAGCGAGC TATGGGGCGA CTGGCTCCAG

      1450        1460        1470        1480        1490        1500
        *           *           *           *           *           *
CTGGGCGATC TCGCCCAGCG CAACGTGCCC ACCCAACTGG GCATGTTCGC CAGGATGTTG
GACCCGCTAG AGCGGGTCGC GTTGCACGGG TGGGTTGACC CGTACAAGCG GTCCTACAAC

      1510        1520        1530        1540        1550        1560
        *           *           *           *           *           *
CAGGCCGAGA CGCTGACCGT GGGAGAAATG GGCCGGCGTT ATGGCGCCAA CGTGGGCTTC
GTCCGGCTCT GCGACTGGCA CCCTCTTTAC CCGGCCGCAA TACCGCGGTT GCACCCGAAG

      1570        1580        1590        1600        1610        1620
        *           *           *           *           *           *
GTCCCGCAGT GGGCGGGAAC CCGCGAGCAG ATCGCGGACC TGATCGAGAT CCATTTCAAG
CAGGGCGTCA CCCGCCCTTG GGCGCTCGTC TAGCGCCTGG ACTAGCTCTA GGTAAAGTTC

      1630        1640        1650        1660        1670        1680
        *           *           *           *           *           *
GCCGGCGGCG CCGATGGCTT CATCATCTCG CCGGCGTTCC TGCCCGGATC TTACGAGGAA
CGGCCGCCGC GGCTACCGAA GTAGTAGAGC GGCCGCAAGG ACGGGCCTAG AATGCTCCTT

      1690        1700        1710        1720        1730        1740
        *           *           *           *           *           *
TTCGTCGATC AGGTGGTGCC CATCCTGCAG CACCGCGGAC TGTTCCGCAC TGATTACGAA
AAGCAGCTAG TCCACCACGG GTAGGACGTC GTGGCGCCTG ACAAGGCGTG ACTAATGCTT

      1750        1760        1770        1780        1790        1800
        *           *           *           *           *           *
GGCCGCACCC TGCGCAGCCA TCTGGGACTG CGTGAACCCG CATACCTGGG AGAGTACGCA
CCGGCGTGGG ACGCGTCGGT AGACCCTGAC GCACTTGGGC GTATGGACCC TCTCATGCGT
```

FIGURE 6C

Sphingomonas dsz sequence (page 4)

```
      1810        1820        1830        1840        1850        1860
        *           *           *           *           *           *
TGACGACAGA CATCCACCCG GCGAGCGCCG CATCGTCGCC GGCGGCGCGC GCGACGATCA
ACTGCTGTCT GTAGGTGGGC CGCTCGCGGC GTAGCAGCGG CCGCCGCGCG CGCTGCTAGT

      1870        1880        1890        1900        1910        1920
        *           *           *           *           *           *
CCTACAGCAA CTGCCCCGTG CCTAATGCCC TGCTCGCCGC GCTCGGCTCA GGTATTCTGG
GGATGTCGTT GACGGGGCAC GGATTACGGG ACGAGCGGCG CGAGCCGAGT CCATAAGACC

      1930        1940        1950        1960        1970        1980
        *           *           *           *           *           *
ACAGTGCCGG GATCACACTT GCCCTGCTGA CCGGAAAGCA GGGCGAGGTG CACTTCACCT
TGTCACGGCC CTAGTGTGAA CGGGACGACT GGCCTTTCGT CCCGCTCCAC GTGAAGTGGA

      1990        2000        2010        2020        2030        2040
        *           *           *           *           *           *
ACGACCGAGA TGACTACACC CGCTTCGGCG GCGAGATTCC GCCGCTGGTC AGCGAGGGAC
TGCTGGCTCT ACTGATGTGG GCGAAGCCGC CGCTCTAAGG CGGCGACCAG TCGCTCCCTG

      2050        2060        2070        2080        2090        2100
        *           *           *           *           *           *
TGCGTGCGCC GGGGCGGACC CGCCTGCTGG GACTGACGCC GGTGCTGGGC CGCTGGGGCT
ACGCACGCGG CCCCGCCTGG GCGGACGACC CTGACTGCGG CCACGACCCG GCGACCCCGA

      2110        2120        2130        2140        2150        2160
        *           *           *           *           *           *
ACTTCGTCCG GGGCGACAGC GCGATCCGCA CCCCGGCCGA TCTTGCCGGC CGCCGCGTCG
TGAAGCAGGC CCCGCTGTCG CGCTAGGCGT GGGGCCGGCT AGAACGGCCG GCGGCGCAGC

      2170        2180        2190        2200        2210        2220
        *           *           *           *           *           *
GAGTATCCGA TTCGGCCAGG AGGATATTGA CCGGAAGGCT GGGCGACTAC CGCGAACTTG
CTCATAGGCT AAGCCGGTCC TCCTATAACT GGCCTTCCGA CCCGCTGATG GCGCTTGAAC

      2230        2240        2250        2260        2270        2280
        *           *           *           *           *           *
ATCCCTGGCG GCAGACCCTG GTCGCGCTGG GGACATGGGA GGCGCGTGCC TTGCTGAGCA
TAGGGACCGC CGTCTGGGAC CAGCGCGACC CCTGTACCCT CCGCGCACGG AACGACTCGT

      2290        2300        2310        2320        2330        2340
        *           *           *           *           *           *
CGCTCGAGAC GGCGGGGCTT GGCGTCGGCG ACGTCGAGCT GACGCGCATC GAGAACCCGT
GCGAGCTCTG CCGCCCCGAA CCGCAGCCGC TGCAGCTCGA CTGCGCGTAG CTCTTGGGCA

      2350        2360        2370        2380        2390        2400
        *           *           *           *           *           *
TCGTCGACGT GCCGACCGAA CGACTGCATG CCGCCGGCTC GCTCAAAGGA ACCGACCTGT
AGCAGCTGCA CGGCTGGCTT GCTGACGTAC GGCGGCCGAG CGAGTTTCCT TGGCTGGACA
```

FIGURE 6D

Sphingomonas dsz sequence (page 5)

