

## RESEARCH ARTICLE OPEN ACCESS

# $\alpha/\beta$ Hydrolases: Toward Unraveling Entangled Classification

Fatih Ozhelvaci | Kamil Steczkiewicz 

Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warszawa, Poland

**Correspondence:** Kamil Steczkiewicz ([ksteczk@ibb.waw.pl](mailto:ksteczk@ibb.waw.pl))**Received:** 7 June 2024 | **Revised:** 14 November 2024 | **Accepted:** 19 November 2024**Funding:** This project was supported by Narodowe Centrum Nauki NCN grant 2019/35/D/NZ2/03411 to K.S.**Keywords:**  $\alpha/\beta$  hydrolases | enzyme structure | homology modeling | MEROPS | peptidase | protease | protein family | S09 | S33 | serine protease**ABSTRACT**

$\alpha/\beta$  Hydrolase-like enzymes form a large and functionally diverse superfamily of proteins. Despite retaining a conserved structural core consisting of an eight-stranded, central  $\beta$ -sheet flanked with six  $\alpha$ -helices, they display a modular architecture allowing them to perform a variety of functions, like esterases, lipases, peptidases, epoxidases, lyases, and others. At the same time, many  $\alpha/\beta$  hydrolase-like families, even enzymatically distinct, share a high degree of sequence similarity. This imposes several problems for their annotation and classification, because available definitions of particular  $\alpha/\beta$  hydrolase-like families overlap significantly, so the unambiguous functional assignment of these superfamily members remains a challenging task. For instance, two large and important peptidase families, namely S9 and S33, blend with lipases, epoxidases, esterases, and other enzymes unrelated to proteolysis, which hinders automatic annotations in high-throughput projects. With the use of thorough sequence and structure analyses, we newly annotate three protein families as  $\alpha/\beta$  hydrolase-like and revise current classifications of the realm of  $\alpha/\beta$  hydrolase-like superfamily. Based on manually curated structural superimpositions and multiple sequence and structure alignments, we comprehensively demonstrate structural conservation and diversity across the whole superfamily. Eventually, after detailed pairwise sequence similarity assessments, we develop a new clustering of the  $\alpha/\beta$  hydrolases and provide a set of family profiles allowing for detailed, reliable, and automatic functional annotations of the superfamily members.

**1 | Introduction**

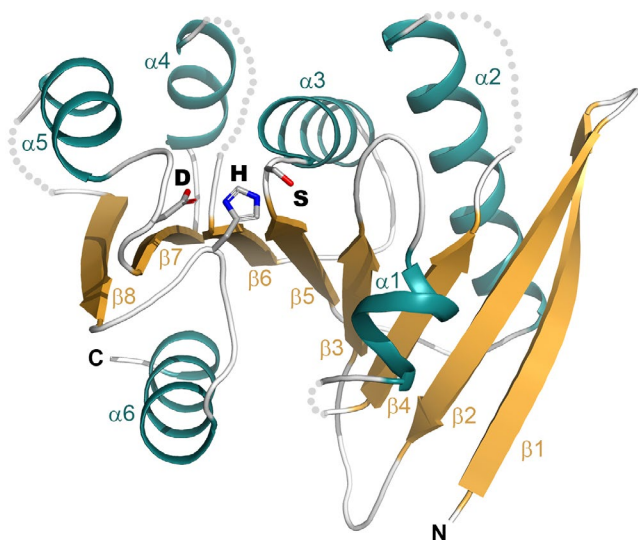
Proteins classified as  $\alpha/\beta$  hydrolases are found in a wide diversity of organisms representing all kingdoms of life, including viruses. They function as lipases, deacetylases, proteases, lactonases, epoxide hydrolases, peroxidases, lactamases, reductases, dehalogenases, lyases, or esterases of many substrates [1, 2]. Almost none of these functions is incidental within this superfamily and each appears within multiple separate protein families. We know many protein superfamilies displaying extreme sequence and structure diversity, for example, RNaseH-like [3] or papain-like [4], but they usually have some major, coherent functional theme like nuclease or peptidase.  $\alpha/\beta$  hydrolases, even at their very definition by Ollis and collaborators

in 1992 were being described as highly flexible regarding their catalytic potency [5]. The first structures commonly defining  $\alpha/\beta$  hydrolases were lipase, dehalogenase, lactone hydrolase, peptidase, and acetylcholine esterase [5]. All of them retain a common structural core of eight-stranded, twisted  $\beta$ -sheet in 12435678 topology, flanked by six helices on both sides, as well as the spatially conserved catalytic triad of the nucleophile, acid, and histidine residues [5–8] (Figure 1).

Although the structural core may differ in  $\beta$ -sheet's curvature or helices alignment against the  $\beta$ -sheet, it remains rather conserved and provides a firm basis for additional structural elements that ultimately shape the active site environment and substrate specificity [1, 10]. The catalytic site architecture is

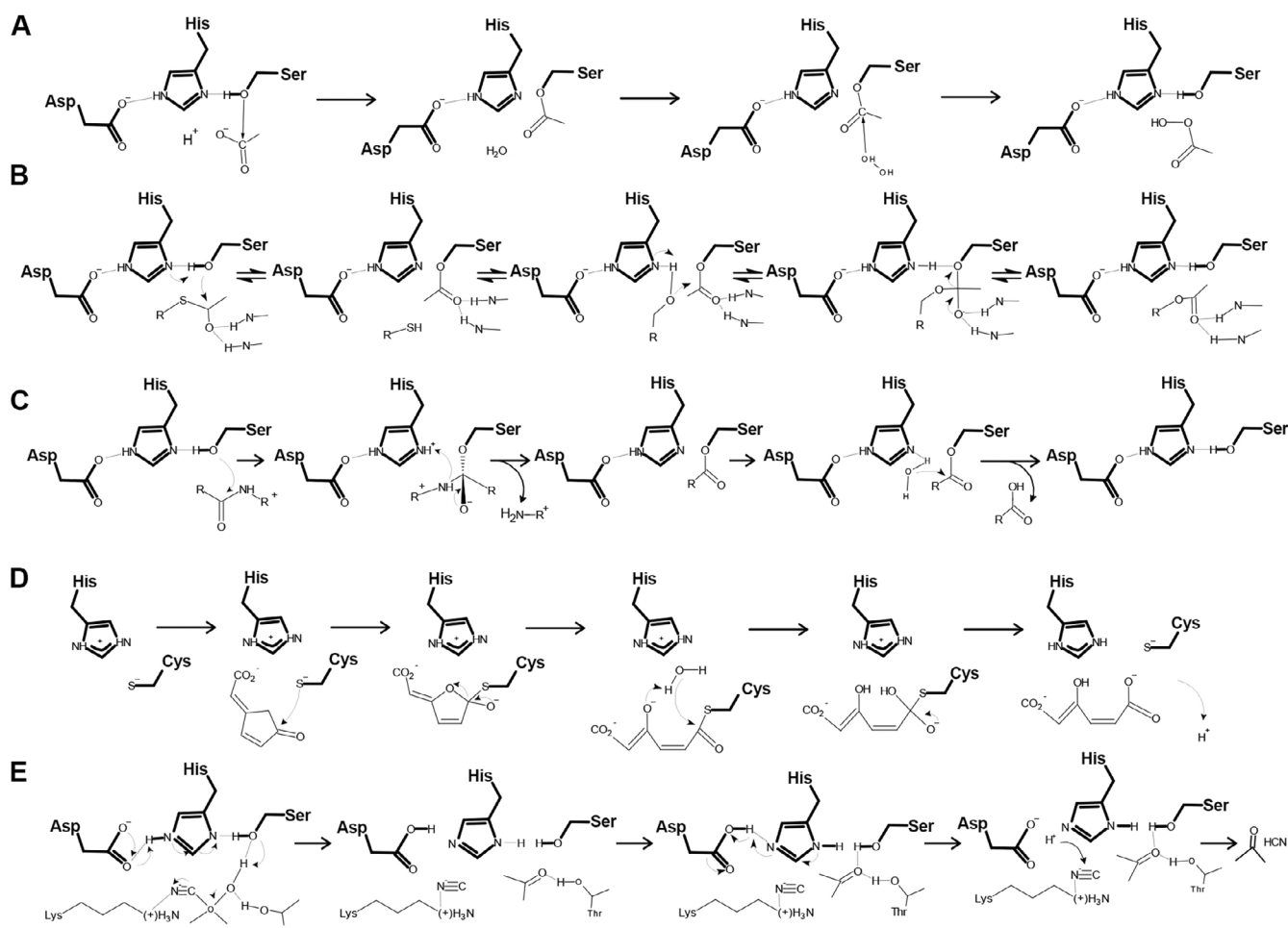
This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *PROTEINS: Structure, Function, and Bioinformatics* published by Wiley Periodicals LLC.



**FIGURE 1** | Structural core of  $\alpha/\beta$  hydrolase-like proteins on the example of wheat serine carboxypeptidase II from *Triticum aestivum* (pd-b13sc2 [9]). Only the core secondary structure elements are shown, for clarity. Catalytic triad residues are rendered with sticks.

also conserved despite sequence and structure diversity. The nucleophile, mostly serine, is located at “nucleophile elbow,” a tight turn between strand  $\beta 5$  and helix  $\alpha 3$ ; the acidic residue, usually aspartic or glutamic acid, follows strand  $\beta 7$ ; and totally conserved histidine localizes to a variable loop after strand  $\beta 8$ . “Nucleophile elbow” also provides one of the oxyanion-binding sites formed by the residue immediately following the catalytic nucleophile: either as the backbone nitrogen or residue sidechain [11], while the rest of the oxyanion hole comes from residues located usually after strand  $\beta 3$  [10]. As initially pointed out by Holmquist, the diversity of substrates of  $\alpha/\beta$  hydrolases may be classified down to three general categories: (i) peptide-, oxyester-, and thioester-bond, (ii) C–halogen or C–O bond, and (iii) C–C bond [2]. Based on subsequent fruitful 15 years of structural studies this classification was detailed by Rauwerdink and Kazlauskas into 17 different reaction mechanisms catalyzed by Ser-His-Asp triad alone [11]. Eventually,  $\alpha/\beta$  hydrolases are classified into multiple enzymatic classes, differing even by the first EC number (Figure 2). Despite the function, the cleavage requires a nucleophilic attack by activated serine/cysteine (or even aspartate for C–halogen or C–O substrate) to form an enzyme–substrate intermediate, which is then cleaved by water [2, 12]. Only lyases catalyzing C–C bond breakage, despite



**FIGURE 2** |  $\alpha/\beta$  hydrolase-like enzymes are classified into multiple EC categories. (A) Chloroperoxidase, EC 1.11.1.10, after Hofmann et al. [14], (B) acetyltransferase, EC 2.3.1.175, adopted from Lejon et al. [15], (C) peptidase, EC 3.4.14.5, based on Roppongi et al. [16], (D) dienelactone hydrolase, EC 3.1.8.1, based on Cheah et al. [17], and (E) hydroxynitrile lyase, EC 4.2.1.37, adopted from Sharma et al. [18].

having identical catalytic triad display different mechanisms of action: they skip acyl–enzyme intermediate, lack oxyanion hole but use serine as a donor of hydrogen bonds instead [13], or even do not require serine at all [11]. Nonetheless, the geometry of the catalytic site remains conserved so that with few mutations it is possible to switch between esterase and lyase, at least in certain  $\alpha/\beta$  hydrolases [13].

The access to the active site's cavity as well as substrate specificity is regulated by additional structural elements inserted into the core fold, namely caps, lids, or flaps, as well as by supplementary protein domains covering the entrance [19]. For instance, lipases remain almost inactive with the catalytic site blocked by the lid when in the aqueous solution unless they reach the lipid phase which interacts with the hydrophobic surface of the lid, opens the catalytic cleft, and activates the enzyme [20, 21]. Active sites of proline exopeptidases on the other hand are covered by a seven-bladed  $\beta$ -propeller allowing specific substrates to pass through its central canal into the peptidase active site [22]. However, the majority of  $\alpha/\beta$  hydrolase enzymes bind their substrates with smaller structural insertions, loops, or secondary structure elements extending their central  $\beta$ -strands [1].

