**The modular structure of α/β-hydrolases**

Tabea L. Bauer, Patrick C. F. Buchholz and Jürgen Pleiss

Institute of Biochemistry and Technical Biochemistry, University of Stuttgart, Germany

**Keywords**
Lipase Engineering Database; scale-free network; sequence–structure–function relationships

**Correspondence**
J. Pleiss, Institute of Biochemistry and Technical Biochemistry, University of Stuttgart, Allmandring 31, 70569 Stuttgart, Germany
Tel: +49 711 685-63191
E-mail: Juergen.Pleiss@itb.uni-stuttgart.de

(Received 25 June 2019, revised 15 August 2019, accepted 19 September 2019)
doi:10.1111/febs.15071

The α/β-hydrolase fold family is highly diverse in sequence, structure and biochemical function. To investigate the sequence–structure–function relationships, the Lipase Engineering Database (https://led.biocatnet.de) was updated. Overall, 280 638 protein sequences and 1557 protein structures were analysed. All α/β-hydrolases consist of the catalytically active core domain, but they might also contain additional structural modules, resulting in 12 different architectures: core domain only, additional lids at three different positions, three different caps, additional N- or C-terminal domains and combinations of N- and C-terminal domains with caps and lids respectively. In addition, the α/β-hydrolases were distinguished by their oxyanion hole signature (GX-, GGGX- and Y-types). The N-terminal domains show two different folds, the Rossmann fold or the β-propeller fold. The C-terminal domains show a β-sandwich fold. The N-terminal β-propeller domain and the C-terminal β-sandwich domain are structurally similar to carbohydrate-binding proteins such as lectins. The classification was applied to the newly discovered polyethylene terephthalate (PET)-degrading PETases and MHETases, which are core domain α/β-hydrolases of the GX- and the GGGX-type respectively.

**Introduction**

α/β-hydrolases represent a rapidly growing enzyme family with a common fold and a similar active site. The α/β-hydrolase fold consists of a central β-sheet packed between two layers of α-helices [1]. The catalytic triad consists of a nucleophile (serine, aspartate or cysteine), a histidine, a catalytic acid (aspartate or glutamate), and two or three amino acids which form the oxyanion hole. In addition to this conserved, catalytically active core domain, most α/β-hydrolases contain further modules: a lid, a cap, an N-terminal

**Abbreviations**
EH, epoxide hydrolase; GID1, gibberellin insensitive dwarf1; HMM, hidden Markov model; LED, Lipase Engineering Database; MHET, mono(2-hydroxyethyl) terephthalate; PDB, Protein Data Bank; PET, polyethylene terephthalate; rmsd, root-mean-square deviation.
domain, or a C-terminal domain. Despite the high structural similarity of the core domain and the catalytic machinery, \(\alpha/\beta\)-hydrolases show a high sequence diversity and include enzymes with a broad variety of catalytic activities such as acetylcholinesterases, acyltransferases, amidas, dehalogenases, dienelactone hydrolases, epoxide hydrolases (EH), esterases, hydroxynitrile lyases, lipases, peroxidases, proteases and thioesterases [2–5]. It would have been expected that enzymes with different catalytic activities are separated in sequence space and form separate subfamilies. However, the catalytic activities are overlapping rather than clearly separated in sequence space. Many \(\alpha/\beta\)-hydrolases have multiple catalytic activities, and single amino acid positions discriminate between activities such as lipase and amidase [5]. The exchange of single amino acids switched a lipase into an EH [6] or a catalyst for Michael additions [7], a hydrolase into an acyltransferase [8], an esterase into an amida [9,10], an esterase into an EH [11] and an esterase into a hydroxynitrile lyase [12].

Lipases and other \(\alpha/\beta\)-hydrolases are widely used in biocatalysis because of their high catalytic activity, their regio- and stereoselectivity, their stability in non-polar media and their ability to catalyse alcoholysis, esterification and transesterification reactions [13,14]. Metagenomic screenings have discovered an ever increasing stream of new sequences [15,16]. Therefore, there is a growing need for the classification of newly discovered sequences into protein families and for the prediction of biochemical and biophysical properties from sequence information [17]. Protein family databases on \(\alpha/\beta\)-hydrolases are versatile tools to systematically compare protein sequences, to annotate functionally relevant amino acids and to assign new sequences to already known subfamilies. Protein family databases such as the Lipase Engineering Database (LED) [18], the database of the \(\alpha/\beta\)-hydrolase fold superfamily ESTHER [19] and the \(\alpha/\beta\)-Hydrolase Fold 3DM Database ABHDB [20] provide a large collection of sequences and structures of \(\alpha/\beta\)-hydrolases, group them into protein families, perform multiple sequence alignments, and derive sequence profiles.

Although sequence alignment allows for a systematic analysis of close homologues, it misses structural and functional relationships between remote homologues. Therefore, a classification based on structural and functional properties is needed. The systematic comparison of the amino acids forming the oxyanion hole allowed to assign \(\alpha/\beta\)-hydrolases to two classes, the GX- and GGGX-types [18]. In GX-types, the first part of the oxyanion hole is formed by the backbone N-H of an amino acid (X) in a GX\(_\infty\) motif, whereas in GGGX-types it is formed by the backbone N-H of the third glycine in a GGGX-motif. Later, a third type of oxyanion hole (Y-type) was discovered [21]. In Y-type hydrolases, the oxyanion hole is formed by the side-chain of a bulky amino acid, mainly tyrosine or aspartate. The classification by GX- and GGGX-types has high predictive value: in contrast to GX-type, most of the GGGX-types are active toward tertiary alcohols [22]. A systematic comparison of the structures of EHs revealed the modular structure of this subfamily of \(\alpha/\beta\)-hydrolases with a high similarity within the core domains and within the caps [23]. Cytosolic and microsomal EHs mainly differ by the length of a single loop inside the cap and by their N-terminal domains. Because of their modular structure, the global sequence similarity is misleading, resulting in an over-estimation of diversity. Using this classification, the structures of core domains and caps of all EHs and the cap loops of most of the EHs could be reliably modelled [23]. Thus, modularity is key to sequence classification, structure modelling and prediction of function. This analysis was extended to the families of haloalkane dehalogenases and prolyl iminopeptidases, which are homologous to EH despite their different catalytic activities [24]. In the meantime, additional domains were identified in other hydrolases, and their functional role was discussed, such as the N-terminal \(\beta\)-propeller domain in acyl aminocarboxylases [25], the role of subdomains in the catalytic behaviour of lipases and acyltransferases [26], the stabilizing N-terminal domain in the murine liver EH (MLEH) [27] or the importance of the C-terminal domain in colipase binding of pancreatic lipases [28].

In this paper, we analysed the architectures of all known \(\alpha/\beta\)-hydrolase structures. The modular structure of \(\alpha/\beta\)-hydrolases can be described as a combination of three core domains (GX, GGGX and Y-types) with caps, N-terminal or C-terminal domains. In addition, a mobile element (the lid) is located at five different positions in lipases. For the analysis, the LED was updated and now contains 280 638 sequences and 1557 structures (https://led.biocatnet.de).

**Results**

**Lipase Engineering Database**

**Update of the LED**

The LED (https://led.biocatnet.de) was updated starting from the previous version (release 3.0, December 2009) which contained 24 783 sequences from 18 585 proteins and 1117 structures. All structures from this
release together with newly identified structures of MHETases [29] were clustered. The resulting centroids were then used as seed sequences (query sequences) for a BLAST search in the NCBI non-redundant protein database and in the Protein Data Bank (PDB), which led to the identification of more than 450 000 putative α/β-hydrolase sequences. One homologous family was created for every centroid sequence, and sequence homologues were added. The centroids of each homologous family and the unclassified structures were analysed visually, revealing a similar core structure which contains the active site, the oxyanion hole residues, and 10 additional modules that can be attached to the core domain: a lid between β-strands β_{-1} and β_{+2}, β_{-1} and β_{0}, β_{-4} and β_{-3}, β_{+3} and β_{+4}, or between the N terminus and β_{+3}, a single cap, a double cap, an N-terminal cap, two N-terminal domains or a C-terminal domain. These additional modules determine the architecture of the α/β-hydrolase and interestingly some of these elements, such as the lid between β_{-3} and β_{+4} and the lid between the N terminus and β_{-3}, only occur in combination with a C-terminal domain (Fig. 1). One superfamily was created for each architecture. The newly classified sequences were clustered with a similarity threshold of 60%, and new homologous families were created and assigned to their respective superfamilies. Superfamilies were named according to the architecture, whereas the homologous families were named according to their centroid sequence, and groups were formed by the oxyanion hole types (GX, GGGX and Y). Similarly, the superfamilies were assigned to five groups (core, lid, cap, one additional domain or two additional domains). The update resulted in a new version of the LED, which comprises 280 638 individual sequence entries assigned to 198 844 protein entries by a threshold of

