Protein Engineering, Design & Selection vol. 25 no. 11 pp. 689–697, 2012
Published online October 4, 2012 doi:10.1093/protein/gzs068

Bioinformatic analysis of alpha/beta-hydrolase fold enzymes reveals subfamily-specific positions responsible for discrimination of amidase and lipase activities†

D.A.Suplatov[1], W.Besenmatter[2], V.K.Švedas[1,3] and A.Svendsen[2,3]

[1]Lomonosov Moscow State University, Faculty of Bioengineering and Bioinformatics and Belozersky Institute of Physicochemical Biology, Lenin Hills 1/73, Moscow 119991, Russia and [2]Novozymes A/S, Krogshoejvej 36, 2880 Bagsvaerd, Denmark

[3]To whom correspondence should be addressed. ASv@novozymes.com (A.S.); vytas@belozersky.msu.ru (V.S.)

Received July 6, 2012; revised August 17, 2012; accepted August 20, 2012

Edited by Andrew Wang

Superfamily of alpha-beta hydrolases is one of the largest groups of structurally related enzymes with diverse catalytic functions. Bioinformatic analysis was used to study how lipase and amidase catalytic activities are implemented into the same structural framework. Subfamily-specific positions—conserved within lipases and peptidases but different between them—that were supposed to be responsible for functional discrimination have been identified. Mutations at subfamily-specific positions were used to introduce amidase activity into *Candida antarctica* lipase B (CALB). Molecular modeling was implemented to evaluate influence of selected residues on binding and catalytic conversion of amide substrate by corresponding library of mutants. *In silico* screening was applied to select reactive enzyme-substrate complexes that satisfy knowledge-based criteria of amidase catalytic activity. Selected CALB variants with substitutions at subfamily-specific positions Gly39, Thr103, Trp104, and Leu278 were produced and showed significant improvement of experimentally measured amidase activity. Based on these results, we suggest that value of subfamily-specific positions should be further explored in order to develop a systematic tool to study structure-function relationship in enzymes and to use this information for rational enzyme engineering.

Keywords: bioinformatic analysis/CALB/catalytic promiscuity/α/β-hydrolases/subfamily-specific positions

Introduction

The α/β-hydrolase fold superfamily is one of the largest groups of enzymes with diverse catalytic function that share common structural framework and appear to be related by divergent evolution (Holmquist, 2000). Members of this group lost sequence similarity during natural selection and specialization from a common ancestor to efficiently operate on various substrates with different structure and physicochemical properties. While subunit size and oligomeric organization of family members can vary they all share conserved structural core and contain three topologically similar residues that form a catalytic triad (a nucleophile, catalytic acid and histidine) (Ollis *et al.*, 1992; Carr and Ollis, 2009). Thus, α/β-hydrolases provide an excellent model to study structure–function relationship of multiple catalytic capabilities within active sites of common structural organization.

Lipases are members of α/β-hydrolase fold superfamily that catalyze hydrolysis of poorly water-soluble long-chain acylglycerols (Rubin and Dennis, 1997). Lipases are highly flexible biocatalysts used for the acylation and deacylation of a broad range of natural and unnatural substrates with high regioselectivity and enantioselectivity (Kazlauskas and Bornscheuer, 1998), stable under extreme reaction conditions in both aqueous and organic media (Zaks and Klibanov, 1984). They have found a widespread application in organic synthesis, food industry and biotechnology (Schmid and Verger, 1998). In addition, use of lipases for enantioselective acylation of chiral amines has been attracting a particular attention in recent years (Van Rantwijk and Sheldon, 2004). However, lipases were shown to be very poor catalysts for hydrolysis of amides (Henke and Bornscheuer, 2003) despite wide spread of protease activities within the α/β-hydrolase fold. It is thus important from both fundamental and practical points of view to study molecular mechanisms of lipase and amidase catalytic promiscuity within the same structural context.

