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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 June 2017 (15.06.2017)

(10) International Publication Number
WO 2017/100240 A1

(51) International Patent Classification:
C12N 9/20 (2006.01) *G06F 19/16* (2011.01) *C12P 21/02* (2006.01)

(21) International Application Number:
PCT/US2016/065255

(22) International Filing Date:
7 December 2016 (07.12.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
14/961,237 7 December 2015 (07.12.2015) US

(71) Applicants: EASTMAN CHEMICAL COMPANY [US/US]; 200 South Wilcox Drive, Kingsport, TN 37663 (US). NORTH CAROLINA STATE UNIVERSITY [US/US]; Office Of Technology Transfer, 1021 Main Campus Drive, 2nd Floor, Raleigh, NC 27606 (US).

(72) Inventors: CLENDENNEN, Stephanie, Kay; 148 Lebanon Road, Kingsport, TN 37663 (US). YINGLING, Yaroslava, Georgievna; 111 West Laurenbrook Court, Cary, NC 27518 (US). KIM, Hoshin; 4811 Smith Creek Pkwy., Pat. 303, Raleigh, NC 27612 (US).

(74) Agents: TENG, Sally, P. et al.; Lee & Hayes, PLLC, 601 W. Riverside Ave, Suite 1400, Spokane, WA 99201 (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, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, 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, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

[Continued on next page]

(54) Title: CALB VARIANTS

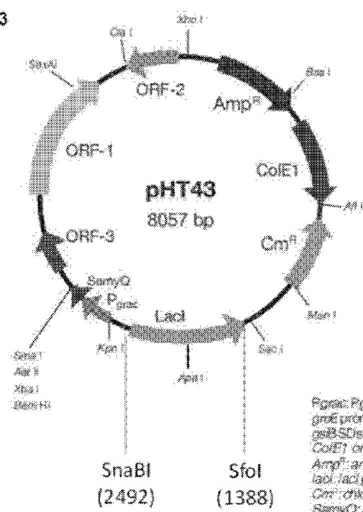


Fig.1

(57) Abstract: The invention relates to amino acid sequence variants of a lipase with improved activity for catalyzing synthesis reactions and methods of preparing the variants. The methods include predicting amino acid sites for change based on computational models of the protein structure in non-aqueous conditions, and expressing the protein in a prokaryotic host for subsequent purification and use. The enzyme sequence variants described have a three to nine-fold improvement in synthesis activity over the parent protein sequence.

WO 2017/100240 A1

— *with sequence listing part of description (Rule 5.2(a))*

CALB Variants

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application claims the benefit of U.S. Application Serial No. 14/961,237 filed December 7, 2015, which is incorporated herein by reference in its entirety.

SEQUENCE LISTING INFORMATION

[0002] A computer readable textfile, entitled CALB Variants "E023-0054US-Sequence Listing.txt," created on or about November 10, 2015, with a file size of about 52 KB, contains the sequence listing for this application and is hereby incorporated by reference in its entirety.

FIELD

[0003] The present disclosure relates to variants of *Candida antarctica* lipase B (CALB) having improved lipase activity as compared to wild type CALB, methods of identifying the variants, and methods of using the variants.

BACKGROUND

[0004] The *Candida antarctica* lipase B (CALB) is able to catalyze synthesis reactions, with esterification reactions catalyzed by CALB being particularly well-studied. While CALB is able to react with a wide variety of alcohol substrates to form esters, it is more limited in the type of acid substrate it recognizes, with preference for straight-chain fatty acids.

[0005] Eukaryotic hosts like fungi and the yeasts *Saccharomyces cerevisiae*, *Yarrowia lipolytica* and *Pichia pastoris* have been engineered to produce a secreted form of CALB. CALB is also supplied commercially, expressed and secreted in a recombinant fungal or yeast host (Novozymes and cLecta product data sheets). However, low transformation efficiency and long growth periods make these eukaryotic systems difficult to use for high-throughput screening of large numbers of enzyme variants. Typically, DNA constructs must pass or “shuttle” through an *E. coli* or other bacterial host prior to introduction into the eukaryotic host, and the cells must be grown for several days, often in the presence of an inducer to stimulate

expression. In addition, a generally useful secretion system for extracellular lipase expression in yeast is lacking, and efficient recovery of the recombinant lipase requires the lipase to be active outside the host cell in a cell-free system. The expression of the eukaryotic CALB enzyme has been accomplished in the common bacterial (prokaryotic) host E. coli, but not secretion, and so a subsequent cell-lysis step is required to liberate the lipase for characterization. Frequently only the hydrolytic activity of these recombinant lipases was confirmed, but not synthesis activity. A recent publication summarizes the difficulty of expressing CALB in a heterologous host, especially a bacterial host (Larsen et al., 2008). The authors hypothesize that incorrect protein folding in *E. coli* is a limitation in expression of CALB in this bacterial host.

[0006] A CALB variant with improved activity for synthesis reactions would improve the efficiency of esterification, amidation and transesterification reactions and permit the economic manufacture of compounds using an enzyme catalyst. A CALB variant with improved activity for synthesis reactions would also permit the use of an enzyme catalyst to synthesize derivatives of hindered substrates. While some CALB variants having improved hydrolytic activity have been prepared, these variants are irrelevant to improving synthetic activity, which occurs in the absence of water. In order to identify amino acid changes in the native CALB sequence to target for change and to measure and understand the impact of amino acid changes on structure and function of the enzyme, it is also necessary to devise methods for predicting the protein structure in synthesis conditions, that is in the absence of water, and also to devise a method for expressing the enzyme variants and isolating them in a form suitable for synthesis reactions.

SUMMARY

[0007] The present disclosure provides CALB variants having improved synthesis activity. Moreover, the present disclosure provides a new bacterial expression system using *Bacillus subtilis* (Bsub) for expressing CALB. The bacterial expression system is suitable for high-throughput screening of enzyme variants. Additionally, the present invention provides an improved method for molecular dynamic simulation analysis to accurately determine amino acid residues for alteration to obtain CALB variants with improved functional activity.

[0008] The present disclosure provides *Candida antarctica* lipase B (CALB) variants having about two fold to about fifteen fold improved synthetic activity as compared to a wild type (WT) CALB. In embodiments, the CALB variant has an amino acid sequence having one or more modifications and the one or more modifications can be at position 141, 146, 188, 189, 223, 227, or 235 of SEQ ID NO: 2. In other embodiments, the CALB variant has two to seven amino acid substitutions and the amino acid substitutions can be at position 141, 146, 188, 189, 223, 227, or 235 of SEQ ID NO: 2. In further embodiments, the amino acid substitutions for the CALB variants include one or more of the following: A141T, A146T, E188D, I189V, D223G, S227T, or V235A.

[0009] The present disclosure provides CALB variants having an amino acid sequence as set forth in SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20 or SEQ ID NO: 22.

[0010] The present disclosure also provides nucleic acids encoding the CALB variants disclosed herein. In an aspect, the nucleic acid encoding a CALB variant is in a vector, such as an expression vector. The expression vector comprises regulatory elements including a secretion signal.

[0011] The present disclosure provides methods for expressing a CALB variant disclosed herein, wherein the method comprises transfecting an expression vector comprising the nucleic acid encoding a CALB variant into a host cell and culturing the host cell under conditions allowing expression of the CALB variant. The present disclosure also provides a method for preparing a CALB variant comprising expressing a CALB variant and obtaining the supernatant from the culture medium. The method can further comprise concentrating the supernatant comprising a CALB variant. The host cell for expressing the CALB variant can be a strain of *Bacillus subtilis.*

[0012] The present disclosure provides an expression vector for expressing CALB or CALB variant in a strain of *Bacillus,* wherein the expression vector comprises a nucleic acid encoding CALB or a CALB variant, one or more origins of replication for replication in *Bacillus*, a promoter, a secretion signal, and optionally a selectable marker. The promoter can be a constitutive promoter. In embodiments, the present disclosure provides an expression system comprising the expression vector, a host strain of *Bacillus*, and a culture medium.

[0013] The present disclosure provides culture medium comprising a non-carbohydrate micronutrient source, a buffering agent for maintaining pH of the medium at a range of about 5 to about 9, a non-hydrolyzable nonionic surfactant, and a nitrogen source. In embodiments, the culture medium comprises: about 0.1% to about 5% of a non-carbohydrate micronutrient source, relative to the total weight of the composition; a buffering agent for maintaining the pH at about 6 to about 8; about 0.01% to about 1% of a non-hydrolyzable nonionic surfactant, relative to the total weight of the composition; and about 0.1% to about 5% of a nitrogen source, relative to the total weight of the composition. In embodiments, the culture medium comprises: yeast extract, a buffering agent for maintaining the pH of the medium at 7, a block copolymer, and a nitrogen source comprising tryptone, hydrolyzed casein, casamino acids, peptone, soy peptone, nutrient broth, or meat extract.

[0014] The present disclosure provides a method of making CALB, WT or CALB variant, comprising transfecting the expression vector for expression in *Bacillus* into a *Bacillus*, and cultivating the *Bacillus* in a culture medium under conditions that allow expression and secretion of the polypeptide in the culture medium. In an aspect, the culture medium comprises a non-carbohydrate micronutrient source, a buffering agent for maintaining pH of the medium at a range of about 5 to about 9, a non-hydrolyzable nonionic surfactant, and a nitrogen source. In other aspects, the method comprises cultivating the *Bacillus* at a temperature of about 30°C to about 42°C, about 35°C to about 40°C, or about 37°C. In a further aspect, the method further comprises adding an additive for selection to the culture medium in the presence of a vector containing a selectable marker. The strain of *Bacillus* for expression of CALB, WT or CALB variant, can be *B. subtilis, B. cereus, B. brevis, B. licheniformis, B. stearothermophilus, B. pumilis, B.amyloliquefaciens, B. clusii,* or *B. megaterium.*

[0015] The method of making CALB or CALB variant disclosed herein can further comprise obtaining supernatant from the culture medium and concentrating the supernatant comprising CALB or CALB variant. The supernatant can be concentrated by ultrafiltration.

[0016] The present disclosure further comprises methods for identifying amino acid mutations in CALB that alter lipase activity, wherein the method comprises (a) obtaining crystal structure of a wild type (WT) CALB; (b) introducing one or more amino acid mutations into the WT CALB to obtain a CALB variant; (c) obtaining

crystal structure of the CALB variant; (d) solvating the crystal structures of the WT CALB and the CALB variant into an implicit solvent; (e) performing molecular dynamic simulation on the structures to obtain resultant structures; (f) solvating the resultant structures into an explicit solvent; (g) performing molecular dynamic simulation on the resultant structures in the explicit solvent medium to obtain refined structures; (h) obtaining structural data for the refined structures of WT CALB and CALB variant; and (i) comparing the structural data obtained for the refined structures of the WT CALB and the CALB variant, to identify one or more amino acid mutations in CALB that alter lipase activity. In an aspect, the method disclosed herein comprises obtaining structural data for cavity volume of an active site of the WT CALB and the CALB variant. In other aspects, the method disclosed herein comprises obtaining structural data for solvent accessible surface area (SASA) of an active site of WT CALB and CALB variant. In a further aspect, the method comprises obtaining structural data for distance between center of mass of two residues of interest. The method disclosed herein further comprises correlating SASA, cavity volume, with the structure of CALB variant. The method further comprises correlating SASA of the CALB variant and its synthetic activity.

[0017] The method disclosed herein further comprises obtaining SASA of one or more amino acid positions of a catalytic triad of the WT CALB and obtaining SASA of one or more positions of a catalytic triad of the CALB variant, comparing the obtained SASA of an amino acid position of the catalytic triad of the WT CALB with the obtained SASA of a corresponding amino acid position of the catalytic triad of the CALB variant, and identifying a CALB variant that enhances lipase activity, such as its synthetic activity. The lipase activity can be a synthetic activity.

[0018] The present disclosure provides methods of catalyzing synthesis of a carboxylic acid ester, wherein one or more alcohols and one or more carboxylic acids or one or more carboxylic esters are reacted in the presence of a CALB variant disclosed herein to form a carboxylic acid ester. In embodiments, at least one or more carboxylic acids is a branched carboxylic acid, or at least one or more carboxylic acid esters are a branched carboxylic acid ester.

[0019] The present disclosure provides methods of catalyzing synthesis of an amide, wherein one or more carboxylic acids or carboxylic acid esters and one or more amines are reacted in the presence of a CALB variant disclosed herein to form an amide. In embodiments, at least one of the carboxylic acids is a branched

carboxylic acid or at least one of the carboxylic acid esters is a branched carboxylic acid ester.

BRIEF DESCRIPTION OF THE DRAWINGS

[0020] Figure 1 shows the vector map of pHT43. The vector includes: Pgrac promoter comprising groE promoter, lacO operator and gsiBSD sequence; ColE1 origin (ColE1 ori); ampicillin resistance (Amp^R), lac repressor (lacI gene); chloramphenicol resistance (Cm^R); and amyQ signal sequence (SamyQ).

[0021] Figures 2A and 2B show: (A) The comparison between new approach (case 3) and two conventional computation approaches for MD simulations (case 1 and case 2). The three dimensional structure of CALB and (B) cavity volume and solvent accessible surface area of catalytic triad of variant 554 resulting from MD simulations using case 1, 2 and 3.

[0022] Figure 3 shows the alignment of amino acid sequences of the native CALB sequence (22delta, SEQ ID NO: 2) with variants 529 (D223G and S227T), 578 (D223G, SEQ ID NO: 29), and 1001 (S227T, SEQ ID NO: 30).

[0023] Figures 4A and 4B show final simulation snapshots after MD refinement of WT and 529 structure.

[0024] Figure 5 shows final simulation snapshot of Variant 554. Dotted line is drawn around the catalytic triad.

DETAILED DESCRIPTION

[0025] *Candida antarctica* lipase B (CALB) is an enzyme able to catalyze hydrolysis of esters, and more significantly, synthesis reactions, such as direct esterification, transesterification, and amidation of a wide range of alcohols and acids. The native CALB synthesis activity needs improvement for use at a large scale or for non-specialty products, especially for hindered substrates. The terms “wild type” and “native” when referring to CALB are used interchangeably throughout the present disclosure.

[0026] The present disclosure is based in part on the identification of specific amino acid changes in the wild type (WT) or parent protein sequence of CALB that lead to an increase in the synthesis activity of the enzyme of greater than about two fold, about 2 to 15 fold, or about 3 to 12 fold over the WT CALB. Additionally, the present disclosure is based on the development of methods that allow for the

generation and characterization of these enzyme variants using novel computational methods and effective enzyme expression in a prokaryotic host, such as *Bacillus*.

[0027] The crystal structure of CALB has been solved experimentally and deposited in the RCSB Protein Data Bank (PDB) as structure 1TCA. 1TCA represents the structure of the native amino acid sequence of CALB in an aqueous environment, which would occur during synthesis and secretion of the enzyme by a microbial host, and also when the enzyme is catalyzing a hydrolytic reaction. However, when the enzyme is isolated, dried, and used to catalyze a condensation reaction, such as esterification or amidation, then an implicit “aqueous” solvent model is not applicable.

[0028] The active site of CALB is a triad consisting of residues S105, D187 and H224. The terms “active site” and “catalytic site” are used interchangeably throughout the present disclosure. The present disclosure is based in part on the discovery that the structural changes that accompany specific amino acid sequence changes, especially near the active site residues, result in a significant change to the protein structure when modelled in explicit solvent, mimicking the organic esterification reaction mixture. Provided herein are various CALB variants comprising amino acid alterations. As an example, the structure of the novel CALB variants 529 (D223G, S227T) and 554 (E188D, D223G, S227T) predict a more open substrate cavity than the WT CALB or the CALB variant with D223G, with better access to the active site in non-aqueous conditions, such as during synthesis reactions.

[0029] As used herein, the term CALB includes both WT CALB and CALB variants.

[0030] Protein engineering requires expression of a catalytically active protein in a microbial system. However, CALB is not expressed effectively in prokaryotic expression systems as a secreted protein. To address the limitation of eukaryotic and *E. coli* expression hosts, the present disclosure provides an improved expression system for CALB using *Bacillus subtilis* (Bsub) as the host. This expression system is suitable for high-throughput screening of enzyme variants. While there are no published reports of CALB expression in Bsub, the components of a suitable expression system can be assembled or built from components available from commercial suppliers, published reports and database sequences, or

requested from academic institutions and culture collections. Provided herein is a novel bacterial expression system for expressing large quantities of active CALB.