$\alpha/\beta$  Hydrolases are classified in multiple publicly available databases of differing scopes. PFAM, a general protein families database (ver. 35.0) [23], identifies 75 families within the AB\_hydrolase clan (CL0028) covering over 2 million protein sequences. ESTHER database (ESTerases and alpha/beta-Hydrolase Enzymes and Relatives) [24], provides a more detailed classification of  $\alpha/\beta$  hydrolases at the level of their molecular function.  $\alpha/\beta$  Hydrolases are represented also in the CAZy database of carbohydrate-active enzymes [25]. The Lipase Engineering Database (LED) in turn was neatly designed to describe  $\alpha/\beta$  hydrolase diversity based on their highly modular structure [19], and catalogs 200 000 proteins and almost 1600 Protein Data Bank (PDB) structures in a thoroughly curated manner. Representatives of  $\alpha/\beta$  hydrolases are also included in MEROPS, a peptidase reference database that defines a curated hierarchy of clans and families of experimentally studied proteolytic enzymes [26]. Seven MEROPS families, namely S09, S10, S15, S28, S33, S37, and S82 belong to  $\alpha/\beta$  hydrolases.

The very dense sequence space of  $\alpha/\beta$  hydrolases and the ability of highly similar enzymes to catalyze different reactions (reviewed in [19]) impose several problems regarding their detailed classification. Substrate specificity of  $\alpha/\beta$  hydrolases is shaped by structural add-ons to the protein's catalytic core, sometimes extensive caps or big additional domains, but many times just loops or lids made of one or two helices [19]. Such minute traits are likely to disappear in statistical definitions for protein families (PFAM, ESTHER) so that multiple functional families of  $\alpha/\beta$  hydrolases massively overlap. For instance, S09 and S33 families, even in MEROPS itself, despite peptidases containing thioesterases, lipases, methylesterases, carboxylesterases, epoxide hydrolases, and even hydroxynitrile lyases. In consequence, genuine peptidases are in the minority in these families so high throughput genome screening for these proteins would provide false positive results, which we experienced previously [27].

Although many previous studies already addressed the observed diversity of structures, catalytic mechanisms, and functions of  $\alpha/\beta$  hydrolases, the problem of discerning between particular families and functions still remains open. The aim of this work is to identify new  $\alpha/\beta$  hydrolase families, describe structural diversity within this superfamily, and propose supplementary family definitions that might help in reliable, automatic annotations. Using remote sequence homology detection methods we identify new PFAM families, PDB representatives (PDB90), and human proteins previously not classified as  $\alpha/\beta$  hydrolases. Using sequence similarity networks between superfamily representative sequences we define clusters of functionally similar proteins and develop HMM profiles allowing for immediate and more detailed classification of any  $\alpha/\beta$  hydrolase-like protein. Based on manually curated structural comparisons we provide structure-based multiple sequence alignment covering core structure elements for all defined clusters of proteins belonging to  $\alpha/\beta$  hydrolase superfamily. Eventually, by combining sequence- and structure-based analyses with additional data, like domain architectures and genomic contexts, we hypothesize on the potential functions of previously uncharacterized  $\alpha/\beta$  hydrolase families.

## 2 | Results

### 2.1 | Identification of New $\alpha/\beta$ Hydrolases

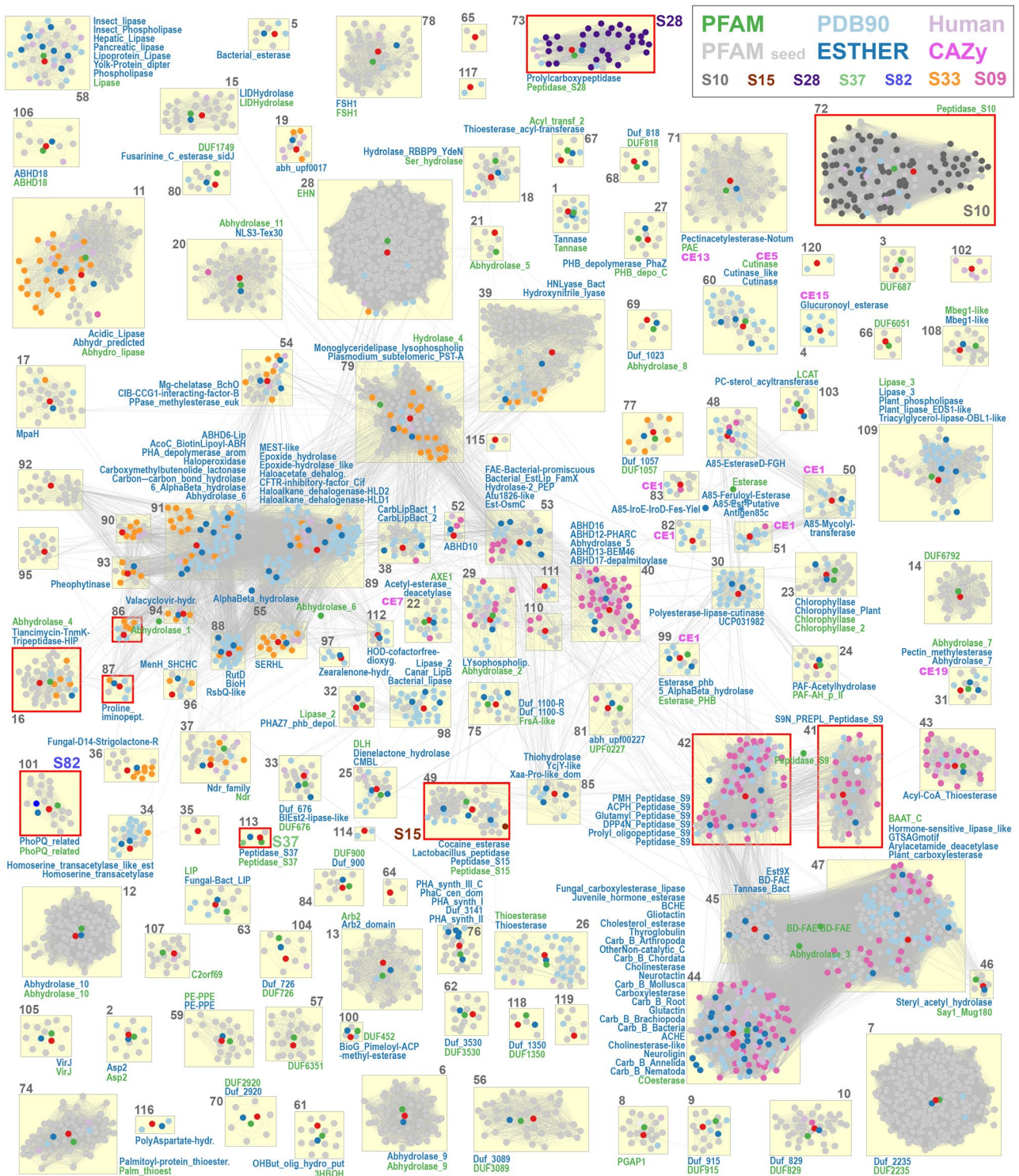
Starting from the initial 75 PFAM families (classified to PFAM AB\_hydrolase clan) and PDB90 representatives belonging to these families, with the use of sequence-based remote homology detection method HHSEARCH we identified as  $\alpha/\beta$  hydrolases additional 5 PFAM families which were not assigned to any superfamily. In the newest release of PFAM (ver.36.0) published during the preparation of this manuscript, two families: PF20591 (DUF6792) and PF10561 (C2orf69) became  $\alpha/\beta$  hydrolases clan members, yet still no functional annotation is available for them. Three PFAM families previously classified to AB\_hydrolase clan were discarded from our curated dataset: PF07167 (PhaC\_N), which covers barely three initial  $\beta$ -strands and overlaps with PF00561 (Abhydrolase\_1); PF07176 (DUF1400)—domain preceding actual  $\alpha/\beta$  hydrolase; and PF03893 (Lipase3\_N) standing for a region N-terminal to lipase domain. Eventually, the final dataset covers 77 PFAM families, 672 PDB90 structures, and 130 human proteins. Identified families span over 3.7 million proteins originating from all kingdoms of life.

### 2.2 | Clustering of $\alpha/\beta$ Hydrolases

The easiest way to categorize identified sequences within a superfamily would be to map them into already existing families. However, due to overlaps between  $\alpha/\beta$  hydrolase families, the usage of general HMM profiles would not allow for unambiguous assignments. Hence, based on pairwise sequence similarities evaluated with BLASTP scores we developed a supplementary clustering of  $\alpha/\beta$  hydrolases sequence space covering representative sequences of all identified PFAM families (seed sequences), as well as MEROPS holotypes, PDB90, and human proteins. Eventually,  $\alpha/\beta$  hydrolases

were categorized into 120 clusters displaying structural and functional, with minor exceptions discussed below, consistency (Table S1 and Figure 3). One hundred and nine clusters contain at least one PFAM seed sequence, 80 include a PDB structure, and human proteins were present in 49 clusters. Two PFAM families are exceptionally spread across defined

clusters: PF12697 (Abhydrolase\_6) representatives may be found in 27 clusters, and PF00561 (Abhydrolase\_1)—in 19; the next three families are present in five (PF00756, Esterase) and two (PF00326, Peptidase\_S9; PF01674, Lipase\_2) clusters. Out of seven MEROPS peptidase families S10, S15, S28, S37, and S82 localize within well-separated single clusters,



**FIGURE 3** | 2D sequence similarity-based clustering of  $\alpha/\beta$  hydrolase-like superfamily, including representatives of PFAM (gray) and MEROPS (multiple colors, see legend) families, PDB90 structures (light blue) and human proteome (pink). PFAM and ESTHER HMM profiles are marked with green and blue points, respectively. Lines connecting the points denote BLASTP mappings. Clusters are tagged with abbreviated PFAM, CAZy, and ESTHER family names written in green, pink, and blue font, respectively. Please notice the wide distribution of S09 and S33 peptidase families.

whereas S09 and S33 representatives are widely found in 19 and 22 clusters, respectively.

## 2.3 | Development of HMM Profiles

For each cluster, we calculated the HMM profile and estimated family-specific reliability cutoffs (encoded as “trusted cutoffs” in each profile). To test the performance of newly developed profiles we scanned 3 781 468 sequences of  $\alpha/\beta$  hydrolase-like proteins (collected using local PSI-BLAST searches, see Section 4) with both original PFAM and newly developed HMM profiles. PFAM profiles identified  $\alpha/\beta$  hydrolase domain in 3 487 071 protein sequences (92.2%) within the score trusted cutoff, and 3 753 025 (99.2%) if the score was not taken into account. However, 2 502 192 sequences (66.2%) were mapped into at least two PFAM families within trusted cutoff, and 1 177 into 10 or more families. New, clustering-based profiles identified 2 627 258 protein sequences (69.5%) within the defined trusted cutoff score (65 852 sequences, 1.7%, mapped to more than one cluster), and 3 765 265 (99.6%) without using the threshold.

There are regions in our clustering map (Figure 3) that still overlap, although not as massively as for original PFAM HMM profiles. The first region includes prolyl oligopeptidases (Cluster 41) and isopeptidases (Cluster 42) along with less overlapping feruloyl and acetyl xylan esterases (Cluster 45) and multifunctional Cluster 47 (vibrilactone cyclase, heroin esterase, and hormone-sensitive lipase). The second region covers Clusters 89 (epoxide hydrolase, dehalogenase, and oxidoreductase) and 91 (chloroperoxidase, lactamase, and hydrolase) with less overlapping Cluster 93 (chlorophyll dephytylase) and Cluster 39 (esterase/lyase).

## 2.4 | New Clusters of $\alpha/\beta$ Hydrolases

Out of 120 clusters, 106 are represented in the PFAM AB\_hydrolase clan, 17 of which are annotated as domains of unknown function (DUF). Ten clusters contain at least one PDB90 representative but no PFAM seeds, and one cluster consists of six human lipases only (DDHD1 phospholipase and its homologs). Eventually, three clusters cover protein families not included in the PFAM clan: PF19519 (DUF6051), PF05095 (DUF687), and PF09757 (Arb2). The latter family of histone deacetylases is structurally studied (pdb|5ikk\_A [28]) and although it lacks a catalytic triad it retains structural features characteristic of this superfamily.

## 2.5 | Uncharacterized Families

### 2.5.1 | DUF687 (PF05095)

Protein family of uncharacterized proteins found in *Chlamydia*. All family members lack catalytic residues and probably are either involved in substrate binding, or scaffolding within protein complexes. These proteins have a conserved additional C-terminal bundle of six transmembrane helices similar to the domain in Abhydrolase\_9 (PF10081) alleged lipases [29]. Hence, DUF786 members might function at the membrane as inactive homologs of lipids processing enzymes.

