

Thioesterase enzyme families: Functions, structures, and mechanisms

Benjamin T. Caswell¹ | Caio C. de Carvalho¹ | Hung Nguyen² |
Monikrishna Roy² | Tin Nguyen² | David C. Cantu¹ 

¹Department of Chemical and Materials Engineering, University of Nevada, Reno, Reno, Nevada, USA

²Department of Computer Science and Engineering, University of Nevada, Reno, Reno, Nevada, USA

Correspondence

David C. Cantu, Department of Chemical and Materials Engineering, University of Nevada, Reno, Reno, NV 89557, USA.
Email: dcantu@unr.edu

Funding information

U.S. National Science Foundation, Grant/Award Number: 2001385

Abstract

Thioesterases are enzymes that hydrolyze thioester bonds in numerous biochemical pathways, for example in fatty acid synthesis. This work reports known functions, structures, and mechanisms of updated thioesterase enzyme families, which are classified into 35 families based on sequence similarity. Each thioesterase family is based on at least one experimentally characterized enzyme, and most families have enzymes that have been crystallized and their tertiary structure resolved. Classifying thioesterases into families allows to predict tertiary structures and infer catalytic residues and mechanisms of all sequences in a family, which is particularly useful because the majority of known protein sequence have no experimental characterization. Phylogenetic analysis of experimentally characterized thioesterases that have structures with the two main structural folds reveal convergent and divergent evolution. Based on tertiary structure superimposition, catalytic residues are predicted.

KEYWORDS

catalytic residue prediction, enzyme family, structure superimposition, thioesterase

1 | INTRODUCTION

Thioesterases (TEs) hydrolyze thioester bonds and catalyze reactions in many different pathways such as fatty acid synthesis, polyketide synthesis, and non-ribosomal peptide synthesis. TEs are enzymes used in the biological production of tailored fatty acids and other medically relevant compounds such as macrolide antibiotics.^{1–4} TEs catalyze the hydrolysis of a wide variety of thioesters; for example, acyl-coenzyme A (CoA) hydrolysis occurs in the biological production of 3-hydroxybutyrate,⁵ in fatty acid β -oxidation,^{6,7} in vitamin K biosynthesis,⁸ and in 4-chlorobenzoate dehalogenation,⁹ among multiple pathways. TEs are also medically important, for example,

protein palmitoylation plays a role in malaria pathogenesis,¹⁰ and acyl-CoA thioesterases (ACOTs) are involved with fatty acid metabolism that affects obesity, diabetes, and nonalcoholic fatty liver disease in humans.¹¹

Classifying enzymes by primary structure (amino acid sequence) into families allows to predict the tertiary structure of all enzymes in a family as well as to identify catalytic residues and mechanisms. In 2010, the TE enzymes were classified into 23 families,¹² and placed in the publicly available Thioester-active enZYmes (ThYme) database.¹³ This is particularly useful since known protein sequences vastly outnumber enzymes whose function has been experimentally characterized or whose structure has been experimentally determined.

Enzyme family classification allows to infer the structure and function of an uncharacterized sequence in an

Benjamin T. Caswell and Caio C. de Carvalho contributed equally to this study.

organism of interest, based on a single enzyme with a known function and structure in a family. For example, structural knowledge of bacterial enzymes in TE family 14 (TE14) led to understanding substrate-protein interactions in algal TEs,¹⁴ as well as to structure prediction and analysis of plant sequences in the same family.¹⁵ Further, structural predictions and analysis of plant sequences in TE14, combined with site-directed mutagenesis, resulted in identifying the catalytic residues of the *Cuphea viscosissima* acyl-ACP TE, relevant for the biological production of tailored fatty acids.¹⁶ More recently, knowledge of enzyme sequences and their substrate specificity was used to predict function from structure, as recently done with acyl-ACP TEs.¹⁷

Since we first classified the TEs into families, the number of known protein sequences has increased by about three orders of magnitude, and more TEs have been experimentally characterized. New TE substrate specificities have been determined: as examples, (a) in TE4, a preference toward short chain fatty acids was observed in ACOT¹⁸; (b) RpaL, a TesB-like TE4 enzyme from *Rhodopseudomonas palustris*, was found to be active on aromatic and long and short aliphatic molecules bound to CoA¹⁹; (c) in TE6, YciA enzymes from *Methylobacterium extorquens* were shown to be hydrolyze ethylmalonyl-CoA for dicarboxylic acid production²⁰; and, (d) aryl-CoA substrate specificity was observed for enzymes in TE13.²¹

More TEs have been identified since we first classified TEs into families, some which form part of existing families. As examples, (a) guanosine diphosphate regulation TEs from *Neisseria meningitidis* appear in TE6²²; (b) acyl-lipid thioesterase from *Arabidopsis thaliana* in TE9²³; (c) methylketone synthases,²⁴ which were originally characterized from tomato prior to the ThYme database, have also been found in *Solanum melongena* and *Glycine max* and form part of TE9^{25,26}; (d) *Shewanella oneidensis* YbgC, which was found to primarily hydrolyze short chain acyl-CoA thioesters, also forms part of TE9²⁷; (e) BorB, required for borrelidin biosynthesis, is a member of TE18²⁸; and (f) the *Isochrysis galbana* thioesterase/carboxylesterase (IgTeCe) is in TE21.²⁹

Structural knowledge about how enzymes perform thioester hydrolysis has increased; an insightful, recent review describes TE structures, with a particularly useful and clear connection of catalytic residues with enzyme topology.³⁰ Since we first classified TEs into families, new TE structures have been resolved, as examples: (a) in TE4, the TesB enzyme in *Yersinia pestis* was crystallized³¹; as were (b) the TesB enzymes in mycobacteria³²; (c) in TE6, the human ACOT12 enzyme structure was obtained³³; (d) in TE11, the tertiary structure of the TE involved with azinomycin biosynthesis was determined³⁴;

and (e) in TE12, the *Synechocystis* 1,4-dihydroxy-2-naphthoyl-CoA TE was crystallized.³⁵

Given the increase in known sequences, structures, and experimental characterization, TE families were updated. In this work, we report 35 TE families: their functions and mechanisms described, their structures analyzed, catalytic residues predicted, as well as showing the phylogenetic analysis of TE enzymes with the main structural folds. The updated TE families are available in the new, updated ThYme database (<http://thyme.engr.unr.edu>).

2 | RESULTS AND DISCUSSION

Based on sequence similarity, following the approach described in Section 4, we identified 35 TE families almost completely unrelated by primary structure. In the following sections, we discuss their functions (Section 2.1), tertiary structures and catalytic residues (Section 2.2), and phylogeny (Section 2.3). All the TE families are based on experimentally characterized enzymes, and most include tertiary structures from crystallization.

2.1 | TE families and their functions

Enzymes in families TE1 to TE13, TE24 to TE26, TE28, and TE31 to TE35 hydrolyze substrates with various functionalities bound by a thioester to CoA. Those in TE14 to TE19, and TE30 add a water to break the thioester bonds between acyl groups and an acyl carrier protein (ACP). The enzymes in TE20, TE21, TE27, and TE29 cleave the bonds between acyl groups and other proteins. Members of TE22 and TE23 break bonds between acyl groups and glutathione and its derivatives. The thioester-carrying moiety in CoA and ACP is a pantetheine residue, while glutathione itself carries the sulfur moiety, and in non-ACP proteins, the sulfur-carrying moiety is built up mainly from a cysteine residue.

For most TE families, the main function of their enzymes is thioester hydrolysis; however, TE is not the main activity for TE33–TE35. All the reported TE families have at least one member that has been experimentally confirmed to have TE function; however, some families have members that catalyze other reactions besides TE.

Some TE families include enzymes that are the TE domains of larger, multimodular proteins such as fatty acid synthases (FASs), polyketide synthases (PKSs), or non-ribosomal peptide synthases (NRPs). FASs, PKSs, and NRPs are large enzymes with multiple domains each

having different functions. Only the TE domains were used to identify TE family members.

The functions of enzymes in families TE1–TE23 are described in detail in our previous work,¹² and those of families TE24–TE35 are described here. Table 1 includes common names and genes, their overall function, known substrate specificities, and references for all TE families.

Enzymes in family TE24, assigned to EC 3.1.2.2, are able to hydrolyze fatty acyl-CoA molecules with varying chain lengths (C₄–C₁₈), but they usually show a preference for long chain fatty acyl groups.⁸⁰ TE24 members from *Mycobacterium tuberculosis* are involved in the synthesis of mycolic acids, which are used by the organism to form a protective layer around pathogens.⁸¹

Members of TE25, which include EC 3.1.2.29 among others, are able to breakdown fluoroacetyl-CoA, suggesting a key metabolic step in the resistance mechanism of *Streptomyces cattley* to fluoroacetate, a well-known toxic substance produced by plants as a biodefense.^{83,84}

Family TE26 includes structures ybfF enzymes that hydrolyze palmitoyl-CoA and malonyl-CoA.⁸⁵ TE26 also includes alcohol acetyl transferases which could produce industrially relevant esters. The yeast *Wickerhamomyces anomalus* showed alcohol acetyltransferase (AATase) activity with ethanol and acetyl-CoA, releasing free CoA under high acetyl-CoA concentration. Although thioester hydrolysis is not the main function of the AATases in TE26, free CoA in the absence of ethanol was also reported, confirming TE activity by acetyl-CoA hydrolysis.⁸⁶

Enzymes in TE27 (EC 3.1.2.22), described as mitochondrial palmitoyl-protein TEs, present in mammals, include the α/β hydrolases 10 (ABHD10) enzymes. ABHD10 enzymes are related with S-palmytoilation, a reversible lipid posttranslational modification.⁸⁷

Enzymes in TE28 include mpaH, responsible for making mycophenolic acid from mycophenoyl-CoA, a natural antibiotic produced in the *Penicillium brevicompactum* peroxisome. These enzymes have a C-terminal cyclase/TE domain that catalyzes the cyclization and release of the polyketide.^{88,89}

Family TE29 (EC 3.1.2.22) includes acyl-protein thioesterases (APTs). APT enzymes are known to remove palmitate from cytosolic cysteine residues, such as S-hexadecanoyl-L-cysteinyl, in the Golgi complex of *Homo sapiens*.⁹⁰

Enzymes in TE30 (EC 3.1.2.-) are known to be involved in the biosynthesis of citrinin, a mycotoxin, in *Penicillium* and *Monascus* species. Multi-domain PKSs are associated in citrinin biosynthesis. Type I and type VII PKS enzymes have a TE domain (CitA) involved in

hydrolysis of thioester bond tethered with an ACP, releasing a free ACP and an aldehyde.⁹¹

Family TE31 (EC 3.1.2.2) has TEs that break down long-chain acyl-CoA molecules, releasing acyl chains used for reacylation of precursors of cardiolipin, a mitochondrial phospholipid found in *H. sapiens* and other mammals.⁹²

Among enzymes from TE32 (EC 3.1.2.32), those from *Pseudomonas aeruginosa* can hydrolyze 2-aminobenzoylacetyl-CoA to form 2-aminobenzoylacetate and CoA, a reaction in the signaling system for the expression of virulence genes that affect the cell density.^{93,94}

TE33 (EC 2.3.1.84 and EC 3.1.2.20) includes AATase enzymes, also known as alcohol-O-acyltransferase, that in *Saccharomyces cerevisiae* hydrolyze thioesters, but whose main function is not TE activity. These enzymes promote the esterification of isoamyl alcohol by acetyl-CoA. TE33 members, which prefer long- and straight-chain alcohol substrates over those with short and branched-chains, transfer the acyl group from an acyl-CoA donor to an acceptor alcohol, releasing acyl esters that can be applied in the food and beverage industry as flavoring agents. Some acetate ester products are: ethyl acetate, isoamyl acetate, isobutyl acetate, butyl acetate, hexyl acetate, heptyl acetate, and octyl acetate.^{95,96}

Family TE34 includes citramalyl-CoA lyase (EC 2.3.3.9 or EC 3.1.2.30), a human mitochondrial enzyme involved in vitamin B₁₂ metabolism that is expressed from polymorphic human genes known as CLYBL, which turns malyl-CoA into malate and free CoA.⁹⁷ Also present in TE34 are malyl-CoA lyase enzymes, which are structurally similar to CitE enzymes,¹⁰³ were described as a multifunctional enzyme that plays a role in autotrophic CO₂ fixation by *Chloroflexus aurantiacus*. These enzymes catalyze steps to generate (S)-malyl-CoA and β -methylmalyl-CoA in the 3-hydroxipropionate pathway.

Family TE35 (EC 3.1.1.4 and EC 3.1.2.2) includes enzymes encoded by the PLA2G6 human gene. Also known as VIA calcium-independent phospholipase A2 (iPLA₂ β), they perform SN-2 acyl chain hydrolysis, producing free fatty acids and lysophospholipids. Also, although not their main function, these enzymes can hydrolyze the thioester bonds from saturated long-chain fatty acyl-CoAs.^{101,102}

Other enzymes that have TE function, but were not classified into a family, include human mitochondrial 3-ketoacyl-CoA thiolases that are active on short, medium, or long-chain substrates to release free CoA, with the fastest rate being attributed to butyryl-CoA.¹⁰⁴ The main function of thiolases is a condensation of acyl groups, and not TE. Ubiquitin carboxyl-terminal hydrolases¹⁰⁵ were not classified into TE families because

TABLE 1 Thioesterase families, gene and enzyme names, functions, and substrate specificities

Family	Genes and/or enzyme names	General function	Known substrate specificities	References
TE1	Ach1	Acyl-CoA hydrolase	Acetyl-CoA	36,37
TE2	Acot1–Acot6 BAAT thioesterase	Acyl-CoA hydrolase	Palmitoyl-CoA Bile-acid-CoA	38,39
TE3	<i>tesA</i> <i>estA</i> Acyl-CoA thioesterase I Protease I Lysophospholipase L1	Acyl-CoA hydrolase	Medium- to long-chain acyl-CoA	40,41
TE4	<i>tesB</i> Acyl-CoA thioesterase II Acot8	Acyl-CoA hydrolase	Short-chain acyl-CoA Short- to long-chain acyl-CoA Palmitoyl-CoA Choloyl-CoA	18,42,43
TE5	<i>tesC</i> (<i>ybaW</i>) Acyl-CoA thioesterase III	Acyl-CoA hydrolase	Long-chain acyl-CoA 3,5-tetradecadienoyl-CoA	44
TE6	Acot7 (BACH) Acot11 (BFIT, Them1) Acot12 (CACH) YciA	Acyl-CoA hydrolase	Short- to long-chain acyl-CoA Ethylmalonyl-CoA	20,33,45–49
TE7	Acot9 Acot10	Acyl-CoA hydrolase	Short- to long-chain acyl-CoA	50,51
TE8	Acot13 (Them2)	Acyl-CoA hydrolase	Short- to long-chain acyl-CoA	52
TE9	YbgC ALT MKS	Acyl-CoA hydrolase	Short-chain acyl-CoA Short- to long-chain acyl-CoA 4-Hydroxybenzoyl-CoA	23,27,53–55
TE10	4HBT-I	Acyl-CoA hydrolase	4-Hydroxybenzoyl-CoA	56
TE11	4HBT-II EntH (YbdB) menI DHNAT1 1,4-Dihydroxy-2-naphthoyl-CoA hydrolase AziG	Acyl-CoA hydrolase	4-Hydroxybenzoyl-CoA	34,57
TE12	1,4-Dihydroxy-2-naphthoyl-CoA hydrolase	Acyl-CoA hydrolase	1,4-Dihydroxy-2-naphthoyl-CoA	58
TE13	<i>paaI</i> <i>paaD</i>	Acyl-CoA hydrolase	Short and medium-chain acyl-CoA Hydroxyphenylacetyl-CoA aryl-CoA	21,59
TE14	FatA FatB	Acyl-ACP hydrolase	Short- to long-chain acyl-ACP	60,61
TE15	Thioesterase CalE7	Acyl-ACP hydrolase	—	62
TE16	Thioesterase I Type I thioesterase TE domain of FAS TE domain of PKS or NRP	Acyl-ACP hydrolase	Long-chain acyl-ACP Polyketides Non-ribosomal peptides	63–65
TE17	TE domain of PKS	Acyl-ACP hydrolase	Polyketides	64
TE18	Thioesterase II Type II thioesterase (TE II) <i>tesA</i> <i>rifR</i> OLAH SAST	Acyl-ACP hydrolase	Medium-chain acyl-ACP Polyketides Non-ribosomal peptides	66–70