```
      2410        2420        2430        2440        2450        2460
        *           *           *           *           *           *
TCCCCGACGT GACCAGCCAG CAGGCCGCAG TCCTTGAGGA TGAGCGCGCC GACGCCCTGT
AGGGGCTGCA CTGGTCGGTC GTCCGGCGTC AGGAACTCCT ACTCGCGCGG CTGCGGGACA

      2470        2480        2490        2500        2510        2520
        *           *           *           *           *           *
TCGCGTGGCT TCCCTGGGCG GCCGAGCTCG AGACCCGCAT CGGTGCACGG CCGGTCCTAG
AGCGCACCGA AGGGACCCGC CGGCTCGAGC TCTGGGCGTA GCCACGTGCC GGCCAGGATC

      2530        2540        2550        2560        2570        2580
        *           *           *           *           *           *
ACCTCAGCGC AGACGACCGC AATGCCTATG CGAGCACCTG GACGGTGAGC GCCGAGCTGG
TGGAGTCGCG TCTGCTGGCG TTACGGATAC GCTCGTGGAC CTGCCACTCG CGGCTCGACC

      2590        2600        2610        2620        2630        2640
        *           *           *           *           *           *
TGGACCGGCA GCCCGAACTG GTGCAGCGGC TCGTCGATGC CGTGGTGGAT GCAGGGCGGT
ACCTGGCCGT CGGGCTTGAC CACGTCGCCG AGCAGCTACG GCACCACCTA CGTCCCGCCA

      2650        2660        2670        2680        2690        2700
        *           *           *           *           *           *
GGGCCGAGGC CAATGGCGAT GTCGTCTCCC GCCTGCACGC CGATAACCTC GGTGTCAGTC
CCCGGCTCCG GTTACCGCTA CAGCAGAGGG CGGACGTGCG GCTATTGGAG CCACAGTCAG

      2710        2720        2730        2740        2750        2760
        *           *           *           *           *           *
CCGAAAGCGT CCGCCAGGGA TTCGGAGCCG ATTTTCACCG CCGCCTGACG CCGCGGCTCG
GGCTTTCGCA GGCGGTCCCT AAGCCTCGGC TAAAAGTGGC GGCGGACTGC GGCGCCGAGC

      2770        2780        2790        2800        2810        2820
        *           *           *           *           *           *
ACAGCGATGC TATCGCCATC CTGGAGCGTA CTCAGCGGTT CCTGAAGGAT GCGAACCTGA
TGTCGCTACG ATAGCGGTAG GACCTCGCAT GAGTCGCCAA GGACTTCCTA CGCTTGGACT

      2830        2840        2850        2860        2870        2880
        *           *           *           *           *           *
TCGATCGGTC GTTGGCGCTC GATCGGTGGG CTGCACCTGA ATTCCTCGAA CAAAGTCTCT
AGCTAGCCAG CAACCGCGAG CTAGCCACCC GACGTGGACT TAAGGAGCTT GTTTCAGAGA

      2890        2900        2910        2920        2930        2940
        *           *           *           *           *           *
CACGCCAGGT CGAAGGGCAG ATAGCATGAA CGAACTCGTC AAAGATCTCG GCCTCAATCG
GTGCGGTCCA GCTTCCCGTC TATCGTACTT GCTTGAGCAG TTTCTAGAGC CGGAGTTAGC

      2950        2960        2970        2980        2990        3000
        *           *           *           *           *           *
ATCCGATCCG ATCGGCGCTG TGCGGCGACT GGCCGCGCAG TGGGGGGCCA CCGCTGTTGA
TAGGCTAGGC TAGCCGCGAC ACGCCGCTGA CCGGCGCGTC ACCCCCCGGT GGCGACAACT
```

FIGURE 6E

Sphingomonas dsz sequence (page 6)

```
      3010       3020       3030       3040       3050       3060
        *          *          *          *          *          *
TCGGGACCGG GCCGGCGGAT CGGCAACCGC CGAACTCGAT CAACTGCGCG GCAGCGGCCT
AGCCCTGGCC CGGCCGCCTA GCCGTTGGCG GCTTGAGCTA GTTGACGCGC CGTCGCCGGA

      3070       3080       3090       3100       3110       3120
        *          *          *          *          *          *
GCTCTCGCTG TCCATTCCCG CCGCATATGG CGGCTGGGGC GCCGACTGGC CAACGACTCT
CGAGAGCGAC AGGTAAGGGC GGCGTATACC GCCGACCCCG CGGCTGACCG GTTGCTGAGA

      3130       3140       3150       3160       3170       3180
        *          *          *          *          *          *
GGAAGTTATC CGCGAAGTCG CAACGGTGGA CGGATCGCTG GCGCATCTAT TCGGCTACCA
CCTTCAATAG GCGCTTCAGC GTTGCCACCT GCCTAGCGAC CGCGTAGATA AGCCGATGGT

      3190       3200       3210       3220       3230       3240
        *          *          *          *          *          *
CCTCGGCTGC GTACCGATGA TCGAGCTGTT CGGCTCGGCG CCACAAAAGG AACGGCTGTA
GGAGCCGACG CATGGCTACT AGCTCGACAA GCCGAGCCGC GGTGTTTTCC TTGCCGACAT

      3250       3260       3270       3280       3290       3300
        *          *          *          *          *          *
CCGCCAGATC GCAAGCCATG ATTGGCGGGT CGGGAATGCG TCGAGCGAAA ACAACAGCCA
GGCGGTCTAG CGTTCGGTAC TAACCGCCCA GCCCTTACGC AGCTCGCTTT TGTTGTCGGT

      3310       3320       3330       3340       3350       3360
        *          *          *          *          *          *
CGTGCTCGAG TGGAAGCTTG CCGCCACCGC CGTCGATGAT GGCGGGTTCG TCCTCAACGG
GCACGAGCTC ACCTTCGAAC GGCGGTGGCG GCAGCTACTA CCGCCCAAGC AGGAGTTGCC