### 2.5.2 | DUF6051 (PF19519)

The bacterial family found predominantly in Bacteroidales, *Chryseobacterium*, Bacteroidetes, and other CFB group bacteria species. Family members retain catalytic triad as well as two structural insertions after strands  $\beta 4$  and  $\beta 6$ . These additional elements interact with each other above the active site and form a double cap-like structure charged positively from the active site cleft. Genes coding for DUF6051 proteins co-occur with genes encoding 2TM and LytTR domains related to developing antigenic variation [30] and virulence regulation [31], respectively, as well as with long-chain-fatty-acid-CoA ligase, and PadR transcription regulator sensing environmental conditions [32]. Although *Chryseobacterium* species are rarely pathogenic, some species display potency to degrade demanding structures like cuticular exoskeleton [33]. Taking into account that many  $\alpha/\beta$  hydrolase-like families group lipases, DUF6051 proteins might function as enzymes cleaving specific membrane parts during the infection process.

### 2.5.3 | DUF6792 (PF20591)

The bacterial family is found in Firmicutes, mostly in *Bacillus*, *Priestia*, and *Alkalihalobacillus*. Members of this family have a small, mostly unstructured lid located after strand  $\beta 4$ . The  $\alpha/\beta$  hydrolase-like domain is surrounded by two helical regions: N-terminal comprising two long helices and another, four-helix bundle, inserted in-between strand  $\beta 8$  and catalytic histidine. None of these additional helices are predicted to be transmembrane and both display patches of negative charge on their surfaces, probably for interaction with other either proteins or macromolecules. Insertion of the whole domain (helical bundle, 229 residues) immediately preceding catalytic histidine, located on the loop, is unique in the whole  $\alpha/\beta$  hydrolase-like superfamily, and such arrangement might lead to catalytic site stabilization only upon substrate binding. Genes encoding DUF6792 representatives co-occur with genes encoding PsbP-like domains (PF18933); PsbP proteins are associated with photosystem II functioning and optimize the availability of  $\text{Ca}^{2+}$  and  $\text{Cl}^-$  ions in higher plants [34]. However, the more detailed function of DUF6792 proteins remains unknown.

### 2.5.4 | C2orf69 (PF10561)

Eukaryotic family of catalytically active  $\alpha/\beta$  hydrolases found in animals. Human representative, C2orf69, is essential for mitochondrial respiratory chain functioning [35], however, no more details regarding its functions are known. This protein has a 29 residue-long insertion after strand  $\beta 4$  modeled as rather unstructured, as well as an extensive unstructured region between helix  $\alpha B$  and strand  $\beta 5$ , unusually located on the opposite side of the protein than the catalytic pocket.

### 2.5.5 | DUF915 (PF06028)

Although structures of three representatives of DUF915 family have been solved (lin2722 products from *Listeria innocua* (pdb|3ds8), SE\_1780 protein of unknown function from

*Staphylococcus epidermidis* (pdb13fle), putative cell surface hydrolase from *Lactobacillus plantarum* WCFS1 (pdb13lp5)), their function remains unknown. Family members are present in Firmicutes, for example, in *Staphylococcus*, *Listeria*, *Streptococcus*, and *Enterococcus*. In *Listeria* genomes, DUF915 genes are surrounded by CynX/NimT family MFS cyanate transporter, FadR/GntR fatty acid-responsive transcription factor which binds acyl-coA [36], and AI-2E family transporter of quorum-sensing signal autoinducer 2 [37], whereas in *Staphylococci* it co-occurs with AmaP (alkaline shock response membrane anchor protein), BCCT family transporter, zinc-binding alcohol dehydrogenase, oxidoreductase, 6-phospho- $\beta$ -galactosidase, PTS transporter subunits EIIC and IIA. The genomic context might suggest functions related to cell membranes, but no more specific prediction might be made for now.

### 2.5.6 | DUF829 (PF05705)

Protein family widely present in Eukaryota. Human representative, TMEM53 (Q6P2H8) blocks cytoplasm–nucleus translocation of Smad proteins in osteoblasts [38] and has an inhibitory effect against SARS-CoV by disrupting NSP8–NSP12 interaction for viral RNA synthesis [39]. It has a hydrophobic, two-helix lid covering the catalytic site, suggesting that this protein becomes activated after reaching the lipid phase. The yeast genome encodes several DUF829 proteins; for instance, ICT1 protein (YLR099C) is an acyltransferase involved in membrane remodeling [40], LPX1 (YOR084W) is a peroxisomal matrix-localized lipase [41], and LDH1 (YBR204C) is an esterase and triacylglycerol lipase for lipid homeostasis [42]. Therefore, besides the intriguing antiviral role of TMEM53, in general, DUF829 proteins might function as esterases/lipases at the cell's membranes.

### 2.5.7 | DUF3089 (PF11288)

This bacterial family is present mainly in  $\alpha$ -proteobacteria, CFB bacteria,  $\gamma$ -proteobacteria, and others. In  $\alpha$ -proteobacteria DUF3089 genes are surrounded by Holliday junction resolvase (RuvX), holo-(acyl-carrier-protein) synthase, and Asp-tRNA(Asn)/Glu-tRNA(Gln) amidotransferase subunits GatA, GatB and GatC. Modeled DUF3089 proteins from Paracoccaceae bacterium (A0A2D5TGJ5,  $\alpha$ -proteobacteria) and Bacteroidota bacterium (A0A3M1JDV7, CFB bacteria) have positively charged catalytic clefts, which suggests binding negatively charged substrates. Together with co-occurrence with genes encoding nucleic acids processing enzymes, this might suggest that DUF3089 proteins are also engaged therein as acyltransferases.

### 2.5.8 | DUF3530 (PF12048)

Protein family found in *Pseudomonas* species. Due to the lack of a catalytic triad, family members probably play non-enzymatic roles. Genomic neighborhood of DUF3530 genes is very conserved and includes LPS-assembly protein LptD involved in the assembly of lipopolysaccharide at the surface of the outer membrane, phosphotransferase, *N*-acetylmuramate

alpha-1-phosphate uridylyltransferase MurU, TerB family tellurite resistance protein, sensor histidine kinase, response regulator transcription factor, and ABC transporter ATP binding protein. Modeled structure of DUF3530 family protein (Q9I5T9, PA0599) from *Pseudomonas aeruginosa* shows minimal  $\alpha/\beta$  hydrolase domain with a negatively charged shallow cleft in place of the catalytic site. According to the STRING database, PA0599 protein co-occurs with PA1463 (CheW chemotaxis protein), PA3353 (flagellar brake-like protein), and PA5037 (ATPase) to list only a few. Besides the connection to the bacterial membrane, details about DUF3530 function remain elusive.

### 2.5.9 | DUF818 (PF05677)

Small protein family found exclusively in *Chlamydia* species. Their catalytic cleft is flanked with a small, two-helix lid-like structure inserted after strand  $\beta$ 6, and three longer N-terminal helices of which the second one shows a non-zero probability of being transmembrane.

### 2.5.10 | DUF2920 (PF11144)

Protein family specific to *Campylobacter* species. In bacterial genomes, DUF2920 genes are flanked with genes encoding HAD-superfamily hydrolase (haloacid dehalogenase-like hydrolase), acyl carrier protein (transporting fatty acid chains between fatty acid synthases), pseudaminic acid cytidyltransferase (catalyzing a step in the biosynthesis of pseudaminic acid used for flagellin modification), and UDP-2,4-diacetamido-2,4,6-trideoxy- $\beta$ -L-altropyranose hydrolase (also involved in pseudaminic acid synthesis). DUF2920 proteins have a small additional domain assembled from two insertions: first after strand  $\beta$ 4 (inserted two  $\beta$ -strands and two helices) and the second after strand  $\beta$ 6 (two  $\beta$ -strands), together forming a curved  $\beta$ -sheet with peripheral  $\alpha$ -helices. This domain does not resemble other known protein structures to provide any hint about its functional implications.

### 2.5.11 | DUF1057 (PF06342)

Mostly eukaryotic family of proteins present in, for example, nematodes, mites, oomycetes, spiders, and gastropods. DUF1057 protein from *Caenorhabditis elegans* (O16919) are predicted to localize in mitochondria (DeepLoc: 0.78; WoLF PSORT: mito 17, cyto 15). AlphaFold model of Hydrolase\_4 domain-containing protein (O16919) shows a three-helix lid inserted after strand  $\beta$ 6. However, we could find no hints for the potential function of these proteins.

### 2.5.12 | DUF900 (PF05990)

Bacterial family found in  $\alpha$ -proteobacteria, predominantly in *Rhizobium*, but also in *Sinorhizobium*, *Agrobacterium*, *Mesorhizobium*, and others. TMHMM identified an N-terminal transmembrane helix in *Rhizobium meliloti* DUF900 protein (Q92YA4) and SignalP annotated it as lipoprotein signal peptide (Sec/SPII), although with low probability (0.44). DUF900

proteins have extensive, 11-strand  $\beta$ -sheet, and their catalytic site is buried between longer helix  $\alpha$ B and an additional helix inserted after strand  $\beta$ 7. Their function is unknown.

### 2.5.13 | DUF726 (PF05277)

A family of eukaryotic proteins, including human transmembrane and coiled-coil domain-containing protein 4 (TMCO4, Q5TGY1). Despite its name, TMCO4 does not seem to have any transmembrane element according to DeepTMHMM prediction, which is consistent with the 3D model of the protein. The majority of family members contain only one,  $\alpha/\beta$  hydrolase-like domain, with few exceptions: *Aspergillus* species have additional glutamyl-tRNA amidotransferase complex subunit Gta3 (PF20978); *Alternaria* (major plant pathogens) have Pex14 N-terminal domain (PF04695, peroxisomal membrane anchor binding PTS1 receptor); *Trichinella* parasitic roundworms have TRAM (PF01938), radical SAM (PF04055), and fucosyltransferase (PF17039 and PF00852, transfers fucose from GDP-fucose to GlcNAc).

### 2.5.14 | DUF1350 (PF07082)

A family predominantly present in Cyanobacteria and eudicots, but also in monocots, algae, and plants. In Synechococcales, its representatives located in genomic proximity to acetate-CoA ligase, solanesyl diphosphate synthase (ubiquinone synthesis [43]), HAD family phosphatase (housekeeping detoxification [44]), and peroxiredoxin (oxidative stress protection [45]), suggesting potential involvement of DUF1350 proteins in ROS stress maintenance in photosynthetic organisms. Yet, the detailed roles of these proteins remain to be uncovered.

### 2.5.15 | Cluster 64

A part of Abhydrolase\_6 PFAM family (PF12697) found in high G+C Gram-positive bacteria, mainly in *Streptomyces*, *Mycolicibacterium*, *Mycobacterium*, *Pseudonocardia*, and *Phodococcus* species. The detailed function of this family remains unknown, however, its members share a non-canonical catalytic triad with serine residue substituted with aspartic acid, which makes them candidates for epoxide hydrolases. In the genomes of *Streptomyces*, genes coding for these proteins are surrounded by genes encoding NADPH-dependent FMN reductase, LLM class flavin-dependent oxidoreductase, NtaA/DmoA family FMN-dependent monooxygenase, as well as multiple ABC transporter proteins: ATP-binding, permease, substrate-binding. Oxidoreductases and dehydrogenases, along with MFS transporters, are all encoded in genomic proximity in Mycolibacteria and Mycobacteria. This might suggest involvement in degradation pathways of indoles or styrenes which link the action of oxidoreductase with epoxidase [46].