Fig. 1. Sketches of the different architectures found in α/β-hydrolases. The core domain is displayed in grey with the N terminus on the left and the C terminus on the right. The position of the active site is marked by a yellow star. (A) Core domain α/β-hydrolases possess no additional modules. (B–D) Lids cover the active site and can emerge from different positions. Lids are displayed in green. (E–G) α/β-Hydrolases can have single caps, double caps or an N-terminal cap. The single cap and the bottom cap of the double cap protein are very similar and displayed in red. The cap of proteins with an N-terminal cap and the secondary N-terminal cap of double cap proteins are likewise similar and displayed in pale red. (H, I) Additional domains are displayed in blue and they can be attached to the N-terminal (H) or C-terminal (I) end of the protein. Other than most modules, they contain a lot of β-strands and are hence displayed as squares rather than ovals. (J, K) α/β-Hydrolases can also have a lid and an additional C-terminal domain. The lids can emerge from different positions in the central β-sheet, but always cover the active side from the C terminus of the protein. The additional C-terminal domains of these proteins are similar to the other C-terminal domains. (J, L) α/β-Hydrolases can also have a single cap and an additional N-terminal domain. The additional domain of these proteins is more similar to the core domain than to the other N-terminal domains, and hence is displayed in the same shape as the core.
98% sequence identity and contains 1557 protein structures in 2772 homologous families and 12 superfamilies. The largest superfamilies in the current version of the LED are the N-terminal domain superfamily (GX- and Y-types), containing 25% of the α/β-hydrolases, followed by the single cap superfamily (GX-, GGGX- and Y-types, 23%) the N-terminal cap family (GX- and GGGX-types, 21%), and the core domain family (GX- and GGGX-types, 13%). The smallest protein in the LED with known structure is a lipase of *Bacillus subtilis* (LED sequence ID 60, PDB-ID: 2QXT and 2QXU) with a length of 179 amino acids and a molecular weight of 19 kDa, the largest protein a dipeptidyl peptidase 8 (LED sequence ID 456274, PDB-ID: 6EOP) with a sequence length of 900 amino acids and a molecular weight of 104 kDa.

Of each homologous family, one protein with available structure was visually analysed, and the catalytic triad, the oxyanion hole residues and structural modules such as lids, single caps, double caps, N-terminal caps, N-terminal domains and C-terminal domains were annotated. The annotation information was then transferred from the centroids to the other members of a homologous family based on a multiple sequence alignment. In total, 162 737 sequences were annotated (58% of all sequences of the LED). Based on all annotation entries in the LED, the lengths of the additional modules and domains were analysed. Lids are the shortest modules with a length of 20 ± 7 amino acids (mean ± standard deviation), N-terminal caps and the N-terminal part of the double caps are similar in size (40 ± 8 and 53 ± 14 amino acids respectively), single caps are larger (75 ± 18 amino acids), and the additional C- and N-terminal domains are largest and vary considerably in size (367 ± 129 amino acids).

**Cluster size distribution**

Pairs of homologous sequences were derived for 280 638 sequence entries from the updated LED (release 4.0). The sequences were assigned to different clusters (communities), by setting a threshold of global sequence identity to 60%, 70%, 80% or 90%. The number of clusters \(N(s)\) with a cluster size \(s\) decreased with increasing \(s\) following a power-law distribution \(N(s) \sim s^{-\tau_h}\) (Fig. 2). Logarithmic histograms were formed for \(s \geq 2\) and \(s \leq 10\), \(s \geq 11\) and \(s \leq 100\), ..., \(s \geq 1001\) and \(s \leq 10000\) sequences [30]. \(\tau_h\) was determined from the slope of the histograms as 0.8, 1.0, 1.2 and 1.4 at a threshold of global sequence identity of 60%, 70%,
80% and 90%, respectively (Table 1), and the underlying \( \tau \) was modelled as described previously [30]. For comparison, \( \tau \) was also determined by linear regression of the actual distribution for \( s \leq 100 \) (Fig. S1), and both methods resulted in the same values of \( \tau \), ranging from 1.7 to 2.3 between 60% and 90% global sequence identity. \( \tau \) was then extrapolated to 100% global sequence identity, yielding \( \tau_{100} = 2.6 \) (Fig. S2), which is identical to the value derived previously for \( \alpha/\beta \)-hydrolases [30].

### Degree distribution

As further property of the network topology, the number of neighbouring sequences (degrees), was analysed. Neighbouring sequences were defined by a threshold of global sequence identity of 95%. The number \( N(n) \) of sequences having \( n \) neighbours followed a power law distribution \( N(n) \sim n^{-\gamma} \). Although many sequences had few neighbours, a small number of sequences had a high degree and were located in a hub region. Linear regression was performed for \( n \leq 80 \) resulting in a scaling exponent \( \gamma = 1.4 \) (Fig. 3), which is similar to the values of \( \gamma \) between 1.1 and 1.3 determined previously for five different protein families [31]. All sequences in the hub region, (\( n \geq 300 \)) belong to a single homologous family, the protease 2 homologous family number 1011, which are Y-type \( \alpha/\beta \)-hydrolases from the N-terminal domain superfamily number 8 (Table S1).

### Modular structure of \( \alpha/\beta \)-hydrolases

One representative \( \alpha/\beta \)-hydrolase structure for each architecture and oxyanion hole type was visually analysed. The structures of centroids were selected if they were fully resolved and had a resolution of at least 3 Å. Otherwise, a structure with high similarity to the centroid was selected. All \( \alpha/\beta \)-hydrolases have a similar core structure, the \( \alpha/\beta \)-hydrolase fold, which contains the catalytic triad and the oxyanion hole residues (Fig. 1). The \( \beta \)-strands of the central \( \beta \)-sheet were numbered starting from \( \beta_0 \), the \( \beta \)-strand preceding the nucleophilic elbow, where the catalytic serine is located.

All \( \beta \)-strands in the direction of the N terminus were assigned negative numbers (\( \ldots, \beta_{-3}, \beta_{-2}, \beta_{-1} \)), whereas the \( \beta \)-strands in the direction of the C terminus were assigned positive numbers (\( \beta_{1}, \beta_{2}, \beta_{3}, \ldots \)) (Figs S3–S9). As an additional module, \( \alpha/\beta \)-hydrolases might contain a mobile lid consisting of one or two \( \alpha \)-helices. The lid can only be unambiguously assigned if open and closed conformations of the same or two homologous proteins have been crystallized. Therefore, in the absence of both conformations, \( \alpha/\beta \)-hydrolases with a lid might be misclassified as core domain \( \alpha/\beta \)-hydrolases. The lid is located at different positions within the central \( \beta \)-sheet: between \( \beta \)-strands \( \beta_{1} \) and \( \beta_{2} \), \( \beta_{1} \) and \( \beta_{0} \), \( \beta_{-4} \) and \( \beta_{-3} \), \( \beta_{-3} \) and \( \beta_{-4} \), or between the N terminus of the protein and \( \beta_{-3} \). In contrast to the mobile lid, a cap is immobile and consists of more than two \( \alpha \)-helices which are stacked on top of the core structure. Three cap arrangements were found: a single cap located between \( \beta \)-strands \( \beta_{1} \) and \( \beta_{2} \), a double cap where an additional N-terminal cap is stacked on top of the single cap, and an architecture with an N-terminal cap only. In addition, \( \alpha/\beta \)-hydrolases might contain an N-terminal domain with a \( \beta \)-propeller fold or a Rossmann fold, or a C-terminal domain with a \( \beta \)-sandwich fold. The N-terminal Rossmann fold domain and C-terminal domains can be combined with a cap or a lid, respectively. Thus, in total 12 different architectures were identified (Fig. 1). In addition, \( \alpha/\beta \)-hydrolases can be distinguished by their oxyanion hole (Table 2). Ten of the 12 architectures contain GX-types, but only four architectures contain GGGX-types (core, core with lid

---

**Table 1.** Slopes of the cluster size distributions \( N(n) \) at different thresholds of global sequence identity, given as \( \tau \) for logarithmic histograms (Fig. 2) and as \( \tau \) fitted for distributions with \( s \leq 100 \) (Fig. S1).

| Threshold (%) | \( \tau_h \) | \( \tau (s \leq 100) \) |
|--------------|-------------|-----------------|
| 60           | 0.8         | 1.7             |
| 70           | 1.0         | 2.0             |
| 80           | 1.2         | 2.2             |
| 90           | 1.4         | 2.3             |

---

**Fig. 3.** Degree distribution. Degree distribution \( N(n) \) at a threshold of 95% global sequence identity with linear regression for degrees \( n \leq 80 \).
between \( \beta_{-3} \) and \( \beta_{-4} \), core with a single cap or with an N-terminal cap) and three architectures contain Y-types (core with a single cap, core with an N-terminal \( \beta \)-propeller domain or core with a C-terminal \( \beta \)-sandwich domain).

### Core domain \( \alpha/\beta \)-hydrolases

The core domain of \( \alpha/\beta \)-hydrolases is composed of up to 13 \( \beta \)-strands forming a central parallel \( \beta \)-sheet. The \( \beta \)-strands alternate with varying numbers of \( \alpha \)-helices, and the \( \beta \)-sheet is packed between two layers of \( \alpha \)-helices. The central \( \beta \)-strand \( \beta_0 \) is connected to the subsequent \( \alpha \)-helix by a sharp turn, the nucleophilic elbow, which harbours the catalytic serine and one of the oxyanion hole residues (Fig. S3). \( \alpha/\beta \)-Hydrolases consisting only of the core domain were found in the GX-types (29 712 sequences, representative structure: lipase A from \( B. subtilis \), BSLA) and in the GGGX-types (5376 sequences, representative structure: carboxylesterase from \( Lactobacillus plantarum \), LPC). As of now, no Y-type \( \alpha/\beta \)-hydrolases were found that consist exclusively of the core domain (Table 2).