      3370       3380       3390       3400       3410       3420
        *          *          *          *          *          *
CGCGAAGCAC TTCTGCAGCG GCGCCAAAAG CTCCGACCTG CTCATCGTGT TCGGCGTGAT
GCGCTTCGTG AAGACGTCGC CGCGGTTTTC GAGGCTGGAC GAGTAGCACA AGCCGCACTA

      3430       3440       3450       3460       3470       3480
        *          *          *          *          *          *
CCAGGACGAA TCCCCCCTGC GCGGCGCGAT CATCACCGCG GTCATTCCCA CCGACCGGGC
GGTCCTGCTT AGGGGGGACG CGCCGCGCTA GTAGTGGCGC CAGTAAGGGT GGCTGGCCCG

      3490       3500       3510       3520       3530       3540
        *          *          *          *          *          *
CGGTGTTCAG ATCAATGACG ACTGGCGCGC AATCGGGATG CGCCAGACCG ACAGCGGCAG
GCCACAAGTC TAGTTACTGC TGACCGCGCG TTAGCCCTAC GCGGTCTGGC TGTCGCCGTC

      3550       3560       3570       3580       3590       3600
        *          *          *          *          *          *
CGCCGAATTT CGCGACGTCC GAGTCTACCC AGACGAGATC TTGGGGGCAC CAAACTCAGT
GCGGCTTAAA GCGCTGCAGG CTCAGATGGG TCTGCTCTAG AACCCCCGTG GTTTGAGTCA
```

FIGURE 6F

Sphingomonas dsz sequence (page 7)

```
      3610        3620        3630        3640        3650        3660
        *           *           *           *           *           *
CGTTGAGGCG TTCGTGACAA GCAACCGCGG CAGCCTGTGG ACGCCGGCGA TTCAGTCGAT
GCAACTCCGC AAGCACTGTT CGTTGGCGCC GTCGGACACC TGCGGCCGCT AAGTCAGCTA

      3670        3680        3690        3700        3710        3720
        *           *           *           *           *           *
CTTCTCGAAC GTTTATCTGG GGCTCGCGCG TGGCGCGCTC GAGGCGGCAG CGGATTACAC
GAAGAGCTTG CAAATAGACC CCGAGCGCGC ACCGCGCGAG CTCCGCCGTC GCCTAATGTG

      3730        3740        3750        3760        3770        3780
        *           *           *           *           *           *
CCGGACCCAG AGCCGCCCCT GGACACCCGC CGGCGTGGCG AAGGCGACAG AGGATCCCCA
GGCCTGGGTC TCGGCGGGGA CCTGTGGGCG GCCGCACCGC TTCCGCTGTC TCCTAGGGGT

      3790        3800        3810        3820        3830        3840
        *           *           *           *           *           *
CATCATCGCC ACCTACGGTG AACTGGCGAT CGCGCTCCAG GGCGCCGAGG CGGCCGCGCG
GTAGTAGCGG TGGATGCCAC TTGACCGCTA GCGCGAGGTC CCGCGGCTCC GCCGGCGCGC

      3850        3860        3870        3880        3890        3900
        *           *           *           *           *           *
CGAGGTCGCG GCCCTGTTGC AACAGGCGTG GGACAAGGGC GATGCGGTGA CGCCCGAAGA
GCTCCAGCGC CGGGACAACG TTGTCCGCAC CCTGTTCCCG CTACGCCACT GCGGGCTTCT

      3910        3920        3930        3940        3950        3960
        *           *           *           *           *           *
GCGCGGCCAG CTGATGGTGA AGGTTTCGGG TGTGAAGGCC CTCTCGACGA AGGCCGCCCT
CGCGCCGGTC GACTACCACT TCCAAAGCCC ACACTTCCGG GAGAGCTGCT TCCGGCGGGA

      3970        3980        3990        4000        4010        4020
        *           *           *           *           *           *
CGACATCACC AGCCGTATTT TCGAGACAAC GGGCTCGCGA TCGACGCATC CCAGATACGG
GCTGTAGTGG TCGGCATAAA AGCTCTGTTG CCCGAGCGCT AGCTGCGTAG GGTCTATGCC

      4030        4040        4050        4060        4070        4080
        *           *           *           *           *           *
ATTCGATCGG TTCTGGCGTA ACATCCGGAC TCATACGCTG CACGATCCGG TATCGTATAA
TAAGCTAGCC AAGACCGCAT TGTAGGCCTG AGTATGCGAC GTGCTAGGCC ATAGCATATT

      4090        4100        4110        4120        4130        4140
        *           *           *           *           *           *
AATCGTCGAT GTGGGGAACT ACACGCTCAA CGGGACATTC CCGGTTCCCG GATTTACGTC
TTAGCAGCTA CACCCCTTGA TGTGCGAGTT GCCCTGTAAG GGCCAAGGGC CTAAATGCAG

ATGA
TACT
```

FIGURE 6G

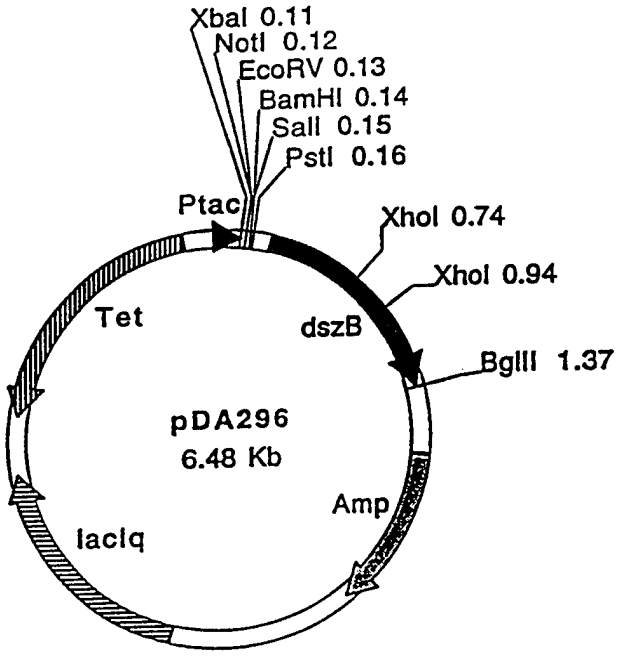

XbaI 0.11
NotI 0.12
EcoRV 0.13
BamHI 0.14
SalI 0.15
PstI 0.16
Ptac
XhoI 0.74
XhoI 0.94
Tet
dszB
BglII 1.37
pDA296
6.48 Kb
Amp
lacIq

FIGURE 7