Comparative bioinformatic analysis of α/β-hydrolases with lipase and protease activities was performed and high structural similarity of active sites was observed despite different types of catalyzed chemical transformations, low sequence and full-structure identity. While completely conserved positions in α/β-hydrolases define properties that are common for the entire family (for example, have a direct role in enzyme catalytic machinery) they do not explain functional diversity. Thus, recently developed bioinformatic analysis method (Suplatov *et al.*, 2012) was used to identify subfamily-specific positions (SSPs)—conserved only within protein subfamilies, but different between subfamilies—that were supposed to be responsible for functional discrimination between lipases and peptidases. These hotspots were used to construct *in silico* library of *Candida antarctica* lipase B (CALB) mutants. Molecular modeling was applied to screen the variants by evaluating their ability to stabilize amide substrate in the active site. Structural filtration was implemented

†This paper is continuation of the **INPEC 2012 meeting** talk 'Introduction of amidase activity into a lipase B from *Candida antarctica*' presented by Allan Svendsen in Taipei.

to select productive enzyme–substrate complexes that satisfy knowledge-based criteria of amidase catalytic activity: substrate interaction with the catalytic nucleophile (distance, angle and dihedral angle of nucleophilic attack trajectory) and the oxyanion hole residues. Selected CALB mutants were then produced and shown experimentally to possess improved amidase activity.

Materials and methods

Bioinformatic analysis

SSM program (Krissinel and Henrick, 2004) was used to search for significant structure matches between the query structure 1TCB of CALB and target peptidase structures available in the PDB databank. The following non-redundant set (not more than 95% of pairwise sequence identity) of structures was finally selected: CALB (1TCB), wheat serine carboxypeptidase (1WHS), human dipeptidyl peptidases 4 and 7 (1J2E and 3N0T) and proline iminopeptidase (1AZW). Local structure alignment of active sites was done as recently described (Suplatov *et al.*, 2011). Then, sequence similarity search with PSI-BLAST v.2.2.18 (Altschul *et al.*, 1997) against Swiss-Prot database ('nr' database in case of CALB) was applied independently to each selected structure. Significance *e*-value threshold for reconstruction of PSSM search matrix was set to 10^{-5}. Results from different PSI-BLAST iterations were combined to realize both the benefits of low corruption from the early rounds and the benefits of high sensitivity from the later rounds (Lee *et al.*, 2008). The resulting set was filtered to attain non-redundancy on 95% level, incomplete sequences were removed. T-coffee (Notredame *et al.*, 2000) was used to align sequences to corresponding structures. Resulting structure-based sequence alignment contained 192 proteins.

Bioinformatic analysis algorithm (available at http://biokinet.belozersky.msu.ru/zebra) was used to calculate and rank SSPs in lipases and peptidases. Alignment columns with >35% of gaps were dismissed. The most statistically significant classification was identified with six subfamilies: carboxypeptidases and dipeptidyl peptidases were represented by two subfamilies each, proline iminopeptidases and CALB-like proteins—by one subfamily each.

Modeling amidase activity of wild-type CALB

No binding of the amide substrate (2-chloro-*N*-benzylacetamide) in near-to-attack conformation was identified by molecular docking in the crystallographic structure of CALB (1TCB). Thus, wild-type CALB was simulated for 5.7 ns in water box and 2000 coordinate structure files were retrieved from unconstrained part of the run with 2.5 ps interval. Then, 2-chloro-*N*-benzylacetamide was docked into every frame and the structure with positive binding and smallest root mean square deviation versus initial crystallographic structure was selected (0.41 Å for Cα atoms and 0.42 Å for non-hydrogen atoms). This enzyme-substrate configuration was relaxed using 2500 steps of conjugant-gradient energy minimization algorithm. Molecular docking and dynamics were carried out as explained below.

