

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
14 February 2008 (14.02.2008)

PCT

(10) International Publication Number
WO 2008/019069 A2

(51) International Patent Classification:
C12N 9/00 (2006.01) *C12N 9/10* (2006.01)

(74) Agent: JACOBSON, Jill, A.; Danisco Us, Inc., Genencor Division, 925 Page Mill Road, Palo Alto, California 94304 (US).

(21) International Application Number:
PCT/US2007/017351

(22) International Filing Date: 3 August 2007 (03.08.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/835,402 3 August 2006 (03.08.2006) US

(71) Applicant (for all designated States except US):
DANISCO US, INC., GENENCOR DIVISION
[US/US]; 925 Page Mill Road, Palo Alto, California 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **BOTT, Richard, R.** [US/US]; Danisco Us, Inc., Genencor Division, 925 Page Mill Road, Palo Alto, California 94304 (US). **CERVIN, Marguerite, A.** [CA/US]; Danisco Us, Inc., Genencor Division, 925 Page Mill Road, Palo Alto, California 94304 (US). **YAO, Jian** [US/US]; Danisco Us, Inc., Genencor Division, 925 Page Mill Road, Palo Alto, California 94304 (US).

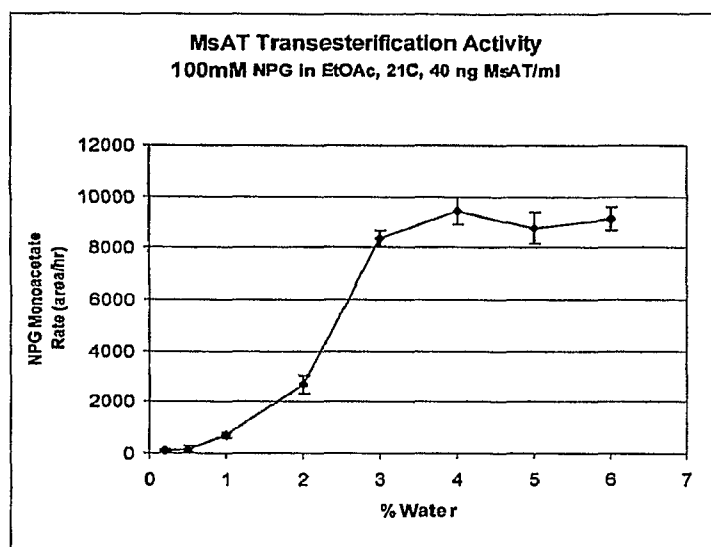
(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN,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

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

(54) Title: ENZYMATIC AQUEOUS ACYLATION



(57) Abstract: The present invention provides methods for engineering enzymes belonging to the class of enzymes known as SGNH hydrolases and alpha/beta hydrolases, to create compositions comprising at least one enzyme suitable for use in enzymatic aqueous acylation and/or perhydrolysis. The present invention further provides compositions comprising at least one perhydrolase enzyme suitable for use in enzymatic aqueous acylation and/or perhydrolysis.

ENZYMATIC AQUEOUS ACYLATION

CROSS-REFERENCE TO RELATED APPLICATIONS

[01] This application claims the benefit of U.S. Provisional Application No. 60/835,402, filed August 3, 2006, which is hereby incorporated by reference herein in its entirety.

FIELD OF THE INVENTION

[02] The present invention provides methods for engineering enzymes belonging to the class of enzymes known as SGNH hydrolases and α/β hydrolases, to create compositions comprising at least one enzyme suitable for use in enzymatic aqueous acylation and/or perhydrolysis. The present invention further provides compositions comprising at least one perhydrolase enzyme suitable for use in enzymatic aqueous acylation and/or perhydrolysis.

BACKGROUND OF THE INVENTION

[03] The ability to catalyze acyl transfer reactions in water has been an unrealized goal of biocatalysis. Such reactions would eliminate the need for protection and deprotection steps in synthesis, and therefore reduce the environmental impact and cost of such chemistry. The opportunity to exploit the selectivity and catalytic efficiency in an economical media could remove cost bottlenecks in the synthesis of many bioproducts, in particular, pharmaceuticals. Previously the best candidates have been found in lipase enzymes, which belong to the α/β hydrolase family of enzymes. Lipases are now widely used for catalytic and stereospecific transesterification reactions, in both academic and industrial laboratories. However, in order to promote the desired alcoholysis, the currently used methods involve the use of anhydrous solvents to prevent hydrolysis of the target ester. Thus, there is a need in the art for methods and compositions that involve use of aqueous conditions, in addition to being economically feasible and attractive, as well as environmentally friendly.

SUMMARY OF THE INVENTION

[04] The present invention provides methods for engineering enzymes belonging to the class of enzymes known as SGNH hydrolases and α/β hydrolases, to create compositions comprising at least one enzyme suitable for use in enzymatic aqueous acylation and/or perhydrolysis. The present invention further provides compositions comprising at least one perhydrolase enzyme suitable for use in enzymatic aqueous acylation and/or perhydrolysis.

[05] In some embodiments, the present invention provides means to identify regions suitable for the introduction of at least one insertion to modify an enzyme. In some particular embodiments, the modified enzyme has the SGNH hydrolase fold or an α/β hydrolase fold. In some embodiments, the modification(s) increase the enzyme's ability to catalyze aqueous acylation and/or perhydrolysis. The present invention further provides compositions comprising at least one enzyme engineered for use in enzymatic aqueous acylation and/or perhydrolysis.

[06] In yet additional embodiments, the present invention provides means for searching homologs of interest, in regard to the presence and/or absence of insertions in proteins. In some embodiments, hidden Markov methods (*e.g.*, HMM1 or HMM2) find use in searching sequence and/or sequence/structure space. In additional embodiments, searches are based on the presence of at least one insertion of 5-11 amino acids, with a space of a defined set of amino acids. By defining the insert length and distance between inserts, homologs with low sequence specificity, but that exhibit the desired activity are identified.

[07] In one aspect, the invention provides polypeptides with enzymatic aqueous acylation and/or perhydrolysis activity, identified in accordance with the methods described herein. In some embodiments, the polypeptide comprises, consists of, or consists essentially of a sequence selected from the group consisting of SEQ ID NOs: 11-17, 19-27, 29-40, and 42-48. The invention also provides polynucleotides encoding the polypeptides, expression vectors comprising the polynucleotides, host cells comprising the polynucleotides and/or expression vectors, and methods for producing the polypeptides by expression from polynucleotides encoding them in a host cell.

DESCRIPTION OF THE FIGURES

[08] Figure 1 provides a graph showing the catalytic activity of MsAcT in the transesterification of an acetate moiety from ethyl acetate to neopentyl glycol (NPG) in the

presence of varying concentrations of water. Reactions were conducted under conditions yielding single phase mixtures of water dissolved in ethyl acetate or ethyl acetate dissolved in water.

[09] Figure 2 provides data showing the selectivity of perhydrolysis to hydrolysis catalyzed by acyltransferases and other enzymes in the presence of 10 mM triacetin and 30 mM hydrogen peroxide. The two acyltransferases show a high degree of selectivity for perhydrolysis over hydrolysis compared with other enzymes known to hydrolyze triacetin.

[10] Figure 3 provides a diagram showing the octameric arrangement of MsAcT. The octamer is found as a tetramer of closely associated dimers. Monomers within each dimer pair are colored gray and dark gray, in order to allow differentiation between them in the dimer. The octamer is characterized as having six surfaces, with the “top” and “bottom” surrounding a large channel, while the “sides” have a smaller crevice which opens into two active sites formed by the association of monomers into dimers.

[11] Figure 4 provides a schematic showing the basic SGNH hydrolase fold having a characteristic central beta sheet usually consisting of five strands designated β 1-5 and connected to intervening helices that cross on one or the other side of this sheet in a conserved pattern. The SGNH hydrolase fold positions residues form the catalytic triad consisting (*i.e.*, Ser 11, Asp 192 and His 195), relative to the characteristic feature of secondary structure, namely β strands and α helices. The catalytic serine is found in a small helical segment following the first β strand (*i.e.*, β 1) and the catalytic Asp and His are found in a loop found between helices 6 and 7.

[12] Figure 5 provides a stereodiagram that compares the monomer of MsAcT (dark) with the *E. coli* thioesterase (light gray). This Figure shows that the overall folding and juxtaposition of secondary features such as sheets (indicated by arrows) and helices (indicated by coils) are highly conserved, supporting the assignment of these enzymes into the same superfamily. The locations of the four large insertions labeled “insertions 1-4” represent distinct departures of the folding pattern that is unique to the MsAcT structure.

[13] Figure 6 provides a schematic that shows the locations of the insertions relative to the conserved secondary feature that is common to all SGNH hydrolases. The locations of these can be easily associated with any SGNH hydrolase, based on the identification of the conserved secondary elements of the basic SGNH hydrolase fold.

[14] Figure 7, Panel A provides a diagram that shows the relative location of insertion 3 of MsACT. This insertion extends from one monomer into the dimer mate. This insertion stabilizes the dimer, along with several conserved residues at the dimer interface.

[15] Figure 7, Panel B provides a diagram that illustrates the interaction of insertion 3 as shown in Figure 7, Panel A, contributed from a dimer mate with insertions 1, 2 and 4, to form an elaborate substrate binding surface, which is associated with the unique synthetic properties of MsAcT.

[16] Figure 8 provides a schematic diagram of the α/β hydrolase fold. This can be compared to the SGNH hydrolase fold in Figure 4 and it is evident that there are numerous similarities particularly in the region where insertions 1-4 occur in MsAct. The same insertion sites exist in the α/β hydrolase fold. In this case, insertion 1 occurs after strand 3 and before helix A; insertion 2 occurs after strand 4 and before helix B; insertion 3 occurs after helix C and before strand 6; and insertion 4 occurs after strand 6 and before helix D.

[17] Figure 9 provides a sequence comparison of five sequences (SEQ ID NOS:1, 2, 3, 4, and 5).

[18] Figure 10 is a stereoview of aligned structures of *E. coli* thioesterase (light gray) and *M. smegmatis* aryl esterase (dark gray). Segments of the common SGNH hydrolase secondary structure are highlighted as stick figures showing all main chain atoms.

[19] Figure 11 is a stereoview that is orthogonal to the view shown in Figure 10.

DETAILED DESCRIPTION

[20] As indicated above, the ability to catalyze aqueous acyl transfer reactions has been an unrealized goal of biocatalysis. Such reactions eliminate the need for protection and deprotection steps in synthesis, which leads to a reduction in the environmental impact and cost of such chemistry. It is contemplated that the present invention will provide opportunities to exploit selectivity and catalytic efficiency in an economical media will eliminate cost bottlenecks in the synthesis of bioproducts, including but not limited to pharmaceuticals. Previously, the best candidates identified for such reactions are lipase enzymes, which belong to the α/β hydrolase family of enzymes. Indeed, lipases are currently widely used for catalytic and stereospecific transesterification in both academic and industrial laboratories. These reactions are alcoholysis reactions run in anhydrous solvents and have found application in chiral

synthesis, regioselective protection, and enantiomeric resolution (*See*, Klibanov, *Nature* 409:241-246 [2001]). In order to promote the desired alcoholysis, these reactions are conducted in anhydrous solvents to prevent the hydrolysis of the target ester. Those of skill in the art know how to select the enzyme, solvent, reaction conditions, as well as utilize suitable substrate specificity, chiral selectivity, immobilization techniques, and reaction kinetics.

[21] The present invention provides methods for engineering enzymes belonging to the class of enzymes known as SGNH hydrolases and α/β hydrolases, to create compositions comprising at least one enzyme suitable for use in enzymatic aqueous acylation and/or perhydrolysis. The present invention further provides compositions comprising at least one enzyme suitable for use in enzymatic aqueous acylation and/or perhydrolysis.

[22] In some embodiments, the present invention provides means to identify regions suitable for the introduction of at least one insertion to modify an enzyme. In some particular embodiments, the modified enzyme has the SGNH hydrolase fold or an α/β hydrolase fold. In some embodiments, the modification(s) increase the enzyme's ability to catalyze aqueous acylation and/or perhydrolysis. The present invention further provides compositions comprising at least one enzyme engineered for use in enzymatic aqueous acylation and/or perhydrolysis. The invention also provides methods for identification of polypeptides with enzymatic aqueous acylation and/or perhydrolysis activity and polypeptides identified in accordance with these methods.

[23] The headings provided herein are not limitations of the various aspects or embodiments of the invention which can be had by reference to the specification as a whole. Accordingly, the terms defined immediately below are more fully defined by reference to the specification as a whole. Nonetheless, in order to facilitate understanding of the invention, a number of terms are defined below.

Definitions

[24] Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. For example, Singleton and Sainsbury, *Dictionary of Microbiology and Molecular Biology*, 2d Ed., John Wiley and Sons, NY (1994); and Hale and Marham, *The Harper Collins Dictionary of Biology*, Harper Perennial, NY (1991) provide those of skill in the art with a general dictionaries of many of the terms used in the invention. Any methods and materials

similar or equivalent to those described herein may find use in the practice of the present invention. Accordingly, the terms defined immediately below are more fully described by reference to the Specification as a whole. Also, as used herein, the singular terms "a", "an," and "the" include the plural reference unless the context clearly indicates otherwise. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy orientation, respectively. It is to be understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary, depending upon the context they are used by those of skill in the art.

[25] It is intended that every maximum numerical limitation given throughout this specification includes every lower numerical limitation, as if such lower numerical limitations were expressly written herein. Every minimum numerical limitation given throughout this specification will include every higher numerical limitation, as if such higher numerical limitations were expressly written herein. Every numerical range given throughout this specification will include every narrower numerical range that falls within such broader numerical range, as if such narrower numerical ranges were all expressly written herein.

[26] As used herein, the term "SGNH hydrolase" refers to the structurally related superfamily of catalytic serine hydrolase enzymes. These enzymes have a catalytic triad comprising residues in the linear sequence, Ser –Asp –His. These catalytic residues are associated with several blocks of conserved sequence associated with specific secondary features characteristic of all SGNH hydrolases. These features are illustrated in Figure 3, and include a central parallel β sheet that preferably contains five strands.

[27] The first conserved block of sequence is found at the C-terminal of the first β strand and includes the sequence GDS, where the S represents the serine of the catalytic triad. Another conserved block includes the sequence DXXH, which is found in a random coil following the fifth β strand ($\beta 5$ in Figure 1) and preceding a conserved helical segment (H7); this segment includes the Asp and His of the catalytic triad..

[28] Proteins belonging to the SGNH hydrolase superfamily are homologous to other known SGNH hydrolases, including but not limited to such hydrolases as thioesterase I (PDB code: 1ivn), platelet-activating factor (an acetyl hydrolase; PDB code: 1wab), esterase (PDB code: 1esc), rhamnogalacturonan acylesterase (RGAE, PDB code: 1deo), and at least one

hypothetical protein (PDB code: 1vjg), as identified using a structural search engine (*e.g.*, DALI).

[29] As used herein, the term “bleaching” refers to the treatment of a material (*e.g.*, fabric, laundry, pulp, etc.) or surface for a sufficient length of time and under appropriate pH and temperature conditions to effect a brightening (*i.e.*, whitening) and/or cleaning of the material. Examples of chemicals suitable for bleaching include but are not limited to ClO₂, H₂O₂, peracids, NO₂, etc.

[30] As used herein, the term “disinfecting” refers to the removal of contaminants from the surfaces, as well as the inhibition or killing of microbes on the surfaces of items. It is not intended that the present invention be limited to any particular surface, item, or contaminant(s) or microbes to be removed.

[31] As used herein, the term “perhydrolase” refers to an enzyme that is capable of catalyzing a reaction that results in the formation of sufficiently high amounts of peracid suitable for applications such as cleaning, bleaching, and disinfecting. Generally, the perhydrolase enzymes of the present invention produce very high perhydrolysis to hydrolysis ratios. The high perhydrolysis to hydrolysis ratios of these distinct enzymes makes these enzymes suitable for use in a very wide variety of applications. Generally, the perhydrolases of the present invention are characterized by having distinct tertiary structure and primary sequence. Generally, the perhydrolases of the present invention comprises distinct primary and tertiary structures. Generally, the perhydrolases of the present invention comprise distinct quaternary structure. In some embodiments, the perhydrolase is the *M. smegmatis* perhydrolase (SEQ ID NO:1), or a variant or homolog thereof. However, it is not intended that the present invention be limited to this specific *M. smegmatis* perhydrolase, specific variants of this perhydrolase, nor specific homologs of this perhydrolase. In some embodiments, a monomeric hydrolase is engineered to produce a multimeric enzyme that has better perhydrolase activity than the monomer. In some embodiments, a perhydrolase of the invention is a polypeptide identified by methods described herein (see, *e.g.*, Example 7), such as, but not limited to, any of the polypeptides having the sequences set forth in SEQ ID NOs:1 and 10-48.

[32] As used herein, the term “multimer” refers to two or more proteins or peptides that are covalently or non-covalently associated and exist as a complex in solution. A “dimer” is a multimer that contains two proteins or peptides; a “trimer” contains three proteins or peptides, etc. As used herein, “octamer” refers to a multimer of eight proteins or peptides.

[33] As used herein, the phrase “perhydrolysis to hydrolysis ratio” is the ratio of the amount of enzymatically produced peracid to that of enzymatically produced acid by the perhydrolase, under defined conditions and within a defined time. In some embodiments, the assays provided herein are used to determine the amounts of peracid and acid produced by the enzyme.

[34] As used herein, “personal care products” means products used in the cleaning, bleaching and/or disinfecting of hair, skin, scalp, and teeth, including, but not limited to shampoos, body lotions, shower gels, topical moisturizers, toothpaste, and/or other topical cleansers. In some embodiments, these products are utilized on humans, while in other embodiments, these products find use with non-human animals (*e.g.*, in veterinary applications).

[35] As used herein, “pharmaceutically-acceptable” means that drugs, medicaments and/or inert ingredients which the term describes are suitable for use in contact with the tissues of humans and other animals without undue toxicity, incompatibility, instability, irritation, allergic response, and the like, commensurate with a reasonable benefit/risk ratio.

[36] As used herein, “cleaning compositions” and “cleaning formulations” refer to compositions that find use in the removal of undesired compounds from items to be cleaned, such as fabric, dishes, contact lenses, other solid substrates, hair (shampoos), skin (soaps and creams), teeth (mouthwashes, toothpastes), etc. The term encompasses any materials/compounds selected for the particular type of cleaning composition desired and the form of the product (*e.g.*, liquid, gel, granule, or spray composition), as long as the composition is compatible with the perhydrolase and other enzyme(s) used in the composition. The specific selection of cleaning composition materials are readily made by considering the surface, item or fabric to be cleaned, and the desired form of the composition for the cleaning conditions during use.

[37] The terms further refer to any composition that is suited for cleaning, bleaching, disinfecting, and/or sterilizing any object and/or surface. It is intended that the terms include, but are not limited to detergent compositions (*e.g.*, liquid and/or solid laundry detergents and fine fabric detergents; hard surface cleaning formulations, such as for glass, wood, ceramic and metal counter tops and windows; carpet cleaners; oven cleaners; fabric fresheners; fabric softeners; and textile and laundry pre-spotter, as well as dish detergents).

[38] Indeed, the term “cleaning composition” as used herein, includes unless otherwise indicated, granular or powder-form all-purpose or heavy-duty washing agents, especially

cleaning detergents; liquid, gel or paste-form all-purpose washing agents, especially the so-called heavy-duty liquid (HDL) types; liquid fine-fabric detergents; hand dishwashing agents or light duty dishwashing agents, especially those of the high-foaming type; machine dishwashing agents, including the various tablet, granular, liquid and rinse-aid types for household and institutional use; liquid cleaning and disinfecting agents, including antibacterial hand-wash types, cleaning bars, mouthwashes, denture cleaners, car or carpet shampoos, bathroom cleaners; hair shampoos and hair-rinses; shower gels and foam baths and metal cleaners; as well as cleaning auxiliaries such as bleach additives and "stain-stick" or pre-treat types.

[39] As used herein, the terms "detergent composition" and "detergent formulation" are used in reference to mixtures which are intended for use in a wash medium for the cleaning of soiled objects. In some embodiments, the term is used in reference to laundering fabrics and/or garments (*e.g.*, "laundry detergents"). In alternative embodiments, the term refers to other detergents, such as those used to clean dishes, cutlery, etc. (*e.g.*, "dishwashing detergents"). It is not intended that the present invention be limited to any particular detergent formulation or composition. Indeed, it is intended that in addition to perhydrolase, the term encompasses detergents that contain surfactants, transferase(s), hydrolytic enzymes, oxido reductases, builders, bleaching agents, bleach activators, bluing agents and fluorescent dyes, caking inhibitors, masking agents, enzyme activators, antioxidants, and solubilizers.

[40] As used herein, "enhanced performance" in a detergent is defined as increasing cleaning of bleach-sensitive stains (*e.g.*, grass, tea, wine, blood, dingy, etc.), as determined by usual evaluation after a standard wash cycle. In particular embodiments, the perhydrolase of the present invention provides enhanced performance in the oxidation and removal of colored stains and soils. In further embodiments, the perhydrolase of the present invention provides enhanced performance in the removal and/or decolorization of stains. In yet additional embodiments, the perhydrolase of the present invention provides enhanced performance in the removal of lipid-based stains and soils. In still further embodiments, the perhydrolase of the present invention provides enhanced performance in removing soils and stains from dishes and other items.

[41] As used herein the term "hard surface cleaning composition," refers to detergent compositions for cleaning hard surfaces such as floors, walls, tile, bath and kitchen fixtures, and the like. Such compositions are provided in any form, including but not limited to solids, liquids, emulsions, etc.

[42] As used herein, "dishwashing composition" refers to all forms for compositions for cleaning dishes, including but not limited to granular and liquid forms.

[43] As used herein, "fabric cleaning composition" refers to all forms of detergent compositions for cleaning fabrics, including but not limited to, granular, liquid and bar forms.

[44] As used herein, "textile" refers to woven fabrics, as well as staple fibers and filaments suitable for conversion to or use as yarns, woven, knit, and non-woven fabrics. The term encompasses yarns made from natural, as well as synthetic (*e.g.*, manufactured) fibers.

[45] As used herein, "textile materials" is a general term for fibers, yarn intermediates, yarn, fabrics, and products made from fabrics (*e.g.*, garments and other articles).

[46] As used herein, "fabric" encompasses any textile material. Thus, it is intended that the term encompass garments, as well as fabrics, yarns, fibers, non-woven materials, natural materials, synthetic materials, and any other textile material.

[47] As used herein, the term "compatible," means that the cleaning composition materials do not reduce the enzymatic activity of the perhydrolase to such an extent that the perhydrolase is not effective as desired during normal use situations. Specific cleaning composition materials are exemplified in detail hereinafter.

[48] As used herein, "effective amount of perhydrolase enzyme" refers to the quantity of perhydrolase enzyme necessary to achieve the enzymatic activity required in the specific application (*e.g.*, personal care product, cleaning composition, etc.). Such effective amounts are readily ascertained by one of ordinary skill in the art and are based on many factors, such as the particular enzyme variant used, the cleaning application, the specific composition of the cleaning composition, and whether a liquid or dry (*e.g.*, granular, bar) composition is required, and the like.

[49] As used herein, "non-fabric cleaning compositions" encompass hard surface cleaning compositions, dishwashing compositions, personal care cleaning compositions (*e.g.*, oral cleaning compositions, denture cleaning compositions, personal cleansing compositions, etc.), and compositions suitable for use in the pulp and paper industry.

[50] As used herein, "oral cleaning compositions" refers to dentifrices, toothpastes, toothgels, toothpowders, mouthwashes, mouth sprays, mouth gels, chewing gums, lozenges, sachets, tablets, biogels, prophylaxis pastes, dental treatment solutions, and the like. Oral care compositions that find use in conjunction with the perhydrolases of the present invention are

well known in the art (*See e.g.*, U.S. Patent Nos. 5,601,750, 6,379,653, and 5,989,526, all of which are incorporated herein by reference).

[51] As used herein, "pulp treatment compositions" refers to the use of the present perhydrolase enzymes in compositions suitable for use in papermaking. It is intended that the term encompass compositions suitable for the treatment of any pulp material, including wood, as well as non-wood materials, such as "agricultural residues" and "fiber crops," including but not limited to wheat straw, rice straw, corn stalks, bagasse (sugar cane), rye grass straw, seed flax straw, flax straw, kenaf, industrial hemp, sisal, textile flat straw, hesperaloe, etc. Thus, the present invention also encompasses the use of the perhydrolases of the present invention in pulp treatment methods.

[52] As used herein, "oxidizing chemical" refers to a chemical that has the capability of bleaching pulp or any other material. The oxidizing chemical is present at an amount, pH and temperature suitable for bleaching. The term includes, but is not limited to hydrogen peroxide and peracids.

[53] As used herein, "acyl" is the general name for organic acid groups, which are the residues of carboxylic acids after removal of the -OH group (*e.g.*, ethanoyl chloride, $\text{CH}_3\text{CO}-\text{Cl}$, is the acyl chloride formed from ethanoic acid, $\text{CH}_3\text{COO}-\text{H}$). The names of the individual acyl groups are formed by replacing the "-ic" of the acid by "-yl."

[54] As used herein, the term "acylation" refers to the chemical transformation which substitutes the acyl ($\text{RCO}-$) group into a molecule, generally for an active hydrogen of an -OH group.

[55] As used herein, the term "transferase" refers to an enzyme that catalyzes the transfer of functional compounds to a range of substrates.

[56] As used herein, "leaving group" refers to the nucleophile which is cleaved from the acyl donor upon substitution by another nucleophile.

[57] As used herein, the term "enzymatic conversion" refers to the modification of a substrate to an intermediate or the modification of an intermediate to an end-product by contacting the substrate or intermediate with an enzyme. In some embodiments, contact is made by directly exposing the substrate or intermediate to the appropriate enzyme. In other embodiments, contacting comprises exposing the substrate or intermediate to an organism that expresses and/or

excretes the enzyme, and/or metabolizes the desired substrate and/or intermediate to the desired intermediate and/or end-product, respectively.

[58] As used herein, the phrase “detergent stability” refers to the stability of a detergent composition. In some embodiments, the stability is assessed during the use of the detergent, while in other embodiments, the term refers to the stability of a detergent composition during storage.

[59] As used herein, the phrase, “stability to proteolysis” refers to the ability of a protein (*e.g.*, an enzyme) to withstand proteolysis. It is not intended that the term be limited to the use of any particular protease to assess the stability of a protein.

[60] As used herein, “oxidative stability” refers to the ability of a protein to function under oxidative conditions. In particular, the term refers to the ability of a protein to function in the presence of various concentrations of H₂O₂ and/or peracid. Stability under various oxidative conditions can be measured either by standard procedures known to those in the art and/or by the methods described herein. A substantial change in oxidative stability is evidenced by at least about a 5% or greater increase or decrease (in most embodiments, it is preferably an increase) in the half-life of the enzymatic activity, as compared to the enzymatic activity present in the absence of oxidative compounds.

[61] As used herein, “pH stability” refers to the ability of a protein to function at a particular pH. In general, most enzymes have a finite pH range at which they will function. In addition to enzymes that function in mid-range pHs (*i.e.*, around pH 7), there are enzymes that are capable of working under conditions with very high or very low pHs. Stability at various pHs can be measured either by standard procedures known to those in the art and/or by the methods described herein. A substantial change in pH stability is evidenced by at least about 5% or greater increase or decrease (in most embodiments, it is preferably an increase) in the half-life of the enzymatic activity, as compared to the enzymatic activity at the enzyme’s optimum pH. However, it is not intended that the present invention be limited to any pH stability level nor pH range.

[62] As used herein, “thermal stability” refers to the ability of a protein to function at a particular temperature. In general, most enzymes have a finite range of temperatures at which they will function. In addition to enzymes that work in mid-range temperatures (*e.g.*, room temperature), there are enzymes that are capable of working in very high or very low

temperatures. Thermal stability can be measured either by known procedures or by the methods described herein. A substantial change in thermal stability is evidenced by at least about 5% or greater increase or decrease (in most embodiments, it is preferably an increase) in the half-life of the catalytic activity of a mutant when exposed to a different temperature (*i.e.*, higher or lower) than optimum temperature for enzymatic activity. However, it is not intended that the present invention be limited to any temperature stability level nor temperature range.