## 2.6 | Structural Diversity

In general,  $\alpha/\beta$  hydrolases are conservative regarding their structural core. Central  $\beta$ -sheet retains its characteristic twisted shape

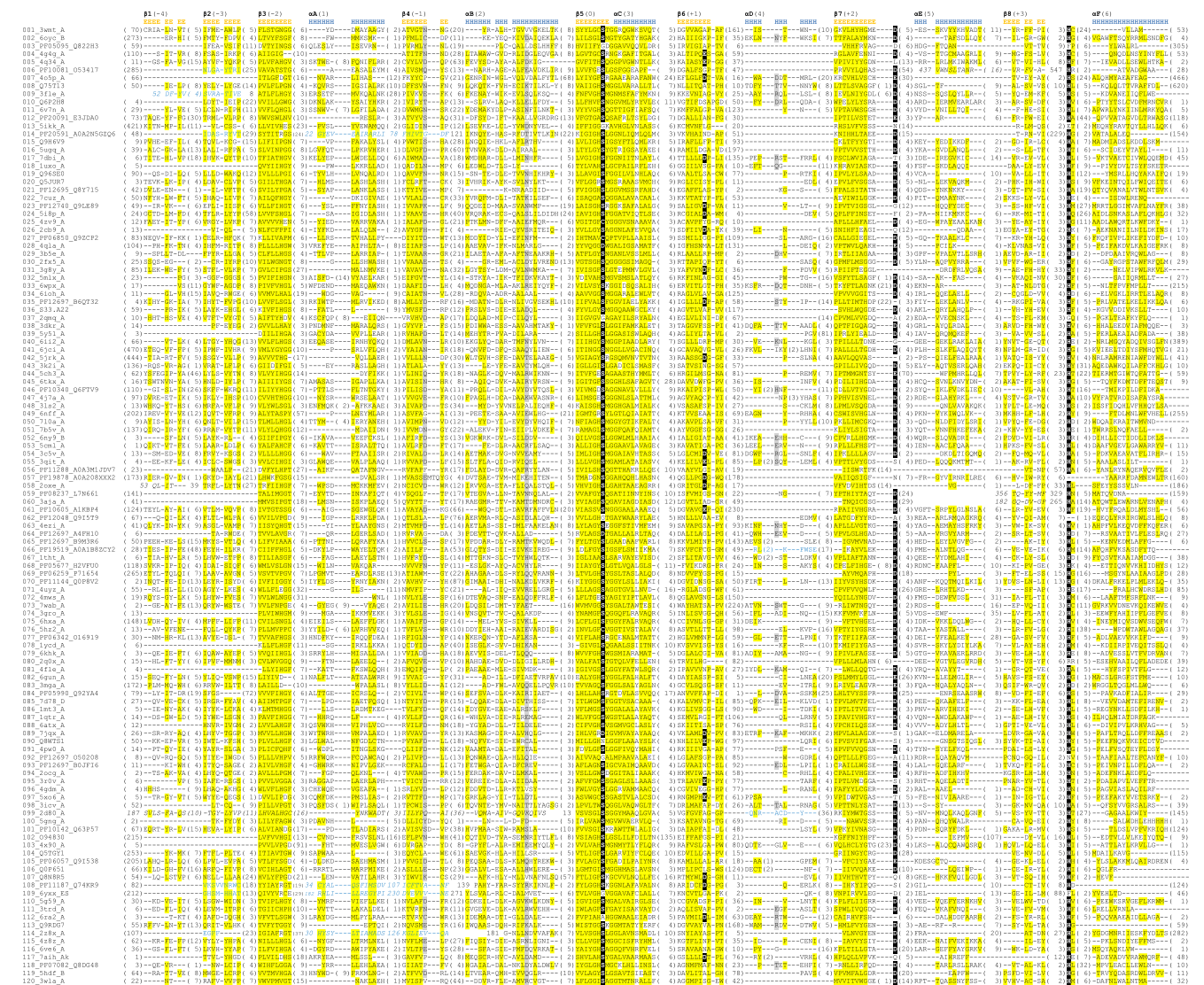
wrapping around the helices  $\alpha$ A and  $\alpha$ F across the whole superfamily. The most conserved are central hydrophobic  $\beta$ -strands, especially strand 5 which together with helix  $\alpha$ C form a hydrophilic elbow at the heart of the catalytic site. Peripheral elements like strands  $\beta$ 1,  $\beta$ 2, and  $\beta$ 8 are more diverse; some enzymes lack strands  $\beta$ 1 and  $\beta$ 2 (lipase from *P. aeruginosa* (pdb14o5p) [47] or human lysosomal phospholipase A2 [pdb14x90] [48]) or strand 8 (ribosomal protein mL73 from *Leishmania major*, pdb17aih [49]). In SE\_1780 protein of unknown function from *S. epidermidis* (pdb13fle) strands  $\beta$ 1 and  $\beta$ 2 are permuted and reversed in direction. Similar permutation and strand direction reversal may be observed in at least two lipases (extracellular lipase from *Pseudomonas* pdb12z8x [50], lipase II from *Rhizopus niveus*, pdb11lgy [51]), as well as in Mbeg1-like (PF11187) and DUF6792 (PF20591) families. The most variable helix  $\alpha$ D connects strands  $\beta$ 6 and  $\beta$ 7—in many structures it is either spatially displaced or deteriorated into a loop (see Figure 4).

## 2.7 | Active Site Variation

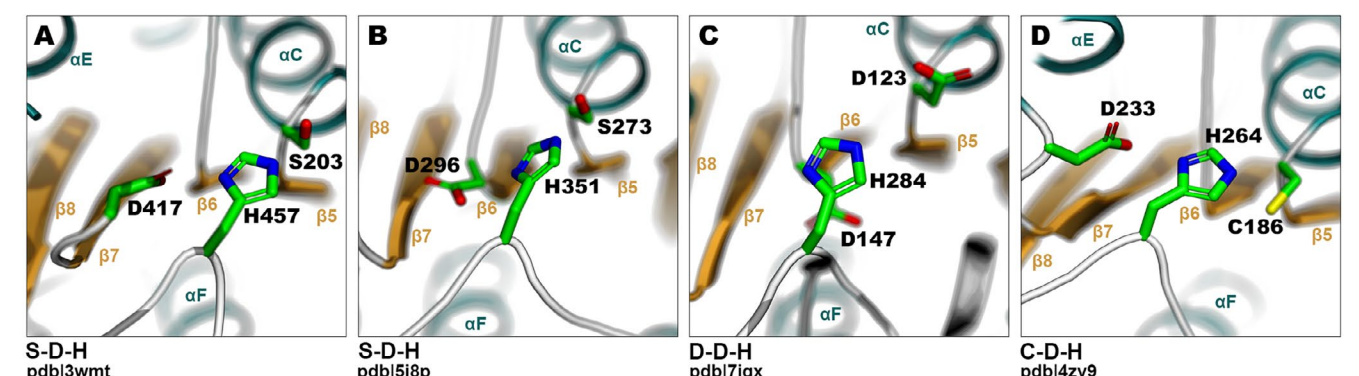
Catalytic sites of  $\alpha/\beta$  hydrolases are conserved regarding their spatial location within the structural core. Serine residue is always harbored at the nucleophilic elbow, and the histidine reaches the catalytic site from the loop connecting strand  $\beta$ 8 and helix  $\alpha$ F. Third, acidic residue in most cases is located at the C-terminus of strand  $\beta$ 7 (Figure 5A), but in many clusters it migrates to the strand  $\beta$ 6 (Figure 5B). This reconfiguration is not specific to any particular structural or functional traits (see Figure 4 and Table S1). On the other hand, the composition of the catalytic triad may vary depending on the catalyzed reaction. The most common is the canonical S-D/E-H triad represented by the majority of clusters. Epoxide hydrolases change serine to aspartic acid residue resulting in D-D/E-H triad (clusters: 28, juvenile hormone epoxide hydrolase [52]; 89, epoxide hydrolase, pdb17jqx [53], and 64 which is a part of Abhydrolase\_6 family, Figure 5C). C-D-H triad, with cysteine instead of serine (Figure 5D), is present in diene lactone hydrolase, 4zi5 [54] (Cluster 25) but also in poly(3-hydroxyalkanoate) depolymerase PhaZ, F8GXT6 [55] (Cluster 27), and polyhydroxyalkanoate (PHA) synthase, 5hz2 [56] (Cluster 76, DUF3141). Extraordinarily, some  $\alpha/\beta$  hydrolases, like carboxy lyases, function with a D/E-H dyad, without the hydrophilic serine (with some variations, discussed in [11]) (Cluster 39). Coenzyme A-dependent lysophosphatidic acid acyltransferase, cgi-58 (Q8WTS1, Cluster 90) also lacks catalytic serine and functions as lysophosphatidic acid acyltransferase [57], as well as a cofactor for the activity of lipases [58]. However, the exact catalytic site of this enzyme remains unknown.

## 2.8 | Caps and Lids

The versatility of  $\alpha/\beta$  hydrolases comes from their diversity in structural additions to the core, often referred to as lids or caps. Lids tend to be smaller and made of few helices at most but predominantly are mobile in managing access to the catalytic site depending on circumstances (Figure 6A). Caps in turn are bigger and immobile, and statically shape the access to the active site pocket which allows only specific substrates to enter (Figure 6B). Disregarding their size, the ultimate difference between lid and cap lies in their mechanics—for the lid

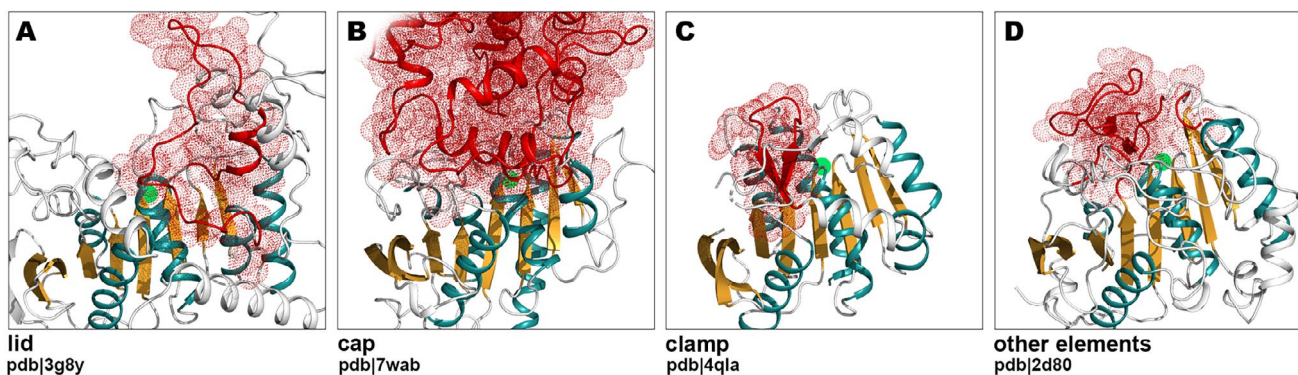


**FIGURE 4** | Structure-guided multiple sequence alignment of structural core elements for  $\alpha/\beta$  hydrolase representatives. Each sequence is tagged with cluster number and Uniprot/PDB identifier. The numbers provided in parentheses denote residues omitted from the alignment for clarity. Sequences written in italics indicate permutation and are preceded by the number of the first permuted residue, whereas sequences in turquoise are written backward and correspond to elements of reversed direction. Sequence conservation is marked with highlights following the scheme: Yellow—non-polar, gray—charged, and black—confirmed or predicted active site residues.



**FIGURE 5** | Examples of catalytic residues in  $\alpha/\beta$  hydrolase-like enzymes. (A) The most common triad in feruloyl esterase B from *Aspergillus oryzae* (pdb3wmt [59]). (B) Example of acidic residue migration to strand  $\beta_6$  in human Platelet-activating factor acetylhydrolase (pdb5i8p [60]). (C) D-D/E-H triad is common in epoxide hydrolases, here in cif-like epoxide hydrolase from *Burkholderia cenocepacia* (pdb7jqx [53]). (D) *Escherichia coli* dienelactone hydrolase catalytic triad having serine substituted with cysteine (pdb4zv9, no publication available).





**FIGURE 6** | Common structural add-ons to  $\alpha/\beta$  hydrolase-like fold core. (A) Lids, a small, mobile insertions managing the access to the catalytic site depending on conditions, putative hydrolase from *Phocaecicola vulgatus* (pdb|3g8y, no publication available). (B) Caps are much bigger and rigid structures covering active site cleft to allow only certain substrates to enter, prolyl endoprotease from *Aspergillus niger* (pdb|7wab [70]). (C) Clamps, the characteristic lid-like structures of  $\beta$ -hairpin, esterase domain from *Rhodothermus marinus* (pdb|5cml [69]). (D) Besides lids and caps, there is a diversity of other structural insertions lining the rim of the catalytic pocket, polyhydroxybutyrate depolymerase from *Penicillium funiculosum* (pdb|2d80 [71]). Discussed structural additions are rendered in red; the core  $\alpha/\beta$  hydrolase-like fold is colored in orange and teal; localizations of catalytic sites are marked with green dots.

two conformations should be observed, open, and closed [19]. Bauer et al. reported lids to be inserted into the structural core at five spots:  $\beta_{+1}/\beta_{+2}$  (after  $\beta_6$  according to our numbering),  $\beta_{-1}/\beta_0$  (after  $\beta_4$ ),  $\beta_{-4}/\beta_{-3}$  (after  $\beta_1$ ),  $\beta_{+3}/\beta_{+4}$  (after  $\beta_8$ ), and  $N_{\text{term}}/\beta_{-3}$  (before  $\beta_2$ ). The most prevalent localization of the lid (over 40 clusters), is between strand  $\beta_6$  and helix  $\alpha_D$ . Its immediate proximity to the catalytic site couples substrate binding with the locking catalytic site in its active conformation. Consistently, the remaining observed lid insertion sites include: strand  $\beta_8$  (polyhydroxybutyrate hydrolase [pdb|5mtx] [61] and pancreatic lipase [pdb|20xe] [62]), strand  $\beta_4$  (yeast FSH1 phospholipase [pdb|1ycd] [63, 64], bacterial esterase [pdb|1qlw] [65], and human phospholipase DDHD2 [O94830] [66], as well as within DUF6792 [PF20591] and Abhydrolase\_8 [PF06259] [67] families), strand  $\beta_2$  (lipoprotein-associated phospholipase A2, [pdb|5i8p] [60]), and strand  $\beta_1$  (Family L3 lipase [pdb|2z8x] [50]). Ferulic acid esterases (pdb|7b5v, pdb|5cml) have a lid-like structure, named clamp, consisting of a single  $\beta$ -hairpin located after  $\beta$ -strand 6 [68, 69] (Figure 6C).

