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

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 December 2008 (31.12.2008)

PCT

(10) International Publication Number
WO 2009/002480 A9

(51) International Patent Classification:
*C07K 14/435* (2006.01)

(21) International Application Number:
PCT/US2008/007807

(22) International Filing Date: 23 June 2008 (23.06.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/937,417 26 June 2007 (26.06.2007) 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).

(74) Agent: **JACOBSON, Jill A.**; 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, AO, 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),

*[Continued on next page]*

(54) Title: ACYL TRANSFERASE HAVING ALTERED SUBSTRATE SPECIFICITY

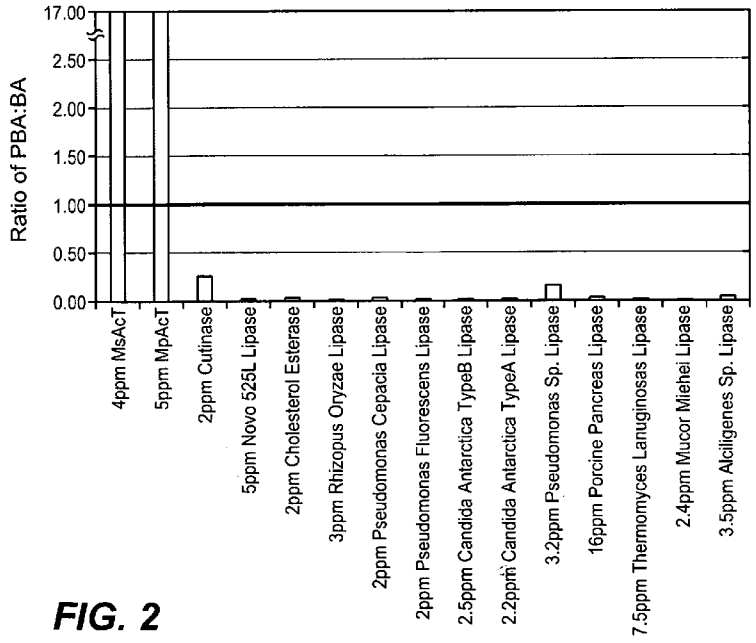


FIG. 2

(57) Abstract: The present invention provides methods for engineering the M. smegmatis acyltransferase 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.

WO 2009/002480 A9

European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, 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:**
— *with international search report*

**(88) Date of publication of the international search report:**
19 February 2009

**(48) Date of publication of this corrected version:**
9 April 2009

**(15) Information about Correction:**
see Notice of 9 April 2009

# ACYL TRANSFERASE HAVING ALTERED SUBSTRATE SPECIFICITY

## CROSS-REFERENCE TO RELATED APPLICATIONS

[01] This application claims the benefit of U.S. Provisional Application No. 60/937,417, filed on June 26, 2007, which is incorporated by reference herein in its entirety.

## BACKGROUND OF THE INVENTION

[02] 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

[03] The present invention provides methods for engineering the *M. smegmatis* acyltransferase 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. The subject enzymes may be employed in a variety of applications, including, for example, in detergents, textile manufacturing, in foodstuffs, and for the manufacture of chemicals.

[04] 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 particularly preferred 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.

**[05]** 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 Markhov 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.

**[06]** Certain aspects of the invention described herein relate to segments of SGNH and α/β hydrolases which, in the tertiary structure, determine the substrate specificity of those enzymes.

**[07]** In certain cases, altering the length of a particular segment of a wild-type or variant *M. smegmatis* acyltransferase - the "substrate chain length specificity determining segment" - can increase or decrease the length of acyl chain that can be transferred by the enzyme. Alterations in the length of the substrate chain length specificity determining segment, alone or in combination with alterations of the other segments of the perhydrolase, can alter the overall dimensions and characteristics of the substrate binding surface of the enzyme. Such alterations alter the substrate specificity of those enzymes.

**[08]** The present invention also provides methods useful in obtaining optimum loop conformation, including but not limited to mutagenesis of the preceding and following approximately two residues that bracket the insertion, along with the amino acid sequence of the insertion itself.

**[09]** In these embodiments and in general terms, the length of the substrate chain length specificity determining segment of a parental *M. smegmatis* acyltransferase may be increased to provide an altered enzyme that has specificity for longer substrates. In other embodiments, the length of the substrate chain length specificity determining segment of a parental perhydrolase may be decreased to provide an altered enzyme that has specificity for shorter substrates.

## DESCRIPTION OF THE FIGURES

**[10]** This patent or application file contains at least one drawing executed in color. Copies of this patent application publication with color drawing(s) will be provided by the U.S. Patent and Trademark Office upon request and payment of the necessary fee.

**[11]** 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.

**[12]** 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.

**[13]** 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.

**[14]** 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.*, β1) and the catalytic Asp and His are found in a loop found between helices 6 and 7.

**[15]** 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.

[16] 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 SGHN hydrolase fold.

[17] 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.

[18] 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.

[19] 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.

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

[21] Figure 10 illustrates the tertiary structures of two representative SGNH hydrolases.

## DESCRIPTION OF THE INVENTION

[22] The present invention provides methods for engineering the *M. smegmatis* acyltransferase, 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.

[23] 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 particularly preferred 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.

**[24]** 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**

**[25]** 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. Although any methods and materials similar or equivalent to those described herein find use in the practice of the present invention, the preferred methods and materials are described herein. 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.

**[26]** 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.

**[27]** 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.

**[28]** The first conserved block of sequence is found at the C-terminal region of the first β strand and includes the sequence GDS, where 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 (β5 in Figure 1) and preceding a conserved helical segment (H7); this segment includes the Asp and His of the catalytic triad.

**[29]** 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 acetylesterase (RGAE, PDB code: 1deo), and at least one hypothetical protein (PDB code: 1vjg), as identified using a structural search engine (*e.g.*, DALI). SGNH hydrolase-type esterase domains contain a unique hydrogen bond network that stabilizes their catalytic centers and a conserved Ser/Asp/His catalytic triad. SGNH acyltransferases are also described in accession number cd01839.3 in Genbank's conserved domain database, which database entry is incorporated by reference.

**[30]** As used herein, the term "α/β hydrolase" refers to a hydrolase having a canonical α/β fold structure, as described in Ollis et al (Protein Eng. 1992 5:197-211), Nardini et al (Curr. Opin. Struct. Biol. 1999 9 732-737) and Schrag et al (Methods Enzymol. 1997 284: 85-107). Although structurally related, SGNH hydrolases and α/β hydrolases are distinct families of hydrolases.

**[31]** As used herein, the term "hydrolase" is used to generically describe SGNH hydrolases and α/β hydrolases. In particular embodiments, a subject hydrolase catalyzes acyl transfer from an acyl-containing substrate to a suitable recipient molecule in an aqueous environment. A subject hydrolase may catalyze transfer of an acyl group from an acyl-containing substrate (e.g.,

an acyl ester) onto a recipient hydrogen (via hydrolysis), peroxyl (via perhydrolysis to produce a peracid), or alcohol (via alcoholysis to produce an ester), for example.

**[32]** As used herein, the term "acyl substrate" refers to the molecule that donates an acyl group in an acyltransferase reaction.

**[33]** As used herein, the term "substrate chain length specificity determining segment" refers to a sequence of contiguous amino acids that is positionally conserved in the tertiary structures of SGNH and α/β hydrolases and determines the length of acyl chain that can be transferred by the hydrolase. The substrate chain length specificity determining segment of an SGNH hydrolase is a region of contiguous amino acids that lies between the β5 β-strand of the enzyme and the Asp residue of the catalytic triad of that enzyme. The substrate chain length specificity determining segment of an α/β hydrolase lies in an equivalent position to the substrate chain length specificity determining segment of an SGNH hydrolase, and is present between the β-strand 6 and α-helix D of that enzyme, using the structural definitions shown in Fig. 8 and discussed in Nardini et al (Curr. Opin. Struct. Biol. 1999 9 732-737) and Schrag et al (Methods Enzymol. 1997 284: 85-107).

**[34]** As used herein, the term "aqueous", as used in the phrases "aqueous composition" and "aqueous environment" refers to a composition that is made up of at least 50% water. An aqueous composition may contain at least 50% water, e.g., at least 60% water, at least 70% water, at least 80% water, at least 90% water, at least 95% water, at least 97% or at least 99% water. In certain cases, some of the remainder of an aqueous composition may be an alcohol.

**[35]** In particular embodiments, the term "aqueous", refers to a composition having a water activity ($A_w$) of at least 0.75, e.g., at least 0.8, at least 0.9 or at least 0.95 as compared to distilled water.

**[36]** As used herein, the term "parental hydrolase" is a hydrolase that is the target of amino acid changes. In certain embodiments, amino acids, e.g., a substrate chain length specificity determining segment, may be "donated" by a "donor" hydrolase to the parent hydrolase to produce an altered hydrolase. A parental hydrolase may be a wild-type hydrolase or a variant of a wild-type hydrolase that has acyltransferase activity.

**[37]** As used herein, the term "altered hydrolase" is a hydrolase having an altered amino acid sequence relative to a parental hydrolase. As will be described in greater detail below, an altered hydrolase may have a substrate chain length specificity determining segment that is a different length to the substrate chain length specificity determining segment of a parental hydrolase.

[38] As used herein, the term "altered substrate specificity" a difference in the acyl chains that can be transferred by two hydrolases (e.g., a parental hydrolase and an altered hydrolase). In certain case, the range of lengths of acyl chains that can be transferred by two hydrolases, e.g., a parental and an altered hydrolase may be different. In certain cases, the optimum acyl chain transferred by two hydrolases, e.g., a parental and an altered hydrolase, may be different.

[39] As used herein, a "long acyl chain substrate" is an acyl substrate containing a chain of at least 6 carbon atoms that is transferred by a subject hydrolase.

[40] As used herein, a "short acyl chain substrate" is an acyl substrate containing a chain of less than 6 carbon atoms that is transferred by a subject hydrolase.

[41] 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 $ClO_2$, $H_2O_2$, peracids, $NO_2$, etc.

[42] 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.

[43] 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. In particularly preferred embodiments, 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. In additional preferred embodiments, the perhydrolases of the present invention are characterized by having distinct tertiary structure and primary sequence. In particularly preferred embodiments, the perhydrolases of the present invention comprises distinct primary and tertiary structures. In some particularly preferred embodiments, the perhydrolases of the present invention comprise distinct quaternary structure. In some preferred embodiments, the perhydrolase of the present invention is the *M. smegmatis* perhydrolase, while in alternative embodiments, the perhydrolase is a variant of this perhydrolase, while in still further embodiments, the perhydrolase is a homolog of this perhydrolase. In further preferred

embodiments, a monomeric hydrolase is engineered to produce a multimeric enzyme that has better perhydrolase activity than the monomer. 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.

**[44]** 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.

**[45]** 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 preferred embodiments, the assays provided herein are used to determine the amounts of peracid and acid produced by the enzyme.

**[46]** 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 particularly preferred embodiments, these products are utilized on humans, while in other embodiments, these products find use with non-human animals (*e.g.*, in veterinary applications).

**[47]** 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.

**[48]** 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.

[49] 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-spotters, as well as dish detergents).

[50] 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.

[51] 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 preferred 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.

[52] 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.

[53] 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.

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

[55] 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.

[56] 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.

[57] 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).

[58] 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.

[59] 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.

[60] 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.

[61] 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.

**[62]** 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).

**[63]** 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.

**[64]** 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.

**[65]** 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, $CH_3CO$-Cl, is the acyl chloride formed from ethanoic acid, $CH_3COO$-H). The names of the individual acyl groups are formed by replacing the "-ic" of the acid by "-yl."

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

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

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

[69] 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.

[70] 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.

[71] 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.

[72] 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 $H_2O_2$ 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.

[73] 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.

[74] 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.

[75] 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.

[76] 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.

[77] 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.

**[78]** 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.

**[79]** As used herein, the phrase "is independently 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:

**[80]** 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).

**[81]** 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, alkyldienyl, 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.

**[82]** 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.

**[83]** 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.

**[84]** 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.

**[85]** 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 "protein," "peptide" and polypeptide are used interchangeably herein. Wherein a peptide is a portion of a protein, those skilled in the art understand the use of the term in context.

**[86]** 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. In most particularly preferred embodiments, the related proteins of the present invention very high ratios of perhydrolysis to hydrolysis.

**[87]** 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.

**[88]** Related (and derivative) proteins comprise "variant proteins." In some preferred 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 preferred embodiments, the number of different amino acids between variants is between 1 and 10. In some particularly preferred 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.