[0031] Expression system components include a host strain of *Bacillus*, an expression vector, expression media and growth conditions. The ideal expression system results in constitutive, high level expression of catalytically active protein secreted into the media without a deleterious impact on the host or the need to shuttle DNA between multiple hosts. The host should have a high transformation efficiency using the expression vector.

[0032] Any suitable *Bacillus* host may direct the expression of CALB, the WT or the variant form. In embodiments, enzyme expression in the host strain Bsub WB800N (MoBiTec) benefits from the deletion of eight extracellular proteases from the genome. In other embodiments, the CALB was expressed in strain BGSC 1S141. The expression vector can include a nucleic acid encoding CALB, a *Bacillus* origin of replication, a promoter, a secretion signal, and optionally a selection marker. As an example, the expression vector pHT43 allows inducible expression of the target protein as a translational fusion with an extracellular amylase, which is secreted via the sec system in *Bacillus*. Constitutive expression is preferred, as it eliminates the need to induce expression, and avoids any variation in expression caused by induction timing or conditions. Various methods for *B. subtilis* transformation are available, including the protocol supplied with the host strain, protocols based on electroporation and protoplast transformation, and protocols based on natural competence of *B. subtilis*. The latter method takes advantage of the natural competence of *B. subtilis* to incorporate DNA, and both closed circular and linear DNA can be successfully introduced. Long linear repeats of vector and insert can be made and assembled via PCR, and introduced directly into Bsub competent cells. In this case, the shuttle vector features (coliform origin of replication and antibiotic resistance marker for *E. coli* host) would be non-essential vector components.

[0033] Factors that impact CALB expression in the *Bacillus* system include media and conditions such as growth temperature, nitrogen source and content, the presence of a non-hydrolyzable surfactant, and the biomass density of the inoculum. Surprisingly, the presence of a carbohydrate carbon source or dense inoculum reduces lipase expression/activity in B subtilis. The secreted lipase is readily isolated from the culture supernatant and used to catalyze synthesis reactions.

Provided herein are novel culture media comprising a non-carbohydrate micronutrient source, a buffering agent for maintaining pH of the media at a range of about 5 to about 9, a non-hydrolyzable nonionic surfactant, and a nitrogen source

[0034] The disclosure is also based on the discovery of a new strategy to elucidate the key mutation sites via molecular dynamic (MD) simulations. This strategy provides a two-step approach: (1), MD simulation of a protein and/or its mutants are performed in implicit aqueous solvent condition and, followed by (2) a simulation using explicit solvents which are equivalent to the solvents used in experiment. Crystal structure of CALB that was used in the simulations was obtained from Protein Data Bank (PDB, Code: 1TCA) and mutations of residues in CALB were introduced using Discovery Studio 4.0 software (Accelrys Software Inc.). This approach permitted the discovery of front (I189 and I285) and side (E188 and L278) gates that directly control the access to the catalytic triad. It was also discovered that the distance between these gating residues can be controlled through specific mutations described herein. The mutations change the size of the catalytic cavity and accessibility of the catalytic triad and consequently control the activity. The analysis of the simulations that assisted this discovery include solvent accessible surface area of the catalytic amino acids, cavity volume, the distance between gating residues and stability of the CALB structure.

[0035] Provided herein are new computational protocols that enable accurate exploration of the effect of mutations and solvent environment on the structure of CALB that can be used for prediction of new mutation sites and obtaining novel variants with improved enzyme activity.

[0036] The present disclosure provides CALB variants having an amino acid sequence that is different from the WT CALB. The terms “CALB variant,” “modified CALB,” and “CALB mutant” are used interchangeably throughout the disclosure to refer to a CALB with an amino acid sequence that is different from the WT CALB. The CALB variant can have altered physical and functional activity as compared to a WT CALB. Physical activity can include stability, such as thermostability. Functional activity can include lipase activity, such as hydrolytic activity and synthetic activity, for example, catalyzing the amidation, direct esterification, and transesterification. The CALB variants can have improved activity for catalyzing synthesis reactions as compared to a WT CALB. The CALB variants can have any increase in activity for catalyzing synthesis reactions. The CALB variants can have approximately a 2 fold,

3 fold, 4 fold, 5 fold, 6 fold, 7 fold, 8 fold, 9 fold, 10 fold, 15 fold, or 20 fold increase in activity for catalyzing reactions, such as synthesis reactions. The CALB variants provided herein catalyzes synthesis reactions, such as amidation, direct esterification, and transesterification. The lipase activity of CALB variants can be measured by any synthesis assay that can determine the reactants or products of a condensation reaction is suitable. The esterification reaction to generate esters or amides such as octyl benzoate, octyl octanoate or octyl octanamide, where the depletion of reactants and appearance of products can be detected by liquid or gas chromatography. The depletion of reactants can also be monitored by titration or hydroxyl number. The synthesis activity can be measured as PLU or propyl laurate units (Chow et al., PLoS One 2012, 7(10), e47665). There is no external reference for the benzoate ester assay. The synthesis activity of CALB can be measured by the benzoic acid (BZA) esterification assay.

[0037] The term "CALB variant" as used herein refers to a CALB that has been modified to comprise an alteration, such as a substitution, insertion, and/or deletion, of one or more amino acid residues at one or more specific positions of the polypeptide of SEQ ID No: 2 (WT CALB). The nucleic acid or polynucleotide encoding the CALB variant can be obtained through human intervention by modification of the polypeptide coding sequence disclosed in SEQ ID No: 1 (WT CALB nucleic acid). The amino acid substitution, insertion, and/or deletion can be conservative or non-conservative. The CALB variant has an activity that is different from the WT CALB. The activity can be for catalyzing synthesis reactions.

[0038] It was a surprising discovery that amino acid alterations near the active triad site, S105, D187, and H224 result in CALB variants having altered activity or stability. The variants provided herein can have one or more, two or more, three or more, four or more, five or more, six or more, seven or more, eight or more, nine or more, or ten or more amino acid alterations. Provided herein are CALB variants having one or more amino acid alterations at position 141, 146, 188, 189, 223, 227, or 235 of SEQ ID NO: 2. In an aspect, the one or more amino acid substitutions is: a substitution at A141 to a threonine (T); a substitution at A146 to a threonine (T); a substitution at E188 to an aspartic acid (D); a substitution at I189 to a valine (V); a substitution at D223 to a glycine (G); a substitution at S227 to a threonine (T); and a substitution at V235 to an alanine (A). In other aspect, the CALB variants provided

herein have two to seven amino substitutions, wherein the substitutions are at position 141, 146, 188, 189, 223, 227, or 235 of SEQ ID NO: 2.

[0039] The present disclosure provides CALB variants as shown in Table 1.

Table 1: CALB Variants

Variant	Amino Acid Positions							SEQ ID NO:
	A141	A146	E188	I189	D223	S227	V235	
529					G	T		4
554			D		G	T		6
857	T		D		G	T		8
984	T	T	D		G	T		10
940	T		D	V	G	T		12
880	T		D		G	T	A	14
959	T	T	D		G	T	A	16
965	T	T	D	V	G	T		18
953	T		D	V	G	T	A	20
963	T	T	D	V	G	T	A	22

[0040] In embodiments, the present disclosure provides fusion proteins comprising a CALB variant disclosed herein and a heterologous peptide or polypeptide. The heterologous sequences can, for example, include sequences designed to facilitate purification, e.g. histidine tags, and/or visualization of recombinantly-expressed proteins. Other non-limiting examples of fusion proteins include those that permit display of the CALB variant on the surface of a phage or a cell, fusions to intrinsically fluorescent proteins, such as green fluorescent protein (GFP), fusion to signal peptides to direct polypeptide processing and export, fusion to cellulose binding module(s), fusion to dockerin domain(s), fusion to cohesion domain(s), fusion to fibronectin-like domain(s), fusions to the IgG Fc region, and the

like. The fusions can be direct or can be by way of intervening peptide linker regions/domains.

[0041] The present disclosure provides peptides and polypeptides that are fragments of CALB variants of the CALB variants disclosed herein. The peptides and polypeptides of the CALB variants disclosed herein are functionally active and have synthesis activity. In embodiments, these peptides include at least 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, or more contiguous amino acids of a CALB variant disclosed herein. As an example, the first 34 amino acids can be removed from a CALB variant disclosed herein. The peptides and polypeptides include at least the catalytic domain of a CALB variant and/or possess functional activity, such as synthesis activity. In certain embodiments, the peptides and polypeptides comprising the catalytic domain of the variant cellulolytic enzyme are provided and possess functional activity, such as synthesis activity. The functional activity of the peptides can be the same as the full length CALB variant.

[0042] Also disclosed herein are nucleic acids encoding CALB variants having a 2 fold, 3 fold, 4 fold, 5 fold, 6 fold, 7 fold, 8 fold, 9 fold, 10 fold, 15 fold, or 20 fold increase in activity for catalyzing synthesis reactions, such as esterification, amidation, and transesterification. The nucleic acid or polynucleotide can be a DNA or RNA. The DNA can be a cDNA.

[0043] In embodiments, the nucleic acids provided herein encode a CALB variant having an amino acid sequence as set forth in SEQ ID NO: 4, 6, 8, 10, 12, 14, 16, 18, 20, or 22. In other embodiments, the nucleic acids provided herein encoding a CALB variant has a nucleic acid sequence as set forth in SEQ ID NO: 3, 5, 7, 9, 11, 13, 15, 17, 19, or 21.

[0044] The CALB variants provided herein have at least about 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity with the amino acid sequence of the WT CALB, wherein the CALB variants have a 2 fold, 3 fold, 4 fold, 5 fold, 6 fold, 7 fold, 8 fold, 9 fold, 10 fold, 15 fold, or 20 fold increase in activity for catalyzing reactions, such as synthesis reactions. The CALB variants provided herein catalyzes synthesis reactions, such as esterification, amidation, and transesterification.

[0045] The relatedness between two amino acid sequences or between two nucleotide sequences is described by the parameter "identity". For purposes of the present disclosure, the degree of identity between two amino acid sequences is

determined using the Needleman-Wunsch algorithm as implemented in the Needle program of the EMBOSS package (EMBOSS: The European Molecular Biology Open Software Suite), preferably version 3.0.0 or later. The optional parameters used are gap open penalty of 10, gap extension penalty of 0.5, and the EBLOSUM62 (EMBOSS version of BLOSUM62) substitution matrix. The output of Needle labeled "longest identity" (obtained using the--nobrief option) is used as the percent identity and is calculated as follows:

(Identical Residues X 100)/(Length of Alignment-Total Number of Gaps in Alignment)

[0046] In embodiments, the nucleic acids encoding CALB variants provided herein have at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity with the nucleic acid encoding the WT CALB (SEQ ID NO: 1). Moreover, the nucleic acids encoding the CALB variants hybridize at least under low stringency conditions to SEQ ID NO: 1.

[0047] For purposes of the present disclosure, the degree of identity between two deoxyribonucleotide sequences is determined using the Needleman-Wunsch algorithm as implemented in the Needle program of the EMBOSS package, preferably version 3.0.0 or later. The optional parameters used are gap open penalty of 10, gap extension penalty of 0.5, and the EDNAFULL (EMBOSS version of NCBI NUC4.4) substitution matrix. The output of Needle labeled "longest identity" (obtained using the--nobrief option) is used as the percent identity and is calculated as follows:

(Identical Deoxyribonucleotides X 100)/(Length of Alignment-Total Number of Gaps in Alignment)

[0048] As disclosed herein, the CALB variants disclosed herein have an amino acid sequence comprising two or more amino acid substitutions at a position in a sequence corresponding to any of positions 141, 146, 188, 189, 223, 227 or 235 of SEQ ID NO: 2, wherein (a) the variant has about 2 fold to 20 fold increase in activity for catalyzing synthesis reactions such as esterification, amidation, and transesterification; (b) the variant has an amino acid sequence having at least 80%

sequence identity with SEQ ID NO: 2 (WT CALB); (c) the nucleic acid encoding the CALB variant hybridizes under at least low stringency conditions with the nucleic acid encoding WT CALB (SEQ ID NO: 1) or its complementary strand; and (d) the variant is encoded by a nucleic acid comprising a sequence having at least 80% identity with SEQ ID NO: 1.

[0049] In an aspect, the CALB variant is encoded by a nucleic acid sequence that hybridizes under very low stringency conditions, low stringency conditions, medium stringency conditions, medium-high stringency conditions, high stringency conditions, or very high stringency conditions with SEQ ID NO: 1 or its complementary strand. The stringency conditions are provided in Sambrook et al. (J. Sambrook, E. F. Fritsch, and T. Maniatis, 1989, Molecular Cloning, A Laboratory Manual, 2d edition, Cold Spring Harbor, N.Y.), which is incorporated by reference in its entirety.

[0050] In an aspect, the preparation of a variant CALB can be achieved by preparing a nucleic acid sequence by modifying the nucleic acid sequence encoding the WT CALB, transforming the nucleic acid sequence into a suitable host, and expressing the modified sequence to form the variant CALB. In other aspects, the variant CALB can be prepared by expressing the nucleic acid sequence encoding the WT CALB in an in vitro expression system without the need for a host. In embodiments the variant CALB can be chemically synthesized.

[0051] CALB variants can be prepared according to any mutagenesis procedure known in the art, such as site-directed mutagenesis, synthetic gene construction, semi-synthetic gene construction, random mutagenesis, shuffling, etc.

[0052] Site-directed mutagenesis is a technique in which one or several mutations are created at a defined site in a nucleic acid molecule encoding the WT CALB polypeptide. The technique can be performed in vitro or in vivo.

[0053] Site-directed mutagenesis can be accomplished in vitro by PCR involving the use of oligonucleotide primers containing the desired mutation. Site-directed mutagenesis can also be performed in vitro by cassette mutagenesis involving the cleavage by a restriction enzyme at a site in the plasmid comprising a nucleic acid encoding the WT CALB polypeptide and subsequent ligation of an oligonucleotide containing the mutation in the polynucleotide. Usually the restriction enzyme that digests at the plasmid and the oligonucleotide is the same, permitting sticky ends of the plasmid and insert to ligate to one another. See, for example, Scherer & Davis

(1979 PNAS USA 76:4951-4955); and Barton et al. (1990 Nucleic Acids Research 18:7349-4966).

[0054] Site-directed mutagenesis can be accomplished in vivo by methods known in the art. See, for example, U.S. Patent Application Publication 2004/0171154; Storici et al. (2001 Nature Biotechnology 19:773-776); Kren et al. (1998 Nat. Med. 4:285-290); and Calissano & Macino (1996 Fungal Genet. Newslett. 43:15-16). Any site-directed mutagenesis procedure can be used for preparing the CALB variants. There are many commercial kits available that can be used to prepare variants of a WT CALB.

[0055] Single or multiple amino acid substitutions, deletions, and/or insertions can be made and tested using known methods of mutagenesis, recombination, and/or shuffling, followed by a relevant screening procedure, such as those disclosed by Reidhaar-Olson & Sauer (1988 Science 241:53-57); Bowie & Sauer (1989 PNAS USA 86:2152-2156); WO95/17413; or WO95/22625. Other methods that can be used include error-prone PCR, phage display (Lowman et al. 1991 Biochem. 30:10832-10837; U.S. Pat. No. 5,223,409; WO92/06204) and region-directed mutagenesis (Derbyshire et al., 1986, Gene 46:145; Ner et al., 1988, DNA 7:127).

[0056] In various embodiments, nucleic acids encoding one or more of the CALB variants described herein are inserted into vectors suitable for expressing the CALB variants in a host cell. In such vectors, the nucleic acid sequence encoding the CALB variant is operably linked to one or more promoters and/or other regulatory sequences.

[0057] The term "operably linked" refers herein to a configuration in which a control sequence is appropriately placed at a position relative to the nucleic acid encoding the variant CALB sequence such that the control sequence influences the expression of a variant CALB polypeptide.

[0058] As used herein, the term "expression" includes any step involved in the production of the polypeptide including, but not limited to, transcription, post-transcriptional modification, translation, post-translational modification, and secretion.