Caps, similarly to lids, are often located after  $\beta$ -strand 6, in clusters representing various functions, for example, lysosomal acid lipase (pdb|6v7n) [72], chlorophyllase (pdb|8fjd) [73], polyneuridine aldehyde esterase (pdb|2wfl) [74], hydroxynitrile lyase (pdb|1dwo) [75], methyl salicylate esterase (pdb|1xkl) [76], as well as in multiple peptidases: serine protease Hip-1 (pdb|5uqg) [77], protease cathepsin A (pdb|4mws) [78], prolyl endoprotease (pdb|7wab) [70], tricorn-interacting aminopeptidase F1 (pdb|1mt3) [79], prolyl aminopeptidase (pdb|1qtr) [80], secreted tripeptidyl aminopeptidase, Q9RDG7, and PhoPQ-activated pathogenicity-related proteins [81]. Additionally, juvenile hormone epoxide hydrolase (pdb|4qla) [52], hormone-sensitive lipase (pdb|4j7a) [82], and DUF6051 (PF19519) have double caps—two structural insertions placed above catalytic site one on top of the other. However, the distinction between cap and non-cap insertion may be approximate, because many  $\alpha/\beta$  hydrolases, although devoid of lid/cap, still have core extensions shaping entrance to the active site, like PHB depolymerase (pdb|2d80) [71] with a crevice

formed at enzyme's surface allowing single polymer molecule to enter catalytic site (Figure 6D).

## 2.9 | Additional Domains

Representatives of many  $\alpha/\beta$  hydrolase clusters possess family-conserved additional protein domains, corresponding to their biological functions. For instance, fungal glucuronol esterases (part of Cluster 4), found mainly in Basidiomycota and Ascomycota, have, among the others, additional fungal cellulose binding domain (PF00734, CBM\_1). Abhydrolase\_9 family members (PF10081, Cluster 6) are often fused with Abhydrolase\_9\_N domain (PF15420), which is a transmembrane domain found in lipases [29]. Similarly, poly- $\beta$ -hydroxybutyrate polymerases (DUF3141, PF11339, Cluster 76), in many cases, have additional N-terminal add-on—PhaC\_N helical domain (PF07176); thioesterases (PF08840, Cluster 43) tend to be preceded by acyl-CoA thioester hydrolase/BAAT N-terminal region (PF04775, Bile\_Hydr\_Trans) of  $\beta$ -sandwich structure; histone deacetylases have N-terminal catalytic domain (PF00850) followed by  $\alpha/\beta$  hydrolase-like inactive Arb2 domain (PF09757, Cluster 13) facilitating protein-protein interactions [83]. In bacterial thioesterases (Cluster 26, PF00975)  $\alpha/\beta$  hydrolase can be preceded by several domains coherently functioning in peptide antibiotics biosynthesis: Condensation domain (PF00668), AMP-binding (PF00501), AMP-binding\_C (PF13193), and additional phosphopantetheine attachment site domain (PF00550, PP-binding). Finally, in eukaryotic phospholipases (e.g., phospholipase DDHD2, O94830, Cluster 102)  $\alpha/\beta$  hydrolase domain is often fused to WWE (PF02825), SAM\_1 (PF00536), and DDHD (PF02862) domains for interaction with other proteins.

Three peptidase families also have obligatory additional domains: prolyl oligopeptidases (PF00326, Cluster 41) require N-terminal Peptidase\_S9\_N (PF02897)  $\beta$ -propeller structure, isopeptidases (PF00326, Cluster 42) have N-terminal WD40-like Beta Propeller Repeat (PD40, PF07676), whereas bacterial

dipeptidyl peptidases (PF02129, Cluster 49) have C-terminal  $\beta$ -sandwich domain (PepX\_C, PF08530) and, in firmicutes, N-terminal helical domain (PepX\_N, PF09168) mediating dimerization.

## 2.10 | $\alpha/\beta$ Hydrolase-Like Peptidases

There are seven MEROPS peptidase families within the  $\alpha/\beta$  hydrolase-like superfamily, all belonging to the MEROPS SC clan: S09, S10, S15, S28, S33, S37, and S82. S10 family overlaps with the definition of PFAM PF00450 (Peptidase\_S10 serine carboxypeptidase) and includes multiple eukaryotic carboxypeptidases: membrane peptidases (e.g., yeast's KEX1 processing toxin precursors and  $\alpha$ -factor [84], and facilitating cell fusion during mating [85]), vacuolar peptidases (e.g., carboxypeptidase Y [86]), or lysosomal protective peptidases (e.g., human CTSA [87]). S15 (PF02129, Peptidase\_S15 X-Pro dipeptidyl-peptidase) is present in bacteria (Firmicutes,  $\beta$ -proteobacteria, high G+C Gram-positive bacteria, CFB bacteria) are peptidases removing N-terminal dipeptides from their targets [88]. S15 includes also bacterial cocaine esterases enabling bacteria inhabiting coca plants to harness cocaine as a carbon source [89]. S28 (PF05577, Peptidase\_S28 serine carboxypeptidase S28) is mostly fungal (Ascomycota, Basidiomycota) and plant family, but its representatives are found also in animals, including humans. This family includes carboxypeptidases (e.g., human lysosomal Pro-X carboxypeptidase PRCP) and proline-specific dipeptidyl peptidases (e.g., dipeptidyl peptidase 2, DPP7 [90]) important for peptide hormone signaling [91]. In plants, S28 peptidases may regulate gametogenesis [91]. S37 (PF05576, Peptidase\_S37 PS-10 peptidase S37) can be found predominantly in *Streptomyces*, *Bacterioides*, *Prevotella* species; S37 representative from *Streptomyces mobaraensis*, ptp (prolyl tri/tetrapeptidyl aminopeptidase) takes part in the maturation of transglutaminase by cleaving away N-terminal tetrapeptide [92]. S82 (PF10142, PhoPQ-related PhoPQ-activated pathogenicity-related protein) members are barely characterized. Its representatives are involved in virulence [81] and can inhibit cell proliferation in *Dictyostelium* [93]. These five peptidase families are well-defined and clearly discernible from the remaining  $\alpha/\beta$  hydrolase-like clusters, which is not the case for S09 and S33.

MEROPS representatives of families S09 (PF00326, Peptidase\_S9, and other PFAM families) and S33 (PF00561, Abhydrolase\_1, and others) are found in 19 and 22 clusters, respectively. Of these, two S09 clusters and three S33 clusters contain genuine peptidases. Taking into account the number of protein sequences assigned to each cluster, S9 bona fide peptidases account for 34.7% of all sequence space covered by all 19 clusters (321 741/927 482), while for S33—14.2% (102 992/727 819).

The S09 family covers two discernible peptidase clusters. The first includes prolyl oligopeptidases cleaving a variety of peptides (like neuropeptides or peptide hormones) after proline residue (S9A MEROPS subfamily, e.g. S09.001; *Novosphingobium capsulatum* prolyl oligopeptidase, pdbl1yr2 [94]), oligopeptidases B (e.g., S09.010; bacterial oligopeptidase B, pdbl6tf5 [95]) but also peptide cyclases (e.g., S09.078; PCY1, pdbl5uzw [96]) catalyzing transamidation instead of peptide bond hydrolysis.

The second S09 cluster contains among the others isopeptidases targeting lariat knotted peptides (e.g., S09.032; sphingopyxin I lasso peptide isopeptidase, pdbl5jrk [97]), dipeptidyl aminopeptidases degrading bioactive peptides (S9B subfamily, e.g., S09.003; dipeptidyl aminopeptidase IV, pdbl2ecf [98]), and aminoacyl peptidases cleaving acylated amino acids for clearing cytotoxic denatured proteins (S9C subfamily, e.g., S09.004; pdbl518s [99]), glutamyl endopeptidases (S9D subfamily, e.g., S09.021).

S33 peptidases fall into three clusters. The first contains prolyl aminopeptidases removing N-terminal proline from specific peptides (e.g., S33.001; pdbl1qtr [80]). The second includes tripeptidyl-peptidases (e.g., S33.002), and Hip1 peptidase cleaving GroEL2 (pdbl5ugq [77]). Finally, the third S33 cluster groups proline iminopeptidases (pdbl3nwo [100]), aminopeptidase grinding short peptides coming from tricorner degradation down into single amino acids (pdbl1mt3 [79]). The remaining clusters containing S09 and S33 non-peptidase members represent a diversity of functions, including epoxide hydrolases, haloalkane dehydrogenases, lipases, esterases, lyases, thioesterases, and others (see Table S1).

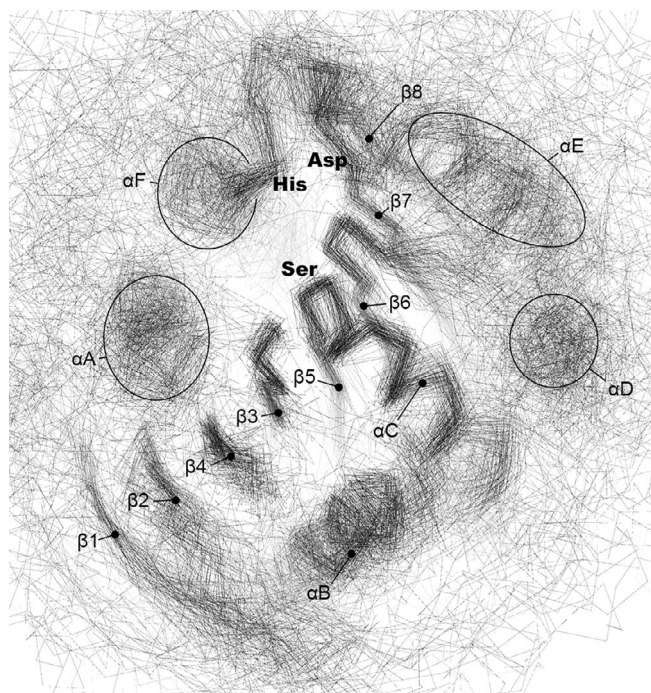
## 3 | Discussion

$\alpha/\beta$  Hydrolase-like proteins form one of the more extraordinary superfamilies. Its members, numbering over 3.7 million, share a well-defined structural core organized around an extensive, eight-stranded  $\beta$ -sheet of characteristic twist conserved across all families, as well as a catalytic triad tethered at the same location within the structure. However, relatively high structure conservation goes along with a high diversity of functions and one dominant setting of the active site is capable of catalyzing different enzymatic reactions. Moreover, switching between distinct enzymatic activities may require as little as two amino acid substitutions [13]. This phenomenon may be explained by the modular character of these proteins [19], whose function depends on small changes of residues lining the active site pocket, and structural elements added to the core. In consequence, similarities within conserved parts of these proteins may mask smaller differences to blend functionally distinct enzymes into the same families. Hence,  $\alpha/\beta$  hydrolases cover 77 PFAM families which are highly diverse in terms of mutual sequence similarities; families like Abhydrolase\_6 (PF12697), Abhydrolase\_1 (PF00561), and Hydrolase\_4 (serine aminopeptidase S33, PF12146) not only contain multiple, functionally unrelated proteins but also overlap regarding sequences belonging to them. On the other hand, many  $\alpha/\beta$  hydrolase families are evolutionarily distinct and separate from other superfamily representatives.

Considering all the above, the development of the curated set of families/clusters definitions exposing a more detailed structure of  $\alpha/\beta$  hydrolase superfamily must not rely solely on the automatic construction of sequence profiles; rather it should include the assessment of pairwise sequence similarities between representative proteins to define collectively similar groups. Such an initial set of clusters may be further extended to cover as many superfamily members as possible, yet with special attention paid to minimizing the overlaps.