**[89]** 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.

**[90]** In particularly preferred embodiments, homologous proteins are engineered to produce enzymes with the desired activity(ies). In some particularly preferred embodiments, the engineered proteins are included within the SGNH-hydrolase family of proteins. In some most preferred embodiments, the engineered proteins comprise at least one or a combination of the following conserved residues: 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 preferred embodiments, the engineered proteins exhibit a perhydrolysis to hydrolysis ratio that is greater than 1.

**[91]** 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).

**[92]** 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 preferred embodiments, alignment of conserved residues conserves 100% of such residues. However, alignment of greater than 75% or as little as 50% of conserved residues is also adequate to define equivalent residues. In preferred 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).

**[93]** 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 sequence that are modified at these positions. Indeed, it is intended that the present invention encompass various modifications and combinations of modifications.

**[94]** 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 with 0.13 nm of the corresponding side chain atoms of *M. smegmatis* perhydrolase. The coordinates of the three dimensional structure of *M. smegmatis* perhydrolase were determined and are set forth in US Pat. Appln. Ser. No. 10/526,764, hereby incorporated by reference in its entirety, find use in determining equivalent residues on the level of tertiary structure.

**[95]** 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 nucleic acid that comprises a signal sequence. In some embodiments of 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.

**[96]** 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.

**[97]** 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.

**[98]** In some preferred 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.

**[99]** 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.

**[100]** 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.

**[101]** 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.

**[102]** 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.

**[103]** As used herein, a "corresponding region" that is present in two proteins refers to a region that is present at an analogous position in the two proteins. In the hydrolases described herein, corresponding regions are regions found in connecting segments between common secondary structural elements (β strands and α helicies) in the tertiary structure of aligned SGNH hydrolases, as shown in the figures herein. Common secondary elements are those that are seen to overlap when the tertiary structures of two SGNH hydrolases are aligned. Overlapping elements are those in which a consecutive series of alpha carbon atoms are found to share a

common small (<2 Ang) pairwise displacement (usually expressed as an rms deviation). In certain cases, a "corresponding region" may be referred to as a "homologous region".

**[104]** Alignment of two SGNH hydrolases can be accomplished using commercial molecular modeling software program such as MOE supplied by Chemical Computing Group in Montreal Canada. If, the tertiary structure has been determined, the program Moe can be used to perform a structure based alignment including constraints to align the serine residues found in the DGS(L) block sequence common to all SGNH hydrolase along with the Asp and His residues found in the DXXH block sequence that is also common to all SGNH hydrolases. If the tertiary structure has not been determined, then the program can perform a sequence based alignment to a known SGNH hydrolase structure having suitable sequence homology (at least 60% homology, preferably at least 70% and most preferably >80% homology) using known methods. The E. coli thioesterase (PDB entry code 1ivn) may be used for comparisons.

**[105]** 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.

**[106]** 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.*, a 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 preferred 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.

**[107]** 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."

**[108]** 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.

**[109]** 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.

**[110]** 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.

**[111]** 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]).

**[112]** 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, M'5, N'-4, and a comparison of both strands.

**[113]** 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.

**[114]** 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.

**[115]** 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 (Tm) of the nucleic acid binding complex or probe. For example, "maximum stringency" typically occurs at about Tm-5° C (5° below the Tm of the probe); "high stringency" at about 5-10° below the Tm; "intermediate stringency" at about 10-20° below the Tm of the probe; and "low stringency" at about 20-25° below the Tm. 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, 6xSSC = very low stringency; 3xSSC = low to medium stringency; 1xSSC = medium stringency; and 0.5xSSC = 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.

**[116]** 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).

**[117]** 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% identity, more preferable at least about 50% identity, yet more preferably at least about 60% identity, preferably at least about 75% identity, more preferably at least about 80% identity, yet more preferably at least about 90%, still more preferably about 95%, most preferably about 97% identity, sometimes as much as about 98%

and about 99% sequence identity, compared to the 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).

**[118]** As used herein, "equivalent residues" refers to proteins that share particular amino acid residues. For example, equivalent resides 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.

**[119]** 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 preferred embodiments, these parental proteins are homologs of one another. For example, in some embodiments, a preferred hybrid perhydrolase or fusion protein contains the N-terminus of a protein and the C-terminus of a homolog of the protein. In some preferred embodiment, the two terminal ends are combined to correspond to the full-length active protein.

**[120]** 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.

**[121]** 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.

**[122]** 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.

**[123]** 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.

**[124]** 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.

**[125]** 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.

**[126]** "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.

**[127]** 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.

**[128]** 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.

**[129]** 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."

**[130]** 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.

**[131]** "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 Qβ 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.

**[132]** 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.

**[133]** 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.

**[134]** 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.

**[135]** 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".

**[136]** 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.).

**[137]** 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.

**[138]** 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.

**[139]** 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.

## DETAILED DESCRIPTION OF THE INVENTION

**[140]** The present invention provides methods for engineering the *M. smegmatis* acyltransferase enzyme. In certain embodiments, a composition comprising at least one enzyme suitable for use in enzymatic aqueous acylation and/or perhydrolysis is provided. The present invention further provides compositions comprising at least one perhydrolase enzyme suitable for use in enzymatic aqueous acylation and/or perhydrolysis. As will be described in greater detail below, a subject enzyme may be employed in a variety of applications, including, but not limited to: in cleaning compositions (e.g., detergents), in textile manufacture, in food products, and in the manufacture chemicals, for example.

**[141]** 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.

**[142]** *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.

**[143]** 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 (Figure1A). Further it was found that the acylation reaction was still fully active (Kcat=$10^5$) in aqueous reactions containing just millimolar concentrations of ethyl acetate.

**[144]** *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*, $MC_2155$. 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.*"

**[145]** 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.

**[146]** 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.

**[147]** MsAcT represents an enzyme with two very desirable characteristics, including being able to perform both alcholysis 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.

**[148]** 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. radiobacter* 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 acetylesterase, did not show any activity in the acyltransferase or perhydrolysis assays, as indicated in Table 1, below.

**Table 1. Characterization of Acyltransferase and Perhydrolase Activity of MsAcT and Related Enzymes**

| Enzyme | Acyltransferase Activity | Perhydrolase Activity | Soluble Aggregate Form | Sequence Identity (%) | Structure |
|---|---|---|---|---|---|
| MsAcT | ++++ | ++++ | Octamer | 100 | SGNH hydrolase |
| *S. meliloti* RSM02162 | n.d. | ++ | Apparent octamer | 63.3 | Unknown |
| *A. radiobacter* 7-ACA | + | + | Tetramer | 42.5 | Unknown |
| *E. coli* thioesterase | - | - | Monomer | 14 | SGNH hydrolase |

| Table 1. Characterization of Acyltransferase and Perhydrolase Activity of MsAcT and Related Enzymes | | | | | |
|---|---|---|---|---|---|
| **Enzyme** | **Acyltransferase Activity** | **Perhydrolase Activity** | **Soluble Aggregate Form** | **Sequence Identity (%)** | **Structure** |
| *A. aculeatus* rhamnogalacturonan acetylesterase | - | - | Monomer | 13.8 | SGNH hydrolase |

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

**[150]** 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.

**[151]** 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**

| Space group | P4 | | | |
|---|---|---|---|---|
| Unit cell parameters | a = 98.163 Å, b = 98.163 Å, c = 229.896 Å, α=β=γ=90° | | | |
| Data Collection | $\lambda_0$MADSe | $\lambda_1$MADSe | $\lambda_2$MADSe | $\lambda_3$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/□(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 | $\lambda_0$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.

$R_{sym} = \Sigma|I_i-\langle I_i\rangle| / \Sigma|I_i|$ where $I_i$ is the scaled intensity of the $i^{th}$ measurement, and $\langle I_i\rangle$ is the mean intensity for that reflection.

$R_{cryst} = \Sigma| |F_{obs}|-|F_{calc}| | / \Sigma|F_{obs}|$ 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**

| | | | |
|---|---|---|---|
| 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 | $\lambda_0$ 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 | $\lambda_0$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.
$R_{sym} = \Sigma|I_i - \langle I_i \rangle| / \Sigma|I_i|$ where $I_i$ is the scaled intensity of the $i^{th}$ measurement, and $\langle I_i \rangle$ is the mean intensity for that reflection.
$R_{cryst} = \Sigma| |F_{obs}| - |F_{calc}| | / \Sigma|F_{obs}|$ 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.

**[152]** 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 acetylesterase (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.

**[153]** 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.

**[154]** 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™.

**[155]** 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.

**[156]** 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.

**[157]** 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* _ RSM02162 and *A. radiobacter* 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.

**[158]** 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.

**[159]** 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 β2 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.

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

***Methods for altering substrate specificity***

**[161]** As noted above, a method for altering the substrate specificity of the *M. smegmatis* acyltransferase, is provided. In certain embodiments, the method includes altering the length of a substrate chain length specificity determining segment that lies immediately N-terminal to the Asp residue of the catalytic triad of a parent SGNH hydrolase that is a wild-type or enzymatically active variant of the *M. smegmatis* acyltransferase to produce an altered SGNH hydrolase that has an altered substrate specificity relative to the parent SGNH hydrolase.

**[162]** The method may also include altering other segments of a hydrolase, e.g., those segments described elsewhere in this disclosure, to alter the substrate specificity of the hydrolase.

*Substrate chain length specificity determining segments*

**[163]** The substrate chain length specificity determining segment of the *M. smegmatis* acyltransferase is a region of contiguous amino acids that lies between the β5 β-strand of the

enzyme and the Asp residue of the catalytic triad of that enzyme (the Asp residue being part of the Asp-Xaa-Xaa-His motif found in all SGNH hydrolases).

**[164]** The tertiary structures of two exemplary SGNH hydrolases: the *M smegmatis* MsACT (deposited as NCBI's Genbank database as accession number YP_890535; GID: 118468600, as also described in WO05/056782), and the E. coli thioesterase (deposited as NCBI's Genbank database as accession number 1IVN_A; GID:33357066), each showing a signature three-layer alpha/beta/alpha structure, where the beta-sheets are composed of five parallel strands, is known. Akoh et al (Prog. Lipid Res. 2004 43: 534-552), Wei et al (Nat. Struct. Biol. 1995 2: 218-223) and Lo (J. Mol. Biol. 2003 330:539-51) provide a detailed discussion of the structure of SGNH hydrolases. For reference, the substrate chain length specificity determining segments of each of the enzyme structures are shown on Fig. 10.

**[165]** The substrate chain length specificity determining segment of the *M. smegmatis* acyltransferase lies immediately N-terminal to the Asp residue of the catalytic triad of the enzyme. However, the length of the substrate chain length specificity determining segment may vary according to the distance between the Asp residue and the β5 β-strand of the enzyme. For example, the substrate chain length specificity determining segments of MsACT and the *E. coli* thioesterase are about 13 amino, and about 19 amino acids in length, respectively. As such, depending on the hydrolase, a substrate chain length specificity determining segment may be in the range of 10 to 70 amino acids in length, e.g., in the range of 10 to 30 amino acids in length, 30 to 50 amino acids in length, or 50 to 70 amino acids.

**[166]** Table 7 below provides exemplary sequences for the substrate chain length specificity determining segment of a representative selection of SGNH hydrolases.

[167] Table 7:

| | |
|---|---|
| *M. smegmatis* MsAcT | |
| PFFDAGSVISTDGV | SEQ ID NO:10 |
| *S. meliloti* RSM02162 | |
| EFFAAGDCISTDVI | SEQ ID NO:11 |
| *A. radiobacter* 7-ACA | |
| GFFDAGSVARTTPV | SEQ ID NO:12 |
| *E. coli* thioesterase | |
| PLLPFFMEEVLYKPQWMQD | SEQ ID NO:13 |
| *A. aculeatus* rhamno-galacturonan acetylesterase | |
| EYVDHWSYVDSIYETLGNATVNSYFPI | SEQ ID NO:14 |

[168] The substrate chain length specificity determining segment of an α/β hydrolase lies in an equivalent position to the substrate chain length specificity determining segment of an SGNH hydrolase when the structures of the proteins are superimposed using their common elements of secondary structure. The substrate chain length specificity determining segment of an α/β hydrolase lies between the β-strand 6 and α-helix D of that enzyme, using the structural definitions shown in Fig. 8 and described in Nardini et al (Curr. Opin. Struct. Biol. 1999 9 732-737) and Schrag et al (Methods Enzymol. 1997) 284: 85-107.