The active site of core domain \( \alpha/\beta \)-hydrolases is completely exposed to the solvent (Fig. S10A). In contrast to GX-types, additional \( \alpha \)-helices restrict the substrate access to a specific direction in most GGGX-types, thus contributing to substrate specificity (Figs S11A and S12A), because GGGX-types are larger (200–500 amino acids, 8–13 \( \beta \)-strands) than GX-types (180–300 amino acids, 5–8 \( \beta \)-strands) and thus have more helices close to the substrate access tunnel that can shield the active site (Table S3).

### Lids as an opening and closing mechanism for the active site

A lid consists of one or two mobile \( \alpha \)-helices. It blocks the entrance to the active site in the closed conformation and allows substrate access in the open conformation (Figs 4 and 5). Interestingly, the position of the lid is not conserved. GX-type \( \alpha/\beta \)-hydrolases can have a lid between \( \beta_{+1} \) and \( \beta_{+2} \) (9610 sequences, representative structures: lipase from \( B. cepacia \) in open and in closed conformation, \( BCL_{\text{open}} \) and \( BGL_{\text{closed}} \) (Fig. S4A, B and Table 2). GGGX-type \( \alpha/\beta \)-hydrolases can have a lid between \( \beta_{-4} \) and \( \beta_{-3} \) (9079 sequences, representative structure: lipase from \( C. rugosa \) in open and closed conformation,

---

**Table 2.** Basic architectures of \( \alpha/\beta \)-hydrolases. Representative structures were chosen for the different oxyanion types and architectures. Abbreviations of protein names can be found in Table S2. In addition to the architecture, the position of the additional module in the central \( \beta \)-sheet is indicated. The letter in brackets following the protein name indicates the amino acid of the active site nucleophile.

| Basic architectures | GX-type | GGGX-type | Y-type |
|--------------------|---------|-----------|-------|
| Core               | BSLA (S)(PDB-ID: 116W) | LPC (S)(PDB-ID: 4BZW) | –     |
| Core + lid         | –       | –         | –     |
| \( \beta_{+1} \rightarrow \beta_{+2} \) | BCL\text{\textsubscript{open}} and BGL\text{\textsubscript{closed}} (S) (PDB-ID: 3LIP and 1TAH) | – | – |
| \( \beta_{-1} \rightarrow \beta_{0} \) | RML\text{\textsubscript{open}} and RML\text{\textsubscript{closed}} (S) (PDB-ID: 3TGL and 4TGL) | – | – |
| \( \beta_{-4} \rightarrow \beta_{-3} \) | – | CRL\text{\textsubscript{open}} and CRL\text{\textsubscript{closed}} (S) (PDB-ID: 1CRL and 1TRH) | – |
| Core + single cap \( \beta_{+1} \rightarrow \beta_{+2} \) | XAH (D)(PDB-ID: 1EDB) | MTH (S)(PDB-ID: 2VF2) | CALA (S)(PDB-ID: 2VEO) |
| Core + double cap \( \beta_{+1} \rightarrow \beta_{+2} \) | SCEH (D)(PDB-ID: 4QA9) | – | – |
| N terminus \( \rightarrow \beta_{+4} \) | TRC (S)(PDB-ID: 4PSG) | PCE (S)(PDB-ID: 3ZWQ) | – |
| Core + N-terminal cap | – | – | – |
| N terminus \( \rightarrow \beta_{+4} \) | – | – | – |
| Core + N-terminal \( \beta \)-propeller domain | APAP (S)(PDB-ID: 2HU5) | – | DPPV (S)(PDB-ID: 1NU8) |
| Core + C-terminal \( \beta \)-sandwich domain | – | RCE (S)(PDB-ID: 1DU4) | – |
| Core + lid \( \beta_{+1} \rightarrow \beta_{+4} \) + C-terminal \( \beta \)-sandwich domain | HPL\text{\textsubscript{open}} and HPLRP\text{\textsubscript{closed}} (S) (PDB-ID: 1LPB and 2PPL) | – | – |
| Core + lid \( \beta_{+1} \rightarrow \beta_{+4} \) + C-terminal \( \beta \)-sandwich domain | PSML\text{\textsubscript{open}} and PSML\text{\textsubscript{closed}} (S) (PDB-ID: 2Z8X and 2ZVD) | – | – |
| Core + cap \( \beta_{+1} \rightarrow \beta_{+2} \) + \( \beta \)-sandwich domain | MLEH (D)(PDB-ID: 1CQZ) | – | – |

---

1040 The FEBS Journal 287 (2020) 1035–1053 © 2019 The Authors. The FEBS Journal published by John Wiley & Sons Ltd on behalf of Federation of European Biochemical Societies.
No Y-type α/β-hydrolases with a lid has been found, yet. Regardless of where the lid emerges, its position in respect to the active site is almost identical. In its closed conformation, it covers the active site from the N-terminal side of the protein and moves towards the N terminus of the protein upon opening (Figs 1B–D, S10 and S11B–D). When the lid is in the closed conformation, the active site residues are not visible and the contact area of the lid surrounds the active site completely. This suggests that the lid covers the active site in the closed conformation, making the active site inaccessible to solvent or substrate molecules. Upon opening of the lid, the contact area of the lid moves away from the active site, thus making the active site fully accessible to the solvent and thereby allowing substrate molecules to bind to the active site (Figs 4 and 5).

Caps covering the active site
Caps are immobile elements, which cover the active site either partly or completely. α/β-Hydrolases can have single caps, N-terminal caps or double caps added to the core domain (Fig. 1E–G). Single caps consist of three or more α-helices on top of the proteins (Figs S5A and S11E). N-terminal caps consist of two or more long α-helices formed by the N terminus folding back over the active site (Figs S5B and S11F). In α/β-hydrolases with a double cap, two caps are stacked on top of the active site. The lower cap is similar to the single cap, the upper cap to the N-terminal cap (Figs S5C and S11G).

α/β-Hydrolases with a single cap were found as GX-types (61 596 sequences, representative structure: haloalkane dehalogenase from Xanthobacter autotrophicus, XAHD), GGGX-types (2581 sequences, representative structure: HsaD from Mycobacterium tuberculosis, MTH), and Y-types (22 sequences, lipase A from Candida antarctica, CALA) (Table 2), and is the only architecture that can be found for all oxyanion hole types. The position of the single cap is conserved and is always located between strand β₁⁺ and β₁⁻ (Fig. S5A). The active site of Y-type α/β-hydrolases with a single cap is completely covered by the cap, thereby restricting

![Image of lids opening and closing in GX-types](image-url)
access to the active site. In GX-types, the active site is mostly covered, however, the catalytic triad histidine is partially exposed to the solvent. In GGGX-types, the entrance to the active site from the top is completely blocked by the cap, but there is a narrow tunnel between the cap and the core domain which leads to the active site completely without leaving a tunnel between the caps (Figs. S10G, S12C and S13B).

In N-terminal caps, the helices restrict the access of solvent or substrate molecules to the active site (Figs 6B, S10F and S122B). \(\alpha/\beta\)-Hydrolases with an N-terminal cap could be found as GX-types (585 sequences, representative structure: cutinase from Trichoderma reesei, TRC) and as GGGX-types (57 900 sequences, representative structure: cocaine esterase from Aeropyrum pernix K1, APAP) and Y-types (53 669 sequences, representative structure: human dipeptidyl peptidase IV [DPPIV]) (Table 2). The \(\beta\)-propeller domain consists of seven to eight blades, each blade is formed by a four-stranded \(\beta\)-sheet. (Figs 7A and S6). The \(\beta\)-propeller domain is stacked on top of the core domain, and the blades are arranged to form a channel through the middle of the propeller towards the active site (Fig. S10I). In Y-types, the propeller domain does not block the active site completely, suggesting there might be an additional substrate access between the propeller and the core domains (Figs 7A and S13C).

\(\alpha/\beta\)-Hydrolases with an additional C-terminal \(\beta\)-sandwich domain were found in the Y-types only (13 624 sequences, representative structure: cocaine esterase from Rhodococcus sp. MB1, RCE) and not in GX-types or GGGX-types (Figs 7B, S7 and Table 2). In the CATH database, the C-terminal \(\beta\)-sandwich domain is annotated as galactose-binding domain-like, CATH superfamily 2.60.120.260 [32]. The C-terminal \(\beta\)-sandwich domain does not interfere with the active site and is located at the side of the protein. Therefore, \(\alpha/\beta\)-hydrolases with a C-terminal \(\beta\)-sandwich domain have a freely accessible active site (Figs 7B, S10H and S13D).

Proteins with two additional modules

\(\alpha/\beta\)-Hydrolases with more than one additional module are rare and only found for GX-types (Fig. 1J,L): \(\alpha/\beta\)-hydrolases with a C-terminal domain and a lid.
The C-terminal domains of HPL/HPLRP1 and PSML are β-sandwich domains (Figs 8A and S11H). The β-sandwich domains of HPL and HPLRP1 are smaller than that of RCE and are annotated as a PLAT/LH2 domain in the CATH database (CATH Superfamily 2.60.60.20) [32]. The β-sandwich domain of PSML is similar in size to RCE and annotated as a C-terminal serralysin-like metalloprotease domain in the CATH database (CATH Superfamily 2.150.10.10) [32]. The C-terminal β-sandwich domains of HPL/HPLRP1 and PSML are attached to the side of the protein and therefore do not interfere with the accessibility of the active site (Figs 8A and S13E). The lids of HPL/HPLRP1 and PSML consist of a single α-helix (Figs 8A and S11H) and were found in an open (HPL_open, PSML_open) or a closed conformation (HPLRP1_closed, PSML_closed) (Figs S10J,K and S12E, F). In the closed conformation, the active site is blocked. Upon opening, the lid helix moves away thereby uncovering the active site (Figs 8A and S13E). Although the opening/closing transition and the size of the lids are similar, the location of the lid differs. In HPL/HPLRP1, it is located between strands β_3 and β_4, whereas in PSML it is located between the N terminus and β_3 (Fig. S8A,B). Interestingly, although the lids emerge from different locations within the central β-sheet, they are situated in a similar position covering the active site from the C terminus of the protein, rather than from the N terminus as observed for other α/β-hydrolases (Fig. 1).