HMM profiles developed in this project cover the majority of  $\alpha/\beta$  hydrolase-like protein sequences while maintaining distinction between the clusters. Although the superfamily coverage of newly provided profiles within the estimated, family-specific score thresholds, is lower compared to PFAM profiles (69.5% vs. 92.2%), they are still able to identify 99.6% of collected  $\alpha/\beta$  hydrolases sequences; in the latter case, additional inspection of mappings may be needed to decide on an eventual family membership. Although general HMM profiles cover more protein sequences within reliability thresholds, they are not conclusive about the family assignment for 66.2% of the superfamily members which simultaneously satisfy reliability thresholds for two or more profiles. It should not be excluded that the non-covered realm of the superfamily, or at least some part of it, would form additional clusters, not represented within the initial set of PFAM, PDB90, MEROPS, and human protein sequences.

Despite careful picking of each cluster member and reliability threshold estimation, some clusters display minor overlaps adding up to 1.7% of the superfamily members. The overlaps were present mainly between Clusters 41 and 42 (both peptidases), and within the group of Clusters 89, 91 (both containing proteins of diverse functions), 93 (chlorophyll dephytylase), and 39 (esterases/lyases). Observed, the functional diversity of Clusters 89 and 91 could not be separated with the presented approach. However, many broadly defined families were successfully split into smaller, more functionally coherent clusters, with extreme cases of Abhydrolase\_6 (PF12697) distributed within 27 clusters and Abhydrolase\_1 (PF00561) whose members may be found in 19 clusters.



**FIGURE 7** | Superimposition of 3D structures representing all 120 defined clusters of  $\alpha/\beta$  hydrolase-like proteins. While strands  $\beta 3$ ,  $\beta 4$ ,  $\beta 5$ ,  $\beta 6$ ,  $\beta 7$ , and helices  $\alpha B$  and  $\alpha C$  are well preserved, the remaining helices are much less pronounced and therefore seem blurred.

In spite of the conservation of the structural core (Figure 7) and catalytic residues signature, new  $\alpha/\beta$  hydrolase families may still be hard to identify. This broad superfamily covers proteins from across the tree of life, diverged in amino acid sequences. Moreover, insertions of bigger elements to the structural core may disrupt multiple sequence alignments used for homology detection algorithms, hindering correct annotation. On the other hand, thanks to the contemporary advancements in computational methods, many families can now be classified as clans/superfamily members [23, 101]. Yet, such a rough assignment barely sheds light on the potential functions of so-called “DUFs.” To hypothesize on their roles, one might need to take into account additional family traits, like domain architectures, genomic neighborhoods, or additional structural elements. However, functional prediction remains challenging for  $\alpha/\beta$  hydrolases, because, as stated above, their detailed functions rely on small sequential or structural traits.

Additionally, despite the conserved structural core, the superimposition of PDB structures representing all superfamily members is not trivial in many cases. Although all structures retain nearly identical elements surrounding the hydrophilic elbow: hydrophobic  $\beta$ -strands folding around helices  $\alpha A$  and  $\alpha F$ ,  $\beta$ -strands located at both rims of the central  $\beta$ -sheet as well as helices may display greater distortions. Eventually, peripheral strands, like  $\beta 1$  and  $\beta 2$  may be missing, circularly permuted or even have reversed directions suggesting deeper evolutionary changes. The most structurally diverse element is helix  $\alpha D$  connecting strands  $\beta 6$  and  $\beta 7$ . The overwhelming majority of  $\alpha/\beta$  hydrolases have a cap or lid inserted in-between strand  $\beta 6$  and helix  $\alpha D$  so that the latter functions as an elastic part of the hinge mechanism controlling the cap/lid functioning, explaining the extraordinary diversity of the helix.

Catalytic triad: Ser-Asp/Glu-His—one of the hallmarks of  $\alpha/\beta$  hydrolases, although canonical in the majority of the clusters, may vary to some extent regarding catalyzed reaction and substrate processing. Hydrophilic serine is substituted with acidic residue in epoxide hydrolases, with cysteine in lactone hydrolases, or is missing at all in carboxy lyases. The acidic residue may migrate from strand  $\beta 7$  to immediately adjacent strand  $\beta 6$  while maintaining catalytic triad geometry. Nevertheless, all analyzed structures retain characteristic hydrophobic elbow (Figure 7) allowing for the immediate distinction between  $\alpha/\beta$  hydrolases and other, similar protein folds, for example, Rossmann-like with its crossover helix corresponding to the helix  $\alpha B$ .

Superfamily-wide structural comparisons provide a unique view of conserved and diverse features of  $\alpha/\beta$  hydrolase-like proteins. They also allow for more precise designation of protein domain borders, which further affects subsequent sequence-based analyses. Based on pairwise sequence similarities between  $\alpha/\beta$  hydrolase-like domains in the initial set of representatives, using semi-manual clustering we defined 120 protein clusters, with minor exceptions discussed above, coherent structurally and functionally. By developing HMM profiles for each cluster, we provide a ready-to-use database for further research projects requiring detailed annotation of  $\alpha/\beta$  hydrolases.

## 4 | Experimental Procedures

### 4.1 | Preparing Initial Dataset

The initial set of previously known  $\alpha/\beta$  hydrolase-like proteins contained members of AB\_hydrolase PFAM [23] clan (CL0028) and corresponding PDB90 representatives (PDB structures clustered to 90% of sequence identity to reduce redundancy), as well as peptidases from MEROPS database [26] and classified to the SC clan (families: S09, S10, S15, S28, S33, S37, and S82).

### 4.2 | Identification of New $\alpha/\beta$ Hydrolase Superfamily Members

New members of the  $\alpha/\beta$  superfamily were identified using the sequence-based remote homology detection method HHSEARCH. HHSEARCH, a part of HHSUITE package [102], is a remote homology detection program which, instead of seeking for similarities between raw sequences, or even sequence profiles, boosts its sensitivity and specificity by comparing meta-profiles—statistical objects encoding information about residue conservation enriched with predicted secondary structure patterns. HHSEARCH results were additionally verified according to the conservation of essential structural core elements and catalytic residues.

### 4.3 | Clustering the Realm of $\alpha/\beta$ Hydrolases

Obtained set of protein sequences consisting of PFAM seed sequences, PDB90 and MEROPS representatives, and human proteins, was manually clustered based on pairwise sequence similarities (BLASTP [103], *E*-value threshold  $1 \times 10^{-4}$ ) with the use of Cytoscape program [104], and AutoAnnotate Cytoscape app [105].

### 4.4 | Preparing HMM Profiles

For each cluster, sequences of its members were aligned using MAFFT [106] (localpair, maxiterate 1000). Based on multiple sequence alignments, HMM profiles were built using HMMBUILD [107]. Since the initial set of HMM profiles was developed based on a very limited count of sequences which might affect their performance, we extended our analysis to the whole superfamily. For that, we collected all superfamily members using PSI-BLAST [108] runs against the nr non-redundant database (downloaded on Feb 24th, 2024) with one representative sequence per cluster as a query. The resulting dataset of 3 781 468 unique sequences was scanned with newly developed HMM profiles using HMMSCAN [107]. Based on the resulting bitscore distributions, for each profile, we selected a new set of protein sequences, clustered them with CD-HIT [109] to 90% sequence identity, and repeated the procedure of profile calculation to include a wider set of homologs. This step differed from the initial HMM creation procedure in multiple sequence alignment options—instead of an exhaustive localpair iterative approach we used “—auto” flag to MAFFT to reduce computation time for bigger alignments. Eventually, the final scanning of 3 781 468 superfamily members with the updated HMM profiles

allowed us to estimate reliability thresholds that balance profile's sensitivity and selectivity.

### 4.5 | Additional Analyses

Domain architecture analyses were based on HMMSCAN searches against the PFAM 35.0 database. Cluster mappings to the CAZy database were performed using HMMSCAN against our newly developed HMM database with CAZy representative sequences used as a query. The genomic neighborhood was derived from the NCBI genomic data with the use of the E-UTILS package. Taxonomic distributions were obtained from the NCBI taxonomy database provided along with the nr sequence database. Transmembrane elements detection was made with DeepTMHMM [110]. 3D structures for proteins representing clusters without known PDB structure were modeled using AlphaFold2 [111]. All structures were superimposed with the use of SPDBV [112], and manually curated multiple structure alignment was prepared in Discovery Studio Visualizer. Alignment manipulations/preparations were made in SEAVIEW [113]. All structures were visualized using the PYMOL program (<https://pymol.org>).

#### Author Contributions

**Fatih Ozhelvaci:** investigation, writing – original draft, formal analysis. **Kamil Steczkiewicz:** conceptualization, investigation, funding acquisition, writing – original draft, methodology, validation, visualization, writing – review and editing, software, formal analysis, project administration, supervision, resources, data curation.

#### Acknowledgments

We thank Anna Muszewska for critical reading of this article.

#### Conflicts of Interest

The authors declare no conflicts of interest.

#### Data Availability Statement

Additional data, including HMM profiles and superimposed protein structures are available through Zenodo repository under DOI [10.5281/zenodo.11182431](https://doi.org/10.5281/zenodo.11182431).

#### References

1. D. L. Ollis and P. D. Carr, “ $\alpha/\beta$  Hydrolase Fold: An Update,” *Protein and Peptide Letters* 16 (2009): 1137–1148.
2. M. Holmquist, “Alpha Beta-Hydrolase Fold Enzymes Structures, Functions and Mechanisms,” *Current Protein & Peptide Science* 1 (2000): 209–235.
3. K. A. Majorek, S. Dunin-Horkawicz, K. Steczkiewicz, et al., “The RNase H-Like Superfamily: New Members, Comparative Structural Analysis and Evolutionary Classification,” *Nucleic Acids Research* 42 (2014): 4160–4179.
4. F. Ozhelvaci and K. Steczkiewicz, “Identification and Classification of Papain-Like Cysteine Proteinases,” *Journal of Biological Chemistry* 299 (2023): 104801.
5. D. L. Ollis, E. Cheah, M. Cygler, et al., “The Alpha/Beta Hydrolase Fold,” *Protein Engineering* 5 (1992): 197–211.