[169] In certain embodiments, the amino acid sequence of a substrate chain length specificity determining segment may or may not be the amino acid sequence of a wild-type *M. smegmatis* acyltransferase. In certain embodiments, the substrate chain length specificity determining segment may have an amino acid sequence that is at least 70%, e.g., at least 80%, at least 90% or at least 95% identical to the substrate chain length specificity determining segment of a wild type hydrolase.

*Parental hydrolases*

[170] A parental hydrolase may be any SGNH or α/β hydrolase, which, in particular embodiments, may be a wild type *M. smegmatis* acyltransferase or a catalytically active variant thereof. Exemplary parental SGNH acyltransferase hydrolases that may be employed in the methods described herein include the wild-type SGNH acyltransferases deposited in NCBI's

Genbank database as accession numbers: YP_890535 (GID: 118468600 as also described in WO05/056782; *M. smegmatis*), NP_436338.1 (GID: 16263545; *Sinorhizobium meliloti*); ZP_01549788.1 (GID: 118592396; *Stappia aggregate*); NP_066659.1 (GID: 10954724; *Agrobacterium rhizogenes*); YP_368715.1 (GID: 78065946; *Burkholderia sp.*); YP_674187.1 (GID: 110633979; *Mesorhizobium sp.*), NP_532123.1 (GID: 17935333; *Agrobacterium tumefaciens*), *Agrobacterium rhizogenes* (Q9KWA6), *A. rhizogenes* (Q9KWB1), *A. tumefaciens* (Q8UFG4), *A. tumefaciens* (Q8UAC0), *A. tumefaciens* (Q9ZI09), *A. tumefaciens* (ACA), *Prosthecobacter dejongeii* (RVM04532), *Rhizobium. loti* (Q98MY5), *R. meliloti* (Q92XZ1), *R. meliloti* (Q9EV56), *R. rhizogenes* (NF006), *R. rhizogenes* (NF00602875), *R. solanacerarum* (Q8XQI0), *Sinorhizobium meliloti* (RSM02162), *S. meliloti* (RSM05666), *Mesorhizobium loti* (RMLO00301), *A. rhizogenes* (Q9KWA6), *A. rhizogenes* (Q9KWB1), *Agrobacterium tumefaciens* (AAD02335), *Mesorhizobium loti* (Q98MY5), *Mesorhizobium loti* (ZP00197751), *Ralstonia solanacearum* (Q8XQI0), *Ralstonia eutropha* (ZP00166901), *Moraxella bovis* (AAK53448), *Burkholderia cepacia* (ZP00216984), *Chromobacterium violaceum* (Q7NRP5), *Pirellula sp.* (NP_865746), *Vibrio vulnificus* (AA007232), *Salmonella typhimurium* (AAC38796), *Sinorhizobium meliloti* (SMa1993), *Sinorhizobium meliloti* (Q92XZ1) and *Sinorhizobium meliloti* (Q9EV56), and wild-type orthologs and homologs thereof, and variants thereof that have an amino acid sequence that is at least 70% identical, at least 80% identical, at least 90% identical, at least 95% identical, or at least at least 98% identical to any of those wild-type enzymes. These Genbank accessions are incorporated by reference in their entirety, including the nucleic acid and protein sequences therein and the annotation of those sequences, as of the earliest filing date of this patent application. Further examples of such enzymes may be obtained by performing sequence homology-based searches of NCBI's Genbank database using standard sequence comparison methods, e.g., BLAST, etc. In particular embodiments, the acyltransferase used has an amino acid sequence that is at least 70% identical to the amino acid sequence set forth in Genbank entry YP_890535 (GID: 11846860 as also described in WO05/056782; *M. smegmatis*). The amino acid sequences of these proteins, the sequence alignments, and all other information relating to the above is incorporated by reference herein for all purposes from WO05/056782. In a particular embodiment, the parental hydrolase is a wild-type *M. smegmatis* acyltransferase (YP_890535; GID: 118468600) or a variant thereof (e.g., a hydrolase having an amino acid sequence that is at least 95% identical to the wild-type *M.*

*smegmatis* acyltransferase) that has an ability to transfer acyl groups via perhydrolysis or alcoholysis in an aqueous environment.

**[171]** Several examples of such enzymes have been crystallized, and many exemplary amino acid substitutions that may be made in such enzymes that retain or alter their activity are known, e.g., in WO05/056782, which is incorporated by reference for disclosure of such substitutions. For example, lists of hundreds of amino acid substitutions that are tolerated by and, in certain embodiments, may alter the hydrolytic activity, perhydrolytic activity, peracid degradation activity or stability of the *M. smegmatis* acyltransferase are set forth in table 10-3, 10-4, 10-5, 10-6, 10-7, 10-8 and 10-9 of WO05/056782. Each of the amino acid substitutions described in WO05/056782, and the amino acid sequences produced by those substitutions, is incorporated by reference herein.

**[172]** In certain embodiments, the substrate chain length specificity determining segment of a parental hydrolase is the same length as a wild-type hydrolase. In certain embodiments, a parental hydrolase may have other modifications, e.g., insertions, such as those that allow the enzyme to efficiently perhydrolyze or alcoholyze in an aqueous environment. Such insertions are described throughout this disclosure.

*Altered hydrolases*

**[173]** An altered hydrolase is a hydrolase that is otherwise identical to a parental hydrolase, except the substrate chain length specificity determining segment of the hydrolase has an altered length relative to the parental hydrolase. As would be readily apparent, altered hydrolases may be made by inserting amino acids into or deleting amino acids from the substrate chain length specificity determining segment of a parental hydrolase. In certain embodiments, an insertion or deletion may be done by substituting the substrate chain length specificity determining segment of a parental hydrolase with a longer or shorter substrate chain length specificity determining segment of another hydrolase.

**[174]** The substrate chain length specificity determining segment of an altered hydrolase may contain an insertion or deletion of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11-15, 16-20, 21-30 or 31 or more amino acids. If the substrate chain length specificity determining segment of a parental hydrolase is substituted with that of a donor hydrolase, the length of the substrate chain length specificity determining segment substituted into the parent hydrolase from the donor hydrolase may be 1, 2,

3, 4, 5, 6, 7, 8, 9, 10, 11-15, 16-20, 21-30 or 31 or more amino acids longer or shorter than the substrate chain length specificity determining segment of the parental hydrolase.

**[175]** The hydrolase may be produced and isolated using conventional methods (see generally, Ausubel, et al., Short Protocols in Molecular Biology, 3rd ed., Wiley & Sons 1995 and Sambrook et al., Molecular Cloning: A Laboratory Manual, Third Edition, 2001 Cold Spring Harbor, N.Y.). In particular embodiments, production of the hydrolase may be done using recombinant methods using a non-native host, which may produce the hydrolase intracellularly, or secrete the hydrolase. In some embodiments, a signal sequence is added to the enzyme, which facilitates expression of the enzyme by secretion into the periplasm (*i.e.*, in Gram-negative organisms, such as *E. coli*), or into the extracellular space (*i.e.*, in Gram-positive organisms, such as *Bacillus* and *Actinomycetes*), or eukaryotic hosts (*e.g.*, *Trichoderma*, *Aspergillus*, *Saccharomyces*, and *Pichia*). These are just a few examples of possible prokaryotic and eukaryotic hosts. It is not intended that any aspect the present invention be limited to these specific hosts, as various other organisms find use as expression hosts.

*Altered substrate specificity*

**[176]** An above-described altered hydrolase has altered substrate specificity as compared to the parental hydrolase from which it was made. In general terms, in certain embodiments, an overall reduction in the length of the substrate chain length specificity determining segment decreases the length of acyl chain that is transferred by the hydrolase, and an overall increase in the length of the substrate chain length specificity determining segment increases the length of acyl chain that is transferred by the hydrolase. In certain cases, acyl transfer may occur via perhydrolysis or alcoholysis in an aqueous environment.

**[177]** In certain embodiments, the length of acyl chain that is most efficiently transferred by an altered hydrolase, or the upper limit of acyl chain that may be detectably transferred by an altered hydrolase, may be increased or decreased by at least 1 carbon, e.g., by 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11-15, 16-20, 20-30 or more than 30 carbons.

**[178]** In certain cases, a parent hydrolase and altered hydrolase may differ in substrate specificity in that only one of those enzymes can transfer long acyl chains, e.g., an acyl chain of at least 6 carbon atoms. Exemplary long chain acyl contain a $C_6$, $C_7$, $C_8$, $C_9$, $C_{10}$, $C_{11}$, $C_{12}$, $C_{13}$, $C_{14}$, $C_{15}$, $C_{16}$, $C_{17}$, $C_{18}$, $C_{19}$, $C_{20}$, $C_{21}$ or $C_{22}$ carbon chain. Exemplary long chain acyl substrates include: caproic acid ester, caprylic acid ester, nonanoic acid ester, decanoic acid ester,

dodecanoic acid ester, myristic acid ester, palmitic acid ester, stearic acid ester, oleic acid ester, and longer chain acid esters. In general terms, suitable long acyl chain substrates are of the formula $R_1C(=O)OR_2$, where $R_1$ is any transferable acyl moiety that contains a chain of at least 6 carbon atoms and $R_2$ is any organic moiety. Exemplary long chain acyl esters include those that contain a carbon chain of 6-10 carbon atoms (i.e., a $C_6$ - $C_{10}$ carbon chain) or a carbon chain of at least 11 carbon atoms (i.e., a $C_{11+}$ carbon chain). Short acyl chains have a transferable acyl chain containing 5 or less carbons, e.g., a $C_3$-$C_5$ carbon chain.

*Screening methods*

**[179]** Also provided herein is a screening method. In certain embodiment, the method includes altering the length of the substrate chain length specificity determining segment of a parent hydrolase, to produce an altered hydrolase, as described above, and testing the altered SGNH hydrolase for an altered substrate specificity.

**[180]** A variety of assays may be employed to evaluate the substrate specificity of an altered SGNH hydrolase. Many suitable assays are described in PCT publication WO05/056782, which is incorporated by reference for disclosure of such assays. Such assays include: a) plate assays in which cells producing a hydrolase are grown on substrate-containing plates (e.g., agar plates containing 0.25% tricaproin or other substrate) and enzyme activity is evaluated by production of a halo; b) colorigenic assays in which hydrolase activity is evaluated using a substrate (e.g., p-nitrophenylcaproate (pNC6) and p-nitrophenyloctanoate (pNC8)) that is colorigenic upon cleavage, and c) assays in which reaction products are evaluated using mass spectrometry.

**[181]** In certain embodiments, an assay may contain the following components: a test acyl substrate (e.g., an acyl ester) having a carbon chain of a particular length, a recipient molecule, e.g., peroxide or an alcohol, and a hydrolase enzyme, where the assay detects transfer of an acyl group from the acyl substrate to a recipient molecule in an aqueous environment.

**[182]** In certain cases, a library of altered hydrolases may be made, and the members of the library tested for altered substrate specificity. Altered hydrolases having altered substrate specificity may be identified using such methods.

## Compositions

**[183]** Compositions comprising an altered hydrolase are also provided. In certain embodiments the altered hydrolase comprises an amino acid sequence that at least 70% identical to a wild-type

*M. smegmatis* acyltransferase, wherein the length of a substrate determining region of the altered hydrolase is altered relative to the wild-type hydrolase. In certain embodiments, an altered hydrolase may be generally characterized in that it transfers long acyl chains in aqueous environment.

**[184]** The composition may be a food product, e.g., an edible product for human or animal consumption, or an intermediate in the manufacture of an edible food product, or a cleaning composition, for example.

**[185]** As noted above, a cleaning composition comprising a subject hydrolase enzyme is provided. In certain embodiments, the cleaning composition may comprise the following components: a) a subject altered hydrolase, as described above, b) a long chain ester substrate which, in certain embodiments, may be of the formula $R_1C(=O)OR_2$, where $R_1$ comprises a substituted or unsubstituted carbon chain of at least 5 carbon atoms and $R_2$ is any organic moiety; and c) a source of hydrogen peroxide. A variety of other compounds may be present in a subject cleaning composition. A subject altered hydrolase has at least an altered substrate chain length specificity determining segment and, in certain embodiments, other alterations, e.g., those described elsewhere in this disclosure, to alter the substrate specificity of the hydrolase.

**[186]** A subject cleaning composition may be employed for example, in laundry applications, hard surface cleaning, automatic dishwashing applications, as well as cosmetic applications such as dentures, teeth, hair and skin. However, due to their unique properties of increased effectiveness in lower temperature solutions and the superior color-safety profile, the subject enzymes are ideally suited for laundry applications such as the bleaching of fabrics. Furthermore, the enzymes of the present invention find use in both granular and liquid compositions.