The N-terminal domain of MLEH consists of a Rossmann fold formed by alternating α-helices and β-strands, annotated by CATH classification as HAD superfamily/HAD-like domain, or CATH Superfamily 3.40.50.1000 [32]. Other than the N-terminal β-propeller

Fig. 6. Caps cover the active site and thereby restrict substrate access. (A) Surface view of α/β-hydrolases with an additional single cap. GX-types are represented by XAHD, GGGX-types by MTH and Y-types by CALA. (B) Structure of a GGGX-type α/β-hydrolase with an additional N-terminal cap. GGGX-types with an N-terminal cap are represented by PCE. The top panel shows the surface view of the whole protein. In the bottom panel, proteins were disassembled in their subdomains. Both upper and lower panel show the N terminus on the left and the C terminus on the right. In the lower panel of B, the structure was turned 180° so that the N terminus is shown on the right and the C terminus on the left. The side chain of the active site serine, aspartate and histidine, as well as the backbone of the oxyanion hole residues are coloured yellow. The cap is coloured in red, the contact area of the cap and the core domain in brown, and the rest of the protein in grey.
domains and similar to the C-terminal β-sandwich domains, this N-terminal Rossmann fold domain is attached to the side of the protein and does not interfere with substrate access to the active site (Figs 8B and S12D). The cap of MLEH is comparable to other single caps and is located between β+1 and β+2 (Figs 8B and S9). Although it covers the active site, there seems to be a tunnel underneath the cap, which allows access of substrates to the active site (Figs 8B and S10L).

**Similarity of the N- and C-terminal domains to other proteins**

In order to analyse whether the N- and C-terminal domains in superfamilies 8–12 are similar to domains in other proteins, one representative protein from each of these superfamilies and each oxyanion hole type was selected: from the N-terminal domain superfamily number 8, the GX-type *A. pernix* acylaminoacyl peptidase (APAP) and the Y-type human DPPIV, from the C-terminal domain superfamily number 9, the Y-type *Rhodococcus* sp. cocaine esterase (RCE), from the lid (β+3 and β+4) and C-terminal domain superfamily number 10, the GX-type HPL, from the lid (N-terminal) and C-terminal domain superfamily number 11, the GX-type *Pseudomonas* sp. MIS38 lipase (PSML), and from the cap and N-terminal domain superfamily number 12, the GX-type human epoxide hydrolase (HEP) (Table S4).

The N-terminal β-propeller domains of DPPIV and APAP are structurally similar, as well as the β-sandwich domains of RCE and HPL. All α/β-hydrolase core domains are similar to each other. In addition, they are similar to the additional N-terminal Rossmann fold domain of HEP (Fig. S14).

In a second step, the PDB was searched for domains with structural similarity to the N- or C-terminal domains (https://doi.org/10.18419/darus-458).

The search with the β-propeller domains of DPPIV and APAP resulted in 103 and 173 structurally similar proteins, respectively, which consist of a single propeller, two fused propellers, or fusion proteins of a β-propeller and other domains. They cover a broad range of biological functions, such as transcription and translation initiation factors, DNA damage-binding proteins, elongation complexes, export factors, splicing factors, ribosomal proteins, proteins involved in cell cycle, apoptosis, and intracellular transport, and also β-propeller lectins, although the carbohydrate-binding site does not seem to be conserved.

---

**Fig. 7.** Additional N-terminal β-propeller domains cover the active site, whereas C-terminal β-sandwich domains are attached to the side of the protein. Structure and surface view of α/β-hydrolases with an additional domain. Both panels show the N terminus on the left and the C terminus on the right. (A) α/β-Hydrolases with an additional N-terminal domain are represented by the GX-type APAP and the Y-type DPPIV. (B) α/β-Hydrolases with an additional C-terminal domain are represented by the Y-type RCE. The side chain of the active site serine and histidine, as well as the backbone of the oxyanion hole residues are coloured yellow with a yellow sphere through the oxygen of the OH group of the active site serine. The domains are coloured in blue, the contact area of the domain with the core domain in dark blue, and the rest of the protein in grey. Red arrows show possible entrance sites for the substrate to the active site.
The search with the Rossmann fold domain of HEP resulted in 154 hits of structurally similar proteins. Because of its similarity to the α/β-hydrolase core domain, 54% of all hits belonged to α/β-hydrolases, mostly single cap α/β-hydrolases from LED superfamily 5. All other hits were mainly proteins consisting of a single Rossmann fold domain, two fused Rossmann fold domains or fusion proteins of Rossmann fold with other domains. They cover a broad range of enzymatic functions, such as phosphatases, haloacid dehalogenases, phosphoglucomutases, reductases and glycosyltransferases, but also regulators of transcription or replication.

The search with the β-sandwich domains of RCE, PSML and HPL resulted in 41, 35 and 10 proteins, respectively, which consist of a single β-sandwich or are fusion proteins of β-sandwich domains with mainly α-helical domains. Proteins with structural similarity to HPL cover a broad range of annotations such as lipoxigenases, glucanases, xylanases, the GTPase Rab6 and carbohydrate-binding modules. Proteins with structural similarity to RCE are involved in DNA repair, flagellar biosynthesis or receptor binding, but also lectins and enzymes involved in carbohydrate degradation, such as xylanases, glycoside hydrolases, mannanases and endoglucanases. Although the structural similarity suggests a carbohydrate-binding function, the carbohydrate-binding site of lectins is not present in the β-sandwich domains of RCE and HPL. Proteins with structural similarity to PSML include hemagglutinin, adhesins, toxin A and antifreeze proteins.

**Discussion**

**Naming conventions**

Despite their high variability in sequence and structure, α/β-hydrolases can be assigned to a small number of superfamily groups, based on the characteristic Rossmann fold core domain. The Rossmann fold is a β-sheet sandwich structure with multiple catalytic loops, and its recognition is crucial for understanding the functional diversity of these enzymes.
of classes. Two criteria were used: the sequence motif of the oxyanion hole and the presence of lids, caps and N- or C-terminal domains. However, there is no consistent naming of these structural modules in literature, especially for the definition of lids and caps. The definition of a ‘cap’ as a large immobile module covering the active site is widely accepted [33–36]. However, in some publications it is referred to as ‘lid’ [2,37,38]. Similarly, small mobile elements covering the active site are mostly referred to as ‘lid’ [35,39–43], but they might also be called ‘flap’ [44–47]. To improve communication about α/β-hydrolases, we suggest the following naming conventions:

A ‘lid’ is a small mobile structure comprised of one or two α-helices, which can undergo a conformational transition, thereby opening and closing the entrance to the active site. Lids are located between β-strands β₁⁻₃ and β₄₋₅, β₀ and β₁⁻₂, β₁⁻₃ and β₄₋₅ or between the N-terminus and β₅⁻₆.

A ‘cap’ is an immobile module and consists of three or more α-helices, which cover the active site. Caps are located between strands β₁⁻₂ and β₅₋₆. In proteins with double caps, a second cap consisting of two or three α-helices is located at the N terminus. This second cap is similar to N-terminal caps. These are formed by two or more α-helices emerging from the N terminus of the protein, which fold over the active site and thereby cover it.

N-terminal domains have either a β-propeller fold or a Rossmann fold. The C-terminal domains have a β-sandwich fold. These additional domains can be located on top of the core domain covering the active site, on its side or separated by a long loop.

Although caps and the N-terminal or C-terminal domains can be easily identified in protein structures, the unambiguous identification of one or two α-helices as lid requires the presence of an open and a closed conformation, such as in C. rugosa lipase (PDB-ID: 1CRL and 1TRH), Rhizomucor miehei triacylglyceride lipase (PDB-ID: 3TGL and 4TGL), or Pseudomonas sp. MIS38 lipase (PDB-ID: 2Z8X and 2ZVD).

The suggested classification by oxyanion hole motif and by the presence of lids, caps and N- or C-terminal domains is based on the analysis of protein structures. However, a few putative α/β-hydrolases with known structure could not be assigned because of deviations of their oxyanion hole sequence from the conserved motifs, which did not allow us to assign them to GX-, GGGX- or Y-types: The N-acyl homoserine lactone degrading enzyme (PDB-ID: 5EGN) and a hydrolase from Pseudomonas aeruginosa PA01 (PDB-ID: 3OM8) had a cap, but their oxyanion hole motifs were PF and SI respectively. Similarly, the putative dienelactone hydrolase from Klebsiella pneumoniae (PDB-ID: 3F67) and an uncharacterized protein from Escherichia coli (PDB-ID: 4ZV9) are core domain α/β-hydrolases, however, their oxyanion hole motif was EX.

In addition, the database update resulted in ≈170,000 discarded sequences which could not be assigned to one of the superfamilies because of their low sequence similarity to proteins with known structure, which made it impossible to decide about the presence of modules.