```
DszA (S)    1 MTDPRQLHLAGFFCAGNVTHAHGAWRHADDSNGFLTKEYYQQIARTLERG 50
DszA (R)    1 MTQQRQMHLAGFFSAGNVTHAHGAWRHTDASNDFLSGKYYQHIARTLERG 50

DszA (S)   51 KFDLLFLPDALAVWDSYGDNLETGLRYGGQGAVMLEPGVVIAAMASVTEH 100
DszA (R)   51 KFDLLFLPDGLAVEDSYGDNLDTGVGLGGQGAVALEPASVVATMAAVTEH 100

DszA (S)  101 LGLGATISTTYYPPYHVARVVASLDQLSSGRVSWNVVTSLSNAEARNFGF 150
DszA (R)  101 LGLGATISATYYPPYHVARVFATLDQLSGGRVSWNVVTSLNDAEARNFGI 150

DszA (S)  151 DEHLDHDARYDRADEFLEVVRKLWNSWDRDALTLDKATGQFADPAKVRYI 200
DszA (R)  151 NQHLEHDARYDRADEFLEAVKKLWNSWDEDALVLDKAAGVFADPAKVHYV 200

DszA (S)  201 DHRGEWLNVRGPLQVPRSPQGEPVILQAGLSARGKRFAGRWADAVFTISP 250
DszA (R)  201 DHHGEWLNVRGPLQVPRSPQGEPVILQAGLSPRGRRFAGKWAEAVFSLAP 250

DszA (S)  251 NLDIMQATYRDIKAQVEAAGRDPEQVKVFAAVMPILGETEAIARQRLEYI 300
DszA (R)  251 NLEVMQATYQGIKAEVDAAGRDPDQTKIFTAVMPVLGESQAVAQERLEYL 300

DszA (S)  301 NSLVHPEVGLSTLSSHVGVNLADYSLDTPLTEVLGDLAQRNVPTQLGMFA 350
DszA (R)  301 NSLVHPEVGLSTLSSHTGINLAAYPLDTPIKDILRDLQDRNVPTQLHMFA 350

DszA (S)  351 RMLQAETLTVGEMGRRYGANVGFVPQWAGTREQIADLIEIHFKAGGADGF 400
DszA (R)  351 AATHSEELTLAEMGRRYGTNVGFVPQWAGTGEQIADELIRHFEGGAADGF 400

DszA (S)  401 IISPAFLPGSYEEFVDQVVPILQHRGLFRTDYEGRTLRSHLGLREPAYLG 450
DszA (R)  401 IISPAFLPGSYDEFVDQVVPVLQDRGYFRTEYQGNTLRDHLGLRVPQLQG 450