**[187]** The subject enzyme also finds use in cleaning additive products. The subject cleaning additive products are ideally suited for inclusion in wash processes where additional bleaching effectiveness is desired. Such instances include, but is not limited to, low temperature solution cleaning applications. The additive product may be, in its simplest form, one or more of the enzymes of the present invention. Such additive may be packaged in dosage form for addition to a cleaning process where a source of peroxygen is employed and increased bleaching effectiveness is desired. Such single dosage form may comprise a pill, tablet, gelcap or other single dosage unit such as pre-measured powders or liquids. A filler or carrier material may be included to increase the volume of such composition. Suitable filler or carrier materials include,

but are not limited to, various salts of sulfate, carbonate and silicate as well as talc, clay and the like. Filler or carrier materials for liquid compositions may be water or low molecular weight primary and secondary alcohols including polyols and diols. Examples of such alcohols include, but are not limited to, methanol, ethanol, propanol and isopropanol. The compositions may contain from about 5% to about 90% of such materials. Acidic fillers can be used to reduce pH. Alternatively, the cleaning additive may include activated peroxygen source such as esters of alcohols, esters of diols, or esters of polyols or as defined below or the adjunct ingredients as also defined below.

**[188]** The cleaning compositions and cleaning additives of the present invention require an effective amount of the enzyme provided by the present invention. Typically a cleaning composition of the present invention comprise at least 0.0001 weight percent, from about 0.0001 to about 1, from about 0.001 to about 0.5, or even from about 0.01 to about 0.1 weight percent of at least one enzyme of the present invention.

**[189]** The above-described composition may be employed in a variety of methods. In general terms, the method may include contacting a subject altered hydrolase with a substrate under conditions suitable for the hydrolase to transfer an acyl group from said substrate onto an acceptor molecule. The contacting may be done in aqueous conditions.

The methods may be employed in food manufacture, cleaning products, biocatalysis (e.g., to produce emulsifying agents and/or surfactants, etc.) and a variety of other applications.

**[190]** Further uses for the above-described enzyme are described in, for example, the following published patent applications: US20070026106, 20060078648, 20050196766 and WO2005066347, which patent applications are incorporated by reference for disclosure of those uses.

**[191]** In addition to the above-described applications, the above-described enzyme may be employed in a variety of food applications. For example, the enzyme may be present in, or may be used to make, a foodstuff, where a foodstuff is any substance which is suitable for human and/or animal consumption. By way of example, the term foodstuff encompasses both baked goods produced from dough as well as the dough used in the preparation of said baked goods. In certain cases, a foodstuff may be a water-containing foodstuff. An exemplary water-containing foodstuff comprise 10-98% water, e.g. 14-98%, 18-98%, 20-98%, 40-98%, 50-98%, 70-98%, 75-98% water, excluding solid components.

**[192]** In certain embodiments, a foodstuff may be selected from one or more of the following: eggs, egg-based products, including but not limited to mayonnaise, salad dressings, sauces, ice creams, egg powder, modified egg yolk and products made therefrom; baked goods, including breads, cakes, sweet dough products, laminated doughs, liquid batters, muffins, doughnuts, biscuits, crackers and cookies; confectionery, including chocolate, candies, caramels, halawa, gums, including sugar free and sugar sweetened gums, bubble gum, soft bubble gum, chewing gum and puddings; frozen products including sorbets, preferably frozen dairy products, including ice cream and ice milk; dairy products, including cheese, butter, milk, coffee cream, whipped cream, custard cream, milk drinks and yoghurts; mousses, whipped vegetable creams, meat products, including processed meat products; edible oils and fats, aerated and non-aerated whipped products, oil-in-water emulsions, water-in-oil emulsions, margarine, shortening and spreads including low fat and very low fat spreads; dressings, mayonnaise, dips, cream based sauces, cream based soups, beverages, spice emulsions and sauces. So-called "fine foods", including cakes, pastry, confectionery, chocolates, fudge and the like, are also types of foodstuff.

**[193]** In one embodiment the foodstuff in accordance with the present invention may be a dough product or a baked product, such as bread, a fried product, a snack, cakes, pies, brownies, cookies, noodles, snack items such as crackers, graham crackers, pretzels, and potato chips, and pasta. In other embodiments, a foodstuff may be a flour, pre-mix, oil, fat, cocoa butter, coffee whitener, salad dressing, margarine, spread, peanut butter, shortenings, ice cream or cooking oil.

**[194]** In another aspect, the foodstuff in accordance with the present invention may be a dairy product, including butter, milk, cream, cheese such as natural, processed, and imitation cheeses in a variety of forms (including shredded, block, slices or grated), cream cheese, ice cream, frozen desserts, yoghurt, yoghurt drinks, butter fat, anhydrous milk fat, other dairy products. The enzyme according to the present invention may improve fat stability in dairy products.

**[195]** In particular embodiments a method of preparing a foodstuff is provided. This method generally comprises adding an above-described enzyme to the foodstuff or an ingredient thereof. In other embodiments, a foodstuff comprising a subject enzyme is provided. For example, enzyme may be employed in the following methods: in situ production of an emulsifier without an increase in free fatty acids; a reduction in the accumulation of free fatty acids in the foodstuff; a reduction in free cholesterol levels in the foodstuff; an increase in sterol esters and/or stanol esters; a reduction in blood serum cholesterol and/or low density lipoproteins; an increase in carbohydrate esters; a reduction in unwanted free carbohydrates.

**[196]** In one example, the acyl acceptor molecule in the foodstuff may be any compound containing a hydroxy group (--OH), such as for example, polyvalent alcohols, including glycerol; sterol; stanols; carbohydrates; hydroxy acids including fruit acids, citric acid, tartaric acid, lactic acid and ascorbic acid; proteins or a sub-unit thereof, such as amino acids, protein hydrolysates and peptides (partly hydrolysed protein) for example; and mixtures and derivatives thereof. In certain cases, the acyl acceptor is not water.

**[197]** A sterol and/or stanol may comprise one or more of the following structural features: i) a 3-beta hydroxy group or a 3-alpha hydroxy group; and/or ii) A:B rings in the cis position or A:B rings in the trans position or $C_5$-$C_6$ is unsaturated. Suitable sterol acyl acceptors include cholesterol and phytosterols, for example alpha-sitosterol, beta-sitosterol, stigmasterol, ergosterol, campesterol, 5,6-dihydrosterol, brassicasterol, alpha-spinasterol, beta-spinasterol, gamma-spinasterol, deltaspinasterol, fucosterol, dimosterol, ascosterol, serebisterol, episterol, anasterol, hyposterol, chondrillasterol, desmosterol, chalinosterol, poriferasterol, clionasterol, sterol glycosides, and other natural or synthetic isomeric forms and derivatives.

**[198]** In one aspect of the present invention suitably more than one sterol and/or stanol may act as the acyl acceptor, suitably more than two sterols and/or stanols may act as the acyl acceptor. In other words, in one aspect of the present invention, suitably more than one sterol ester and/or stanol ester may be produced. Suitably, when cholesterol is the acyl acceptor one or more further sterols or one or more stanols may also act as the acyl acceptor. Thus, in one aspect, the present invention provides a method for the in situ production of both a cholesterol ester and at least one sterol or stanol ester in combination. In other words, the lipid acyltransferase for some aspects of the present invention may transfer an acyl group from a lipid to both cholesterol and at least one further sterol and/or at least one stanol.

**[199]** In one embodiment, the sterol acyl acceptor may be cholesterol. In these embodiments, the amount of free cholesterol in the foodstuff may be reduced as compared with the foodstuff prior to exposure to the enzyme and/or as compared with an equivalent foodstuff which has not been treated with the enzyme.

**[200]** In other embodiment, a subject enzyme may be used in the production of an egg-based product. As such, a method that includes contacting a subject enzyme with an egg or egg-based product is provided. An egg-based product comprising a subject enzyme is also provided.

**[201]** In particular, the presence of sugars, in particular glucose, in eggs and egg products is often seen as disadvantageous. Egg yolk may comprise up to 1% glucose and, egg or egg based

products may be treated with glucose oxidase to remove some or all of this glucose. However, in accordance with certain aspects of the present invention this unwanted sugar can be readily removed by "esterifying" the sugar to form a sugar ester.

**[202]** In particular cases, a carbohydrate ester can function as an emulsifier in foodstuffs. Thus, in certain cases, the enzyme can be employed to transfer an acyl group to a sugar, the invention encompasses the production of an emulsifier, in situ, in the foodstuff. In these cases, a subject enzyme may utilize a sterol and/or stanol and a carbohydrate as an acyl acceptor, which method is particularly useful for the production of foodstuffs containing eggs or egg products. In other embodiments the ester produced (e.g., stanol ester or the sterol ester) may be a flavouring and/or texturiser agent.

**[203]** In another embodiment, a subject enzyme may be added to dough, for example, as part of a baking method. The method may also include baking dough containing the enzyme to make a baked product from the dough. When used in preparation of a dough or baked product a subject enzyme may result in one or more of the following technical effects in dough and/or baked products: an improved specific volume of either the dough or the baked products (for example of bread and/or of cake); an improved dough stability; an improved crust score (for example a thinner and/or crispier bread crust), an improved crumb score (for example a more homogenous crumb distribution and/or a finer crumb structure and/or a softer crumb); an improved appearance (for example a smooth surface without blisters or holes or substantially without blisters or holes); a reduced staling; an enhanced softness; an improved odour; an improved taste.

**[204]** In other embodiments, a subject enzyme may be employed to degum (i.e., reduce the amount of polar lipid, e.g., phospholipids and/or glycolipid such as lecithin, i.e., phosphatidylcholine and cephalin) in vegetable or edible oils. In these embodiment, a subject enzyme may be contacted with the oil so as to hydrolyse a polar lipids in the oil. In particular embodiments a subject enzyme may be employed to reduce phospholipid in an edible oil, comprising treating the oil with a subject enzyme so as to hydrolyse a major part of the phospholipid, and separating an aqueous phase containing the hydrolysed phospholipid from the oil.

**[205]** In other embodiments, a subject enzyme may be employed to convert polar lipids (e.g. glycolipids) into a higher value product, such as carbohydrate esters, protein esters (e.g., via reaction with a serine, threonine, tyrosine, or cysteine residue), and a hydroxy acid ester. Thus, a

subject enzyme may be employed to transfer any acyl chain to onto a sterol, a stanol, a carbohydrate, a protein, or glycerol, for example.

**[206]** In certain embodiments, an emulsifier may be prepared in situ in the foodstuff without an increase in the free fatty acid content of the foodstuff. In certain cases, the production of free fatty acids can be detrimental to foodstuffs. In particular, free fatty acids have been linked with off-odours and/or off-flavours in foodstuffs, as well other detrimental effects, including a soapy taste in cheese for instance. In certain case, this method results in the in situ preparation of an emulsifier(s) wherein the accumulation of free fatty acids is reduced and/or eliminated. In cases, the fatty acid that is removed from the lipid is transferred by the enzyme to an acyl acceptor, for example a sterol and/or a stanol. Thus, unlike similar methods that might employ other lipases (e.g., those having an activity defined by E.C. 3.1.1.x such as a phospholipase A enzyme), the instant method may result in no significant increase in the level of free fatty acids in the foodstuff. Such methods may be particularly employed on foodstuffs containing eggs.

**[207]** A lipid substrate upon which the subject enzyme acts may be a phospholipid, such as a lecithin, e.g. phosphatidylcholine, a triacylglyceride, a cardiolipin, a diglyceride, or a glycolipid, such as digalactosyldiglyceride (DGDG) for example. This lipid substrate may be referred to herein as the "lipid acyl donor". The term lecithin as used herein encompasses phosphatidylcholine, phosphatidylethanolamine, phosphatidylinositol, phosphatidylserine and phosphatidylglycerol. A lipid acyl donor or substrate may be one or more lipids present in one or more of the following substrates: fats, including lard, tallow and butter fat; oils including oils extracted from or derived from palm oil, sunflower oil, soya bean oil, safflower oil, cotton seed oil, ground nut oil, corn oil, olive oil, peanut oil, coconut oil, and rape seed oil. Lecithin from soya, rape seed or egg yolk is also a suitable lipid substrate. The lipid substrate may be an oat lipid or other plant based material containing galactolipids. A lipid may have a fatty acid chain length of from 8 to 22 carbons, e.g., 16 to 22 carbons or 16 to 20 carbons, or no greater than 14 carbons, e.g., 4 to 14 carbons, 4 to 10 carbons or 4 to 8 carbons.