[0059] An "expression vector" refers to a nucleic acid construct comprising a nucleic acid sequence (e.g., DNA sequence) that is operably linked to a suitable control sequence capable of effecting the expression of the nucleic acid in a suitable host. Such control sequences may include a promoter to effect transcription, an

optional operator sequence to control transcription, a sequence encoding suitable ribosome-binding sites on the mRNA, and sequences that control termination of transcription and translation. Different cell types are typically used with different expression vectors. For example, an illustrative promoter for vectors used in *Bacillus spp.* is the groE, AprE, or Pgrac promoter; an illustrative promoter used in *E. coli* is the Lac promoter, an illustrative promoter used in *Saccharomyces spp.* is PGK1, an illustrative promoter used in *Aspergillus spp.* glaA, and an illustrative promoter for *Trichoderma spp.* is cbhI. In embodiments 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, under suitable conditions, integrate into the genome itself. Expression techniques are known in the art and are described generally in, for example, Sambrook.

[0060] Nucleic acids encoding the CALB variants described herein can be incorporated into any of a variety of expression vectors suitable for expressing a polypeptide. Any vector that expresses genetic material into a cell can be used.

[0061] When incorporated into an expression vector, the nucleic acid sequence encoding the desired CALB variant is operatively linked to an appropriate transcription control sequence (promoter) to direct mRNA synthesis, e.g., T5 promoter. Examples of such transcription control sequences include the cauliflower mosaic virus (CaMV) and figwort mosaic virus (FMV), SV40 promoter, *E. coli* lac or trp promoter, phage lambda P_L promoter, tac promoter, T7 promoter, and the like. In bacterial host cells, suitable promoters include the promoters obtained from the *E. coli* lac operon, *Streptomyces coelicolor* agarase gene (dagA), a gene from a *Bacillus* sp., such as, for example, the *Bacillus subtilis* levansucranse gene (sacB) or gro E gene, the *Bacillus licheniformis* alpha-amylase gene (amyl), the *Bacillus* megaterium InhA gene, the *Bacillus stearothermophilus* maltogenic amylase gene (amyM), the *Bacillus amyloliquefaciens* alpha-amylase gene (amyQ), *Bacillus subtilis* xylA and xylB genes, the xylose promoter (Pxyl) from *Bacillus* megaterium, the promoter obtained from the prokaryotic beta-lactamase gene, and so forth.

[0062] In various embodiments, an expression vector optionally contains a ribosome binding site for translation initiation, and a transcription terminator, such as PinII. The vector also optionally includes appropriate sequences for amplifying expression, such as an enhancer. The vector also includes regions for vector replication, as examples, ORF-1, ORF-2, and ORF-3 for replication in *bacillus.*

[0063] In various embodiments the vector or DNA construct may also generally include a signal peptide coding region that codes for an amino acid sequence linked to the amino terminus of a polypeptide and which directs the encoded polypeptide into the cell's secretory pathway. Suitable signal peptides include, but are not limited to the *Bacillus* megaterium penicillin G acylase signal peptide sequence.

[0064] Other illustrative signal peptide coding regions for bacterial host cells may be obtained from the genes of *Bacillus* NCIB 11837 maltogenic amylase, *B. stearothermophilus* alpha-amylase, *B. licheniformis* subtilisin, B. beta-lactamase, B. *stearothermophilus* neutral proteases (nprT, nprS, nprM) and *B. subtilis* prsS. Further illustrative signal sequences are described in Simonen and Palva (1993, Microbiological Reviews 57: 109-137). Effective signal peptide coding regions for filamentous fungal host cells include but are not limited to the signal peptide coding regions obtained from *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* neutral amylase, *Aspergillus niger* glucoamylase, *Rhizomucor miehei* asparatic proteinase, *Humicola insolens* cellulase and *Humicola lanuginosa* lipase. Variants of these signal peptides and other signal peptides are suitable, as well as expression mutants thereof having one or more silent mutations.

[0065] In various embodiments the expression vectors optionally contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells. Suitable marker genes include those coding for antibiotic resistance such as, ampicillin, kanamycin, chloramphenicol, or tetracycline resistance. Further examples include the antibiotic spectinomycin or streptomycin (e.g., the aada gene), the streptomycin phosphotransferase (SPT) gene coding for streptomycin resistance, the neomycin phosphotransferase (NPTII) gene encoding kanamycin or geneticin resistance, the hygromycin phosphotransferase (HPT) gene coding for hygromycin resistance. Additional selectable marker genes include dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, and tetracycline or ampicillin resistance in E. coli.

[0066] The vector may further contain genetic elements that facilitate integration by either homologous or non-homologous recombination. Genetic elements that facilitate integration by homologous recombination have sequence homology to targeted integration sites in the genomic sequence of the desired expression host cell. Genetic elements or techniques which facilitate integration by non-homologous

recombination include restriction enzyme-mediated integration (REMI), transposon-mediated integration, and other elements and methods that are well known in the art.

[0067] In embodiments, the nucleic acid sequence encoding the CALB variants described herein can also be fused, for example, in-frame to nucleic acids encoding a secretion/localization sequence, to target polypeptide expression to a desired cellular compartment, membrane, or organelle of a cell, or to direct polypeptide secretion to the periplasmic space, to the cell membrane or cell wall, or into the cell culture media. Such sequences are known to those of skill, and include secretion leader peptides, organelle targeting sequences (e.g., nuclear localization sequences, endoplasmic reticulum (ER) retention signals, mitochondrial transit sequences, peroxisomal transit sequences, and chloroplast transit sequences), membrane localization/anchor sequences (e.g., stop transfer sequences, GPI anchor sequences), and the like.

[0068] The expression vector includes elements for inducible or constitutive expression of the CALB variant.

[0069] "Host strain" or "host cell" means a suitable host for expressing nucleic acids comprising an expression vector as described herein. Illustrative host cells include prokaryotic or eukaryotic hosts, including any transformable microorganism in which expression can be achieved. The host cell may be chosen from eukaryotic or prokaryotic systems, such as for example bacterial cells, (Gram negative or Gram positive), yeast cells (for example, Saccharomyces cereviseae or Pichia pastoris), animal cells (such as Chinese hamster ovary (CHO) cells), plant cells, and/or insect cells using baculovirus vectors. In some embodiments, the host cells for expression of the polypeptides include, and are not limited to, those taught in U.S. Pat. Nos. 6,319,691, 6,277,375, 5,643,570, or 5,565,335, each of which is incorporated by reference in its entirety, including all references cited within each respective patent.

[0070] Examples of host strains include, but are not limited to, *Bacillus, Escherichia coli, Trichoderma reesei, Saccharomyces cerevisiae, Aspergillus niger,* and the like. Specific examples of *Bacillus* include *B. subtilis, B. cereus, B. brevis, B. licheniformis, B. stearothermophilus, B. pumilis, B.amyloliquefaciens, B. clusii,* or *B. megaterium.*

[0071] Recombinant expression vectors can be introduced into host cells to produce a transformed host cell. These cells are useful experimental systems. Accordingly, the present disclosure provides a host cell comprising a recombinant

expression vector for expression of the CALB variants disclosed herein. The term "transformed host cell" is intended to include prokaryotic and eukaryotic cells which have been transformed or transfected with a recombinant expression vector of the invention. Prokaryotic cells can be transformed with nucleic acid by, for example, electroporation or calcium-chloride mediated transformation. Nucleic acid can be introduced into mammalian cells via conventional techniques such as calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofectin, electroporation or microinjection. Suitable methods for transforming and transfecting host cells can be found in Sambrook et al., and other such laboratory textbooks. Suitable host cells include a wide variety of prokaryotic and eukaryotic host cells. For example, the peptides of the invention may be expressed in bacterial cells such as *E. coli, Pseudomonas, Bacillus subtillus*, insect cells (using baculovirus), yeast cells or mammalian cells. Other known suitable host cells can also be used.

[0072] The nucleic acids encoding the CALB variant can be codon optimized for the host used to express the CALB variant.

[0073] The present disclosure also provides a novel expression system for expressing WT CALB and CALB variant. The expression system is based on a bacterial expression system using a *Bacillus* host, for expressing large quantities of active CALB polypeptide. Examples of *Bacillus* host include but are not limited to *B. subtilis, B. cereus, B. brevis, B. licheniformis, B. stearothermophilus, B. pumilis, B.amyloliquefaciens, B. clusii,* and *B. megaterium.*

[0074] The bacterial expression system disclosed herein provides the following benefits over other known expression systems for eukaryotic lipases like CALB:

- There is no need to codon-optimize the protein coding region for the host.
- There is no need to shuttle expression constructs through an intermediate host like E. coli.
- The lipase is effectively expressed and secreted directly into the media, both liquid and solid, so there is no need to include a cell lysis step.
- The lipase can be constitutively expressed, eliminating the need for induction.
- The lipase is correctly folded into a form able to catalyze hydrolysis and synthesis reactions, including transesterification, direct esterification. and amidation reactions.

- Maximum expression levels can be achieved in as little as 18 hours of liquid culture.

[0075] The new bacterial system is able to express the lipase under a variety of conditions, and in a relatively simple and inexpensive media. Factors that impact lipase expression in the system include growth temperature, nitrogen source and content, the presence of a non-hydrolyzable surfactant, and the biomass density of the inoculum. Surprisingly, the presence of a carbohydrate carbon source or dense inoculum reduces lipase expression/activity.

[0076] For expressing secreted recombinant CALB in a prokaryotic host of the genus Bacillus, it was discovered that the expression media should contain a source of micronutrients and vitamins, a complex nitrogen source, a non-hydrolyzable nonionic surfactant, and optionally an additive for selection, if a selectable marker is present on the expression vector. It was also discovered that a carbohydrate source, such as glucose, sucrose, glycerol, and the like, should not be present in the expression media.

[0077] The expression media base disclosed herein comprises a source of micronutrients and vitamins, such as yeast extract. The concentration of yeast extract can vary from about 0.1% to about 5%, about 0.2% to about 4%, about 0.3% to about 3%, about 0.4% to about 2% or about 0.5% to about 1%. Synthetic defined media may also be sufficient as a base, such as that described in Demain (Minimal media for quantitative studies with Bacillus subtilis, 1956, Journal of Bacteriology, vol 75 p. 517).

[0078] The expression media disclosed herein comprises a complex nitrogen source. Complex nitrogen sources include yeast extract, tryptone, hydrolyzed casein, casamino acids, peptone, soy peptone, nutrient broth, and meat extract. A defined mixture of amino acids can also be used. The concentration of nitrogen source can vary from about 0.1% to 5%, about 0.2% to about 4%, about 0.3% to about 3%, or about 0.5% to about 2%.

[0079] The expression media disclosed herein comprises a non-hydrolyzable nonionic surfactant. Typical surfactants used in biological systems, such as Span and Tween, contain ester bonds that can be hydrolyzed by the expressed lipase, rendering them non-functional as a surfactant. Instead, the surfactant additive must not be hydrolysable, but must also not denature the expressed enzyme. Such surfactants include the block-copolymer surfactants, also known as poloxamers

(Pluronics™). Examples of poloxamer-type block copolymer include but are not limited to Pluronic L31™, L35™, and F68™. The poloxamer surfactants differ in monomer composition and molecular weight. The poloxamers tested improved recovered activity of secreted lipase. The concentration of the non-hydrolyzable surfactant in the media is about 0.01% to about 1%, about 0.02% to about 0.8%, about 0.03% to about 0.6%, about 0.04 % to about 0.4%, or from 0.05 to about 0.2%.

[0080] An additive for selection may be included in the media, if a selectable marker is present on the expression vector.

[0081] A fermentable carbohydrate source (glucose, sucrose, glycerol, etc) should be absent from the expression media.

[0082] Provided herein is an expression media comprising a source of micronutrients and vitamins, a complex nitrogen source, a non-hydrolyzable nonionic surfactant, and optionally an additive for selection.

[0083] It was also discovered that the optimal expression conditions include buffering the expression media to a neutral pH and maintaining the expression system at a temperature of about 30°C to about 42°C.

[0084] In an aspect, the expression system comprises an expression media buffered at between about pH 5 and about pH 9, between about pH 6 and about pH 8, or at about pH 7.

[0085] In other aspects, the expression system is maintained at a temperature of between about 30°C to about 42°C, about 35°C to about 40°C, or about 37°C for expression of CALB.

[0086] The present disclosure provides a method for expressing a CALB comprising the use of a novel expression system for optimal expression of CALB. The expression system comprises an expression vector for expressing a CALB in *Bacillus*. The CALB is expressed under optimal conditions such as in a novel culture medium buffered to a neutral pH and at a temperature of between about 30°C to about 42°C.

[0087] Provided herein are expression vectors for expressing a CALB in *Bacillus* comprising a nucleic acid encoding a CALB, one or more origins of replication for replication in *Bacillus*, a promoter, a secretion signal, and optionally a selectable marker.

[0088] Provided herein are culture media comprising about 0.1% to about 5% of a non-carbohydrate micronutrient source and vitamins, a buffering agent for maintaining the pH at about 6 to about 8, about 0.01% to about 1% of a non-hydrolyzable nonionic surfactant, and about 0.1% to about 5% of a nitrogen source, All amounts of the components of the culture media are relative to the total weight of the composition.

[0089] In embodiments, the methods for expressing CALB provided herein further comprise concentrating the expressed CALB. In an aspect, the expressed CALB can be concentrated by precipitation or chromatography. In other aspects, the expressed CALB is concentrated by ultrafiltration through a filter with a molecular weight cut-off smaller than the size of the enzyme.

[0090] The present disclosure provides compositions comprising CALB variants and a carrier. The composition can be a pharmaceutical composition, in which case the carrier is a pharmaceutically acceptable carrier.

[0091] The embodiments of any of the products (CALB variants), compositions, or methods disclosed herein can consist of or consist essentially of--rather than comprise/include/contain/have--any of the described steps, elements, and/or features. Thus, in any of the claims, the term "consisting of" or "consisting essentially of" can be substituted for any of the open-ended linking verbs recited above, in order to change the scope of a given claim from what it would otherwise be using the open-ended linking verb. As used herein, the term “consisting essentially of” indicates that the product or composition necessarily includes the listed ingredients and is open to unlisted ingredients that do not materially affect the basic and novel properties of the invention. As an example, peptides of CALB variants can consist of or consist essentially of a number of contiguous amino acids of a full length amino acid sequence.

[0092] The present disclosure provides methods for analyzing molecular interactions in solution for accurate determination of the effect of mutations and solvent environment on the structure of CALB. The information obtained from the analysis can be used to predict new mutation sites for improving enzyme activity.

[0093] In embodiments, the present invention provides computational tools and structure-function analysis such as molecular dynamics (MD) simulations for obtaining structural data including cavity volume of a group of amino acids and solvent accessibility of functional groups (SASA) in a protein, such as CALB and its

variants. MD numerically solve Newton's equations of motion of atoms in the molecular system to obtain information on its time-dependent properties, which gives an insight into conformational changes of bio-materials, such as proteins and DNA. For protein engineering, MD simulations have been widely used as tools for evaluating the structural properties of residues in proteins and for the selection of key mutation sites for better enzyme activity and stability.

[0094] In an aspect, the approach employed involves combining structural information of the CALB polypeptide derived from X-ray crystallography with computational modeling and simulation of the wild type and variant forms of the catalytic triad domain. This approach enabled obtaining structural and functional data of the potential role of individual amino acids and their cooperative action in the catalytic triad domain. The structural and functional data obtained from multiple variants along with the empirical data of stability and other chemical and physical properties of CALB and its variants permitted a determination of the correlation of structure and function and the identification of specific amino acid responsible for structural changes around the catalytic pocket of CALB.

[0095] Previous computational studies have mainly used water as a solvent to explain the effect of mutations in organic solvents or to suggest mutation sites for the experiment which has to be performed in non-aqueous systems, such as trans-esterification reactions. Same approach has been used for MD simulations on *Candida Antarctica* Lipase B (CALB) and its mutants: simulations have been always performed in aqueous solutions regardless of the solvents used in experiment. Even though water is a major contributor to a protein's 3-D structure and therefore this conventional approach has helped in the guidance of mutation studies, it is expected that using water as a solvent for MD simulations cannot represent the rational structures of proteins or their mutants in non-aqueous solutions, thus leading to inaccurate selections of important mutation sites.

[0096] The present disclosure provides a novel approach to MD simulation of CALB and its variants. The method provided herein involves a two step approach to accurately determine important mutation sites for improved CALB activity. MD simulation of CALB and its variants are first performed in an implicit aqueous solvent condition. Subsequently, a MD simulation is performed using explicit solvents, which are equivalent to the solvents used in synthesis reactions that CALB catalyzes, such as esterification, amidation, and transesterification.