DszA (S)  451 EYA 453

DszA (R)  451 QPS 453
```

FIGURE 8

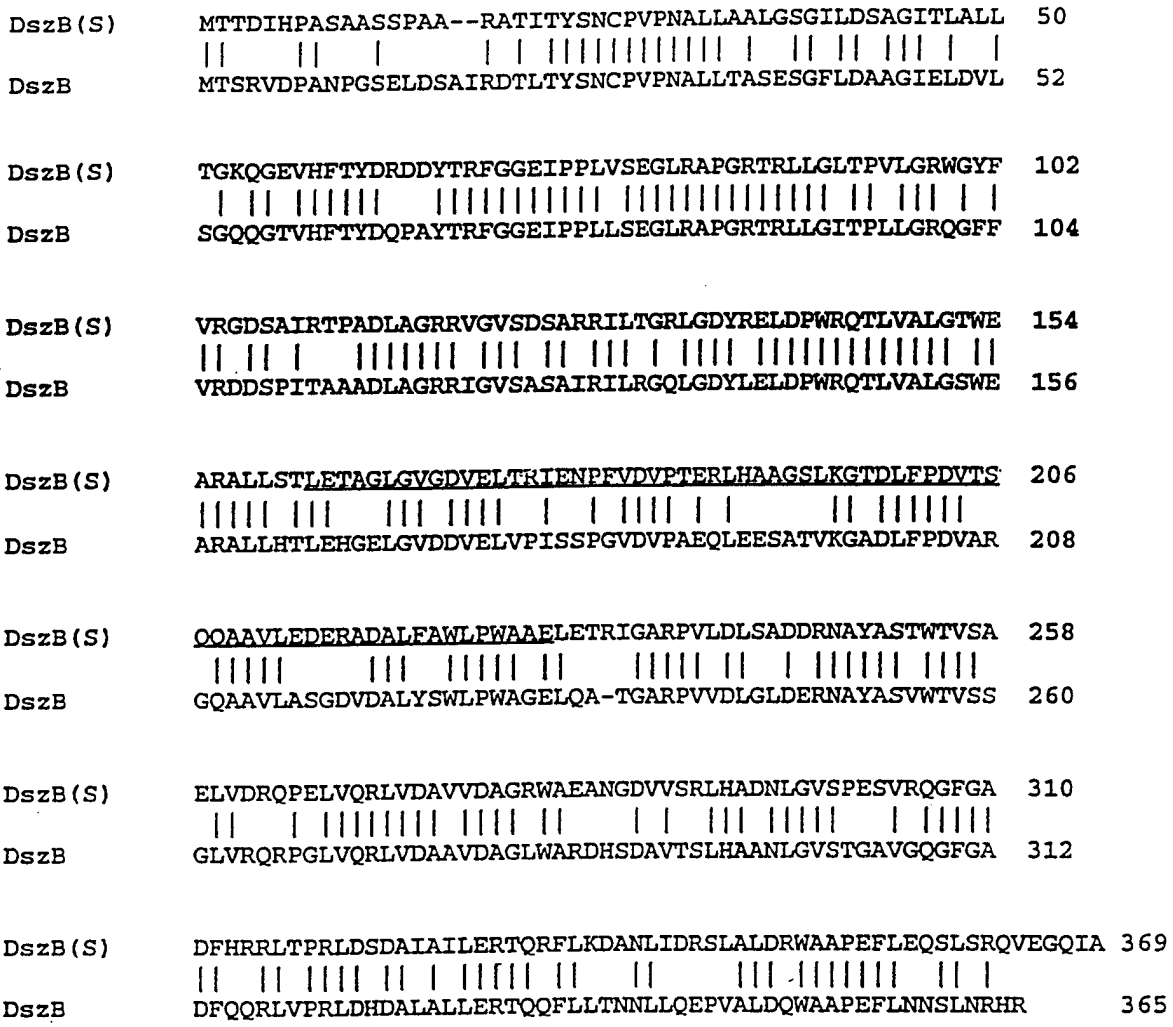

DszB(S) MTTDIHPASAASSPAA--RATITYSNCPVPNALLAALGSGILDSAGITLALL 50
DszB MTSRVDPANPGSELDSAIRDTLTYSNCPVPNALLTASESGFLDAAGIELDVL 52
DszB(S) TGKQGEVHFTYDRDDYTRFGGEIPPLVSEGLRAPGRTRLLGLTPVLGRWGYF 102
DszB SGQQGTVHFTYDQPAYTRFGGEIPPLLSEGLRAPGRTRLLGITPLLGRQGFF 104
DszB(S) VRGDSAIRTPADLAGRRVGVSDSARRILTGRLGDYRELDPWRQTLVALGTWE 154
DszB VRDDSPITAAADLAGRRIGVSASAIRILRGQLGDYLELDPWRQTLVALGSWE 156
DszB(S) ARALLSTLETAGLGVGDVELTRIENPFVDVPTERLHAAGSLKGTDLFPDVTS 206
DszB ARALLHTLEHGELGVDDVELVPISSPGVDVPAEQLEESATVKGADLFPDVAR 208
DszB(S) QQAAVLEDERADALFAWLPWAAELETRIGARPVLDLSADDRNAYASTWTVSA 258
DszB GQAAVLASGDVDALYSWLPWAGELQA-TGARPVVDLGLDERNAYASVWTVSS 260
DszB(S) ELVDRQPELVQRLVDAVVDAGRWAEANGDVVSRLHADNLGVSPESVRQGFGA 310
DszB GLVRQRPGLVQRLVDAAVDAGLWARDHSDAVTSLHAANLGVSTGAVGQGFGA 312
DszB(S) DFHRRLTPRLDSDAIAILERTQRFLKDANLIDRSLALDRWAAPEFLEQSLSRQVEGQIA 369
DszB DFQQRLVPRLDHDALALLERTQQFLLTNNLLQEPVALDQWAAPEFLNNSLNRHR 365

FIGURE 9

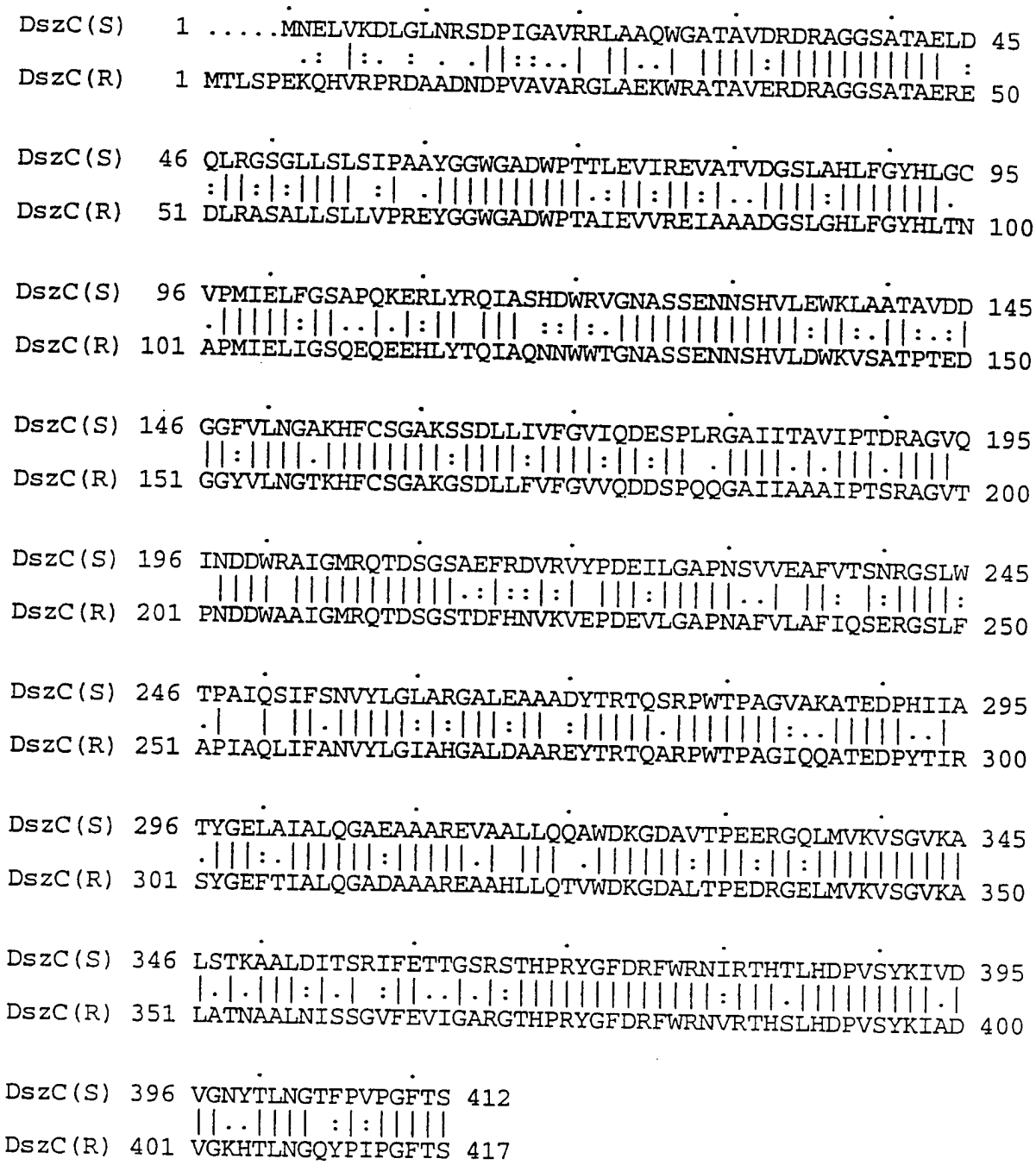

DszC(S) 1 .....MNELVKDLGLNRSDPIGAVRRLAAQWGATAVDRDRAGGSATAELD 45
DszC(R) 1 MTLSPEKQHVRPRDAADNDPVAVARGLAEKWRATAVERDRAGGSATAERE 50
DszC(S) 46 QLRGSGLLSLSIPAAYGGWGADWPTTLEVIREVATVDGSLAHLFGYHLGC 95
DszC(R) 51 DLRASALLSLLVPREYGGWGADWPTAIEVVREIAAADGSLGHLFGYHLTN 100
DszC(S) 96 VPMIELFGSAPQKERLYRQIASHDWRVGNASSENNSHVLEWKLAATAVDD 145