**[208]** For some aspects of the present invention, the variant lipid acyltransferase according to the present invention may utilize a protein as the acyl acceptor. Suitably, the protein may be one or more of the proteins found in a food product, for example in a dairy product and/or a meat product. By way of example only, suitable proteins may be those found in curd or whey, such as lactoglobulin. Other suitable proteins include ovalbumin from egg, gliadin, glutenin, puroindoline, lipid transfer proteins from grains, and myosin from meat.

**[209]** In addition to its applications in detergents, the present invention provides methods and compositions for the use of peracids in textile bleaching and in various other applications. In some embodiments, the present invention provides one-step methods for textile processing applications, including but not limited to one-step desizing, scouring and bleaching processes (*See e.g.,* EP WO 03002810, EP 1255888, WO 0164993, and US 20020007516, all of which are hereby incorporated by reference). As described in greater detail herein, in some embodiments, bleaching involves processing textile material before it is dyed and/or after it is incorporated into textile goods. However, it is not intended that the present invention be limited to any particular regimen of use nor any particular textile material.

## EXPERIMENTAL

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

**[211]** In the experimental disclosure which follows, the following abbreviations apply: °C (degrees Centigrade); rpm (revolutions per minute); $H_2O$ (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); $MgCl_2$ (magnesium chloride); NaCl (sodium chloride); $OD_{280}$ (optical density at 280 nm); $OD_{600}$ (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'-tetraacetylethylenediamine); 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).

**[212]** 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), $K_2HPO_4$ (10 g), and d $H_2O$ (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($H_2O$) (2.4 g), Biospringer yeast extract (6 g), $(NH_4)2SO_4$ (2.4 g), $MgSO_4 \cdot 7\ H_2O$ (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, $CaCl_2 \cdot 2\ H_2O$ (2 mls of 100 mg/ml solution), $KH_2PO_4$ (200 ml of a 13% (w/v) solution at pH6.9), and 20 mls of a 50% glucose solution were added to the medium.

**[213]** 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**

**[214]** 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 ($\lambda 1$), low energy remote ($\lambda 2$), and the peak ($\lambda 3$) of a selenium MAD experiment.

Later, a data set (λ0) 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.

**[215]** 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**

**[216]** 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 (λ0) 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.

**[217]** 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*

**[218]** 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, GTRRILSFGDSLTWGWIPV (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 *Nco*I site at the ATG start codon and add a *Bam*HI site after the stop codon.

**[219]** 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).

**[220]** 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 *Eco*RI to release the fragment, and sequencing using primers supplied by the pCR2.1 manufacturer (Invitrogen). The plasmid was designated pMSATNcoI. The plasmid pMSATNco1 was digested with *NcoI/Bam*HI (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/Bam*HI 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/Bam*HI. 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 MSATNco1-1.

## EXAMPLE 2

### Expression of MsAcT

**[221]** Production of MsAcT for enzymatic analysis was performed by inoculating 5 ml of LB with carbenicillin (100 μg/ml) with a single colony of MSATNco1-1 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 $OD_{600}$ of 0.1. The cultures were grown at 30°C with shaking at 200 rpm until they reached an $OD_{600}$ 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 $KPO_4$, 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 $KPO_4$ 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

**[222]** 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; $Na_2HPO_4$, 6 g/L; $KH_2PO_4$, 3 g/L; $NH_4Cl$, 1 g/L; NaCl, 0.5 g/L; thiamine, 5 mg/L; $MgSO_4$, 2 mM; $CaCl_2$, 100 uM; Citric acid•$H_2O$, 40 mg/L; $MnSO_4$•$H_2O$, 30 mg/L; NaCl, 10 mg/L; $FeSO_4$•$7H_2O$, 1 mg/L; $CoCl_2$•$6H_2O$, 1 mg/L; $ZnSO_4$•$7H_2O$, 1 mg/L; $CuSO_4$•$5H_2O$, 100 ug/L; $H_3BO_3$•$5H_2O$, 100 ug/L; and $NaMoO_4$•$2H_2O$, 100 ug/L; and supplemented with carbenicillin, 100 μg/L.

**[223]** 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 $OD_{600}$ 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 $KPO_4$ buffer, pH 6.8. The yield was 28.5 gm wet weight of cells to which was added 114 ml of 100 mM $KPO_4$ buffer, pH 8.2 and 5 mg of DNase. This mixture was frozen at -80°C and thawed twice.

**[224]** 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 $(NH_4)_2SO_4$ and loaded on to a 10 cm X 11.6 cm Fast Flow Phenyl Sepharose (Pharmacia) column equilibrated in100 mM $KPO_4$ buffer, pH 6.8, containing 20% saturation (109 g/L) $(NH_4)_2SO_4$. 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 $(NH_4)_2SO_4$, 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

**[225]** 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**

**[226]** 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 um diameter, 250 nm film thickness). The results were reported as area of the NPG monoester/hr.

**Determination of Hydrolysis**

**[227]** 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 um diameter, 250 nm film thickness). A 3 µL 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**

**[228]** 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 μL 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 μL 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.

**[229]** 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)

**[230]** 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.

**[231]** 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.

**[232]** Having described the preferred 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.

**[233]** 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 of preferred 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.

## EXAMPLE 5

### Designing an Insertion to Enlarge Acyl Accepting Cavity in *M smegmatis* ACT.

**[234]** One approach is for enlarging the acyl accepting cavity of MsACT is to replace the nucleic acid segment encoding amino acid residues 178-191 corresponding to the sequence:

P F F D A G S V I S T D G V (SEQ ID NO: 10)

with a nucleic acid segment encoding amino acid residues 135-154 of *E coli* thioesterase corresponding to the amino acid sequence :

P L L P F F M E E V L Y K P Q W M Q D (SEQ ID NO:13).

**[235]** When the tertiary structures of the E. coli thioesterase and the MsACT enzymes are aligned, the loop corresponding to amino acid residues 135-154 on the thioesterase starts and returns to the same topological fold as the segment corresponding to amino acid residues 178-191 of *M. smegmatis* ACT.

[236] 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 preferred 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.

[237] 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 method comprising:

altering the length of a substrate chain length specificity determining segment that lies immediately N-terminal to the Asp residue of the catalytic triad of a parent SGNH hydrolase that has an amino acid sequence that is at least 70% identical to the *M. smegmatis* acyltransferase, to produce an altered SGNH hydrolase that has an altered substrate specificity relative to said parent SGNH hydrolase.

2. The method of claim 1, wherein said altering comprises making an amino acid insertion or deletion in said substrate chain length specificity determining segment.

3. The method of claim 1, wherein said altering comprises substituting said substrate chain length specificity determining segment of said parent SGNH hydrolase with the substrate chain length specificity determining segment of a different SGNH hydrolase to produce said altered SGNH hydrolase.

4. The method of claim 1, wherein said altering increases the length of acyl chain that can be transferred by said SGNH hydrolase.

5. The method of claim 4, wherein said acyl chain transfer occurs in an aqueous environment.

6. A method comprising:

altering the length of a substrate chain length specificity determining segment that lies immediately N-terminal to the Asp residue of the catalytic triad of a parent SGNH hydrolase that has an amino acid sequence that is at least 70% identical to the *M. smegmatis* acyltransferase to produce an altered SGNH hydrolase; and

testing said altered SGNH hydrolase for an altered substrate specificity.

7. The method of claim 6, wherein said altering includes making an amino acid insertion or deletion in said substrate chain length specificity determining segment.

8. The method of claim 6, wherein said method includes inserting a random sequence of amino acids into said loop.

9. The method of claim 6, wherein said testing includes evaluating the ability of said altered SGNH hydrolase to transfer an acyl group from a substrate to a recipient molecule in an aqueous environment.

10. The method of claim 9, wherein said testing includes evaluating the ability of said altered SGNH hydrolase and said parent SGNH hydrolase to transfer acyl chains of different lengths.

11. An altered SGNH hydrolase comprising an amino acid sequence that is at least 70% identical to a wild-type *M. smegmatis* acyltransferase, wherein a substrate chain length specificity determining segment that lies immediately N-terminal of the Asp residue of the catalytic triad of said altered SGNH hydrolase has an altered length relative to said wild-type *M. smegmatis* acyltransferase.

12. The altered SGNH hydrolase of claim 11, wherein said altered SGNH hydrolase comprises an amino acid sequence that is at least 90% identical to a wild-type *M. smegmatis* acyltransferase.

13. A food product comprising the altered SGNH hydrolase of claim 11.

14. A cleaning composition comprising the altered SGNH hydrolase of claim 11.

15. A method comprising:

contacting the altered SGNH hydrolase of claim 11 with a substrate under conditions suitable for said SGNH hydrolase to transfer an acyl group from said substrate onto a recipient.

16. The method of claim 15, wherein said contacting is done in an aqueous environment.

17. The method of claim 15, wherein said substrate is a long acyl chain substrate.

18. The method of claim 15, wherein said recipient is an alcohol or peroxide.

19 The method of claim 15, wherein altered SGNH hydrolase is contacted with a food product that contains said substrate.

20. The method of claim 15, wherein said altered SGNH hydrolase transfers an acyl group from said substrate to a recipient to produce an emulsifier or surfactant.

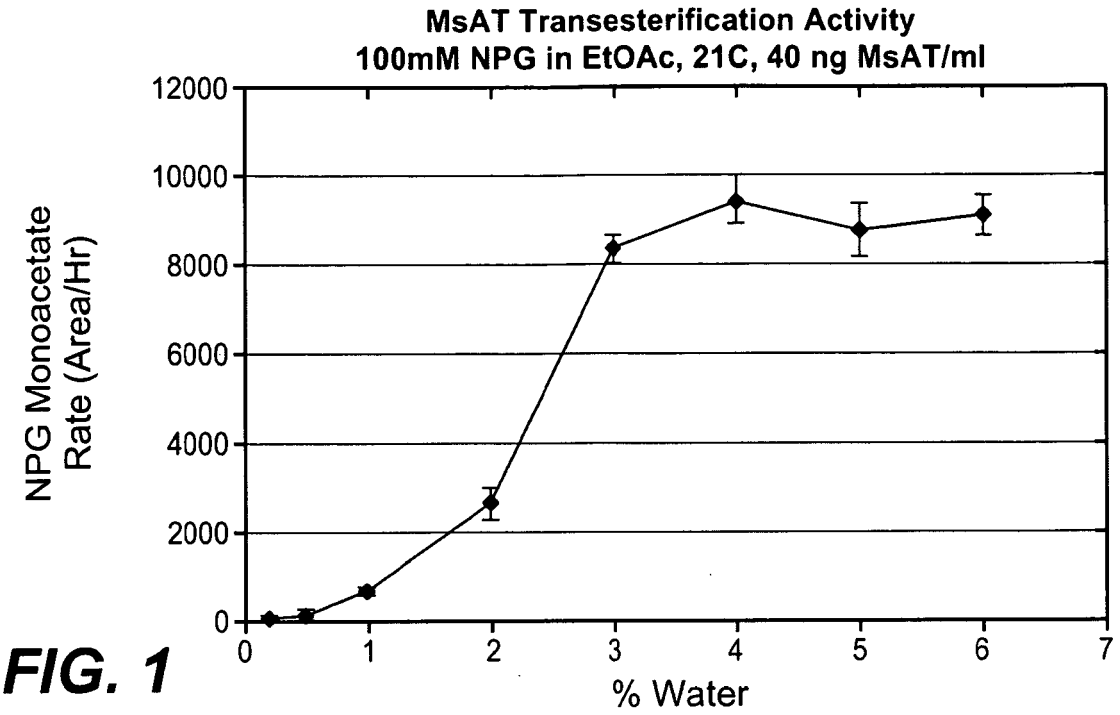
MsAT Transesterification Activity
100mM NPG in EtOAc, 21C, 40 ng MsAT/ml
NPG Monoacetate Rate (Area/Hr)
12000
10000
8000
6000
4000
2000
0
0
1
2
3
4
5
6
7
% Water