[63] As used herein, the term “chemical stability” refers to the stability of a protein (*e.g.*, an enzyme) towards chemicals that adversely affect its activity. In some embodiments, such chemicals include, but are not limited to hydrogen peroxide, peracids, anionic detergents, cationic detergents, non-ionic detergents, chelants, etc. However, it is not intended that the present invention be limited to any particular chemical stability level nor range of chemical stability.

[64] As used herein, the phrase “perhydrolase activity improvement” refers to the relative improvement of perhydrolase activity, in comparison with a standard enzyme. In some embodiments, the term refers to an improved rate of perhydrolysis product, while in other embodiments, the term encompasses perhydrolase compositions that produce less hydrolysis product. In additional embodiments, the term refers to perhydrolase compositions with altered substrate specificity.

[65] As used herein, the phrase “alteration in substrate specificity” refers to changes in the substrate specificity of an enzyme. In some embodiments, a change in substrate specificity is defined as a difference between the K_{cat}/K_m ratio observed with an enzyme compared to enzyme variants or other enzyme compositions. Enzyme substrate specificities vary, depending upon the substrate tested. The substrate specificity of an enzyme is determined by comparing the catalytic efficiencies it exhibits with different substrates. These determinations find particular use in assessing the efficiency of mutant enzymes, as it is generally desired to produce variant enzymes that exhibit greater ratios for particular substrates of interest. For example, the perhydrolase enzymes of the present invention are more efficient in producing peracid from an ester substrate than enzymes currently being used in cleaning, bleaching and disinfecting applications. Another example of the present invention is a perhydrolase with a lower activity on peracid degradation compared to the wild type. Another example of the present invention is a perhydrolase with higher activity on more hydrophobic acyl groups than acetic acid. However, it

is not intended that the present invention be limited to any particular substrate composition nor any specific substrate specificity.

[66] As used herein, “surface property” is used in reference to an electrostatic charge, as well as properties such as the hydrophobicity and/or hydrophilicity exhibited by the surface of a protein.

[67] As used herein, the phrase “is independently selected from the group consisting of . . .” or “selected from the group consisting of . . .” means that moieties or elements that are selected from the referenced *Markush* group can be the same, can be different or any mixture of elements as indicated in the following example:

[68] In reference to chemical compositions, the term “substituted” as used herein, means that the organic composition or radical to which the term is applied is:

- (a) made unsaturated by the elimination of at least one element or radical; or
- (b) at least one hydrogen in the compound or radical is replaced with a moiety containing one or more (i) carbon, (ii) oxygen, (iii) sulfur, (iv) nitrogen or (v) halogen atoms; or
- (c) both (a) and (b).

[69] Moieties which may replace hydrogen as described in (b) immediately above, that contain only carbon and hydrogen atoms, are hydrocarbon moieties including, but not limited to, alkyl, alkenyl, alkynyl, alkylidenyl, cycloalkyl, phenyl, alkyl phenyl, naphthyl, anthryl, phenanthryl, fluoryl, steroid groups, and combinations of these groups with each other and with polyvalent hydrocarbon groups such as alkylene, alkylidene and alkylidyne groups. Moieties containing oxygen atoms that may replace hydrogen as described in (b) immediately above include, but are not limited to, hydroxy, acyl or keto, ether, epoxy, carboxy, and ester containing groups. Moieties containing sulfur atoms that may replace hydrogen as described in (b) immediately above include, but are not limited to, the sulfur-containing acids and acid ester groups, thioether groups, mercapto groups and thioketo groups. Moieties containing nitrogen atoms that may replace hydrogen as described in (b) immediately above include, but are not limited to, amino groups, the nitro group, azo groups, ammonium groups, amide groups, azido groups, isocyanate groups, cyano groups and nitrile groups. Moieties containing halogen atoms that may replace hydrogen as described in (b) immediately above include chloro, bromo, fluoro, iodo groups and any of the moieties previously described where a hydrogen or a pendant alkyl group is substituted by a halo group to form a stable substituted moiety.

[70] It is understood that any of the above moieties (b)(i) through (b)(v) can be substituted into each other in either a monovalent substitution or by loss of hydrogen in a polyvalent substitution to form another monovalent moiety that can replace hydrogen in the organic compound or radical.

[71] As used herein, the terms "purified" and "isolated" refer to the removal of contaminants from a sample. For example, perhydrolases are purified by removal of contaminating proteins and other compounds within a solution or preparation that are not perhydrolases. In some embodiments, recombinant perhydrolases are expressed in bacterial or fungal host cells and these recombinant perhydrolases are purified by the removal of other host cell constituents; the percent of recombinant perhydrolase polypeptides is thereby increased in the sample.

[72] As used herein, "protein of interest," refers to a protein (*e.g.*, an enzyme or "enzyme of interest") which is being analyzed, identified and/or modified. Naturally-occurring, as well as recombinant proteins find use in the present invention.

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

[74] As used herein, functionally and/or structurally similar proteins are considered to be "related proteins." In some embodiments, these proteins are derived from a different genus and/or species, including differences between classes of organisms (*e.g.*, a bacterial protein and a fungal protein). In some embodiments, these proteins are derived from a different genus and/or species, including differences between classes of organisms (*e.g.*, a bacterial enzyme and a fungal enzyme). In additional embodiments, related proteins are provided from the same species. Indeed, it is not intended that the present invention be limited to related proteins from

any particular source(s). In addition, the term “related proteins” encompasses tertiary structural homologs and primary sequence homologs (*e.g.*, the perhydrolase of the present invention). In further embodiments, the term encompasses proteins that are immunologically cross-reactive. Generally, the related proteins of the present invention very high ratios of perhydrolysis to hydrolysis.

[75] As used herein, the term “derivative” refers to a protein which is derived from a protein by addition of one or more amino acids to either or both the C- and N-terminal end(s), substitution of one or more amino acids at one or a number of different sites in the amino acid sequence, and/or deletion of one or more amino acids at either or both ends of the protein or at one or more sites in the amino acid sequence, and/or insertion of one or more amino acids at one or more sites in the amino acid sequence. The preparation of a protein derivative is preferably achieved by modifying a DNA sequence which encodes for the native protein, transformation of that DNA sequence into a suitable host, and expression of the modified DNA sequence to form the derivative protein.

[76] Related (and derivative) proteins comprise “variant proteins.” In some embodiments, variant proteins differ from a parent protein and one another by a small number of amino acid residues. The number of differing amino acid residues may be one or more, preferably 1, 2, 3, 4, 5, 10, 15, 20, 30, 40, 50, or more amino acid residues. In some embodiments, the number of different amino acids between variants is between 1 and 10. In some embodiments, related proteins and particularly variant proteins comprise at least 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, or 99% amino acid sequence identity. Additionally, a related protein or a variant protein as used herein, refers to a protein that differs from another related protein or a parent protein in the number of prominent regions. For example, in some embodiments, variant proteins have 1, 2, 3, 4, 5, or 10 corresponding prominent regions that differ from the parent protein.

[77] Several methods are known in the art that are suitable for generating variants of the perhydrolase enzymes of the present invention, including but not limited to site-saturation mutagenesis, scanning mutagenesis, insertional mutagenesis, random mutagenesis, site-directed mutagenesis, and directed-evolution, as well as various other recombinatorial approaches.

[78] In some embodiments, homologous proteins are engineered to produce enzymes with the desired activity(ies). In some embodiments, the engineered proteins are included within the SGNH-hydrolase family of proteins. In some embodiments, the engineered proteins comprise at

least one or a combination of the following conserved residues in reference to SEQ ID NO:1: L6, W14, W34, L38, R56, D62, L74, L78, H81, P83, M90, K97, G110, L114, L135, F180, G205. In alternative embodiments, these engineered proteins comprise the GDSL-GRTT and/or ARTT motifs. In further embodiments, the enzymes are multimers, including but not limited to dimers, octamers, and tetramers. In yet additional embodiments, the engineered proteins exhibit a perhydrolysis to hydrolysis ratio that is greater than 1.

[79] An amino acid residue of a perhydrolase is equivalent to a residue of *M. smegmatis* perhydrolase if it is either homologous (*i.e.*, having a corresponding position in either the primary and/or tertiary structure) or analogous to a specific residue or portion of that residue in *M. smegmatis* perhydrolase (*i.e.*, having the same or similar functional capacity to combine, react, and/or chemically interact).

[80] In some embodiments, in order to establish homology to primary structure, the amino acid sequence of a perhydrolase is directly compared to the *M. smegmatis* perhydrolase primary sequence and particularly to a set of residues known to be invariant in all perhydrolases for which sequence is known. After aligning the conserved residues, allowing for necessary insertions and deletions in order to maintain alignment (*i.e.*, avoiding the elimination of conserved residues through arbitrary deletion and insertion), the residues equivalent to particular amino acids in the primary sequence of *M. smegmatis* perhydrolase are defined. In some embodiments, alignment of conserved residues conserves 100% of such residues. However, alignment of greater than 75% or as little as 50% of conserved residues are also adequate to define equivalent residues. In some embodiments, conservation of the catalytic serine and histidine residues are maintained.

Conserved residues are used to define the corresponding equivalent amino acid residues of *M. smegmatis* perhydrolase in other perhydrolases (*e.g.*, perhydrolases from other *Mycobacterium* species, as well as any other organisms).

[81] In some embodiments of the present invention, the DNA sequence encoding *M. smegmatis* perhydrolase is modified. In some embodiments, the following residues are modified: Cys7, Asp10, Ser11, Leu12, Thr13, Trp14, Trp16, Pro24, Thr25, Leu53, Ser54, Ala55, Thr64, Asp65, Arg67, Cys77, Thr91, Asn94, Asp95, Tyr99, Val125, Pro138, Leu140, Pro146, Pro148, Trp149, Phe150, Ile153, Phe154, Thr159, Thr186, Ile192, Ile194, and Phe196. However, it is not intended that the present invention be limited to sequences that are modified

at these positions. Indeed, it is intended that the present invention encompass various modifications and combinations of modifications.

[82] In additional embodiments, equivalent residues are defined by determining homology at the level of tertiary structure for a perhydrolase whose tertiary structure has been determined by x-ray crystallography. In this context, "equivalent residues" are defined as those for which the atomic coordinates of two or more of the main chain atoms of a particular amino acid residue of the carbonyl hydrolase and *M. smegmatis* perhydrolase (N on N, CA on CA, C on C, and O on O) are within 0.13nm and preferably 0.1 nm after alignment. Alignment is achieved after the best model has been oriented and positioned to give the maximum overlap of atomic coordinates of non-hydrogen protein atoms of the perhydrolase in question to the *M. smegmatis* perhydrolase. As known in the art, the best model is the crystallographic model giving the lowest R factor for experimental diffraction data at the highest resolution available. Equivalent residues which are functionally and/or structurally analogous to a specific residue of *M. smegmatis* perhydrolase are defined as those amino acids of the perhydrolases that preferentially adopt a conformation such that they either alter, modify or modulate the protein structure, to effect changes in substrate binding and/or catalysis in a manner defined and attributed to a specific residue of the *M. smegmatis* perhydrolase. Further, they are those residues of the perhydrolase (in cases where a tertiary structure has been obtained by x-ray crystallography), which occupy an analogous position to the extent that although the main chain atoms of the given residue may not satisfy the criteria of equivalence on the basis of occupying a homologous position, the atomic coordinates of at least two of the side chain atoms of the residue lie within 0.13 nm of the corresponding side chain atoms of *M. smegmatis* perhydrolase. The coordinates of the three dimensional structure of *M. smegmatis* perhydrolase have been determined and are set forth in WO 2005/056782, hereby incorporated by reference in its entirety, and find use in determining equivalent residues on the level of tertiary structure.

[83] In some embodiments, some of the residues identified for substitution, insertion or deletion are conserved residues whereas others are not. The perhydrolase mutants of the present invention include various mutants, including those encoded by a nucleic acid that comprises a signal sequence. In some embodiments, perhydrolase mutants that are encoded by such a sequence are secreted by an expression host. In some further embodiments, the nucleic acid sequence comprises a homolog having a secretion signal.

[84] Characterization of wild-type and mutant proteins is accomplished via any means suitable and is preferably based on the assessment of properties of interest. For example, pH and/or temperature, as well as detergent and /or oxidative stability is/are determined in some embodiments of the present invention. Indeed, it is contemplated that enzymes having various degrees of stability in one or more of these characteristics (pH, temperature, proteolytic stability, detergent stability, and/or oxidative stability) will find use. In still other embodiments, perhydrolases with low peracid degradation activity are selected.

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

[86] In some embodiments, the perhydrolase gene is ligated into an appropriate expression plasmid. The cloned perhydrolase gene is then used to transform or transfect a host cell in order to express the perhydrolase gene. This plasmid may replicate in hosts in the sense that it contains the well-known elements necessary for plasmid replication or the plasmid may be designed to integrate into the host chromosome. The necessary elements are provided for efficient gene expression (*e.g.*, a promoter operably linked to the gene of interest). In some embodiments, these necessary elements are supplied as the gene's own homologous promoter if it is recognized, (*i.e.*, transcribed, by the host), a transcription terminator (a polyadenylation region for eukaryotic host cells) which is exogenous or is supplied by the endogenous terminator region of the perhydrolase gene. In some embodiments, a selection gene such as an antibiotic resistance gene that enables continuous cultural maintenance of plasmid-infected host cells by growth in antimicrobial-containing media is also included.

[87] The following cassette mutagenesis method may be used to facilitate the construction of the perhydrolase variants of the present invention, although other methods may be used.

[88] First, as described herein, a naturally-occurring gene encoding the perhydrolase is obtained and sequenced in whole or in part. Then, the sequence is scanned for a point at which it is desired to make a mutation (deletion, insertion or substitution) of one or more amino acids in the encoded perhydrolase. The sequences flanking this point are evaluated for the presence of restriction sites for replacing a short segment of the gene with an oligonucleotide pool which when expressed will encode various mutants. Such restriction sites are preferably unique sites within the protein gene so as to facilitate the replacement of the gene segment. However, any convenient restriction site which is not overly redundant in the perhydrolase gene may be used, provided the gene fragments generated by restriction digestion can be reassembled in proper sequence. If restriction sites are not present at locations within a convenient distance from the selected point (from 10 to 15 nucleotides), such sites are generated by substituting nucleotides in the gene in such a fashion that neither the reading frame nor the amino acids encoded are changed in the final construction. Mutation of the gene in order to change its sequence to conform to the desired sequence is accomplished by M13 primer extension in accord with generally known methods. The task of locating suitable flanking regions and evaluating the needed changes to arrive at two convenient restriction site sequences is made routine by the redundancy of the genetic code, a restriction enzyme map of the gene and the large number of different restriction enzymes. Note that if a convenient flanking restriction site is available, the above method need be used only in connection with the flanking region which does not contain a site.

[89] Once the naturally-occurring DNA and/or synthetic DNA is cloned, the restriction sites flanking the positions to be mutated are digested with the cognate restriction enzymes and a plurality of end termini-complementary oligonucleotide cassettes are ligated into the gene. The mutagenesis is simplified by this method because all of the oligonucleotides can be synthesized so as to have the same restriction sites, and no synthetic linkers are necessary to create the restriction sites.

[90] As used herein, "corresponding to," refers to a residue at the enumerated position in a protein or peptide, or a residue that is homologous, or equivalent to an enumerated residue in a protein or peptide.

[91] As used herein, "corresponding region," generally refers to an analogous position along related proteins or a parent protein.

[92] The terms "nucleic acid molecule encoding," "nucleic acid sequence encoding," "DNA sequence encoding," and "DNA encoding" refer to the order or sequence of deoxyribonucleotides along a strand of deoxyribonucleic acid. The order of these deoxyribonucleotides determines the order of amino acids along the polypeptide (protein) chain. The DNA sequence thus codes for the amino acid sequence.

[93] As used herein, "homologous protein" refers to a protein (*e.g.*, perhydrolase) that has similar action and/or structure, as a protein of interest (*e.g.*, an perhydrolase from another source). It is not intended that homologs be necessarily related evolutionarily. Thus, it is intended that the term encompass the same or similar enzyme(s) (*i.e.*, in terms of structure and function) obtained from different species. In some some embodiments, it is desirable to identify a homolog that has a quaternary, tertiary and/or primary structure similar to the protein of interest, as replacement for the segment or fragment in the protein of interest with an analogous segment from the homolog will reduce the disruptiveness of the change.

[94] As used herein, "homologous genes" refers to at least a pair of genes from different species, which genes correspond to each other and which are identical or very similar to each other. The term encompasses genes that are separated by speciation (*i.e.*, the development of new species) (*e.g.*, orthologous genes), as well as genes that have been separated by genetic duplication (*e.g.*, paralogous genes). These genes encode "homologous proteins."

[95] The degree of homology between sequences may be determined using any suitable method known in the art (*See e.g.*, Smith and Waterman, *Adv. Appl. Math.*, 2:482 [1981]; Needleman and Wunsch, *J. Mol. Biol.*, 48:443 [1970]; Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444 [1988]; programs such as GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package (Genetics Computer Group, Madison, WI); and Devereux *et al.*, *Nucl. Acid Res.*, 12:387-395 [1984]).

[96] For example, PILEUP is a useful program to determine sequence homology levels. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments. It can also plot a tree showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng and Doolittle, (Feng and Doolittle, *J. Mol. Evol.*, 35:351-360 [1987]). The method is

similar to that described by Higgins and Sharp (Higgins and Sharp, CABIOS 5:151-153 [1989]).

Useful PILEUP parameters including a default gap weight of 3.00, a default gap length weight of 0.10, and weighted end gaps. Another example of a useful algorithm is the BLAST algorithm, described by Altschul *et al.*, (Altschul *et al.*, J. Mol. Biol., 215:403-410, [1990]; and Karlin *et al.*, Proc. Natl. Acad. Sci. USA 90:5873-5787 [1993]). One particularly useful BLAST program is the WU-BLAST-2 program (See, Altschul *et al.*, Meth. Enzymol., 266:460-480 [1996]). parameters "W," "T," and "X" determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a wordlength (W) of 11, the BLOSUM62 scoring matrix (See, Henikoff and Henikoff, Proc. Natl. Acad. Sci. USA 89:10915 [1989]) alignments (B) of 50, expectation (E) of 10, M⁵, N⁻⁴, and a comparison of both strands.

[97] As used herein, "wild-type" and "native" proteins are those found in nature. The terms "wild-type sequence," and "wild-type gene" are used interchangeably herein, to refer to a sequence that is native or naturally occurring in a host cell. In some embodiments, the wild-type sequence refers to a sequence of interest that is the starting point of a protein engineering project. The genes encoding the naturally-occurring protein may be obtained in accord with the general methods known to those skilled in the art. The methods generally comprise synthesizing labeled probes having putative sequences encoding regions of the protein of interest, preparing genomic libraries from organisms expressing the protein, and screening the libraries for the gene of interest by hybridization to the probes. Positively hybridizing clones are then mapped and sequenced.

[98] The term "recombinant DNA molecule" as used herein refers to a DNA molecule that is comprised of segments of DNA joined together by means of molecular biological techniques.

[99] The term "recombinant oligonucleotide" refers to an oligonucleotide created using molecular biological manipulations, including but not limited to, the ligation of two or more oligonucleotide sequences generated by restriction enzyme digestion of a polynucleotide sequence, the synthesis of oligonucleotides (*e.g.*, the synthesis of primers or oligonucleotides) and the like.

[100] As used herein, "percent (%) nucleic acid sequence identity" is defined as the percentage of nucleotide residues in a candidate sequence that are identical with the nucleotide residues of the sequence.

[101] As used herein, the term "hybridization" refers to the process by which a strand of nucleic acid joins with a complementary strand through base pairing, as known in the art.

[102] As used herein, the phrase "hybridization conditions" refers to the conditions under which hybridization reactions are conducted. These conditions are typically classified by degree of "stringency" of the conditions under which hybridization is measured. The degree of stringency can be based, for example, on the melting temperature (T_m) of the nucleic acid binding complex or probe. For example, "maximum stringency" typically occurs at about $T_m - 5^\circ\text{C}$ (5° below the T_m of the probe); "high stringency" at about $5 - 10^\circ$ below the T_m ; "intermediate stringency" at about $10 - 20^\circ$ below the T_m of the probe; and "low stringency" at about $20 - 25^\circ$ below the T_m . Alternatively, or in addition, hybridization conditions can be based upon the salt or ionic strength conditions of hybridization and/or one or more stringency washes. For example, $6\times\text{SSC}$ = very low stringency; $3\times\text{SSC}$ = low to medium stringency; $1\times\text{SSC}$ = medium stringency; and $0.5\times\text{SSC}$ = high stringency. Functionally, maximum stringency conditions may be used to identify nucleic acid sequences having strict identity or near-strict identity with the hybridization probe; while high stringency conditions are used to identify nucleic acid sequences having about 80% or more sequence identity with the probe.

[103] For applications requiring high selectivity, it is typically desirable to use relatively stringent conditions to form the hybrids (e.g., relatively low salt and/or high temperature conditions are used).

[104] The phrases "substantially similar and "substantially identical" in the context of at least two nucleic acids or polypeptides typically means that a polynucleotide or polypeptide comprises a sequence that has at least about 40%, 50%, 60%, 75%, 80%, 90%, 95%, 97%, 98%, 99%, or 99.5% sequence identity, compared to a reference (*i.e.*, wild-type) sequence. Sequence identity may be determined using known programs such as BLAST, ALIGN, and CLUSTAL using standard parameters. (See e.g., Altschul, *et al.*, J. Mol. Biol. 215:403-410 [1990]; Henikoff *et al.*, Proc. Natl. Acad. Sci. USA 89:10915 [1989]; Karin *et al.*, Proc. Natl. Acad. Sci. USA 90:5873 [1993]; and Higgins *et al.*, Gene 73:237 - 244 [1988]). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. Also, databases may be searched using FASTA (Pearson *et al.*, Proc. Natl. Acad. Sci. USA 85:2444-2448 [1988]). One indication that two polypeptides are substantially identical is that the first polypeptide is immunologically cross-reactive with the second polypeptide. Typically, polypeptides that differ by conservative amino acid substitutions are immunologically cross-

reactive. Thus, a polypeptide is substantially identical to a second polypeptide, for example, where the two peptides differ only by a conservative substitution. Another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions (*e.g.*, within a range of medium to high stringency).

[105] As used herein, "equivalent residues" refers to proteins that share particular amino acid residues. For example, equivalent residues may be identified by determining homology at the level of tertiary structure for a protein (*e.g.*, perhydrolase) whose tertiary structure has been determined by x-ray crystallography. Equivalent residues are defined as those for which the atomic coordinates of two or more of the main chain atoms of a particular amino acid residue of the protein having putative equivalent residues and the protein of interest (N on N, CA on CA, C on C and O on O) are within 0.13 nm and preferably 0.1 nm after alignment. Alignment is achieved after the best model has been oriented and positioned to give the maximum overlap of atomic coordinates of non-hydrogen protein atoms of the proteins analyzed. The preferred model is the crystallographic model giving the lowest R factor for experimental diffraction data at the highest resolution available, determined using methods known to those skilled in the art of crystallography and protein characterization/analysis.

[106] As used herein, the terms "hybrid perhydrolases" and "fusion perhydrolases" refer to proteins that are engineered from at least two different or "parental" proteins. In some embodiments, these parental proteins are homologs of one another. For example, in some embodiments, a hybrid perhydrolase or fusion protein contains the N-terminus of a protein and the C-terminus of a homolog of the protein. In some embodiments, the two terminal ends are combined to correspond to the full-length active protein.

[107] The term "regulatory element" as used herein refers to a genetic element that controls some aspect of the expression of nucleic acid sequences. For example, a promoter is a regulatory element which facilitates the initiation of transcription of an operably linked coding region. Additional regulatory elements include splicing signals, polyadenylation signals and termination signals.

[108] As used herein, "host cells" are generally prokaryotic or eukaryotic hosts which are transformed or transfected with vectors constructed using recombinant DNA techniques known in the art. Transformed host cells are capable of either replicating vectors encoding the protein variants or expressing the desired protein variant. In the case of vectors which encode the pre- or

prepro-form of the protein variant, such variants, when expressed, are typically secreted from the host cell into the host cell medium.

[109] The term "introduced" in the context of inserting a nucleic acid sequence into a cell, means transformation, transduction or transfection. Means of transformation include protoplast transformation, calcium chloride precipitation, electroporation, naked DNA and the like as known in the art. (See, Chang and Cohen, *Mol. Gen. Genet.*, 168:111 - 115 [1979]; Smith *et al.*, *Appl. Env. Microbiol.*, 51:634 [1986]; and the review article by Ferrari *et al.*, in Harwood, Bacillus, Plenum Publishing Corporation, pp. 57-72 [1989]). The presence of "splicing signals" on an expression vector often results in higher levels of expression of the recombinant transcript. Splicing signals mediate the removal of introns from the primary RNA transcript and consist of a splice donor and acceptor site (Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, New York [1989], pp. 16.7-16.8). A commonly used splice donor and acceptor site is the splice junction from the 16S RNA of SV40.

[110] The term "stable transfection" or "stably transfected" refers to the introduction and integration of foreign DNA into the genome of the transfected cell. The term "stable transfectant" refers to a cell which has stably integrated foreign or exogenous DNA into the genomic DNA of the transfected cell.

[111] The terms "selectable marker" or "selectable gene product" as used herein refer to the use of a gene which encodes an enzymatic activity that confers resistance to an antibiotic or drug upon the cell in which the selectable marker is expressed.

[112] As used herein, the terms "amplification" and "gene amplification" refer to a process by which specific DNA sequences are disproportionately replicated such that the amplified gene becomes present in a higher copy number than was initially present in the genome. In some embodiments, selection of cells by growth in the presence of a drug (*e.g.*, an inhibitor of an inhibitable enzyme) results in the amplification of either the endogenous gene encoding the gene product required for growth in the presence of the drug or by amplification of exogenous (*i.e.*, input) sequences encoding this gene product, or both. Selection of cells by growth in the presence of a drug (*e.g.*, an inhibitor of an inhibitable enzyme) may result in the amplification of either the endogenous gene encoding the gene product required for growth in the presence of the drug or by amplification of exogenous (*i.e.*, input) sequences encoding this gene product, or both.

[113] "Amplification" is a special case of nucleic acid replication involving template specificity. It is to be contrasted with non-specific template replication (*i.e.*, replication that is template-dependent but not dependent on a specific template). Template specificity is here distinguished from fidelity of replication (*i.e.*, synthesis of the proper polynucleotide sequence) and nucleotide (ribo- or deoxyribo-) specificity. Template specificity is frequently described in terms of "target" specificity. Target sequences are "targets" in the sense that they are sought to be sorted out from other nucleic acid. Amplification techniques have been designed primarily for this sorting out.

[114] As used herein, the term "co-amplification" refers to the introduction into a single cell of an amplifiable marker in conjunction with other gene sequences (*i.e.*, comprising one or more non-selectable genes such as those contained within an expression vector) and the application of appropriate selective pressure such that the cell amplifies both the amplifiable marker and the other, non-selectable gene sequences. The amplifiable marker may be physically linked to the other gene sequences or alternatively two separate pieces of DNA, one containing the amplifiable marker and the other containing the non-selectable marker, may be introduced into the same cell.

[115] As used herein, the terms "amplifiable marker," "amplifiable gene," and "amplification vector" refer to a marker, gene or a vector encoding a gene which permits the amplification of that gene under appropriate growth conditions.

[116] As used herein, the term "amplifiable nucleic acid" refers to nucleic acids which may be amplified by any amplification method. It is contemplated that "amplifiable nucleic acid" will usually comprise "sample template."