[0097] The present method employs MD to simulate the CALB structure, to evaluate the intrinsic dynamic nature of the structure, especially the active site, in an aqueous environment and in the organic (explicit) solvents. Molecular dynamics simulations track the dynamic trajectory of a molecule resulting from motions arising out of interactions and transient forces acting between all the atomic entities in the protein and its local environment, in this case the atoms constituting the CALB or its variants and their surrounding water (implicit environment) and/or organic (explicit environment) molecules. This analysis provides an understanding of the differences of each CALB variant as compared to the WT CALB, with respect to properties, such as solvent accessible surface area (SASA) of the catalytic amino acids, cavity volume, distance between gating residues, and stability of the CALB structure, thus allowing identification of amino acids for mutation to improve functional activity, for example, synthesis reaction of CALB.

[0098] The method disclosed herein is based on extensive expertise in modeling CALB and its variants and the use of molecular dynamics (MD) simulations to evaluate the influence of solvents on WT CALB activity. The studies performed identified specific amino acids responsible for structural changes around the catalytic pocket of CALB and correlated these changes with the experimental observations. As an example, it was shown that access to the catalytic site and the volume of the cavity is correlated with the activity of the enzyme.

[0099] CALB is an enzyme catalyst used for the production of esters. Using the method disclosed herein, variants having lipase activity, such as synthesis activity, were produced. These variants were shown, as an example, to esterify benzoic and/or 2-ethylhexanoic acid. The methods disclosed herein provides identification of more variants with improved activity or substrate preference more easily.

[00100] Accordingly, understanding the structure/function relationship of CALB variants is important to predicting their utility in making esters. Experimental data collected for the active variants based on model esterification reactions, and the sequence variations can be mapped onto the native protein structure. The method disclosed herein provides an understanding of the basis of the activity changes, and ultimately to more accurately predict other variants for study.

[00101] The method provided herein is an improved method for performing molecular dynamics (MD) simulations of WT CALB and its variants. The method

provides more accurate information on structure/function relationship CALB which enables discovery of CALB variants with improved synthesis activity.

[00102] The data obtained from MD simulations are analyzed and compared. In an aspect, the MD simulations provides solvent accessible surface area (SASA) of the catalytic amino acids, cavity volume, the distance between gating residues, and stability of the CALB structure for correlating structure/function relationship of CALB. In other aspects, information provided by the MD simulation assists in the determination of how an amino acid alteration affects the size of the catalytic cavity and accessibility of the catalytic triad which ultimately affects the activity of CALB.

[00103] The methods provided herein show that the MD simulations based on the novel approach of performing MD simulation in first the implicit solvent followed by MD simulation in the explicit solvent are useful for investigating structure/function relationship of CALB and its variants. Amino acid residues can be accurately predicted for mutagenesis based on the results from the simulations. The methods provided herein allows the accurate identification of specific amino acid sites for alteration and the production of other CALB variants with improved synthesis activity.

[00104] Further, the expression systems for CALB provided herein enables the experimental validation of the mutants obtained based on the data obtained by the simulations.

[00105] The following examples illustrate exemplary methods provided herein. These examples are not intended, nor are they to be construed, as limiting the scope of the disclosure. It will be clear that the methods can be practiced otherwise than as particularly described herein. Numerous modifications and variations are possible in view of the teachings herein and, therefore, are within the scope of the disclosure.

EXAMPLES

Example 1. Expression of CALB and variants in a prokaryotic host

[00106] An expression system for the eukaryotic enzyme CALB was developed and optimized in a *Bacillus subtilis* host. The lipase is expressed well under a variety of conditions, and in a relatively simple and inexpensive media. Factors that impact lipase expression in the system include growth temperature, nitrogen source and content, the presence of a non-hydrolyzable surfactant, and the biomass density of

the inoculum. Surprisingly, the presence of a carbohydrate carbon source or dense inoculum reduces lipase expression/activity.

[00107] The secreted lipase is readily isolated from the culture supernatant by absorption and retains synthesis activity. It is possible to concentrate the media by ultrafiltration and retain >90% of the lipase activity, suggesting that a concentrated enzyme solution is feasible for commercially relevant biocatalyst production. Other options for improving expression include replacing the Pgrac inducible promoter with a strong constitutive promoter that is active after the culture enters stationary phase. Fed batch or chemostat culture conditions can also be used to maintain the culture in the optimum growth phase for enzyme production.

Detecting CALB lipase expression in Bacillus subtilis

[00108] The native *Candida antarctica* lipase B (CALB) coding region was codon-optimized for yeast, synthesized, and cloned as a translation fusion into the pHT43 expression vector (Fig. 1). The *Bacillus subtilis* host strain WB800N (MoBiTec) was used throughout. Genotype: *trpC2* (Trp -) *nprE aprE epr bpr mpr :: ble nprB :: bsr .vpr wprA :: hyg cm :: neo;* NeoR

[00109] Secreted lipase expression in *Bacillus subtilis* was verified on plates and in liquid culture. On plates, cells expressing an active lipase (CALB) grew and generated a zone of clearing when plated on a thin overlay of an opaque ester emulsion. Suitable indicator esters for reporting lipase activity include medium chain triglycerides (MCT), octyl octanoate (1%) and Tween 80 (0.1 to 1%) added to solid media (LB agar, 50mM phosphate buffer pH 7, 6 mg/L chloramphenicol). In vector pHT43, it is necessary to induce protein expression from the construct with IPTG, at 0.5 to 1mM (or 25 ul per plate of 100mg/ml IPTG added to cells before plating).

Optimization of inducible CALB lipase expression in Bacillus subtilis

[00110] In a series of experiments, a WT CALB lipase expression construct 22, which is sequence-identical to the mature native CALB, was expressed in *Bacillus subtilis*. The inoculum, growth media and induction regime were varied. When LB agar (LBA) is indicated in the inoculum column, the inoculum was a cell suspension made from a bacterial lawn grown on solid media overnight. The inoculated liquid cultures were allowed to grow 2-6h (155 rpm, 37°C, baffled flask) before lipase expression was induced with the addition of 1mM IPTG. The cells were separated

from the expression medium 20-26h post induction by centrifugation (10 min., 10,000Xg, 10ºC). Soluble lipase in the culture supernatant was adsorbed to the resin as described. The initial esterification reaction used a 10% enzyme loading (100 mg dry resin for 1g of reactant mix; a 1:1 molar ratio of 2-ethylhexanol:oleic acid), stirred in a closed vial at 55ºC, 800 rpm. Active enzyme preparations were also analyzed in reactions at lower enzyme loadings (5% and 2% resin). The reaction mix was analyzed by GC, and 2-ethylhexyl oleate (EHO) formation was estimated as peak area%.

[00111] A review of the results in Table 1.1 suggests that including the non-hydrolyzable surfactant Pluronic L61 at 0.1% and buffering the media at pH 7 with 50mM phosphate buffer for the inoculum and optionally for the expression media leads to the highest lipase activity (Rows 1, 8-11). Surprisingly, glycerol as a supplemental carbon source at 2 or 0.5% did not promote lipase expression (Rows 3, 4, 13, 15, 17). Addition of other surfactants or esters to either the inoculum or the expression medium had no apparent benefit. A buffered growth medium containing 1% yeast extract and 1% tryptone (YT) gave slightly better results than LB (0.5% yeast extract, 1% tryptone, 1% NaCl) routinely used for propagation of *E. coli* and Bsub (Row 16 vs 12). Media containing 1% yeast extract and 1% casamino acids (YC) was also suitable for lipase expression (Rows 14 & 18).

Table 1.1. Survey of growth conditions for inducible lipase expression in *Bacillus subtilis*

Row	Inoculum	Growth media	Induced at	Grown for	ml /g resin	% EHO at 24h, resin loading	
						10 %	5 %
1	LBA chlor6 + 50mM PO4 +0.1% L61	LB chlor6 +0.1% L61	4h	24h	20	**72**	**50**
2	LBA chlor6 + 50mM PO4	LB chlor6 +0.1% L61 +25mM oct-oct	4h	24h	20	**59**	

	+0.1% L61						
3	LBA chlor6 + 50mM PO4 +0.1% L61	LB chlor6 +0.1% L61 +2% glycerol	4h	24h	20	**16**	
4	LBA chlor6 + 50mM PO4 +0.1% L61	LB chlor6 +0.1% L61 +2% glycerol +25mM oct-oct	4h	24h	20	**13**	
5	LB chlor6 (overnight culture)	LB chlor6 +0.1% L61	6h	20h	26	**71**	**52**
6	LB chlor6 (overnight culture)	LB chlor6 +0.1% L61 +25mM oct-oct	6h	20h	26	**29**	
7	LBA chlor6	LB chlor6 +0.1% L61	2h	22h	28	**24**	
8	LBA chlor6 + 50mM PO4 +0.1% L61	LB chlor6 +0.1% L61	2h	22h	28	**79**	**64**
9	LBA chlor6 + 50mM PO4 +0.1% Tween80 +25mM oct-oct	LB chlor6 +0.1% L61	2h	22h	28	**74**	
10	LBA chlor6 + 50mM PO4 +0.1% L61 +0.1% Tween80 +25mM oct-oct	LB chlor6 +0.1% L61	2h	22h	28	**76**	
11	LBA chlor6 + 50mM PO4 +0.1% L61	LB chlor6 +0.1% L61 +50mM PO4	2h	22h	28	**79**	**56**
12	LBA chlor6 + 50mM PO4 +0.1% L61	LB chlor6 +0.1% L61 +50mM PO4	4h	20h	26		**61**
13	LBA chlor6 +	LB chlor6 +0.1% L61	4h	20h	26		**42**

	50mM PO4 +0.1% L61	+50mM PO4 +0.5% glycerol					
14	LBA chlor6 + 50mM PO4 +0.1% L61	YC chlor6 +0.1% L61 +50mM PO4	4h	20h	26		**68**
15	LBA chlor6 + 50mM PO4 +0.1% L61	YC chlor6 +0.1% L61 +50mM PO4 +0.5% glycerol	4h	20h	26		**19**
16	LBA chlor6 + 50mM PO4 +0.1% L61	YT chlor6 +0.1% L61 +50mM PO4	4h	20h	26		**73**
17	LBA chlor6 + 50mM PO4 +0.1% L61	YT chlor6 +0.1% L61 +50mM PO4 +0.5% glycerol	4h	20h	26		**65**
18	LBA chlor6 + 50mM PO4 +0.1% L61	YC chlor6 +0.1% L61 +50mM PO4	4h	20h	25		**66**
19	LBA chlor6 + 50mM PO4 +0.1% L61	YC chlor6 +0.1% L61 +50mM PO4 + 100mg/L tryptophan	4h	20h	25		**65**

Constitutive expression of CALB lipase in Bacillus subtilis

[00112] Changes to the pHT43 expression vector were made to simplify lipase expression. In the pHT43 vector, expression of the target gene (lipase) is controlled by the Pgrac promoter. Pgrac is an artificial promoter consisting of *Bacillus subtilis* groE promoter, lac operator and gsiB ribosome binding site. This is a strong IPTG-inducible promoter for *Bacillus subtilis*. Pgrac is repressed by the lacI repressor encoded on the pHT43 plasmid vector, and requires induction by the lactose analog IPTG. To allow constitutive lipase expression in *Bacillus subtilis*, the lacI repressor was deleted from the construct using existing restriction sites SnaBI and SfoI. The pHT43 expression vector (MobiTec) was modified for constitutive protein expression by deleting the laqI repressor encoded on the vector to generate pHT43D. The

1.1kb SnaBI / SfoI fragment containing the laqI gene in pHT43 was deleted and confirmed by PCR. The repressor deletion expression constructs in pHT43D were readily propagated in *B. subtilis* with no obvious growth impairment, and the lipase was expressed constitutively from the Pgrac promoter without the need for IPTG induction. The lacI deletion construct could be propagated in *E. coli* strains that contain an episomal copy of the lacI repressor. The lacI repressor was deleted from a construct expressing the native CALB sequence (strain 22) and verified by restriction mapping. The lacI deletion construct containing the native CALB coding region (22Δ) was transformed into *Bacillus subtilis*.

[00113] The repressor deletion WT lipase construct (22Δ) was expressed in *Bacillus subtilis* and compared to the original lipase construct (22). The culture with the repressor deletion construct (22Δ) was uninduced, while the culture with the original construct 22 was induced with 1 mM IPTG at 2h. In both cases, the inoculum was a cell suspension made from a lawn of the strain grown on LB agar + 50mM phosphate buffer, pH 7.0 and 0.1% Pluronic L61. After a total of 24h of growth, lipase in the culture supernatant was immobilized on resin (1g Lewatit VP OC1600 resin/25ml media) and 2-ethylhexyl oleate (EHO) esterification reactions were performed as described previously. Results of the optimization of lipase expression in the repressor deletion construct are summarized in Table 1.2. The repressor deletion construct resulted in lipase expression identical to that obtained from the original construct after induction by 1mM IPTG. The lacI deletion vector was a great improvement over the inducible system because it eliminates a process step, and the inducer IPTG is a costly and hazardous reagent.

Table 1.2. Lipase activity expressed from the expression construct with the lacI repressor (22) and without (22Δ).

Expression construct (NB p. 100, 104)	Relative activity (%EHO at 24h 2% resin)
22, 1mM IPTG @ 2h	100%
22Δ (n=6) uninduced	97-104

Conditions: 1% Yeast extract; 1% Tryptone; 50mM phosphate buffer pH7 (Hydrion). Sterile filter; add 0.1% Pluronic L61 and 6 mg/L chloramphenicol. Inoculate and grow 24h (37 deg C, 160 rpm). Immobilize: 25 ml culture supernatant/g resin.

Optimizing growth media for CALB expression in Bacillus

[00114] Focusing on the constitutive WT CALB expression construct with the lacI deletion (22Δ), growth conditions for the expression culture were investigated. The standard media recipe contains only two nutrient sources: yeast extract and tryptone (an enzymatic digest of casein, a milk protein), plus phosphate buffer, antibiotic and non-hydrolyzable surfactant. The relative amounts of each nutrient were varied, and the pre-mixed phosphate buffer preparation was replaced with its individual components. The growth temperature was also varied. Table 1.3 summarizes the results of the growth conditions on relative lipase esterification activity.

[00115] Surprisingly, addition of the carbon sources glycerol (Table 1.1) or glucose (Table1.3) to the media reduced lipase expression or activity. An inverse relationship between cell biomass and lipase activity is evident when the inoculum density is increased **(Table 1.3),** with a higher biomass inoculum leading to lower lipase expression/activity. Reducing tryptone to 0.5% or eliminating yeast extract reduced the lipase expression/activity, as did changing the temperature from 37°C. Reducing the yeast extract content to 0.5% had no impact on lipase activity, but would reduce the cost of the media.

Table 1.3. Media and temperature effects on lipase activity from 22Δ

Row	Media/growth condition change*	Relative synthesis activity (%EHO at 24h, 2% resin)
1	Standard conditions*	100%
2	Substitute Pluronic L31 for L61	117
3	Substitute Pluronic L35 for L61	99-121
4	Substitute Pluronic F68 for L61	104
5	10mM calcium nitrate	93
6	10mM magnesium sulfate	103

7	Glucose, 1%	65
8	Glucose 1%, Sodium citrate 1%, MgSO4 10mM	85
9	Standard inoculum	100
10	2X inoculum	65
11	5X inoculum	48
12	1% Yeast extract, 1% Tryptone	100
13	0.5% Yeast extract, 0.5% Tryptone	74
14	0.5% Yeast extract, 1% Tryptone	103
15	0.2% Yeast extract, 1% Tryptone	90
16	0% Yeast extract, 1% Tryptone	75
17	Grow liquid culture at 37°C	100
18	Grow liquid culture at 30°C	80
19	Grow liquid culture at 42°C	62

* Standard conditions: Expression media: 1% Yeast extract; 1% Tryptone; 50mM phosphate buffer pH7 (Hydrion). Sterile filter; add 0.1% Pluronic L61 and 6 mg/L chloramphenicol. Inoculate using a cell suspension from a lawn grown on solid media (LBA chlor6 + 50mM PO4 +0.1% L61) and grow 24h (37°C, 160 rpm). Immobilize: 25 ml culture supernatant/g resin.