FIG. 1

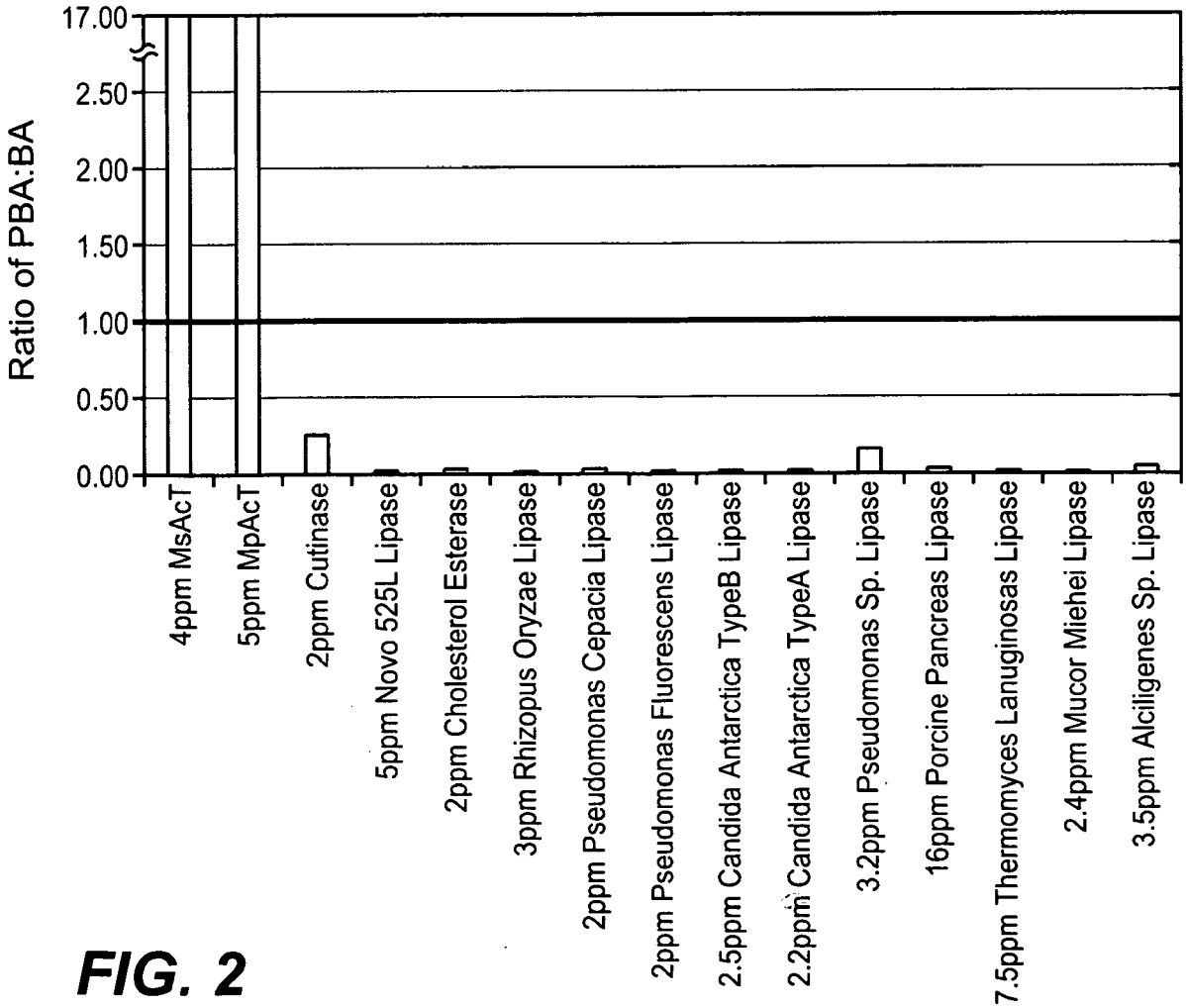
Ratio of PBA:BA
17.00
2.50
2.00
1.50
1.00
0.50
0.00
4ppm MsAcT
5ppm MpAcT
2ppm Cutinase
5ppm Novo 525L Lipase
2ppm Cholesterol Esterase
3ppm Rhizopus Oryzae Lipase
2ppm Pseudomonas Cepacia Lipase
2ppm Pseudomonas Fluorescens Lipase
2.5ppm Candida Antarctica TypeB Lipase
2.2ppm Candida Antarctica TypeA Lipase
3.2ppm Pseudomonas Sp. Lipase
16ppm Porcine Pancreas Lipase
7.5ppm Thermomyces Lanuginosas Lipase
2.4ppm Mucor Miehei Lipase
3.5ppm Alciligenes Sp. Lipase

FIG. 2

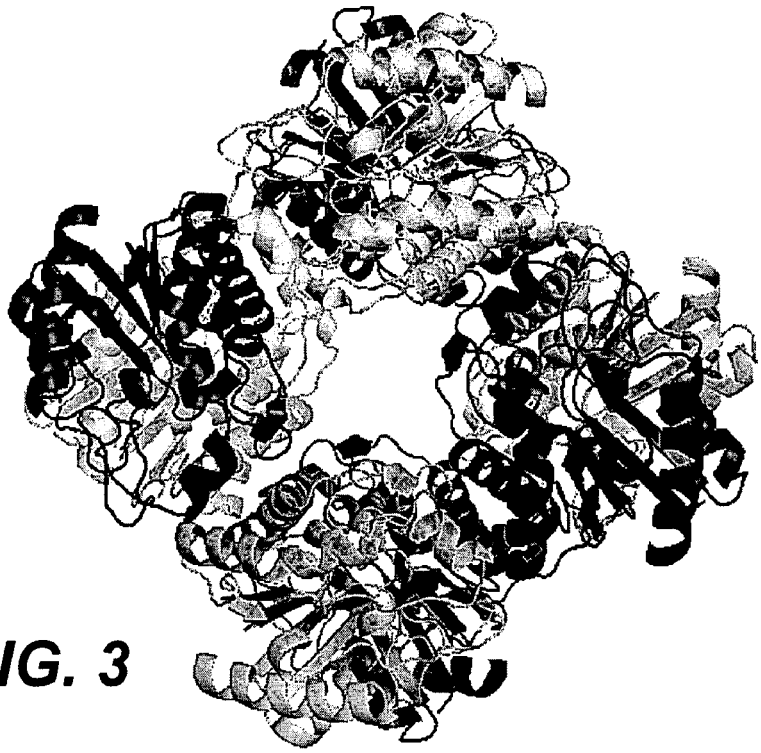

FIG. 3

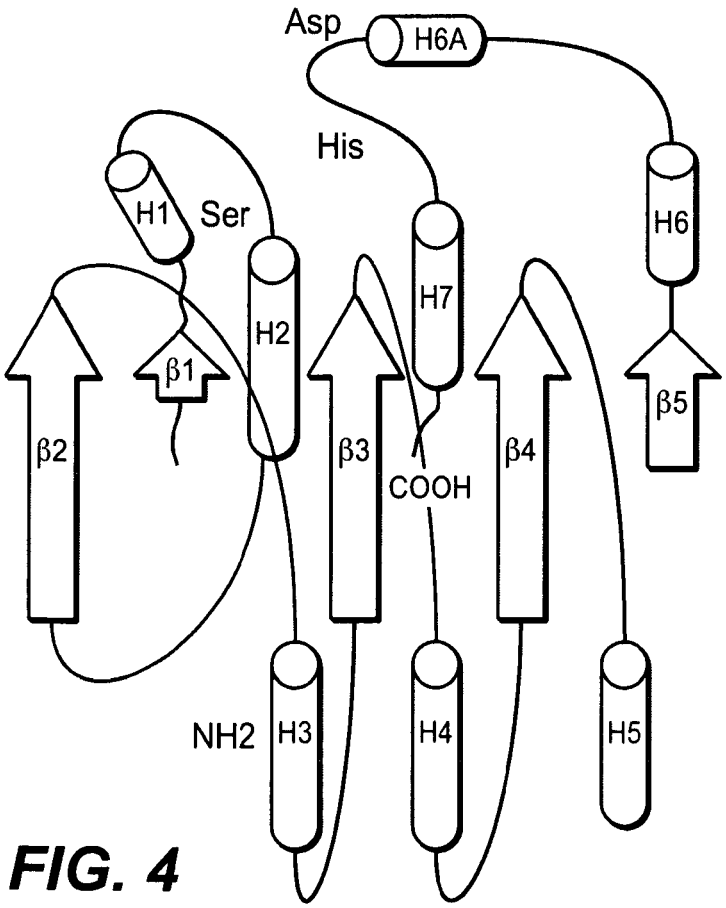


FIG. 4

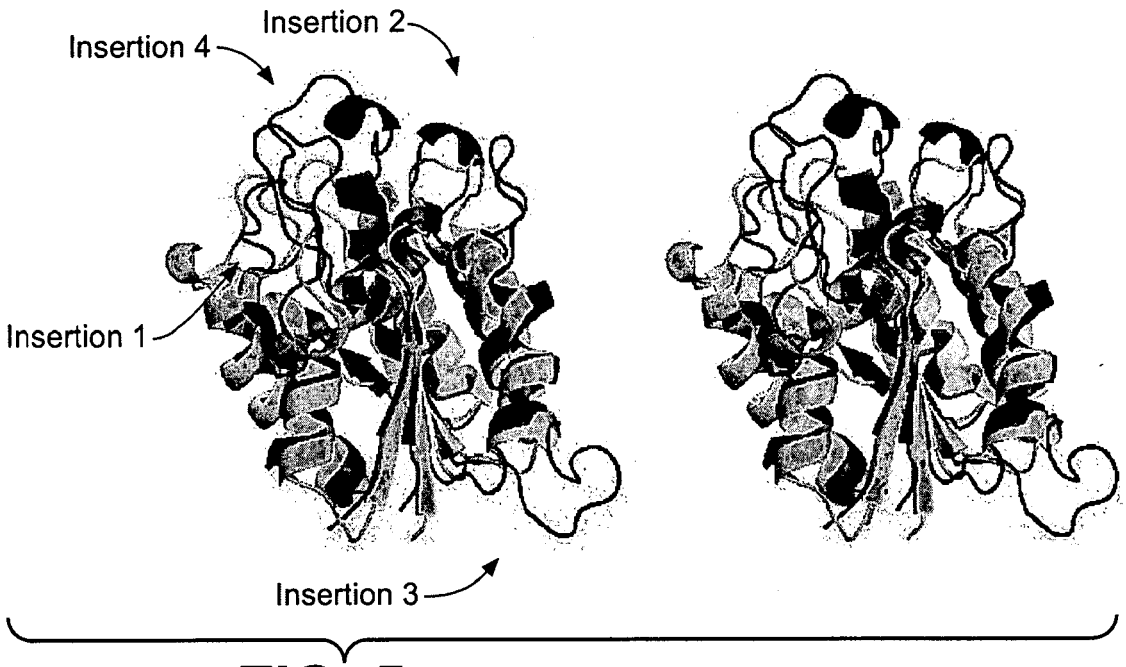
Insertion 2
Insertion 4
Insertion 1
Insertion 3

FIG. 5

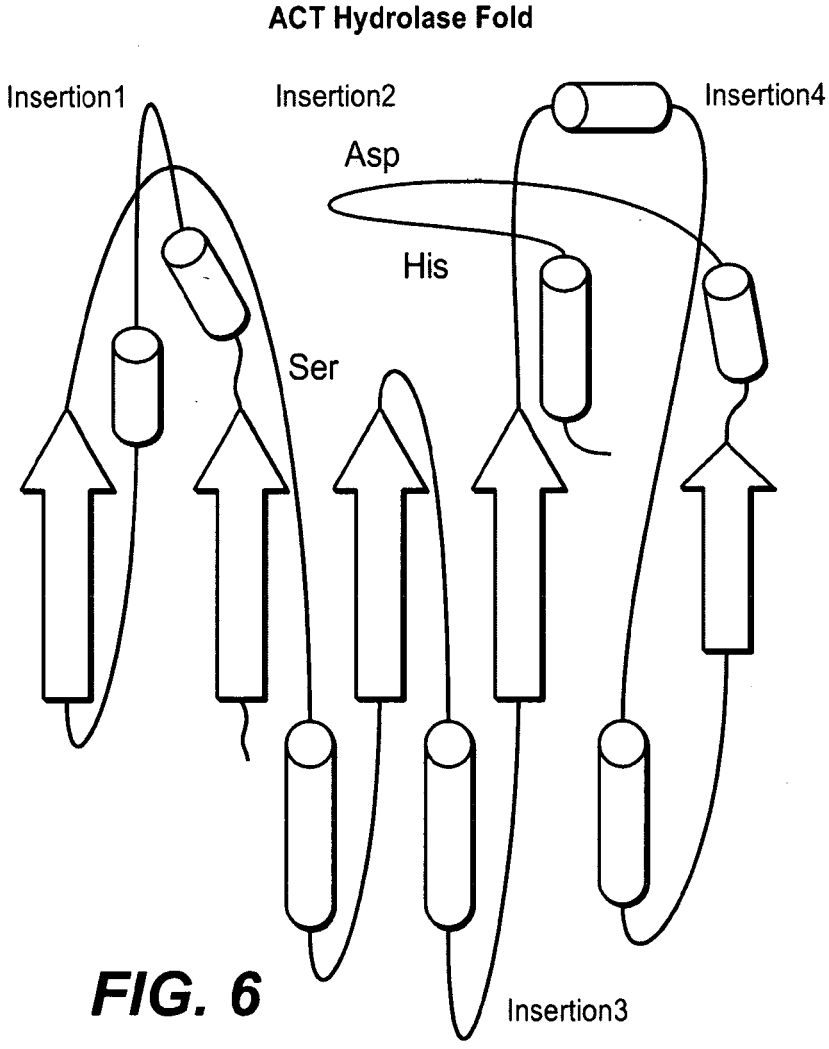
ACT Hydrolase Fold
Insertion1
Insertion2
Insertion4
Asp
His
Ser
Insertion3