TABLE 1 (Continued)

Family	Genes and/or enzyme names	General function	Known substrate specificities	References
TE19	luxD	Acyl-ACP hydrolase	Myristoyl-ACP	71
TE20	ppt1 ppt2 Palmitoyl-protein thioesterase	Acyl-protein hydrolase	Palmitoyl-protein	72–74
TE21	apt1 apt2 Acyl-protein thioesterase Phospholipase Carboxylesterase	Acyl-protein hydrolase	Thioacylate proteins Palmitoyl-protein	75,76
TE22	S-formylglutathione hydrolase Esterase A Esterase D	Glutathione hydrolase	S-formylglutathione	77
TE23	Hydroxyglutathione hydrolase Glyoxalase II	Glutathione hydrolase	D-Lactoylglutathione	78,79
TE24	Fcot-like thioesterase Type III thioesterase CmiS1	Acyl-CoA hydrolase	Palmitoyl-CoA Stearoyl-CoA Lauroyl-CoA Hexanoyl-CoA	80–82
TE25	Fluoroacetyl-CoA thioesterase	Acyl-CoA hydrolase	Fluoroacetyl-CoA	83,84
TE26	EAT1 ybfF	Acyl-CoA hydrolase	Acetyl-CoA Palmitoyl-CoA Malonyl-CoA	85,86
TE27	ABHD10 Palmitoyl-protein thioesterase	Acyl-protein hydrolase	S-palmitoyl-protein	87
TE28	mpaH Type I acyl-CoA thioesterase	Acyl-CoA hydrolase	Malonyl-CoA	88,89
TE29	ABHD17A ABHD17B ABHD17C	Acyl-protein hydrolase	S-hexadecanoyl-L-cysteinyl	90
TE30	citA lovG mlcF mpL1 afoC mokD	Acyl-ACP hydrolase	Malonyl-ACP Acetoacetyl-ACP	91
TE31	Them4 Them5	Acyl-CoA hydrolase	Long-chain acyl-CoA	92
TE32	ACAA2 3-Ketoacyl-CoA thiolase	Acyl-CoA hydrolase	2-Aminobenzoylacetyl-CoA	93,94
TE33	ATF1 Alcohol-O-acetyltransferase	Alcohol acetyl transferases	Acyl-CoA	95,96
TE34	CLYBL Citramalyl-CoA lyase citE RipC	Citramalyl-CoA lyase	Malyl-CoA	97–100
TE35	PLA2G6 Calcium-independent phospholipase A2	Calcium-independent phospholipase	Long-chain fatty acyl-CoAs	101,102

Abbreviations: ALT, acyl-lipid thioesterase; CoA, coenzyme A; MKS, methylketone synthases.

peptidase activity is their main function, and they can be found in the MEROPS database.¹⁰⁶

2.2 | TE families and their structures, catalytic residues, and mechanisms

The tertiary structures in each TE family were superimposed to confirm structural similarity. Each family that underwent this analysis exhibits members very highly similar in tertiary structure; their cores are nearly identical and their overall resemblance is high. This structural similarity is shown by RMSD_{ave} values of <1.4 Å and P_{ave} values of >77% in all families (see Section 4 for definitions). Table 2 reports the structural fold of the enzymes in each family, as well as the RMSD_{ave} and P_{ave} values for families with more than two known tertiary structures. Table 3 describes the catalytic residues, and their corresponding literature, of the structures in each TE family. We predicted catalytic residues from tertiary structure superimposition as those which spatially correspond with known catalytic residues in superimposed structures, also reported in Table 3. Figures 1 and 2 show how catalytic residues were predicted, based on structure superimposition and spatial correspondence, for TEs with HotDog fold (TE25) and an α/β -Hydrolase fold (TE20), respectively. Enzymes in TE23 and TE32 have available tertiary structures; however, their catalytic residues have not been proposed, and therefore predictions based on structural superimpositions were not done. Other families do not have any known tertiary structures: TE7, TE28, TE29, TE30, and TE33. Predicting catalytic residues was not necessary for TE13, TE14, TE17, TE18, TE19, TE24, TE26, and TE31 as every structure in these families has published literature indicating the catalytic residues (see Table 3). Within each of these families, the catalytic residues are suitably conserved between structures, with the exception of TE19 and TE26, which each only have single known structures.

2.2.1 | HotDog catalytic residues and mechanisms

Families with HotDog^{160,161} fold structures (TE4–TE15, TE24, TE25, TE31) have highly similar tertiary structures, indicated by the consistently low RMSD_{ave} and high P_{ave} values.

HotDog-fold enzymes lack defined non-solvated binding pockets and conserved catalytic residues,⁴⁵ thus a variety of catalytic residues and mechanisms exist.

In TE4, *Mycobacterium marinum* TesB2 (3U0A) catalytic residues were predicted to be Asp194-Ser216-Gln266,

based on comparison to an *Escherichia coli* TE II enzyme (1C8U) in which Asp204-Gln278-Thr228 orient a water molecule for nucleophilic attack on the substrate.¹¹⁵ This is consistent with the catalytic residues found in *Y. pestis* TesB (4QFW, 4R4U); a structure that presents an octameric quaternary structure, unique among HotDog families.³¹ A *S. cerevisiae* TE I structure (1TBU) contains only residues from the N-terminal domain that does not include the residues that could be compared to the catalytic triad. Catalytic residues for the remaining family members were predicted (see Table 3). Of note in these predictions are *Mycobacterium avium* MAV2540 (3RD7) and MAP1729c (4R9Z); these inactive TesB enzymes contain a mutation in which the highly conserved Asp residue is substituted for an Ala residue. Within TesB TEs, this mutation appears to be unique to *Mycobacterium* species.³²

In TE6, *Mus musculus* Acot7 N-terminal domain (2V1O) and C-terminal domain (2Q2B) catalytic residues are reported as Asn24 and Asp213, respectively.¹¹⁶ The structures for human Acot12 (3B7K, 4MOB, 4MOC) and *M. musculus* Acot7 (4ZV3, 6Vfy) contain both N and C-terminal domains. Our alignment placed both 2V1O and 2Q2B over the C-terminal of these structures confirm catalytic residues in the C-terminal domain. Using this molecular symmetry, the N-terminal catalytic residues were predicted as well. This follows with literature which indicates that these structures form a functioning active site when joined as a dimer.³³ A study on *N. meningitidis* TE 12 (5SZU) supported these findings, pointing to a covalent disulfide bond dimer linkage that is requisite for enzymatic activity.²² The Asn-Asp catalytic motif is highly consistent in this family, recently supported by findings on a *Bacillus cereus* TE (7CZ3).¹¹⁹ Unique among the family is a *S. aureus* TE (4NCP) that also relies on a Thr residue for catalysis.¹¹⁸ Also in TE6, YciA structures have and aspartic acid catalytic residues in the same structural position as those in *Campylobacter jejuni* Cj0915 (3D6L) and *Haemophilus influenzae* Rd KW20 HI0827 (1YLI, 3BJK).^{46,117}

Although TE7 has no known crystal structures, sequence analysis with other ACOT enzyme suggests that Asp120 and Asn305 are catalytic residues in the mouse ACOT9 enzyme.⁵⁰

It was proposed for TE8 enzymes, based on the crystal structure of a human Them2 enzyme, that Gly57 and Asn50 bind and polarize the thioester carbonyl group while Asp65 and Ser85 orient and activate the water nucleophile.^{120,121} It was later proposed, based on mixed quantum mechanics/molecular mechanics simulations of the same human enzyme, that a His-Ser pair acts as the acid proton donor in a concerted mechanism where the Asp residue activates the water molecule.¹²² Based on superimposition with the crystal structure of the human

TABLE 2 Thioesterase folds and structure superimposition

Family	Fold	RMSD _{ave} , Å	P _{ave} , %	Structures in the PDB
TE1	NagB	0.92	95.7	2G39, 2NVV, 4EU3, 4EU4, 4EU5, 4EU6, 4EU7, 4EU8, 4EU9, 4EUA, 4EUB, 4EUC, 4EUD, 5DDK, 5DW4, 5DW5, 5DW6, 5E5H
TE2	α/β-Hydrolase	0.86	94.6	3HLK, 3K2I
TE3	SGNH	0.92	87.4	1IVN, 1J00, 1JRL, 1U8U, 1V2G, 3HP4, 4JGG, 5TIC, 5TID, 5TIE, 5TIF, 6IQ9, 6IQA, 6IQB, 6LFB, 6LFC, 7C23, 7C29, 7C2A, 7C82, 7C84
TE4	HotDog	1.09	81.6	1C8U, 1TBU, 3RD7, 3U0A, 4QFW, 4R4U, 4R9Z
TE5	HotDog	—	—	1NJK
TE6	HotDog	1.09	86.9	1YLI, 2EIS, 2G6S, 2Q2B, 2QQ2, 2V1O, 3B7K, 3BJK, 3D6L, 4IEN, 4MOB, 4MOC, 4ZV3, 5DM5, 5SZU, 5SZV, 5SZY, 5SZZ, 5T02, 5V3A, 4NCP, 5EGJ, 5EGK, 5EGL, 5HWF, 5HZ4, 6Vfy, 7CZ3
TE7	Putative HotDog	—	—	—
TE8	HotDog	0.56	97.7	2CY9, 2F0X, 2H4U, 3F5O, 4ORD
TE9	HotDog	0.48	96.7	1S5U, 2PZH, 5KL9, 5T06, 5T07
TE10	HotDog	1.01	94.2	1BVQ, 1LO7, 1LO8, 1LO9, 1Z54, 5WH9
TE11	HotDog	0.90	98.4	1Q4S, 1Q4T, 1Q4U, 1SBK, 1SC0, 1VH5, 1VH9, 1VI8, 2B6E, 3LZ7, 3R32, 3R34, 3R35, 3R36, 3R37, 3R3A, 3R3B, 3R3C, 3R3D, 3R3F, 3S4K, 3TEA, 4K02, 4K49, 4K4A, 4K4B, 4K4C, 4K4D, 4M20, 4QD7, 4QD8, 4QD9, 4QDA, 4QDB, 4YBV, 5EP5, 5HMB, 5HMC
TE12	HotDog	0.92	88.3	2HX5, 4K00
TE13	HotDog	0.49	98.8	1J1Y, 1PSU, 1WLU, 1WLV, 1WM6, 1WN3, 2DSL, 2FS2
TE14	HotDog	1.36	81.3	2ESS, 2OWN, 4GAK, 5X04
TE15	HotDog	0.85	96.2	2W3X, 2XEM, 2XFL, 4I4J, 5VPJ
TE16	α/β-Hydrolase	1.40	64.5	1JMK, 1XKT, 2CB9, 2CBG, 2K2Q, 2PX6, 3ILS, 3TJM, 4Z49, 4ZXH, 4ZXL, 5V3W, 5V3X, 5V3Y, 5V3Z, 5V40, 5V41, 5V42, 6OJC, 6OJD
TE17	α/β-Hydrolase	1.23	79.2	1KEZ, 1MN6, 1MNA, 1MNQ, 1MO2, 2H7X, 2H7Y, 2HFJ, 2HFK, 3LCR, 5D3K, 5D3Z, 6MLK
TE18	α/β-Hydrolase	1.16	77.0	3FLA, 3FLB, 3QMV, 3QMW, 4XJV, 5UGZ, 6BA8, 6BA9, 6FVJ, 6FW5, 6VAP
TE19	α/β-Hydrolase	—	—	1THT
TE20	α/β-Hydrolase	0.69	90.6	1EH5, 1EI9, 1EXW, 1PJA, 3GRO
TE21	α/β-Hydrolase	1.03	85.6	1AUO, 1AUR, 1FJ2, 3CN7, 3CN9, 3U0V, 4F21, 4FHZ, 4FTW, 5KRE, 5SYM, 5SYN, 6AVV, 6AVW, 6AVX, 6AVY, 6BJE, 6QGN, 6QGO, 6QGQ, 6QGS
TE22	α/β-Hydrolase	0.90	95.6	1PV1, 3C6B, 3E4D, 3FCX, 3I6Y, 3LS2, 3S8Y, 4B6G, 4FLM, 4FOL, 6JZL
TE23	Lactamase	1.24	82.6	1QH3, 1QH5, 1XM8, 2Q42, 2QED, 3TP9, 4YSB, 6RZ0, 6SOI
TE24	HotDog	0.85	93.1	2PFC, 3B18, 5WSX, 5WSY
TE25	HotDog	0.71	99.5	2CWZ, 3KUV, 3KUW, 3KV7, 3KV8, 3KVI, 3KVU, 3KVZ, 3KW1, 3KX7, 3KX8, 3P2Q, 3P2R, 3P2S, 3P3F, 3P3I
TE26	α/β-Hydrolase	— ^a	—	3BF7, 3BF8
TE27	α/β-Hydrolase	1.06	85.2	3LLC, 6NY9
TE28	Putative α/β-Hydrolase	—	—	—

(Continues)

TABLE 2 (Continued)

Family	Fold	RMSD _{ave} , Å	P _{ave} , %	Structures in the PDB
TE29	Putative α/β-Hydrolase	—	—	—
TE30	Putative α/β-Hydrolase	—	—	—
TE31	HotDog	0.54	98.5	4AE7, 4AE8, 4GAH
TE32	Lactamase	0.31	1.00	2Q0I, 2Q0J, 2VW8, 3DH8, 5HIO, 5HIP, 5HIQ, 5HIS
TE33	—	—	—	—
TE34	Beta-hairpin (C-terminal) TIM barrel (N-terminal)	1.15	87.4	1SGJ, 1U5H, 1U5V, 1Z6K, 3QLL, 4L9Y, 4L9Z, 5UGR, 5VXC, 5VXO, 5VXS, 6AQ4
TE35	—	—	—	6AUN

Abbreviations: PDB, Protein Data Bank; RMSD, root mean square distance.

^aRMSD and P_{ave} for TE26 were not calculated because the two PDB entries are of the same protein structure.