Concentrating Bacillus subtilis-expressed lipase activity

[00116] Commercial enzyme solutions are typically concentrated, such as by precipitation or chromatography, and stabilized prior to distribution. On large scale, ultrafiltration can be used to concentrate an enzyme solution through a filter with a molecular weight cut-off smaller than the size of the enzyme. In theory, if the *Bacillus subtilis* culture supernatant containing the secreted lipase can be concentrated, then the immobilized activity on a resin dry weight basis can be increased. A 50 ml culture of *Bacillus subtilis* constitutively expressing CALB (22Δ) was grown using the standard conditions. After 24h, 25 ml of culture supernatant was used directly for immobilization, as for the standard protocol. The other 25ml was first cleared through a 0.2um filter to remove any cells, then concentrated to a final volume of 2ml through a 10,000 MW cutoff PES filter (Corning Spin-X UF 20, 7500Xg, 15°C, 80 minutes). The concentrate was diluted in water to 25ml before

adding resin. The culture supernatant concentrate retained 91% of the lipase activity, while the filtrate had no activity above background.

EXAMPLE 2: Comparison of native CALB structure in implicit and explicit solvent

[00117] In this Example, MD simulations are used to study the CALB variants that are already isolated.

[00118] This Example include a comparison between newly developed methods (case 3) and conventional methods (case 1 and case 2). See Figures 2A and 2B.

[00119] Implicit aqueous solvent models can represent the protein structure in an aqueous environment. For implicit solvents (case 2 and case 3), all starting structures (wild type, 529 and 554) were subjected to 10,000 steps minimization followed by heating the system to 300 K in 100 ps using Langevin thermostat. The MD simulations were carried out for 12 ns and temperature was maintained at 300 K with Berendsen thermostat.

[00120] Explicit solvent models (case 1 and case 3) represents the protein structure in an environment that mimics the esterification reaction mix (3:1:1:1 molar ratio of octanol : octanoic acid : 2-ethylhexanoic acid : benzoic acid). All solvent structures (octanol, octanoic acid, 2-ehtylhexanoic acid, and benzoic acid) were made by Discovery Studio Visualizer 4.0 and their partial charges were obtained using Antechamber with AM1-BCC method in AMBER 12 package. The solvent box (3:1:1:1 reaction mix) for each variant were made via Xleap in Ambertools. As a starting structure in explicit solvent, all CALB (wild type, 529, 554) were extracted from a last frame of simulations using implicit solvent model and solvated into aforementioned solvent box. In all cases, the systems were carefully minimized and an equilibrated by 11 stages starting from minimization for 10,000 steps while CALB was restrained for 200 kcal/mol. In the same restrained state, temperature was gradually increased to 300 K in 40 ps. A short NPT MD simulation was performed for 200 ps under the 200 kcal/mol constraint on CALB. Then another minimization for 10,000 steps followed with 20 kcal/mol restraint on CALB. Additional short NPT simulation was carried out for 20 ps with same constraint energy. Four consecutive minimization stages were performed for 1,000 steps by gradually decreasing constraint of CALB from 20 kcal/mol to 0 kcal/mol. As a final equilibration step, the system was reheated to 300 K without any constraints of CALB for 40 ps,

respectively. After careful minimization and equilibration steps, production MD runs were performed under the NPT-ensemble for 60 ns at 300 K, 1 atm.

[00121] As shown in Figures 2A and 2B, the structure of Variant 554 changed as a function of the method used. In case 1, where variant 554 was only simulated in explicit solvent (reaction mixture), no significant structural changes near the catalytic cavity were observed and entire structure was almost identical to the wild type. In case 2, where variant 554 was simulated in only aqueous solution, conformational changes near catalytic cavity resulted in narrower cavity entrance with strong interactions between E188D and L278. In case 3, when MD simulations followed the experimental procedures, Variant 554 showed a wide open cavity entrance and high solvent accessible surface area of catalytic triad. This serves as an example illustrating that key mutation sites for better enzyme activity, such as E188D can be overlooked if conventional simulation approaches (using only water or organic solvents as a solvent) were used in this study.

Example 3. Isolation of variant 529 with two amino acid changes (D223G, S227T) & comparative variants with single amino acid changes only (D223G) or (S227T)

[00122] Random changes within 3 amino acids of the active site residues were targeted in the mature native CALB amino acid sequence. A very large pool of sequence variants was screened for an active lipase phenotype on plates containing an indicator ester. A single variant showed a significant increase in benzoic acid esterification: variant 529, with two amino acid changes from the native CALB sequence.

[00123] Figure 3 shows the amino acid sequences of CALB variants aligned with the native CALB sequence (22Δ). Variant 529 contains two residue changes (D223G and S227T) near the active site residue H224. Based on the sequence of variant 529, comparative variants were generated that contained only one amino acid change, either D223G (variant 578) or S227T (variant 1001) alone.

[00124] The variants were expressed in the *Bacillus subtilis* WB800N host grown at 25ml scale, and immobilized for synthesis activity. The relative benzoic acid esterification activity (BZA) and the 2-ethylhexyl palmitate synthesis activity (2-EHP) were compared between native and variant expressed sequences. The synthesis

activity of variant 529 with two amino acid substitutions is greater than either single amino acid substitution.

Variant	Screening activity, BZA	Screening activity, 2-EHP
Empty vector	0	4
WT (22Δ)	1.0	10
1001	2.4	18
578	2.5	19
529	3.0	24

[00125] BZA Esterification Activity: The standard benzoic acid (BZA) synthesis screening reaction mix contained 3 mole equivalents of octanol, and 1 equivalent each of octanoic acid, 2-ethylhexanoic acid and benzoic acid. Dry resin (50 mg) was weighed into a 3 dram glass vial, and 1g of reaction mix was added along with a stir bar. The reactions were stirred (800rpm) at 55ºC for 24h, then 25ul was sampled into 1ml of methanol for analysis by GC. Relative conversion was estimated by integrating only the octanol and three ester peaks. Total esterification was expressed as the % peak area of the combined esters to the total peak area, while the esterification of 2-ethylhexanoic acid and benzoic acid was expressed relative to octanoic acid (ester peak areas only).

[00126] The plasmid from variants 529, 578, and 1001 was isolated and re-transformed into the *Bacillus subtilis* WB800N host. At least 4 independent colonies were expressed in cultures at 25ml scale, and immobilized separately for synthesis activity. The benzoic acid esterification activity (BZA) and the 2-ethylhexyl palmitate synthesis activity (2-EHP) were compared. The synthesis activity of variant 529 with two amino acid substitutions is consistent among multiple independent transformants and greater than either single amino acid substitution.

Variant	Avg Synthesis Activity BZA (n)

1001	2.0 (4)
578	2.6 (6)
529	3.4 (8)

[00127] Methods: The template for mutagenic PCR was a WT CALB coding region cloned into the modified pHT43Δ vector. Vector primers 2619_F and 2990_R were used along with the appropriate mutagenic primers to generate the short overlapping fragments containing the targeted amino acid changes, using Taq DNA polymerase (NEB). Fragments were amplified using internal and vector primer pairs. The pHT43Δ vector fragment (with the laqI deletion for constitutive expression) was amplified with primers (SIGSEQ_R and G341_F) using a high-fidelity polymerase blend (LongAmp Taq, NEB) to generate a linear vector fragment that included a sequence that overlapped the sequence-modified lipase gene fragments. Then the vector and insert fragments were combined in a multimerization reaction (LongAmp Taq, no added primers), in which the overlapping ends of the vector and insert fragments served to prime the extension reaction, creating long linear repeats of alternating insert and vector. These multimers appeared as very high molecular weight DNA on an agarose gel and could be used to transform *B. subtilis* directly.

Primer sequences, 5' to 3'

2619_F	GCTTGGTACCAGCTATTGTAACATAATCG (SEQ ID NO: 23)
2990_R	CAGACAAAGATCTCCATGGACGCGTG (SEQ ID NO: 24)
G341_F	GGTAAGAGAACTTGTTCTGGTATTGTTACTCCATAATAACCC (SEQ ID NO: 25)
SIGSEQ_R	CAGCGTGCACATAAGCACAAGTCTGAACGAAACTGTCCGC (SEQ ID NO: 26)

[00128] The PCR-assembled multimers were effective for direct transformation of *B. subtilis* competent cells, strain WB800N, and selected on plates containing chloramphenicol and an indicator ester to visualize colonies expressing a hydrolytically active lipase. Indicator plates contained LB agar, 50mM phosphate buffer pH 7.0, 6mg/L chloramphenicol, 0.2% Tween 80, and 0.05% PEG6000 distearate. Active lipase appeared as a white or opaque halo surrounding a colony.

Plasmid was isolated from any leads strains and the lipase gene amplified from the vector using primers pHT2619_F and pHT2990_R. The 1.2 kb fragment was G-50 column purified and submitted for sequencing with the amplification primers (Eurofins MWG Operon). The DNA sequences were analyzed (trimmed, translated and aligned) using the programs in the DNASTAR core suite.

[00129] For lipase expression and immobilization at 25ml scale, colonies were picked into a small volume of Expression Media, and the cell suspension was first plated on L61 plates (LB agar, 50mM phosphate buffer pH 7.0, 0.1% Pluronic L61, 6 mg/L chloramphenicol) and incubated overnight at 37ºC. Expression Media (~2 ml) was added to the plate surface and the cells scraped off to create a dense cell suspension. This cell suspension was used to inoculate 25 ml of expression media. After 24h of growth (37ºC, 160 rpm), the cells were pelleted by centrifugation (10 min, 10,000xg, 10ºC), and the supernatant was decanted to an 8-dram glass vial. To immobilize the lipase from the cleared broth, 0.85g of resin (Purolite Lifetech™ ECR 1030M) was added to each vial, and the vials were rotated horizontally overnight at room temperature. The resin was transferred to a paper filter and rinsed twice with water. The resin was air-dried, and used in synthesis reactions to determine relative synthesis activity.

[00130] The standard benzoic acid (BZA) synthesis screening reaction mix contained 3 mole equivalents of octanol, and 1 equivalent each of octanoic acid, 2-ethylhexanoic acid and benzoic acid. Dry resin (50 mg) was weighed into a 3 dram glass vial, and 1g of reaction mix was added along with a stir bar. The reactions were stirred (800rpm) at 55ºC for 24h, then 25ul was sampled into 1ml of methanol for analysis by GC. Relative conversion was estimated by integrating only the octanol and three ester peaks. Total esterification was expressed as the % peak area of the combined esters to the total peak area, while the esterification of 2-ethylhexanoic acid and benzoic acid was expressed relative to octanoic acid (ester peak areas only).

[00131] The 2-ethylhexyl palmitate (2-EHP) reaction contained 5 grams of palmitic acid and 2.5 grams of 2-ethylhexanol and 50 mg of dry resin. The reactants were melted at 72ºC for 15 minutes, then the reaction proceeded at 72ºC for 4h with stirring, then 25ul was sampled into 1ml of methanol for analysis by GC.

[00132] GC Analysis: Analyzed on an Agilent 6890 GC using an Agilent DB-5 column (#122-5032; 30m x 0.25mm x 0.25uM) with flame ionization detection and

split injection with 2 uL injection volume. Temperature program: 100°C for 10 min, then 100-250°C at a rate of 25°C/min, hold at 250°C for 9 min.

EXAMPLE 4: Explicit solvent model of Variant 529 and identification of E188 as a target for variation

[00133] A crystal structure of CALB was obtained from Protein Data Bank (PDB, Code: 1TCA) and mutations of amino acids were introduced via Discovery Studio Visualizer 4.0 software (Accelrys, USA) (*Discovery Studio Modeling Environment*, (2007) Accelrys Software Inc., San Diego). In Variant 529, two amino acids, D223 and S227, were replaced with Glycine and Threonine, respectively. Two CALB structures, wild type (WT) and Variant 529 were solvated into implicit solvent, which represents the protein structure in an aqueous environment and simulated for 12 ns. The resultant structures of the wild type and Variant 529 were solvated with explicit reaction mixtures which is considered identical reaction media as used in aforementioned experiments. Then MD simulations in explicit reaction media were performed for 100 ns to refine the structure of the wild type and 529 variant in a reaction mix. AMBER 12 with FF12SB (for proteins) and GAFF (for reaction mixtures) force fields was used for MD simulations.

Table 4.1. Summary of structural analysis of WT and Variant 529 after MD refinement with implicit aqueous solvent and explicit reaction mixture.

Variant	Cavity volume (Å^3)	Solvent accessible surface area (Å^2)			Enzyme activity (Ratio to WT)
		SER-105	ASP-187	HIS-224	
WT	356.8	5.45 (±1.23)	0.00 (±0.01)	4.87 (±2.05)	1.0
529	490.5	6.73 (±6.25)	0.01 (±0.03)	17.91 (±7.90)	3.5

[00134] Since reaction mixtures in this study were directly used as substrates for synthesis reactions, cavity volume and solvent accessibility of catalytic triad can be important factors for enzyme activity.

[00135] From the refined 3-D structure of two cases, the volume of catalytic cavity was measured via Caver catalyst software. As shown in Table 4.1, Variant 529 has

a larger cavity volume (356.8 Å^3) as compared to the cavity volume of WT (490.5 Å^3).

[00136] Solvent accessible surface area (SASA) was defined as the surface area of molecules that is accessible to solvent molecules and the SASA of the catalytic triad residues was calculated through GETAREA (University of Texas Medical Branch, USA). Simulation results showed that SASA of H224 in Variant 529 case was approximately 4-fold greater than that of the WT.

[00137] A larger cavity volume and higher SASA of catalytic triad residues illustrated that the catalytic triad of Variant 529 has a higher chance to interact with solvent molecules in the reaction mixture thus leading to greater enzyme activity as compared to WT.

[00138] Final simulation snapshots indicated that two amino acids, E188 and L278 tend to interact with each other in WT case but this interaction was partially broken in Variant 529 which resulted in the exposure of catalytic triad to the solvent as well as made a larger cavity volume (Figures 4A and 4B**)**. In both cases, E188 and L278 acted as the gate residues: closed cavity conformations in WT case and open cavity structures in Variant 529. The interactions between E188 and L278 and its conformations were quantified via analysis of the average distance between center of mass (COM) of two residues (Table 4.2). This analysis was performed by CPPTRAJ module in AMBER 12 package and the last 10 ns of simulation trajectories were used for obtaining average distance between these residues with the standard deviations. As a reference, the same analysis was performed for CALB the crystal structure. Even though an open cavity structure was observed in Variant 529, it was found that two catalytic residues, H224 and D187 were structurally screened by E188. Also L278 can frequently interact with E188 during the simulation of Variant 529 forming a closed cavity conformation. For these reasons, E188 was selected as a target for the variation.

Table 4.2. The relation between catalytic gate opening and distance between E188 and L278 amino acids.

Case	Distance between COM of 188 and 278 (with a standard deviation) (Å)	Gate structure
Crystal	8.77 (n/a)	Closed

CALB		
WT	8.54 (± 0.61)	Closed
529	9.96 (±1.07)	Open

Example 5. Isolation of variant 554 (E188D, D223G, S227T)

[00139] Position E188 was predicted to be a good candidate for change based on the structural models in Example 4. This residue was allowed to vary to any amino acid using degenerate primers. Only lipase variants with a further E188D change resulted in an increase in synthesis activity over the parent variant 529.

[00140] Variant 554 was expressed in the *Bacillus subtilis* WB800N host grown at 25ml scale, and immobilized for synthesis activity. The relative benzoic acid esterification activity (BZA) and the 2-ethylhexyl palmitate synthesis activity (2-EHP) were compared between native and variant expressed sequences. The synthesis activity of variant 554 with three amino acid substitutions is greater than the parent sequence with two amino acid changes.

Variant	Screening activity (BZA)	2-EHP
Empty vector	0	4
WT (22Δ)	1.0	10
529	3.0	24
554	6.0	33

[00141] The plasmid from variant 554 was isolated and re-transformed into the *Bacillus subtilis* WB800N host. At least 4 independent colonies were expressed in cultures at 25ml scale, and immobilized separately for synthesis activity. The benzoic acid esterification activity (BZA) was determined. The synthesis activity of variant 529 with two amino acid substitutions is consistent among multiple independent transformants and greater than either single amino acid substitution.

Variant	Avg BZA

	(n)
554	5.7 (8)

[00142] Cloning, expression and analysis were performed as in Example 3. The template for mutagenic PCR was variant 529. Fragments were amplified using primer pairs (E222X_F plus 2990R) and (2619_F plus L233wt_R), where E188 was varied to any amino acid. The pHT43Δ vector fragment was amplified with primers (SIGSEQ_R and G341_F) using a high-fidelity polymerase blend (LongAmp Taq, NEB) to generate a linear vector fragment that included a sequence that overlapped the sequence-modified lipase gene fragments. Then the vector and insert fragments were combined in a multimerization reaction with no added primers, and used to transform *B. subtilis* WB800N directly.