FIG. 6

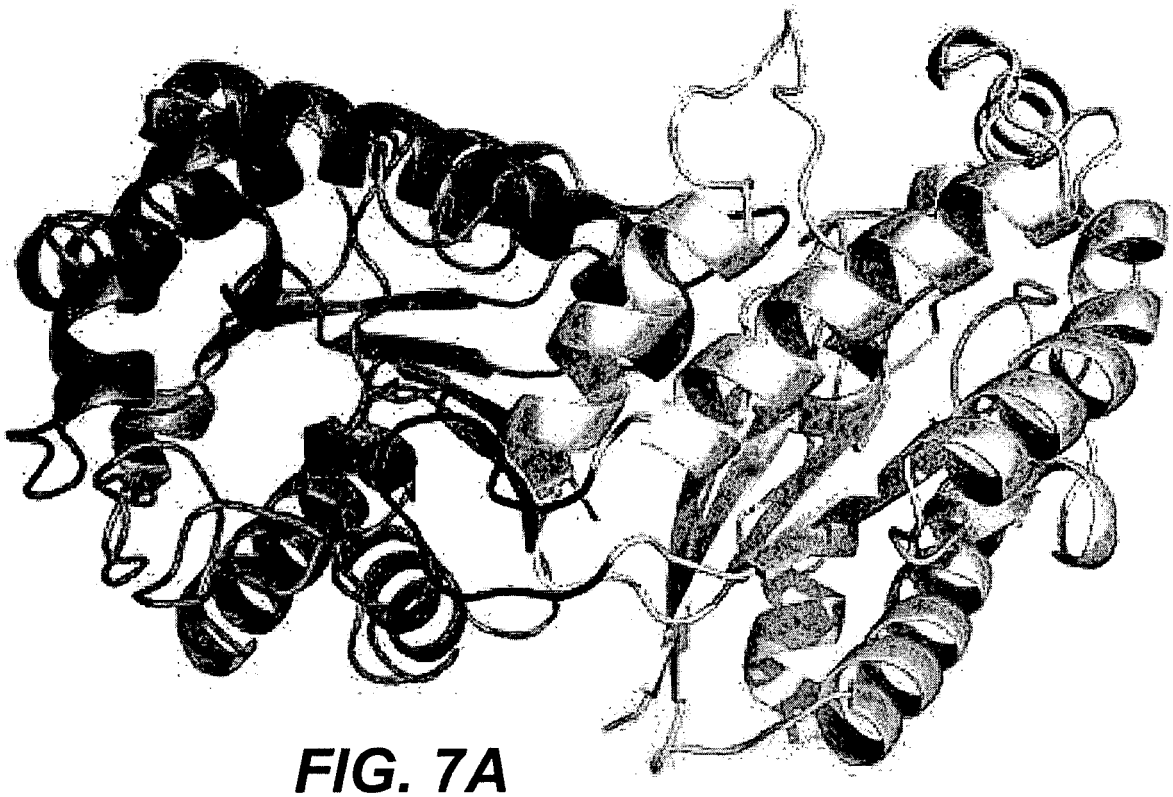

FIG. 7A

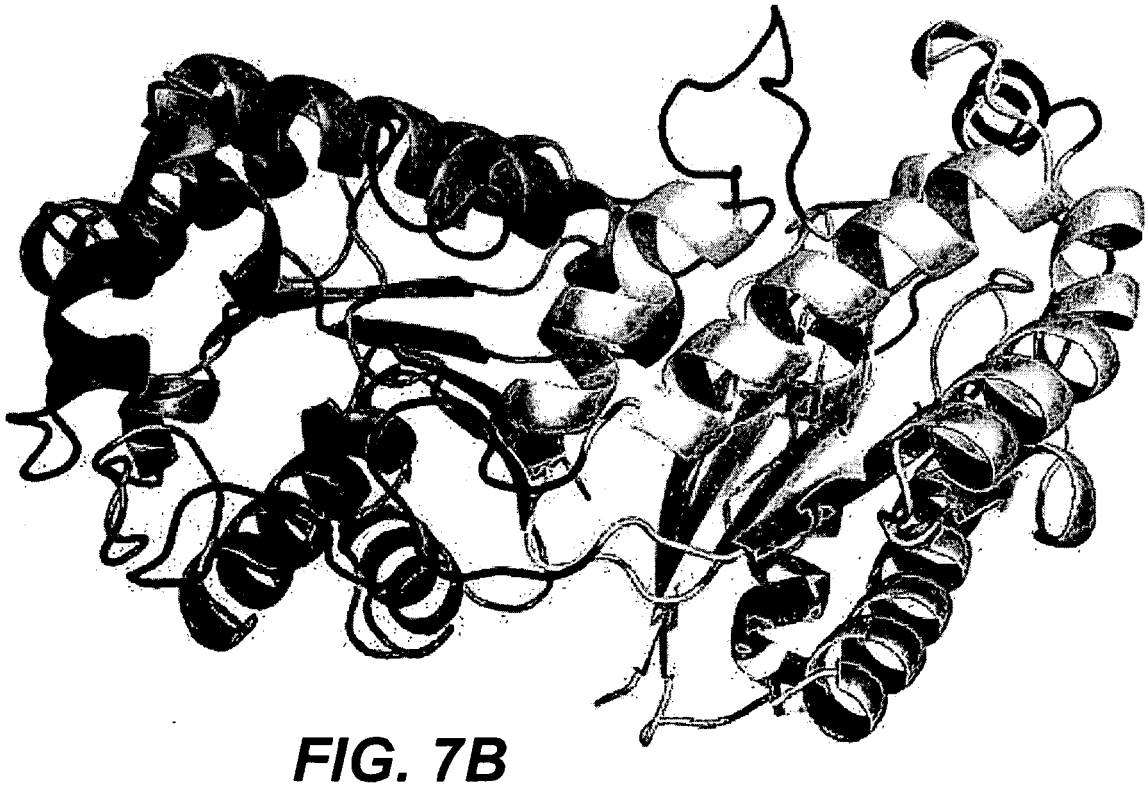

FIG. 7B

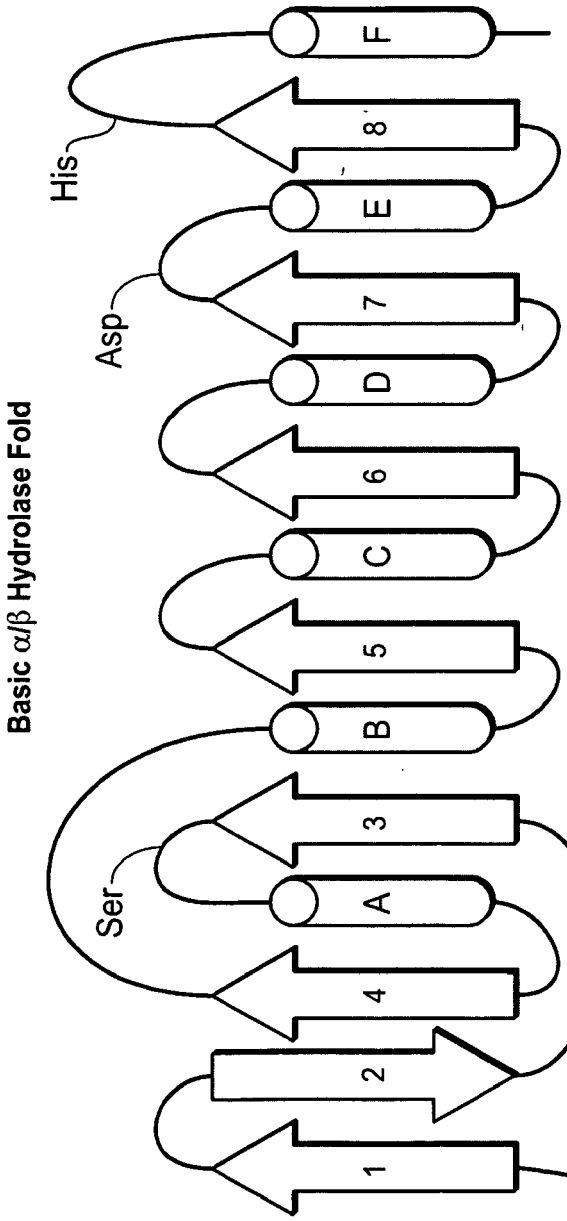


FIG. 8

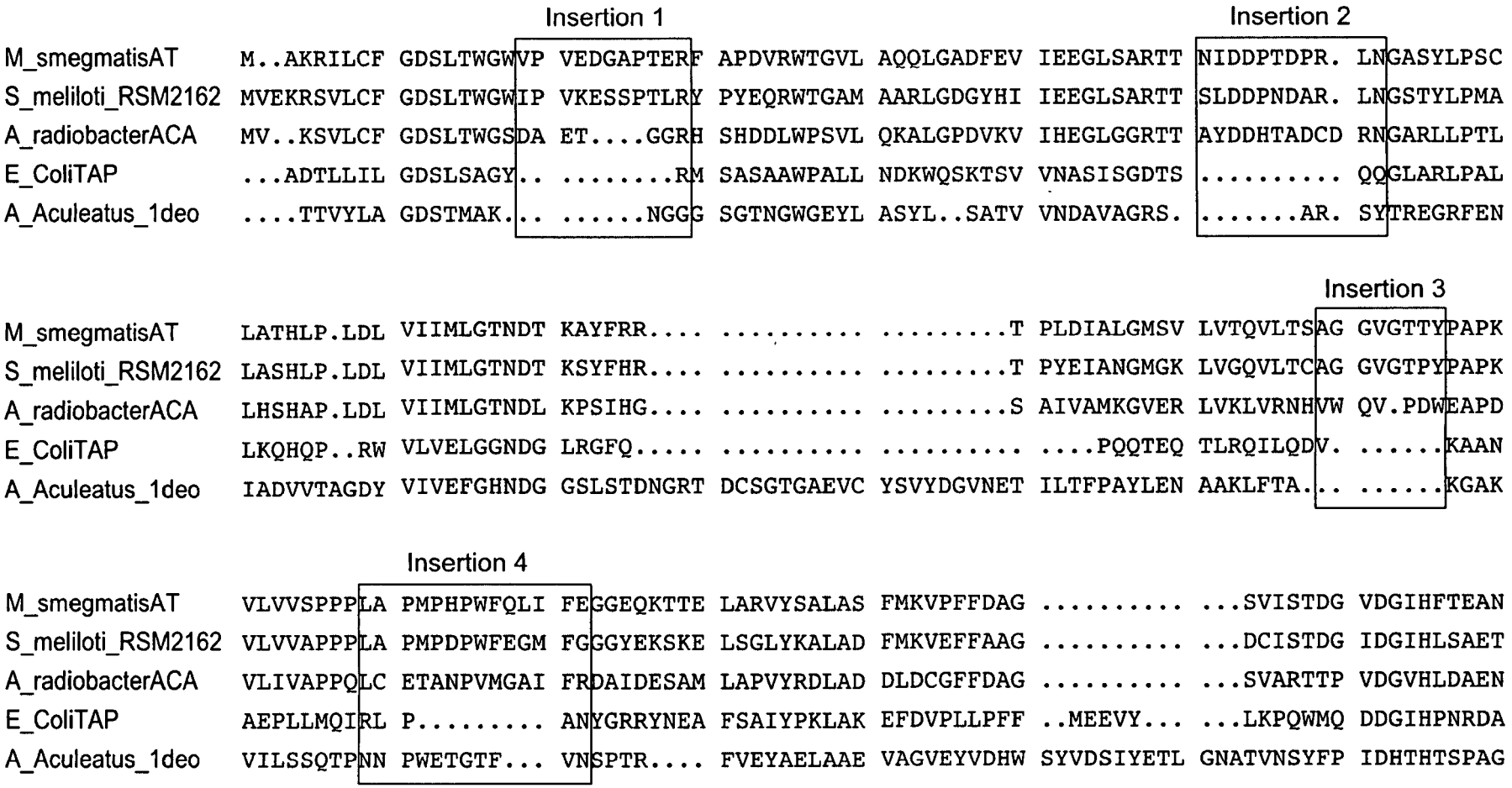
Insertion 1
Insertion 2
Insertion 3
Insertion 4
M_smegmatisAT M..AKRILCF GDSLTWGWVP VEDGAPTERF APDVRWTGVL AQQLGADFEV IEEGLSARTT NIDDPTDPR. LNGASYLPSC
S_meliloti_RSM2162 MVEKRSVLCF GDSLTWGWIP VKESSPTLRY PYEQRWTGAM AARLGDGYHI IEEGLSARTT SLDDPNDAR. LNGSTYLPMA
A_radiobacterACA MV..KSVLCF GDSLTWGSDA ET....GGRH SHDDLWPSVL QKALGPDVKV IHEGLGGRTT AYDDHTADCD RNGARLLPTL
E_ColiTAP ...ADTLLIL GDSLSAGY.. ........RM SASAAWPALL NDKWQSKTSV VNASISGDTS .......... QQGLARLPAL
A_Aculeatus_1deo ....TTVYLA GDSTMAK... ......NGGG SGTNGWGEYL ASYL..SATV VNDAVAGRS. .......AR. SYTREGRFEN
M_smegmatisAT LATHLP.LDL VIIMLGTNDT KAYFRR.... .......... .........T PLDIALGMSV LVTQVLTSAG GVGTTYPAPK
S_meliloti_RSM2162 LASHLP.LDL VIIMLGTNDT KSYFHR.... .......... .........T PYEIANGMGK LVGQVLTCAG GVGTPYPAPK
A_radiobacterACA LHSHAP.LDL VIIMLGTNDL KPSIHG.... .......... .........S AIVAMKGVER LVKLVRNHVW QV.PDWEAPD
E_ColiTAP LKQHQP..RW VLVELGGNDG LRGFQ..... .......... .......... ....PQQTEQ TLRQILQDV. ......KAAN
A_Aculeatus_1deo IADVVTAGDY VIVEFGHNDG GSLSTDNGRT DCSGTGAEVC YSVYDGVNET ILTFPAYLEN AAKLFTA... ......KGAK
M_smegmatisAT VLVVSPPPLA PMPHPWFQLI FEGGEQKTTE LARVYSALAS FMKVPFFDAG .......... ...SVISTDG VDGIHFTEAN
S_meliloti_RSM2162 VLVVAPPPLA PMPDPWFEGM FGGGYEKSKE LSGLYKALAD FMKVEFFAAG .......... ...DCISTDG IDGIHLSAET
A_radiobacterACA VLIVAPPQLC ETANPVMGAI FRDAIDESAM LAPVYRDLAD DLDCGFFDAG .......... ...SVARTTP VDGVHLDAEN
E_ColiTAP AEPLLMQIRL P......... ANYGRRYNEA FSAIYPKLAK EFDVPLLPFF ..MEEVY... ...LKPQWMQ DDGIHPNRDA
A_Aculeatus_1deo VILSSQTPNN PWETGTF... VNSPTR.... FVEYAELAAE VAGVEYVDHW SYVDSIYETL GNATVNSYFP IDHTHTSPAG
M_smegmatisAT NRDLGVALAE QVRSLL.... .......... ...
S_meliloti_RSM2162 NIRLGHAIAD KVAALF.... .......... ...
A_radiobacterACA TRAIGRGLEP VVRMMLGL.. .......... ...
E_ColiTAP QPFIADWMAK QLQPLVNHDS .......... ...
A_Aculeatus_1deo AEVVAEAFLK AVVCTGTSLK SVLTTTSFEG TCL

**FIG. 9**

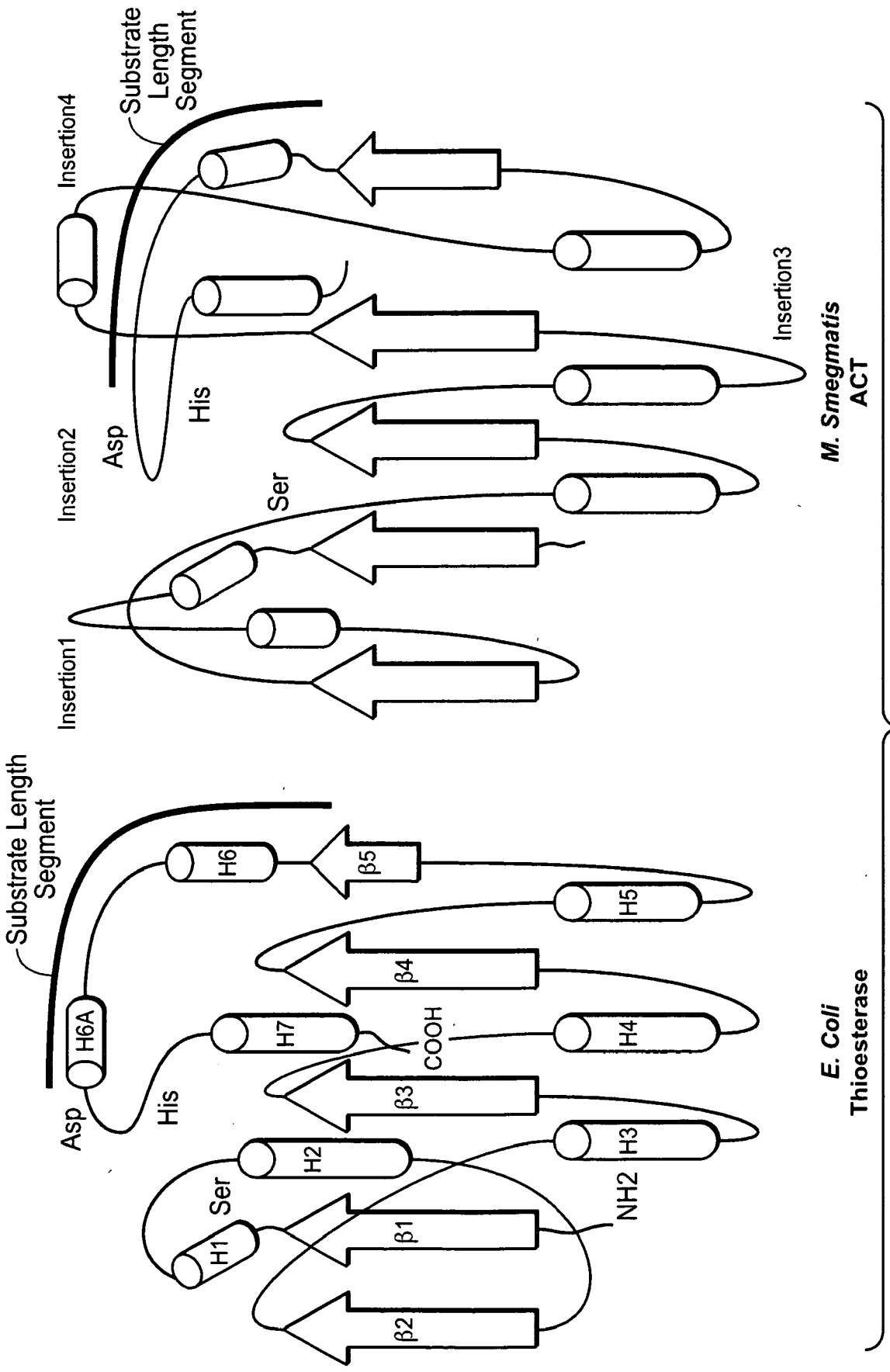
Substrate Length Segment
Asp
H6A
His
H7
H6
β5
H5
β4
H4
COOH
β3
H2
Ser
H1
β1
β2
H3
NH2
E. Coli Thioesterase
Insertion1
Insertion2
Asp
His
Ser
Insertion4
Substrate Length Segment
Insertion3
M. Smegmatis ACT

FIG. 10

# INTERNATIONAL SEARCH REPORT

International application No
PCT/US2008/007807

A. CLASSIFICATION OF SUBJECT MATTER
INV. C07K14/435

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C07K

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

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, Sequence Search, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | WO 2005/056782 A (GENENCOR INT [US]; PROCTER & GAMBLE [US]; AMIN NEELAM S [US]; BOSTON M) 23 June 2005 (2005-06-23) Page 80-83: Structure of M, smegmatis Perhydrolase, Figures 3 and 4, Example 13. Claims & DATABASE Geneseq [Online] 25 August 2005 (2005-08-25), "Perhydrolase homolog SEQ ID NO 695." retrieved from EBI accession no. GSP:AEA43769 Database accession no. AEA43769 | 11,12 |