[117] As used herein, the term "sample template" refers to nucleic acid originating from a sample which is analyzed for the presence of "target" (defined below). In contrast, "background template" is used in reference to nucleic acid other than sample template which may or may not be present in a sample. Background template is most often inadvertent. It may be the result of carryover, or it may be due to the presence of nucleic acid contaminants sought to be purified away from the sample. For example, nucleic acids from organisms other than those to be detected may be present as background in a test sample.

[118] "Template specificity" is achieved in most amplification techniques by the choice of enzyme. Amplification enzymes are enzymes that, under conditions they are used, will process

only specific sequences of nucleic acid in a heterogeneous mixture of nucleic acid. For example, in the case of Q β replicase, MDV-1 RNA is the specific template for the replicase (*See e.g.*, Kacian *et al.*, Proc. Natl. Acad. Sci. USA 69:3038 [1972]). Other nucleic acids are not replicated by this amplification enzyme. Similarly, in the case of T7 RNA polymerase, this amplification enzyme has a stringent specificity for its own promoters (*See*, Chamberlin *et al.*, Nature 228:227 [1970]). In the case of T4 DNA ligase, the enzyme will not ligate the two oligonucleotides or polynucleotides, where there is a mismatch between the oligonucleotide or polynucleotide substrate and the template at the ligation junction (*See*, Wu and Wallace, Genomics 4:560 [1989]). Finally, *Taq* and *Pfu* polymerases, by virtue of their ability to function at high temperature, are found to display high specificity for the sequences bounded and thus defined by the primers; the high temperature results in thermodynamic conditions that favor primer hybridization with the target sequences and not hybridization with non-target sequences.

[119] As used herein, the term "primer" refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, (*i.e.*, in the presence of nucleotides and an inducing agent such as DNA polymerase and at a suitable temperature and pH). The primer is preferably single stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the inducing agent. The exact lengths of the primers will depend on many factors, including temperature, source of primer and the use of the method.

[120] As used herein, the term "probe" refers to an oligonucleotide (*i.e.*, a sequence of nucleotides), whether occurring naturally as in a purified restriction digest or produced synthetically, recombinantly or by PCR amplification, which is capable of hybridizing to another oligonucleotide of interest. A probe may be single-stranded or double-stranded. Probes are useful in the detection, identification and isolation of particular gene sequences. It is contemplated that any probe used in the present invention will be labeled with any "reporter molecule," so that is detectable in any detection system, including, but not limited to enzyme (*e.g.*, ELISA, as well as enzyme-based histochemical assays), fluorescent, radioactive, and

luminescent systems. It is not intended that the present invention be limited to any particular detection system or label.

[121] As used herein, the term "target," when used in reference to amplification methods (*e.g.*, the polymerase chain reaction), refers to the region of nucleic acid bounded by the primers used for polymerase chain reaction. Thus, the "target" is sought to be sorted out from other nucleic acid sequences. A "segment" is defined as a region of nucleic acid within the target sequence.

[122] As used herein, the term "polymerase chain reaction" ("PCR") refers to the methods of U.S. Patent Nos. 4,683,195, 4,683,202, and 4,965,188, hereby incorporated by reference, which include methods for increasing the concentration of a segment of a target sequence in a mixture of genomic DNA without cloning or purification. This process for amplifying the target sequence consists of introducing a large excess of two oligonucleotide primers to the DNA mixture containing the desired target sequence, followed by a precise sequence of thermal cycling in the presence of a DNA polymerase. The two primers are complementary to their respective strands of the double stranded target sequence. To effect amplification, the mixture is denatured and the primers then annealed to their complementary sequences within the target molecule. Following annealing, the primers are extended with a polymerase so as to form a new pair of complementary strands. The steps of denaturation, primer annealing and polymerase extension can be repeated many times (*i.e.*, denaturation, annealing and extension constitute one "cycle"; there can be numerous "cycles") to obtain a high concentration of an amplified segment of the desired target sequence. The length of the amplified segment of the desired target sequence is determined by the relative positions of the primers with respect to each other, and therefore, this length is a controllable parameter. By virtue of the repeating aspect of the process, the method is referred to as the "polymerase chain reaction" (hereinafter "PCR"). Because the desired amplified segments of the target sequence become the predominant sequences (in terms of concentration) in the mixture, they are said to be "PCR amplified".

[123] As used herein, the term "amplification reagents" refers to those reagents (deoxyribonucleotide triphosphates, buffer, etc.), needed for amplification except for primers, nucleic acid template and the amplification enzyme. Typically, amplification reagents along with other reaction components are placed and contained in a reaction vessel (test tube, microwell, etc.).

[124] With PCR, it is possible to amplify a single copy of a specific target sequence in genomic DNA to a level detectable by several different methodologies (*e.g.*, hybridization with a labeled

probe; incorporation of biotinylated primers followed by avidin-enzyme conjugate detection; incorporation of ^{32}P -labeled deoxynucleotide triphosphates, such as dCTP or dATP, into the amplified segment). In addition to genomic DNA, any oligonucleotide or polynucleotide sequence can be amplified with the appropriate set of primer molecules. In particular, the amplified segments created by the PCR process itself are, themselves, efficient templates for subsequent PCR amplifications.

[125] As used herein, the terms "PCR product," "PCR fragment," and "amplification product" refer to the resultant mixture of compounds after two or more cycles of the PCR steps of denaturation, annealing and extension are complete. These terms encompass the case where there has been amplification of one or more segments of one or more target sequences.

[126] As used herein, the terms "restriction endonucleases" and "restriction enzymes" refer to bacterial enzymes, each of which cut double-stranded DNA at or near a specific nucleotide sequence.

[127] As used herein, the term "polynucleotide" refers to a polymeric form of nucleotides of any length and any three-dimensional structure and single- or multi-stranded (e.g., single-stranded, double-stranded, triple-helical, etc.), which contain deoxyribonucleotides, ribonucleotides, and/or analogs or modified forms of deoxyribonucleotides or ribonucleotides, including modified nucleotides or bases or their analogs. Any type of modified nucleotide or nucleotide analog may be used, so long as the polynucleotide retains the desired functionality under conditions of use, including modifications that increase nuclease resistance (e.g., deoxy, 2'-O--Me, phosphorothioates, etc.). Labels may also be incorporated for purposes of detection or capture, for example, radioactive or nonradioactive labels or anchors, e.g., biotin. The term polynucleotide also includes peptide nucleic acids (PNA). Polynucleotides may be naturally occurring or non-naturally occurring. The terms "polynucleotide" and "nucleic acid" and "oligonucleotide" as used herein are used interchangeably. Polynucleotides of the invention may contain RNA, DNA, or both, and/or modified forms and/or analogs thereof. A sequence of nucleotides may be interrupted by non-nucleotide components. One or more phosphodiester linkages may be replaced by alternative linking groups. These alternative linking groups include, but are not limited to, embodiments wherein phosphate is replaced by P(O)S ("thioate"), P(S)S ("dithioate"), (O)NR.sub.2 ("amidate"), P(O)R, P(O)OR', CO or CH.sub.2 ("formacetal"), in which each R or R' is independently H or substituted or unsubstituted alkyl (1-20 C) optionally containing an ether (--O--) linkage, aryl, alkenyl, cycloalkyl, cycloalkenyl or araldyl. Not all

linkages in a polynucleotide need be identical. Polynucleotides may be linear or circular or comprise a combination of linear and circular portions. The terms "polynucleotide" and "nucleic acid" and "oligonucleotide" as used herein are used interchangeably.

Methods

[128] The present invention provides methods for engineering enzymes belonging to the class of enzymes known as SGNH hydrolases and α/β hydrolases, to create compositions comprising at least one enzyme suitable for use in enzymatic aqueous acylation and/or perhydrolysis. The invention further provides methods for identification of enzymes comprising aqueous acylation and/or perhydrolase activity. An example of such a method of identification is provided in Example 7.

Polypeptides

[129] The invention also provides polypeptides having enzymatic aqueous acylation and/or perhydrolase activity, identified by the methods described herein. The present invention further provides compositions comprising at least one perhydrolase enzyme suitable for use in enzymatic aqueous acylation and/or perhydrolysis. In various embodiments, polypeptides of the invention comprise, consist of, or consist essentially of a sequence selected from the group consisting of SEQ ID NOs: 11-17, 19-27, 29-40, and 42-48 or a variant thereof that has any of at least about 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 99.5% amino acid sequence identity, wherein the polypeptide has enzymatic aqueous acylation and/or perhydrolysis activity.

[130] It will be apparent to one of skill in the art that modifications may be made to a polypeptide without diminishing its biological activity. Some modifications may be made to facilitate the cloning and/or expression of the subject molecule(s). Such modifications are well known to those of skill in the art and include, for example, a methionine added at the amino terminus to provide an initiation site, or additional amino acids introduced as a linker to provide a protease cleavage site, etc.

[131] Polypeptides of the invention may be used in applications in which bleaching by peracids and/or hydrogen peroxide are desired, for example, laundry, fabric treatment (e.g., textile bleaching, before textiles are dyed and/or after incorporation into textile goods), pulp and paper processing, personal care, disinfection, and cleaning of hard surfaces. Polypeptides of the

invention may be used in compositions and methods for sterilization/disinfection of various objects, including but not limited to medical devices, medical equipment, industrial equipment, and fermenters, as well as any additional object that needs to be sterilized or disinfected. The polypeptides may be used in compositions and methods for biofilm control, for example, in cooling towers.

Production of Polypeptides

De novo Chemical Synthesis

[132] Polypeptides of the invention can be synthesized using standard chemical peptide synthesis techniques that are well known to those of skill in the art. In embodiments in which the amino acid sequences are relatively short, the molecule can be synthesized as a single contiguous polypeptide. Where larger molecules are desired, subsequences can be synthesized separately (in one or more units) and then fused by condensation of the amino terminus of one molecule with the carboxyl terminus of the other molecule thereby forming a peptide bond.

[133] Solid phase synthesis in which the C-terminal amino acid of the sequence is attached to an insoluble support followed by sequential addition of the remaining amino acids in the sequence may be used for the chemical synthesis of the polypeptides of this invention.

Techniques for solid phase synthesis are described, for example, by Barany and Merrifield, Solid-Phase Peptide Synthesis; pp. 3-284 in The Peptides: Analysis, Synthesis, Biology. Vol. 2: Special Methods in Peptide Synthesis, Part A., Merrifield, et al. (1963) *J. Am. Chem. Soc.*, 85: 2149-2156, and Stewart et al. (1984) Solid Phase Peptide Synthesis, 2nd ed., Pierce Chem. Co., Rockford, Ill.

Recombinant Expression

[134] In some embodiments, polypeptides of the invention are synthesized using recombinant expression systems. Generally this involves creating a nucleic acid (e.g., DNA) sequence that encodes a polypeptide, placing the DNA in an expression vector under the control of a promoter, and expressing the protein in a host cell.

[135] A nucleic acid encoding a polypeptide of the invention may be cloned, or amplified by *in vitro* methods, such as, for example, the polymerase chain reaction (PCR), the ligase chain reaction (LCR), the transcription-based amplification system (TAS), the self-sustained sequence replication system (SSR). A wide variety of cloning and *in vitro* amplification methodologies are

well-known to persons of skill. Examples of techniques sufficient to direct persons of skill through *in vitro* amplification methods are found in Mullis et al., (1987) U.S. Pat. No.4,683,202; PCR Protocols A Guide to Methods and Applications (Innis et al. eds) Academic Press Inc. San Diego, Calif. (1990) (Innis); Arnheim & Levinson (Oct. 1, 1990) *C&EN* 3647; *The Journal OF NIH Research* (1991) 3: 81-94; (Kwoh et al. (1989) *Proc. Natl. Acad. Sci. USA* 86: 1173; Guatelli et al. (1990) *Proc. Natl. Acad. Sci. USA* 87, 1874; Lomell et al. (1989) *J. Clin. Chem.*, 35: 1826; Landegren et al., (1988) *Science*, 241: 1077-1080; Van Brunt (1990) *Biotechnology*, 8: 291-294; Wu and Wallace, (1989) *Gene*, 4: 560; and Barringer et al. (1990) *Gene*, 89: 117.

[136] In one embodiment, the nucleic acids of this invention can be cloned using DNA amplification methods such as polymerase chain reaction (PCR). Thus, for example, the nucleic acid sequence or subsequence is PCR amplified, using a sense primer containing one restriction site (e.g., NdeI) and an antisense primer containing another restriction site (e.g., HindIII). This will produce a nucleic acid encoding the desired sequence or subsequence and having terminal restriction sites. This nucleic acid can then be easily ligated into a vector containing a nucleic acid having the appropriate corresponding restriction sites. Suitable PCR primers can be determined by one of skill in the art using the sequence information. Appropriate restriction sites can also be added to the nucleic acid encoding the desired protein or protein subsequence by site-directed mutagenesis.

[137] In addition, DNA encoding desired fusion protein sequences may be prepared synthetically using methods that are well known to those of skill in the art, including, for example, direct chemical synthesis by methods such as the phosphotriester method of Narang et al. (1979) *Meth. Enzymol.* 68: 90-99, the phosphodiester method of Brown et al. (1979) *Meth. Enzymol.* 68: 109-151, the diethylphosphoramidite method of Beaucage et al. (1981) *Tetra. Lett.*, 22: 1859-1862, or the solid support method of U.S. Pat. No. 4,458,066.

[138] Chemical synthesis produces a single-stranded oligonucleotide. This may be converted into double stranded DNA by hybridization with a complementary sequence, or by polymerization with a DNA polymerase using the single strand as a template. One of skill would recognize that while chemical synthesis of DNA is generally limited to sequences of about 100 bases, longer sequences can be obtained by the ligation of shorter sequences. Alternatively, subsequences may be cloned, and cleaved using appropriate restriction enzymes. The fragments may then be ligated together to produce the desired nucleic acid sequence.

[139] A nucleic acid of the invention encoding a polypeptide having enzymatic aqueous acylation and/or perhydrolysis activity, as described herein, can be incorporated into a recombinant expression vector in a form suitable for expression in a host cell. As used herein, an "expression vector" is a nucleic acid which, when introduced into an appropriate host cell, can be transcribed and translated into a polypeptide. The terminology "in a form suitable for expression of the fusion protein in a host cell" is intended to mean that the recombinant expression vector includes one or more regulatory sequences operably linked to the nucleic acid encoding the enzyme(s) in a manner that allows for transcription of the nucleic acid into mRNA and translation of the mRNA into the subject protein(s). The term "regulatory sequence" is art-recognized and intended to include promoters, and/or enhancers and/or other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are known to those skilled in the art (see, e.g., Goeddel (1990) Gene Expression Technology, *Meth. Enzymol.* 185, Academic Press, San Diego, Calif.; Berger and Kimmel, Guide to Molecular Cloning Techniques, *Methods in Enzymology* 152 Academic Press, Inc., San Diego, Calif.; Sambrook et al. (1989) Molecular Cloning-A Laboratory Manual (2nd ed.) Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor Press, N.Y., etc.).

[140] The design of an expression vector for use in the invention herein can depend on such factors as the choice of the host cell to be transfected and/or particular polypeptide(s) to be expressed.

[141] It will be appreciated that desired polypeptides can be operably linked to constitutive promoters for high level, continuous expression. Alternatively, inducible and/or tissue-specific promoters can be utilized. When used in mammalian cells, a recombinant expression vector's control functions are often provided by a promoter, often of viral origin. Promoters include, but are not limited to CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Use of appropriate regulatory elements can allow for high level expression of the polypeptide(s) in a variety of host cells.

[142] In some embodiments, a recombinant expression vector for production of a polypeptide of the invention is a plasmid or cosmid. In other embodiments, the expression vector is a virus, or portion thereof, that allows for expression of a nucleic acid introduced into the viral nucleic acid. For example, replication defective retroviruses, adenoviruses and adeno-associated viruses can be used.

[143] Expression vectors may be derived from bacteriophage, including all DNA and RNA phage (e.g., cosmids), or viral vectors derived from all eukaryotic viruses, such as baculoviruses and retroviruses, adenoviruses and adeno-associated viruses, Herpes viruses, Vaccinia viruses and all single-stranded, double-stranded, and partially double-stranded DNA viruses, all positive and negative stranded RNA viruses, and replication defective retroviruses. Another example of an expression vector is a yeast artificial chromosome (YAC), which contains both a centromere and two telomeres, allowing YACs to replicate as small linear chromosomes. A number of suitable expression systems are commercially available and can be modified to produce the vectors of this invention. Illustrative expression systems include, but are not limited to baculovirus expression vectors (see, e.g., O'Reilly et al. (1992) Baculovirus Expression Vectors: A Laboratory Manual, Stockton Press) for expression in insect (e.g. SF9) cells, a wide variety of expression vectors for mammalian cells (see, e.g., pCMV-Script.RTM. Vector, pCMV-Tag1, from Stratagene), vectors for yeast (see, e.g., pYepSec1, Baldari et al. (1987) *EMBO J.* 6: 229-234, pMFa (Ktujan and Herskowitz, (1982) *Cell* 30: 933-943), pJRY88 (Schultz et al., (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and the like), prokaryotic vectors (see, e.g., arabinose-regulated promoter (Invitrogen pBAD Vector), T7 Expression Systems (Novagen, Promega, Stratagene), Trc/Tac Promoter Systems (Clontech, Invitrogen, Kodak, Life Technologies, MBI Fermentas, New England BioLabs, Pharmacia Biotech, Promega), PL Promoters (Invitrogen pLEX and pTrxFus Vectors), Lambda PR Promoter (Pharmacia pRIT2T Vector), Phage T5 Promoter (QIAGEN), tetA Promoter (Biometra pASK75 Vector), and the like.

Host Cells

[144] Polypeptides of this invention can be expressed in a host cell. As used herein, the term "host cell" is intended to include any cell or cell line into which a recombinant expression vector for production of a polypeptide having enzymatic acylation and/or perhydrolysis activity, as described herein, may be transfected for expression of the polypeptide. Host cells include progeny of a single host cell, and the progeny may not necessarily be completely identical (in morphology or in total genomic DNA complement) to the original parent cell due to natural, accidental, or deliberate mutation. A host cell includes cells transfected or transformed *in vivo* with an expression vector.

[145] Suitable host cells include, but are not limited to, to algal cells, bacterial cells (*e.g.*, *E. coli*), yeast cells (*e.g.*, *S. cerevisiae*, *S. pombe*, *P. pastoris*, *K. lactis*, *H. polymorpha*, see, *e.g.*, Fleer (1992) *Curr Opin. Biotech.* 3(5): 486-496), fungal cells, plant cells (*e.g.*, *Arabidopsis*), invertebrate cells (*e.g.*, insect cells such as SF9 cells, and the like), and vertebrate cells including mammalian cells. Non-limiting examples of mammalian cell lines which can be used include CHO cells (Urlaub and Chasin (1980) *Proc. Natl. Acad. Sci. USA* 77: 4216-4220), 293 cells (Graham et al. (1977) *J. Gen. Virol.* 36: 59), or myeloma cells like (*e.g.*, SP2 or NS0, see Galfre and Milstein (1981) *Meth. Enzymol.* 73(B):346).

[146] An expression vector encoding a polypeptide of the invention can be transfected into a host cell using standard techniques. "Transfection" or "transformation" refers to the insertion of an exogenous polynucleotide into a host cell. The exogenous polynucleotide may be maintained as a non-integrated vector, for example, a plasmid, or alternatively, may be integrated into the host cell genome. The term "transfecting" or "transfection" is intended to encompass all conventional techniques for introducing nucleic acid into host cells. Examples of transfection techniques include, but are not limited to, calcium phosphate precipitation, DEAE-dextran-mediated transfection, lipofection, electroporation, and microinjection. Suitable methods for transfecting host cells can be found in Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory press, and other laboratory textbooks. Nucleic acid can also be transferred into cells via a delivery mechanism suitable for introduction of nucleic acid into cells *in vivo*, such as via a retroviral vector (see *e.g.*, Ferry et al. (1991) *Proc. Natl. Acad. Sci., USA*, 88: 8377-8381; and Kay et al. (1992) Human Gene Therapy 3: 641-647), an adenoviral vector (see, *e.g.*, Rosenfeld (1992) *Cell* 68: 143-155; and Herz and Gerard (1993) *Proc. Natl. Acad. Sci., USA*, 90:2812-2816), receptor-mediated DNA uptake (see *e.g.*, Wu, and Wu (1988) *J. Biol. Chem.* 263: 14621; Wilson et al. (1992) *J. Biol. Chem.* 267: 963-967; and U.S. Pat. No. 5,166,320), direct injection of DNA (see, *e.g.*, Acsadi et al. (1991) *Nature* 332: 815-818; and Wolff et al. (1990) *Science* 247:1465-1468) or particle bombardment (biolistics) (see *e.g.*, Cheng et al. (1993) *Proc. Natl. Acad. Sci., USA*, 90:4455-4459; and Zelenin et al. (1993) *FEBS Letts.* 315: 29-32).

[147] Certain vectors integrate into host cells at a low frequency. In order to identify these integrants, in some embodiments a gene that contains a selectable marker (*e.g.*, drug resistance) is introduced into the host cells along with the nucleic acid of interest. Examples of selectable markers include those which confer resistance to certain drugs, such as G418 and hygromycin.

Selectable markers can be introduced on a separate vector from the nucleic acid of interest or on the same vector. Transfected host cells can then be identified by selecting for cells using the selectable marker. For example, if the selectable marker encodes a gene conferring neomycin resistance, host cells which have taken up nucleic acid can be identified by their growth in the presence of G418. Cells that have incorporated the selectable marker gene will survive, while the other cells die.

[148] Once expressed, a polypeptide of the invention can be purified according to standard procedures of the art, including, but not limited to affinity purification, ammonium sulfate precipitation, ion exchange chromatography, or gel electrophoresis (see generally, R. Scopes, (1982) *Protein Purification*, Springer-Verlag, N.Y.; Deutscher (1990) *Methods in Enzymology* Vol. 182: *Guide to Protein Purification*, Academic Press, Inc. N.Y.).

Cleaning and Detergent Formulations

[149] Polypeptides of the invention may be incorporated into cleaning and detergent compositions. The detergent compositions of the present invention are provided in any suitable form, including for example, as a liquid diluent, in granules, in emulsions, in gels, and pastes. When a solid detergent composition is employed, the detergent is preferably formulated as granules. Preferably, the granules are formulated to additionally contain a protecting agent (*See e.g.*, U.S. Appln. Ser. No. 07/642,669 filed January 17, 1991, incorporated herein by reference). Likewise, in some embodiments, the granules are formulated so as to contain materials to reduce the rate of dissolution of the granule into the wash medium (*See e.g.*, U.S. Patent No. 5,254,283, incorporated herein by reference in its entirety). In addition, the enzymes of the present invention having aqueous acylation and/or perhydrolase activity find use in formulations in which substrate and enzyme are present in the same granule. Thus, in some embodiments, the efficacy of the enzyme is increased by the provision of high local concentrations of enzyme and substrate (*See e.g.*, U.S. Patent Application Publication US2003/0191033, herein incorporated by reference).

[150] A number of known compounds are suitable surfactants useful in compositions comprising the polypeptides of the invention. These include nonionic, anionic, cationic, anionic or zwitterionic detergents (*See e.g.*, U.S. Patent Nos 4,404,128 and 4,261,868). A suitable detergent formulation is that described in U.S. Patent No. 5,204,015 (previously incorporated by reference). Some surfactants suitable for use in the present invention are described in British

Patent Application No. 2 094 826 A, incorporated herein by reference. In some embodiments, mixtures of surfactants are used in the present invention.

[151] Suitable anionic surfactants for use in the detergent compositions of the present invention include linear or branched alkylbenzene sulfonates; alkyl or alkenyl ether sulfates having linear or branched alkyl groups or alkenyl groups; alkyl or alkenyl sulfates; olefin sulfonates; alkane sulfonates and the like. Suitable counter ions for anionic surfactants include alkali metal ions such as sodium and potassium; alkaline earth metal ions such as calcium and magnesium; ammonium ion; and alkanolamines having 1 to 3 alkanol groups of carbon number 2 or 3.

[152] Ampholytic surfactants that find use in the present invention include quaternary ammonium salt sulfonates, betaine-type ampholytic surfactants, and the like. Such ampholytic surfactants have both the positive and negative charged groups in the same molecule.

[153] Nonionic surfactants that find use in the present invention generally comprise polyoxyalkylene ethers, as well as higher fatty acid alkanolamides or alkylene oxide adduct thereof, fatty acid glycerine monoesters, and the like.

[154] In some embodiments, the surfactant or surfactant mixture included in the detergent compositions of the present invention is provided in an amount from about 1 weight percent to about 95 weight percent of the total detergent composition and preferably from about 5 weight percent to about 45 weight percent of the total detergent composition. In various embodiments, numerous other components are included in the compositions of the present invention. Many of these are described below. It is not intended that the present invention be limited to these specific examples. Indeed, it is contemplated that additional compounds will find use in the present invention. The descriptions below merely illustrate some optional components.

[155] The polypeptides of the invention can be formulated into known powdered and liquid detergents having pH between 3 and 12.0, at levels of about .001 to about 5% (preferably 0.1% to 0.5%) by weight. In some embodiments, these detergent cleaning compositions further include other enzymes such as proteases, amylases, mannanases, peroxidases, oxido reductases, cellulases, lipases, cutinases, pectinases, pectin lyases, xylanases, and/or endoglycosidases, as well as builders and/or stabilizers.

[156] Polypeptides of the invention can be used, for example, in bar and liquid soap applications, dishcare formulations, surface cleaning applications, contact lens cleaning solutions or products, waste treatment, textile applications, pulp-bleaching, disinfectants, skin care, oral

care, hair care, etc. Indeed, it is not intended that the polypeptides of the present invention be limited to any particular use.

[157] The addition of proteins to conventional cleaning compositions does not create any special use limitations. In other words, any temperature and pH suitable for the detergent are also suitable for the present compositions, as long as the pH is within the range in which the enzyme(s) is/are active, and the temperature is below the described protein's denaturing temperature. In addition, proteins of the invention find use in cleaning, bleaching, and disinfecting compositions without detergents, again either alone or in combination with a source of hydrogen peroxide, an ester substrate (*e.g.*, either added or inherent in the system utilized, such as with stains that contain esters, pulp that contains esters etc), other enzymes, surfactants, builders, stabilizers, etc. Indeed it is not intended that the present invention be limited to any particular formulation or application.

Substrates

[158] In some embodiments, esters comprising aliphatic and/or aromatic carboxylic acids and alcohols are utilized with the polypeptides of the invention in the detergent formulations described herein. In some embodiments, the substrates are selected from one or more of the following: formic acid, acetic acid, propionic acid, butyric acid, valeric acid, caproic acid, caprylic acid, nonanoic acid, decanoic acid, dodecanoic acid, myristic acid, palmitic acid, stearic acid, and oleic acid. Thus, in some embodiments, detergents comprising at least one polypeptide of the invention, at least one hydrogen peroxide source, and at least one ester acid are provided.

Hydrolases

[159] In addition to the polypeptides of the invention described herein, various hydrolases find use in the present invention, including but not limited to carboxylate ester hydrolase, thioester hydrolase, phosphate monoester hydrolase, and phosphate diester hydrolase which act on ester bonds; a thioether hydrolase which acts on ether bonds; and α -amino-acyl-peptide hydrolase, peptidyl-amino acid hydrolase, acyl-amino acid hydrolase, dipeptide hydrolase, and peptidyl-peptide hydrolase which act on peptide bonds, all these enzymes having high perhydrolysis to hydrolysis ratios (*e.g.*, >1). Preferable among them are carboxylate ester hydrolase, and peptidyl-peptide hydrolase. Suitable hydrolases include: (1) proteases belonging to the peptidyl-peptide hydrolase class (*e.g.*, pepsin, pepsin B, rennin, trypsin, chymotrypsin A, chymotrypsin B,

elastase, enterokinase, cathepsin C, papain, chymopapain, ficin, thrombin, fibrinolysin, renin, subtilisin, aspergillopeptidase A, collagenase, clostridiopeptidase B, kallikrein, gastrisin, cathepsin D, bromelin, keratinase, chymotrypsin C, pepsin C, aspergillopeptidase B, urokinase, carboxypeptidase A and B, and aminopeptidase); (2) carboxylate ester hydrolase including carboxyl esterase, lipase, pectin esterase, and chlorophyllase; and (3) enzymes having high perhydrolysis to hydrolysis ratios. Especially effective among them are lipases, as well as esterases that exhibit high perhydrolysis to hydrolysis ratios, as well as protein engineered esterases, cutinases, and lipases, using the primary, secondary, tertiary, and/or quaternary structural features of the perhydrolases of the present invention.