TABLE 3 Thioesterase families and catalytic residues

Family	Catalytic residues	Corresponding structure	Producing organism	Reference
TE1	Val270, Glu294, Asn347, Gly388	4EU3, 4EU4, 4EU5, 4EU6, 4EU7, 4EU8, 4EU9, 4EUA, 4EUB, 4EUC, 4EUD	<i>Acetobacter aceti</i>	107
	Val270, Glu294, Asn347, Gly388	5DDK, 5DW4, 5DW5, 5DW6, 5E5H	<i>A. aceti</i>	108
	Val259, Glu284 , Asn337, Gly378	2NVV	<i>Porphyromonas gingivalis</i>	Predicted in this work
	Ile264, Glu288, Asn341, Gly382	2G39	<i>Pseudomonas aeruginosa</i>	Predicted in this work
TE2	Ser294, His422, Asp388	3HLK	<i>Homo sapiens</i>	109
	Ser232, His360, Asp326	3K2I	<i>H. sapiens</i>	Predicted in this work
TE3	Ser10, Asp154, His157	1IVN, 1JRL, 1J00, 1U8U, 1V2G	<i>Escherichia coli</i>	110
	Ser11, Asp158, His161	3HP4	<i>Pseudoalteromonas sp.</i>	111
	Ser9, Asp156, His159	4JGG	<i>P. aeruginosa</i>	112
	Ser10, Asp154, His157	5TIC, 5TID, 5TIE, 5TIF	<i>E. coli</i>	113
	Ser10, Asp154, His157	6LFB, 6LFC	<i>E. coli</i>	Predicted in this work
	Ser29, Asp178, His181	7C23, 7C29, 7C2A, 7C82, 7C84	<i>Croceicoccus marinus</i>	114
	Ser13, Asp162, His165	6IQ9, 6IQA, 6IQB	<i>Altericroceibacterium indicum</i>	Predicted in this work
TE4	Asp204, Thr228, Gln278	1C8U	<i>E. coli</i>	115
	Asp194, Ser216, Gln266	3U0A	<i>Mycobacterium marinum M</i>	Predicted in this work
	Asp204, Thr228, Gln278	4QFW, 4R4U	<i>Yersinia pestis</i>	31
	—	1TBU	<i>Saccharomyces cerevisiae</i>	—
	Ala202, Leu225, Gln275 ^a	3RD7	<i>Mycobacterium avium 104</i>	Predicted in this work
Ala197, Gln 216, Gln266 ^a	4R9Z	<i>M. avium subsp. paratuberculosis K-10</i>	Predicted in this work	
TE5	—	1NJK	<i>E. coli</i>	—
TE6	Asp213	2Q2B	<i>Mus musculus</i>	116
	Asn24	2V1O		
	Asp44	1YLI, 3BJK	<i>Haemophilus influenzae Rd KW20</i>	117

TABLE 3 (Continued)

Family	Catalytic residues	Corresponding structure	Producing organism	Reference
	Asp34	3D6L	<i>Campylobacter jejuni</i>	46
	Asp36, Asn195	3B7K, 4MOB, 4MOC	<i>H. sapiens</i>	Predicted in this work
	Asp245	2QQ2	<i>H. sapiens</i>	Predicted in this work
	Asp46	5DM5	<i>Yersinia pestis</i>	Predicted in this work
	Asp31	2EIS	<i>Thermus thermophilus</i>	Predicted in this work
	Asn70, Asp259	4ZV3, 6VFY	<i>M. musculus</i>	Predicted in this work
	Asn24, Asp39	4IEN, 5SZU, 5SZV, 5SZY, 5SZZ, 5T02, 5V3A	<i>Neisseria meningitidis</i>	22
	Asn28, Asp43, Thr60	4NCP, 5EGJ, 5EGK, 5EGL, 5HWF, 5HZ4	<i>Staphylococcus aureus, subsp. aureus Mu50</i>	118
	Asn23, Asp38	7CZ3	<i>Bacillus cereus ATCC 14579</i>	119
TE7	—	—	—	—
TE8	Asn50, His56, Gly57, Asp65	2F0X, 3F5O, 2H4U	<i>H. sapiens</i>	120,121
	Asn50, His56, Gly57, Asp65	2CY9	<i>M. musculus</i>	Predicted in this work
	Asn51, His57, Gly58, Asp66	4ORD	<i>Danio rerio</i>	Predicted in this work
	Asp65, Ser83, His134	Simulation ^b	<i>H. sapiens</i>	122
TE9	Tyr7, Asp11, His18	2PZH	<i>Helicobacter pylori</i>	53
	Tyr14, Asp18, His25	1S5U, 5KL9, 5T06, 5T07	<i>E. coli</i>	Predicted in this work
TE10	Asp17	1BVQ, 1LO7, 1LO8, 1LO9	<i>Pseudomonas sp.</i>	123
	Asp16	5WH9	<i>Alkalihalobacillus halodurans C-125</i>	Predicted in this work
TE11	Gly65, Glu73	1Q4S, 1Q4T, 1Q4U	<i>Arthrobacter sp.</i>	124
	Gly55, Glu63	1VH9, 1VH5, 1VI8, 1SBK	<i>E. coli</i>	Predicted in this work
	Gly55, Glu63	2B6E, 1SC0, 3LZ7	<i>Haemophilus influenzae</i>	Predicted in this work
	Gly39, Glu47	4M20, 4YBV, 5EP5	<i>Staphylococcus aureus, subsp. aureus Mu50</i>	Predicted in this work
	Gly65, Ala73	3R32, 3R34, 3R35, 3R36, 3R37, 3R3A, 3R3B, 3R3C, 3R3D, 3R3F, 3TEA	<i>Arthrobacter sp.</i>	Predicted in this work
	Gly52, Glu60	3S4K	<i>Mycobacterium tuberculosis</i>	Predicted in this work
	Gly55, Glu63	4K49, 4K4A, 4K4B, 4K4C, 4K4D	<i>E. coli K-12</i>	125
	Gly56, Glu64	4QD7, 4QD8, 4QD9, 4QDA, 4QDB	<i>P. aeruginosa</i>	Predicted in this work
	Gly49, Glu57	5HMB, 5HMC	<i>Streptomyces sahachiroi</i>	Predicted in this work
	Gly49, Glu57	4K02	<i>Arabidopsis thaliana</i>	35

(Continues)

TABLE 3 (Continued)

Family	Catalytic residues	Corresponding structure	Producing organism	Reference
TE12	Asp16	2HX5	<i>Prochlorococcus marinus</i>	Predicted in this work
	Asp16	4K00	<i>Synechocystis</i> sp. PCC 6803 substr. Kazusa	35
TE13	Gly40, Asp48	1WLU, 1J1Y, 1WM6, 1WLV, 1WN3, 2DSL ^c	<i>Thermus thermophilus</i>	126
	Gly53, Asp61	2FS2, 1PSU	<i>E. coli</i>	127
TE14	Asp281, Asn283, His285, Glu319	2ESS	<i>Bacteroides thetaiotaomicron</i> VPI-5482	128
	Asp281, Asn283, His285, Glu319	2OWN	<i>Lactiplantibacillus plantarum</i>	128
	Asp281, Asn283, His285, Glu319	4GAK	<i>Spirosoma linguale</i> DSM 74	128
	Asp281, Asn283, His285, Glu319	5X04	<i>Umbellulaia californica</i>	128
TE15	Asn19, Tyr29, Arg37	2W3X	<i>Micromonospora echinospora</i>	62
	Asn23, Tyr33, Arg41	2XEM, 2XFL	<i>Micromonospora chersina</i>	Predicted in this work
	Asn21, Tyr31, Arg39	4I4J	<i>Streptomyces globisporus</i>	Predicted in this work
	Asn17, Tyr27, Arg35	5VPJ	<i>Actinomadura verrucosospora</i>	Predicted in this work
TE16	Ser2308, Asp2338, His2481	1XKT, 2PX6, 3TJM, 4Z49	<i>H. sapiens</i>	129
	Ser80, Asp107, His207	1JMK	<i>Bacillus subtilis</i>	130
	Ser84, Asp111, His201	2CB9, 2CBG	<i>B. subtilis</i>	131
	Ser1937, Asp1964, His2088	3ILS	<i>Aspergillus parasiticus</i>	132
	Cys1135, Asp1162, His1295	4ZXH, 4ZXI	<i>Acinetobacter baumannii</i> AB307-0294	Predicted in this work
	Ser1533, Asp1560, His1699	5V3W, 5V3X, 5V3Y, 5V3Z, 5V40, 5V41, 5V42	<i>M. tuberculosis</i>	133
	Ser1790, Asp1806, His1901	6OJC, 6OJD	<i>Nocardia uniformis</i> subsp. <i>tsuyamanensis</i>	134
TE17	Ser142, Asp169, His259	1KEZ, 1MO2, 5D3K, 5D3Z, 6MLK	<i>Saccaropolyspora erythaea</i>	135
	Ser148, Asp176, His268	1MN6, 1MNA, 1MNQ, 2H7X, 2H7Y, 2HFJ, 2HFK	<i>Streptomyces venezuelae</i>	136
	Ser132, Asp159, His255	3LCR	<i>Streptomyces</i> sp. CK4412	137
TE18	Ser86, Asp189, His216	2K2Q, 2RON	<i>Brevibacillus parabrevis</i> , <i>B. subtilis</i>	138
	Ser94, Asp200, His228	3FLA, 3FLB	<i>Amycolatopsis mediterranei</i>	67
	Ser107, Asp213, His241	3QMV, 3QMW	<i>Streptomyces coelicolor</i>	139
	Ser101, Asp212, His237	4XJV	<i>H. sapiens</i>	140
	Ser78, Asp186, His215	5UGZ	<i>E. coli</i>	141
	Ser89, Asp197, His225	6BA8, 6BA9	<i>E. coli</i>	142
	Ser104, Asp208, His236	6FVJ, 6FW5	<i>M. tuberculosis</i>	66
	Ser98, Asp204, His232	6VAP	<i>Streptomyces</i> sp. WAC02707	28
TE19	Ser114, Asp211, His241	1THT	<i>Vibrio harveyi</i>	143
TE20	Ser115, Asp233, His289	1EH5, 1EI9, 1EXW	<i>Bos taurus</i>	144
	Ser111, Asp228, His283	1PJA, 3GRO	<i>H. sapiens</i>	Predicted in this work
TE21	Ser114, Asp168, His199	1AUO, 1AUR	<i>Pseudomonas fluorescens</i>	145
	Ser114, Asp169, His203	1FJ2	<i>H. sapiens</i>	146

TABLE 3 (Continued)

Family	Catalytic residues	Corresponding structure	Producing organism	Reference
	Ser113, Asp166, His197	3CN7, 3CN9	<i>P. aeruginosa</i>	147
	Ser124, Asp179, Glu212	3U0V	<i>H. sapiens</i>	Predicted in this work
	Ser116, Asp170, His202	4F21	<i>Francisella tularensis subsp. tularensis SCHU S4</i>	148
	Ser165, Asp216, His248	4FHZ, 4FTW	<i>Cereibacter sphaeroides</i>	Predicted in this work
	Ser119, Asp174, His209	5SYM	<i>H. sapiens</i>	149
	Ser122, Asp176, His210	5SYN	<i>H. sapiens</i>	
	Ser106, Asp160, His192	6AVV, 6AVW, 6AVX	<i>A. thaliana</i>	Predicted in this work
	Ser126, Asp197, His230	6AVY	<i>Zea mays</i>	Predicted in this work
	Ser122, Asp176, His210	6BJE	<i>H. sapiens</i>	150
	Ser119, Asp174, His208	6QGN, 6QGO, 6QGQ, 6QGS	<i>H. sapiens</i>	Predicted in this work
TE22	Ser161, Asp241, His276	1PV1, 3C6B	<i>S. cerevisiae</i>	151
	Ser147, Asp223, His256	3E4D	<i>Agrobacterium fabrum str. C58</i>	152
	Ser153, Asp230, His264	3FCX	<i>H. sapiens</i>	153
	Ser148, Asp224, His257	3I6Y, 3S8Y	<i>Oleispira antarctica</i>	154
	Ser147, Asp225, His258	3LS2	<i>Pseudoalteromonas translucida TAC125</i>	155
	Ser145, Asp221, His254	4B6G	<i>N. meningitidis MC58</i>	156
	Ser161, Asp241, His276	4FLM, 4FOL	<i>Saccaromyces cerevisia</i>	Predicted in this work
	Ser148, Asp224, His257	6JZL	<i>Shewanella frigidimarina</i>	157
TE23	—	— ^c	—	—
TE24	Asn83, Tyr87, Tyr33, Met118 (subunit A), and Tyr66, Thr70, His72, Asn74 (subunit B)	2PFC, 3B18	<i>M. tuberculosis</i>	80
	Tyr53, Ile54, His59, Asn61, Ser62 (subunit A), and Tyr20, Asn70, Met73, Tyr74, Ile107 (subunit B)	5WSX, 5WSY	<i>Streptomyces avermitilis MA-4680 = NBRC 14893</i>	82
TE25	Thr42, Glu50, His76	3KUV, 3KUW, 3KV7, 3KV8, 3KVI, 3KVU, 3KVZ, 3KW1, 3KX7, 3KX8	<i>Streptomyces cattleya</i>	158
	Thr36, Glu44, His70	2CWZ	<i>T. thermophilus HB8</i>	Predicted in this work
	Thr42, Glu50, His76	3P2Q, 3P2R, 3P2S, 3P3F, 3P3I	<i>S. cattleya</i>	84
TE26	Ser89, Asp113, Ser206, His234	3BF7, 3BF8	<i>E. coli</i>	85
TE27	Ser100, Asp197, His227	6NY9	<i>M. musculus</i>	87
	Ser113, Asp216, His246	3LLC	<i>Agrobacterium vitis S4</i>	Predicted in this work
TE28	—	—	—	—
TE29	—	—	—	—
TE30	—	—	—	—
TE31	Thr308, Ser473	4AE7, 4AE8, 4GAH	<i>H. sapiens</i>	92
TE32	—	— ^c	—	—

(Continues)

TABLE 3 (Continued)

Family	Catalytic residues	Corresponding structure	Producing organism	Reference
TE33	—	—	—	—
TE34	Asp320	5VXS, 5VXC, 5VXO	<i>H. sapiens</i>	97
	—	1SGJ	<i>Deinococcus radiodurans</i>	—
	—	1U5H, 1U5V, 1Z6K	<i>M. tuberculosis</i>	—
	Glu49	6AQ4	<i>M. tuberculosis</i>	99
	—	3QLL	<i>Yersinia pestis</i>	—
	Asp299	4L9Y, 4L9Z	<i>C. sphaeroides 2.4.1</i>	Predicted in this work
Asp304	5UGR	<i>Methylobacterium extorquens</i> AM1	Predicted in this work	
TE35	Ser465, Asp598	6AUN	<i>Cricetulus griseus</i>	159

^aCatalytic residue prediction for 3RD7 was based purely on their high degree of spatial correlation with the catalytic residues of 1C8U and 4QFW. It is noted that these residues do not have a high degree of chemical similarity.

^bPredicted from mixed quantum mechanics/molecular mechanics simulations based on the 3F5O crystal structure.

^cEven though structures are known, catalytic residues have not been determined, so none are predicted.