Substrate access

The knowledge of the structural elements of a protein allows a prediction about the access of substrate to the active site. For α/β-hydrolases that consist only of the core domain, the active site is fully exposed to the solvent. This was already shown for the lipase from B. subtilis, the cutinase from Fusarium solani or the carboxylesterase from Pseudomonas fluorescens [48–50]. In α/β-hydrolases with a lid, the substrate enters between core and lid. In the closed conformation, the active site is covered by the lid, but becomes accessible to the substrate upon a conformational transition of the lid to an open conformation, as shown for the lipases from C. rugosa, P. aeruginosa or Rhizopus niveus [51–53]. In contrast, the cap covers the active site. In order to access to the active site, a substrate molecule has to pass through a tunnel located between the core and the cap, such as in the HEP, the prolyl aminopeptidase from Serratia marcescens, or the C-C hydrolase from M. tuberculosis [54–56], or through a tunnel in the cap, such as in the fluoroacetate dehalogenase from Rhodopsseudomonas palustris or the hydroxynitrile lyase from Hevea brasiliensis [57,58]. Access to the tunnel is controlled by gatekeepers such as the sidechain of Leu262 in the haloalkane dehalogenase from X. autotrophicus [59]. Engineering of the tunnel can result in changes of substrate specificity [60]. Interestingly, in GGGX-type proteins with a single cap or an N-terminal cap, the active site is partially solvent accessible, which, however, might be a crystallization artefact. For Y-type α/β-hydrolases with an N-terminal β-propeller domain, substrate access is under discussion. For dipeptidyl peptidases, the substrate was shown to enter the active site through a side opening, along the interface between N-terminal and core domain [61,62]. The function of the tunnel in these proteins is not clear, but was suggested to aid the release of product from the active site [63,64]. For prolyl oligopeptidases, which are also Y-type α/β-hydrolases with an N-terminal β-propeller domain, a
crystal structure with a tilted propeller domain was identified, resulting in a large opening between the core domain and the propeller domain, similar to dipeptidyl peptidases [37]. It has therefore been discussed whether substrate molecules enter prolyl oligopeptidases through the tunnel of the propeller domain [65,66] or through the entry site that opens upon conformational change of the proteins [37,67,68].

**Sequence space**

Despite the large number of 280 638 sequence entries in the updated LED, we still know only a tiny fraction of the whole extant sequence space of α/β-hydrolases [69]. The limited coverage of sequence space and the even lower coverage of structure space might explain why GX-types were found in combination with most modules and multiple lid positions, but only few combinations and lid positions were found for GGGX- and Y-types.

However, the analysis of the sequence network confirmed previous results on the properties of protein sequence space. The scale-free distribution of the number of neighbors of a protein sequence demonstrates the existence of a few hubs and a large number of loosely connected sequences, as found previously for five different protein families. The scaling exponent of the α/β-hydrolase sequence network ($\gamma = 1.4$) is similar to other protein families which had scaling exponents between 1.1 and 1.3 [31]. The members of the homologous family 1011 (annotated as ‘protease 2’ or ‘oligopeptidase B’) had the highest number of neighbors (Table S1) and thus formed the largest hub region of the α/β-hydrolase network. Since hub sequences have many functional neighbors, they have proven to be highly evolvable with respect to robustness towards mutations [70]. Because mutations might readily induce new functions, sequences with a large number of neighbors are promising starting points in directed evolution experiments.

A second property of the α/β-hydrolase network, the scale-free cluster size distribution, is also similar to other protein families [30]. The extrapolated scaling exponent ($\tau_{100} = 2.6$) was in the range between 2.3 and 3.3 which was previously determined for six protein families [30] and indicates percolation, that is, connectedness, of protein sequence space, despite the fact that extant sequence space covers only a tiny fraction of theoretic sequence space, yet. The connectedness of protein sequence space implies that various evolutionary pathways between two homologous sequences are possible, favoring substrate ambiguity or promiscuity [71–73].

**Structure space**

Although global sequence similarity is a widely used measure of the relationship between proteins, it is misleading. In addition to exchanges, insertions and deletions of single amino acids, there is a second evolutionary mechanism: the recombination of structural modules or domains, which has been observed in many protein families, such as thiamine diphosphate-dependent enzymes [74] or glycoside hydrolases and carbohydrate-binding modules [75]. Although the identification of structural and functional domains and modules in a protein structure might be challenging [76], many proteins which are distant considering their global sequence and structure share highly similar fragments, which resulted in the view that protein structure space is continuous rather than discrete [77].

The sequence similarity between α/β-hydrolases is generally low and their similarity relies on their structure [3]. The similarity of domains can be hidden by the global similarity, especially if different modules are recombined and exist in different orders [74]. Although the global structural similarity is quite low or undetectable between some α/β-hydrolases with additional C- and N-terminal domains, the structural similarity of the core domains to each other is remarkable. This is different for the additional modules and most of them are structurally quite diverse. However, there are exceptions, such as the β-propeller domain of Y-type and GX-type α/β-hydrolases. These are structurally very similar and so are the core domains, pointing to co-evolution. Interestingly, also the core domain and the additional C-terminal β-sandwich domain of HPL and RCE show significant structural similarity. Interestingly, RCE is a Y-type with an additional C-terminal β-sandwich domain, whereas HPL belongs to the GX-types and has an additional C-terminal β-sandwich domain and a lid attached. This suggests that existing modules can be exchanged or newly added to proteins. In fact, the addition or deletion of N- and C-terminal domains to proteins was described as one of the most frequent domain rearrangements observed [78]. The substitution of domains inside a protein was, however, shown to be rare [79], which raises the questions, whether lids and caps can also be exchanged between proteins. Interestingly, the fold of the additional N- and C-terminal domains also occurs in several other proteins with very different functions. Although the sequence similarity between these modules is very low, they share structural similarity, showing how evolution reuses suitable folds for completely new functions.

Continuous protein space is supported by another very interesting finding in the LED that proteins could...
clearly be assigned to one of the homologous families although they are enzymatically inactive and display no hydrolase activity. One such example is the Gibberellin receptor gibberellin insensitive dwarf1 (GID1) (PDB-ID: 2ZSI and 3ED1). This protein can clearly be assigned to GGGX-types with an N-terminal cap, but they miss the catalytic histidine. This is interesting, because the histidine is the most conserved residue of the catalytic triad of \( \alpha/\beta \)-hydrolases [3]. Indeed, it was shown that although GID1 is similar to hormone-sensitive lipases, it does not display any hydrolytic activity [80–82]. Another interesting example are neurogins, for example, the extracellular domain of neuroligin 2A from mouse (PDB-ID: 3BL8) or neuroligin-1 from rat (PDB-ID: 3BIW). Both can clearly be assigned to the GGGX-type \( \alpha/\beta \)-hydrolases with an additional lid, however, they are missing the catalytic serine and instead show a GXGXG motif, which leads to an enzymatically inactive protein [83,84]. This demonstrates how the structure of \( \alpha/\beta \)-hydrolases can be exploited and used for other functions.

A few years ago, plastic-degrading enzymes were discovered as an interesting possibility for the bioconversion of polyethylene terephthalate (PET) [85]. These so-called PETases are able to convert PET to mono-(2-hydroxyethyl) terephthalate (MHET), which is then hydrolysed by MHETases to terephthalate and ethylene glycol [29]. PETases and MHETases are members of the \( \alpha/\beta \)-hydrolase family [29,85] and can be found in the LED. PETases are GX-types, whereas MHETases are GGGX-types, but both are core domain \( \alpha/\beta \)-hydrolases. MHETases were suggested to have a lid that confers substrate specificity [29], however, the domain specified by Palm and colleagues only shields, but does not cover the active site. Furthermore, it consists of 8 \( \alpha \)-helices and 242 amino acids and thus is too big to be classified as a lid according to the naming convention suggested in this paper. In the LED, PETases and MHETases are found in superfamiliy number 1. There are 1099 PETases and homologues thereof combined in homologous family number 49 and 13 MHETases and homologues thereof in homologous family number 2923. Therefore, the LED can aid in identifying possible PETases and MHETases. It further allows to find related proteins that could be used in protein engineering to develop highly active plastic-degrading enzymes.

**Materials and methods**

**Setup of the Lipase Engineering Database (LED)**

The sequences of 1117 \( \alpha/\beta \)-hydrolases with structure information in the previous version of the LED (release 3.0, December 2009) were clustered using USEARCH (version 11.0.667) with an identity threshold of 90%, resulting in a list of representative sequences (named centroids) [86]. The structures of the centroids were checked visually for their architecture using the PyMOL Molecular Graphics System (version 1.8.0.5, Schrödinger; LLC, New York, NY, USA). Structures lacking the \( \alpha/\beta \)-hydrolase fold were discarded, and the remaining centroid sequences were reclustered by USEARCH with an identity threshold of 60% to reduce the number of queries for later BLAST searches. In addition, the sequences of three recently identified structures of MHETases, which also belong to the family of \( \alpha/\beta \)-hydrolases, were added to the centroids [29]. Each centroid served as a seed sequence of a homologous family and was assigned to a superfamily [87], which was named by its respective architecture. By using the centroid sequences as query, homologous proteins were searched in the NCBI non-redundant protein database and the PDB by BLAST searches with an E-value cut-off of \( 10^{-10} \) [88,89], resulting in more than 450 000 putative \( \alpha/\beta \)-hydrolase sequences. A threshold of 98% global sequence identity was used to assign individual sequences to proteins. Sequences were sorted into homologous families at a global sequence similarity threshold of 60%. For each of the homologous families, a profile hidden Markov model (HMM) was created using HMMER 3.1b2 (http://hmmer.org/). Iterative HMM searches with these profiles in the remaining unassigned sequences were performed with decreasing E-values from \( 1.5 \times 10^{-5} \) to \( 5 \times 10^{-10} \), and the hits were assigned to the respective superfamily after a visual inspection of their architecture. The unassigned sequences were excluded from further analyses. The newly assigned sequences were clustered with an identity threshold of 60% using USEARCH. For each cluster containing more than nine sequences, a new homologous family was created in the respective superfamily. All sequences in a superfamily, which were not assigned to a homologous family, were summarized in the 'singleton family'.