DszC(R) 101 APMIELIGSQEQEEHLYTQIAQNNWWTGNASSENNSHVLDWKVSATPTED 150
DszC(S) 146 GGFVLNGAKHFCSGAKSSDLLIVFGVIQDESPLRGAIITAVIPTDRAGVQ 195
DszC(R) 151 GGYVLNGTKHFCSGAKGSDLLFVFGVVQDDSPQQGAIIAAAIPTSRAGVT 200
DszC(S) 196 INDDWRAIGMRQTDSGSAEFRDVRVYPDEILGAPNSVVEAFVTSNRGSLW 245
DszC(R) 201 PNDDWAAIGMRQTDSGSTDFHNVKVEPDEVLGAPNAFVLAFIQSERGSLF 250
DszC(S) 246 TPAIQSIFSNVYLGLARGALEAAADYTRTQSRPWTPAGVAKATEDPHIIA 295
DszC(R) 251 APIAQLIFANVYLGIAHGALDAAREYTRTQARPWTPAGIQQATEDPYTIR 300
DszC(S) 296 TYGELAIALQGAEAAAREVAALLQQAWDKGDAVTPEERGQLMVKVSGVKA 345
DszC(R) 301 SYGEFTIALQGADAAAREAAHLLQTVWDKGDALTPEDRGELMVKVSGVKA 350
DszC(S) 346 LSTKAALDITSRIFETTGSRSTHPRYGFDRFWRNIRTHTLHDPVSYKIVD 395
DszC(R) 351 LATNAALNISSGVFEVIGARGTHPRYGFDRFWRNVRTHSLHDPVSYKIAD 400
DszC(S) 396 VGNYTLNGTFPVPGFTS 412
DszC(R) 401 VGKHTLNGQYPIPGFTS 417

FIGURE 10

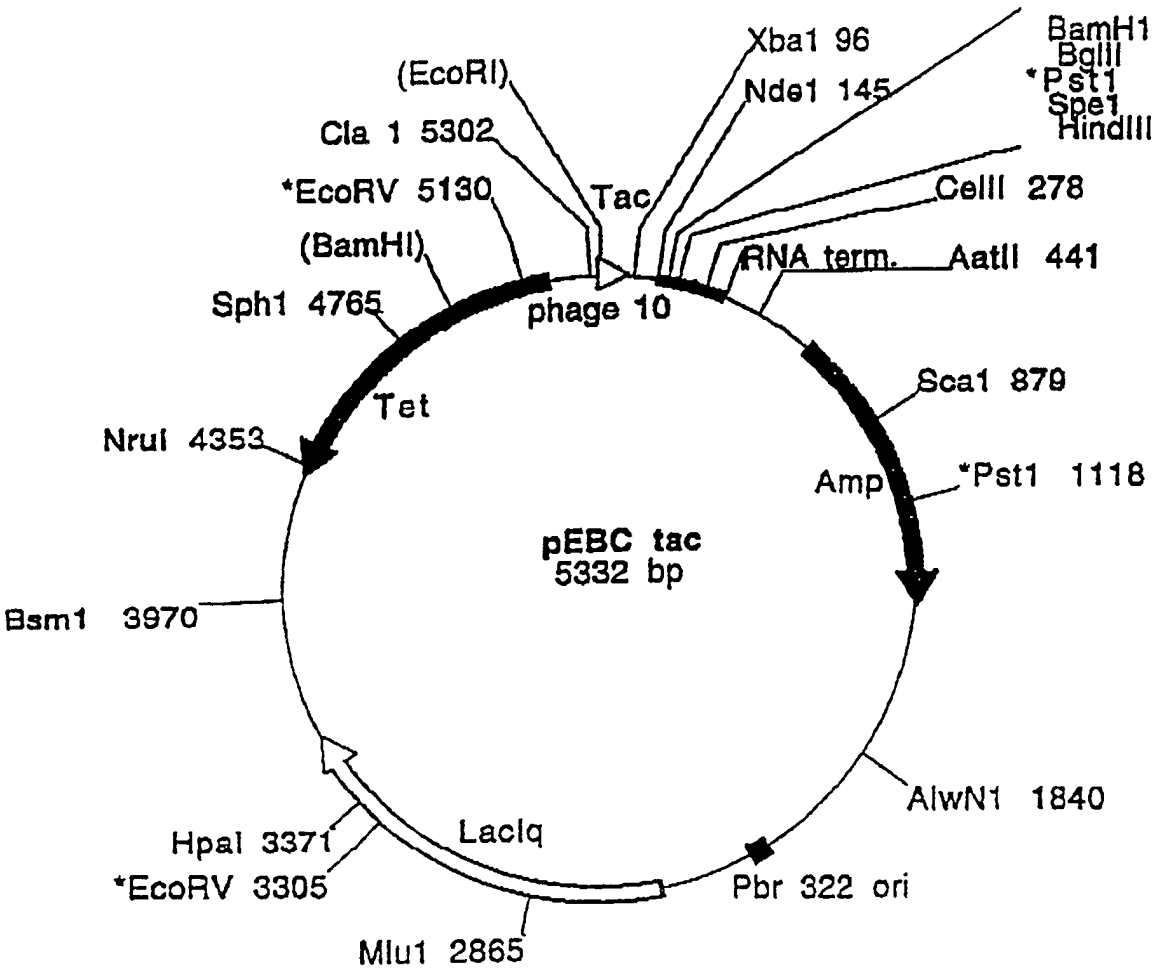

Xba1 96
BamH1
BglII
*Pst1
Spe1
HindIII
Nde1 145
(EcoRI)
Cla 1 5302
*EcoRV 5130
Tac
CelII 278
(BamHI)
RNA term.
AatII 441
Sph1 4765
phage 10
Sca1 879
Tet
NruI 4353
Amp
*Pst1 1118
pEBC tac
5332 bp
Bsm1 3970
AlwN1 1840
HpaI 3371
LacIq
*EcoRV 3305
Pbr 322 ori
MluI 2865

FIGURE 11

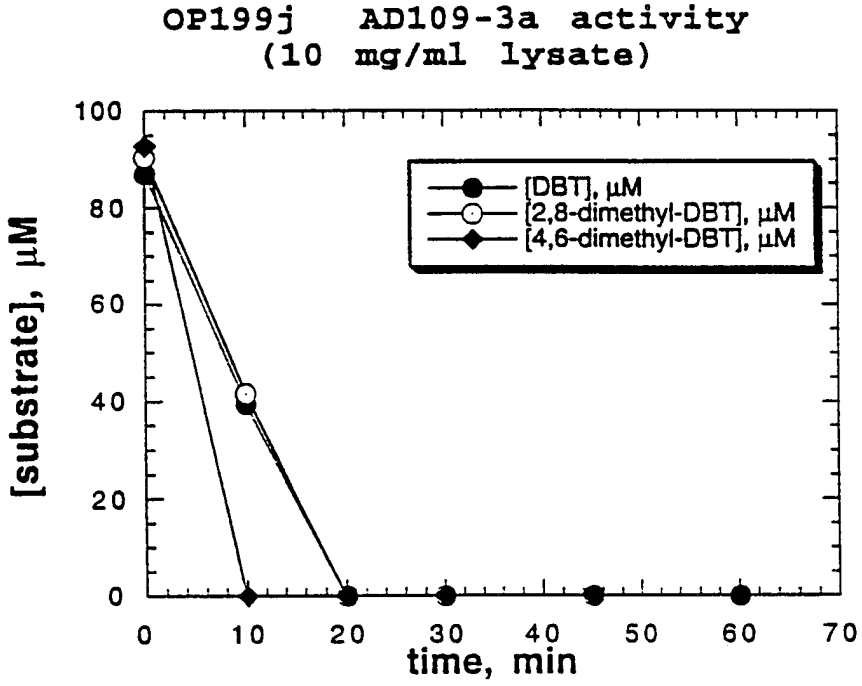

OP199j AD109-3a activity
(10 mg/ml lysate)
[substrate], µM
time, min
[DBT], µM
[2,8-dimethyl-DBT], µM
[4,6-dimethyl-DBT], µM
0
20
40
60
80
100
0
10
20
30
40
50
60
70

FIGURE 12