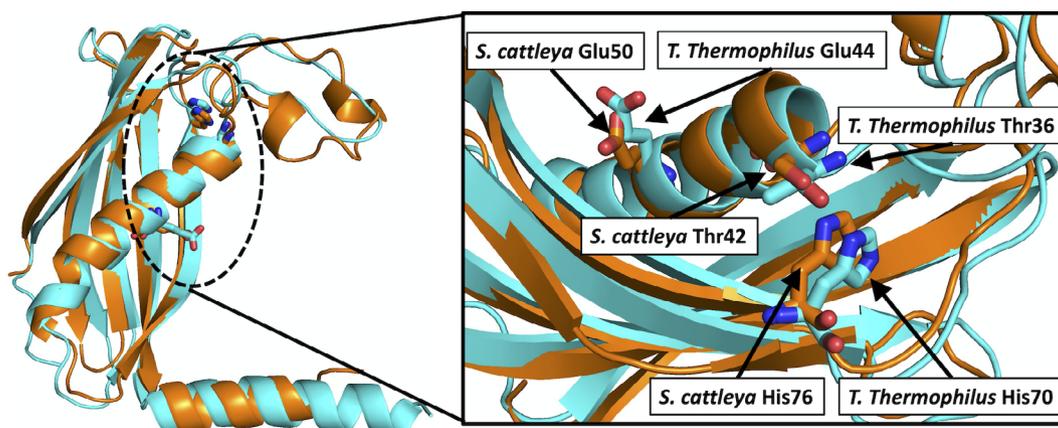


FIGURE 1 The catalytic residues of a HotDog fold enzyme in TE25 from *Thermus thermophilus* (cyan) were predicted based on known residues from a *Streptomyces cattleya* (orange) TE25 enzyme

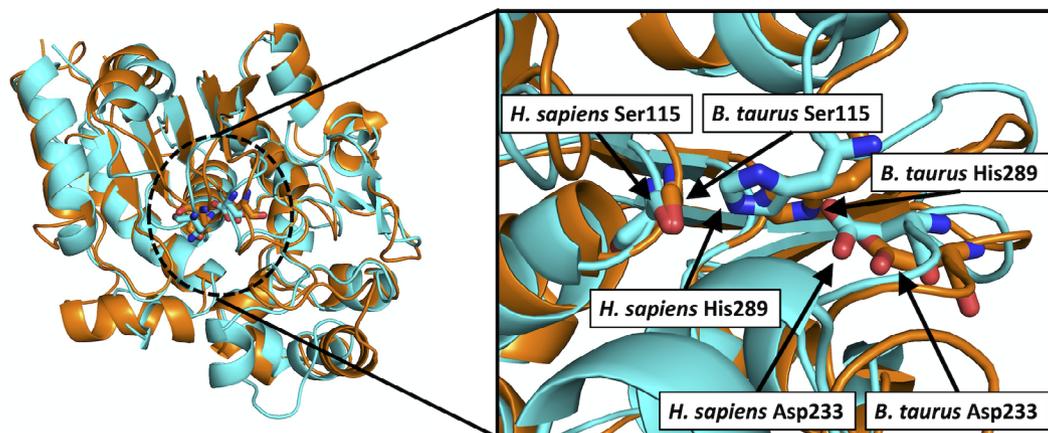


FIGURE 2 The catalytic residues of an α/β -hydrolase fold enzyme in TE20 from *Homo sapiens* (cyan) were predicted based on known residues from a *Bos taurus* (orange) TE20 enzyme

Them2, the structures for *M. musculus* Acot13 (2CY9) and *Danio rerio* Acot13 (4ORD) are predicted to have the same Asn50, His56, Gly57, Asp65 catalytic structure.¹²⁰ The position of these catalytic residues seem to be extremely highly conserved in this family; the position of the catalytic residues in 2CY9 and 2F0X are exactly the same and are only shifted by one position in 4ORD (e.g., Asp65 to Asp66).

In TE9, an *E. coli* enzyme (1S5U) is predicted to have catalytic residues Tyr14-Asp18-His25, based on a strong spatial correlation with the catalytic structure (Tyr7-Asp11-His18) of an *Helicobacter pylori* enzyme (2PZH) in the superimposed structures.⁵³

It was proposed for TE10 4-hydroxybenzoyl-CoA TEs (1LO7, 1LO8, 1LO9) that a helix dipole moment make the thioester carbonyl group more susceptible to a nucleophilic attack by Asp17.¹²³ We predict that Asp16 in an *Alkalihalobacillus halodurans* enzyme (5WH9) is catalytic, based on the Asp17 residue of a *Pseudomonas* TE (1BVQ).⁵⁶

TE11 TEs in *Arthrobacter* (1Q4S, 1Q4T, 1Q4U), *E. coli* K-12 (4K49), and *A. thaliana* At2g48320 (4K02) all have nearly identically positioned glycine and glutamic acid catalytic residues.^{35,124} The crystal structures of other members of this family spatially align well, and are predicted to have the same Gly-Glu catalytic residues (Table 3). Members of TE11 may also act as chain elongation and cyclization domains in certain synthetic pathways.³⁴

TE12 enzymes from *Synechocystis* (4K00) and *Prochlorococcus* (2HX5) bacteria have been crystallized. In 4K00, Asp16 was proposed to act as a nucleophile, while it is also possible that it acts as a base to attack the thioester through activation of a water molecule. The thioester oxygen atom could be stabilized by the amide hydrogen on Phe23. Also, Pro57, which positioned above the substrate moiety, may contribute to substrate specificity.³⁵

From the structures 1WLU, 1J1Y, 1WM6, 1WLV, and 1WN3, a study proposed that TE13 *Thermus thermophilus* PaaI TE hydrolyze substrates with an Asp48-activated water nucleophile.¹²⁶ By comparison, an *E. coli* PaaI structure (2FS2) with the *Arthrobacter* TE11 structures and site-directed mutagenesis, a mechanism similar to that in TE11 was proposed: Gly53 prepares the thioester for a nucleophilic attack from Asp61.¹²⁷

TE14, which has many bacterial sequences that have been less characterized than their plant counterparts, has a surprising breadth of substrate specificity.⁶⁰ In TE14, a site-directed mutagenesis study on a FatB enzyme from *A. thaliana* pointed to a Cys264, His229, and Asn227 papain-like catalytic triad.¹⁶² Another site-directed mutagenesis study on a FatB enzyme, from *Umbellulaia*

californica, proposed a catalytic network of Asp281, Asn283, His285, and Glu319.¹²⁸ More recently, structural predictions and site-directed mutagenesis resulted in identifying the catalytic residues of the *C. viscosissima* acyl-ACP TE.¹⁶

In TE15, a mechanism based on CalE7 enzyme (2W3X), which has no acidic residues in the catalytic region, was proposed: Asn19 and Arg37 hold the substrate while a water molecule or hydroxide anion acts as a nucleophile, and Tyr29 assists in decarboxylation.⁶² Asn, Arg, and Tyr residues in a *Micromonospora chersina* tebC (2XEM, 2XFL), as well as *Streptomyces globisporus* (4I4J) and *Actinomyces verrucosuspora* (5VPJ) TEs are predicted to be catalytic based on spatial correspondence with the superimposed *M. echinospora* structure (2W3X).

The crystal structure for TE24 is represented by Protein Data Bank (PDB) 2PFC and 3B18. The quaternary structure is formed by three dimers and has a long and narrow substrate-binding site. The catalytic site is formed by Asn83, Tyr87, Tyr33, and Met118 for subunit A and Tyr66, Thr70, His72, and Asn74 for subunit B.⁸⁰ Notably, the active site lacks acidic residues common to HotDog TEs, which is also observed in a TE24 *Streptomyces* enzyme.⁸²

In TE25, a *T. thermophilus* TE (2CWZ) is predicted to have Thr36, Glu44, and His70 as catalytic residues (see Figure 1) based on the spatial superimposition with the catalytic residues in *Streptomyces cattleya* fIK (3KUV).¹⁵⁸ The specificity for fluorine-containing compounds could arise from substrate binding through a hydrophobic pocket formed by a helical lid structure (side chains of Val46 and Val54), as well as by Val23, Leu26, Phe33, and Phe36 in *S. cattleya* fIK.⁸⁴

Family TE31 has Them4 and Them5 isoforms, which have been crystallized and are reported by the 4AE8 and 4AE7 structures, respectively, forming a homodimer unity. Their structures consist of a long central alpha helix surrounded by a six-stranded curved antiparallel beta-sheets. Both isoforms are formed by two active sites per homodimer at the end of each HotDog helix: His152, Gly153, Gly154/His158, Gly159, Gly160 (active site one), and Asp161, Thr177/Asp167, Thr183 (active site two).⁹²

2.2.2 | α/β hydrolase catalytic residues and mechanisms

The α/β -hydrolase fold,¹⁶³ found in TE2, TE16 to TE22, and TE26 to TE28, shows higher variation in RMSD_{ave} and P_{ave} values than the HotDog fold. Most α/β -hydrolase fold proteins, not only TEs, are present in the ESTHER database.¹⁶⁴ Two families, TE29 and TE30, based on sequence similarity, are likely to have α/β -hydrolase-like

folds; however, there are no available structures to confirm. α/β hydrolases have conserved catalytic residues: a nucleophile–histidine–acid triad.¹⁶³ Serine, cysteine, or aspartate can act as the nucleophile. There is a large variation of fold architecture and binding sites in α/β hydrolases.¹⁶⁵ In their catalytic mechanism, the acid stabilizes the histidine, which acts as a base by accepting a proton from the nucleophile, which forms a substrate intermediate that attacked by water. In PKSs or NRPs that make cyclic products, for example, in erythromycin biosynthesis,¹⁶⁶ a hydroxyl group from the substrate chain is used instead of a water molecule. Different cyclization mechanisms lead to a wide variety of PKS or NRP products.¹⁶⁷

The structure of TE2 is represented by 3HLK, which comes from human ACOT2, and 3K2I, which comes from human ACOT4. These structures are somewhat unique for this fold: in the primary structure for these enzymes the Asp residue precedes the His residue, where in all other α/β hydrolase TEs the His residue precedes the Asp residue.¹⁰⁹ The catalytic residues of 3K2I (Table 3) are predicted based on alignment with 3HLK.

In TE16, most structures show a consistent Ser-Asp-His catalytic triad: seen in the human FAS TE domain,^{129,168–170} the TE domain in *Bacillus* NRPSs surfactin and fengycin synthetases,^{130,131} the TE domain of the *Aspergillus* aflatoxin PKS,¹³² the TE domain of *Mycobacterium* PKSs involved in making mycolic acids,¹³³ and in the TE domain of NocB enzyme in *Nocardia*.¹³⁴ However, based on structural superimposition with TE16 structures with identified catalytic residues, we predict that the TE domain of an *Acinetobacter baumannii* NRPS enzyme (4ZXH, 4ZXI)¹⁷¹ has a Cys-Asp-His catalytic triad (Table 3).

TE17 has enzymes, which are the TE domain of macrocycle-forming PKSs, such as of 6-deoxyerythronolide B synthase from *S. erythraea*,^{135,136,172,173} picromycin synthase from *S. venezuelae*,^{136,174,175} and tautomycin synthase.¹³⁷ They all show a consistent Ser-Asp-His catalytic triad.

Member of TE18 with crystal structures are type II TEs, a class of enzyme responsible for a variety of functions, primarily maintenance of biosynthetic pathways through release of undesired intermediates from carrier protein domains.^{28,66,67,139–142,176} A lid-flip conformational change is present in these enzymes and the Ser-Asp-His catalytic triad is conserved. This can be seen in the surfactin synthase from *Bacillus subtilis*,¹⁷⁷ from the rifamycin biosynthetic cluster from *A. mediterranei*,⁶⁷ the borrelidin biosynthetic cluster from *Streptomyces*,²⁸ in the prodiginine biosynthetic pathway in *Streptomyces coelicolor*,¹³⁹ and in ClbQ and YbtT enzymes in *E. coli*.^{141,142} This also holds true in a human TE II and in a TesA from *M. tuberculosis*.^{66,140}

In family TE19, a single structure is known, that of a *Vibrio harveyi* TE, which also has the Ser-Asp-His catalytic triad.¹⁴³

Families TE20, TE21, and TE22 all share the characteristic Ser-Asp-His catalytic triad. Comparison of tertiary structures within each family leads us to predict that this Ser-Asp-His catalytic triad is consistent for all structures (see Table 3 and Figure 2).

TE21 includes mainly eukaryotic acyl-protein hydrolases, as well as enzymes with different functions. The carboxylesterase from *P. fluorescens* has very little activity on triacylglycerides with fatty acids longer than four carbons, likely due to the loops constraining the active-site cleft.¹⁴⁵ A closely related human enzyme, hAPT1, originally thought to be a lysophospholipase, has been shown to have stronger TE activity.¹⁴⁶ Another APT, from *Francisella tularensis*, has a similar substrate specificity profile to both of the aforementioned enzymes, though unlike *P. fluorescens*, it lacks a lid domain.¹⁴⁸ This was confirmed by another study that examined the mechanism of isoform-selective inhibitors on human APT1.¹⁴⁹ The carboxylesterase from *P. aeruginosa* was shown to have no activity on triacylglycerols, and a preference for eight-carbon acyl substrates. The human lysophospholipase A2 is a cystolic serine hydrolase partially responsible for lysophospholipid metabolism.¹⁵⁰ All of these structures follow the Ser-Asp-His catalytic motif.

Members of TE22 are involved in glutathione-dependent formaldehyde detoxification, and many of the crystal structures in this family are of S-formylglutathione hydrolase (SFGH) enzymes. These have been studied in a variety of species: *S. cerevisiae*,¹⁵¹ *Agrobacterium fabrum* str. C58,¹⁵² *P. translucida* TAC125,¹⁵⁵ *Shewanella frigidimarina*,¹⁵⁷ and *N. meningitidis* MC58.¹⁵⁶ Other functions are present in this family as well: (a) a human esterase has been studied because it is relevant to retinoblastoma,¹⁵³ and (b) an oil-degrading bacterium, *O. antarctica*, expresses an enzyme with carboxylesterase and TE activity.¹⁵⁴ TE22 enzymes have the characteristic Ser-Asp-His catalytic triad. Based on this, the catalytic structure of a *S. cerevisiae* SFGH (4FLM) is predicted as Ser161-Asp241-His276 (Table 3).

A study on the only crystal structures found for this family, ybF from *E. coli* (3BF7, 3BF8), suggests that this family is unique within the α/β hydrolase TEs: rather than the typical Ser-Asp-His catalytic triad, this family seems to have a Ser89-Asp113-Ser206-His234 catalytic tetrad. The α/β hydrolase domain of these structures gives good alignment with other canonical α/β hydrolases. However, the Asp113 residue, which normally lies above or parallel to the His234 imidazole rings, is located in the lower section of the His imidazole ring. The expected position for the Asp113 residue is instead

occupied by Ser206, which is well conserved in the ybff enzymes.⁸⁵

The structure of TE27 enzymes is described by a *M. musculus* *ABHD10*, which shows a Ser-His-Asp catalytic triad. The location of the catalytic serine residue suggests a hydrophobic interaction between the lipid substrate and the interior surface of the protein. A “cap domain” above the catalytic triad forms a binding pocket and affects substrate accessibility.⁸⁷ We predict that Ser113-Asp216-His246 is the catalytic triad in an *A. vitis* enzyme based on comparison to the *M. musculus* *ABHD10* enzyme.⁸⁷

Families TE28 and TE29 have no crystal structures. TE28 shows sequence similarity with a putative α/β hydrolase fold enzyme, and their structure and mechanisms still unknown despite a close relationship with FASs.⁸⁸ TE29 may also have an α/β hydrolase fold, as was predicted from gene *ABHD17C*.⁸⁹

The structure of an CitA enzyme in TE30, predicted by homology from a co-expression of the PKS gene, suggests a Ser122-His235-Asp207 as catalytic triad.⁹¹

2.2.3 | Catalytic residues and mechanisms in other folds

TEs are found in the NagB (TE1) and SGNH (TE3) folds.^{110–114} In TE1, which also includes acyl-CoA transferases, we predict that the catalytic residues of a putative acetyl-CoA hydrolase from *Porphyromonas gingivalis* (2NVV) and a CoA transferase from *P. aeruginosa* (2G39) are Val259-Glu284-Asn337-Gly378 and Ile264-Glu288-Asn341-Gly382, respectively, based on those known from *A. acetii* AarCH6 structures (4EU3, 5DDK).^{107,108}

In TE3, comparison to available structures—*E. coli* tesA (e.g., 1IVN, 1JRL)¹¹⁰ and *Pseudoalteromonas* estA (3HP4)¹¹¹—reveals the likely catalytic residues for an *E. coli* TE (6LFB, 6LFC) and *A. indicum* AlinE4 esterase (6IQ9, 6IQA, 6IQB) are Ser10-Asp154-His157 and Ser13-Asp162-His165, respectively. TesA enzymes were found to have a Ser-His-Asp catalytic triad similar to those in α/β hydrolases.¹¹⁰ The crystal structure of TesA from *E. coli* was found to be particularly compact and rigid, which likely pushes the substrate specificity toward smaller chain lengths.¹¹² It has also proved to be a useful candidate for attempts at engineering TEs to produce specific lengths of free fatty acids.¹¹³ Other SGNH fold TEs, CrmE10 and AlinE4 were similarly susceptible to engineering for increased enzymatic activity.¹¹⁴