6. P. Heikinheimo, A. Goldman, C. Jeffries, and D. L. Ollis, "Of Barn Owls and Bankers: A Lush Variety of  $\alpha/\beta$  Hydrolases," *Structure* 7 (1999): R141–R146.
7. P. S. Dimitriou, A. Denesyuk, S. Takahashi, et al., "Alpha/Beta-Hydrolases: A Unique Structural Motif Coordinates Catalytic Acid Residue in 40 Protein Fold Families," *Proteins* 85 (2017): 1845–1855.
8. P. S. Dimitriou, A. I. Denesyuk, T. Nakayama, M. S. Johnson, and K. Denessiouk, "Distinctive Structural Motifs Co-Ordinate the Catalytic Nucleophile and the Residues of the Oxyanion Hole in the Alpha/Beta-Hydrolase Fold Enzymes," *Protein Science* 28 (2019): 344–364.
9. D. I. Liao, K. Breddam, R. M. Sweet, T. Bullock, and S. J. Remington, "Refined Atomic Model of Wheat Serine Carboxypeptidase II at 2.2-Å Resolution," *Biochemistry* 31 (1992): 9796–9812.
10. M. Nardini and B. W. Dijkstra, "Alpha/Beta Hydrolase Fold Enzymes: The Family Keeps Growing," *Current Opinion in Structural Biology* 9 (1999): 732–737.
11. A. Rauwerdink and R. J. Kazlauskas, "How the Same Core Catalytic Machinery Catalyzes 17 Different Reactions: The Serine-Histidine-Aspartate Catalytic Triad of  $\alpha/\beta$ -Hydrolase Fold Enzymes," *ACS Catalysis* 5 (2015): 6153–6176.
12. T. D. H. Bugg, "Diverse Catalytic Activities in the Alphabeta-Hydrolase Family of Enzymes: Activation of  $H_2O$ ,  $HCN$ ,  $H_2O_2$ , and  $O_2$ ," *Bioorganic Chemistry* 32 (2004): 367–375.
13. S. K. Padhi, R. Fujii, G. A. Legatt, S. L. Fossum, R. Berchtold, and R. J. Kazlauskas, "Switching From an Esterase to a Hydroxynitrile Lyase Mechanism Requires Only Two Amino Acid Substitutions," *Chemistry & Biology* 17 (2010): 863–871.
14. B. Hofmann, S. Tölzer, I. Pelletier, J. Altenbuchner, K. H. van Pée, and H. J. Hecht, "Structural Investigation of the Cofactor-Free Chloroperoxidases," *Journal of Molecular Biology* 279 (1998): 889–900.
15. S. Lejon, J. Ellis, and K. Valegård, "The Last Step in Cephalosporin C Formation Revealed: Crystal Structures of Deacetylcephalosporin C Acetyltransferase From *Acremonium Chrysogenum* in Complexes With Reaction Intermediates," *Journal of Molecular Biology* 377 (2008): 935–944.
16. S. Roppongi, Y. Suzuki, C. Tateoka, et al., "Crystal Structures of a Bacterial Dipeptidyl Peptidase IV Reveal a Novel Substrate Recognition Mechanism Distinct From That of Mammalian Orthologues," *Scientific Reports* 8 (2018): 2714.
17. E. Cheah, G. W. Ashley, J. Gary, and D. Ollis, "Catalysis by Dienelactone Hydrolase: A Variation on the Protease Mechanism," *Proteins* 16 (1993): 64–78.
18. M. Sharma, N. N. Sharma, and T. C. Bhalla, "Hydroxynitrile Lyases: At the Interface of Biology and Chemistry," *Enzyme and Microbial Technology* 37 (2005): 279–294.
19. T. L. Bauer, P. C. F. Buchholz, and J. Pleiss, "The Modular Structure of  $\alpha/\beta$ -Hydrolases," *FEBS Journal* 287 (2020): 1035–1053.
20. U. Derewenda, A. M. Brzozowski, D. M. Lawson, and Z. S. Derewenda, "Catalysis at the Interface: The Anatomy of a Conformational Change in a Triglyceride Lipase," *Biochemistry* 31 (1992): 1532–1541.
21. F. I. Khan, D. Lan, R. Durrani, W. Huan, Z. Zhao, and Y. Wang, "The Lid Domain in Lipases: Structural and Functional Determinant of Enzymatic Properties," *Frontiers in Bioengineering and Biotechnology* 5 (2017): 16.
22. R. Thoma, B. Löffler, M. Stihle, W. Huber, A. Ruf, and M. Hennig, "Structural Basis of Proline-Specific Exopeptidase Activity as Observed in Human Dipeptidyl Peptidase-IV," *Structure* 11 (2003): 947–959.
23. J. Mistry, S. Chuguransky, L. Williams, et al., "Pfam: The Protein Families Database in 2021," *Nucleic Acids Research* 49 (2021): D412–D419.
24. N. Lenfant, T. Hotelier, E. Velluet, Y. Bourne, P. Marchot, and A. Chatonnet, "ESTHER, the Database of the  $\alpha/\beta$ -Hydrolase Fold Superfamily of Proteins: Tools to Explore Diversity of Functions," *Nucleic Acids Research* 41 (2013): D423–D429.
25. V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho, and B. Henriksat, "The Carbohydrate-Active Enzymes Database (CAZY) in 2013," *Nucleic Acids Research* 42 (2014): D490–D495.
26. N. D. Rawlings, A. J. Barrett, P. D. Thomas, X. Huang, A. Bateman, and R. D. Finn, "The MEROPS Database of Proteolytic Enzymes, Their Substrates and Inhibitors in 2017 and a Comparison With Peptidases in the PANTHER Database," *Nucleic Acids Research* 46 (2018): D624–D632.
27. A. Muszewska, M. M. Stepniewska-Dziubinska, K. Steczkiewicz, J. Pawlowska, A. Dziedzic, and K. Ginalski, "Fungal Lifestyle Reflected in Serine Protease Repertoire," *Scientific Reports* 7 (2017): 9147.
28. G. Job, C. Brugger, T. Xu, et al., "SHREC Silences Heterochromatin via Distinct Remodeling and Deacetylation Modules," *Molecular Cell* 62 (2016): 207–221.
29. M. Lazniewski, K. Steczkiewicz, L. Knizewski, I. Wawer, and K. Ginalski, "Novel Transmembrane Lipases of Alpha/Beta Hydrolase Fold," *FEBS Letters* 585 (2011): 870–874.
30. J. Kaur and R. Hora, "'2TM Proteins': An Antigenically Diverse Superfamily With Variable Functions and Export Pathways," *PeerJ* 6 (2018): e4757.
31. A. N. Nikolskaya and M. Y. Galperin, "A Novel Type of Conserved DNA-Binding Domain in the Transcriptional Regulators of the AlgR/AgrA/LyTR Family," *Nucleic Acids Research* 30 (2002): 2453–2459.
32. T. K. C. Nguyen, N. P. Tran, and J.-F. Cavin, "Genetic and Biochemical Analysis of PadR-padC Promoter Interactions During the Phenolic Acid Stress Response in *Bacillus subtilis* 168," *Journal of Bacteriology* 193 (2011): 4180–4191.
33. A. P. Page, M. Roberts, M.-A. Félix, D. Pickard, A. Page, and W. Weir, "The Golden Death *Bacillus Chryseobacterium nematophagum* Is a Novel Matrix Digesting Pathogen of Nematodes," *BMC Biology* 17 (2019): 10.
34. S. Ishihara, A. Takabayashi, K. Ido, T. Endo, K. Ifuku, and F. Sato, "Distinct Functions for the Two PsbP-Like Proteins PPL1 and PPL2 in the Chloroplast Thylakoid Lumen of Arabidopsis," *Plant Physiology* 145 (2007): 668–679.
35. E. Lausberg, S. Gießelmann, J. P. Dewulf, et al., "C2orf69 Mutations Disrupt Mitochondrial Function and Cause a Multisystem Human Disorder With Recurring Autoinflammation," *Journal of Clinical Investigation* 131 (2021): e143078.
36. D. M. van Aalten, C. C. DiRusso, J. Knudsen, and R. K. Wierenga, "Crystal Structure of FadR, a Fatty Acid-Responsive Transcription Factor With a Novel Acyl Coenzyme A-Binding Fold," *EMBO Journal* 19 (2000): 5167–5177.
37. R. E. Rettner and M. H. Saier, Jr., "The Autoinducer-2 Exporter Superfamily," *Journal of Molecular Microbiology and Biotechnology* 18 (2010): 195–205.
38. L. Guo, A. Iida, G. S. Bhavani, et al., "Deficiency of TMEM53 Causes a Previously Unknown Sclerosing Bone Disorder by Dysregulation of BMP-SMAD Signaling," *Nature Communications* 12 (2021): 2046.
39. Y.-L. Yao, Y. Luo, Q. Wang, et al., "Identification of TMEM53 as a Novel SARS-CoV Restriction Factor That Targets Viral RNA-Dependent RNA Polymerase," *Emerging Microbes & Infections* 12 (2023): 2249120.
40. A. K. Ghosh, G. Ramakrishnan, and R. Rajasekharan, "YLR099C (ICT1) Encodes a Soluble Acyl-CoA-Dependent Lyso-phosphatidic Acid Acyltransferase Responsible for Enhanced Phospholipid Synthesis on Organic Solvent Stress in *Saccharomyces cerevisiae*," *Journal of Biological Chemistry* 283 (2008): 9768–9775.

41. S. Thoms, M. O. Debelyy, K. Nau, H. E. Meyer, and R. Erdmann, "Lpx1p Is a Peroxisomal Lipase Required for Normal Peroxisome Morphology," *FEBS Journal* 275 (2008): 504–514.
42. M. O. Debelyy, S. Thoms, M. Connerth, G. Daum, and R. Erdmann, "Involvement of the *Saccharomyces cerevisiae* Hydrolase Ldh1p in Lipid Homeostasis," *Eukaryotic Cell* 10 (2011): 776–781.
43. K. Hirooka, Y. Izumi, C.-I. An, Y. Nakazawa, E.-I. Fukusaki, and A. Kobayashi, "Functional Analysis of Two Solanesyl Diphosphate Synthases From *Arabidopsis thaliana*," *Bioscience, Biotechnology, and Biochemistry* 69 (2005): 592–601.
44. J. A. Caparrós-Martín, I. McCarthy-Suárez, and F. A. Culiñez-Macià, "HAD Hydrolase Function Unveiled by Substrate Screening: Enzymatic Characterization of *Arabidopsis thaliana* Subclass I Phosphosugar Phosphatase AtSgpp," *Planta* 237 (2013): 943–954.
45. A. Perkins, K. J. Nelson, D. Parsonage, L. B. Poole, and P. A. Karplus, "Peroxiredoxins: Guardians Against Oxidative Stress and Modulators of Peroxide Signaling," *Trends in Biochemical Sciences* 40 (2015): 435–445.
46. T. Heine, W. J. H. van Berkel, G. Gassner, K.-H. van Pée, and D. Tischler, "Two-Component FAD-Dependent Monooxygenases: Current Knowledge and Biotechnological Opportunities," *Biology* 7 (2018): 42.
47. H. Hu, H. Zhang, Z. Gao, et al., "Structure of the Type VI Secretion Phospholipase Effector Tle1 Provides Insight Into Its Hydrolysis and Membrane Targeting," *Acta Crystallographica. Section D, Biological Crystallography* 70 (2014): 2175–2185.
48. A. Glukhova, V. Hinkovska-Galcheva, R. Kelly, A. Abe, J. A. Shayman, and J. G. Tesmer, "Structure and Function of Lysosomal Phospholipase A2 and Lecithin:Cholesterol Acyltransferase," *Nature Communications* 6 (2015): 6250.
49. H. Soufari, F. Waltz, C. Parrot, et al., "Structure of the Mature Kinetoplasts Mitoribosome and Insights Into Its Large Subunit Biogenesis," *Proceedings of the National Academy of Sciences of the United States of America* 117 (2020): 29851–29861.
50. C. Angkawidjaja, D.-J. You, H. Matsumura, et al., "Crystal Structure of a Family I.3 Lipase From *Pseudomonas* sp. MIS38 in a Closed Conformation," *FEBS Letters* 581 (2007): 5060–5064.
51. M. Kohno, J. Funatsu, B. Mikami, W. Kugimiya, T. Matsuo, and Y. Morita, "The Crystal Structure of Lipase II From *Rhizopus niveus* at 2.2 Å Resolution," *Journal of Biochemistry* 120 (1996): 505–510.
52. K. Zhou, N. Jia, C. Hu, et al., "Crystal Structure of Juvenile Hormone Epoxide Hydrolase From the Silkworm *Bombyx mori*," *Proteins* 82 (2014): 3224–3229.
53. N. M. Taher, K. L. Hvorecny, C. M. Burke, et al., "Biochemical and Structural Characterization of Two Cif-Like Epoxide Hydrolases From *Burkholderia cenocepacia*," *Current Research in Structural Biology* 3 (2021): 72–84.
54. P.-Y. Colin, B. Kintsjes, F. Gielen, et al., "Ultra-high-Throughput Discovery of Promiscuous Enzymes by Picodroplet Functional Metagenomics," *Nature Communications* 6 (2015): 10008.
55. G. M. York, J. Lupberger, J. Tian, A. G. Lawrence, J. Stubbe, and A. J. Sinsky, "*Ralstonia eutropha* H16 Encodes Two and Possibly Three Intracellular Poly[D-(−)-3-hydroxybutyrate] Depolymerase Genes," *Journal of Bacteriology* 185 (2003): 3788–3794.
56. J. Kim, Y.-J. Kim, S. Y. Choi, S. Y. Lee, and K.-J. Kim, "Crystal Structure of *Ralstonia eutropha* Polyhydroxyalkanoate Synthase C-Terminal Domain and Reaction Mechanisms," *Biotechnology Journal* 12 (2017): 1600648.
57. A. K. Ghosh, G. Ramakrishnan, C. Chandramohan, and R. Rajasekharan, "CGI-58, the Causative Gene for Chanarin-Dorfman Syndrome, Mediates Acylation of Lysophosphatidic Acid," *Journal of Biological Chemistry* 283 (2008): 24525–24533.
58. J. Guerrero-Santoro, M. Morizane, S.-Y. Oh, et al., "The Lipase Cofactor CGI58 Controls Placental Lipolysis," *JCI Insight* 8 (2023): e168717.
59. K. Suzuki, A. Hori, K. Kawamoto, et al., "Crystal Structure of a Feruloyl Esterase Belonging to the Tannase Family: A Disulfide Bond Near a Catalytic Triad," *Proteins* 82 (2014): 2857–2867.
60. Q. Liu, X. Chen, W. Chen, et al., "Structural and Thermodynamic Characterization of Protein–Ligand Interactions Formed Between Lipoprotein-Associated Phospholipase A2 and Inhibitors," *Journal of Medicinal Chemistry* 59 (2016): 5115–5120.
61. T. F. Kellici, T. Mavromoustakos, D. Jendrossek, and A. C. Papageorgiou, "Crystal Structure Analysis, Covalent Docking, and Molecular Dynamics Calculations Reveal a Conformational Switch in PhaZ7 PHB Depolymerase," *Proteins* 85 (2017): 1351–1361.
62. C. Eydoux, S. Spinelli, T. L. Davis, et al., "Structure of Human Pancreatic Lipase-Related Protein 2 With the Lid in an Open Conformation," *Biochemistry* 47 (2008): 9553–9564.
63. G. Ramachandran, R. Chidambaram, and V. Nachiappan, "FSH1 Encodes Lysophospholipase Activity in *Saccharomyces cerevisiae*," *Biotechnology Letters* 43 (2021): 279–286.
64. S. Quevillon-Cheruel, N. Leulliot, M. Graille, et al., "Crystal Structure of Yeast YHR049W/FSH1, a Member of the Serine Hydrolase Family," *Protein Science* 14 (2005): 1350–1356.
65. P. C. Bourne, M. N. Isupov, and J. A. Littlechild, "The Atomic-Resolution Structure of a Novel Bacterial Esterase," *Structure* 8 (2000): 143–151.
66. K.-I. Nakajima, H. Sonoda, T. Mizoguchi, et al., "A Novel Phospholipase A1 With Sequence Homology to a Mammalian Sec23p-Interacting Protein, p125," *Journal of Biological Chemistry* 277 (2002): 11329–11335.
67. M. Zheng, K. Ginalski, L. Rychlewski, and N. V. Grishin, "Protein Domain of Unknown Function DUF1023 Is an Alpha/Beta Hydrolase," *Proteins* 59 (2005): 1–6.
68. C. Kmezik, S. Mazurkewich, T. Meents, et al., "A Polysaccharide Utilization Locus From the Gut Bacterium *Dysgonomonas mossii* Encodes Functionally Distinct Carbohydrate Esterases," *Journal of Biological Chemistry* 296 (2021): 100500.
69. M.-B. V. Jensen, L. E. Horsfall, C. Wardrope, P. D. Togneri, J. Marles-Wright, and S. J. Rosser, "Characterisation of a New Family of Carboxyl Esterases With an OsmC Domain," *PLoS One* 11 (2016): e0166128.
70. K.-I. Miyazono, K. Kubota, K. Takahashi, and M. Tanokura, "Crystal Structure and Substrate Recognition Mechanism of the Prolyl Endoprotease PEP From *Aspergillus niger*," *Biochemical and Biophysical Research Communications* 591 (2022): 76–81.
71. T. Hisano, K.-I. Kasuya, Y. Tezuka, et al., "The Crystal Structure of Polyhydroxybutyrate Depolymerase From *Penicillium funiculosum* Provides Insights Into the Recognition and Degradation of Biopolyesters," *Journal of Molecular Biology* 356 (2006): 993–1004.
72. F. Rajamohan, A. R. Reyes, M. Tu, et al., "Crystal Structure of Human Lysosomal Acid Lipase and Its Implications in Cholesteryl Ester Storage Disease," *Journal of Lipid Research* 61 (2020): 1192–1202.
73. M. Jo, M. Knapp, D. G. Boggs, M. Brimberry, P. H. Donnan, and J. Bridwell-Rabb, "A Structure-Function Analysis of Chlorophyllase Reveals a Mechanism for Activity Regulation Dependent on Disulfide Bonds," *Journal of Biological Chemistry* 299 (2023): 102958.
74. L. Yang, M. Hill, M. Wang, S. Panjikar, and J. Stöckigt, "Structural Basis and Enzymatic Mechanism of the Biosynthesis of C9- From C10-Monoterpenoid Indole Alkaloids," *Angewandte Chemie, International Edition* 48 (2009): 5211–5213.
75. H. Lauble, S. Förster, B. Miehlisch, H. Wajant, and F. Effenberger, "Structure of Hydroxynitrile Lyase From *Manihot Esculenta* in