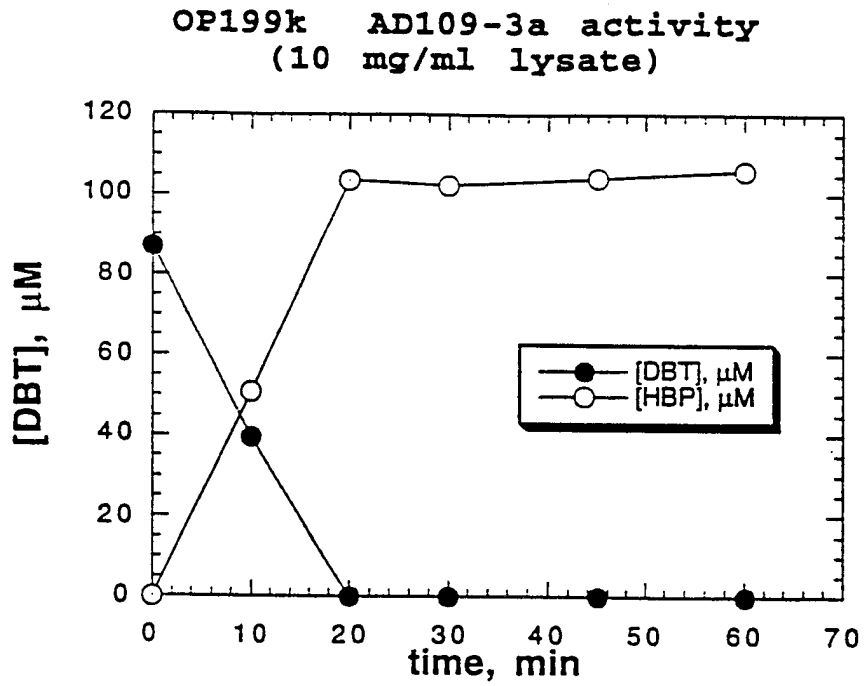

OP199k AD109-3a activity
(10 mg/ml lysate)
[DBT], µM
120
100
80
60
40
20
0
0
10
20
30
40
50
60
70
time, min
[DBT], µM
[HBP], µM

FIGURE 13

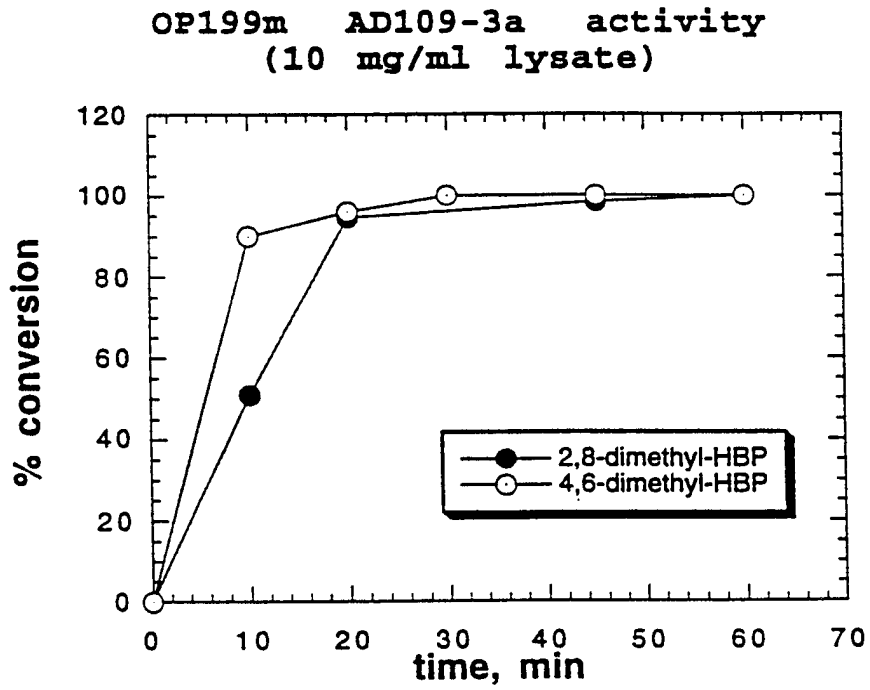

OP199m AD109-3a activity
(10 mg/ml lysate)
% conversion
time, min
2,8-dimethyl-HBP
4,6-dimethyl-HBP
0
20
40
60
80
100
120
0
10
20
30
40
50
60
70

FIGURE 14

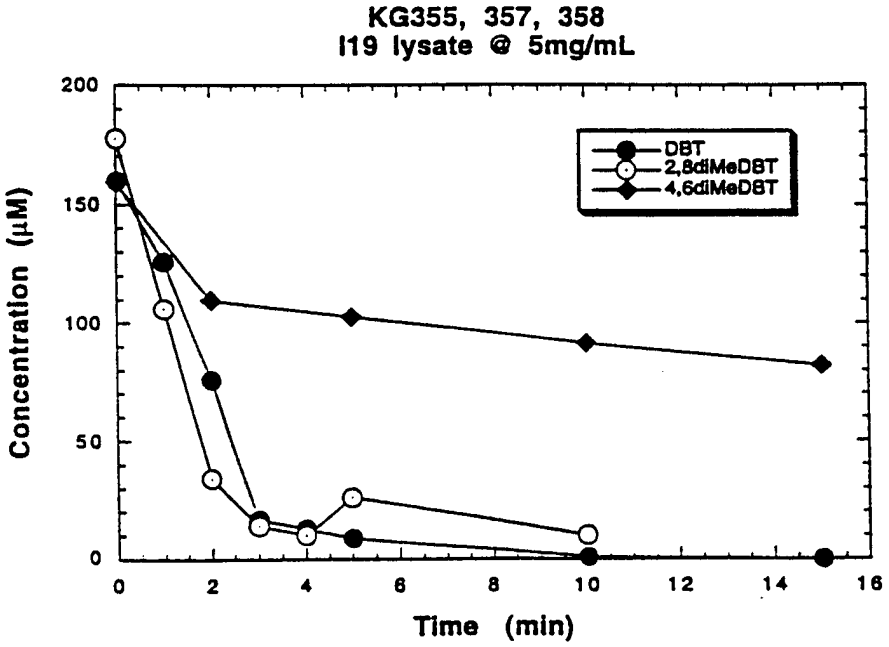

KG355, 357, 358
I19 lysate @ 5mg/mL
Concentration (μM)
Time (min)
DBT
2,8diMeDBT
4,6diMeDBT
0
50
100
150
200
0
2
4
6
8
10
12
14
16

FIGURE 15

INTERNATIONAL SEARCH REPORT

national Application No
PCT/US 98/06684

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/52 C12N9/02 C12N1/21 C12P11/00 C12S1/02
C10G32/00

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C12P C12S C10G

Documentation searched other than minimumdocumentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 94 01563 A (ENERGY BIOSYSTEMS CORP ;RAMBOSEK JOHN (US); PIDDINGTON CHRIS S (US) 20 January 1994 cited in the application see abstract see Seq. ID 1-5 ---	9,12,15, 18-23
A	EP 0 218 734 A (ATLANTIC RES CORP) 22 April 1987 see the whole document --- -/--	1-75

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publicationdate of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of theinternational search	Date of mailing of the international search report
31 July 1998	13/08/1998