Two families have the β -lactamase fold: TE23 and TE32. The structures in TE23 are significantly less well conserved than those in TE32. TE23 hydroxyglutathione hydrolases, which include glyoxalase II enzymes, have a

metallo- β -lactamase fold, and their mechanisms are very different from the rest of TEs that do not have catalytic metal ions. Crystal structures of human glyoxalase II (1QH3, 1QH5) reveal two zinc ions with octahedral coordination, interacting with His and Asp residues. Based on this, a study proposed that a hydroxide ion bonded with both ions attacks the carbonyl carbon atom of the glutathione thioester substrate, forming a tetrahedral intermediate, followed by breakage of the C—S bond.¹⁷⁸ In mitochondrial glyoxalase II from *A. thaliana* (1XM8, 2Q42), the zinc ions were also coordinated by His and Asp residues, but were in trigonal bipyramidal and tetrahedral geometries.¹⁷⁹ Another glyoxylase II enzyme, from *Salmonella typhimurium* (2QED), was proposed to have an uncommon metal affinity: a diiron, dimanganese, or hybrid Fe/Mn.¹⁸⁰ A unique member of the family, a persulfide dioxygenase from *Myxococcus xanthus* (4YSB), has a single ion in the active site with a two-His and one-carboxylate triad coordination pattern.¹⁸¹

Enzymes in TE32 have monomeric metallo- β -lactamase fold structures, with an Fe(II)Fe(III) center in the active site and an $\alpha\beta/\alpha\beta$ sandwich core. All the resolved structures in this family are PqsE enzymes from *P. aeruginosa*, a human pathogen of particular interest due to its tendency for antibiotic resistance.¹⁸² The active center of the enzyme is covered by a lid formed by two α -helices in the C-terminal region, affecting substrate access.⁹⁴ It has also been demonstrated that PqsE has a role in alkylquinolone biosynthesis.¹⁸³

Although TE33 includes no crystal structures, a mechanism has been proposed, which shows an active site His acting as a base, with the substrate hydroxyl forming a hydrogen bond with a histidine residue.^{184–186} A nucleophilic attack from a deprotonated hydroxyl at the carbonyl of an acyl-CoA thioester was described, as was the involvement of an Asp residue in the stabilization of the structure within the active site.^{96,184,185,187}

Crystal structure 5VXS represents a member from TE34 and reveals a homotrimer with a substrate-bound cavity located between the N-terminal from one subunit and the C-terminal from the subsequent subunit. The N-terminal forms a $\beta_8\alpha_8$ -TIM barrel fold and the C-terminal is characterized by a lid-domain consisting of two helices connect by a β -hairpin loop. The β -hairpin loop presents a highly conserved Asp320 that removes a proton from the substrate during the catalytic activity.^{97,103,188,189} In TE34, the catalytic residues for a *M. tuberculosis* (6AQ4), *Cereibacter sphaeroides* (4L9Y, 4L9Z), and *M. extorquens* (5UGR) enzymes are predicted to be Asp261, Asp299, and Asp304, respectively, based on comparison to human CLYBL structure (5VXS).⁹⁷ The catalytic residues for the remaining family members could not be confidently

predicted by structural comparison. Two of these are CitE proteins from *M. tuberculosis*: one study (1U5H) predicts that the catalytic site is in a hydrophobic cavity formed by the C-terminal tips of the TIM β -barrel,¹⁹⁰ while another study (6AQ4) shows that the active site contains an Mg^{2+} ion coordinated by the ligand, Glu112, Asp138, and two water molecules.⁹⁹ Closely related to 1U5H is *Y. pestis* RipC (3QLL), for which the active site is similarly predicted. However, it is also suggested that the active site for 3QLL may be formed through an intermonomer interaction.¹⁰⁰

The structure 6AUN in TE35 is characterized by the presence of an Ankyrin domain, a 33-residue helix-turn-helix structure followed by a hairpin-like loop, and a catalytic domain. Regarding the catalytic mechanisms, a dyad formed by Ser-Asp is responsible for the lipid hydrolysis.^{159,191}

2.3 | TE phylogeny

TE families show convergent evolution because enzymes from different families, with different folds, have the same activity (thioester hydrolysis) despite a wide variety of substrates. Divergent evolution is evidenced by the many substrates that enzymes in single families show activity to, even though they have similar primary and tertiary structures and mechanisms. A phylogenetic analysis of TEs exhibiting the two main folds, α/β hydrolase and HotDog, was performed.

All the amino acid sequences with experimentally confirmed TE activity which are members of TE families with a HotDog fold (TE4–TE15, TE24, TE25, and TE31) were aligned and a phylogenetic tree was constructed, shown in Figure 3.

The HotDog fold cladogram confirms the previously reported TE clans,¹² since families within a clan are grouped in the same clade. TE clans were previously identified with structural superimpositions, not by phylogeny. Figure 3 suggests that TE15 is a part of Clan TE-A, which includes TE5, TE9, TE10, TE12 as well, and is similar to the 4HBT-like SCOP family. TE8, TE11, and TE13 were grouped into clan TE-B in previous work,¹² and form part from same clade in Figure 3. The proximity of sequences from TE25 and TE31 to this clade suggests that they also form part of TE-B. TE14 members present a common ancestor with clan TE-B sequences. However, structural differences and catalytic mechanisms do not support TE14 inclusion in TE-B.

All TE4 members share a common ancestor and present high sequence similarity, forming a single clade in Figure 3. Members of TE6 and TE7 share a common

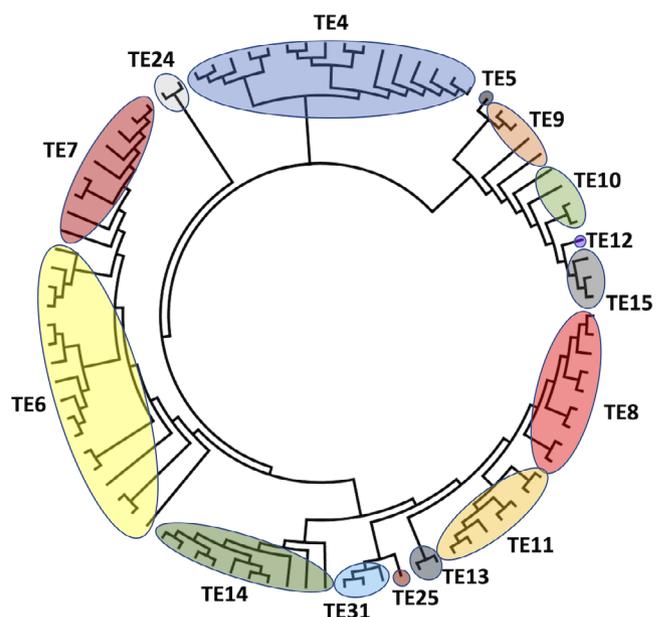


FIGURE 3 Unrooted phylogenetic tree of HotDog fold thioesterases that have been experimentally characterized

ancestor, but the lack of crystal structures in TE7 does not allow for inference of a new clan based on structure superimposition. Enzymes in TE6 and TE7 appear in the same clade, suggesting that ortholog sequences from a speciation event at the branching point. At least TE24, despite belonging to the HotDog fold, seems to diverge from the common ancestor prior to other clades and is represented as an outgroup.

All the amino acid sequences with experimentally confirmed TE activity which are members of TE families with an α/β hydrolase fold (TE2, TE16, TE17, TE18, TE19, TE20, TE21, TE22, TE26, TE27, TE28, TE29 and TE30) were aligned, and a phylogenetic tree was constructed, shown in Figure 4.

The TE families were grouped previously in two clans: TE-C (TE16, TE17, and TE18) and TE-D (TE20 and TE21).¹² Unlike for clans with TE HotDog enzymes, they are not grouped in the same clade, despite structural and functional similarity, suggesting a convergent evolution event.

Members of TE2 are phylogenetically close to TE16, TE17, and TE18, TE19 is close from TE20, and as TE30 is close to TE21, but with not enough structural criteria for it to form part of clan TE-C. The sequences in TE22, TE26, TE27, and TE28 share a common ancestor, forming a well-defined clade that is closer to TE21, TE29, and TE30 than any TE-C family member. Apparently, the α/β hydrolase fold facilitates nonrestricted acyl-ACP hydrolase or acyl-CoA hydrolase activity, increasing the variety of substrate options for this group.

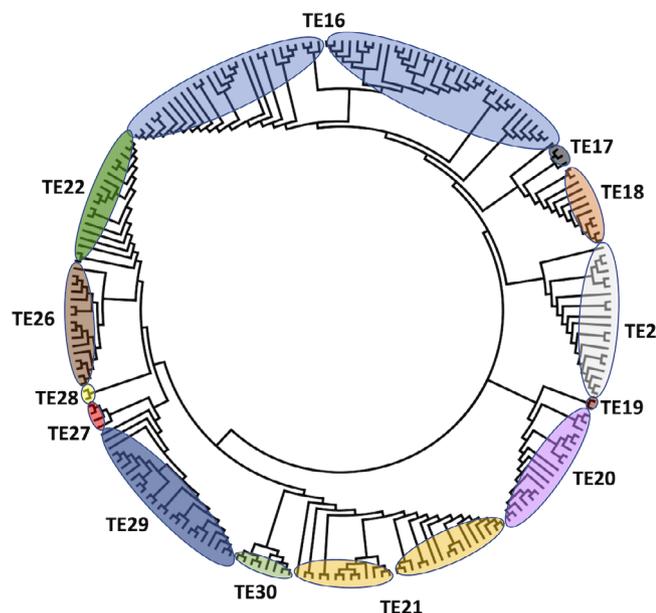


FIGURE 4 Unrooted phylogenetic tree of α/β hydrolase fold thioesterases that have been experimentally characterized

2.4 | Updated ThYme database

All the sequences and structures in the TE families described here appear in the ThYme database,¹³ which is in the process of being completely updated and has a new home at the University of Nevada, Reno (<http://thyme.engr.unr.edu>). Families, their member sequences, taxonomical data, accession codes, and protein names can be viewed using the ThYme database online interface. The database has links to UniProt,¹⁹² GenBank,¹⁹³ and PDB¹⁹⁴ databases. Although the content of families will be updated automatically, human judgment will still be necessary for adding, merging, or deleting families.

In the new ThYme website, each enzyme class (e.g., TEs) will have an interactive interface where users can view content of a single family or multiple families. Each unique sequence is displayed as a row containing: the family, the organism, protein names, protein identifiers, protein evidence information, crystal structures, gene names, as well as gene and pathway identifiers. Each entry will display, at the minimum, the family and a protein identifier; all other fields will be populated if suitable data is known. The content has multiple search fields such as name, identifier, or sequence in FASTA format. Results can be narrowed to show only entries with evidence at protein level or known crystal structures.

3 | CONCLUSIONS

TE families have been updated through analysis of the primary structures of all known TE sequences. New

families have been proposed, and all sequences and structures are classified into new, or previously identified, families. This system of classification provides a standardized nomenclature and a means to predict the tertiary structure, function, and mechanism of a TE sequence that has not been experimentally characterized. These assertions are supported by family members displaying a high degree of primary and tertiary structural similarity, highly conserved active sites and catalytic residues, and consistent mechanisms. Examination of families that share a fold reveals some similarity in primary and tertiary structures, catalytic residues and active sites, and mechanisms. Convergent and divergent evolution is suggested from phylogenetic analyses of TEs whose structures have the two main structural folds.

4 | MATERIALS AND METHODS

For a sequence to be considered a member of a family it must have a strong sequence similarity ($\sim 30\%$), a nearly identical tertiary structure to other structures in the family, and catalytic residues in the same locations as the other members of that family.

The protocol by which the new TE families were identified is described: (a) enzyme sequences experimentally confirmed to have TE activity are gathered and those present in a previously existing family (TE1–TE23) were discarded; (b) each of the remaining TE sequences are independently processed by the Basic Local Alignment Search Tool (BLAST)¹⁹⁵ and results were compared with the other sequences' results to identify the representative sequences that will originate new families; (c) the catalytic domains of the representative sequences were processed by BLAST to populate potential new families; (d) the number of shared sequences were counted for all permutations of pairs of potential new families, highly similar families ($>15\%$ sequences in common) were merged; (e) intra-family congruity and inter-family uniqueness were confirmed by tertiary structure superimposition, comparison of catalytic residue position and identity, multiple sequence alignments (MSAs), and final examination of shared sequences between all possible pairs of families; and (f) sequences common to multiple families are assigned to the family with the highest sequence similarity.

4.1 | Sequence selection and BLAST searches

Enzyme sequences experimentally confirmed to have TE activity were extracted from the Swiss-Prot database in

Uniprot, which contains only reviewed sequences and has a higher level of annotation. Possible TEs were identified by a label of EC 3.1.2.1 to EC 3.1.2.32, EC 3.1.2.–, or having “TE” in the description, as well as having “evidence at protein level.” Less stringently verified sequences, like those with “evidence at transcript level” or “inferred from homology,” as well as fragments or theoretical proteins, were disregarded. The primary sequences meeting the criteria, and not in TE1–TE23, were collected, resulting in ~200 new query sequence candidates.

Each of these sequences was subjected to a BLAST search against the National Center for Bio-technology Information's (NCBI) GenBank nr peptide sequence database using the protein–protein algorithm.¹⁹⁶ These BLAST searches were completed using a local instance of blast-2.9.0-2 and the nr database, both downloaded from NCBI on a Unix system. Previously, an *E*-value cutoff of 1×10^{-3} was used¹²; however, due to the growth of the nr database by ~3 orders of magnitude, an *E*-value of 1×10^{-7} was used to capture as many sequences with the required similarity as possible while minimizing the number of redundant sequences. The highest Max Target Sequences was used to capture all sequences within an *E*-value of 1×10^{-7} . Other parameters were left at default settings.

BLAST results were compared against each other to check for common sequences and identify the representative sequences that results in the lowest number of BLAST results with no overlapping, common sequences. The query sequences of unique, nonredundant BLAST results become the representative sequences that will originate new families from all confirmed TE sequences. The referenced literature in Uniport is checked to confirm experimental TE activity. The catalytic domain of each of the new representative sequences, identified in Pfam-A,¹⁹⁷ were used to populate the prospective new families with BLAST as described above.

4.2 | Comparison of tertiary structures

All known tertiary structures in each family was obtained from the Research Collaboratory for Structural Bioinformatics PDB.¹⁹⁴ Enzyme tertiary structures were reviewed to exclude fragments, putative proteins, and non-TE domains from multidomain proteins from any structural comparisons.

All monomer structures were extracted, and for each family a reference structure was selected, which served as the pivot around which other monomers were superimposed. The shortest monomer in each family was selected as the pivot to ensure consistent alignment of

the core structure and allow for uniform structural similarity calculations. All monomers within each family were superimposed using MultiProt¹⁹⁸ with OnlyRefMol set to 1, Scoring set to 2, and all other parameters left at default.

A root mean square distance (RMSD) of the superimposed tertiary structures in each family with more than one structure was done to quantify structural similarity. For RMSD calculations, the distances between corresponding alpha carbon atoms (C_{α}) from two superimposed structures (pivot and subject) were calculated. A cutoff distance, calculated as the average distance between sequential C_{α} s in the pivot structure, was used to determine corresponding C_{α} s between the pivot and subject structures. Any pairs more distant than the cutoff were not considered to be corresponding and were not used in the RMSD calculation. The percentage value (*P*) of C_{α} s used to calculate the RMSD implies the significance of the RMSD calculation. For a given family, the pivot structure was superimposed to all other structures, resulting in $n - 1$ calculations, where n is the number of monomers being compared within that family. For families where $n > 2$, the average RMSD and *p* values (RMSD_{ave} and *P*_{ave}, respectively) were calculated.