| X | -& DATABASE Geneseq [Online] 25 August 2005 (2005-08-25), "M. gilvum perhydrolase homolog 14B." XP002505225 retrieved from EBI accession no. GSP:AEA43714 Database accession no. AEA43714 | 11,12 |

-/--

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

[X] See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 25 November 2008 | 15/12/2008 |
| Name and mailing address of the ISA/<br>European Patent Office, P.B. 5818 Patentlaan 2<br>NL – 2280 HV Rijswijk<br>Tel. (+31-70) 340-2040,<br>Fax: (+31-70) 340-3016 | Authorized officer<br>Simm, Mariola |

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | -& DATABASE Geneseq [Online]<br>25 August 2005 (2005-08-25), "M. parafortuitum perhydrolase homolog 20A."<br>XP002505226<br>retrieved from EBI accession no. GSP:AEA43716<br>Database accession no. AEA43716 | 11,12 |
| X,P | WO 2007/133263 A (GENENCOR INT [US]; CERVIN MARGUERITE A [US]; WHITED GREGG [US]) 22 November 2007 (2007-11-22)<br>the whole document | 11,12 |
| T | MATHEWS IRIMPAN ET AL: "Structure of a novel enzyme that catalyzes acyl transfer to alcohols in aqueous conditions"<br>BIOCHEMISTRY, AMERICAN CHEMICAL SOCIETY, EASTON, PA.; US,<br>vol. 46, no. 31,<br>1 August 2007 (2007-08-01), pages 8969-8979, XP002465298<br>ISSN: 0006-2960<br>the whole document | |
| A | BERNHARDT P ET AL: "MOLECULAR BASIS OF PERHYDROLASE ACTIVITY IN SERINE HYDROLASES"<br>ANGEWANDTE CHEMIE. INTERNATIONAL EDITION, VCH VERLAG, WEINHEIM, DE,<br>vol. 44, no. 18,<br>29 April 2005 (2005-04-29), pages 2742-2746, XP008047813<br>ISSN: 0570-0833 | |
| P,X | WO 2008/019069 A (DANISCO US INC GENENCOR DIVISI [US]; BOTT RICHARD R [US]; CERVIN MARGU) 14 February 2008 (2008-02-14)<br>the whole document<br>& DATABASE Geneseq [Online]<br>24 July 2008 (2008-07-24), "Mycobacterium smegmatis arylesterase SEQ ID 1."<br>retrieved from EBI accession no. GSP:AQX98615<br>Database accession no. AQX98615<br>abstract | 11,12 |
| P,X | WO 2008/063400 A (DANISCO US INC GENENCOR DIVISI [US]; AMIN NEELAM S [US]; BOTT RICHARD) 29 May 2008 (2008-05-29)<br>Paragraphs 10, 12, 63,132. | 11,12 |

-/--

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| | & DATABASE Geneseq [Online]<br>24 July 2008 (2008-07-24), "Mycobacterium smegmatis mutant perhydrolase (L12G,F154G)."<br>retrieved from EBI accession no. GSP:ARW66032<br>Database accession no. ARW66032<br>abstract<br>----- | |

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2008/007807

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| WO 2005056782 A | 23-06-2005 | BR PI0417233 A | 06-03-2007 |
| | | CA 2547709 A1 | 23-06-2005 |
| | | CN 1981035 A | 13-06-2007 |
| | | EP 1689859 A2 | 16-08-2006 |
| | | MX PA06005652 A | 17-08-2006 |
| | | US 2008145353 A1 | 19-06-2008 |
| WO 2007133263 A | 22-11-2007 | AU 2006343548 A1 | 22-11-2007 |
| | | CA 2633849 A1 | 22-11-2007 |
| | | EP 1954814 A2 | 13-08-2008 |
| WO 2008019069 A | 14-02-2008 | NONE | |
| WO 2008063400 A | 29-05-2008 | US 2007167344 A1 | 19-07-2007 |