[160] The hydrolase is incorporated into the detergent composition as much as required according to the purpose. It should preferably be incorporated in an amount of 0.0001 to 5 weight percent, and more preferably 0.02 to 3 weight percent. This enzyme should be used in the form of granules made of crude enzyme alone or in combination with other enzymes and/or components in the detergent composition. Granules of crude enzyme are used in such an amount that the purified enzyme is 0.001 to 50 weight percent in the granules. The granules are used in an amount of 0.002 to 20 and preferably 0.1 to 10 weight percent. In some embodiments, the granules are formulated so as to contain an enzyme protecting agent and a dissolution retardant material (*i.e.*, material that regulates the dissolution of granules during use).

Cationic Surfactants and Long-Chain Fatty Acid Salts

[161] Such cationic surfactants and long-chain fatty acid salts include saturated or fatty acid salts, alkyl or alkenyl ether carboxylic acid salts, α -sulfofatty acid salts or esters, amino acid-type surfactants, phosphate ester surfactants, quaternary ammonium salts including those having 3 to 4 alkyl substituents and up to 1 phenyl substituted alkyl substituents. Suitable cationic surfactants and long-chain fatty acid salts include those disclosed in British Patent Application No. 2 094 826 A, the disclosure of which is incorporated herein by reference. The composition may contain from about 1 to about 20 weight percent of such cationic surfactants and long-chain fatty acid salts.

Builders

[162] In some embodiments of the present invention, the composition contains from about 0 to about 50 weight percent of one or more builder components selected from the group consisting

of alkali metal salts and alkanolamine salts of the following compounds: phosphates, phosphonates, phosphonocarboxylates, salts of amino acids, aminopolyacetates high molecular electrolytes, non-dissociating polymers, salts of dicarboxylic acids, and aluminosilicate salts. Examples of suitable divalent sequestering agents are disclosed in British Patent Application No. 2 094 826 A, the disclosure of which is incorporated herein by reference.

[163] In additional embodiments, compositions of the present invention contain from about 1 to about 50 weight percent, preferably from about 5 to about 30 weight percent, based on the composition of one or more alkali metal salts of the following compounds as the alkalis or inorganic electrolytes: silicates, carbonates and sulfates as well as organic alkalis such as triethanolamine, diethanolamine, monoethanolamine and triisopropanolamine.

Anti-Redeposition Agents

[164] In yet additional embodiments of the present invention, the compositions contain from about 0.1 to about 5 weight percent of one or more of the following compounds as antiredeposition agents: polyethylene glycol, polyvinyl alcohol, polyvinylpyrrolidone and carboxymethylcellulose. In some embodiments, a combination of carboxymethyl-cellulose and/or polyethylene glycol are utilized with the composition of the present invention as useful dirt removing compositions.

Bleaching Agents

[165] The use of the perhydrolases of the present invention in combination with additional bleaching agent(s) such as sodium percarbonate, sodium perborate, sodium sulfate/hydrogen peroxide adduct and sodium chloride/hydrogen peroxide adduct and/or a photo-sensitive bleaching dye such as zinc or aluminum salt of sulfonated phthalocyanine further improves the detergent effects. In additional embodiments, the perhydrolases of the present invention are used in combination with bleach boosters (e.g., TAED and/or NOBS).

Bluing Agents and Fluorescent Dyes

[166] In some embodiments of the present invention, bluing agents and fluorescent dyes are incorporated in the composition. Examples of suitable bluing agents and fluorescent dyes are disclosed in British Patent Application No. 2 094 826 A, the disclosure of which is incorporated herein by reference.

Caking Inhibitors

[167] In some embodiments of the present invention in which the composition is powdered or solid, caking inhibitors are incorporated in the composition. Examples of suitable caking inhibitors include p-toluenesulfonic acid salts, xylenesulfonic acid salts, acetic acid salts, sulfosuccinic acid salts, talc, finely pulverized silica, clay, calcium silicate (*e.g.*, Micro-Cell by Johns Manville Co.), calcium carbonate and magnesium oxide.

Antioxidants

[168] Antioxidants which may be incorporated into the compositions of the invention include, for example, tert-butyl-hydroxytoluene, 4,4'-butylidenebis(6-tert-butyl-3-methylphenol), 2,2'-butylidenebis(6-tert-butyl-4-methylphenol), monostyrenated cresol, distyrenated cresol, monostyrenated phenol, distyrenated phenol and 1,1-bis(4-hydroxy-phenyl)cyclohexane.

Solubilizers

[169] In some embodiments, the compositions of the present invention also include solubilizers, including but not limited to lower alcohols (*e.g.*, ethanol, benzenesulfonate salts, and lower alkylbenzenesulfonate salts such as p-toluenesulfonate salts), glycols such as propylene glycol, acetylbenzene-sulfonate salts, acetamides, pyridinedicarboxylic acid amides, benzoate salts and urea.

[170] In some embodiments, the detergent compositions of the present invention are used in a broad pH range of from acidic to alkaline pH. In some embodiments, the detergent composition of the present invention is used in mildly acidic, neutral or alkaline detergent wash media having a pH of from above 4 to no more than about 12.

[171] In addition to the ingredients described above, perfumes, buffers, preservatives, dyes and the like also find use with the present invention. These components are provided in concentrations and forms known to those in the art.

[172] In some embodiments, the powdered detergent bases of the present invention are prepared by any known preparation methods including a spray-drying method and a granulation method. A detergent base may be obtained by the spray-drying method and/or spray-drying granulation method. The detergent base obtained by the spray-drying method is not restricted

with respect to preparation conditions. The detergent base obtained by the spray-drying method is hollow granules which are obtained by spraying an aqueous slurry of heat-resistant ingredients, such as surface active agents and builders, into a hot space. After the spray-drying, perfumes, enzymes, bleaching agents, inorganic alkaline builders may be added. With a highly dense, granular detergent base obtained such as by the spray-drying-granulation method, various ingredients may also be added after the preparation of the base.

[173] When the detergent base is a liquid, it may be either a homogeneous solution or an inhomogeneous dispersion.

[174] The detergent compositions of this invention may be incubated with fabric, for example soiled fabrics, in industrial and household uses at temperatures, reaction times and liquor ratios conventionally employed in these environments. The incubation conditions (*i.e.*, the conditions effective for treating materials with detergent compositions according to the present invention), are readily ascertainable by those of skill in the art. Accordingly, the appropriate conditions effective for treatment with the present detergents correspond to those using similar detergent compositions which include wild-type perhydrolase.

[175] As indicated above, detergents according to the present invention may additionally be formulated as a pre-wash in the appropriate solution at an intermediate pH where sufficient activity exists to provide desired improvements softening, depilling, pilling prevention, surface fiber removal or cleaning. When the detergent composition is a pre-soak (*e.g.*, pre-wash or pre-treatment) composition, either as a liquid, spray, gel or paste composition, the perhydrolase enzyme is generally employed from about 0.00001% to about 5% weight percent based on the total weight of the pre-soak or pre-treatment composition. In such compositions, a surfactant may optionally be employed and when employed, is generally present at a concentration of from about 0.0005 to about 1 weight percent based on the total weight of the pre-soak. The remainder of the composition comprises conventional components used in the pre-soak (*e.g.*, diluent, buffers, other enzymes (proteases), etc.) at their conventional concentrations.

Decontamination

[176] The present invention provides an enzyme system that efficiently generates peracetic acid in aqueous solution for use in decontamination applications. In some embodiments, the present invention provides a system that comprises an ester substrate, a hydrogen peroxide, and at least one polypeptide as described herein with aqueous

acylation and/or perhydrolysis activity. However, it is not intended that the present invention be limited to peracetic acid, as any peracid (*e.g.*, pernonanoic acid, as well as peracids made from long chain fatty acids C10-C18 or longer chains) find use in the present invention. In addition, peracids made from short-chain fatty acids find use in the present invention. Indeed, a variety of peracids find use in the present invention. In some embodiments, the present invention provides an enzyme system with an additional enzyme that forms hydrogen peroxide. In some additional embodiments, the present invention provides enzyme systems that contain additional compounds that generate hydrogen peroxide, including but not limited to such compounds as sodium percarbonate, glucose oxidase, urea, and various others, including but not limited to those described in U.S. Pat. Appln. Ser. No. 10/581,014. In some embodiments, the ester substrate is a stable alcohol ester, although it is not intended that the present invention be limited to any particular ester substrate(s). In some embodiments, the present invention provides a system for enzyme-assisted perhydrolysis in aqueous solutions (*e.g.*, more than about 90% water, although it is not intended that the present invention be limited to any particular percentage of water) comprising at least one ester and at least one peroxide. Indeed, it is contemplated that the present invention will find use in various aqueous systems, including those that have a large percentage of water (*e.g.*, more than about 85%, more than about 95% or more than about 95% water), as well as those with lower percentages of water (*e.g.*, less than about 85%).

[177] In some additional embodiments, the system further comprises at least one surfactant. Thus, in some embodiments, the system comprises at least one polypeptide as described herein, at least one hydrogen peroxide source, and at least one ester substrate in a buffer. In some further embodiments, the system also comprises at least one detergent, while in still further embodiments, the system also comprises at least one surfactant. Thus, various formulations are contemplated to find use in the present invention. In addition, in some embodiments, the present formulations are neutral in pH, but in some embodiments, the enzyme systems also function in alkaline and slightly acidic environments (*e.g.*, pHs from about 6 to about 10).

[178] It is contemplated that the enzyme system for decontamination will find use in various forms, including liquids, granules, foams, emulsions, etc., designed to fit the need at hand. Indeed, it is not intended that the present invention be limited to any particular format. In yet further embodiments, additional enzymes are included, including but not limited to proteases, amylases, cellulases, etc.

[179] The enzyme system finds particular use in decontamination involving a wide variety of chemical and biological warfare materials, as well as for general surface cleaning and decontamination.

[180] In some embodiments, the present invention finds use in decontamination of materials contaminated by various toxic and/or pathogenic entities, including but not limited to toxic chemicals, mustard, VX, *B. anthracis* spores, *Y. pestis*, *F. tularensis*, fungi, and toxins (*e.g.*, botulinum toxin, ricin, mycotoxins, etc.), as well as cells infected with infective virions (*e.g.*, flaviviruses, orthomyxoviruses, paramyxoviruses, arenaviruses, rhabdoviruses, arboviruses, enteroviruses, bunyaviruses, etc.). In some embodiments, the present invention provides a system that is capable of functioning over a wide temperature range (*e.g.*, from about 16°C to about 60°C). In yet additional embodiments, the system provides a small chemical footprint and is stable during short and/or long-term storage. Indeed, it is intended that the system of the present invention will find use in numerous applications.

[181] In still further embodiments, the present invention finds use in decontamination of food and/or feed, including but not limited to vegetables, fruits, and other food and/or feed items. Indeed, it is contemplated that the present invention will find use in the surface cleaning of fruits, vegetables, eggs, meats, etc. Indeed, it is intended that the present invention will find use in the food and/or feed industries to remove contaminants from various food and/or feed items. In some embodiments, methods for food and/or feed decontamination set forth by the Food and Drug Administration and/or other food safety entities, as known to those of skill in the art find use with the present invention.

[182] As indicated herein, the present invention provides enzyme systems for generation of peracid in aqueous solution, suitable for use in decontamination. In some embodiments, the system comprises at least one ester substrate, at least one hydrogen peroxide source, and at least one polypeptide as described herein having aqueous acylation and/or perhydrolysis activity. In some embodiments, the peracid is selected from peracetic acid, pernonanoic acid, perpropionic, perbutanoic, perpentanoic, perhexanoic acid, peracids made from long chain fatty acids, and peracids made from short chain fatty acids. In some alternative embodiments, the system further comprises at least one chemical hydrogen peroxide generation system, wherein the chemical hydrogen peroxide generation system comprises at least one chemical selected from sodium percarbonate, perborate, and urea hydrogen peroxide. In some embodiments, the system further comprises at least one enzymatic hydrogen peroxide generation system selected from oxidases

and their corresponding substrates. In some additional embodiments, the system further comprises at least one enzymatic hydrogen peroxide generation system, wherein the enzymatic hydrogen peroxide generation system comprises at least one enzyme selected from glucose oxidase, sorbitol oxidase, hexose oxidase, choline oxidase, alcohol oxidase, glycerol oxidase, cholesterol oxidase, pyranose oxidase, carboxyalcohol oxidase, L-amino acid oxidase, glycine oxidase, pyruvate oxidase, glutamate oxidase, sarcosine oxidase, lysine oxidase, lactate oxidase, vanillyl oxidase, glycolate oxidase, galactose oxidase, uricase, oxalate oxidase, xanthine oxidase, and wherein said the enzymatic hydrogen peroxide generating system further comprises at least one suitable substrate for the at least one enzyme. In some still additional embodiments, the system further comprises at least one additional enzyme. In some embodiments, the at least one additional enzyme is selected from proteases, cellulases, amylases, and microbial cell wall degrading enzymes. In some further embodiments, the at least one ester substrate is an alcohol ester. In some yet additional embodiments, the system further comprises at least one surfactant. In some embodiments, the system further comprises at least one detergent. In some additional embodiments, the system is in a form selected from liquids, granules, foams, and emulsions.

[183] The present invention also provides methods for decontamination comprising the steps of: providing an item in need of decontamination, and at least one system for generation of peracid in aqueous solution, suitable for use in decontamination; and exposing the item to the system under conditions such that the item is decontaminated. In some embodiments, the exposing comprises exposing the item to the system under alkaline or acid pH conditions. In some alternative embodiments, the exposing comprises exposing the item to the system under neutral pH conditions. In some still further embodiments, the exposing comprises exposing the item at high temperature. In some embodiments, the high temperature is about 60°C or higher. However, it is not intended that the present invention be limited to any particular temperature, as various temperatures find use in the methods of the present invention. In some embodiments, the system is in a form selected from liquids, granules, foams, and emulsions. In some yet further embodiments, the system comprises at least one ester substrate, at least one hydrogen peroxide source, and at least one acyl transferase. In some embodiments, the peracid is selected from peracetic acid, pernonanoic acid, perpropionic, perbutanoic, perpentanoic, perhexanoic acid, peracids made from long chain fatty acids, and peracids made from short chain fatty acids. In some alternative embodiments, the method further comprises at least one chemical hydrogen peroxide generation system selected from sodium percarbonate, perborate, and urea hydrogen

peroxide. In some additional alternative embodiments, the method further comprises at least one enzymatic hydrogen peroxide generation system selected from oxidases and their corresponding substrates. In some embodiments, the system comprises at least one enzymatic hydrogen peroxide generation system selected from glucose oxidase, sorbitol oxidase, hexose oxidase, choline oxidase, alcohol oxidase, glycerol oxidase, cholesterol oxidase, pyranose oxidase, carboxyalcohol oxidase, L-amino acid oxidase, glycine oxidase, pyruvate oxidase, glutamate oxidase, sarcosine oxidase, lysine oxidase, lactate oxidase, vanillyl oxidase, glycolate oxidase, galactose oxidase, uricase, oxalate oxidase, xanthine oxidase, and wherein the enzymatic hydrogen peroxide generating system further comprises at least one suitable substrate for the at least one enzyme. In additional embodiments, the method further comprises at least one enzyme or at least one additional enzyme. In some embodiments, the at least one enzyme is selected from proteases, amylases, cellulases, and microbial cell wall degrading enzymes. In some alternative embodiments, the at least one ester substrate is an alcohol ester. In some additional embodiments, the method further comprises at least one surfactant. In some embodiments, decontamination comprises decontaminating items contaminated by at least one toxin and/or at least one pathogen. In some embodiments, the toxin is selected from botulinum toxin, anthracis toxin, ricin, scombroid toxin, ciguatoxin, tetrodotoxin, and mycotoxins. In further embodiments, the pathogen is selected from bacteria, viruses, fungi, parasites, and prions. In some embodiments, the at least one pathogen is selected from *Bacillus spp.*, *B. anthracis*, *Clostridium spp.*, *C. botulinum*, *C. perfringens*, *Listeria spp.*, *Staphylococcus spp.*, *Streptococcus spp.*, *Salmonella spp.*, *Shigella ssp.*, *E. coli*, *Yersinia spp.*, *Y. pestis*, *Francisella spp.*, *F. tularensis*, *Campylobacter ssp.*, *Vibrio spp.*, *Brucella spp.*, *Cryptosporidium spp.*, *Giardia spp.*, *Cyclospora spp.*, and *Trichinella spp.* In still further embodiments, the item in need of decontamination is selected from hard surfaces, fabrics, food, feed, apparel, rugs, carpets, textiles, medical instruments, and veterinary instruments. In some embodiments, the food is selected from fruits, vegetables, fish, seafood, and meat. In some still further embodiments, the hard surfaces are selected from household surfaces and industrial surfaces. In some embodiments, the household surfaces are selected from kitchen countertops, sinks, cupboards, cutting boards, tables, shelving, food preparation storage areas, bathroom fixtures, floors, ceilings, walls, and bedroom areas. In some alternative embodiments, the industrial surfaces are selected from food processing areas, feed processing areas, tables, shelving, floors, ceilings, walls, sinks, cutting boards, airplanes, automobiles, trains, and boats.

[184] The present invention also provides methods for decontamination comprising the steps of: providing an item in need of decontamination, and at least one system for generation of peracid in aqueous solution, suitable for use in decontamination; generating the peracid in aqueous solution; and exposing the item to the peracid in aqueous solution under conditions such that the item is decontaminated. In some embodiments, the exposing comprises exposing the item to the system under alkaline or acid pH conditions. In some alternative embodiments, the exposing comprises exposing the item to the system under neutral pH conditions. In some still further embodiments, the exposing comprises exposing the item at high temperature. In some embodiments, the high temperature is about 60°C or higher. However, it is not intended that the present invention be limited to any particular temperature, as various temperatures find use in the methods of the present invention. In some embodiments, the system is in a form selected from liquids, granules, foams, and emulsions. In some yet further embodiments, the system comprises at least one ester substrate, at least one hydrogen peroxide source, and at least one acyl transferase. In some embodiments, the peracid is selected from peracetic acid, pernonanoic acid, perpropionic, perbutanoic, perpentanoic, perhexanoic acid, peracids made from long chain fatty acids, and peracids made from short chain fatty acids. In some alternative embodiments, the method further comprises at least one chemical hydrogen peroxide generation system selected from sodium percarbonate, perborate, and urea hydrogen peroxide. In some additional alternative embodiments, the method further comprises at least one enzymatic hydrogen peroxide generation system selected from oxidases and their corresponding substrates. In some embodiments, the system comprises at least one enzymatic hydrogen peroxide generation system selected from glucose oxidase, sorbitol oxidase, hexose oxidase, choline oxidase, alcohol oxidase, glycerol oxidase, cholesterol oxidase, pyranose oxidase, carboxyalcohol oxidase, L-amino acid oxidase, glycine oxidase, pyruvate oxidase, glutamate oxidase, sarcosine oxidase, lysine oxidase, lactate oxidase, vanillyl oxidase, glycolate oxidase, galactose oxidase, uricase, oxalate oxidase, xanthine oxidase, and wherein the enzymatic hydrogen peroxide generating system further comprises at least one suitable substrate for the at least one enzyme. In additional embodiments, the method further comprises at least one enzyme or at least one additional enzyme. In some embodiments, the at least one enzyme is selected from proteases, amylases, cellulases, and microbial cell wall degrading enzymes. In some alternative embodiments, the at least one ester substrate is an alcohol ester. In some additional embodiments, the method further comprises at least one surfactant. In some embodiments, decontamination comprises

decontaminating items contaminated by at least one toxin and/or at least one pathogen. In some embodiments, the toxin is selected from botulinum toxin, anthracis toxin, ricin, scombroid toxin, ciguatoxin, tetrodotoxin, and mycotoxins. In further embodiments, the pathogen is selected from bacteria, viruses, fungi, parasites, and prions. In some embodiments, the at least one pathogen is selected from *Bacillus spp.*, *B. anthracis*, *Clostridium spp.*, *C. botulinum*, *C. perfringens*, *Listeria spp.*, *Staphylococcus spp.*, *Streptococcus spp.*, *Salmonella spp.*, *Shigella spp.*, *E. coli*, *Yersinia spp.*, *Y. pestis*, *Francisella spp.*, *F. tularensis*, *Campylobacter spp.*, *Vibrio spp.*, *Brucella spp.*, *Cryptosporidium spp.*, *Giardia spp.*, *Cyclospora spp.*, and *Trichinella spp.* In still further embodiments, the item in need of decontamination is selected from hard surfaces, fabrics, food, feed, apparel, rugs, carpets, textiles, medical instruments, and veterinary instruments. In some embodiments, the food is selected from fruits, vegetables, fish, seafood, and meat. In some still further embodiments, the hard surfaces are selected from household surfaces and industrial surfaces. In some embodiments, the household surfaces are selected from kitchen countertops, sinks, cupboards, cutting boards, tables, shelving, food preparation storage areas, bathroom fixtures, floors, ceilings, walls, and bedroom areas. In some alternative embodiments, the industrial surfaces are selected from food processing areas, feed processing areas, tables, shelving, floors, ceilings, walls, sinks, cutting boards, airplanes, automobiles, trains, and boats.

Oral Care Compositions

[185] The present invention provides compositions and methods for the use of perhydrolase to whiten teeth. In some embodiments, any suitable peracid finds use in the teeth whitening and/or cleaning methods and/or compositions of the present invention.

[186] The present invention provides oral care compositions comprising at least one polypeptide as described herein having aqueous acylation and/or perhydrolase activity. In some embodiments, the oral compositions are oral care products selected from dentifrices, toothpastes, tooth powders, mouth washes, pre-rinses, teeth whitening products, and denture cleaning agents. In some embodiments, the composition comprises an amount of at least one polypeptide of the invention sufficient to whiten teeth. In additional embodiments, the composition further comprises a hydrogen peroxide generating system. In still further embodiments, the composition further comprises hydrogen peroxide. In additional embodiments, the composition further comprises a peracid generating system. In some additional embodiments, the composition further comprises an acid selected from peracetic acid and acetic acid.

[187] The present invention also provides methods for bleaching teeth comprising the contacting teeth with the oral care composition comprising a polypeptide of the invention, under conditions suitable for bleaching teeth. In some embodiments, the oral compositions are oral care products selected from dentifrices, toothpastes, tooth powders, mouth washes, pre-rinses, teeth whitening products, and denture cleaning agents. In some embodiments, the composition comprises an amount of at least one perhydrolase sufficient to whiten teeth. In additional embodiments, the composition further comprises a hydrogen peroxide generating system. In still further embodiments, the composition further comprises hydrogen peroxide. In yet additional embodiments, the composition further comprises a peracid generating system. In some additional embodiments, the composition further comprises an acid selected from peracetic acid and acetic acid.

Treatment of Textiles

[188] Methods and compositions are provided for the one-step enzymatic treatment of textiles. Textiles that can be treated by the methods and compositions described herein are cellulosic or cellulosic-containing textiles, such as cotton and cotton blends, but the treatment is not limited to cellulose.

[189] In an embodiment, the method comprises the enzymatic bleaching of textiles by contacting a textile in need of bleaching with an enzymatic bleaching composition comprising an ester source, a polypeptide as described herein having enzymatic aqueous acylation and/or perhydrolysis activity, and a hydrogen peroxide source for a length of time and under conditions suitable to permit the measurable whitening of the textile. The ester source may be any suitable acetate ester. The ester source is present in the bleaching liquor at a concentration of between about 100 ppm to 10,000 ppm, between about 1000 ppm to 5000 ppm or between about 2000 ppm to 4000 ppm.

[190] A suitable acetate ester is selected from propylene glycol diacetate, ethylene glycol diacetate, triacetin, ethyl acetate, tributyrin and the like. Combinations of the foregoing acetate esters are also contemplated.

[191] The concentration of the polypeptide of the invention in the bleaching liquor is between about 0.005 ppm to 100 ppm, between about 0.01 to 50 ppm or between 0.05 to 10 ppm, and the perhydrolysis to hydrolysis ratio of the enzyme is greater than 1.

[192] The hydrogen peroxide may be added from an exogenous source. Alternatively, the hydrogen peroxide can be enzymatically generated in situ by a hydrogen peroxide generating oxidase and a suitable substrate. The hydrogen peroxide generating oxidase can be a carbohydrate oxidase such as glucose oxidase. The suitable substrate can be glucose. The concentration of the hydrogen peroxide in the bleaching liquor is between about 100 to 5000 ppm, a concentration of between about 500 to 4000 ppm or a concentration of between about 1000 to 3000 ppm.

[193] The suitable conditions will depend on the enzymes and processing method (e.g., continuous vs batch vs pad-batch) used but is contemplated to comprise varying temperatures, pHs, processing time and the like.

[194] Suitable pH conditions comprise a pH of between about 5 – 11, a pH between about 6 and 10, and a pH between 6 and 8. Suitable time conditions for the enzymatic bleaching of the textile are between about preferably 5 minutes and 24 hours, a time between about 15 minutes and 12 hours, or a time between about 30 minutes and 6 hours.

[195] Suitable temperature conditions comprise a temperature of between about 15°C and 90°C, a temperature of between about 24°C and 80°C or a temperature of between about 40°C and 60°C.

[196] In an embodiment, methods for the treatment of textiles with a one-step treatment composition comprise contacting a textile in need of processing with a one-step treatment composition for a length of time and under conditions sufficient to permit desizing, scouring and bleaching of the textile.

[197] The one-step treatment composition preferably comprises i) one or more bioscouring enzymes, ii) one or more desizing enzymes and iii) one or more enzymatic bleaching system. The one-step treatment composition may further comprise one or more auxiliary components selected from surfactants, emulsifiers, chelating agents and/or stabilizers.

[198] The enzymatic bleaching system, the suitable conditions and length of time for this embodiment are as described for the first embodiment.

[199] The bioscouring enzyme is a pectinase, which includes but is not limited to pectate lyases, pectin esterases, polygalacturonases, etc. as described by J.R. Whitaker (Microbial pectolytic enzymes, (1990) p . 133-176 . *In* W . M . Fogarty and C . T . Kelly (ed.), Microbial enzymes and biotechnology. Elsevier Science Publishers, Barking, United Kingdom) or

combination of pectinase and other enzymes such as cutinases, cellulases, proteases, lipases, and hemicellulases. In one embodiment, the pectinase is a pectate lyase.

[200] The desizing enzyme is selected from a group consisting of amylases and mannanases. A specific amylase that finds use as a desizing enzyme is an alpha-amylase.

[201] The one-step treatment composition may further comprise auxiliary components selected from surfactants, emulsifiers, chelating agents, and/or stabilizers. The surfactant may be a non-ionic surfactant or a combination of non-ionic and anionic surfactants.

[202] A chemical bleaching agent may be used in conjunction with the one-step treatment composition. Suitable chemical bleaching agent(s) may be selected from oxidative bleaches, sodium peroxide, sodium perborate, potassium permanganate, sodium hypochlorite, calcium hypochlorite and sodium dichloroisocyanurate.

[203] In a composition embodiment, the one-step treatment composition comprises i) one or more bioscouring enzymes and ii) an enzymatic bleaching system. In one aspect the composition may include one or more desizing enzymes. The one-step treatment composition may further comprise one or more auxiliary components selected from surfactants, emulsifiers, chelating agents and/or stabilizers.