4.3 | Multiple sequence alignments and phylogenetic trees

Phylogeny was initialized by a Multiple Sequence Alignment by MUSCLE v3.8.31¹⁹⁹ with default parameters using the amino acid sequences from the TE catalytic domain. A unrooted dendrogram was built using MEGA X²⁰⁰ with maximum likelihood as statistical method. The reliability of the tree was estimated by the bootstrap method with 1,000 replicates. The tree was visualized and edited using the FigTree v1.4.4.

ACKNOWLEDGMENT

This work was supported by the U.S. National Science Foundation (Award 2001385).

AUTHOR CONTRIBUTIONS

Benjamin T. Caswell: Conceptualization (equal); data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); software (equal); validation (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal). **Caio C. de Carvalho:** Conceptualization (equal); data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); validation (equal); visualization (equal); writing – original draft (equal). **Hung Nguyen:** Data curation (equal); investigation (equal); methodology

(equal); software (equal); writing – original draft (equal). **Monikrishna Roy:** Data curation (equal); investigation (equal); methodology (equal); software (equal); writing – original draft (equal). **Tin Nguyen:** Data curation (equal); funding acquisition (equal); investigation (equal); methodology (equal); project administration (equal); software (equal); supervision (equal); writing – original draft (equal). **David C. Cantu:** Conceptualization (equal); data curation (equal); formal analysis (equal); funding acquisition (equal); investigation (equal); methodology (equal); project administration (equal); supervision (equal); validation (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal).

ORCID

David C. Cantu  <https://orcid.org/0000-0001-9584-5062>

REFERENCES

- Lennen RM, Pflieger BF. Engineering *Escherichia coli* to synthesize free fatty acids. *Trends Biotechnol.* 2012;30:659–667.
- Zhang X, Li M, Agrawal A, San K-Y. Efficient free fatty acid production in *Escherichia coli* using plant acyl-ACP thioesterases. *Metab Eng.* 2011;13:713–722.
- Tang M-C, Fischer CR, Chari JV, et al. Thioesterase-catalyzed aminoacylation and thiolation of polyketides in fungi. *J Am Chem Soc.* 2019;141:8198–8206.
- Paiva P, Medina FE, Viegas M, et al. Animal fatty acid synthase: A chemical nanofactory. *Chem Rev.* 2021;121:9502–9553.
- Guevara-Martínez M, Perez-Zabaleta M, Gustavsson M, Quillaguamán J, Larsson G, van Maris AJA. The role of the acyl-CoA thioesterase “YciA” in the production of (R)-3-hydroxybutyrate by recombinant *Escherichia coli*. *Appl Microbiol Biotechnol.* 2019;103:3693–3704.
- De Marcos LC, van Roermund CWT, Postis VLG, et al. Intrinsic acyl-CoA thioesterase activity of a peroxisomal ATP binding cassette transporter is required for transport and metabolism of fatty acids. *Proc Natl Acad Sci.* 2013;110:1279–1284.
- Hunt MC, Siponen MI, Alexson SEH. The emerging role of acyl-CoA thioesterases and acyltransferases in regulating peroxisomal lipid metabolism. *Biochimica et Biophysica Acta.* 2012;1822:1397–1410.
- Widhalm JR, Ducluzeau A-L, Buller NE, Elowsky CG, Olsen LJ, Basset GJC. Phylloquinone (vitamin K1) biosynthesis in plants: Two peroxisomal thioesterases of lactobacillales origin hydrolyze 1,4-dihydroxy-2-naphthoyl-coa. *Plant J.* 2012;71:205–215.
- Song F, Thoden JB, Zhuang Z, et al. The catalytic mechanism of the hotdog-fold enzyme superfamily 4-hydroxybenzoyl-CoA thioesterase from *Arthrobacter* sp. strain SU. *Biochemistry.* 2012;51:7000–7016.
- Jones ML, Collins MO, Goulding D, Choudhary JS, Rayner JC. Analysis of protein palmitoylation reveals a pervasive role in plasmodium development and pathogenesis. *Cell Host Microbe.* 2012;12:246–258.
- Tillander V, Alexson SEH, Cohen DE. Deactivating fatty acids: Acyl-CoA thioesterase-mediated control of lipid metabolism. *Trends Endocrinol Metab.* 2017;28:473–484.
- Cantu DC, Chen Y, Reilly PJ. Thioesterases: A new perspective based on their primary and tertiary structures. *Protein Sci.* 2010;19:1281–1295.
- Cantu DC, Chen Y, Lemons ML, Reilly PJ. ThYme: A database for thioester-active enzymes. *Nucleic Acids Res.* 2011;39:D342–D346.
- Blatti JL, Beld J, Behnke CA, Mendez M, Mayfield SP, Burkart MD. Manipulating fatty acid biosynthesis in microalgae for biofuel through protein-protein interactions. *PLoS One.* 2012;7:1–12.
- Marika Z, Nathan R, Aashna S, et al. Chimeric fatty acyl-acyl carrier protein thioesterases provide mechanistic insight into enzyme specificity and expression. *Appl Environ Microbiol.* 2018;84:e02868–e02817.
- Jing F, Yandean-Nelson MD, Nikolau BJ. Identification of active site residues implies a two-step catalytic mechanism for acyl-ACP thioesterase. *Biochem J.* 2018;475:3861–3873.
- Banerjee D, Jindra MA, Linot AJ, Pflieger BF, Maranas CD. EnZymClass: Substrate specificity prediction tool of plant acyl-ACP thioesterases based on ensemble learning. *Curr Res Biotechnol.* 2022;4:1–9.
- McMahon MD, Prather KL. Functional screening and in vitro analysis reveal thioesterases with enhanced substrate specificity profiles that improve short-chain fatty acid production in *Escherichia coli*. *Appl Environ Microbiol.* 2014;80:1042–1050.
- Hickman TWP, Baud D, Benhamou L, Hailes HC, Ward JM. Characterisation of four hotdog-fold thioesterases for their implementation in a novel organic acid production system. *Appl Microbiol Biotechnol.* 2020;104:4397–4406.
- Sonntag F, Buchhaupt M, Schrader J. Thioesterases for ethylmalonyl-CoA pathway derived dicarboxylic acid production in *Methylobacterium extorquens* AM1. *Appl Microbiol Biotechnol.* 2014;98:4533–4544.
- Sánchez-Reyez A, Batista-García RA, Valdés-García G, et al. A family 13 thioesterase isolated from an activated sludge metagenome: Insights into aromatic compounds metabolism. *Proteins Struct Funct Bioinform.* 2017;85:1222–1237.
- Khandokar YB, Srivastava P, Cowieson N, et al. Structural insights into GDP-mediated regulation of a bacterial acyl-CoA thioesterase. *J Biol Chem.* 2017;292:20461–20471.
- Pulsifer IP, Lowe C, Narayanan SA, et al. Acyl-lipid thioesterase1-4 from *Arabidopsis thaliana* form a novel family of fatty acyl-acyl carrier protein thioesterases with divergent expression patterns and substrate specificities. *Plant Mol Biol.* 2014;84:549–563.
- Yu G, Nguyen TTH, Guo Y, et al. Enzymatic functions of wild tomato methylketone synthases 1 and 2. *Plant Physiol.* 2010;154:67–77.
- Khuat VLU, Bui VTT, Tran HTD, et al. Characterization of *Solanum melongena* thioesterases related to tomato methylketone synthase 2. *Genes.* 2019;10:549.
- Tran HT, Le NT, Khuat VL, Nguyen TT. Identification and functional characterization of a soybean (*Glycine max*) thioesterase that acts on intermediates of fatty acid biosynthesis. *Plants.* 2019;8:397.

27. Gao T, Meng Q, Gao H. Thioesterase YbgC affects motility by modulating c-di-GMP levels in *Shewanella oneidensis*. *Sci Rep*. 2017;7:3932.
28. Curran SC, Pereira JH, Baluyot MJ, et al. Structure and function of BorB, the type II thioesterase from the Borrelidin biosynthetic gene cluster. *Biochemistry*. 2020;59:1630–1639.
29. Kerviel V, Hérault J, Dumur J, Ergan F, Poisson L, Loiseau C. Cloning and expression of a gene from *Isochrysis galbana* modifying fatty acid profiles in *Escherichia coli*. *J Appl Phycol*. 2014;26:2109–2115.
30. Swarbrick CMD, Nanson JD, Patterson EI, Forwood JK. Structure, function, and regulation of thioesterases. *Prog Lipid Res*. 2020;79:101036.
31. Swarbrick CMD, Perugini MA, Cowieson N, Forwood JK. Structural and functional characterization of TesB from *Yersinia pestis* reveals a unique octameric arrangement of hot-dog domains. *Acta Crystallogr Sect D Biol Crystallogr*. 2015; 71:986–995.
32. Swarbrick CMD, Bythrow GV, Aragao D, Germain GA, Quadri LEN, Forwood JK. Mycobacteria encode active and inactive classes of TesB fatty-acyl CoA thioesterases revealed through structural and functional analysis. *Biochemistry*. 2017;56:1460–1472.
33. Swarbrick CMD, Roman N, Cowieson N, et al. Structural basis for regulation of the human acetyl-CoA thioesterase 12 and interactions with the steroidogenic acute regulatory protein-related lipid transfer (START) domain. *J Biol Chem*. 2014;289:24263–24274.
34. Mori S, Simkhada D, Zhang H, et al. Polyketide ring expansion mediated by a thioesterase, chain elongation and cyclization domain, in azinomycin biosynthesis: Characterization of AziB and AziG. *Biochemistry*. 2016;55:704–714.
35. Furt F, Allen WJ, Widhalm JR, et al. Functional convergence of structurally distinct thioesterases from cyanobacteria and plants involved in phyloquinone biosynthesis. *Acta Crystallogr Sect D Biol Crystallogr*. 2013;69:1876–1888.
36. Lee FJ, Lin LW, Smith JA. A glucose-repressible gene encodes acetyl-CoA hydrolase from *Saccharomyces cerevisiae*. *J Biol Chem*. 1990;265:7413–7418.
37. Fleck CB, Brock M. Re-characterisation of *Saccharomyces cerevisiae* Ach1p: Fungal CoA-transferases are involved in acetic acid detoxification. *Fungal Genet Biol*. 2009;46: 473–485.
38. Hunt MC, Rautanen A, Westin MAK, Svensson LT, Alexson SEH. Analysis of the mouse and human acyl-CoA thioesterase (ACOT) gene clusters shows that convergent, functional evolution results in a reduced number of human peroxisomal ACOTs. *FASEB J*. 2006;20:1855–1864.
39. Johnson MR, Barnes S, Kwakye JB, Diasio RB. Purification and characterization of bile acid-CoA:amino acid N-acyltransferase from human liver. *J Biol Chem*. 1991;266: 10227–10233.
40. Cho H, Cronan JE. Defective export of a periplasmic enzyme disrupts regulation of fatty acid synthesis. *J Biol Chem*. 1995; 270:4216–4219.
41. Karasawa K, Yokoyama K, Setaka M, Nojima S. The *Escherichia coli* pldC gene encoding lysophospholipase L1 is identical to the apeA and tesA genes encoding protease I and thioesterase I, respectively. *J Biochem*. 1999;126:445–448.
42. Jones JM, Nau K, Geraghty MT, Erdmann R, Gould SJ. Identification of peroxisomal acyl-CoA thioesterases in yeast and humans. *J Biol Chem*. 1999;274:9216–9223.
43. Naggert J, Narasimhan ML, DeVeaux L, et al. Cloning, sequencing, and characterization of *Escherichia coli* thioesterase II. *J Biol Chem*. 1991;266:11044–11050.
44. Nie L, Ren Y, Schulz H. Identification and characterization of *Escherichia coli* thioesterase III that functions in fatty acid β -oxidation. *Biochemistry*. 2008;47:7744–7751.
45. Zhuang Z, Song F, Zhao H, et al. Divergence of function in the hot dog fold enzyme superfamily: The bacterial thioesterase YciA. *Biochemistry*. 2008;47:2789–2796.
46. Yokoyama T, Choi K-J, Bosch AM, Yeo H-J. Structure and function of a *Campylobacter jejuni* thioesterase Cj0915, a hexameric hot dog fold enzyme. *Biochim Biophys Acta - Proteins Proteom*. 2009;1794:1073–1081.
47. Yamada J, Furihata T, Tamura H, Watanabe T, Suga T. Long-chain acyl-CoA hydrolase from rat brain cytosol: Purification, characterization, and immunohistochemical localization. *Arch Biochem Biophys*. 1996;326:106–114.
48. Adams SH, Chui C, Schilbach SL, et al. BFIT, a unique acyl-CoA thioesterase induced in thermogenic brown adipose tissue: Cloning, organization of the human gene and assessment of a potential link to obesity. *Biochem J*. 2001;360:135–142.
49. Suematsu N, Isohashi F. Molecular cloning and functional expression of human cytosolic acetyl-CoA hydrolase. *Acta Biochim Pol*. 2006;53:553–561.
50. Tillander V, Arvidsson Nordström E, Reilly J, et al. Acyl-CoA thioesterase 9 (ACOT9) in mouse may provide a novel link between fatty acid and amino acid metabolism in mitochondria. *Cell Mol Life Sci*. 2014;71:933–948.
51. Poupon V, Bègue B, Gagnon J, Dautry-Varsat A, Cerf-Bensussan N, Benmerah A. Molecular cloning and characterization of MT-ACT48, a novel mitochondrial acyl-CoA thioesterase. *J Biol Chem*. 1999;274:19188–19194.
52. Wei J, Kang HW, Cohen DE. thioesterase superfamily member 2 (Them2)/acyl-CoA thioesterase 13 (Acot13): A homotetrameric hotdog fold thioesterase with selectivity for long-chain fatty acyl-CoAs. *Biochem J*. 2009;421:311–322.
53. Angelini A, Cendron L, Goncalves S, Zanotti G, Terradot L. Structural and enzymatic characterization of HP0496, a YbgC thioesterase from *Helicobacter pylori*. *Proteins Struct Funct Bioinform*. 2008;72:1212–1221.
54. Zhuang Z, Song F, Martin BM, Dunaway-Mariano D. The YbgC protein encoded by the ybgC gene of the tol-pal gene cluster of *Haemophilus influenzae* catalyzes acyl-coenzyme A thioester hydrolysis. *FEBS Lett*. 2002;516:161–163.
55. Ben-Israel I, Yu G, Austin MB, et al. Multiple biochemical and morphological factors underlie the production of methylketones in tomato trichomes. *Plant Physiol*. 2009;151: 1952–1964.
56. Benning MM, Wesenberg G, Liu R, Taylor KL, Dunaway-Mariano D, Holden HM. The three-dimensional structure of 4-hydroxybenzoyl-CoA thioesterase from *Pseudomonas* sp. strain CBS-3. *J Biol Chem*. 1998;273:33572–33579.
57. Damien L, Aurélie B, Emmanuelle B. The hotdog thioesterase EntH (YbdB) plays a role in vivo in optimal enterobactin biosynthesis by interacting with the ArCP domain of EntB. *J Bacteriol*. 2007;189:7112–7126.