Preparation of in silico *mutant library and molecular screening*

Three-dimensional structures of lipase mutants were constructed using Mutate v.1.2.1 (Stroganov *et al.*, 2008). Full-atomic models of the mutants were generated by assigning protonation states of arginine, lysine, histidine, aspartate and glutamate residues according to side-chain pKa's determined using PropKa (Olsson *et al.*, 2011). PDB2PQR (Dolinsky *et al.*, 2007) was used to add missing hydrogens, resolve steric hindrances and optimize hydrogen bond network. Coordinate structure file of 2-chloro-*N*-benzylacetamide was created with ChemSketch (ACD/ChemSketch Freeware, version 10.00, Advanced Chemistry Development, Inc., Toronto, ON, Canada, www.acdlabs.com, 2012). Molecular docking calculations were performed using Autodock 4 (Huey *et al.*, 2007). The protein was kept rigid while three active torsions of the substrate were allowed to vary. Hybrid global–local search (Lamarckian genetic algorithm) was applied in 100 independent docking simulations with 150 individuals in a population, 2 500 000 maximum energy evaluations and 300 iterations of strict Solis and Wets local search algorithm. The most energetically favorable substrate positions of each generation were passed to the next generation with mutation and crossover rates set to 0.02 and 0.8, respectively. A cluster analysis was performed on docking results with 1.0 Å rmsd tolerance. Structural filtration was used as a post-docking tool to evaluate near-to-attack conformation of the substrate in active site. Distances of crucial catalytic interactions were restricted as discussed in the text to describe trajectory of the nucleophilic attack. Docking results out of range were dismissed. Two parameters were further used to describe docking results: estimated free energy of binding and frequency of the correct binding geometry. The latter corresponds to how many times the near-to-attack conformation was observed after 100 independent docking runs. Both parameters were taken from Autodock output file. The PyMOL Molecular Graphics System version 1.5.0.1 (Schrödinger, LLC) was used to visualize results and prepare illustrations.

Molecular dynamics simulation

Long-time simulations of wild type and mutant forms of CALB were carried out using NAMD version 2.8 (Phillips *et al.*, 2005). Calculations were performed in a cubic cell filled with TIP3P water (Jorgensen *et al.*, 1983) with standard periodic boundary conditions. Minimal distance between any atom of the protein and box edges was 15 Å. Constant ionic strength 0.1 M was generated by randomly inserting Na+/Cl^- pairs into the simulation box. Total size of a CALB system in water box was ~52 500 atoms. All atoms were considered explicitly and defined in Amber force field (DePaul *et al.*, 2010) using AmberTools package version 1.3. Simulations were carried out in canonical (constant number of particles, volume, and temperature) ensemble. The temperature was controlled by Langevin dynamics with a damping factor of 5 ps^{-1}. Long-range electrostatic interactions were computed using particle mesh Ewald method (Darden *et al.*, 1993) with 12 Å distance cutoff for direct electrostatic and van der Waals interactions. Switching function was set to 10 Å and smoothed the van der Waals potentials to 0 by the cutoff distance. Geometry of water

molecules as well as bonds between hydrogens and attached atoms were constrained to the length and angles defined by the force field. SETTLE was used to keep the water rigid. Thus, 2 fs integration time-step was used for calculations. The initial configuration was relaxed using 2500 steps of conjugant-gradient energy minimization algorithm. Then, all atoms of protein and substrate were constrained using a harmonic energy constraint function with a spring constant of 3 kcal/mol/Å^2 and system was heated for 120 ps from 0 to 300 K. During the next 580 ps of equilibration run constraints were gradually removed in 29 steps (20 ps each), after which a free (unconstrained) run followed for 5 ns. Structures were saved for analysis every 5 ps. Equilibration run allowed for water molecules to fill empty cavities in enzyme structure and optimized substrate orientation and interactions with the active site residues. Unconstrained run was used to calculate statistics of structure movement and binging geometry with visual molecular dynamics (Humphrey *et al.*, 1996).

Identification of hydrogen bonds

A hydrogen bond was counted if it was present for at least 50% of time during unconstrained trajectory run within 3.5 Å donor–acceptor distance and 150–180° donor–hydrogen–acceptor angle (Finkelstein and Ptitsyn, 2002; Tina *et al.*, 2007).