**Visual analysis of \( \alpha/\beta \)-hydrolase structures**

The PyMOL Molecular Graphics System, (version 1.8.0.5, Schrödinger; LLC) was used for the visual analysis of \( \alpha/\beta \)-hydrolases and the generation of figures. Contact areas of the modules with the core domains were calculated using a PyMOL script by Martin Christen, in which the contact radius was increased to 6 Å (Martin Christen, 2013, Contact Surface, v.3.0, https://pymolwiki.org/index.php/ContactSurface). Explosion graphs were created by disassembling the structures and pulling the modules (lids, caps, N- or C-terminal domains) away the from the core protein.

**Analysis of protein sequence networks: clusters and degrees**

Pairwise global sequence alignments for the 280 638 sequence entries from the updated LED were used to derive
edge weights of pairwise sequence identity for protein sequence networks. Instead of aligning all sequence pairs, the calc_distmx command from USEARCH (version 11.0.667) was used to heuristically determine pairs of homologues, thereby reducing computational effort [86]. Subsequently, the sequence pairs were aligned using the Needleman-Wunsch algorithm implemented in the EMBOSS software suite (version 6.6.0, EMBL-EBI, Hinxton, UK) with gap opening and gap extension penalties of 10 and 0.5 respectively [90,91].

Protein sequence networks were constructed as undirected graphs with edge weights of global sequence identity. Thresholds of global sequence identity were applied to form clusters (communities) of homologous sequences. The number of nodes \( N(s) \) of a cluster size \( s \) was fitted by a power law \( N(s) \sim s^{-\tau} \), and the Fisher exponent \( \tau \) was determined from the slope of the distribution in a log-log plot [30]. Logarithmic histograms were formed for \( s \geq 2 \) and \( s \leq 10, s \geq 11 \) and \( s \leq 100, \ldots, s \geq 1001 \) and \( s \leq 10 \, 000 \) sequences. The slopes of these histograms, \( \tau_0 \), were determined at varying thresholds of global sequence identity. The slopes of the actual distribution, \( \tau \), were derived by fits of \( \tau_0 \) against a power-law model distribution as described previously [30]. The number of neighbouring nodes \( n \) at a given threshold is called the degree of a node. The number of nodes \( N(n) \) having a degree of \( n \) was fitted by a power law \( N(n) \sim n^{-\gamma} \), and the scaling exponent \( \gamma \) was derived from a log-log plot [31]. Distributions of cluster sizes and degrees were analysed by linear fitting via the fitlm function from the Statistics and Machine Learning Toolbox (version 11.5) in MATLAB (version R2019a The MathWorks, Natick, MA, USA).

Annotation of protein sequences

If available, a representative protein with structure information was selected from each homologous family, and the catalytic nucleophile (serine, aspartate or cysteine), histidine and the catalytic acid (aspartate or glutamate) were annotated, as well as the residues forming the oxyanion hole, the lid, single cap, secondary cap (N-terminal part of double caps), N-terminal cap, N-terminal domain and C-terminal domain. For each homologous family, a multiple sequence alignment was generated using Clustal Omega (version 1.2.1) [92]. The annotations were then transferred to the respective positions in the aligned sequences. Thus, the sequences of all homologous families that contain at least one protein structure were annotated in the updated release of the LED (release 4.0).

Structural similarities of the N- and C-terminal domains

To analyse the structural similarities of the N- and C-terminal domains, representative protein structures of each oxyanion hole type from the proteins of superfamilies 8–12 were selected. The structures of the N- and C-terminal domains and the core domains were extracted and saved in separate.pdb files, which were then used to compare the domains by the ‘all against all’ structure comparison tool on the Dali server [93]. In a second step, the PDB search tool on the Dali server was used to compare the N- and C-terminal domains against all structures in PDB25, a subset resulting from clustering the whole PDB database with an identity threshold of 25%. The Dali server has different measures to determine similarity between proteins. The Z-Score is a measure for structural similarity, where a Z-score \( > 2 \) implies significant structural similarities and thus a similar fold [94]. Besides the Z-Score, the output also contains information about the rmsd and the sequence identity.

Acknowledgements

The authors acknowledge Catharina Zeil for kindly providing a workflow to setup the database and scripts for the protein annotation. Furthermore, the authors acknowledge the Deutsche Forschungsgemeinschaft (grants PL145/16-1 and EXC2075) and Bundesministerium für Bildung und Forschung (grant 031B0571A) for funding.

Author contributions

TLB performed the experiments, analysed data and contributed to writing the manuscript. PCFB contributed methods and resources, performed analyses and contributed to writing the manuscript. JP planned the experiments, supervised the project and wrote the manuscript.

Conflict of interest

The authors declare no conflict of interest.

References

1 Ollis DL, Cheah E, Cygler M, Dijkstra B, Frolow F, Franken SM, Harel M, Remington SJ, Silman I, Schrag J et al. (1992) The \( \alpha/\beta \) hydrolase fold. Protein Eng Des Sel 5, 197–211.
2 Mindrebo JT, Narrey CM, Seto Y, Burkart MD & Noel JP (2016) Unveiling the functional diversity of the alpha/beta hydrolase superfamily in the plant kingdom. Curr Opin Struct Biol 41, 233–246.
3 Nardini M & Dijkstra BW (1999) Alpha/beta hydrolase fold enzymes: the family keeps growing. Curr Opin Struct Biol 9, 732–737.
4 Vijayakumar A & Rajasekharan R (2017) Distinct roles of alpha/beta hydrolase domain containing proteins. Biochem Mol Biol J 02, 1–3.
5 Suplatov DA, Besenmatter W, Svedas VK & Svendsen A (2012) Bioinformatic analysis of alpha/beta-hydrolase fold enzymes reveals subfamily-specific positions responsible for discrimination of amidase and lipase activities. Proc Natl Acad Sci USA 109, 17988–17993.

6 Svedendahl M, Carlqvist P, Branneby C, Allnér O, Frise A, Hult K, Berglund P & Brinck T (2008) Direct epoxidation in Candida antarctica Lipase B studied by experiment and theory. ChemBioChem 9, 2443–2451.

7 Svedendahl M, Hult K & Berglund P (2005) Fast carbon–carbon bond formation by a promiscuous lipase. J Am Chem Soc 127, 17988–17993.

8 Larsen MW, Ziellinska DF, Martinelle M, Hidalgo A, Jensen LJ, Bornscheuer UT & Hult K (2010) Suppression of water as a nucleophile in Candida antarctica lipase B catalysis. ChemBioChem 11, 796–801.

9 Kourist R, Bartsch S, Fransson L, Hult K & Bornscheuer UT (2008) Understanding promiscuous amidase activity of an esterase from Bacillus subtilis. ChemBioChem 9, 67–69.

10 Syrén P-O, Hendil-Forsell P, Aumailley L, Besenmatter W, Gounine F, Svendsen A, Martinelle M & Hult K (2012) Esterases with an introduced amidase-like hydrogen bond in the transition state have increased amidase specificity. ChemBioChem 13, 645–648.

11 Jochens H, Stiba K, Savile C, Fujii R, Yu J-G, Gerassennov T, Kazlauskas RJ & Bornscheuer UT (2009) Converting an esterase into an epoxide hydrolase. Angew Chem Int Ed Engl 48, 3532–3535.

12 Padhi SK, Fujii R, Legatt GA, Fossum SL, Berchtold R & Kazlauskas RJ (2010) Switching from an esterase to a hydroxynitrile lyase mechanism requires only two amino acid substitutions. Chem Biol 17, 863–871.

13 Priyanka P, Tan Y, Kinsella GK, Henehan GT & Ryan BJ (2019) Solvent stable microbial lipases: current understanding and biotechnological applications. Biotechnol Lett 41, 203–220.

14 Lotti M, Pleiss J, Valero F & Ferrer P (2018) Enzymatic production of biodiesel: strategies to overcome methanol inactivation. Biotechnol J 13, 1–10.

15 Peña-Garcia C, Martinez-Martinez M, Reyes-Duarte D & Ferrer M (2016) High throughput screening of esterases, lipases and phospholipases in mutant and metagenomic libraries: a review. Comb Chem High Throughput Screen 19, 605–615.

16 Castilla I, Woods D, Reen F & O’Gara F (2018) Harnessing marine biocatalytic reservoirs for green chemistry applications through metagenomic technologies. Mar Drugs 16, 227.

17 Martínez-Martínez M, Coscolín C, Santiago G, Chow J, Stogiós PJ, Bargiela R, Gertler C, Navarro-Fernández J, Bollinger A, Thies S et al. (2018) Determinants and prediction of esterase substrate promiscuity patterns. ACS Chem Biol 13, 225–234.