Production of Fragrant Esters

[204] A polypeptide of the invention having enzymatic aqueous acylation and/or perhydrolysis activity can be employed in a variety of fragrant ester-producing methods that generally involve combining: a) a polypeptide of the invention; b) an alcohol substrate for the acyltransferase; and c) an acyl donor, where, in an aqueous environment, the acyltransferase catalyzes transfer of an acyl group from the acyl donor onto the alcohol substrate to produce the fragrant ester. In particular embodiments, the method may involve rehydrating the components after they are combined. In alternative embodiments, the acyltransferase, the alcohol substrate and the acyl donor are combined in an aqueous environment. As noted above, in particular embodiments, the acyltransferase is an SGNH acyltransferase.

[205] These methods find utility in a variety of processes in which fragrant esters are desirable. For example, the subject composition may be incorporated into foodstuffs to improve or produce flavors or fragrance during consumption, or used in cleaning methods, as described above. The subject composition may be employed in ester manufacturing methods.

[206] In one example, the fragrant ester-producing composition may be incorporated in dried form, e.g., adsorbed onto a substrate, into a foodstuff such as chewing gum or candy. Rehydration of the foodstuff, e.g., during mastication or by the addition of water-containing liquid such as water or milk, initiates the acyltransferase reaction to produce the fragrant ester in situ. Likewise, the methods may be employed to make bulk fragrant esters for the food, perfume and cleaning industries.

[207] In particular embodiments, for example, the alcohol substrate may itself be a fragrant alcohol. As such, in certain cases the odor of the reaction described above may change over time, e.g., from the odor of the fragrant alcohol substrate to the odor of an ester of that alcohol.

[208] Likewise, a fragrant alcohol may be transesterified using a long acyl chain, e.g., a long chain fatty acid to produce a non-fragrant ester. In these embodiments, the non-fragrant ester may be hydrolyzed over time (spontaneously or in the presence of a hydrolase) to re-produce the fragrant alcohol.

EXAMPLES

[209] The following examples are provided in order to demonstrate and further illustrate certain embodiments and aspects of the present invention and are not to be construed as limiting the scope thereof.

[210] In the experimental disclosure which follows, the following abbreviations apply: °C (degrees Centigrade); rpm (revolutions per minute); H₂O (water); HCl (hydrochloric acid); aa (amino acid); bp (base pair); kb (kilobase pair); kD (kilodaltons); gm (grams); µg and ug (micrograms); mg (milligrams); ng (nanograms); µl and ul (microliters); ml (milliliters); mm (millimeters); nm (nanometers); µm and um (micrometer); M (molar); mM (millimolar); µM and uM (micromolar); U (units); V (volts); MW (molecular weight); sec (seconds); min(s) (minute/minutes); hr(s) (hour/hours); MgCl₂ (magnesium chloride); NaCl (sodium chloride); OD₂₈₀ (optical density at 280 nm); OD₆₀₀ (optical density at 600 nm); PAGE (polyacrylamide gel electrophoresis); EtOH (ethanol); PBS (phosphate buffered saline [150 mM NaCl, 10 mM sodium phosphate buffer, pH 7.2]); SDS (sodium dodecyl sulfate); Tris (tris(hydroxymethyl)aminomethane); TAED (N,N,N',N'-tetraacetylenediamine); w/v

(weight to volume); v/v (volume to volume); Per (perhydrolase); *per* (perhydrolase gene); Ms (*M. smegmatis*); MS (mass spectroscopy); BRAIN (BRAIN Biotechnology Research and Information Network, AG, Zwingenberg, Germany); TIGR (The Institute for Genomic Research, Rockville, MD); AATCC (American Association of Textile and Coloring Chemists); WFK (wfk Testgewebe GmbH, Bruggen-Bracht, Germany); Amersham (Amersham Life Science, Inc. Arlington Heights, IL); Millipore (Millipore Corp., Billerica, MA); ICN (ICN Pharmaceuticals, Inc., Costa Mesa, CA); Pierce (Pierce Biotechnology, Rockford, IL); Pharmacia (Pharmacia Corp., Peapack, NJ); EMD Bioscience (EMD Bioscience Inc., San Diego, CA); Boehringer Mannheim (Boehringer Mannheim Corp., Indianapolis, IN); Zebron (Zebron Corp., Newport Beach, CA); Amicon (Amicon, Inc., Beverly, MA); ATCC (American Type Culture Collection, Manassas, VA); Becton Dickinson (Becton Dickinson Labware, Lincoln Park, NJ); BioRad (BioRad, Richmond, CA); Clontech (CLONTECH Laboratories, Palo Alto, CA); Difco (Difco Laboratories, Detroit, MI); GIBCO BRL or Gibco BRL (Life Technologies, Inc., Gaithersburg, MD); Novagen (Novagen, Inc., Madison, WI); Qiagen (Qiagen, Inc., Valencia, CA); Invitrogen (Invitrogen Corp., Carlsbad, CA); Genaissance (Genaissance Pharmaceuticals, Inc., New Haven, CT); DNA 2.0 (DNA 2.0, Menlo Park, CA); MIDI (MIDI Labs, Newark, DE) InvivoGen (InvivoGen, San Diego, CA); Sigma-Aldrich (Sigma-Aldrich, St. Louis, MO); Sorvall (Sorvall Instruments, a subsidiary of DuPont Co., Biotechnology Systems, Wilmington, DE); Stratagene (Stratagene Cloning Systems, La Jolla, CA); Roche (Hoffmann La Roche, Inc., Nutley, NJ); Agilent (Agilent Technologies, Palo Alto, CA); Minolta (Konica Minolta, Ramsey, NJ); and Zeiss (Carl Zeiss, Inc., Thornwood, NY).

[211] In the following Examples, various media were used. "TS" medium (per liter) was prepared using Tryptone (16 g) (Difco), Soytone (4 g) (Difco), Casein hydrolysate (20 g) (Sigma), K₂HPO₄ (10 g), and d H₂O (to 1 L). The medium was sterilized by autoclaving. Then, sterile glucose was added to 1.5% final concentration. Streptomyces Production Medium (per liter) was prepared using citric acid(H₂O) (2.4 g), Biospringer yeast extract (6 g), (NH₄)₂SO₄ (2.4 g), MgSO₄·7 H₂O (2.4 g), Mazu DF204 (5 ml), trace elements (5 ml). The pH was adjusted to 6.9 with NaOH. The medium was then autoclaved to sterilize. After sterilization, CaCl₂·2 H₂O (2 mls of 100 mg/ml solution), KH₂PO₄ (200 ml of a 13% (w/v) solution at pH6.9), and 20 mls of a 50% glucose solution were added to the medium.

[212] In these experiments, a spectrophotometer was used to measure the absorbance of the products formed after the completion of the reactions. A reflectometer was used to measure the

reflectance of the swatches. Unless otherwise indicated, protein concentrations were estimated by Coomassie Plus (Pierce), using BSA as the standard.

Data Collection

[213] Multiwavelength anomalous diffraction data were collected for the apo enzyme at the Advanced Light Source (ALS, Berkeley, USA) on beamline 8.2.1, at wavelengths corresponding to the inflection (λ_1), low energy remote (λ_2), and the peak (λ_3) of a selenium MAD experiment. Later, a data set (λ_0) was collected on beamline 8.2.2 to 1.5 Å resolution. The data sets were collected at 100K using Quantum 210 CCD for the MAD data set and Quantum 315 CCD for the high resolution data set. Data were integrated using Mosflm (Leslie, *Acta Crystallogr.*, D55:1696-1702 [1999]) and scaled with the SCALA program from the CCP4 suite (Collaborative Computational Project, *Acta Crystallogr.*, D50:760-763 [1994]). Data statistics are summarized in Table 2.

[214] Diffraction data for the inhibitor bound form was collected at the Stanford Synchrotron Radiation Laboratory (SSRL, Menlo Park, USA) on beamline 9-1. Crystals diffracted to better than 1.2 Å resolution. However, a complete data set was collected to only 1.5 Å resolution. The data set was collected at 100K using Quantum 315 CCD and processed using HKL2000 program suite (Otwinoski and Minor, *Meth. Enzymol.*, 276:307-326 [1997]). Data statistics are summarized in Table 3.

Structure Solution and Refinement

[215] The initial structure was determined using the 2.5 Å selenium MAD data ($\lambda_{1,2,3}$) using the CCP4 suite and SOLVE/RESOLVE programs (Terwilliger and Berendzen, *Acta Crystallogr.*, D55:849-861 [1999]). Model building was performed using O (Jones *et al.*, *Acta Crystallogr.*, A47:110-119 [1991]). The traced model was then refined with the 1.5 Å dataset (λ_0) using REFMAC (Collaborative Computational Project, *supra*). Refinement statistics are summarized in Table 2. The final model includes a protein octamer, eight phosphate ions, eight glycerol molecules, and 1198 water molecules in the asymmetric unit. No electron density was observed for the first methionine residue in any of the molecules. PROCHECK™ (Laskowski *et al.*, *J. Appl. Crystallogr.*, 26:91-97 [1993]) indicates that 94% of the residues in all of the monomers are located in the core regions of the Ramachandran plot (Ramachandran and Sasisekharan, *Adv.*

Protein Chem., 23:283-437 [1968]), with no residues in the disallowed or generously allowed regions.

[216] The inhibitor structure was solved by molecular replacement with MOLREP (Collaborative Computational Project, *supra*), using the coordinates of the apo enzyme. Refinement statistics are summarized in Table 3. The final model includes a protein octamer, eight inhibitor molecules, one sulfate ion, and 2134 water molecules in the unit cell. No electron density was observed for any of the first methionine residues. PROCHECK™ (Laskowski *et al.*, *supra*) indicates 94% of the residues in the core regions with no residues in the disallowed or generously allowed regions of the Ramachandran plot.

EXAMPLE 1

Cloning of *act* from *M. smegmatis*

[217] An enzyme with acyltransferase activity was purified from *Corynebacterium oxydans* (now *Mycobacterium parafortuitum* ATCC19686; See e.g., WO2005/056782). Two peptide sequences were obtained from the purified protein. The sequence of one peptide, KVPFFDAGSVISTDGVDGI (SEQ ID NO:6), was determined by Edman degradation from cyanogen bromide cleavage of the purified enzyme. The sequence of the second peptide, GTRRILSFGDSL TWGWIPV (SEQ ID NO:7), was determined using N-terminal sequencing. A BLAST search against the TIGR unfinished genome database identified a sequence of potential interest in *Mycobacterium smegmatis*. This gene was amplified from *M. smegmatis* by PCR using primers MsRBSF: 5'-

CTAACAGGAGGAATTAACCATGGCCAAGCGAATTCTGTGTTTCGGTGATTCCCTGACCT-3' (SEQ ID NO:8) and MspetBamR: 5'-

GCGCGCGGATCCGCGCGCTTACAGCAGGCTCCGCACCTGTTCCGCGAGGGCCACCCCGA-3' (SEQ ID NO:9), which create an *NcoI* site at the ATG start codon and add a *BamHI* site after the stop codon.

[218] The amplification of the gene was accomplished by PCR using *Taq* DNA polymerase (Roche) as per the manufacturer's instructions, with approximately 500 ng of chromosomal DNA from *Mycobacterium smegmatis* as the template DNA and the addition of 1% DMSO to the PCR reaction mix. Ten picomoles of each of the primers MsRBSF and MspetBamR were

added to the mix. The amplification cycle was: 30 cycles of (95°C for 1 min, 55°C for 1 min, 72°C for 1 min).

[219] The fragments obtained from the PCR reaction were separated on a 1.2% agarose gel and a single band of the expected size of 651 bp was identified. This band was cloned directly into the pCR2.1 TOPO cloning vector (Invitrogen) and transformed into *E. coli* Top 10 cells (Invitrogen) with selection on L agar containing 100 micrograms/ml carbenicillin and X-gal (20 micrograms/ml, Sigma-Aldrich) for blue/white selection and incubated overnight at 37°C. Plasmid DNA was purified from a culture of one of the transformants using the Quikspin kit (Qiagen). The presence of the correct fragment was determined by restriction enzyme digest with *EcoRI* to release the fragment, and sequencing using primers supplied by the pCR2.1 manufacturer (Invitrogen). The plasmid was designated pMSATNcoI. The plasmid pMSATNcoI was digested with *NcoI/BamHI* (Roche) and the fragment was gel purified using the Qiagen gel purification kit. The fragment was ligated into the expression plasmid, pET16b (Novagen), also digested with *NcoI/BamHI* T4 DNA ligase overnight at 16°C. The ligation reaction was transformed into chemically competent *E. coli* Top 10 cells (Invitrogen) and selected on L agar containing 100 µg/ml carbenicillin overnight at 37°C. Plasmid DNA was prepared from cultures of a transformant using the Qiagen Quikspin kit (Qiagen) and the presence of the correct fragment was determined by restriction enzyme digest with *NcoI/BamHI*. The correct plasmid was designated pMSATNcoI-1. This plasmid was transformed into the *E. coli* strain BL21(λDE3)pLysS (Novagen), with selection on LA containing 100 micrograms/ml carbenicillin. Cells were grown overnight at 37°C, one transformant was selected and designated MSATNcoI-1.

EXAMPLE 2

Expression of MsAcT

[220] Production of MsAcT for enzymatic analysis was performed by inoculating 5 ml of LB with carbenicillin (100 µg/ml) with a single colony of MSATNcoI-1 and grown overnight at 37°C with shaking at 200 rpm. This culture was used to inoculate 100 ml of LB with carbenicillin (100 µg/ml) to an OD₆₀₀ of 0.1. The cultures were grown at 30°C with shaking at 200 rpm until they reached an OD₆₀₀ of 0.4. The expression of the *act* gene was then induced by the addition of 100 µM IPTG and the incubation continued overnight. Cultures were harvested

by centrifugation (10 min at 7000 rpm, Sorvall SS34 rotor), the supernatant was removed and the pellets washed in 50 mM KPO₄, pH 6.8. The cells were centrifuged again, the supernatants removed and the wet weight of the cells was determined. The cells were resuspended in 100 mM KPO₄ in a volume that was 4x the wet weight. The resuspended cells were frozen at -70°C. The cells were thawed and lysed in a French pressure cell. The MsAcT was purified as described below in Example 3, below.

EXAMPLE 3

Seleno-Methionine labeling of MsAcT

[221] A 500 ml preculture of MsAcTNcoI-1 was grown in a baffled 2.8 L Fernbach flask in LB containing 100 µg/ml carbenicillin. After overnight incubation at 37°C, with shaking at 200 rpm, the cells were harvested by centrifugation and resuspended in M9 medium containing: glucose, 2 g/L; Na₂HPO₄, 6 g/L; KH₂PO₄, 3 g/L; NH₄Cl, 1 g/L; NaCl, 0.5 g/L; thiamine, 5 mg/L; MgSO₄, 2 mM; CaCl₂, 100 µM; Citric acid•H₂O, 40 mg/L; MnSO₄•H₂O, 30 mg/L; NaCl, 10 mg/L; FeSO₄•7H₂O, 1 mg/L; CoCl₂•6H₂O, 1 mg/L; ZnSO₄•7H₂O, 1 mg/L; CuSO₄•5H₂O, 100 µg/L; H₃BO₃•5H₂O, 100 µg/L; and NaMoO₄•2H₂O, 100 µg/L; and supplemented with carbenicillin, 100 µg/L.

[222] The resuspended cells were used to inoculate six Fernbach flasks containing 500 ml each of M9 medium supplemented with carbenicillin (100 mg/L). The cultures were incubated at 30°C with shaking at 200 rpm until the OD₆₀₀ reached about 0.7 at which time 100 mg/L of lysine, threonine, and phenylalanine and 50 mg/L of leucine, isoleucine, valine, and selenomethionine (EMD Biosciences) were added. After further incubation for 30 min, IPTG was added to a final concentration of 50 µM. The cultures were then incubated overnight (~15hr) at 30°C with shaking at 200 rpm and harvested by centrifugation. The cell pellet was washed twice with 50 mM KPO₄ buffer, pH 6.8. The yield was 28.5 gm wet weight of cells to which was added 114 ml of 100 mM KPO₄ buffer, pH 8.2 and 5 mg of DNase. This mixture was frozen at -80°C and thawed twice. The thawed cell suspension was lysed by disruption in a French pressure cell at 20K psi. The unbroken cells and cell membrane material were sedimented by centrifugation at 100K xg for 1 hour. Then, 128 ml of the supernatant crude extract (CE), was placed in a 600 ml beaker and stirred for 10 minutes in a 55°C water bath to precipitate unstable proteins. After 10 min., the beaker was stirred in ice water for 1 min.

followed by centrifugation at 15K xg for 15 min. The supernatant from this procedure, ("HT"), contained 118 ml. The HT extract was then made 20% saturating in $(\text{NH}_4)_2\text{SO}_4$ and loaded on to a 10 cm X 11.6 cm Fast Flow Phenyl Sepharose (Pharmacia) column equilibrated in 100 mM KPO_4 buffer, pH 6.8, containing 20% saturation (109 g/L) $(\text{NH}_4)_2\text{SO}_4$. After loading the extract, the column was washed with 1700 ml of starting buffer and eluted with a two-step gradient. The first step was a linear 1900 ml gradient from start buffer to the same buffer without $(\text{NH}_4)_2\text{SO}_4$, the second was a 500 ml elution with 100 mM KPO_4 , pH 6.8 containing 5% EtOH. Active fractions, 241 ml, were pooled, diluted 100 % with water and loaded onto a 1.6 mm X 16 mm Poros HQ strong anion exchange column (Boehringer Mannheim) equilibrated in 100 mM Tris-HCl, pH 7.6. After loading the extract, the column was washed with 5 column volumes of starting buffer. The protein was eluted with a 15 column volume gradient from start buffer to start buffer containing 175 mM KCl. The active fractions were pooled and concentrated using a Centriprep 30 (Millipore) to 740 μl .

EXAMPLE 4

Determination of Perhydrolysis to Hydrolysis Ratio

[223] In this Example, methods used to determine the transesterification ability, as well as the perhydrolysis to hydrolysis ratio of the enzymes herein are described.

Transesterification

[224] Reactions contained 100 mM neopentyl glycol (NPG) in ethyl acetate with the indicated amount of dissolved water. The enzyme was added to a final concentration of 40 ng/ml and incubated with shaking at 21°C. Samples were withdrawn at intervals for up to 1 hour and analyzed by gas chromatography on a nitroterephthalic acid modified polyethylene glycol column (Zebron FFAP; with dimensions: 30 m long, 250 μm diameter, 250 nm film thickness). The results were reported as area of the NPG monoester/hr.

Determination of Hydrolysis

[225] The rate of tributyrin hydrolysis was measured in reactions comprised of 50 mM potassium phosphate pH 7.5, 10 mM tributyrin, 29 mM hydrogen peroxide, and 20 mM potassium chloride in a total volume of 0.99 ml and an amount of enzyme that would generate 20 nmoles of butyric acid per minute at 25°C. Hydrolytic activity was measured by monitoring

the increase of butyric acid generated by the enzyme from tributyrin, using gas chromatography coupled with flame ionization detection. Aliquots were taken at intervals over an hour and quenched with 4 volumes of methanol. The methanol-quenched samples were then analyzed by GC using a nitroterephthalic acid modified polyethylene glycol column (Zebron FFAP; with dimensions: 30 m long, 250 μ m diameter, 250 nm film thickness). A 3 μ L aliquot of sample was applied to the column by a splitless injection under constant a helium flow of 1.0 mL/minute. The inlet was maintained at a temperature of 250°C and was purged of any remaining sample components after 2 minutes. The temperature of the column was maintained at 75°C for 1 minute after injection, increased at a rate of 25°C/minute to 100°C, then increased 15°C/minute to 225°C.

Determination of Perhydrolysis

[226] The perhydrolytic activity assay comprised 50 mM potassium phosphate pH 7.5, 10 mM tributyrin, 29 mM hydrogen peroxide, 20 mM potassium chloride, and 10 mM O-phenylenediamine (OPD). Activity was measured by monitoring the absorbance increase at 458 nm of oxidized OPD by peracid generated with the enzyme. The perhydrolytic activity assay solution was prepared in the same manner as the hydrolytic activity assay solution, except that OPD was added to the assay solution to a final concentration of 10 mM. The OPD solution was prepared immediately before conducting the assay by dissolving 72 mg OPD (Sigma-Aldrich, dihydrochloride) in 19.94 mL of the same buffer and the pH was adjusted by slowly adding 60 μ L of 13.5 M potassium hydroxide. The pH was measured and if needed, small quantities of potassium hydroxide were added to return the pH to the original pH of the buffer. Then, 495 μ L of this OPD solution were added with the other assay components to a final assay volume of 0.990 mL. A quenching solution was prepared by dissolving 36mg OPD in 20 mL 100 mM citric acid and 70% ethanol. The assay was conducted at 25°C and was initiated by the addition of enzyme. Aliquots were taken at intervals over an hour and quenched with 2 volumes of quenching solution at various times, typically 2, 5, 10, 15, 25, 40, and 60 minutes, after adding the enzyme. The quenched assay samples were incubated for 30 minutes to allow any remaining peracid to oxidize the OPD and the absorbance was measured. The concentration of peracid was determined by comparison to a standard curve generated under the above conditions.

[227] Perhydrolysis /Hydrolysis ratio:

perhydrolysis/hydrolysis ratio = perhydrolysis measured in the perhydrolysis assay/(total acid detected in the hydrolysis assay-perhydrolysis measured in the perhydrolysis assay)

[228] In preliminary experiments, the lipases showed good activity on tributyrin, so both the HPLC and OPD/GC assays were conducted using tributyrin with the lipases. As there was no perbutyric acid production by HPLC by the *M. smegmatis* perhydrolase homologues, the HPLC and OPD/GC assays were conducted using triacetin.

EXAMPLE 5

Structural Features of *M. Smegmatis* Aryl Esterase Enzyme

[229] *Mycobacterium smegmatis* produces an enzyme of unknown physiological function that catalyzes acyl transfer reactions in water. In contrast to hydrolases that perform alcoholysis under anhydrous conditions, this acyl transferase (MsACT) demonstrated alcoholysis even in substantially aqueous media. Furthermore in the presence of hydrogen peroxide, MsAcT carried out perhydrolysis to form organic peracids with a perhydrolysis:hydrolysis ratio >50-fold to that of best lipase tested. The crystal structure of MsAcT has been determined using MAD phasing techniques to 1.5 Å resolution. MsAcT was determined to be an octamer in asymmetric units that forms a tightly associated aggregate in solution. MsAcT is a member of a subgroup of the SGNH-hydrolase family, which can be differentiated based on their similar catalytic properties and structural features favoring formation of aggregate that in MsAct, greatly restrict the accessibility and shape of the active site. The structure provides important information in engineering improved enzymes having acyltransferase activity. The crystal structures of the apoenzyme and an inhibitor bound form showed have been determined to 1.5 Å resolution. MsAcT was found to be an octamer in the asymmetric unit and formed a tightly associated aggregate in solution. Relative to other structurally similar monomers, AcT contains several “insertions” that contribute to the oligomerization and greatly restrict the shape of the active site limiting thereby its accessibility. Thus, it is contemplated that the present invention will find use in providing means to convert serine hydrolases to acyltransferases.

[230] While investigating the selective oxidation of various alcohols by lyophilized whole cells of *Mycobacterium parafortuitum*, a facile transesterification reaction was observed when ethyl acetate was used as solvent in the presence of prochiral diols resulting in the stereospecific

acylation of the diols. This reaction was assumed to be the product of a lipase and the protein responsible for the activity was purified. During the characterization of the enzyme, it was found that instead of inhibiting the progress of the acylation reaction, water promoted the reaction and the activity increased as the percentage of water was increased up to saturation in ethyl acetate (Figure 1A). Further it was found that the acylation reaction was still fully active ($K_{cat}=10^5$) in aqueous reactions containing just millimolar concentrations of ethyl acetate.

[231] *Mycobacterium smegmatis* demonstrated the same type of activity as *M. parafortuitum*. The sequence of two peptide fragments obtained from cyanogen bromide cleavage of the acyl transferase from *M. parafortuitum* eventually led to the identification of the complete gene sequence from the unfinished genome sequence database of *Mycobacterium smegmatis*, MC₂155. The protein encoded by this gene was found to have essentially identical physical and catalytic properties to the protein purified from *M. parafortuitum*. The protein is referred to herein as "MsACT," "AcT," "Perhydrolase" or "Per," and the encoding gene is referred to herein as "*Msact*," "*act*," or "*per*."

[232] The purified protein was found to be unaffected by water in a single phase reaction mixture. Instead of inhibiting transesterification, it was found that water promoted the reaction. The enzymatic activity increased with increasing water concentration, dissolved in ethyl acetate as a single phase. Lyophilized whole cells of *Mycobacterium smegmatis* also demonstrated the same type of activity. The acyltransfer reaction described above is a replacement of the usual hydrolytic deacylation with an alcoholytic one. A similar reaction involving perhydrolysis results in the formation of aliphatic peracids and provides an effective source of *in-situ* generated bleaching agents. Since lipases that catalyze the formation of peracids for *in-situ* bleaching have been reported previously (See e.g., United States Patent No. 5,030,240). MsAcT was tested for its ability to catalyze perhydrolysis. As shown in Figure 1B, the ratio of perbutyric acid to butyric acid generated by MsAcT from tributyrin and hydrogen peroxide, was far greater than that observed for other lipases catalyzing perhydrolysis in an aqueous reaction milieu. Thus, MsAcT was identified as a source for *in-situ* generation of peracids for commercial applications.

[233] MsAcT represents an enzyme with two very desirable characteristics, including being able to perform both alcoholysis and perhydrolysis in aqueous media. Several homologous sequences were identified and the proteins tested for these properties (See, Table 1). The additional enzymes were "related" by either sequence or structural homology. The sequence

identity was determined using Vector NTi software (Invitrogen) and full-length protein sequences.

[234] Two proteins, RSM02162 from *Sinorhizobium meliloti* (putative arylesterase) and 7-aminocephalosporanic acid arylesterase (7-ACA) from *Agrobacterium tumefaciens* (*radiobacter*) (Sakai *et al.*, J. Ferment. Bioengineer., 85: 138-143 [1998]) were found to share the same properties. All three enzymes were found to exist as oligomers in solution. While MsAcT is an octamer, *A. tumefaciens* 7-ACA is a tetramer, and *S. meliloti* RSM02162, is an apparent octamer, as determined by gel filtration. Of the enzymes tested, those having at least 40% sequence identity were shown to manifest comparable acyltransferase and perhydrolytic activity. All of these enzymes appear as oligomers in solution. The structural experiments provided herein revealed a SGNH hydrolase fold for the MsAcT enzyme. However two of the enzymes from this family, namely, *E. coli* thioesterase and rhamnogalacturonan acetyesterase, did not show any activity in the acyltransferase or perhydrolysis assays, as indicated in Table 1, below.

Table 1. Characterization of Acyltransferase and Perhydrolyase Activity of MsAcT and Related Enzymes					
Enzyme	Acyltransferase Activity	Perhydrolyase Activity	Soluble Aggregate Form	Sequence Identity (%)	Structure
MsAcT	Yes	Yes	Octamer	100	SGNH Hydrolase
<i>S. meliloti</i> RSM02162	Yes	Yes	Apparent Octamer	63.3	Unknown
<i>A. tumefaciens</i> 7-ACA	Some	Some	Tetramer	42.5	Unknown
<i>E. coli</i> thioesterase	None	None	Monomer	14	SGNH Hydrolase
<i>A. aculeatus</i> rhamno-galacturonan acetyesterase	None	None	Monomer	13.8	SGNH Hydrolase

[235] The sequence alignment of these enzymes (SEQ ID NOS: 1, 2, 3, 4, and 5, respectively) is provided in Figure 9.

[236] MsAcT was crystallized to determine if structural features of the protein contributed to its ability to catalyze the unusual reactions. The structures of the apoenzyme and the enzyme with an inhibitor bound were both determined. Crystals were obtained in the tetragonal space group P4 with eight molecules in the asymmetric unit. The three dimensional structure of the *M. smegmatis* enzyme was determined to 1.5 Å resolution by MAD techniques using selenomethionine (SeMet) labeled protein (32 Se in the asymmetric unit). The crystal structure shows that the enzyme is an octamer with eight identical subunits (216 residues per subunit). Thus, the octamer is a tetramer of closely associated dimers that form a block-like structure of roughly 72 Å x 72 Å x 60 Å dimensions with a large channel in the center running from the “top” to the “bottom” and crevices on the “sides” between pairs of dimers. Each monomer has a five stranded parallel β-sheet structure sandwiched by α-helices on either side. As indicated above, the catalytic triad is composed of Ser11, Asp192, and His195.