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Galli, I

3

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

national Application No
PCT/US 98/06684

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	GRAY K.A. ET AL.: "Molecular mechanisms of biocatalytic desulfurization of fossil fuels." NATURE BIOTECHNOLOGY, vol. 14, 14 December 1996, pages 1705-1709, XP002073201 cited in the application see the whole document	1-75
A	CONSTANTI M. ET AL.: "Desulphurization of dibenzothiopene by bacteria" WORLD J. OF MICROBIOLOGY & BIOTECHNOLOGY, vol. 10, no. 5, 1994, pages 510-516, XP002072523 see the whole document	1-75
A	KLUBEK B. ET AL.: "Characterization of soil bacteria that desulfurize organic sulfur compounds. 1. Classification and growth studies." MICROBIOS, vol. 88, no. 357, 1996, pages 223-236, XP002072524 see abstract	1-75

3

INTERNATIONAL SEARCH REPORT

Information on patent family members

national Application No
PCT/US 98/06684

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 9401563 A	20-01-1994	AU	684253 B	11-12-1997
		AU	4671893 A	31-01-1994
		AU	5835398 A	02-07-1998
		CA	2139876 A	20-01-1994
		CN	1085254 A	13-04-1994
		EP	0651808 A	10-05-1995
		JP	7507691 T	31-08-1995
		MX	9304135 A	31-01-1995
		NO	950081 A	09-03-1995
		US	5356801 A	18-10-1994
		US	5578478 A	26-11-1996
EP 0218734 A	22-04-1987	US	4562156 A	31-12-1985