18 Pleiss J, Fischer M, Peiker M, Thiele C & Schmid RD (2000) Lipase engineering database: understanding and exploiting sequence-structure-function relationships. J Mol Catal B Enzym 10, 491–508.

19 Hotelier T (2004) ESTHER, the database of the alpha/beta-hydrolase fold superfamily of proteins. Nucleic Acids Res 32, D145–147.

20 Kourist R, Jochens H, Bartsch S, Kuipers R, Padhi SK, Gall M, Böttcher D, Joosten HJ & Bornscheuer UT (2010) The alpha/beta-hydrolase fold 3DM database (ABHDB) as a tool for protein engineering. ChemBioChem 11, 1635–1643.

21 Fischer M, Thai QK, Grieb M & Pleiss J (2006) DWFAR - A data warehouse system for analyzing protein families. BMC Bioinformatics 7, 1–10.

22 Henke E, Pleiss J & Bornscheuer UT (2002) Activity of lipases and esterases towards tertiary alcohols: insights into structure-function relationships. Angew Chemie Int Ed 41, 3211–3213.

23 Barth S, Fischer M, Schmid RD & Pleiss J (2004) Sequence and structure of epoxide hydrolases: a systematic analysis. Proteins Struct Funct Genet 55, 846–855.

24 Barth S, Fischer M, Schmid RD & Pleiss J (2004) The database of epoxide hydrolases and haloalkane dehalogenases: one structure, many functions. Bioinformatics 20, 2845–2847.

25 Parravicini F, Natalella A, Papaleo E, De Gioia L, Doglia SM, Lotti M & Brocca S (2013) Reciprocal influence of protein domains in the cold-adapted acyl aminoaeryl peptidase from Sporosarcina psychrophila. PLoS ONE 8, 1–11.

26 Jan AH, Dubreucq É & Subileau M (2017) Revealing the roles of subdomains in the catalytic behavior of lipases/acyltransferases homologous to CplIP2 through rational design of chimeric enzymes. ChemBioChem 18, 941–950.

27 Argiriadi MA, Morisseau C, Hammock BD & Christianson DW (1999) Detoxification of environmental mutagens and carcinogens: structure, mechanism, and evolution of liver epoxide hydrolase. Proc Natl Acad Sci USA 96, 10637–10642.

28 van Tilbeurgh H, Sarda L, Verger R & Cambillau C (2004) The database of epoxide hydrolases and haloalkane dehalogenases: one structure, many functions. Bioinformatics 20, 2845–2847.

29 Carpintero C, Nussinov R & Barequet G (2018) Percolation in protein sequence space. Nat Commun 10, 1–10.

30 Buchholz PCF, Fademrecht S & Pleiss J (2017) Percolation in protein sequence space. PLoS ONE 12, 1–11.

31 Buchholz PCF, Zeil C & Pleiss J (2018) The scale-free nature of protein sequence space. PLoS ONE 13, 1–14.
32 Dawson NL, Lewis TE, Das S, Lees JG, Lee D, Ashford P, Orengo CA & Sillitoe I (2017) CATH: an expanded resource to predict protein function through structure and sequence. *Nucleic Acids Res* **45**, D289–D295.

33 Sun Y, Yin S, Feng Y, Li J, Zhou J, Liu C, Zhu G & Guo Z (2014) Molecular basis of the general base catalysis of an α/β-hydrolase catalytic triad. *J Biol Chem* **289**, 15867–15879.

34 Bornscheuer UT (2013) Enzymes in lipid modification: from classical biocatalysis with commercial enzymes to advanced protein engineering tools. *Oléagineux Corps Gras Lipides* **20**, 45–49.

35 Widmann M, Juhl PB & Pleiss J (2010) Structural classification by the Lipase Engineering Database: a case study of *Candida antarctica* lipase A. *BMC Genom* **11**, 1–8.

36 Mandrich L, Menschke V, Alterio V, De Simone G, Pedone C, Rossi M & Manco G (2007) Functional and structural features of the oxyanion hole in a thermostable esterase from *Alicyclobacillus acidocaldarius*. *Proteins Struct Funct Bioinforma* **71**, 1721–1731.

37 Shan L, Matthews II & Kholas C (2005) Structural and mechanistic analysis of two prolyl endopeptidases: Role of interdomain dynamics in catalysis and specificity. *Proc Natl Acad Sci USA* **102**, 3599–3604.

38 Zou J, Hallberg BM, Bergfors T, Oesch F, Arand M, Mowbray SL & Jones TA (2000) Structure of *Aspergillus niger* epoxide hydrolase at 1.8 Å resolution: implications for the structure and function of the mammalian microsomal class of epoxide hydrolases. *Structure* **8**, 111–122.

39 Schrag JD, Li Y, Cygler M, Lang D, Burgdorf T, Hecht HJ, Schmid R, Schomburg D, Rydel TJ, Oliver JD *et al.* (1997) The open conformation of a Pseudomonas lipase. *Structure* **5**, 187–202.

40 Brzozowski AM, Derewenda U, Derewenda ZS, Dodson GG, Lawson DM, Turkenburg JP, Bjorkling F, Huge-Jensen B, Patkar SA & Thim L (1991) A model for interfacial activation in lipases from the structure of a fungal lipase-inhibitor complex. *Nature* **351**, 491–494.

41 Derewenda U, Brzozowski AM, Lawson DM & Derewenda ZS (1992) Catalysis at the interface: the anatomy of a conformational change in a triglyceride lipase. *Biochemistry* **31**, 1532–1541.

42 Schmid RD & Verger R (2002) Lipases: interfacial enzymes with attractive applications. *Angew Chemie Int Ed* **37**, 1608–1633.

43 Roussel A, Amara S, Nyyssölä A, Mateos-Diaz E, Blangy S, Kontkanen H, Westerholm-Parvinen A, Carrière F & Cambillau C (2014) A cutinase from *Trichoderma reesei* with a lid-covered active site and kinetic properties of true lipases. *J Mol Biol* **426**, 3757–3772.

44 Grochulski P, Li Y, Schrag JD & Cygler M (1994) Two conformational states of *Candida rugosa* lipase. *Protein Sci* **3**, 82–91.

45 Jaeger K-E & Reetz MT (1998) Microbial lipases form versatile tools for biotechnology. *Trends Biotechnol* **16**, 396–403.

46 Wang Y, Wei D-Q & Wang J-F (2010) Molecular dynamics studies on T1 Lipase: insight into a double-flap mechanism. *J Chem Inf Model* **50**, 875–878.

47 Hermoso J, Pignol D, Kerfellec B, Crenon I, Chapus C & Fontecilla-Camps JC (1996) Lipase activation by nonionic detergents. The crystal structure of the porcine lipase-colipase-tetraethylene glycol monoocetyl ether complex. *J Biol Chem* **271**, 18007–18016.

48 Van Poudroyen G, Eggert T, Jaeger KE & Dijkstra BW (2001) The crystal structure of *Bacillus subtilis* lipase: a minimal α/β hydrolase fold enzyme. *J Mol Biol* **309**, 215–226.

49 Kim KK, Song HK, Shin DH, Hwang KY, Choe S, Yoo OJ & Suh SW (1997) Crystal structure of carboxylesterase from *Pseudomonas fluorescens*, an alpha-beta hydrolase with broad substrate specificity. *Structure* **5**, 1571–1584.

50 Martinez C, De Geus P, Lauwereys M, MatthysSENS G & Cambillau C (1992) *Fusarium solani* cutinase is a lipolytic enzyme with a catalytic serine accessible to solvent. *Nature* **356**, 615–618.

51 Grochulski P, Li Y, Schrag JD, Bouthillier F, Smith P, Harrison D, Rubin B & Cygler M (1993) Insights into interfacial activation from an open structure of *Candida rugosa* lipase. *J Biol Chem* **268**, 12843–12847.

52 Nardini M, Lang DA, Liebeton K, Jaeger KE & Dijkstra BW (2000) Crystal structure of *Pseudomonas aeruginosa* lipase in the open conformation. The prototype for family I.I of bacterial lipases. *J Biol Chem* **275**, 31219–31225.

53 Kohno M, Funatsu J, Mikami B, Kugimiya W, Matsuo T & Morita Y (1996) The crystal structure of lipase II from *Rhizopus niveus* at 2.2 Å resolution. *J Biochem* **120**, 505–510.

54 Gomez GA, Morisseau C, Hammock BD & Christianson DW (2004) Structure of human epoxide hydrolase reveals mechanistic inferences on bifunctional catalysis in epoxide and phosphate ester hydrolysis. *Biochemistry* **43**, 4716–4723.

55 Lack NA, Yam KC, Lowe EE, Horsman GP, Owen RL, Sim E & Eltis LD (2010) Characterization of a carbon-carbon hydrolase from *Mycobacterium tuberculosis* involved in cholesterol metabolism. *J Biol Chem* **285**, 434–443.

56 Yoshimoto T, Kabashima T, Uchikawa K, Inoue T, Tanaka N, Nakamura KT, Tsuru M & Ito K (1999) Crystal structure of prolyl aminopeptidase from *Serratia marcescens*. *J Biochem* **126**, 559–565.
57 Chan PWY, Yakunin AF, Edwards EA & Pai EF (2011) Mapping the reaction coordinates of enzymatic defluorination. J Am Chem Soc 133, 7461–7468.