Production of CALB variants

Variants of CALB were generated by polymerase chain reaction (PCR)-based site-directed mutagenesis. The PCR was set up with the proofreading KOD DNA polymerase (from Novagen, Toyobo) and a 7467 basepair *Escherichia coli–Aspergillus* plasmid that was previously methylated with CpG methyltransferase (from NEB). The PCR products were used to transform competent *E. coli* DH5α cells (from TaKaRa). Plasmid DNA was recovered and sequenced to verify the presence of the desired substitution. Confirmed plasmid variants were used to transform an *Aspergillus oryzae* strain that is negative in pyrG (orotidine-5′phosphate decarboxylase) and that is also negative in the proteases pepC (a serine protease homologous to yscB), alp (an alkaline protease) NpI (a neutral metalloprotease I) to avoid degradation of the lipase variants during and after fermentation.

The transformed *Aspergillus* strains were fermented as submerged culture in shake flasks and the lipase variants secreted into the fermentation medium. After the fermentation, the lipase variants were purified from the sterile filtered fermentation medium in a three-step procedure with (i) hydrophobic interaction chromatography on decylamine-agarose, (ii) buffer exchange by gel filtration, and (iii) ion exchange chromatography with cation exchange on SP-sepharose at pH 4.5. The lipase variant solutions were stored frozen.

Experimental determination of amidase activity

The amidase activity was determined with a fluorimetric assay similar to the one described previously (Henke and Bornscheuer, 2003).

The aqueous reaction mixture contained 0.03 mg/ml CALB variant, 5 mM substrate (benzyl chloroacetamide), 25 mM phosphate (potassium salt) pH 7.0, 10% (w/v) tetrahydrofuran and was set up in 96-well microtiter plates.

The concentration of the enzyme stock solution was determined by measuring the absorbance at 280 nm and calculated based on the extinction coefficient of 1.21 for CaLB. The microtiter plate was covered with parafilm and incubated for 18 h at 37°C and 300 rpm in the MTP Thermomixer Comfort (form Eppendorf AG). Afterwards 50 µl of 20 mM 4-chloro-7-nitrobenzofurazan (NBDCl) in 1-hexanol was pipetted to 200 µl reaction mixture and incubated for 1 h at 37°C and 500 rpm.

Afterwards the fluorescence intensity was measured with the fluorimeter Fluostar Optima (from BMG Labtech GmbH) with excitation filter at 485 nm and emission filter at 540 nm. Each enzyme variant was measured in three replications, in three wells. In parallel, each enzyme variant was also set up in three different wells without substrate, to measure the small background fluorescents from each variant. To measure the small autohydrolysis of the substrate, further three wells were set up with substrate and without enzyme. Further three wells were set up without substrate and without enzyme, thus only aqueous medium with buffer and solvent, to measure the small background fluorescents from the medium.

The average value from the measurements with enzyme and with substrate is named ‘es’; the value with enzyme and without substrate is ‘e’; the value without enzyme and with substrate is ‘s’; the value without enzyme and without substrate is ‘o’. To correct for fluorescents from enzyme, medium and autohydrolysis the following subtractions were calculated: $(es-e)-(s-o) = (es-s)-(e-o) = es-e-s+o$ = corrected enzyme variant activity value. $s-o$ = autohydrolysis of substrate. Because these measurements are in arbitrary fluorescents units, these values were normalized to percentage of substrate that reacted. Zero percent means no reaction of substrate, 100% means complete reaction of substrate to products. For this normalization two wells for each enzyme variant were set up in parallel. These wells contained the same composition as the other wells, except that they contained instead of substrate the products at a concentration that corresponds to 10% reaction. Thus, the concentrations in these two wells were 0.03 mg/ml enzyme, 0.5 mM hydrolysis products (i.e. 0.5 mM benzylamine and 0.5 mM chloroacetic acid), 25 mM phosphate pH 7.0, 10% (w/v) tetrahydrofuran. The average value from the measurements with enzyme and with product is named ‘ep’. To correct for background fluorescents from the enzyme variants, the average value from the measurements with enzyme and without product (and without substrate), which was named ‘e’, was subtracted. In short, the normalized and corrected enzyme variant activity was calculated as $10\cdot((es-e)-(s-o))/(ep-e)$.