Primer sequences, 5' to 3'

E222X_F	GTATTCTGCCACCGATNNNATCGTCCAACCACAAGTTTCT (SEQ ID NO: 27)
L233wt_R	CCATTAAACAAGTAAGAAGAATCCAATGGAGAGTTAG (SEQ ID NO: 28)

EXAMPLE 6: Explicit solvent model of Variant 554 with analyses

[00143] Additional E188D mutation in the Variant 529 background (Variant 554) was employed via Discovery Studio Visualizer 4.0 software (*Discovery Studio Modeling Environment*, (2007) Accelrys Software Inc., San Diego). The same simulation procedures described in Example 2 (case 3) were used: MD simulation in implicit solvent condition for 12 ns followed by additional MD simulation in explicit reaction mixtures for 100 ns were used for the structure refinement of Variant 554. As stated in Example 5, E188 was chosen as a next target for the mutation because it can block the accessibility of catalytic triad to the solvent as well as frequently interact with L278 forming closed cavity conformations. Since stability of the positively charged catalytic residue H224 can be strongly affected by charged neighbors, a negatively charged residue near H224, E188, was replaced with a smaller but negatively charged residue, Asp (D).

[00144] Varying E188 to D caused a wide-open catalytic cavity with a completely exposed catalytic triad (Figure 5). These significant structural changes resulted in approximately 7-fold higher SASA of catalytic triad and a 2-fold larger cavity volume of Variant 554 as compared to those of WT CALB (Table 6.1). In Variant 529, the interactions between E188 and L278 were the key factor for open cavity conformation as well as were directly related to the enzyme activity. Since E188 was replaced with the smaller residue in Variant 554, the interaction between D188 and L278 became weaker than that between E188 and L278 in Variant 529. This weak interaction led to a complete separation between these residues with larger distance (Table 6.2) and, therefore, L278 became buried in the enzyme structure due to its hydrophobicity. Analysis of the simulations for other variants indicated a strong correlation between activity and open gate structure or large distance between 188 and 278 (Table 6.2).

Table 6.1. Assessment of structural differences between WT and Variant 554.

Variant	Cavity volume (Å^3)	Solvent accessible surface area (Å^2)			Enzyme activity (Ratio to WT)
		SER-105	ASP-187	HIS-224	
WT	356.8	5.45 (±1.23)	0.00 (±0.01)	4.87 (±2.05)	1.0
554	612.6	5.51 (±3.60)	1.92 (±1.36)	33.25 (±8.98)	5.9

Table 6.2. The relation between catalytic gate opening and distance between E188 and L278 amino acids.

Variant	Distance between 188 and 278 (with a standard deviation) (Å)	Enzyme Activity (Ratio to WT)	Gate Structure
569	4.30 (± 0.92)	<1	Closed
WT	5.22 (± 0.94)	1	Closed
551	7.35 (± 0.91)	2.5	Open / Closed
529	9.48 (± 1.13)	3.0	Open
554	10.63 (± 1.06)	5.9	Open
953	10.13 (± 0.79)	10.0	Open

Example 7. Octyl octanoate synthesis kinetics of WT, 578, 529, 554

[00145] WT CALB and lipase variants 578, 529 and 554 were expressed in *Bacillus subtilis* as described in Example 3, and the secreted lipase was adsorbed from the broth onto Purolite™ ECR1030M resin. The dried resin was used in

synthesis reactions to compare the synthesis activities of the different lipase sequences.

[00146] The following were added to a 3 dram glass vial: stir bar, X grams 1-octanol (see table below) and Y grams octanoic acid (see Table 7.1). Each reaction condition (A, B, C, D) was prepared in triplicate. Vials were stirred and maintained at reaction temperature of 60°C for 15 minutes.

Table 7.1. Reactant mixtures – various equivalent ratios of 1-octanol & octanoic acid

Reaction	Mole Equivalents 1-Octanol	Mole Equivalents Octanoic acid	mmol 1-Octanol	X = g 1-Octanol	mmol 1-Octanoic acid	Y = g 1-Octanoic acid
A	1	2	3.51	0.46	7.02	1.01
B	1	1	3.50	0.46	3.50	0.5
C	2	1	7.02	0.91	3.51	0.51
D	3	1	10.54	1.37	3.51	0.51

[00147] Reactions were initiated by adding carefully pre-weighed dry resin with immobilized lipase (~30mg each) into each reaction vial. Upon addition of lipase to each vial a stopwatch was immediately started. Vials were staggered in 30 second intervals to allow for consistent sampling. 50μl was transferred from each vial to pre-weighed GC vials every 10 minutes for a total of 60 minutes for each reaction vial.

[00148] GC vials with added sample were then weighed and individual sample weights utilized to calculate micromoles octyl octanoate produced at each time point. A calibration curve was generated prior to reaction sampling using 2mg/ml 1-methylnaphthalene in toluene as an internal standard. A calibration curve for octyl octanoate demonstrated linear response on the GC in the concentration range of 1.0 to 40.0 mg/ml octyl octanoate.

[00149] GC analysis: to each GC vial containing the above 50μl sample were added 500μl 2mg/ml 1-methylnaphthalene (in toluene) as internal standard and 1000μl toluene. Vials were capped and mixed by inversion. Each vial was analyzed on an Agilent 6890 gas chromatograph equipped with a 5% diphenyl / 95% dimethyl

polysiloxane capillary (20m length, 0.25mm ID, 0.25µm film thickness, Restek RTX-5 Cat# 10223-124) GC column and a flame ionization detector. Initial temperature was held at 100°C for 3 minutes, then a gradient from 100°C to 325°C over 5.63 minutes. Temperature was held at 325°C for 2.38 minutes for a final run time of 11.0 minutes.

[00150] Reaction rate of esterification was defined as micromoles octyl octanoate produced over time. The slope of each initial esterification rate (defined as reaction velocity) was divided by mg dry resin with immobilized lipase to yield a specific activity: micromoles octyl octanoate minute^{-1} gram resin^{-1}. The initial synthesis rate of lipase variants 529 and 554 is 3.4 to 9.4 times greater than the WT lipase activity. The fastest rate was measured when the reactants were present in a 1:1 molar ratio.

Micromoles octyl octanoate minute^{-1} gram resin^{-1}

A) 1:2 octanol:acid	WT	578	529	554
Avg (st dev)	100.2 (5.7)	279.0 (17.7)	343.7 (31.3)	750.4 (53.1)
Ratio to WT	1.0	2.8	3.4	7.5

B) 1:1 octanol:acid	WT	578	529	554
Avg (st dev)	84.2 (6.0)	320.6 (31.9)	447.4 (47.2)	794.1 (56.3)
Ratio to WT	1.0	3.8	5.3	9.4

C) 2:1 octanol:acid	WT	578	529	554
Avg (st dev)	92.1 (4.5)	284.3 (2.4)	412.7 (26.5)	591.4 (8.6)
Ratio to WT	1.0	3.1	4.5	6.4

D) 3:1 octanol:acid	WT	578	529	554
Avg (st dev)	82.1 (5.6)	232.9 (2.3)	405.3 (3.9)	580.9 (24.8)
Ratio to WT	1.0	2.8	4.9	7.1

Example 8. 2-ethylhexyl palmitate synthesis kinetics of WT, 578, 529, 554

[00151] In a 50ml conical tube, 50mg of dry resin with immobilized lipase and 10 grams of Palmitic Acid were weighed into each tube, 1.03 equivalents (5.23 grams) of 2-Ethylhexanol was added. The tube was heated in 72°C water bath for 45 minutes to melt all starting materials. Once melted each tube was placed in a heat block set at 72°C. A sparge needle was inserted into each tube through a vented cap to mix and nitrogen-strip the reaction. The sparge rate was set to 300mL/min dry nitrogen for each tube.

[00152] Samples were taken at 0, 1h, 2h, 4h, and 24h and accurately weighed for analysis by a wt% GC method. GC analysis: Analyzed on a Hewlett-Packard 5890 GC equipped with a J&W DB-5 column, 30 m x 0.25 mm with flame ionization detection and split injection with 2 uL injection volume. Temperature program: 100°C for 9 min, then 100-300°C at a rate of 25°C/min, hold at 300°C for 10 min.

[00153] The initial rate of conversion (% palmitic acid converted/h) was determined from the slope of the data points from 0 to 4h. The conversion rate relative to WT is also reported. Lipase variants with 2 (529) or 3 (554) amino acid changes have a 2-ethylhexyl palmitate synthesis rate 3.5 to 5.9-fold greater than WT CALB.

	WT	578	529	554
% palmitic acid converted to ester (h-1)	0.72	1.84	2.54	4.27
r squared	0.6	0.99	1	0.93
ratio to WT	1.0	2.6	3.5	5.9

Example 9. Octyl octanamide synthesis kinetics of WT, 554

[00154] The following were added to a 3 dram glass vial: stir bar, 2.26 grams 1-octylamine and 2.77 grams methyl octanoate. Each reaction was performed in duplicate. Vials were stirred and maintained at reaction temperature of 60°C for 15 minutes. A nitrogen sparge (250 ml/min) was used to remove the methanol by-product.

[00155] Reactions were initiated by adding carefully pre-weighed dry resin with immobilized lipase (~30mg each) into each reaction vial. Upon addition of lipase to each vial a stopwatch was immediately started. Vials were staggered in 30 second

intervals to allow for consistent sampling. 50µl was transferred from each vial to pre-weighed GC vials every 30 minutes for a total of 180 minutes for each reaction vial.

[00156] GC vials with added sample were then weighed and individual sample weights utilized to calculate micromoles octyl octanamide produced at each time point. A calibration curve was generated prior to reaction sampling using 2mg/ml 1-methylnaphthalene in toluene as an internal standard. A calibration curve for octyl octanoamide demonstrated linear response on the GC in the concentration range of 1.0 to 40.0 mg/ml octyl octanoamide.

[00157] GC analysis: to each GC vial containing the above 50µl sample, added 500µl 2mg/ml 1-methylnaphthalene (in toluene) as internal standard and 1000µl toluene. Vials were capped and mixed by inversion. Each vial was analyzed on an Agilent 6890 gas chromatograph equipped with a 5% diphenyl / 95% dimethyl polysiloxane capillary (20m length, 0.25mm ID, 0.25µm film thickness, Restek RTX-5 Cat# 10223-124) GC column and a flame ionization detector. Initial temperature was held at 100°C for 3 minutes, then a gradient from 100°C to 325°C over 5.63 minutes. Temperature was held at 325°C for 3.38 minutes for a final run time of 12.0 minutes.

[00158] Reaction rate of esterification was defined as micromoles octyl octanamide produced over time. The slope of each initial amidation rate (defined as reaction velocity) was divided by mg dry resin with immobilized lipase to yield a specific activity: micromoles octyl octanamide minute^{-1} gram resin^{-1}. The initial synthesis rate of WT lipase and variant 554 is 1.9 times greater than the WT lipase activity.

Micromoles octyl octanamide minute^{-1} gram resin^{-1}

n=2	Empty vector	WT	554
Avg	1.4	38.5	73.5
Ratio to WT		1.0	1.9

Example 10: Random mutations combine to improve synthesis activity in 554 background

[00159] Four additional amino acid substitutions were identified following random mutagenesis and screening and were combined in the 554 background to further improve synthesis activity. Changes A141T, A146T, I189V, and V235A were identified by random mutagenesis using the screening protocol described in Example

3. These changes appeared to increase CALB synthesis activity independently and to different degrees in different backgrounds. The systematic combination of random changes in the 554 background (E188D, D223G, S227T) resulted in eight variants with a range of activities (variants 857, 984, 940, 880, 959, 965, 953, and 963). Variant 963 with all 7 amino acid changes shows the highest benzoic acid esterification activity.

Variant	Average Benzoate Synthesis Activity (n=4)	A141	A146	E188	I189	D223	S227	V235
857	7.2	T		D		G	T	
984	7.3	T	T	D		G	T	
940	8.4	T		D	V	G	T	
880	7.8	T		D		G	T	A
959	8.5	T	T	D		G	T	A
965	9.8	T	T	D	V	G	T	
953	10.0	T		D	V	G	T	A
963	11.9	T	T	D	V	G	T	A

[00160] **Random mutagenesis:** The template for mutagenic PCR was a WT CALB or variant coding region cloned into the modified pHT43Δ vector. Vector primers 2619_F and 2990_R were used as primers. Taq DNA polymerase (NEB). The Taq polymerase buffer was supplemented with an additional 2.5mM MgCl2 and 0.1mM MnSO4 to promote nucleotide changes in the coding region. The pHT43Δ vector fragment (with the laqI deletion for constitutive expression) was amplified with primers (SIGSEQ_R and G341_F) using a high-fidelity polymerase blend (LongAmp Taq, NEB), then the vector and mutagenized insert fragments were combined in a multimerization reaction (LongAmp Taq, no added primers), in which the overlapping ends of the vector and insert fragments served to prime the extension reaction, creating long linear repeats of alternating insert and vector. These were used to transform *B. subtilis* strain WB800N directly. Colonies expressing an active lipase

were selected for screening at 2ml scale, and the insert from high-activity leads was sequenced to determine the resulting amino acid sequence.

[00161] **Benzoic acid esterification activity (BZA):** *Bacillus subtilis* WB800N was transformed with the CALB variant expression constructs corresponding to variants 857, 984, 940, 880, 959, 965, 953, and 963. Four separate colonies were picked into expression media for each variant, and expressed at 25ml scale and immobilized as described in Example 3. The standard benzoic acid (BZA) synthesis screening reaction results were determined for the 4 independent transformants, and the average activity is reported in the Table above.

Primer sequences, 5' to 3'

2619_F	GCTTGGTACCAGCTATTGTAACATAATCG (SEQ ID NO: 23)
2990_R	CAGACAAAGATCTCCATGGACGCGTG (SEQ ID NO: 24)
G341_F	GGTAAGAGAACTTGTTCTGGTATTGTTACTCCATAATAACCC (SEQ ID NO: 25)
SIGSEQ_R	CAGCGTGCACATAAGCACAAGTCTGAACGAAACTGTCCGC (SEQ ID NO: 26)

[00162] **Example 11. Octyl octanoate synthesis kinetics of variants 857, 984, 940, 880, 959, 965, 953, and 963**

[00163] CALB variants 857, 984, 940, 880, 959, 965, 953, and 963 were expressed in *Bacillus subtilis* as described in Example 3, and the secreted lipase was adsorbed from the broth onto Purolite ECR1030M resin. The dried resin was used in synthesis reactions to compare the octyl octanoate synthesis activities of the different lipase sequences as described in Example 7.

[00164] The following were added to a 3 dram glass vial: stir bar, 0.46 grams (3.5 mmoles) 1-octanol and 0.5 grams (3.5 mmoles) octanoic acid. The reaction was prepared in triplicate, and run and analyzed as in Example 7.

[00165] Reactions were initiated by adding carefully pre-weighed dry resin with immobilized lipase (~30mg each) into each reaction vial. Upon addition of lipase to each vial a stopwatch was immediately started. Vials were staggered in 30 second

intervals to allow for consistent sampling. 50μl was transferred from each vial to pre-weighed GC vials every 10 minutes for a total of 60 minutes for each reaction vial.

[00166] GC vials with added sample were then weighed and individual sample weights utilized to calculate micromoles octyl octanoate produced at each time point. A calibration curve was generated prior to reaction sampling using 2mg/ml 1-methylnaphthalene in toluene as an internal standard. A calibration curve for octyl octanoate demonstrated linear response on the GC in the concentration range of 1.0 to 40.0 mg/ml octyl octanoate.

[00167] GC analysis: to each GC vial containing the above 50μl sample was added 500μl 2mg/ml 1-methylnaphthalene (in toluene) as internal standard and 1000μl toluene. Vials were capped and mixed by inversion. Each vial was analyzed on an Agilent 6890 gas chromatograph equipped with a 5% diphenyl / 95% dimethyl polysiloxane capillary (20m length, 0.25mm ID, 0.25μm film thickness, Restek RTX-5 Cat# 10223-124) GC column and a flame ionization detector. Initial temperature was held at 100˚C for 3 minutes, then a gradient from 100˚C to 325˚C over 5.63 minutes. Temperature was held at 325˚C for 2.38 minutes for a final run time of 11.0 minutes.