- Complex With Substrates Acetone and Chloroacetone: Implications for the Mechanism of Cyanogenesis," *Acta Crystallographica. Section D, Biological Crystallography* 57 (2001): 194–200.
76. F. Forouhar, Y. Yang, D. Kumar, et al., "Structural and Biochemical Studies Identify Tobacco SABP2 as a Methyl Salicylate Esterase and Implicate It in Plant Innate Immunity," *Proceedings of the National Academy of Sciences of the United States of America* 102 (2005): 1773–1778.
77. J. L. Naffin-Olivos, A. Daab, A. White, et al., "Structure Determination of *Mycobacterium tuberculosis* Serine Protease Hip1 (Rv2224c)," *Biochemistry* 56 (2017): 2304–2314.
78. N. Kolli and S. C. Garman, "Proteolytic Activation of Human Cathepsin A," *Journal of Biological Chemistry* 289 (2014): 11592–11600.
79. P. Goettig, M. Groll, J.-S. Kim, R. Huber, and H. Brandstetter, "Structures of the Tricorn-Interacting Aminopeptidase F1 With Different Ligands Explain Its Catalytic Mechanism," *EMBO Journal* 21 (2002): 5343–5352.
80. T. Yoshimoto, T. Kabashima, K. Uchikawa, et al., "Crystal Structure of Prolyl Aminopeptidase From *Serratia marcescens*," *Journal of Biochemistry* 126 (1999): 559–565.
81. C. B. García-Calderón, J. Casadesús, and F. Ramos-Morales, "Rcs and PhoPQ Regulatory Overlap in the Control of *Salmonella enterica* Virulence," *Journal of Bacteriology* 189 (2007): 6635–6644.
82. T. D. Ngo, B. H. Ryu, H. Ju, et al., "Structural and Functional Analyses of a Bacterial Homologue of Hormone-Sensitive Lipase From a Metagenomic Library," *Acta Crystallographica. Section D, Biological Crystallography* 69 (2013): 1726–1737.
83. D. D. Leipe and D. Landsman, "Histone Deacetylases, Acetoin Utilization Proteins and Acetylpolymine Amidohydrolases Are Members of an Ancient Protein Superfamily," *Nucleic Acids Research* 25 (1997): 3693–3697.
84. A. Dmochowska, D. Dignard, D. Henning, D. Y. Thomas, and H. Bussey, "Yeast KEX1 Gene Encodes a Putative Protease With a Carboxypeptidase B-Like Function Involved in Killer Toxin and Alpha-Factor Precursor Processing," *Cell* 50 (1987): 573–584.
85. M. G. Heiman, A. Engel, and P. Walter, "The Golgi-Resident Protease Kex2 Acts in Conjunction With Prm1 to Facilitate Cell Fusion During Yeast Mating," *Journal of Cell Biology* 176 (2007): 209–222.
86. H. R. Stennicke, U. H. Mortensen, and K. Breddam, "Studies on the Hydrolytic Properties of (Serine) Carboxypeptidase Y," *Biochemistry* 35 (1996): 7131–7141.
87. N. J. Galjart, H. Morreau, R. Willemsen, N. Gillemans, E. J. Bonten, and A. d'Azzo, "Human Lysosomal Protective Protein Has Cathepsin A-Like Activity Distinct From Its Protective Function," *Journal of Biological Chemistry* 266 (1991): 14754–14762.
88. C. Zevaco, V. Monnet, and J.-C. Gripon, "Intracellular X-Prolyl Dipeptidyl Peptidase From *Lactococcus lactis* spp. Lactis: Purification and Properties," *Journal of Applied Bacteriology* 68 (1990): 357–366.
89. J. M. Turner, N. A. Larsen, A. Basran, et al., "Biochemical Characterization and Structural Analysis of a Highly Proficient Cocaine Esterase," *Biochemistry* 41 (2002): 12297–12307.
90. M.-B. Maes, A.-M. Lambeir, K. Gilany, et al., "Kinetic Investigation of Human Dipeptidyl Peptidase II (DPPII)-Mediated Hydrolysis of Dipeptide Derivatives and Its Identification as Quiescent Cell Proline Dipeptidase (QPP)/dipeptidyl Peptidase 7 (DPP7)," *Biochemical Journal* 386 (2005): 315–324.
91. J. S. Rosenblum and J. W. Kozarich, "Prolyl Peptidases: A Serine Protease Subfamily With High Potential for Drug Discovery," *Current Opinion in Chemical Biology* 7 (2003): 496–504.
92. J. Zotzel, R. Pasternack, C. Pelzer, D. Ziegert, M. Mainusch, and H.-L. Fuchsbauer, "Activated Transglutaminase From *Streptomyces mobaraensis* Is Processed by a Tripeptidyl Aminopeptidase in the Final Step," *European Journal of Biochemistry* 270 (2003): 4149–4155.
93. J. M. Choe, D. Bakthavatsalam, J. E. Phillips, and R. H. Gomer, "Dictyostelium Cells Bind a Secreted Autocrine Factor That Represses Cell Proliferation," *BMC Biochemistry* 10 (2009): 4.
94. L. Shan, I. I. Mathews, and C. Khosla, "Structural and Mechanistic Analysis of Two Prolyl Endopeptidases: Role of Interdomain Dynamics in Catalysis and Specificity," *Proceedings of the National Academy of Sciences of the United States of America* 102 (2005): 3599–3604.
95. D. E. Petrenko, V. I. Timofeev, V. V. Britikov, et al., "First Crystal Structure of Bacterial Oligopeptidase B in an Intermediate State: The Roles of the Hinge Region Modification and Spermine," *Biology* 10 (2021): 1021.
96. J. R. Chekan, P. Estrada, P. S. Covello, and S. K. Nair, "Characterization of the Macrocyclase Involved in the Biosynthesis of RiPP Cyclic Peptides in Plants," *Proceedings of the National Academy of Sciences of the United States of America* 114 (2017): 6551–6556.
97. C. D. Fage, J. D. Hegemann, A. J. Nebel, et al., "Structure and Mechanism of the Sphingopyxin I Lasso Peptide Isopeptidase," *Angewandte Chemie, International Edition* 55 (2016): 12717–12721.
98. Y. Nakajima, K. Ito, T. Toshima, et al., "Dipeptidyl Aminopeptidase IV From *Stenotrophomonas maltophilia* Exhibits Activity Against a Substrate Containing a 4-Hydroxyproline Residue," *Journal of Bacteriology* 190 (2008): 7819–7829.
99. S. Brocca, C. Ferrari, A. Barbiroli, A. Pesce, M. Lotti, and M. Nardini, "A Bacterial Acyl Aminoacyl Peptidase Couples Flexibility and Stability as a Result of Cold Adaptation," *FEBS Journal* 283 (2016): 4310–4324.
100. L. Baugh, I. Phan, D. W. Begley, et al., "Increasing the Structural Coverage of Tuberculosis Drug Targets," *Tuberculosis* 95 (2015): 142–148.
101. J. Durairaj, A. M. Waterhouse, T. Mets, et al., "Uncovering New Families and Folds in the Natural Protein Universe," *Nature* 622 (2023): 646–653.
102. M. Steinegger, M. Meier, M. Mirdita, H. Vöhringer, S. J. Haunsberger, and J. Söding, "HH-suite3 for Fast Remote Homology Detection and Deep Protein Annotation," *BMC Bioinformatics* 20 (2019): 473.
103. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic Local Alignment Search Tool," *Journal of Molecular Biology* 215 (1990): 403–410.
104. P. Shannon, A. Markiel, O. Ozier, et al., "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks," *Genome Research* 13 (2003): 2498–2504.
105. M. Kucera, R. Isserlin, A. Arkhangorodsky, and G. D. Bader, "AutoAnnotate: A Cytoscape App for Summarizing Networks With Semantic Annotations," *F1000Res* 5 (2016): 1717.
106. K. Katoh and D. M. Standley, "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability," *Molecular Biology and Evolution* 30 (2013): 772–780.
107. S. R. Eddy, "Accelerated Profile HMM Searches," *PLoS Computational Biology* 7 (2011): e1002195.
108. S. F. Altschul, T. L. Madden, A. A. Schäffer, et al., "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs," *Nucleic Acids Research* 25 (1997): 3389–3402.
109. W. Li and A. Godzik, "Cd-Hit: A Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences," *Bioinformatics* 22 (2006): 1658–1659.
110. J. Hallgren, K. D. Tsigos, M. D. Pedersen, et al., "DeepTMHMM Predicts Alpha and Beta Transmembrane Proteins Using Deep Neural Networks," (2022), bioRxiv, <https://doi.org/10.1101/2022.04.08.487609>.

111. J. Jumper, R. Evans, A. Pritzel, et al., “Highly Accurate Protein Structure Prediction With AlphaFold,” *Nature* 596 (2021): 583–589.
112. N. Guex, “Swiss-PdbViewer: A Fast and Easy-To-Use PDB Viewer for Macintosh and PC,” *Protein Data Bank Quarterly Newsletter* 77 (1996): 7.
113. M. Gouy, S. Guindon, and O. Gascuel, “SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building,” *Molecular Biology and Evolution* 27 (2010): 221–224.

### Supporting Information

Additional supporting information can be found online in the Supporting Information section.