Results

Bioinformatic analysis of α/β-hydrolases with lipase and peptidase activities

CALB represents a special class of lipases as it lacks interfacial activation and appears to be in a constantly open conformation. A small lid is limiting the active site entrance which could account for substrate specificity and high stereospecificity of this lipase (Uppenberg *et al.*, 1994). Interestingly, CALB was previously shown to have modest amidase activity in contrast to interfacially activated lipases that are usually not suitable for hydrolytic cleavage of

amides (Henke and Bornscheuer, 2003). Therefore, CALB presents an interesting example to study why lipases do not hydrolyze amides despite the fact that peptidase activity is available within the α/β-hydrolase fold.

Divergent evolution created multiple catalytic capabilities within active sites of common structural organization. Catalytic versatility observed in large superfamilies indicates deep specialization from a common ancestor accompanied by major loss of sequence similarity. Thus, comparative structure analysis of catalytic sites is expected to provide more clues to understanding protein function than sequence alignment alone.

Structure similarity search against the complete PDB databank was used to identify structural homologs of CALB with annotated peptidase activity—serine carboxypeptidases (EC 3.4.16.6), dipeptidyl peptidases (EC 3.4.14.2 and 3.4.14.5), and proline iminopeptidases (EC 3.4.11.5). Comparative analysis revealed that structurally homologous regions of lipases and proteases include main blocks of the α/β-hydrolase fold (Ollis *et al.*, 1992)—β-strands 3, 4, 5, 6, 7 and 8 of the central β-sheet and α-helixes B, C, E and F. High structural similarity of active site regions was observed near the catalytic triad residues—loops between β5 and αC (so-called nucleophilic elbow), β7 and αE, β8 and αF. Residues of the catalytic triad were identified as completely conserved in the dataset. Those positions not only contained identical amino acid types Ser-Asp-His but also had the same location and orientation in the structure (Fig. 1). A notable difference in orientation of the loop between β3 and αA ('3A-loop') was observed (residues 37–43 in CALB). In wheat serine carboxypeptidase (CPDW) and proline iminopeptidase from *Xanthomonas campestris* (XcPIP) its orientation was similar to CALB but contained a proline insertion (Pro54 and Pro44, respectively) that shifted subsequent conserved Gly55 and Gly45 into an arc 3.5 Å away from main chain of CALB. Contrary, in human dipeptidyl peptidases 4 and 7 (DPP4 and DPP7) both orientation and length of 3A-loop were different compared with CALB. Despite major structural difference the 3A-loop is a host to one of the oxyanion hole residues in all enzymes studied. In CALB, CPDW and XcPIP a negative charge on the tetrahedral intermediate during catalysis can be stabilized by hydrogen bond with main-chain NH group of Thr40, Gly53 and Gly43, respectively, while in DPP4 and DPP7 the same role can be played by side-chain atoms—OH group of Tyr547 and NδH2 group of Asn77. Thus, it seems that similar function is played by analogous residues with different evolutionary origin. The second hydrogen bond with oxyanion can be formed by structurally conserved main-chain NH group of a residue next to the catalytic serine (Gln106 in CALB). Finally, in CALB a third hydrogen bond in oxyanion hole can be formed by Oγ1 atom of Thr40. It may be concluded that the 3A-loop has preserved its function as a host to one of the oxyanion hole residues during evolution despite major structural variability. Located in the core of the active site in a close vicinity of catalytic triad the 3A-loop could play important role in functional discrimination of α/β-hydrolases.