[00168] Reaction rate of esterification was defined as micromoles octyl octanoate produced over time. The slope of each initial esterification rate (defined as reaction velocity) was divided by mg dry resin with immobilized lipase to yield a specific activity: micromoles octyl octanoate minute^{-1} gram resin^{-1}. The initial synthesis rate of lipase variants 857, 984, 940, 880, 959, 965, 953, and 963 is 3.4 to 9.4 times greater than the WT lipase activity.

Micromoles octyl octanoate minute^{-1} gram resin^{-1}

1:1 octanol:acid	WT	857	984	940	880	959	965	953	963
Avg (st dev)	84.2 (6)	364 (15)	320 (9)	370 (22)	405 (1)	441 (24)	532 (47)	526 (19)	645 (4)
Ratio to WT	1.0	4.3	3.8	4.4	4.8	5.2	6.3	6.2	7.7

Example 12. 2-ethylhexyl palmitate synthesis kinetics of variants 857, 984, 940, 880, 959, 965, 953, and 963

[00169] The 2-EHP assay was performed as described in Example 8. In a 50ml conical tube, 50mg of dry resin with immobilized lipase and 10 grams of Palmitic Acid were weighed into each tube, 1.03 equivalents (5.23 grams) of 2-Ethylhexanol was added. The tube was heated in 72ºC water bath for 45 minutes to melt all starting materials. Once melted each tube was placed in a heat block set at 72ºC. A sparge needle was inserted into each tube through a vented cap to mix and nitrogen-strip the reaction. The sparge rate was set to 300mL/min dry nitrogen for each tube.

[00170] Samples were taken at 0, 1h, 2h, 4h, and 24h and accurately weighed for analysis by a wt% GC method. GC analysis : Analyzed on a Hewlett-Packard 5890 GC equipped with a J&W DB-5 column, 30 m x 0.25 mm with flame ionization detection and split injection with 2 uL injection volume. Temperature program: 100ºC for 9 min, then 100-300ºC at a rate of 25ºC/min, hold at 300ºC for 10 min.

[00171] The initial rate of conversion (% palmitic acid converted/h) was determined from the slope of the data points from 0 to 4h. The conversion rate relative to WT is also reported. Lipase variants with 4 to 7 amino acid changes have a 2-ethylhexyl palmitate synthesis rate 2 to 3-fold greater than WT CALB.

	WT	857	984	940	880	959	965	953	963
wt% palmitic acid converted to ester (h-1)	0.72	1.82	1.46	1.87	1.63	2.09	1.84	1.84	2.22
Ratio to WT	1.0	2.5	2.0	2.6	2.3	2.9	2.6	2.6	3.1

Example 13. Octyl methacrylate synthesis kinetics of WT and variants 554, 857 and 963

[00172] WT CALB and variants 554, 857, and 963 were expressed in *Bacillus subtilis* as described in Example 3, and the secreted lipase was adsorbed from the broth onto Purolite™ ECR1030M resin. The dried resin was used in synthesis reactions to compare the octyl methacrylate synthesis activities of the different lipase variants.

[00173] An equimolar mix of octanol and methyl methacrylate was prepared, and each variant was analyzed in triplicate. The following were added to a 3 dram

glass vial: stir bar, 0.5 grams mix. Reactions were initiated by adding carefully pre-weighed dry resin with immobilized lipase (~50mg) into each reaction vial. Upon addition of lipase to each vial a stopwatch was immediately started. Vials were staggered in 30 second intervals to allow for consistent sampling. Ten microliters was transferred from each reaction vial to 0.5ml methanol in GC vials every 60 minutes for a total of 360 minutes for each reaction.

[00174] Each sample was analyzed on an Agilent 6890 gas chromatograph equipped with a 5% diphenyl / 95% dimethyl polysiloxane capillary (20m length, 0.25mm ID, 0.25µm film thickness, Restek RTX-5 Cat# 10223-124) GC column and a flame ionization detector. Initial temperature was held at 100°C for 3 minutes, then a gradient from 100°C to 325°C over 5.63 minutes. Temperature was held at 325°C for 2.38 minutes for a final run time of 11.0 minutes. The octanol and octyl methacrylate peaks were integrated, and reaction rate of esterification was estimated as peak area % octyl methacrylate produced over time based on dry resin weight. The initial synthesis rate of lipase variants 554, 857, and 963 are 2.8 to 5 times greater than the WT lipase activity.

Relative octyl methacrylate hour^{-1} gram resin^{-1}

1:1 octanol:acid	WT	554	857	963
Avg (n=3)	4.8	13.2	16.8	24.0
Ratio to WT	1.0	2.8	3.4	5.0

[00175] The subject matter described above is provided by way of illustration only and should not be construed as limiting. Various modifications and changes may be made to the subject matter described herein without following the example embodiments and applications illustrated and described, and without departing from the true spirit and scope of the present invention, which is set forth in the following claims.

[00176] All publications, patents and patent applications cited in this specification are incorporated herein by reference in their entireties as if each individual publication, patent or patent application were specifically and individually indicated to be incorporated by reference. While the foregoing has been described in terms of

various embodiments, the skilled artisan will appreciate that various modifications, substitutions, omissions, and changes may be made without departing from the spirit thereof.

CLAIMS

1. A *Candida antarctica* lipase B (CALB) variant having about two fold to about fifteen fold improved synthetic activity as compared to a wild type (WT) CALB.

2. The CALB variant of claim 1, wherein the CALB variant has an amino acid sequence having one or more modifications at position 141, 146, 188, 189, 223, 227, or 235 of SEQ ID NO: 2.

3. The CALB variant of claim 2, wherein the amino acid sequence has at least two modifications.

4. The CALB variant of claim 3, wherein the modifications are amino acid substitutions.

5. The CALB variant of claim 4, wherein the amino acid substitutions are A141T, A146T, E188D, I189V, D223G, S227T, or V235A.

6. The CALB variant of claim 5, wherein the CALB variant has an amino acid sequence of SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20 or SEQ ID NO: 22.

7. A nucleic acid encoding the CALB variant of claim 1.

8. A vector comprising the nucleic acid of claim 7.

9. An expression vector comprising the nucleic acid of claim 7.

10. The expression vector of claim 9, wherein the expression vector comprises a secretion signal.

11. A method for expressing a CALB variant, wherein the method comprises transfecting the vector of claim 9 into a host cell, and culturing the host cell under

conditions that allow expression of the CALB variant.

12. A method for preparing a CALB variant, wherein the method comprises transfecting the vector of claim 10 into a host cell, culturing the host cell in a culture medium and under conditions that allow expression of the CALB variant, and obtaining supernatant from the culture medium.

13. The method of claim 12, wherein the method further comprises concentrating the supernatant comprising the CALB variant.

14. The method of claim 12, wherein the host cell is *Bacillus subtilis.*

15. An expression vector expressing CALB in *Bacillus* comprising a nucleic acid encoding CALB, one or more origins of replication for replication in *Bacillus*, a promoter, a secretion signal, and optionally a selectable marker.

16. The expression vector of claim 15, wherein the expression vector comprises a constitutive promoter.

17. An expression system comprising the expression vector of claim 15, a host strain of *Bacillus*, and a culture medium.

18. A culture medium comprising a non-carbohydrate micronutrient source, a buffering agent for maintaining pH of the medium at a range of about 5 to about 9, a non-hydrolyzable nonionic surfactant, and a nitrogen source.

19. The culture medium of claim 18, wherein the culture medium comprises:

about 0.1% to about 5% of a non-carbohydrate micronutrient source, relative to the total weight of the composition,

a buffering agent for maintaining the pH at about 6 to about 8,

about 0.01% to about 1% of a non-hydrolyzable nonionic surfactant, relative to the total weight of the composition,

and about 0.1% to about 5% of a nitrogen source, relative to the total weight of the composition.

20. The culture medium of claim 19, wherein the culture medium comprises:
yeast extract,
a buffering agent for maintaining the pH of the medium at 7,
a block copolymer, and
a nitrogen source comprising tryptone, hydrolyzed casein, casamino acids, peptone, soy peptone, nutrient broth, or meat extract.

21. A method of making CALB comprising transfecting the expression vector of claim 15 into a *Bacillus*, and cultivating the *Bacillus* in a culture medium under conditions that allow expression and secretion of CALB in the culture medium.

22. The method of claim 21, wherein the culture medium comprises a non-carbohydrate micronutrient source, a buffering agent for maintaining pH of the medium at a range of about 5 to about 9, a non-hydrolyzable nonionic surfactant, and a nitrogen source.

23. The method of claim 22, wherein the method comprises cultivating the *Bacillus* at a temperature of about 30°C to about 42°C, about 35°C to about 40°C, or about 37°C.

24. The method of claim 21, wherein in the presence of a selectable marker on the vector, the method further comprises adding an additive for selection to the culture medium.

25. The method of claim 21, wherein the *Bacillus* is *B. subtilis, B. cereus, B. brevis, B. licheniformis, B. stearothermophilus, B. pumilis, B.amyloliquefaciens, B. clusii,* or *B. megaterium.*

26. The method of claim 21, wherein the method further comprises obtaining the supernatant from the culture medium and concentrating the supernatant comprising CALB.

27. The method of claim 26, wherein the supernatant is concentrated by ultrafiltration.

28. A method for identifying amino acid mutations in CALB that alter lipase activity, wherein the method comprises

(a) obtaining crystal structure of a wild type (WT) CALB;

(b) introducing one or more amino acid mutations into the WT CALB to obtain a CALB variant;

(c) obtaining crystal structure of the CALB variant;

(d) solvating the crystal structures of the WT CALB and the CALB variant into an implicit solvent;

(e) performing molecular dynamic simulation on the structures to obtain resultant structures;

(f) solvating the resultant structures into an explicit solvent;

(g) performing molecular dynamic simulation on the resultant structures in the explicit solvent medium to obtain refined structures;

(h) obtaining structural data for the refined structures of WT CALB and CALB variant; and

(i) comparing the structural data obtained for the refined structures of the WT CALB and the CALB variant, to identify one or more amino acid mutations in CALB that alter lipase activity.

29. The method of claim 28, wherein the method comprises obtaining structural data for cavity volume of an active site of the WT CALB and the CALB variant.

30. The method of claim 28, wherein the method comprises obtaining structural data for solvent accessible surface area (SASA) of an active site of WT CALB and CALB variant.

31. The method of claim 28, wherein the method comprises obtaining structural data for distance between center of mass of two residues of interest.

32. The method of claim 28, wherein the method further comprises correlating SASA, cavity volume, with the structure of CALB variant.

33. The method of claim 28, wherein the method further comprises obtaining SASA

of one or more amino acid positions of a catalytic triad the WT CALB and obtaining SASA of one or more positions of a catalytic triad of the CALB variant, comparing the obtained SASA of an amino acid position of the catalytic triad of the WT CALB with the obtained SASA of a corresponding amino acid position of the catalytic triad of the CALB variant, and identifying a CALB variant that enhances synthetic activity of CALB.

34. The method of claim 28, wherein the method further comprises correlating SASA of the CALB variant and its synthetic activity.

35. A method of catalyzing synthesis of a carboxylic acid ester, wherein one or more alcohols and one or more carboxylic acids or one or more carboxylic esters are reacted in the presence of the CALB variant of claim 1 to form a carboxylic acid ester.

36. The method of claim 35, wherein at least one of the carboxylic acids is a branched carboxylic acid, or wherein at least one of the carboxylic acid esters is a branched carboxylic acid ester.

37. A method of catalyzing synthesis of an amide, wherein one or more carboxylic acids or carboxylic acid esters and one or more amines are reacted in the presence of the CALB variant of claim 1 to form an amide.

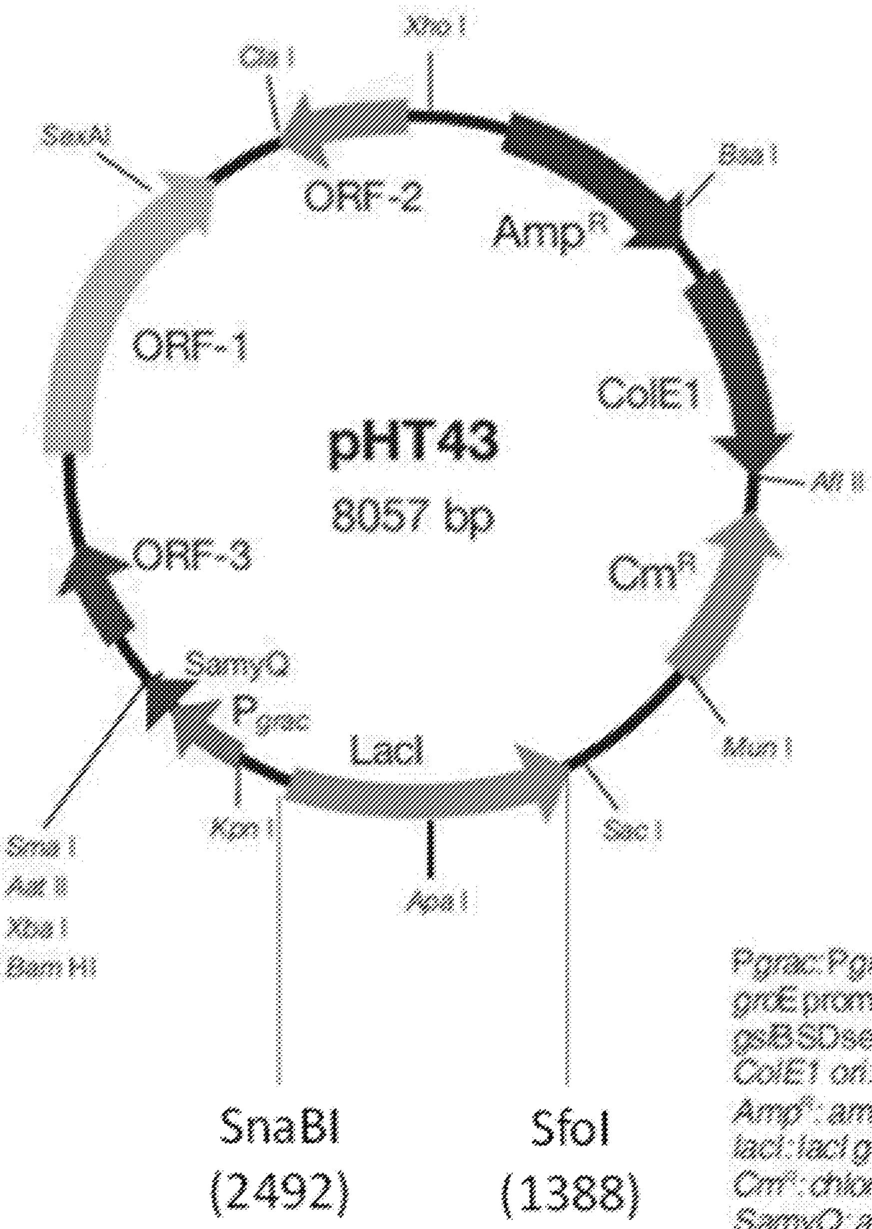
38. The method of claim 37, wherein at least one of the carboxylic acids is a branched carboxylic acid or wherein at least one of the carboxylic acid esters is a branched carboxylic acid ester.