[237] The refined MsAcT octamer contains 1720 residues (residues 2-216 for all monomers), eight sulfate ions, eight glycerol molecules, and 1608 water molecules (See, Table 2). The crystallographic R factor is 17.5% and R_{free} is 19.6% (using all data without any sigma cut off). The percentages of non-glycine residues in the most favored and allowed Ramachandran areas are 94.2% and 5.8%, respectively, as assessed by PROCHECK™ software (Laskowski *et al.*, J. Appl. Crystallogr., 26:91-97 [1993]).

Table 2. Summary of Crystal parameters, Data Collection, and Refinement**Statistics for****the Apo Enzyme**

the Apo Enzyme

Space group	P4			
Unit cell parameters	a = 98.163 Å, b = 98.163 Å, c = 229.896 Å, α=β=γ=90°			
Data Collection	λ ₀ MADSe	λ ₁ MADSe	λ ₂ MADSe	λ ₃ MADSe
Wavelength (Å)	1.0871	0.9796	1.000	0.9795
Resolution range (Å)	29.39 – 1.50	76.7-2.50	76.7-2.50	76.7-2.50
Number of observations	1861,019	270,764	270,480	275,841
Number of reflections	327148	74,120	74,233	74,320
Completeness (%)	94.7(62.8) ⁺	99.4 (97.7)	99.5(98.2) ⁺	99.6(99.0) ⁺
Mean I/σ(I)	14.9 (2.6) ⁺	10.5(8.5) ⁺	10.6(8.3) ⁺	10.4(8.8) ⁺
R _{sym} on I	0.087 (0.208) ⁺	0.081(0.111) ⁺	0.080(0.106) ⁺	0.083(0.110) ⁺
Sigma Cutoff	0.0	0.0	0.0	0.0
Highest resolution shell (Å)	1.54- 1.50	2.57 - 2.50	2.57 - 2.50	2.57 - 2.50
Model and Refinement Statistics				
Resolution range (Å)	30.00 – 1.50		Data set used in refinement	λ ₀ MADSe
No. of reflections (total)	327138		Cutoff criteria	F > 0
No. of reflections (test)	16328		R _{cryst}	0.175
Completeness (% total)	94.7		R _{free}	0.196
Deviation from ideal geometry (rms):				
Bond length		0.015 Å		
Bond angle		1.525 °		
Average B-value protein		13.1 Å ²		
Average B-value SO ₄		16.0 Å ²		
Average B-value glycerol		23.4 Å ²		
Average B-value water		23.9 Å ²		
ESU based on R value		0.07 Å		
Protein residues / atoms		1720/13,072		
Sulfate / atoms		8 / 40		
Glycerol / atoms		8 / 48		
Solvent molecules		1608		

+ highest resolution shell

ESU = Estimated overall coordinate error.

R_{sym} = $\sum |I_i - \langle I_i \rangle| / \sum I_i$ where I_i is the scaled intensity of the i^{th} measurement, and $\langle I_i \rangle$ is the mean intensity for that reflection.R_{cryst} = $\sum ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum |F_{\text{obs}}|$ where F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively.R_{free} = as for R_{cryst}, but for 5.0% of the total reflections chosen at random and omitted from refinement.

Table 3. Summary of Crystal Parameters, Data Collection, and Refinement Statistics for the Inhibitor Bound Form

Summary for the Enrichment Dataset

Space group	P1		
Unit cell parameters	a = 67.754 Å, b = 80.096 Å, c = 85.974 Å, α = 104°, β = 112°, γ = 97°		
Data Collection	λ ₀ MAD Se		
Wavelength (Å)	0.9795		
Resolution range (Å)	50.00 – 1.25		
Number of observations	1623,435		
Number of reflections	302,271		
Completeness (%)	69.4 (14.5) ⁺		
Mean I/σ(I)	30.8 (8.0) ⁺		
R _{sym} on I	0.042 (0.13) ⁺		
Sigma Cutoff	0.0		
Highest resolution shell (Å)	1.29- 1.25		
Model and Refinement Statistics			
Resolution range (Å)	50.00 – 1.50	Data set used in refinement	λ ₀ MADSe
No. of reflections (total)	240,526	Cutoff criteria	F > 0
No. of reflections (test)	11,930	R _{cryst}	0.134
Completeness (% total)	95.4	R _{free}	0.160
Deviation from ideal geometry (rms):			
Bond length		0.016 Å	
Bond angle		1.627°	
Average B-value protein		11.5 Å ²	
Average B-value inhibitor		12.2 Å ²	
Average B-value SO ₄		36.2 Å ²	
Average B-value water		23.1 Å ²	
ESU based on R value		0.06 Å	
Protein residues / atoms		1720/13,07	
Inhibitor / atoms		8 / 72	
Sulfate / atoms		1 / 5	
Solvent molecules		2134	

+ highest resolution shell

ESU = Estimated overall coordinate error.

 $R_{\text{sym}} = \sum |I_i - \langle I_i \rangle| / \sum |I_i|$ where I_i is the scaled intensity of the i^{th} measurement, and $\langle I_i \rangle$ is the mean intensity for that reflection. $R_{\text{cryst}} = \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum |F_{\text{obs}}|$ where F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively. R_{free} = as for R_{cryst} , but for 5.0% of the total reflections chosen at random and omitted from refinement.

[238] A structural-homology search was performed using the program DALI (Holm and Sander, Trends Biochem. Sci., 478-480 [1995]), which is based on a distance criterion and does not use sequence information for the comparison, showed five closely related proteins. These proteins are thioesterase I (PDB code: 1ivn); platelet-activating factor acetyl hydrolase (PDB code: 1wab); a hypothetical protein (PDB code: 1vjg); esterase (PDB code: 1esc); and rhamnogalacturonan acetyl esterase (PDB code: 1deo). All these proteins, along with MsAcT, were found to share a common structural motif, having the five-stranded parallel β -sheet structure sandwiched by α -helices on either side, characteristic of the SGNH-hydrolase fold family (See, Figure 4). Interestingly, the active serine appears on a short helical segment following the first beta strand and the aspartic acid and histidine, forming the catalytic triad, follow a helical segment extending a short beta strand (*i.e.*, B5 in Figure 4). In many cases, the helical segment is part of an elbow bend. The different SGNH-hydrolases can be differentiated by the pattern of insertions and deletion from the basic fold that is best exemplified by the *E. coli* thioesterase. Indeed, using thioesterase as the representative of the SGNH-hydrolase family, structural comparisons were made between MsAcT and thioesterase. As shown in Figure 5, MsAcT is easily superimposed with thioesterase.

[239] As found in other SGNH-hydrolase structures, the nucleophile Ser11 in the catalytic triad of MsAcT is located in the GDS sequence motif on the short helical segment H1 (See, Figure 4), which represents the SGNH block I sequence motif (Dalrymple *et al.*, Microbiol., 143:2605-2516 [1997]). The sulfate group at the active site is well defined by the electron density, as well as the hydrogen bonds to Ser11 and His195. The sulfate oxygen involved in hydrogen bonding interactions with Ser11 also participates in hydrogen bonding with the amide nitrogen of Ala55 and the side chain ND2 of Asn94. Asn94 is the conserved residue present in the block III sequence motif (GXND) of SGNH hydrolase. Unlike the conserved asparagine, Asn 94, MsAcT deviates from the SGNH-hydrolase by having alanine rather than glycine at position 55 which is in the block II region. Both alanine and glycine residues function equally well in contributing the amide N to form the oxyanion hole. Although MsAcT is referred to herein as having the SGNH hydrolase fold, MsAcT is actually a S(G/A)NH hydrolase. The sulfate oxygen in MsAcT occupies a structurally similar location as that found for the sulfate ion of RGAE (See, Molgaard *et al.*, Structure 8:373-383 [2000]). Although the overall topology of MsAcT is identical to that of SGNH-hydrolases, there are several insertions and one deletion in the MsAcT relative to the general SGNH hydrolase-fold (See, Figure 5) as represented by thioesterase. Prominent among

them, are the four insertions highlighted in Figure 4 which is formed by residues 17-27, insertion 1); residues 59-69, insertion 2; residues 122-130, insertion 3 and residues 142-156, insertion 4. MsAcT also has a deletion relative to the thioesterase and the other known SGNH-hydrolases, eliminating a helical elbow preceding the catalytic Asp-X-X-His sequence which completes the catalytic triad with Ser 11.

[240] As shown in Table 1, MsAcT and other homologous enzymes that share the capability to catalyze acyltransfer and perhydrolysis in aqueous media exist as oligomers, while two other known SGNH-hydrolases that were tested (*i.e.*, thioesterase and RGAE), do not share these properties are found as monomers. It is therefore contemplated that the oligimerization state is a significant structural difference between MsAcT and other SGNH-hydrolases. While the present invention is not limited to any particular mechanism, it is contemplated that the oligomeric nature of MsAcT restricts access to and redefines the overall topology of the active site pocket. The nature of this pocket was investigated using an inhibitor, 4'-nitrophenyl-n-hexylcarbamate. The inhibited complex was crystallized in a triclinic space group P1 having an octamer in the asymmetric unit (*See*, Table 3). Diffraction data was collected to 1.25 Å and the current model was refined using all data, to 1.50 Å resolution. This model was found to contain 1720 residues (residues 2-216 for all monomers), eight covalently bound inhibitor molecules, one sulfate ion, and 2134 water molecules (*See*, Table 3). The crystallographic R factor and R_{free} are 13.4% and 16.0%, respectively (using all data between 50.0 – 1.50 Å resolution). The percentages of non-glycine residues in the most favored and allowed Ramachandran areas are 93.8% and 6.2%, respectively, as assessed by PROCHECK™.

[241] Density was observed for the inhibitor, which is covalently bound to the active site residue Ser 11 in all monomers. The inhibitor is bound in a hydrophobic channel, which extends to the exterior of the octamer surface and the position of the alkyl chain indicates the probable direction of substrate approach into the active site. The interior of the channel is formed by the four large loops that arise from the insertions 1-4. Three of these insertions, namely 1, 2 and 4, come from the same monomer, while the loop corresponding to insertion 3 comes from the dimer mate (*See*, Figure 7, Panel A). The substrate binding cavity is completed with inclusion of the dimer mate. Insertion 3 of the dimer mate can be seen to complete the elaborated cavity surrounding the catalytic triad formed by insertions 1, 2 and 4 (*See*, Figure 7, Panel B). Thus, the formation of tight dimer pairs results in the creation of the interior channel that leads to the catalytic triad.

[242] A similar pattern of insertion loop inter-digitation was observed for the loop of insertion 4 with the neighbor dimer pairs (*See*, Figure 8). Here the side chain of Trp 149 of insertion 4 occupies the hydrophobic pocket defined by residues Leu 105, Leu 109 and Phe 174 on the surface of a neighboring dimer pair. The two-fold symmetry of the dimer creates a pattern of linkages where the insertion 4 loops link one dimer pair to both adjoining dimer pairs in the octamer. Within the octamer, the aggregation of dimers into the octamer further elaborates the substrate binding surface and restricts access to the catalytic triad.

[243] Other than the platelet-activating factor, an acetyl hydrolase, which forms a dimer in the crystal structure, all other SGNH-hydrolases have monomeric structures and most of them do not show any acyltransferase activity in water. Several SGNH-hydrolases were tested for the ability to carry out alcoholysis in aqueous conditions and also for their ability to perform perhydrolysis. Representative sequences of SGNH-hydrolases that manifest the ability to catalyze acyltransfer reactions in water have more highly conserved sequences relative to other SGNH-hydrolases that do not catalyze such reactions (*See*, Figure 9). In Figure 9, five sequences are compared, three that show the highest rates of acyltransferase activity in water (MsAcT and two closely homologous enzymes (*S. meliloti* _RSM0g2162 and *A. tumifaciens* 7-ACA); and two SGNH-hydrolases with known structure, that do not catalyze this reaction (thioesterase and RGAE). As indicated by the alignments, the SGNH-hydrolases that catalyze acyltransfer reactions in water share a common pattern of insertions that include insertions 1-4 and deletion, as described above. In addition, several specific residues found at dimer and inter-dimer interfaces are also conserved. Among these are the Glu 51, Tyr 73 and His 81, at the dimer interface; Arg 101 and Asp 106, which form a salt bridge between dimers in the octamer; and Phe 174, which along with Leu 105 and Leu 109, creates a hydrophobic pocket for Trp 149 from the insertion 4 loop. All of the enzymes that catalyze acyltransfer in water share the property of appearing in solution as aggregates, either as tetramers or octamers. Moreover, the residues and loops, which form the dimer and inter-dimer interfaces, are conserved in the sequences of SGNH-hydrolases that show acyltransferase activity in water. Therefore, the restricted access to the active site through the hydrophobic channel is a common feature of these enzymes. Indeed, it is contemplated that these enzymes comprise a subclass of enzymes of the hydrolase family.

[244] The architecture of the MsAcT enzyme provides a structural basis for the control of substrate and the exclusion and partitioning of water that contributes to its ability to catalyze alcoholysis reactions in vast molar excesses of water. Although it is not intended that the

present invention be limited to any particular mechanism, for MsAcT, this appears to arise from an intricate oligomerization resulting in a highly restrictive reactive channel, which favors alcoholysis over hydrolysis, at least under some conditions. The pattern of insertions found in the MsAcT molecule contributes to this in two distinct ways: first, to create a channel leading to the reactive center; and second, by stabilizing the formation of an oligomeric structure that further elaborates the reactive cavity and contributes to the important synthetic capabilities of the enzyme.

[245] During the development of the present invention, it was noted that the SGNH hydrolases have a similar overall folding pattern as the α/β hydrolases, which comprise another large superfamily of enzymes. Figure 8 provides a schematic diagram of the α/β hydrolase fold. As indicated in this Figure, the overall folding pattern is similar to that of the SGNH hydrolases (See, Figure 4). As indicated, there is an identical helix crossover pattern in the vicinity of insertion 1, which occurs between strand $\beta 2$ and helix H3. Comparison of these structures provides an easy way to identify the location of the remaining insertion sites, particularly insertion 2 and insertion 4. Insertions 1, 2 and 4 of 8-18 residues would be introduced preceding the homologous helices HA, HB and HD in the α/β hydrolase fold for any enzyme of this class of superfamily. It is contemplated that this would introduce similar synthetic properties as observed with the SGNH hydrolases.

[246] Thus, the present invention provides means to mimic natural engineering by modifying other SGNH-hydrolases, lipases and/or α/β hydrolases to form new acyltransferases that catalyze transfer reactions in water, thereby creating high value materials (*e.g.*, for biotechnology and the pharmaceutical industries).

EXAMPLE 6

Structural alignment of *M. smegmatis* aryl esterase with *E. coli* thioesterase

[247] Tertiary structures can be aligned to superimpose central five strand, five helix structures of SGNH hydrolases.

[248] Stereoviews showing aligned segments highlighted as sticks showing all main chain atoms for *E. coli* thioesterase (light gray) and *M. smegmatis* aryl esterase (dark gray) are shown in orthogonal stereoviews in Figures 10 and 11. Stereoviews of aligned structural segments are

highlighted in these figures as stick figures showing all main chain atoms. It is important to note the common directionality of the peptide bonds forming the core structure common to all SGNH hydrolases.

[249] Table 4 lists the residues corresponding to common SGNH hydrolase secondary structural elements.

Table 4

	E coli thio esterase	M smegmatis aryl esterase
Strand 1	4-10	5-11
Helix 1	22-30	33-41
Strand 2	35-41	46-52
Helix 2	49-60	69-80
Strand 3	64-72	85-93
Helix 3	88-95	111-118
Strand 4	101-105	133-137
Helix 4	124-129	167-172
Strand 5	134-137	177-180
Helix 5	166-177	204-215

EXAMPLE 7

Identification of Polypeptides Having Enzymatic Aqueous Acylation and/or Perhydrolysis Activity

[250] Amino acid sequences of thioesterase from *E. coli* (Genbank accession no. YP_668488, SEQ ID NO:4), acylesterase from *Agrobacterium tumefaciens* (Genbank accession no. AAD02335, SEQ ID NO:3), rhamnogalacturonan acetylerase from *Aspergillus aculeatus* (Genbank accession no. 1DEO_A., SEQ ID NO:5), acylesterase from *Sinorhizobium meliloti* (Genbank accession no. np_436338, SEQ ID NO:2), and arylesterase from *Mycobacterium smegmatis* (Genbank accession no. YP_890535, SEQ ID NO:1) were multiply aligned automatically using the program MUSCLE (which can be downloaded at <http://www.drive5.com/muscle/>) using default parameters. The obtained multiple alignment was further refined with MUSCLE, again using default parameters.

[251] The alignment was further manually adjusted according to the structural features of the x-ray crystallographic models for the *M. smegmatis* arylesterase and *E. coli* thioesterase.

Structurally, the molecules were aligned on the basis of common secondary structural elements (alpha helices and beta sheet strands). The sequences were then aligned with regard to the residues that share structurally homologous main chain atoms. These are atoms that share common root mean square deviations with directionality of the peptide bond atoms linking the residues following an identical pattern. A similar alignment was performed for the *A. aculeatus* rhamnogalacturonan acylesterase. The sequences for active enzymes from *S. meliloti* and *A. tumifaciens*, which share sufficient sequence homology to allow a conventional sequence alignment relative to the sequence of *M. smegmatis* arylesterase, were then aligned according to the alignment of the *M. smegmatis* sequence relative to the sequences of the *E. coli* acylesterase and the *A. aculeatus* rhamnogalacturonan acylesterase. The results of this alignment revealed that loops linking specific common structural elements were associated with insertions (see Figures 6 and 9).

[252] A Hidden Markov Model (HMM) was then constructed. The HMM is based on a set of states, each of which is associated with a probability distribution. Transitions among the states are governed by a set of probabilities called transition probabilities. In a particular state an outcome or observation can be generated according to the associated probability distribution. It is only the outcome, and not the state, that is visible to an external observer, and therefore states are "hidden" to the outside observer. The extracted model is composed of structured parameters that can then be used to perform further analysis, for example, for pattern recognition applications. (See "Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids" by Durbin et al., 1998, Cambridge University Press.)

[253] To build an HMM, sequences were removed from the above multiple alignment except the sequences from *M. smegmatis*, *S. meliloti*, and *A. tumifaciens*, while the relative alignment was kept. Gaps at positions that remain in all sequences were removed. The resultant multiple alignment was used to build the HMM with the program hmmbuild in the HMMER software package (which can be downloaded at <http://bioweb.pasteur.fr/seqanal/motif/hmmer-uk.html>). The HMM was calibrated with hmmscalibrate with default parameters. The resulting HMM was then used to search the nonredundant protein database obtained from the National Center for Biotechnology Information (NCBI), National Library of Medicine, National Institutes of Health. The search was performed with default parameters. The search results are ordered in the output from the program with expect values (E values). The E-value is a parameter that describes the number of hits one can "expect" to see by chance when searching a database of a particular size.

The lower the value, the more significant the hit is. The lowest value for E is zero. An E value of 0.003 was used as the cutoff to obtain a final list of database sequences that are considered hits. The NCBI accession numbers were extracted from the hit sequence description and used to obtain from NCBI other related information such as organism name from which the sequence was derived. The hits from this search are presented below. One of the hits was the arylesterase sequence from *M. smegmatis* (SEQ ID NO:1. Four of the hits were previously identified in PCT Application No. WO 05/056782 A2 as having sequence homology to *M. smegmatis* arylesterase (SEQ ID NOs: 10, 18, 28, and 41).

[254] >gi|16263545|ref|NP_436338.1| hypothetical protein SMa1993 [Sinorhizobium meliloti 1021]|| gi|14524247|gb|AAK65750.1| Conserved hypothetical protein [Sinorhizobium meliloti 1021]

MTINSHSWRTLMEKRSVLCFGDSLWGWIPVKESSPTLRYPYEQRWGAMAARLGD
GYHIIIEGLSARTTSLDDPNDARLNGSTYLPALASHLPLDLVIIMLGTNDTKSYFHRTPY
EIANGMGKLVGQVLTCAAGVGTPYPAPKVLVAVPPPLAPMPDPWFEGMFGGGYEKSK
ELSGLYKALADFMKVEFFAAGDCISTDGIDGIHLSAETNIRLGHAIAADKVAALF (SEQ ID
NO:10)

[255] >gi|118468600|ref|YP_890535.1| lipolytic enzyme, G-D-S-L [Mycobacterium smegmatis str. MC2 155]|| gi|118169887|gb|ABK70783.1| lipolytic enzyme, G-D-S-L [Mycobacterium smegmatis str. MC2 155]

MAKRILCFGDSLWGWVPVEDGAPTERFAPDVRWTGVLAQQLGADFEVIEGLSARTT
NDDPTDPRLNGASYLPSCATHLPLDLVIIMLGTNDTKAYFRRTPLDIALGMSVLVTQV
LTSAGGVGTTPAPKVLVSPPLAPMPHPWFQLIFEGGEQKTTELARVYSALASFMKV
PFFDAGSVISTDGVDDGIHFTEANNRDLGVALAEQVRSL (SEQ ID NO:1)

[256] >gi|4105272|gb|AAD02335.1| arylesterase [Agrobacterium tumefaciens]
MVKSVLCFGDSLWGSDAETGGRHSHDDLWPSVLQKALGPDVKVIHEGLGGRTTAYD
DHTADCDRNGARLLPTLLHSHAPLDLVIIMLGTNDLKPSIHGSAIVAMKGVERLVKLVR
NHVWQVPDWEAPDVLIVAPPQLCETANPVMGAIFRDAIDESAMLAPVYRDLADDLDCG
FFDAGSVARTTPVDGVHLD AENTRAIGRGLEPVVRMMLGL (SEQ ID NO:11)

[257] >gi|15888758|ref|NP_354439.1| hypothetical protein AGR_C_2642 [Agrobacterium tumefaciens str. C58]|| gi|17935333|ref|NP_532123.1| arylesterase [Agrobacterium tumefaciens str. C58]|| gi|15156506|gb|AAK87224.1| AGR_C_2642p [Agrobacterium tumefaciens str. C58]|| gi|17739852|gb|AAL42439.1| arylesterase [Agrobacterium tumefaciens str. C58]

MVKSVLCFGDSL TWGSNAETGGRHSHDDLWPSVLQKALGSDVHVIHEGLGGRTTAYD
DHTGDCDRNGARLLPTLLHSHAPLDMVIIMLG TNDMKPAIHGSAIVAMKGVERLVKLT
RNHVWQVSDWEAPDVLIVAPPQLCETANPFMGAI FRDAIDESAMLASVYRDLADELDC
GFFDAGSVARTTPVDGVHLDAENTRAIGRGLEPVVRMMLGL (SEQ ID NO:12)

[258] >gi|10954724|ref|NP_066659.1| hypothetical protein pRi1724_p079 [Agrobacterium
rhizogenes]|| gi|8918724|dbj|BAA97789.1| hypothetical protein [Agrobacterium
rhizogenes]|| gi|10567388|dbj|BAB16197.1| hypothetical protein [Agrobacterium rhizogenes]
MAESRSILCFGDSL TWGWIPVPESPTLRYPFQ RWTGAMAAALGDGYSIIIEGLSARTT
SVEDPNDPRLNGSAYLPMALASHLPLDLVILLGTNDTKSYFRRTPYEIANGMGKLAGQ
VLTSAGGIGTPYPAPKLLIVSPPLAPMPDPWFEGMFGGGYEKSLELAKQYKALANFLK
VDFLDAGEFVKTDGCDGIHFSAETNITLGHAI AAKVEAIFSQEKNAAA (SEQ ID NO:13)

[259] >gi|126438117|ref|YP_001073808.1| lipolytic enzyme, G-D-S-L family [Mycobacterium
sp. JLS]|| gi|126237917|gb|ABO01318.1| lipolytic enzyme, G-D-S-L family [Mycobacterium sp.
JLS]

MAMKRVLCFGDSL TWGWVPVEDGVPTERYPRDVRWTGVLADELGADYTVIEEGLSAR
TTSADDPSPRLNGAAYLPACLASHLPLDLVIMLG TNDTKAYFHREPLDIAMGMGILV
TQVLTAAGGVGTVPAPRTL VVAPPPLAQIPHPWFDLIFAGGHEKTAALSSAYSALASH
MKVDFFDAGSVISTDGV DGIHFTEQNNRDLGAALAAQVRTLLS (SEQ ID NO:14)

[260] >gi|108802136|ref|YP_642333.1| lipolytic enzyme, G-D-S-L [Mycobacterium sp.
MCS]|| gi|119871289|ref|YP_941241.1| lipolytic enzyme, G-D-S-L family [Mycobacterium sp.
KMS]|| gi|108772555|gb|ABG11277.1| lipolytic enzyme, G-D-S-L [Mycobacterium sp.
MCS]|| gi|119697378|gb|ABL94451.1| lipolytic enzyme, G-D-S-L family [Mycobacterium sp.
KMS]

MAMKRVLCFGDSL TWGWVPVEDGVPTERYPRDVRWTGVLADELGADYTVIEEGLSAR
TTSADDPSPRLNGAAYLPACLASHLPLDLVIMLG TNDTKAYFHREPLDIAMGMGILV
TQVLTAAGGVGTVPAPRTL VVAPPPLAQIPHPWFDLIFAGGHEKTAALSSAYSALASH
MKVDFFDAGSVISTDGV DGIHFTEQNNRDLGAALAAQVRTLLS (SEQ ID NO:15)

[261] >gi|118592396|ref|ZP_01549788.1| hypothetical protein SIAM614_26251 [Stappia
aggregata IAM 12614]|| gi|118435054|gb|EAV41703.1| hypothetical protein SIAM614_26251
[Stappia aggregata IAM 12614]

MANEKSILCFGDSL TWGWIPVVEGAPTLRYPYEQ RWTGAMAAELGEGYRVIEEGLSAR
TTSADDPNDPRLNGSQYLPSALASHLPLDLAILLGTNDTKPFFNRTPYDIAYGMSKLVG

QVLTSGGGIGTPYPAPKCLVVAPPPLTPMPHPYFQGMFGGGAHEKSAALAEQYRNMAADF
MKVDFLSAADHITTDGCDGIHFTAQNNIDLKAIAGKVREILSRQQAEAA (SEQ ID
NO:16)

[262] >gi|150396299|ref|YP_001326766.1| lipolytic protein G-D-S-L family [Sinorhizobium
medicae WSM419]|| gi|150027814|gb|ABR59931.1| lipolytic protein G-D-S-L family
[Sinorhizobium medicae WSM419]

[263] MKTVLCYGD SLTWGFD AVGSGRHA LEDRWPSVLQKGLGSGTHVIAEGLNGRRTT
AYDDHLADCDRNGARVLPTVLHTHAPLDLIVIMLGSNDMKPIIHGTAFGAVKGIERLVN
LVRRHDWPTPGEDGPDILIVSPPLCETANSFAAMFAGGVEQSAMLAPLYRDLADELD
CGFFDGGSVARTSPIDGVHLDAENTRAIGRGLEPVVRMMLGL (SEQ ID NO:17)

[264] >gi|15965201|ref|NP_385554.1| PROBABLE ARYLESTERASE PROTEIN
[Sinorhizobium meliloti 1021]|| gi|11061691|emb|CAC14575.1| putative arylesterase
[Sinorhizobium meliloti]|| gi|15074381|emb|CAC46027.1| PROBABLE ARYLESTERASE
PROTEIN [Sinorhizobium meliloti]