58. Widhalm JR, van Oostende C, Furt F, Basset GJC. A dedicated thioesterase of the hotdog-fold family is required for the biosynthesis of the naphthoquinone ring of vitamin K₁. *Proc Natl Acad Sci*. 2009;106:5599–5603.
59. Ferrández A, Miñambres B, García B, et al. Catabolism of phenylacetic acid in *Escherichia coli*: Characterization of a new aerobic hybrid pathway. *J Biol Chem*. 1998;273:25974–25986.
60. Jing F, Cantu DC, Tvaruzkova J, et al. Phylogenetic and experimental characterization of an acyl-ACP thioesterase family reveals significant diversity in enzymatic specificity and activity. *BMC Biochem*. 2011;12:44.
61. Jones A, Davies HM, Voelker TA. Palmitoyl-acyl carrier protein (ACP) thioesterase and the evolutionary origin of plant acyl-ACP thioesterases. *Plant Cell*. 1995;7:359–371.
62. Kotaka M, Kong R, Qureshi I, et al. Structure and catalytic mechanism of the thioesterase CalE7 in enediynes biosynthesis. *J Biol Chem*. 2009;284:15739–15749.
63. Wakil SJ. Fatty acid synthase, a proficient multifunctional enzyme. *Biochemistry*. 1989;28:4523–4530.
64. Gokhale RS, Hunziker D, Cane DE, Khosla C. Mechanism and specificity of the terminal thioesterase domain from the erythromycin polyketide synthase. *Chem Biol*. 1999;6:117–125.
65. Kohli RM, Takagi J, Walsh CT. The thioesterase domain from a nonribosomal peptide synthetase as a cyclization catalyst for integrin binding peptides. *Proc Natl Acad Sci*. 2002;99:1247–1252.
66. Nguyen PC, Nguyen VS, Martin BP, et al. Biochemical and structural characterization of TesA, a major thioesterase required for outer-envelope lipid biosynthesis in *Mycobacterium tuberculosis*. *J Mol Biol*. 2018;430:5120–5136.
67. Claxton HB, Akey DL, Silver MK, Admiraal SJ, Smith JL. Structure and functional analysis of RifR, the type II thioesterase from the rifamycin biosynthetic pathway. *J Biol Chem*. 2009;284:5021–5029.
68. Libertini LJ, Smith S. Purification and properties of a thioesterase from lactating rat mammary gland which modifies the product specificity of fatty acid synthetase. *J Biol Chem*. 1978;253:1393–1401.
69. Mikkelsen J, Witkowski A, Smith S. Interaction of rat mammary gland thioesterase II with fatty acid synthetase is dependent on the presence of acyl chains on the synthetase. *J Biol Chem*. 1987;262:1570–1574.
70. Heathcote ML, Staunton J, Leadlay PF. Role of type II thioesterases: Evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units. *Chem Biol*. 2001;8:207–220.
71. Ferri SR, Meighen EA. A lux-specific myristoyl transferase in luminescent bacteria related to eukaryotic serine esterases. *J Biol Chem*. 1991;266:12852–12857.
72. Camp LA, Hofmann SL. Purification and properties of a palmitoyl-protein thioesterase that cleaves palmitate from H-Ras. *J Biol Chem*. 1993;268:22566–22574.
73. Vesa J, Hellsten E, Verkruyse LA, et al. Mutations in the palmitoyl protein thioesterase gene causing infantile neuronal ceroid lipofuscinosis. *Nature*. 1995;376:584–587.
74. Das AK, Bellizzi JJ, Tandel S, Biehl E, Clardy J, Hofmann SL. Structural basis for the insensitivity of a serine enzyme (palmitoyl-protein thioesterase) to phenylmethylsulfonyl fluoride. *J Biol Chem*. 2000;275:23847–23851.
75. Sugimoto H, Hayashi H, Yamashita S. Purification, cDNA cloning, and regulation of lysophospholipase from rat liver. *J Biol Chem*. 1996;271:7705–7711.
76. Duncan JA, Gilman AG. A cytoplasmic acyl-protein thioesterase that removes palmitate from G protein α subunits and p21. *J Biol Chem*. 1998;273:15830–15837.
77. Gonzalez CF, Proudfoot M, Brown G, et al. Molecular basis of formaldehyde detoxification: Characterization of two S-formylglutathione hydrolases from *Escherichia coli*, FrmB and YeiG. *J Biol Chem*. 2006;281:14514–14522.
78. Vander Jagt DL. Glyoxalase II: Molecular characteristics, kinetics and mechanism. *Biochem Soc Trans*. 1993;21:522–527.
79. Carfi A, Pares S, Duée E, et al. The 3-D structure of a zinc metallo-beta-lactamase from *Bacillus cereus* reveals a new type of protein fold. *EMBO J*. 1995;14:4914–4921.
80. Wang F, Langley R, Gulten G, Wang L, Sacchettini JC. Identification of a type III thioesterase reveals the function of an operon crucial for Mtb virulence. *Chem Biol*. 2007;14:543–551.
81. Gurvitz A, Hiltunen JK, Kastaniotis AJ. Heterologous expression of mycobacterial proteins in *Saccharomyces cerevisiae* reveals two physiologically functional 3-hydroxyacyl-thioester dehydratases, HtdX and HtdY, in addition to HadABC and HtdZ. *J Bacteriol*. 2009;191:2683–2690.
82. Chisuga T, Miyanaga A, Kudo F, Eguchi T. Structural analysis of the dual-function thioesterase SAV606 unravels the mechanism of Michael addition of glycine to an α,β -unsaturated thioester. *J Biol Chem*. 2017;292:10926–10937.
83. Huang F, Haydock SF, Spitteller D, et al. The gene cluster for fluorometabolite biosynthesis in *Streptomyces cattleya*: A thioesterase confers resistance to fluoroacetyl-coenzyme A. *Chem Biol*. 2006;13:475–484.
84. Weeks AM, Coyle SM, Jinek M, Doudna JA, Chang MCY. Structural and biochemical studies of a fluoroacetyl-CoA-specific thioesterase reveal a molecular basis for fluorine selectivity. *Biochemistry*. 2010;49:9269–9279.
85. Park SY, Lee SH, Lee J, et al. High-resolution structure of ybfF from *Escherichia coli* K12: A unique substrate-binding crevice generated by domain arrangement. *J Mol Biol*. 2008;376:1426–1437.
86. Kruis AJ, Levisson M, Mars AE, et al. Ethyl acetate production by the elusive alcohol acetyltransferase from yeast. *Metab Eng*. 2017;41:92–101.
87. Cao Y, Qiu T, Kathayat RS, et al. ABHD10 is an S-depalmitoylase affecting redox homeostasis through peroxiredoxin-5. *Nat Chem Biol*. 2019;15:1232–1240.
88. Regueira TB, Kildegaard KR, Hansen BG, Mortensen UH, Hertweck C, Nielsen J. Molecular basis for mycophenolic acid biosynthesis in *Penicillium brevicompactum*. *Appl Environ Microbiol*. 2011;77:3035–3043.
89. Zhang W, Cao S, Qiu L, et al. Functional characterization of MpaG^I, the O-methyltransferase involved in the biosynthesis of mycophenolic acid. *Chembiochem*. 2015;16:565–569.
90. Lin DTS, Conibear E. ABHD17 proteins are novel protein depalmitoylases that regulate N-Ras palmitate turnover and subcellular localization. *Elife*. 2015;4:1–14.

91. Storm PA, Townsend CA. In trans hydrolysis of carrier protein-bound acyl intermediates by CitA during citrinin biosynthesis. *Chem Commun.* 2017;54:50–53.
92. Zhuravleva E, Gut H, Hynx D, et al. Acyl coenzyme A thioesterase Them5/Acot15 is involved in cardiolipin remodeling and fatty liver development. *Mol Cell Biol.* 2012;32:2685–2697.
93. Drees SL, Fetzner S. PqsE of *Pseudomonas aeruginosa* acts as pathway-specific thioesterase in the biosynthesis of alkylquinolone signaling molecules. *Chem Biol.* 2015;22:611–618.
94. Yu S, Jensen V, Seeliger J, et al. Structure elucidation and preliminary assessment of hydrolase activity of PqsE, the *Pseudomonas* quinolone signal (PQS) response protein. *Biochemistry.* 2009;48:10298–10307.
95. Minetoki T, Bogaki T, Iwamatsu A, Fujii T, Hamachi M. The purification, properties and internal peptide sequences of alcohol acetyltransferase isolated from *Saccharomyces cerevisiae* Kyokai No. 7. *Biosci Biotechnol Biochem.* 1993;57:2094–2098.
96. Nancolas B, Bull ID, Stenner R, Dufour V, Curnow P. *Saccharomyces cerevisiae* Atf1p is an alcohol acetyltransferase and a thioesterase in vitro. *Yeast.* 2017;34:239–251.
97. Shen H, Campanello GC, Flicker D, et al. The human knockout gene CLYBL connects itaconate to vitamin B12. *Cell.* 2017;171:771–782.e11.
98. Pidugu LS, Maity K, Ramaswamy K, Suroliya N, Suguna K. Analysis of proteins with the “hot dog” fold: Prediction of function and identification of catalytic residues of hypothetical proteins. *BMC Struct Biol.* 2009;9:37.
99. Wang H, Fedorov AA, Fedorov EV, et al. An essential bifunctional enzyme in *Mycobacterium tuberculosis* for itaconate dissimilation and leucine catabolism. *Proc Natl Acad Sci U S A.* 2019;116:15907–15913.
100. Torres R, Chim N, Sankaran B, Pujol C, Bliska JB, Goulding CW. Structural insights into RipC, a putative citrate lyase beta subunit from a *Yersinia pestis* virulence operon. *Acta Crystallogr Sect F.* 2012;68:2–7.
101. Engel LA, Jing Z, O'Brien DE, Sun M, Kotzbauer PT. Catalytic function of PLA2G6 is impaired by mutations associated with infantile neuroaxonal dystrophy but not dystonia-parkinsonism. *PLoS One.* 2010;5:e12897.
102. Jenkins CM, Yan W, Mancuso DJ, Gross RW. Highly selective hydrolysis of fatty acyl-CoAs by calcium-independent phospholipase A2 β : Enzyme autoacylation and acyl-CoA-mediated reversal of calmodulin inhibition of phospholipase A2 activity. *J Biol Chem.* 2006;281:15615–15624.
103. Zarzycki J, Kerfeld CA. The crystal structures of the tri-functional *Chloroflexus aurantiacus* and bi-functional *Rhodobacter sphaeroides* malyl-CoA lyases and comparison with CitE-like superfamily enzymes and malate synthases. *BMC Struct Biol.* 2013;13:28.
104. Kiema TR, Harijan RK, Strozyk M, Fukao T, Alexson SEH, Wierenga RK. The crystal structure of human mitochondrial 3-ketoacyl-CoA thiolase (T1): Insight into the reaction mechanism of its thiolase and thioesterase activities. *Acta Crystallogr Sect D Biol Crystallogr.* 2014;70:3212–3225.
105. Rose IA, Warms JVB. An enzyme with ubiquitin carboxy-terminal esterase activity from reticulocytes. *Biochemistry.* 1983;22:4234–4237.
106. Rawlings ND, Barrett AJ, Bateman A. MEROPS: The peptidase database. *Nucleic Acids Res.* 2010;38:D227–D233.
107. Mullins EA, Kappock TJ. Crystal structures of *Acetobacter acetii* succinyl-coenzyme A (CoA):Acetate CoA-transferase reveal specificity determinants and illustrate the mechanism used by class I CoA-transferases. *Biochemistry.* 2012;51:8422–8434.
108. Murphy JR, Mullins EA, Kappock TJ. Functional dissection of the bipartite active site of the class I coenzyme A (CoA)-transferase succinyl-CoA: Acetate CoA-transferase. *Front Chem.* 2016;4:23.
109. Mandel CR, Tweel B, Tong L. Crystal structure of human mitochondrial acyl-CoA thioesterase (ACOT2). *Biochem Biophys Res Commun.* 2009;385:630–633.
110. Lo YC, Lin SC, Shaw JF, Liaw YC. Crystal structure of *Escherichia coli* thioesterase I/protease I/lysophospholipase L1: Consensus sequence blocks constitute the catalytic center of SGNH-hydrolases through a conserved hydrogen bond network. *J Mol Biol.* 2003;330:539–551.
111. Brzuszkiewicz A, Nowak E, Dauter Z, et al. Structure of esterase from psychrotrophic *Pseudoalteromonas* sp. 643A covalently inhibited by monoethylphosphonate. *Acta Crystallogr Sect F Struct Biol Cryst Commun.* 2009;65:862–865.
112. Kovačić F, Granzin J, Wilhelm S, Kojić-Prodić B, Batra-Safferling R, Jaeger KE. Structural and functional characterization of TesA—A novel Lysophospholipase A from *Pseudomonas aeruginosa*. *PLoS One.* 2013;8:e69125.
113. Grisewood MJ, Hernández-Lozada NJ, Thoden JB, et al. Computational redesign of acyl-ACP thioesterase with improved selectivity toward medium-chain-length fatty acids. *ACS Catal.* 2017;7:3837–3849.
114. Li Z, Li L, Huo Y, et al. Structure-guided protein engineering increases enzymatic activities of the SGNH family esterases. *Biotechnol Biofuels.* 2020;13:13.
115. Li J, Derewenda U, Dauter Z, Smith S, Derewenda ZS. Crystal structure of the *Escherichia coli* thioesterase II, a homolog of the human Nef binding enzyme. *Nat Struct Biol.* 2000;7:555–559.
116. Forwood JK, Thakur AS, Guncar G, et al. Structural basis for recruitment of tandem hotdog domains in acyl-CoA thioesterase 7 and its role in inflammation. *Proc Natl Acad Sci U S A.* 2007;104:10382–10387.
117. Willis MA, Zhuang Z, Song F, Howard A, Dunaway-Mariano D, Herzberg O. Structure of YciA from *Haemophilus influenzae* (HI0827), a hexameric broad specificity acyl-coenzyme A thioesterase. *Biochemistry.* 2008;47:2797–2805.
118. Khandokar Y, Srivastava P, Raidal S, Sarker S, Forwood JK. Structural basis for disulphide-CoA inhibition of a butyryl-CoA hexameric thioesterase. *J Struct Biol.* 2020;210:107477.
119. Park J, Kim YJ, Lee D, Kim KJ. Structural basis for nucleotide-independent regulation of acyl-CoA thioesterase from *Bacillus cereus* ATCC 14579. *Int J Biol Macromol.* 2021;170:390–396.
120. Cao J, Xu H, Zhao H, Gong W, Dunaway-Mariano D. The mechanisms of human hotdog-fold thioesterase 2 (hTHEM2) substrate recognition and catalysis illuminated by a structure and function based analysis. *Biochemistry.* 2009;48:1293–1304.