39. A fusion protein comprising the CALB variant of claim 1 and a heterologous peptide.

40. The fusion protein of claim 39, wherein the heterologous peptide is a signal peptide.

41. A nucleic acid encoding the fusion protein of claim 39.

Vector Map pHT43



Pgrac: Pgrac promoter (consisting of the groE promoter, the lacO operator and the gsiB SD sequence)
ColE1 ori: ColE1 origin
Amp^R: ampicillin resistance
lacI: lacI gene (lac repressor)
Cm^R: chloramphenicol resistance
SamyQ: amyQ signal sequence

Fig.1

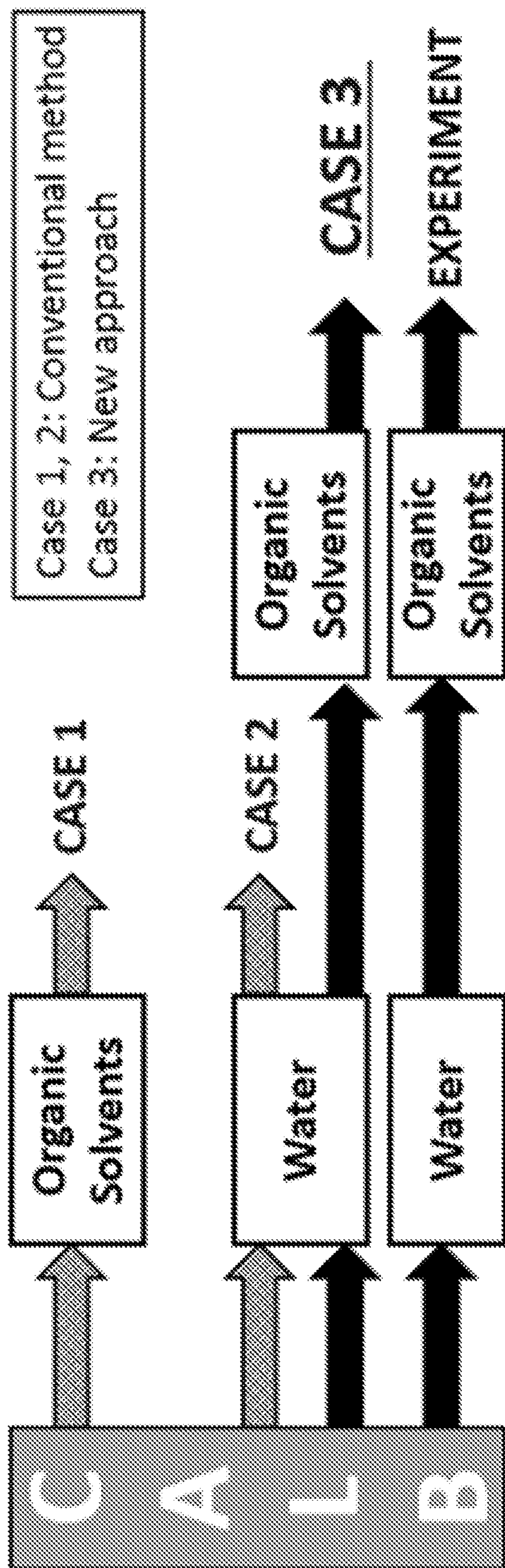
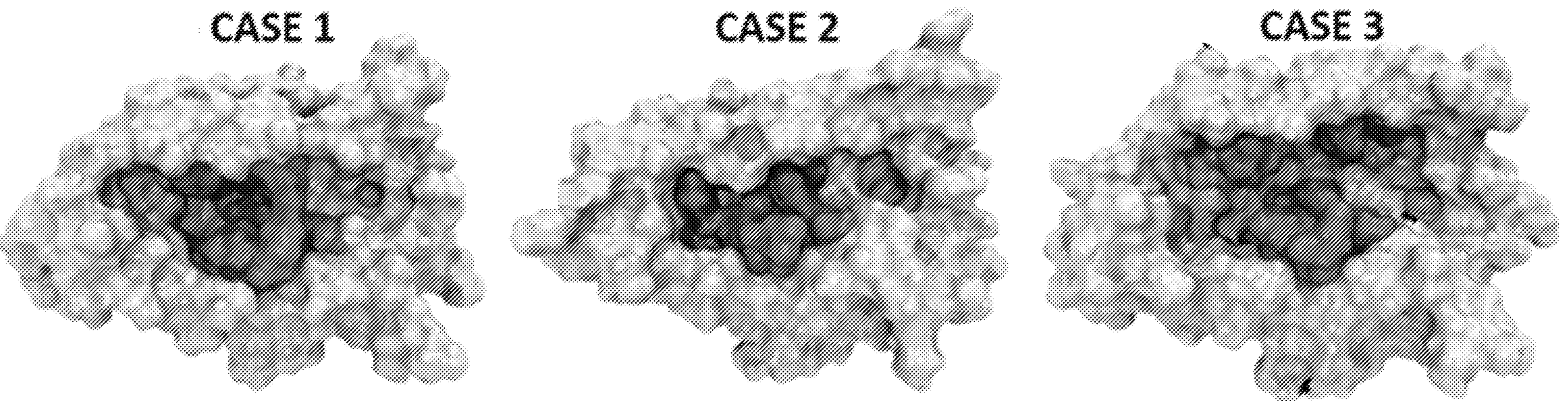


Fig. 2A



Case	Cavity Volume (Å^3)	Solvent Accessible Surface Area (Å^2)		
		105 SER	187 ASP	224 HIS
Case 1	493.30	4.34 (± 1.40)	0.33 (± 0.71)	5.51 (± 2.82)
Case 2	124.40	1.47 (± 1.79)	3.82 (± 2.22)	20.44 (± 9.84)
Case 3	612.60	5.51 (± 3.60)	1.92 (± 1.36)	33.25 (± 8.98)

Fig. 2B

```
                10        20        30        40        50        60        70        80
        ---------+---------+---------+---------+---------+---------+---------+---------+
22 Δ.pro  LPSGSDPAFSQPKSVLDAGLTCQGASPSSVSKPILLVPGTGTTGPQSFDSNWIPLSTQLGYTPCWISPPPFMLNDTQVNT   80
529.pro   ................................................................................   80
578.pro   ................................................................................   80
1001.pro  ................................................................................   80

                90       100       110       120       130       140       150       160
        ---------+---------+---------+---------+---------+---------+---------+---------+
22 Δ.pro  EYMVNAITALYAGSGNNKLPVLTWSQGGLVAQWGLTFFPSIRSKVDRLMAFAPDYKGTVLAGPLDALAVSAPSVWQQTTG  160
529.pro   ................................................................................  160
578.pro   ................................................................................  160
1001.pro  ................................................................................  160

               170       180       190       200       210       220       230       240
        ---------+---------+---------+---------+---------+---------+---------+---------+
22 Δ.pro  SALTTALRNAGGLTQIVPTTNLYSATDEIVQPQVSNSPLDSSYLFNGKNVQAQAVCGPLFVIDHAGSLTSQFSYVVGRSA  240
529.pro   ..............................................................G...T.............  240
578.pro   ..............................................................G.................  240
1001.pro  ..................................................................T.............  240

               250       260       270       280       290       300       310
        ---------+---------+---------+---------+---------+---------+---------+-------
22 Δ.pro  LRSTTGQARSADYGITDCNPLPANDLTPEQKVAAAALLAPAAAAIVAGPKQNCEPDLMPYARPFAVGKRTCSGIVTP     317
529.pro   .............................................................................     317
578.pro   .............................................................................     317
1001.pro  .............................................................................     317
```

Fig.3

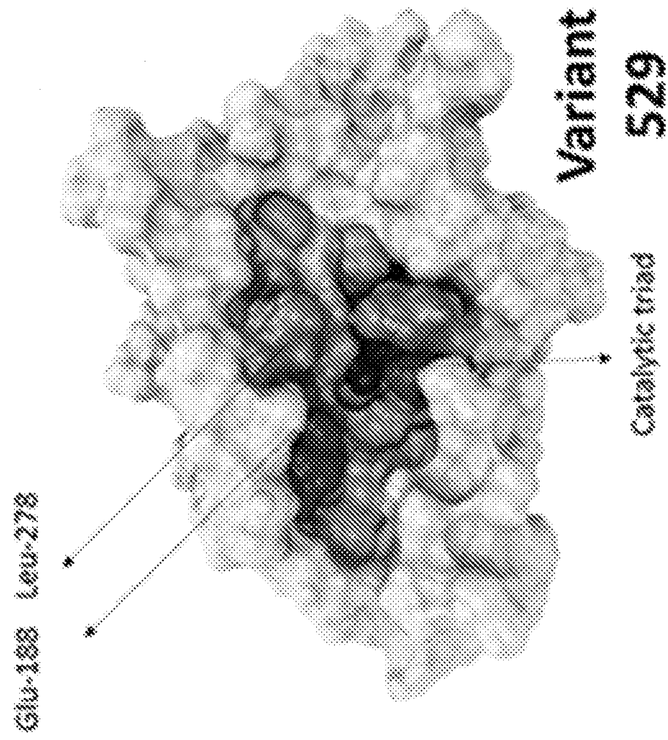


Fig. 4B

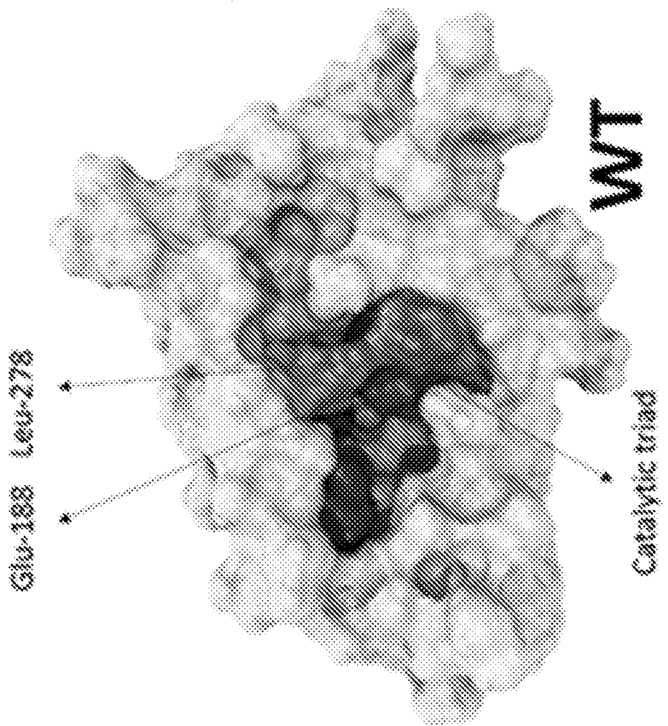


Fig. 4A

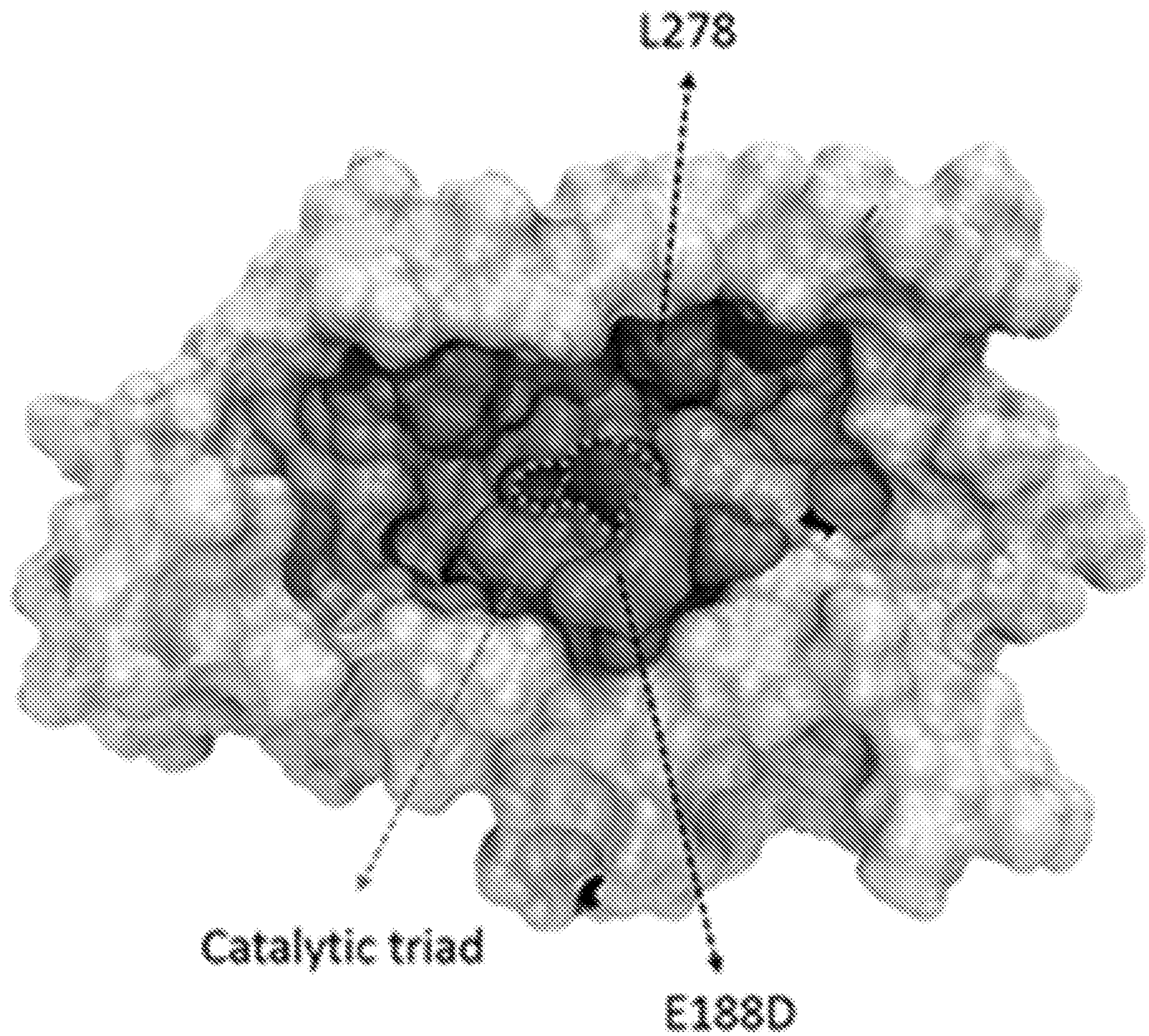


Fig.5

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2016/065255

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N9/20 C12P21/02 G06F19/16
ADD.

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)
C12N C12P G06F

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 practicable, search terms used)

EPO-Internal, WPI Data, CHEM ABS Data, Sequence Search, BIOSIS, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2013/010783 A1 (NOVOZYMES AS [DK]; BESENMATTER WERNER [AT]; SVENDSEN ALLAN [DK]; RANNE) 24 January 2013 (2013-01-24)	1-5, 7-17, 21-27, 35-41
A	lines 31-33 - page 39 -----	6
X	CN 104 745 550 A (UNIV JIANGNAN) 1 July 2015 (2015-07-01)	1-5, 7-17, 21-27, 35-41
A	paragraphs [0012], [0049], [0052]; sequences 1, 4 ----- -/--	6

[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 application or patent 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
16 February 2017	18/04/2017

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Petri, Bernhard

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2016/065255

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 UniProt [Online] 24 July 2013 (2013-07-24), "SubName: Full=Uncharacterized protein {ECO:0000313¦EMBL:GAC93661.1};", XP002767318, retrieved from EBI accession no. UNIPROT:R9NYC6 Database accession no. R9NYC6	1-5, 7-17, 21-27, 35-41
A	sequence -----	6
A	WO 2011/067349 A1 (NOVOZYMES AS [DK]; NOVOZYMES SOUTH ASIA PVT LTD [IN]; SVENDSEN ALLAN [) 9 June 2011 (2011-06-09) page 37, lines 8-9, 12-13; example 5; table 5 -----	1-17, 21-27, 35-41
A	WO 2009/080676 A1 (BASF SE [DE]; HAUER BERNHARD [DE]; KVARNSTROEM BRANNEBY CECILIA [SE];) 2 July 2009 (2009-07-02) page 2, lines 15-20 -----	1-17, 21-27, 35-41

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2016/065255

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-17, 21-27, 35-41

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-17, 21-27, 35-41

   Candida antarctica lipase B variant comprising one or more modifications at position 141, 146, 188, 18923, 227, 235 of Seq.Id No. 2 and having at least about two fold improved synthetic activity as compared to wild type CALB.

   ---

2. claims: 18-20

   Culture medium comprising non-carbohydrate micronutrient source, buffering agent (pH 5-9), non-hydrolyzable nonionic surfactant, nitrogen source

   ---

3. claims: 28-34

   Method for identifying amino acid mutations in CALB that alter lipase activity comprising obtaining crystal structures, introducing mutations, solvating crystal structures and performing molecular dynamic simulation.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2016/065255

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2013010783 A1	24-01-2013	CN 103827298 A	28-05-2014
		EP 2732033 A1	21-05-2014
		WO 2013010783 A1	24-01-2013
CN 104745550 A	01-07-2015	NONE	
WO 2011067349 A1	09-06-2011	EP 2507369 A1	10-10-2012
		US 2013023028 A1	24-01-2013
		WO 2011067349 A1	09-06-2011
WO 2009080676 A1	02-07-2009	DK 2245146 T3	29-06-2015
		EP 2245146 A1	03-11-2010
		US 2010273223 A1	28-10-2010
		WO 2009080676 A1	02-07-2009