MKTVLCYGD SLTWGYDA TGSGRHA LEDRWPSVLQKALGSDAHVIAEGLNGRRTTAYDD
HLADCDRNGARVLPTVLHTHAPLDLIVFMLGSNDMKPIIHGTAFGAVKGIERLVNLVRR
HDWPTETEEGPEILIVSPPLCETANSFAAMFAGGVEQSAMLAPLYRDLADELDCGFFD
GGSVARTTPIDGVHLDAENTRAVGRGLEPVVRMMLGL (SEQ ID NO:18)

[265] >gi|86357557|ref|YP_469449.1| probable arylesterase protein [Rhizobium etli CFN
42]|| gi|86281659|gb|ABC90722.1| probable arylesterase protein [Rhizobium etli CFN 42]
MTKTVLCYGD SLTWGHD PENIGRHEYKNRWPSVLQAALGSEARIIAEGLNGRRTTAFDD
HLADCDRNGARVLPTILQTHAPLDLVILLGTNDMKPVVAGSAFAACQGISRLVRLIRNH
AWPFEDGPEILIVAPPAICATGNVPFAASFPGGIEESAKLATLYRDLADELGCGFFDGNS
VAKTTPIDGIHLDAENTRALGRGLESIVRMMLGI (SEQ ID NO:19)

[266] >gi|116252000|ref|YP_767838.1| putative arylesterase [Rhizobium leguminosarum bv.
viciae 3841]|| gi|115256648|emb|CAK07736.1| putative arylesterase [Rhizobium leguminosarum
bv. viciae 3841]

MTKTVLCYGD SLTWGYDA ETIGRHDYKNRWPSVLQAALGGDARIIEGLNGRRTTAFDD
HLADCDRNGARILPTILQTHAPLDLVILLGTNDMKPVVAGSAFAACQGISRLVRLIRNH
AWPFEDGPEILIVAPPAIRATGNVPFAASFPGGIEESSKLA

TLYRDLADELGCGFFDGNVAKTTPIDGIHLDAENTRALGRGLESIVRMMLGI (SEQ ID NO:20)

[267] >gi|110633979|ref|YP_674187.1| lipolytic enzyme, G-D-S-L [Mesorhizobium sp. BNC1]|| gi|110284963|gb|ABG63022.1| lipolytic enzyme, G-D-S-L [Mesorhizobium sp. BNC1] MKTILCYGDSL TWGYDAVGPSRHAYEDRWPSVLQGRLGSSARVIAEGLCGRTTAFDD WVAGADRNGARILPTLLATHSPLDLVIVMLGTNDMKSFVCGRAIGAKQGMERIVQIIRG QPYSFNYKVPSILLVAPPPLCATENSDFAEIFEGGMAESQKLAPLYAALAQQTGCAFFDA GTVARTTPLDGIHLDAENTRAIGAGLEPVVRQALGL (SEQ ID NO:21)

[268] >gi|13470623|ref|NP_102192.1| arylesterase [Mesorhizobium loti MAFF303099]|| gi|14021365|dbj|BAB47978.1| arylesterase [Mesorhizobium loti MAFF303099] MKTVLCYGDSL TWGYNAEGGRHALED RWPSVLQASLGGGVQVIADGLNGRTTAFDDH LAGADRNGARLLPTALTTHAPIDLIVIMLGANDMKPWIHGNPVA AKQGIQRLIDIVRGHD YPFDWPAPQILIVSPPVVSRTENADFREMFAGGDEASKQLAPQYAALADEVGCGFFDAG TVAQTTPLDGVHLDAENTRNIGKALTSVVRVMLEL (SEQ ID NO:22)

[269] >gi|114704548|ref|ZP_01437456.1| probable arylesterase protein [Fulvimarina pelagi HTCC2506]|| gi|114539333|gb|EAU42453.1| probable arylesterase protein [Fulvimarina pelagi HTCC2506] MKTVLAFGDSL TWGYDPETMLRHDFDRWPNVVG RALGHKVRVISEGLNGRTTAFDD HLVPAERNGVKILPTLLASHQPLDLVVILLGTNDLKRDTGGGRVFESRQGLERLVEICQT FPYQRGYEPPKLLL VAPPPFCSTTEPDFSLLFGHAVEESTHFR SAIKVAEEYGIAMFDAGEVCETSPIDGVHLDAENTRRLGEALIEPIRTLLELDD (SEQ ID NO:23)

[270] >gi|90419626|ref|ZP_01227536.1| arylesterase [Aurantimonas sp. SI85-9A1]|| gi|90336563|gb|EAS50304.1| arylesterase [Aurantimonas sp. SI85-9A1] MRTVLCYGDSL TWGYNPETGLRHDYDSRWPNVLARALGEDVVVITDALNGRTTVFDD QMVMAERNGVKTLPTALGAHQPLDLVVLMLGTNDLKRHTGGGRVFESQGMERLVEIIL TFPYQRGYDAPKILIVAPPMFCATTEPDFTLLFGHALEESQHFRSACTKVAEEYGCAMFD ASEVCETTPLDGIHLSAEHTRALGEALIAPVKALLDETA (SEQ ID NO:24)

[271] >gi|15891484|ref|NP_357156.1| hypothetical protein AGR_L_2749 [Agrobacterium tumefaciens str. C58]|| gi|17937162|ref|NP_533951.1| arylesterase [Agrobacterium tumefaciens

str. C58]] gi|15159896|gb|AAK89941.1| AGR_L_2749p [Agrobacterium tumefaciens str.

C58]] gi|17741852|gb|AAL44267.1| arylesterase [Agrobacterium tumefaciens str. C58]

MKTVLAFGDSL TWGADPATGLRHPVEHRWPDVLEAELAGKAKVHPEGLGGRTTCYDD
HAGPACRNGARALEVALSCHMPLDLVIIMLG TNDIKPVHGGRAEAAVSMRRLAQIVETF
IYKPREAVPKLLIVAPPCVAGPGGEPAGGRDIEQSMRLAPLYRKLAAELGHHFFDAGSV
ASASPDGVHLDASATAAIGRALAAPVRDILG (SEQ ID NO:25)

[272] >gi|118592401|ref|ZP_01549793.1| hypothetical protein SIAM614_26276 [Stappia

aggregata IAM 12614]] gi|118435059|gb|EAV41708.1| hypothetical protein SIAM614_26276

[Stappia aggregata IAM 12614]

MRTILCYGDSNTHGQIPGATPLDRYTLLQRWPGVLARELGTGWHIIEGLSGRTTVHDD
PIEGALKNGRTYLRPCLMSHAPLDLVIIMLG TNDLKARFGQPASEVAMGIGCLIHDIKELS
PGPGGTVP EIMVVSPPPMLDDIKEWENIFKAAQKSH ELALQFEIMSDSLEVHFFDAGSV
STCDPLDGFHINAQA HESLGVALAREVEAIGWK (SEQ ID NO:26)

[273] >gi|10954719|ref|NP_066654.1| hypothetical protein pRi1724_p074 [Agrobacterium

rhizogenes]] gi|8918719|dbj|BAA97784.1| hypothetical protein [Agrobacterium

rhizogenes]] gi|10567383|dbj|BAB16192.1| hypothetical protein [Agrobacterium rhizogenes]

MICHKGGEEMRSVLCYGDSNTHGQIPGGSPLDRYGPNERWPGVLRRELGSQWVIEEGL
SGRTTVRDDPIEGTMKNGRTYLRPCLMSHAILDLVIIMLG TNDLKARFGQPPSEVAMGIG
CLVYDIRELAPGPGGKPPEIMVVAPPPMLDDIKEWEPIFSGAQEKSRRLLALEFEIADSLEV
HFFDAATVASCDPCDGFHINREAHEALGTALAREVEAIGWR (SEQ ID NO:27)

[274] >gi|16263550|ref|NP_436343.1| hypothetical protein SMa2002 [Sinorhizobium meliloti

1021]] gi|14524252|gb|AAK65755.1| Conserved hypothetical protein [Sinorhizobium meliloti
1021]

MEETVARTVLCFGDSNTHGQVPGRGPLDRYRREQRWGGVLQGLLGPNWQVIEGLSGR
TTVHDDPIEGSLKNGRIYLRPCLQSHAPLDLIIMLG TNDLKRRFNMPPSEVAMGIGCLVH
DIRELSPGRTGNDPEIMIVAPPPMLEDLKEWESIFSGAQEKSRLALEFEIMADSLEAHFF
DAGTVCQCSPADGFHIDE DAHRLLG EALAQEVLAIGWPDA (SEQ ID NO:28)

[275] >gi|78065946|ref|YP_368715.1| Lipolytic enzyme, G-D-S-L [Burkholderia sp.

383]] gi|77966691|gb|ABB08071.1| Lipolytic enzyme, G-D-S-L [Burkholderia sp. 383]

MTMTQKTVLCYGDSNTHGTRPMTHAGGLGRFAREERWTGVLAQTLGASWRVIEEGLP
ARTTVHDDPIEGRHKNGLSYLRACVESHLPVDV VVLM LG TNDLKTRFSVTPADIATSVG

VLLAKIAACGAGPSGASPKLVLMAPAPIVEVGFLGEIFAGGAAKSRQLAKRYEQVASDA
GAHFLDAGAIVEVSPVDGVHFAADQHRVLGQRVAALLQQIA (SEQ ID NO:29)

[276] >gi|118591350|ref|ZP_01548748.1| Lipolytic enzyme, G-D-S-L [Stappia aggregata IAM
12614]|| gi|118436022|gb|EAV42665.1| Lipolytic enzyme, G-D-S-L [Stappia aggregata IAM
12614]

MSKTEKEPVTTVVCFGDSL TWGFNPVDKSR YGHDVRWTRLLQKELGDGFYVVEEGVNG
RTTVFEDPVKGDKNGLAHLATVRKTHMPIDILIVMLGTNDLQARFGMNAETIAIAMGRL
LDFARRPTDDVEGRAPKVLLMAPPPLAPLAGTPYAAQFSDQSVAESHRLADCYREKAA
EYGAAFFDTGTVISASPIDAVHFEAEPQADLAKAVAAEVRKLAN (SEQ ID NO:30)

[277] >gi|17549466|ref|NP_522806.1| PUTATIVE ARYLESTERASE PROTEIN [Ralstonia
solanacearum GMI1000]|| gi|17431719|emb|CAD18396.1| putative arylesterase protein
[Ralstonia solanacearum]

MQQILLYSDSLSWGIIIPGTRRRLPFAARWAGVMEHALQAQGHAVRIVEDCLNGRTTVL
DDPARPGRNGLQGLAQRIEHAHAPLALVILMLGTNDFQAIFRHTAQDAAQGV AQLVRAIR
QAPIEPGMPVPPVLIVVPPAITAPAGAMADKFADAQPKCAGLAQAYRATAQTLGCHVF
DANSVTPASRVDGIHL DADQHAQLGRAMAQVVGTL LAQ (SEQ ID NO:31)

[278] >gi|126729853|ref|ZP_01745666.1| lipolytic enzyme, G-D-S-L [Sagittula stellata E-
37]|| gi|126709972|gb|EBA09025.1| lipolytic enzyme, G-D-S-L [Sagittula stellata E-37]

MGIDQARAAGLTASKGGTMAVILCFGDSNTHGTAPLDRPGGQARHPKGQRWPDVLAS
ELGNAHEVIQEGLPGRRTTVHDDPVEGGCRNGQAVLTAILHTRPIDLMLIMLGTNDLKN
RFSVTAWEIARSVRLVIMSRAEAVVSDFMVIAPAPVRECGSLEAPYLGAEARQTGLSE
RLEEMAERQGIGFFDAGHYATVSPKDG VHWEPESHIALGQALAPAVRARLS (SEQ ID
NO:32)

[279] >gi|87120951|ref|ZP_01076843.1| arylesterase [Marinomonas sp.
MED121]|| gi|86163789|gb|EAQ65062.1| arylesterase [Marinomonas sp. MED121]

MATILCYGDSL TWGGSPYGGRFPPHCRWPEILNELLGNHHNVINFGLPGRRTTMWDDPFN
EGRNGAKYVNSAMEIFGPVDLLIVMLGTNDL KRYFNASAYEAAKGVEQIIKKVRQPNE
HEFPTPKILIVAPPTILSPKGEEAEVFKGGVEKSRDLHQAYQKVAIKNDCLFVNAAGLLQ
PSVQDGVHLDTDGNSLLASALCSIVSDSFKG (SEQ ID NO:33)

[280] >gi|83749609|ref|ZP_00946593.1| Hypothetical Protein RRSL_00374 [Ralstonia solanacearum UW551]|| gi|83723716|gb|EAP70910.1| Hypothetical Protein RRSL_00374 [Ralstonia solanacearum UW551]

MASAARLPDMQQVLLYADSLSWGIVPGTRQRLPFAARWAGVMEHALLAQGHAVRLIE
DCLNGRRTTVLDDPVRPGRNGLHGLAQRIEAHAPLALVILMLGTNDFQAIFPHTAQDVAH
GVAQLVRAIRQAPVEPGMPVPPVLIVVPPAIVAPAGAMADKFAGAQSRSAAGLAQAYRA
VAQTLGCHLFDANSVTPASRADGIHLDAHQHARLGRAMAQVVGPLLTD (SEQ ID
NO:34)

[281] >gi|89055632|ref|YP_511083.1| lipolytic enzyme, G-D-S-L [Jannaschia sp. CCS1]|| gi|88865181|gb|ABD56058.1| lipolytic enzyme, G-D-S-L [Jannaschia sp. CCS1]
MKRILIFGDSNSHGTVPRLVLGQSDRFAPGVWPDPVLAQAQTGHSVFTEGLPGRTTVFED
PVDGGARNGAAVLPAVLLSHVPLDAVVILLGTNDLKPRFATSAFDIAKSVERLVGIVRAL
VPAAKVLVVSAPVTETGVLADAFAGAEARQAAGLDGHLEAAALRVGAGFARAGDYAE
VSPVDGVHLEADGHRALAMGLIPHVDALLAAPPKTGGGLPAPDPSAPEPPVTLARAVPP
DWVDYNNHMNEAFYLTAFSDAADQMLHWAGMDADCVAAGASVFTVETHIRHLGEVS
IGDALRVTTTRVVQGGGKKLHLWHEMWVGDALCATGEQLLLHMDLRARTSALPPERIA
DWLGRAKVAHGPLSLPDGFGRSVAQRT (SEQ ID NO:35)

[282] >gi|89067298|ref|ZP_01154811.1| Lipolytic enzyme, G-D-S-L [Oceanicola granulosus HTCC2516]|| gi|89046867|gb|EAR52921.1| Lipolytic enzyme, G-D-S-L [Oceanicola granulosus HTCC2516]
MPTLLCFGDSNTHGSPANGPGPYARYGHGVRWPSVCAGALGERWEVIEEGLPGRTAQ
FDDPGMGAHMNGWPALRIALSHGPLDVLAVMLGTNDVKTRFAPTPARVTAGIAGLLD
IALSEAMQERHGGRLRVLLVAPPKVVLGDLAGEFHGAEEVSAALPDHYRALAEARGCA
FLDANDVISVSPLDGIHFDETAHGALGKAVAEQVTRLA (SEQ ID NO:36)

[283] >gi|139437172|ref|ZP_01771332.1| Hypothetical protein COLAER_00311 [Collinsella aerofaciens ATCC 25986]|| gi|133776819|gb|EBA40639.1| Hypothetical protein COLAER_00311 [Collinsella aerofaciens ATCC 25986]
MKNVLCFGDSNTYGYDPAGMRDGTAVRYAQDVRWCGVAQRDLGEGWHVIEEGLNG
RTTVRDDMCHLDTNLNGIRALPMLLEAHKPLDAIVIMLGTNDCKTVFNVTASDIARGA
MALIRAVRAFPWTDAAAPCPRILLMAPIKIKPQIADVYMTDFDEHSVEASEHFGEYYAHV
AEQFGCDFLNAAEFAEPGDIDYLMMPESHESLGHAVAANKLQEMLGE (SEQ ID NO:37)

[284] >gi|118747166|ref|ZP_01595039.1| lipolytic enzyme, G-D-S-L [Marinomonas sp. MWYL1] || gi|150835082|gb|ABR69058.1| lipolytic protein G-D-S-L family [Marinomonas sp. MWYL1]

MATILCYGDSL TWGRVPNGGRYPKHLRWPTMLNDLLGQQHQVINFGLPGRRTTIWNDPF
LEGRNGLTYLQAALETFGPVDILILMLGTNDLKRHFHIGAYEAAKGIEKLEKSRIPNSHD
FPVPTLVVIAPPNLSPTGSMASFDGAIEKSQHFHQYYQDIAIRNQCIFLNAAGVLQPSD
VDGVHLD TQANEQLANAIYSLIKMEI (SEQ ID NO:38)

[285] >gi|126664421|ref|ZP_01735405.1| hypothetical protein MELB17_01245 [Marinobacter sp. ELB17] || gi|126630747|gb|EBA01361.1| hypothetical protein MELB17_01245 [Marinobacter sp. ELB17]

MHQILVYADSLSWGIIIPDTRGRFRFDQRWPGVLETALLNSGKSVRIEDCLNGRRTVWD
DPYKPGRNGLEGLEQRIEINSPLSLVIMLLGTNDFQSMHNLTAWQSAQGLKALVGAIIR
APIEPGMPVPAIVLVAPPELQKARGNIAPKFESAEDKAVGLAGAIQVVAEECGCYFDA
GSVTPTS RVDGVHLDDEDQHRTLGLALAKIIQPLVQ (SEQ ID NO:39)

[286] >gi|32472752|ref|NP_865746.1| hypothetical protein RB3830 [Rhodopirellula baltica SH 1] || gi|32443989|emb|CAD73431.1| conserved hypothetical protein [Rhodopirellula baltica SH 1]

MHSILYGDLSWSWGIIIPGTRRRFAFHQRWPGVMEIELRQTGIDARVIEDCLNGRRVLEDPI
KPGRNGLDGLQQRIEINSPLSLVVLFLGTNDFQSVHEFHAEQSAQGLALLVDAIRRS PFEP
GMPTPKILLVAPPTVHHPKLDMAAKFQNAETKSTGLADAIRKVSTEHSCEFFDAATVTT
TSVVDGVHLDQE QHQALGTALASTIAEILADC (SEQ ID NO:40)

[287] >gi|73541632|ref|YP_296152.1| putative arylesterase protein [Ralstonia eutropha JMP134] || gi|72119045|gb|AAZ61308.1| putative arylesterase protein [Ralstonia eutropha JMP134]

MPLTAPSEVDPLQILVYADSLSWGIVPGTRRRLPFPVRWPGRLELGLNADGGAPVRIED
CLNGRRTVWDDPFKPGRNGLQGLAQRIEIHSPVALVVLMLGNND FQSMHPHNA WHAA
QGVGALVHAIRTAPIEPGMPVPPI LVVVPPIRTPCGPLAPKFAGGEHKWAGLPEALREL
CATVDCSLFDAGTVIQSSAVDGVHLDADAHVALGDALQP VVRALLAESSGHPS (SEQ
ID NO:41)

[288] >gi|121609289|ref|YP_997096.1| lipolytic enzyme, G-D-S-L family [Verminephrobacter eiseniae EF01-2] gi|121553929|gb|ABM58078.1| lipolytic enzyme, G-D-S-L family [Verminephrobacter eiseniae EF01-2]

MQHILVYADSLSWGIVPGTRQRLAFAQRWPGALELALLRQGLAVRVIEDCLNGRRRTVW
DDPFKPGRNGLQGLAQRIEHSPLALVLLMLGTNDFQSSHQNTAWHSAQGLAAVIDTIR
QAPIEPGLPVPPIVLAPPPITRPRGLMAPKFEGAQARCAGLAAACEALARALGCQFLDT
AGVVTTSSVDGVHLDADQHALLAQALTPLVAARLAAPGAATGG (SEQ ID NO:42)

[289] >gi|113954134|ref|YP_730806.1| hypothetical protein sync_1601 [Synechococcus sp. CC9311] gi|113881485|gb|ABI46443.1| conserved hypothetical protein [Synechococcus sp. CC9311]

MEPKSILCFGDSNTWGMAPDGSGRLPFKTRWP NRLQQILNQNLNHQSWNVIEQGLNS
RTWVMDDPLGAVNYGGDYSCNGRQDLSMILHSCKPLNVVLLALGCNDCKSHLHLSPE
EITSGAKILIHDIRMSYQCGPRYSNHPPTIVLVTPGVIQTTPQSLAWGFKGAAEKSRMLPS
LYRNLAEQESVFFFD TQAI AETSPLDGVHFGADQQDSIAAGLAELITAIS (SEQ ID
NO:43)

[290] >gi|113868100|ref|YP_726589.1| Arylesterase protein [Ralstonia eutropha H16] gi|113526876|emb|CAJ93221.1| Arylesterase protein [Ralstonia eutropha H16]
MTAAPLAETDPLQILVYADSLSWGIVPGTRRRLPFPVRWPGRLELGLNADGGAPVRVIE
DCLNGRRRTVWDDPFKPGRNGLHGLAQ RVEIHSPLALVVLMLGNND FQSMHPHNA WHA
AQGIATLVQAIRSAPIEPGMPVPPVLVVT PPPIRTPRGPLAPKFAGGEHKWTGLPQAVAE
VCTQLGCPLFDAGTVIQSSEVDGVHLDADAHLALGDALRPVVRELLRG (SEQ ID
NO:44)

[291] >gi|149834185|gb|EDM89265.1| hypothetical protein RUMOB_E_00388 [Ruminococcus obeum ATCC 29174]
MNILCFGDSNTWGYKPKDKSGRYDENIRWTGLLQKKLGSGYHIEEGLCGRTTVFQDEL
EGRRLDLIGVTVMHNPIDLMIIMLG TNDCKTRYRASASVIAKGLDQVIRKARQNASR
HFDLLVISPIHLGVGVGDADFDPEFDAASVAVSRNLANEYRKIALQNHA AFLNASDFAA
PSVTDREHMDEKGHAALADAIYNKILALQKGLSHVI (SEQ ID NO:45)

[292] >gi|89093144|ref|ZP_01166094.1| hypothetical protein MED92_03667 [Oceanospirillum sp. MED92] gi|89082440|gb|EAR61662.1| hypothetical protein MED92_03667 [Oceanospirillum sp. MED92]

MEQILVYSDSLTWGIIPDTRKRLTFEQRWPGVCENGLRALGMNVRILENCLNGRRTTWS
DPFKPGRDGSEHLAQVIEMHSPLRLVVLMLGTNDFQNTHNNDAWLSAQGMAKLINIVR
QAPIEPGMPPEIMVIAPPKMVKPKGPIAHKFQKAELRAEGHSDLLKQICQEHNCNFFDSS
SVTDASVVDGIHLDPQHQTLGEAIAKEIAQIIQE (SEQ ID NO:46)

[293] >gi|149753603|gb|EDM63534.1| hypothetical protein DORLON_01200 [Dorea
longicatena DSM 13814]

MKQILCFGDSNTWGLDGETGKRFPWEERWTGILQEKLADRDRIVEEGLCGRTTIFEDPL
RLGRRGTELLPILLEHTHPDAVVLMLGTNDCKTIFGASAEIIGKGIARLLEQINQYADKMK
VLVISPIYLGEKVWQDGYDQEFSPESVEVSKKLEPVYEKVSQKYGKRFMRAADYVSPSE
ADQEHMDGQSHKILADAIYRKLETEIL (SEQ ID NO:47)

[294] >gi|126734521|ref|ZP_01750267.1| Lipolytic enzyme, G-D-S-L [Roseobacter sp.

CCS2]] gi|126715076|gb|EBA11941.1| Lipolytic enzyme, G-D-S-L [Roseobacter sp. CCS2]

MTKTVLTFGDSNTYGTTPAHARGENRRFGPDTRWPTVMQAALGPDWALVENGLPGRT
TCRADPVMGAHMDGQLGLRIALESSGPIDLLVIMLGTNDLQTHHAATVDQVVGGLAGL
LAIARSEPYQLRHNGFDILLIAPPVVLEQGTYRDTLLGAHEKSRDLPFIAIAQLADHWGIAF
LDAGAHINSSPLDGLHFAAEDHITLGQVIAAKIAAL (SEQ ID NO:48)

[295] >gi|77362107|ref|YP_341681.1| hypothetical protein PSHAb0190 [Pseudoalteromonas
haloplanktis TAC125]] gi|76877018|emb|CAI89235.1| conserved protein of unknown function
[Pseudoalteromonas haloplanktis TAC125]

MIKNILCFGDSNTWGLNPQTGKRFSATVRWPAKLAVKLGANFTIIEAGQPNRALVNNPP
FAGELSGVTYLQSYLNELEIDLIIIDLGTNDLKNRFSNLVHDIAAGLTALIDSILNFYEQSN
APGSPKILILSPAHVREVGSYAGVYKGAEQKVVLAAAFANVASIKQCSFYDLARIVSVS
NEDGVHLDEKQHQKISDALETIREF (SEQ ID NO:49)

[296] >gi|148239418|ref|YP_001224805.1| Possible esterase related to lysophospholipase L1

[Synechococcus sp. WH 7803]] gi|147847957|emb|CAK23508.1| Possible esterase related to
lysophospholipase L1 [Synechococcus sp. WH 7803]

MPTTLQDPRTVLCFGDSNTWGFNPDGSGRFPQSTRWPNQLEHALNRTKPALPWRTVEE
GLNSRTWLHDDPIGAANYGGDYSCNGRADLMTALHSHKPIDVVILALGCNDCKGYLNL
SAAQIAAGARILIHDTRRALNCGPRLQEDMAPQIILMTPPAVLITSQSLAWGFEGAALKS
KDVAEHYHRLAAELGVSCFDVQPVAKPSPLDGVHFDAQAQALIAAGLAACIRQNLEQH
PNPDN (SEQ ID NO:50)

[297] >gi|88859727|ref|ZP_01134367.1| hypothetical protein PTD2_22127 [Pseudoalteromonas tunicata D2]|| gi|88818744|gb|EAR28559.1| hypothetical protein PTD2_22127

[Pseudoalteromonas tunicata D2]

MKKIICFGDSNTFGITPQGGRFNANERWPTLLETLLNQTTTSKQFHVIEQGQPNRTL VHN
PPFHGDKSGLRYFKNCLLDHQPDLALIMLGTDNDLKRKFCLDASEIASGLNALITASLALQ
PALKVLIICPPPILEVGSYAQIYLGGAVKSLALESAYQAVAKELNCAFFAASSVVNSCPIE
GIHWSAKQHTQLAQALVKVVIELCQHEF (SEQ ID NO:51)

[298] >gi|119469810|ref|ZP_01612648.1| hypothetical protein ATW7_01937 [Alteromonadales bacterium TW-7]|| gi|119446793|gb|EAW28065.1| hypothetical protein ATW7_01937

[Alteromonadales bacterium TW-7]

MKKSILCFGDSNTWGLNPATGKRFEKVRWPSVLAERLGDQFAIIEAGQPNRTL VNNPP
FNGDL SGVSYLKPLLEAHS�TAIVIALGTNDLKKRFKLTPEQIGNGLANLIDSIEGFYEGY
KQPKLIILSPPYVSCVGQYKHVYEGAPSKNLALSKEFENVACIKNANFYDLQKIVSVSPID
GVHIEANSHAKISGLVCNLINEVDA (SEQ ID NO:52)

[299] >gi|87119604|ref|ZP_01075501.1| hypothetical protein MED121_06685 [Marinomonas sp. MED121]|| gi|86165080|gb|EAQ66348.1| hypothetical protein MED121_06685

[Marinomonas sp. MED121]

MTKILCFGDSNTWGTKPEGGRYDDHERWSFLLGVYLNMPSSDSINQETESKYL VMEAG
QPNRTL IKNAPFDGDKSGLTYLKPLLAHEPELVIIMLGTDNDLKA KFNLS PADIGAGAGTL
IEQIQAFSQAISHSSATKILLVSPPEIKEVPPYVS VYAGGAQKS QLLAHEFATQAKQYQC
DFLDANQVIKSSDVG IHWGAQA HESLALSIAEYIKQRMSFSHSSN (SEQ ID NO:53)

[300] >gi|88808445|ref|ZP_01123955.1| probable arylesterase protein [Synechococcus sp. WH 7805]|| gi|88787433|gb|EAR18590.1| probable arylesterase protein [Synechococcus sp. WH 7805]

MTATIQDPRTILCFGDSNTWGFNPDGSGRFPQPTRWPNQLERALNQTRPDLPWRTVEEG
LNSRTWLHDDAIGTANYGGDYSCSGRSGLM TALHSHKPIDVVILALGCNDCKGYLNLS
AEQITAGARILHDTRRALNCGPRLDPPTSPRIILMTPPAIRITPQSLTWGFEGAELKSQAL
AEHYCQLAAELDVACFDVQPVANPSTLDGVHFDAQAQALIAAGLAACICQTLGQHPSA
DN (SEQ ID NO:54)

[301] >gi|114769131|ref|ZP_01446757.1| arylesterase [alpha proteobacterium HTCC2255]|| gi|114550048|gb|EAU52929.1| arylesterase [alpha proteobacterium HTCC2255]

MSENKRILVYGDSNSWGYLDDGSGERFQDRWPVEMDKHLSYDLSITLIEECLPARTTNL
ADPELGSSFNGEATLEAVLLSHQPLDHVLIMLGTDNDLKAKFDRNINDIAEAIVNLAEIARS
TPAGRGGWFSDQTPNVSVICPLIIGQRVSDLKWDRFEEWVGAYEKSILLVDTLKRACNA
VNINMIDGNQFGSSSELDPIHWDKENHQRFGQKMAKAIQAFLLD (SEQ ID NO:55)

[302] >gi|150270324|gb|EDM97650.1| hypothetical protein BACCAP_04509 [Bacteroides
capillosus ATCC 29799]

[303] MSAVIQSFWQLIEHRHHQEKRKHKMNVICFGDSNTYGYDPRDYFGGRYDGDSR
WVDILAARTEWSVCNMGQNGREIPSAALDFPADTDLLIVMLGTNDLLQGRKPEQAAEK
LGRFLCGISLERSKILLIAPPPVILGDWVPNQQLIDDSRIFARLCQTLA

[304] EQLGIRFADAGKWDISLAYDGVHFTEQGHRAFAAGLLEELK (SEQ ID NO:56)

[305] >gi|149833987|gb|EDM89067.1| hypothetical protein RUMOB_00190 [Ruminococcus
obeum ATCC 29174]

[306] MPGTLCYGDSNTYGYDPHDTNEGRYTKEVRWTGILDTETDWKVENHGVNGRS
IPHTVSTVKFACQQVRDWHRKPNVWLLVMLGTNDLLENPDFTAADVAKRMERFLKR
MMEEAGLASRKMLRLIAPPAMQEGQWVDRPELLTESRNLGKEYKRIAKRLGIAFTDA
SGWEIPTIYDGVHFTKEGHRIFAENIRKELNI (SEQ ID NO:57)

[307] >gi|104774603|ref|YP_619583.1| Putative lipase [Lactobacillus delbrueckii subsp.
bulgaricus ATCC 11842]] gi|116514727|ref|YP_813633.1| Lysophospholipase L1 related
esterase [Lactobacillus delbrueckii subsp. bulgaricus ATCC BAA-

365]] gi|103423684|emb|CAI98651.1| Putative lipase [Lactobacillus delbrueckii subsp.
bulgaricus ATCC 11842]] gi|116094042|gb|ABJ59195.1| Lysophospholipase L1 related esterase
[Lactobacillus delbrueckii subsp. bulgaricus ATCC BAA-365]

MKKVVFLFGDSIFNAYDGQKDTDRLTKALAKRLGDAYEVVNISVSGACAQDVLPRVSSL
PACDILVVEYGTNDAASWGCSAYDYQEGLESLIKQAQKVTGASDTLVLAPSMPLTNP
EMAAAYSLEKLDEYVDIAQRDAGKTNSFFFDLTHKMEKLDLPSFMIADGLHYSEKGIS
WLADVADQINKMA (SEQ ID NO:58)

[308] All patents and publications mentioned in the specification are indicative of the levels of
those skilled in the art to which the present invention pertains. All patents and publications are
herein incorporated by reference to the same extent as if each individual publication was
specifically and individually indicated to be incorporated by reference.