58 Schmidt A, Gruber K, Kratky C & Lamzin VS (2008) Atomic resolution crystal structures and quantum chemistry meet to reveal subtleties of hydroxynitrile lyase catalysis. J Biol Chem 283, 21827–21836.

59 Franken SM, Rozeboom HJ, Kalk KH & Dijkstra BW (1991) Crystal structure of haloalkane dehalogenase: an enzyme to detoxify halogenated alkanes. EMBO J 10, 1297–1302.

60 Brezovsky J, Babkova P, Degtjarik O, Fortova A, Gora A, Iermak I, Rezacova P, Dvorak P, Smatanova IK, Prokop Z et al. (2016) Engineering a de novo transport tunnel. ACS Catal 6, 7597–7610.

61 Aertgeerts K, Ye S, Tennant MG, Kraus ML, Rogers A, Sang B-C, Skene RJ, Webb DR & Prasad GS (2004) Structural basis of proline-specific exopeptidase activity as observed in human dipeptidyl peptidase IV in complex with a decapeptide reveals details on substrate specificity and tetrahedral intermediate formation. Protein Sci 13, 412–421.

62 Ross B, Krapp S, Augustin M, Kierfersauer R, Arciniega M, Geiss-Friedlander R & Huber R (2018) Structures and mechanism of dipeptidyl peptidases 8 and 9, important players in cellular homeostasis and cancer. Proc Natl Acad Sci USA 115, E1437–E1445.

63 Thoma R, Löfler B, Stihle M, Huber W, Ruf A & Hennig M (2003) Structural basis of proline-specific exopeptidase activity as observed in human dipeptidyl peptidase-IV. Structure 11, 947–959.

64 Roppongi S, Suzuki Y, Tateoka C, Fujimoto M, Morisawa S, Iizuka I, Nakamura A, Honma N, Shida Y, Ogasawara W et al. (2018) Crystal structures of a bacterial dipeptidyl peptidase IV reveal a novel substrate recognition mechanism distinct from that of mammalian orthologues. Sci Rep 8, 2714.

65 Fülop V, Szeltner Z & Polgár L (2000) Catalysis of serine oligopeptidases is controlled by a gating filter mechanism. EMBO Rep 1, 277–281.

66 Szeltner Z, Renner V & Polgár L (2000) The noncatalytic β-propeller domain of prolyl oligopeptidase enhances the catalytic capability of the peptidase domain. J Biol Chem 275, 15000–15005.

67 Juhász T, Szeltner Z, Fülop V & Polgár L (2005) Unclosed β-propellers display stable structures: implications for substrate access to the active site of prolyl oligopeptidase. J Mol Biol 346, 907–917.

68 Tarragó T, Martín-Benito J, Sabidó E, Claesen B, Madurga S, Gairi M, Valpuesta JM & Giralt E (2009) A new side opening on prolyl oligopeptidase revealed by electron microscopy. FEBS Lett 583, 3344–3348.

69 Dryden DTF, Thomson AR & White JH (2008) How much of protein sequence space has been explored by life on Earth? J R Soc Interface 5, 953–956.

70 Dellus-Gur E, Toth-Petroczy A, Elias M & Tawfik DS (2013) What makes a protein fold amenable to functional innovation? Fold polarity and stability trade-offs. J Mol Biol 425, 2609–2621.

71 Jensen RA (2003) Enzyme recruitment in evolution of new function. Annu Rev Microbiol 57, 409–425.

72 Khersonsky O, Roedveldt C & Tawfik DS (2006) Enzyme promiscuity: evolutionary and mechanistic aspects. Curr Opin Chem Biol 10, 498–508.

73 Frenkel ZM & Trifonov EN (2007) Walking through protein sequence space. J Theor Biol 244, 77–80.

74 Vogel C & Pleiss J (2014) The modular structure of ThDP-dependent enzymes. Proteins Struct Funct Bioinforma 82, 2523–2537.

75 Boraston AB, Bolam DN, Gilbert HJ & Davies GJ (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem J 382, 769–781.

76 Sergio Hlap J & Blouin C (2015) The semantics of the modular architecture of protein structures. Curr Protein Pept Sci 17, 62–71.

77 Petrey D, Fischer M & Honig B (2009) Structural relationships among proteins with different global topologies and their implications for function annotation strategies. Proc Natl Acad Sci USA 106, 17377–17382.

78 Ekman D, Björklund AK & Elofsson A (2007) Quantification of the elevated rate of domain rearrangements in Metazoa. J Mol Biol 372, 1337–1348.

79 Weiner J, Beauassart F & Bornberg-Bauer E (2006) Domain deletions and substitutions in the modular protein evolution. FEBS J 273, 2037–2047.

80 Ueguchi-Tanaka M, Nakajima M, Katoh E, Ohmiya H, Asano K, Saji S, Hongyu X, Ashikari M, Kitano H, Yamaguchi I et al. (2007) Molecular interactions of a soluble gibberellin receptor, GID1, with a rice DELLA protein, SLR1, and gibberellin. Plant Cell 19, 2140.

81 Ueguchi-Tanaka M, Ashikari M, Nakajima M, Itoh H, Katoh E, Kobayashi M, Chow TY, Hsing YIC, Kitano H, Yamaguchi I et al. (2005) Gibberellin insensitive dwarf1 encodes a soluble receptor for gibberelin. Nature 437, 693–698.

82 Harel M, Cheah E, Sussman JL, Dijkstra B, Ollis DL, Remington SJ, Verschueren KHG, Goldman A, Schrag J, Silman I et al. (2007) The α/β hydrolase fold. Protein Eng Des Sel 5, 197–211.

83 Koehnke J, Jin X, Budreck EC, Posy S, Scheiffele P, Honig B & Shapiro L (2008) Crystal structure of the extracellular cholinesterase-like domain from neuroligin-2. Proc Natl Acad Sci 105, 1873–1878.

84 Araç D, Boucard AA, Ozkan E, Strop P, Newell E, Sündorf TC & Brunger AT (2007) Structures of neuroligin-1 and the neuroligin-1/neurexin-1 beta complex reveal specific protein-protein and protein-Ca2+ interactions. Neuron 56, 992–1003.
85 Han X, Liu W, Huang J-W, Ma J, Zheng Y, Ko T-P, Xu L, Cheng Y-S, Chen C-C & Guo R-T (2017) Structural insight into catalytic mechanism of PET hydrolase. Nat Commun 8, 2106.
86 Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461.
87 Buchholz PCF, Vogel C, Reusch W, Pohl M, Rother D, Spieß AC & Pleiss J (2016) BioCatNet: a database system for the integration of enzyme sequences and biocatalytic experiments. ChemBioChem 17, 2093–2098.
88 Sayers EW, Cavanaugh M, Clark K, Ostell J, Pruitt KD & Karsch-Mizrachi I (2019) GenBank. Nucleic Acids Res 47, D94–D99.
89 Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN & Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28, 235–242.
90 Needleman SB & Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 48, 443–453.
91 Rice P, Longden I & Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16, 276–277.
92 Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7, 539–539.
93 Holm L & Laakso LM (2016) Dali server update. Nucleic Acids Res 44, W351–W355.
94 Holm L, Kääriäinen S, Rosenström P & Schenkel A (2008) Searching protein structure databases with DaliLite vol 3. Bioinformatics 24, 2780–2781.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Exemplary hub sequences with their annotated source organisms and descriptions.

Table S2. List of the proteins names and abbreviations of the representative structures.

Table S3. Sequence length of the representative protein structures retrieved from the PDB.

Table S4. Representative structures used for the structural comparison of additional modules to other proteins using the Dali web server.

Fig S1. Cluster distributions $N(s)$ with linear regressions for cluster sizes $s \leq 100$ at thresholds of global sequence identity of 60% (A), 70% (B), 80% (C) and 90% (D).

Fig S2. Fitted exponents $\tau$ (dots), derived from fits of the slopes of the histograms, $\tau_n$, against the model distribution from $\dot{t}$, determined for different thresholds of global sequence identity.

Fig S3. Topology diagram of the core domain of $\alpha/\beta$-hydrolases.

Fig S4. Topology diagram of $\alpha/\beta$-hydrolases with an additional lid (green).

Fig S5. Topology diagram of $\alpha/\beta$-hydrolases with an additional single cap (red), N-terminal cap (pale red) or double cap.

Fig S6. Topology diagram of $\alpha/\beta$-hydrolases with an additional N-terminal propeller domain (blue).

Fig S7. Topology diagram of $\alpha/\beta$-hydrolases with an additional C-terminal $\beta$-domain (blue).

Fig S8. Topology diagram of $\alpha/\beta$-hydrolases with two additional domains: an additional lid (green) and C-terminal $\beta$-sandwich domain (blue).

Fig S9. Topology diagram of $\alpha/\beta$-hydrolases with an additional cap (red) and an N-terminal $\beta$-domain (blue).

Fig S10. Surface view from above onto the active site of $\alpha/\beta$-hydrolases showing the accessibility of the active site pocket.

Fig S11. Structure view of different $\alpha/\beta$-hydrolases showing the composition of the additional modules.

Fig S12. Surface view of $\alpha/\beta$-hydrolases.

Fig S13. Explosion graphs of $\alpha/\beta$-hydrolases show the coverage of the active site by additional modules.

Fig S14. Heat map of the $Z$-scores from the pairwise structural comparison of each $\alpha/\beta$-hydrolase domain and additional C- and N-terminal domain using the Dali web server.