121. Cheng Z, Song F, Shan X, et al. Crystal structure of human thioesterase superfamily member 2. *Biochem Biophys Res Commun.* 2006;349:172–177.
122. Cantu DC, Ardèvol A, Rovira C, Reilly PJ. Molecular mechanism of a hotdog-fold acyl-CoA thioesterase. *Chem - Eur J.* 2014;20:9045–9051.
123. Thoden JB, Holden HM, Zhuang Z, Dunaway-Mariano D. X-ray crystallographic analyses of inhibitor and substrate complexes of wild-type and mutant 4-hydroxybenzoyl-CoA thioesterase. *J Biol Chem.* 2002;277:27468–27476.
124. Thoden JB, Zhuang Z, Dunaway-Mariano D, Holden HM. The structure of 4-hydroxybenzoyl-CoA thioesterase from *Arthrobacter* sp. strain SU. *J Biol Chem.* 2003;278:43709–43716.
125. Wu R, Latham JA, Chen D, et al. Structure and catalysis in the *Escherichia coli* hotdog-fold thioesterase paralogs YdiI and YbdB. *Biochemistry.* 2014;53:4788–4805.
126. Kunishima N, Asada Y, Sugahara M, et al. A novel induced-fit reaction mechanism of asymmetric hot dog thioesterase PaaI. *J Mol Biol.* 2005;352:212–228.
127. Song F, Zhuang Z, Finci L, et al. Structure, function, and mechanism of the phenylacetate pathway hot dog-fold thioesterase PaaI. *J Biol Chem.* 2006;281:11028–11038.
128. Feng Y, Wang Y, Liu J, Liu Y, Cao X, Xue S. Structural insight into acyl-ACP thioesterase toward substrate specificity design. *ACS Chem Biol.* 2017;12:2830–2836.
129. Chakravarty B, Gu Z, Chirala SS, Wakil SJ, Quijcho FA. Human fatty acid synthase: Structure and substrate selectivity of the thioesterase domain. *Proc Natl Acad Sci U S A.* 2004;101:15567–15572.
130. Bruner SD, Weber T, Kohli RM, et al. Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. *Structure.* 2002;10:301–310.
131. Samel SA, Wagner B, Marahiel MA, Essen LO. The thioesterase domain of the fengycin biosynthesis cluster: A structural base for the macrocyclization of a non-ribosomal lipopeptide. *J Mol Biol.* 2006;359:876–889.
132. Korman TP, Crawford JM, Labonte JW, et al. Structure and function of an iterative polyketide synthase thioesterase domain catalyzing Claisen cyclization in aflatoxin biosynthesis. *Proc Natl Acad Sci U S A.* 2010;107:6246–6251.
133. Aggarwal A, Parai MK, Shetty N, et al. Development of a novel lead that targets *M. tuberculosis* polyketide synthase 13. *Cell.* 2017;170:249–259.e25.
134. Patel KD, d'Andrea FB, Gaudelli NM, Buller AR, Townsend CA, Gulick AM. Structure of a bound peptide phosphonate reveals the mechanism of nocardicin bifunctional thioesterase epimerase-hydrolase half-reactions. *Nat Commun.* 2019;10:3868.
135. Tsai SC, Miercke LJW, Krucinski J, et al. Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: Versatility from a unique substrate channel. *Proc Natl Acad Sci U S A.* 2001;98:14808–14813.
136. Tsai S-C, Lu H, Cane DE, Khosla C, Stroud RM. Insights into channel architecture and substrate specificity from crystal structures of two macrocycle-forming thioesterases of modular polyketide synthases. *Biochemistry.* 2002;41:12598–12606.
137. Scaglione JB, Akey DL, Sullivan R, et al. Biochemical and structural characterization of the tautomycin thioesterase: Analysis of a stereoselective polyketide hydrolase. *Angew Chem - Int Ed.* 2010;49:5726–5730.
138. Koglin A, Löhr F, Bernhard F, et al. Structural basis for the selectivity of the external thioesterase of the surfactin synthetase. *Nature.* 2008;454:907–911.
139. Whicher JR, Florova G, Sydor PK, et al. Structure and function of the RedJ protein, a thioesterase from the prodiginine biosynthetic pathway in *Streptomyces coelicolor*. *J Biol Chem.* 2011;286:22558–22569.
140. Ritchie MK, Johnson LC, Clodfelter JE, et al. Crystal structure and substrate specificity of human thioesterase 2: Insights into the molecular basis for the modulation of fatty acid synthase. *J Biol Chem.* 2016;291:3520–3530.
141. Sandhya Guntaka N, Healy AR, Crawford JM, Herzon SB, Bruner SD. Structure and functional analysis of ClbQ, an unusual intermediate-releasing thioesterase from the colibactin biosynthetic pathway HHS public access. *ACS Chem Biol.* 2017;12:2598–2608.
142. Ohlemacher SI, Xu Y, Kober DL, et al. YbtT is a low-specificity type II thioesterase that maintains production of the metallophore yersiniabactin in pathogenic enterobacteria. *J Biol Chem.* 2018;293:19572–19585.
143. Lawson DM, Derewenda U, Serre L, et al. Structure of a myristoyl-ACP-specific thioesterase from *Vibrio harveyi*. *Biochemistry.* 1994;33:9382–9388.
144. Bellizzi JJ, Widom J, Kemp C, et al. The crystal structure of palmitoyl protein thioesterase 1 and the molecular basis of infantile neuronal ceroid lipofuscinosis. *Proc Natl Acad Sci.* 2000;97:4573–4578.
145. Kim KK, Song HK, Shin DH, et al. Crystal structure of carboxylesterase from *Pseudomonas fluorescens*, an α/β hydrolase with broad substrate specificity. *Structure.* 1997;5:1571–1584.
146. Devedjiev Y, Dauter Z, Kuznetsov SR, Jones TLZ, Derewenda ZS. Crystal structure of the human acyl protein thioesterase I from a single X-ray data set to 1.5 Å. *Structure.* 2000;8:1137–1146.
147. Pesaresi A, Lamba D. Insights into the fatty acid chain length specificity of the carboxylesterase PA3859 from *Pseudomonas aeruginosa*: A combined structural, biochemical and computational study. *Enferm Infecc Microbiol Clin.* 2010;28:1787–1792.
148. Filippova EV, Weston LA, Kuhn ML, et al. Large scale structural rearrangement of a serine hydrolase from *Francisella tularensis* facilitates catalysis. *J Biol Chem.* 2013;288:10522–10535.
149. Won SJ, Davda D, Labby KJ, et al. Molecular mechanism for isoform-selective inhibition of acyl protein thioesterases 1 and 2 (APT1 and APT2). *ACS Chem Biol.* 2016;11:3374–3382.
150. Wepy JA, Galligan JJ, Kingsley PJ, et al. Lysophospholipases cooperate to mediate lipid homeostasis and lysophospholipid signaling. *J Lipid Res.* 2019;60:360–374.
151. Legler PM, Kumaran D, Swaminathan S, Studier FW, Millard CB. Structural characterization and reversal of the natural organophosphate resistance of a D-type esterase, *Saccharomyces cerevisiae* S-formylglutathione hydrolase. *Biochemistry.* 2008;47:9592–9601.

152. Van Straaten KE, Gonzalez CF, Valladares RB, Xu X, Savchenko AV, Sanders DAR. The structure of a putative S-formylglutathione hydrolase from *Agrobacterium tumefaciens*. *Protein Sci.* 2009;18:2196–2202.
153. Wu D, Li Y, Song G, Zhang D, Shaw N, Liu Z. Crystal structure of human esterase D: A potential genetic marker of retinoblastoma. *FASEB J.* 2009;23:1441–1446.
154. Lemak S, Tchigvintsev A, Petit P, et al. Structure and activity of the cold-active and anion-activated carboxyl esterase OLEI01171 from the oil-degrading marine bacterium *Oleispira antarctica*. *Biochem J.* 2012;445:193–203.
155. Alterio V, Aurilia V, Romanelli A, et al. Crystal structure of an S-formylglutathione hydrolase from *Pseudoalteromonas haloplanktis* TAC125. *Biopolymers.* 2010;93:669–677.
156. Chen NH, Couñago RM, Djoko KY, et al. A glutathione-dependent detoxification system is required for formaldehyde resistance and optimal survival of *Neisseria meningitidis* in biofilms. *Antioxid Redox Signal.* 2013;18:743–755.
157. Lee CW, Yoo W, Park SH, et al. Structural and functional characterization of a novel cold-active S-formylglutathione hydrolase (SfSFGH) homolog from *Shewanella frigidimarina*, a psychrophilic bacterium. *Microb Cell Fact.* 2019;18:140.
158. Dias MVB, Huang F, Chirgadze DY, et al. Structural basis for the activity and substrate specificity of fluoroacetyl-CoA thioesterase FIK. *J Biol Chem.* 2010;285:22495–22504.
159. Malley KR, Koroleva O, Miller I, et al. The structure of iPLA2 β reveals dimeric active sites and suggests mechanisms of regulation and localization. *Nat Commun.* 2018;9:1–11.
160. Leesong M, Henderson BS, Gillig JR, Schwab JM, Smith JL. Structure of a dehydratase–isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: Two catalytic activities in one active site. *Structure.* 1996;4:253–264.
161. Dillon SC, Bateman A. The hotdog fold: Wrapping up a superfamily of thioesterases and dehydratases. *BMC Bioinform.* 2004;5:109.
162. Mayer KM, Shanklin J. Identification of amino acid residues involved in substrate specificity of plant acyl-ACP thioesterases using a bioinformatics-guided approach. *BMC Plant Biol.* 2007;7:1.
163. Ollis DL, Cheah E, Cygler M, et al. The α/β hydrolase fold. *Protein Eng Des Sel.* 1992;5:197–211.
164. Lenfant N, Hotelier T, Velluet E, Bourne Y, Marchot P, Chatonnet A. ESTHER, the database of the α/β -hydrolase fold superfamily of proteins: Tools to explore diversity of functions. *Nucleic Acids Res.* 2013;41:D423–D429.
165. Nardini M, Dijkstra BW. α/β Hydrolase fold enzymes: The family keeps growing. *Curr Opin Struct Biol.* 1999;9:732–737.
166. Chen X-P, Shi T, Wang X-L, et al. Theoretical studies on the mechanism of thioesterase-catalyzed macrocyclization in erythromycin biosynthesis. *ACS Catal.* 2016;6:4369–4378.
167. Adrover-Castellano ML, Schmidt JJ, Sherman DH. Biosynthetic cyclization catalysts for the assembly of peptide and polyketide natural products. *ChemCatChem.* 2021;13:2095–2116.
168. Pemble CW, Johnson LC, Kridel SJ, Lowther WT. Crystal structure of the thioesterase domain of human fatty acid synthase inhibited by Orlistat. *Nat Struct Mol Biol.* 2007;14:704–709.
169. Zhang W, Chakravarty B, Zheng F, et al. Crystal structure of FAS thioesterase domain with polyunsaturated fatty acyl adduct and inhibition by dihomog- γ -linolenic acid. *Proc Natl Acad Sci.* 2011;108:15757–15762.
170. Park I-H, Venable JD, Steckler C, et al. Estimation of hydrogen-exchange protection factors from MD simulation based on amide hydrogen bonding analysis. *J Chem Inf Model.* 2015;55:1914–1925.
171. Drake EJ, Miller BR, Shi C, et al. Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature.* 2016;529:235–238.
172. Li X, Sevillano N, La Greca F, et al. Discovery and characterization of a thioesterase-specific monoclonal antibody that recognizes the 6-deoxyerythronolide B synthase. *Biochemistry.* 2018;57:6201–6208.
173. Argyropoulos P, Bergeret F, Pardin C, et al. Towards a characterization of the structural determinants of specificity in the macrocyclizing thioesterase for deoxyerythronolide B biosynthesis. *Biochim Biophys Acta - Gen Subj.* 2016;1860:486–497.
174. Giraldes JW, Akey DL, Kittendorf JD, Sherman DH, Smith JL, Fecik RA. Structural and mechanistic insights into polyketide macrolactonization from polyketide-based affinity labels. *Nat Chem Biol.* 2006;2:531–536.
175. Akey DL, Kittendorf JD, Giraldes JW, Fecik RA, Sherman DH, Smith JL. Structural basis for macrolactonization by the pikromycin thioesterase. *Nat Chem Biol.* 2006;2:537–542.
176. Linne U, Schwarzer D, Schroeder GN, Marahiel MA. Mutational analysis of a type II thioesterase associated with non-ribosomal peptide synthesis. *Eur J Biochem.* 2004;271:1536–1545.
177. Koglin A, Löhr F, Bernhard F, et al. Structural basis for the selectivity of the external thioesterase of the surfactin synthetase. *Nature.* 2008;454:907–911.
178. Cameron AD, Ridderström M, Olin B, Mannervik B. Crystal structure of human glyoxalase II and its complex with a glutathione thioester substrate analogue. *Structure.* 1999;7:1067–1078.
179. Marasinghe GPK, Sander IM, Bennett B, et al. Structural studies on a mitochondrial glyoxalase II. *J Biol Chem.* 2005;280:40668–40675.
180. Campos-Bermudez VA, Leite NR, Krog R, et al. Biochemical and structural characterization of *Salmonella typhimurium* glyoxalase II: New insights into metal ion selectivity. *Biochemistry.* 2007;46:11069–11079.
181. Sattler SA, Wang X, Lewis KM, et al. Characterizations of two bacterial persulfide dioxygenases of the metallo- β -lactamase superfamily. *J Biol Chem.* 2015;290:18914–18923.
182. Oke M, Carter LG, Johnson KA, et al. The scottish structural proteomics facility: Targets, methods and outputs. *J Struct Funct Genomics.* 2010;11:167–180.
183. Zender M, Witzgall F, Drees SL, et al. Dissecting the multiple roles of PqsE in *Pseudomonas aeruginosa* virulence by discovery of small tool compounds. *ACS Chem Biol.* 2016;11:1755–1763.
184. Morales-Quintana L, Nuñez-Tobar MX, Moya-León MA, Herrera R. Molecular dynamics simulation and site-directed mutagenesis of alcohol acyltransferase: A proposed mechanism of catalysis. *J Chem Inf Model.* 2013;53:2689–2700.

185. Navarro-Retamal C, Gaete-Eastman C, Herrera R, Caballero J, Alzate-Morales JH. Structural and affinity determinants in the interaction between alcohol acyltransferase from *F. x ananassa* and several alcohol substrates: A computational study. *PLoS One*. 2016;11:1–14.
186. Galaz S, Morales-Quintana L, Moya-León MA, Herrera R. Structural analysis of the alcohol acyltransferase protein family from *Cucumis melo* shows that enzyme activity depends on an essential solvent channel. *FEBS J*. 2013;280:1344–1357.
187. Kleanthous C, Shaw WV. Analysis of the mechanism of chloramphenicol acetyltransferase by steady-state kinetics. Evidence for a ternary-complex mechanism. *Biochem J*. 1984;223:211–220.
188. Bracken CD, Neighbor AM, Lamlenn KK, et al. Crystal structures of a halophilic archaeal malate synthase from *Haloferax volcanii* and comparisons with isoforms A and G. *BMC Struct Biol*. 2011;11:23.
189. Strittmatter L, Li Y, Nakatsuka NJ, Calvo SE, Grabarek Z, Mootha VK. CLYBL is a polymorphic human enzyme with malate synthase and β -methylmalate synthase activity. *Hum Mol Genet*. 2014;23:2313–2323.
190. Goulding CW, Bowers PM, Segelke B, et al. The structure and computational analysis of *Mycobacterium tuberculosis* protein CitE suggest a novel enzymatic function. *J Mol Biol*. 2007;365:275–283. Available from: .
191. Tang J, Kriz RW, Wolfman N, Shaffer M, Seehra J, Jones SS. A novel cytosolic calcium-independent phospholipase A2 contains eight ankyrin motifs. *J Biol Chem*. 1997;272:8567–8575.
192. The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*. 2019;36:D506–D515.
193. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res*. 2009;37:D26–D31.
194. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000;28:235–242.
195. Altschup SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–410.
196. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res*. 1997;25:3389–3402.
197. El-Gebali S, Mistry J, Bateman A, et al. The Pfam protein families database in 2019. *Nucleic Acids Res*. 2019;47:D427–D432.
198. Shatsky M, Nussinov R, Wolfson HJ. A method for simultaneous alignment of multiple protein structures. *Proteins Struct Funct Genet*. 2004;56:143–156.
199. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32:1792–1797.
200. Stecher G, Tamura K, Kumar S. Molecular evolutionary genetics analysis (MEGA) for macOS. *Mol Biol Evol*. 2020;37:1237–1239.

How to cite this article: Caswell BT, de Carvalho CC, Nguyen H, Roy M, Nguyen T, Cantu DC. Thioesterase enzyme families: Functions, structures, and mechanisms. *Protein Science*. 2022;1–25. <https://doi.org/10.1002/pro.4263>