[309] Having described certain embodiments of the present invention, it will appear to those ordinarily skilled in the art that various modifications may be made to the disclosed embodiments, and that such modifications are intended to be within the scope of the present invention.

[310] Those of skill in the art readily appreciate that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those inherent therein. The compositions and methods described herein are representative embodiments, are exemplary, and are not intended as limitations on the scope of the invention. It is readily apparent to one skilled in the art that varying substitutions and modifications may be made to the invention disclosed herein without departing from the scope and spirit of the invention.

[311] The invention illustratively described herein suitably may be practiced in the absence of any element or elements, limitation or limitations which is not specifically disclosed herein. The terms and expressions which have been employed are used as terms of description and not of limitation, and there is no intention that in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by certain embodiments and optional features, modification and variation of the concepts herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention.

[312] The invention has been described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the generic disclosure also form part of the invention. This includes the generic description of the invention with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein.

CLAIMS

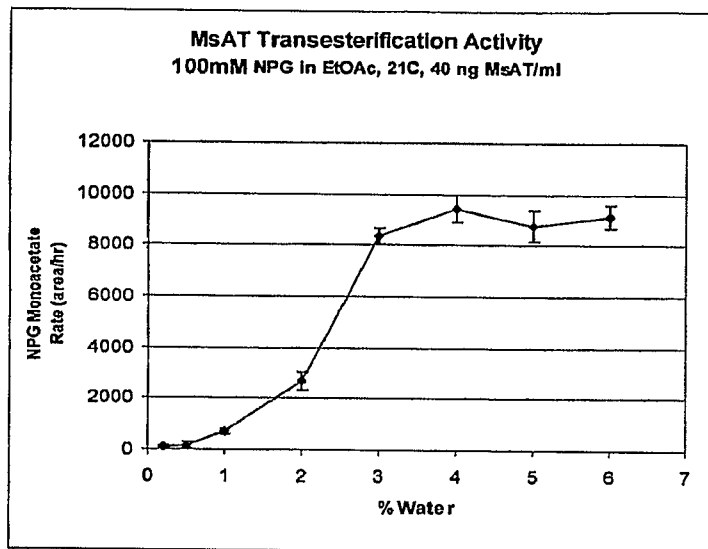
We claim:

1. A polypeptide comprising a sequence selected from the group consisting of SEQ ID NOs: 11-17, 19-27, 29-40, and 42-48, wherein said polypeptide has an enzymatic aqueous acylation and/or perhydrolysis activity.
2. A polynucleotide encoding the polypeptide of claim 1.
3. A polynucleotide according to claim 2, wherein said polynucleotide is comprised within an expression vector.
4. A host cell comprising the polynucleotide of claim 3.
5. A method for producing a polypeptide having enzymatic aqueous acylation and/or perhydrolysis activity, comprising growing the host cell of claim 4 under conditions suitable for expression of the polypeptide.
6. A method for identification of polypeptides having enzymatic aqueous acylation and/or perhydrolysis activity, said method comprising:
 - a. preparing a multiple alignment of the amino acid sequences of thioesterase from *E. coli* (Genbank accession no. YP_668488), acylesterase from *Agrobacterium tumefaciens* (Genbank accession no. AAD02335), rhamnogalacturonan acetylerase from *Aspergillus aculeatus* (Genbank accession no. IDEO_A.), acylesterase from *Sinorhizobium meliloti* (Genbank accession no. np_436338), and arylesterase from *Mycobacterium smegmatis* (Genbank accession no. YP_890535);
 - b. adjusting the alignments for *M. smegmatis* arylesterase, *E. coli* thioesterase, and *A. aculeatus* rhamnogalacturonan acetylerase, wherein the sequences are

aligned with regard to residues that share structurally homologous main chain atoms based on x-ray crystallographic models;

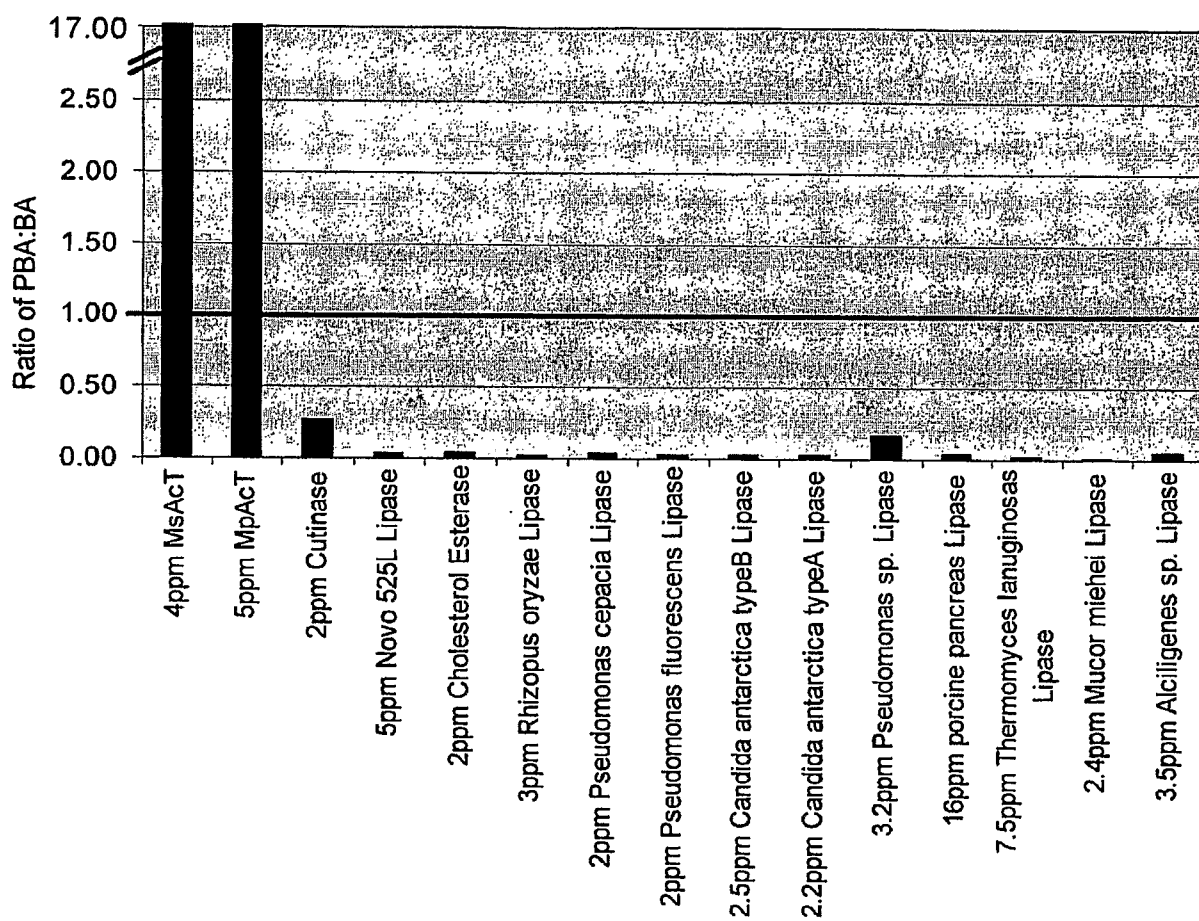
- c. adjusting the alignments of the *S. meliloti* and *A. tumefaciens* enzymes based on sequence homology;
- d. constructing a Hidden Markov Model (HMM); and
- e. using the HMM to search a protein sequence database for polypeptides having enzymatic aqueous acylation and/or hydrolysis activity.

1/11

FIGURE 1

2/11

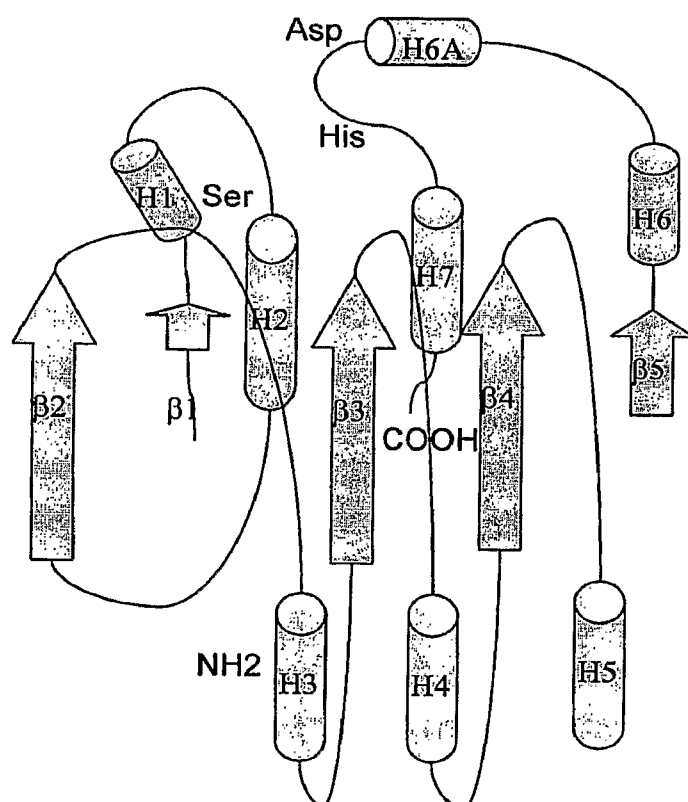
FIGURE 2



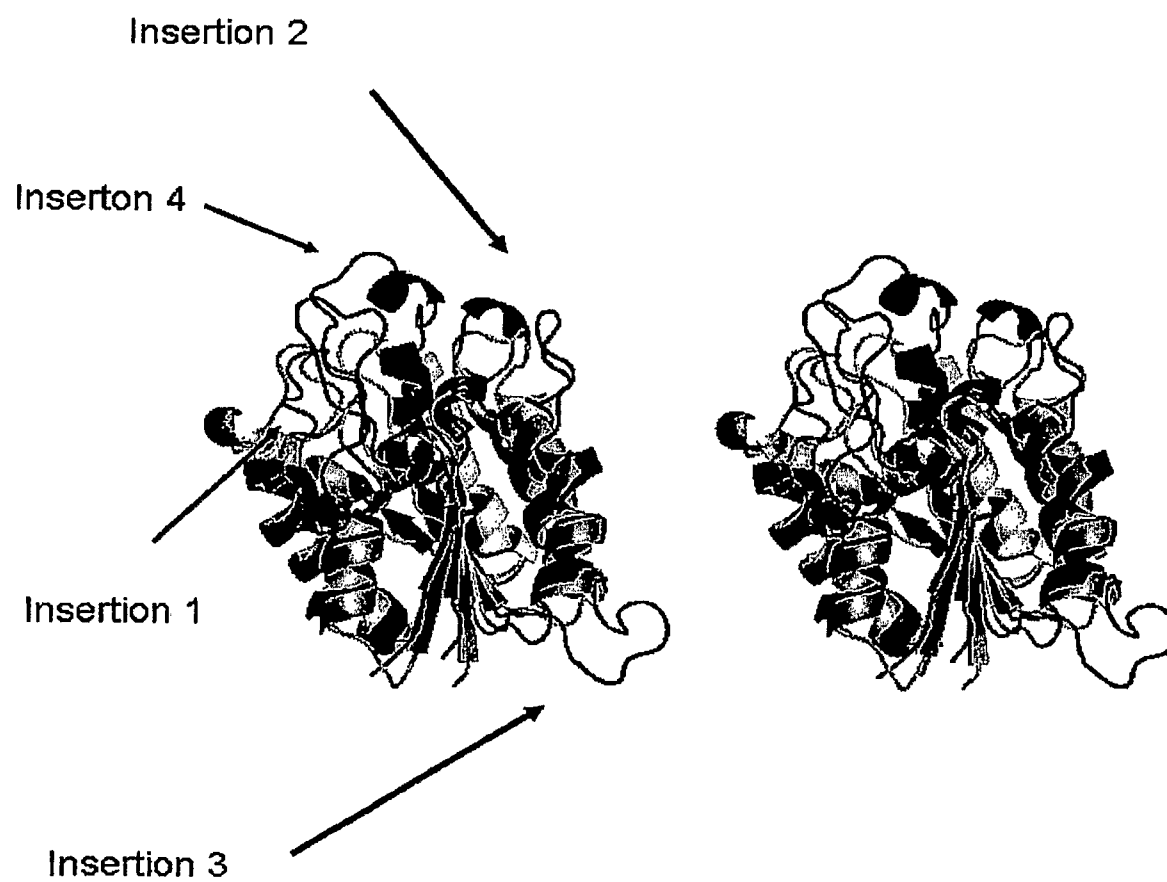
3/11

FIGURE 3



FIGURE 4

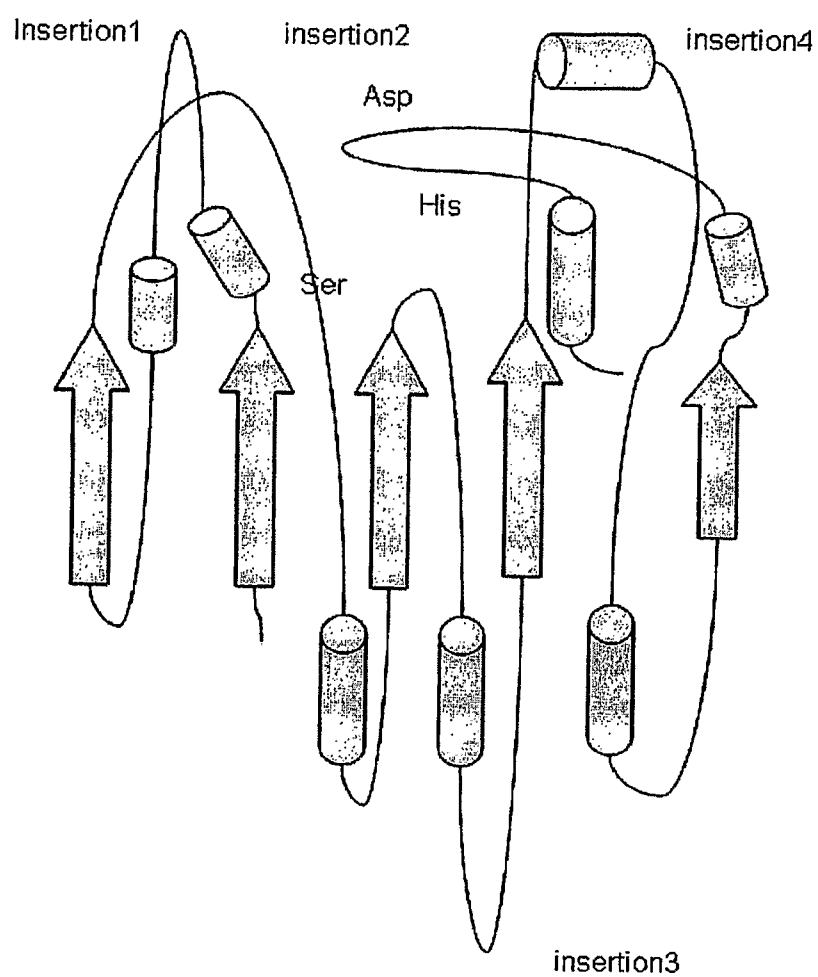
5/11

FIGURE 5

6/11

FIGURE 6

ACT Hydrolase fold



7/11

FIGURE 7A

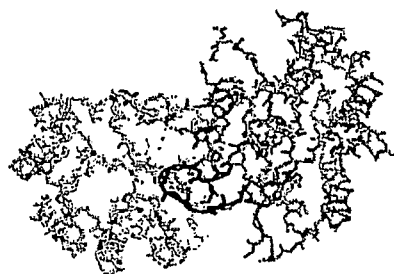
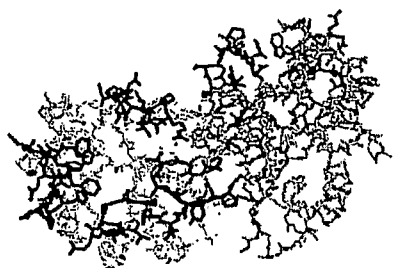


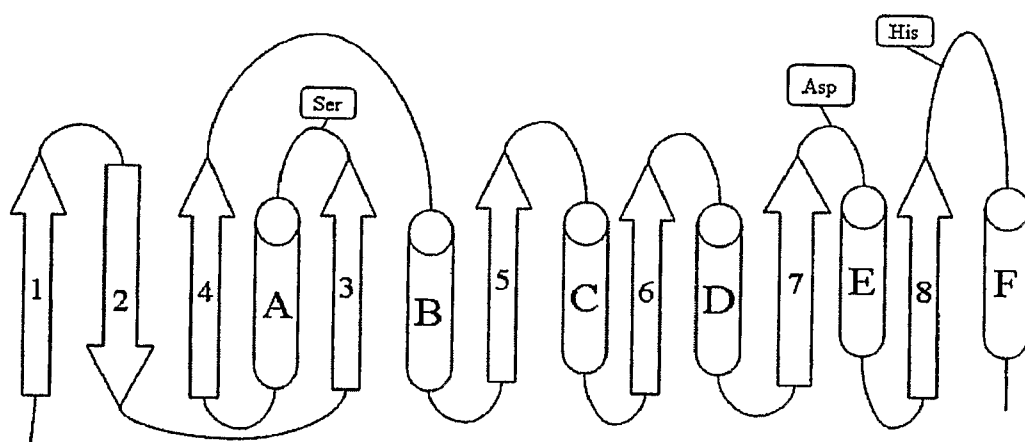
FIGURE 7B



8/11

FIGURE 8

Basic α/β Hydrolase Fold

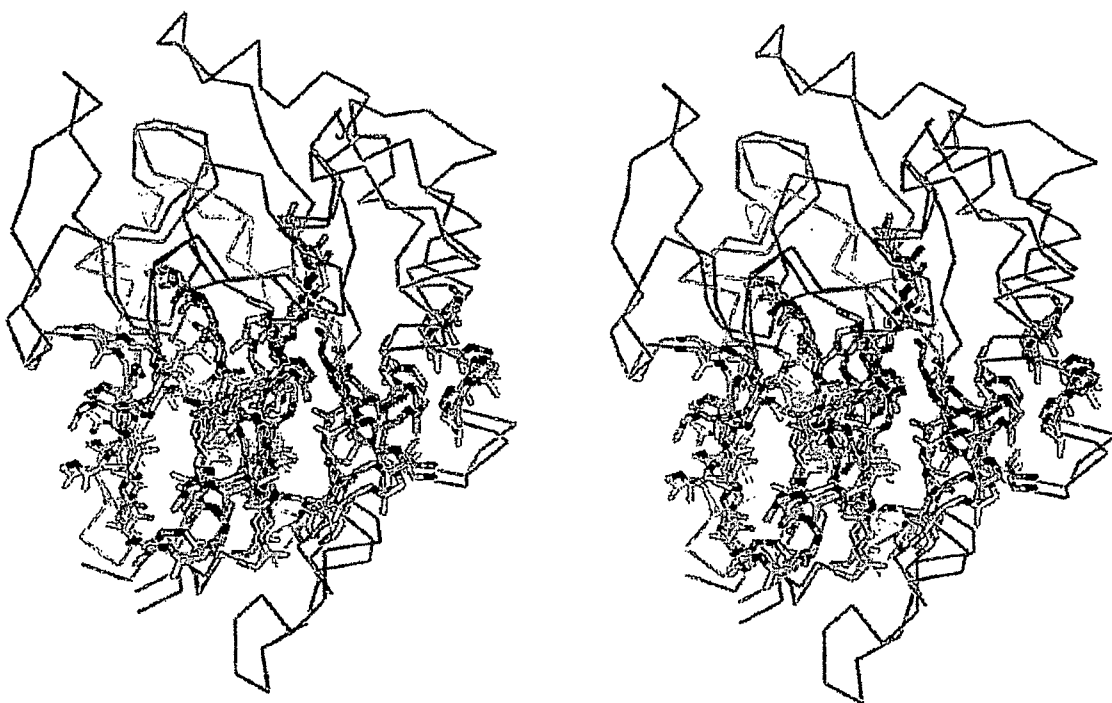


Insertion 1		Insertion 2	
M_smeigmatisAT	M..AKRILCF	GDSLTWGVP	VEDGAPTERF
S_melloti_RSM2162	MVEKRSVLCF	GDSLTWGWP	VKESPTLRM
A_radiobacterACA	MV..KSVLCF	GDSLTWGSDA	ET....GGRH
E_ColiTAP	...ADTLIL	GDLSAGY..	RM
A_Aculeatus_1deo	TTVYLA	GDSTMAK...	NGGG
		APDVRWTGVL	AQQLGADFEV
		IEEGLSARTT	IEEGLSARTT
		IEEGLSARTT	IEEGLSARTT
		IHEGLGGRIT	IHEGLGGRIT
		VNASISGDTIS	VNASISGDTIS
		ASYL...SATV	VNDAVAGRS.
		SGTNGWGEYL	ASYL...SATV
		PLDIALGMSV	LVTOVLTSAG
		PYEIANGMGK	LVGQVLTCAG
		AIIVAMKGVVER	LVKLVRNHVW
		PQQTEQ	TLRQILQDV.
		ILTFPAYLEN	AAKLFTA...
		YSVYDGVNET	KGAK
		DCSGTGAECV	SVISTDG
		LSGLYKALAD	FMKVEFFAAG
		LAPVYRDLAD	DLDCGFFDAG
		FSAIYPKIAK	EFDVPLLPFF
		FVEYAEALAE	VAGVEYVDHW
		VNSPTR....	SVYDSIYETL
		SVYDSIYETL	GNATVNSYFP
		IDHTHTSPAG	IDHTHTSPAG
Insertion 3		Insertion 4	
M_smeigmatisAT	LATHLP.LDL	VIIMLGNDT	KAYFRR....
S_melloti_RSM2162	LASHLP.LDL	VIIMLGNDT	KSYFHR....
A_radiobacterACA	LHSHAP.LDL	VIIMLGNDL	KPSIHG....
E_ColiTAP	LKQHQP...RW	VLVELGGNDG	LRGFQ....
A_Aculeatus_1deo	IADVVTAGDY	VIVEFGHNDG	GSLSTDNGRT
		FMKVFFFDAG	SVISTDG
		FMKVEFFAAG	DCISTDG
		DLDCGFFDAG	SVARTTP
		EFDVPLLPFF	MEEVY...
		VAGVEYVDHW	LKPQWMO
		SVYDSIYETL	DDGIHPNRDA
		GNATVNSYFP	IDHTHTSPAG
Insertion 4		Insertion 5	
M_smeigmatisAT	VLVVSPPPLA	PMPHPWFQLI	FEGGEQKTE
S_melloti_RSM2162	VLVAVAPPPLA	PMPDPWFEGM	FGGGYEKSKE
A_radiobacterACA	VLIVAPPQLC	ETANPVMGAI	FRDAIDESAM
E_ColiTAP	AEPLLMQIRL	P.....	ANVGRRYNEA
A_Aculeatus_1deo	VILSSQTPNN	PWETGTF...	VNSPTR....
		FMKVFFFDAG	SVISTDG
		FMKVEFFAAG	DCISTDG
		DLDCGFFDAG	SVARTTP
		EFDVPLLPFF	MEEVY...
		VAGVEYVDHW	LKPQWMO
		SVYDSIYETL	DDGIHPNRDA
		GNATVNSYFP	IDHTHTSPAG
		SVYDSIYETL	GNATVNSYFP
		IDHTHTSPAG	IDHTHTSPAG
M_smeigmatisAT	NRDLGVALAE	QVRSLL....	
S_melloti_RSM2162	NIRLGHAIAD	KVAALF....	
A_radiobacterACA	TRAIGRGLEP	VVRMMLGL..	
E_ColiTAP	QPFIAQWMAK	QLQPLVNHDS	
A_Aculeatus_1deo	AEVVAEAFIK	AVVCTGTSLK	SVLTITSFEG
		SVLTITSFEG	TCL

FIGURE 9

10/11

FIGURE 10



11/11

FIGURE 11