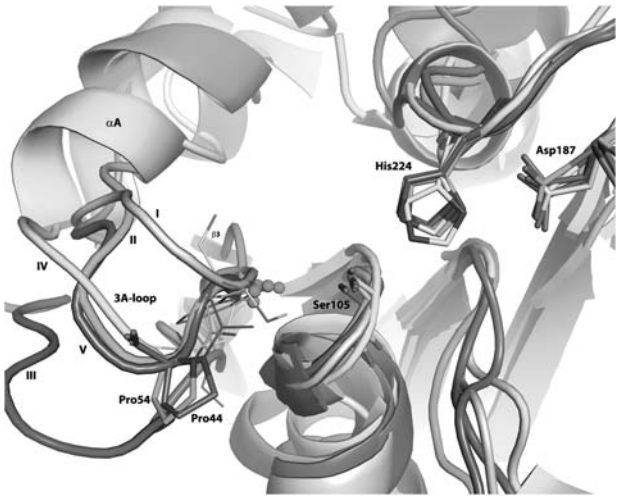

Fig. 1. Structure alignment of *Candida antarctica* lipase B (CALB, I), wheat serine carboxypeptidase (CPDW, II), human dipeptidyl peptidases 4 (DPP4, III) and 7 (DPP7, IV), and proline iminopeptidase (XcPIP, V). Catalytic triad residues in CALB (Ser105, Asp187 and His224) and other proteins are structurally conserved. Notable difference in orientation of the loop between β-strand and αA-helix ("3A-loop") is observed amid overall high structural similarity of active sites. Pro54 (CPDW) and Pro44 (XcPIP) are shown. Hydrogen atoms of main-chain NH groups of Thr40 (CALB), Gly53 (CPDW) and Gly43 (XcPIP), as well as side-chain hydrogens of OH group of Tyr547 (DPP4) and NoH2 group of Asn77 (DPP7) that can participate in hydrogen-bond stabilization of oxyanion during catalysis are shown as spheres.

Multiple structure-based sequence alignment of CALB-like proteins, serine carboxypeptidases, dipeptidyl peptidases and proline iminopeptidases was built (see Methods). Bioinformatic analysis was used to identify SSPs —conserved only within lipase and peptidase subfamilies, but different between them—that are supposed to be responsible for functional discrimination between enzymes with different catalytic activities. Nine statistically significant SSPs—Gly39, Thr40, Thr42, Thr103, Trp104, Gln106, Ile189, Asp223, and Ala225—located within 10 Å from the substrate were selected for further consideration using molecular modeling methods (Table I).

Molecular modeling of amidase activity in CALB

Model of wild type CALB in complex with 2-chloro-N-benzylacetamide was constructed (Fig. 2). Carbonyl carbon atom of the substrate was located within 2.8 Å from the catalytic Ser105, while three potential hydrogen bond donors were positioned around the carbonyl oxygen—main-chain nitrogens of Thr40 (3.2 Å), Gln106 (3.0 Å) and side-chain γ1-oxygen of Thr40 (3.2 Å). In addition, two more geometry parameters of a nucleophilic attack trajectory were measured—the angle $A_{O\cdots C=O}$ defined by the enzyme Ser105 γ-oxygen, the substrate carbonyl carbon, and the substrate carbonyl oxygen; and the dihedral angle $D_{O\cdots C(N)=O}$ between the plane defined by the enzyme Ser105 γ-oxygen, the substrate carbonyl carbon, and the substrate carbonyl oxygen, and the plane defined by the peptide bond. Structural comparison of serine proteases has shown that consensus nucleophilic attack angles should be close to 90° in the reactive Michaelis complexes (Radisky and Koshland, 2002). In our case $D_{O\cdots C(N)=O}$ had the value of 91.1°, but $A_{O\cdots C=O}$ was equal to 74.0°. Thus, the double displacement mechanism via tetrahedral acyl-enzyme intermediate stabilized in the oxyanion hole (formed by Thr40 and Gln106) could be preserved in CALB for the hydrolysis of amides.

Table I. Subfamily-specific positions of lipases and peptidases in the α/β-hydrolase superfamily

Rank	Z score	*P* value	Reference	DPP (56)	DPP (10)	CPD (14)	CPD (55)	PIP (32)	CALB (25)
2	4.8	$5.2*10^{-10}$	189I	W	N	V	I	F (47%) C (28%) V (16%) A (9%)	I (92%) V (8%)
7	4.6	$2.7*10^{-33}$	104W	G	W	E	E	N (47%) G (25%) Q (22%) H (6%)	W
13	4.4	$1.1*10^{-57}$	42T	–	–	C	C	A (50%) S (28%) G (19%) N (3%)	A (36%) T (36%) S (12%) V (12%) D (4%)
14	4.4	$2.0*10^{-62}$	225A	H	S (59%) G (41%)	E	M	W (47%) M (25%) S (22%) V (3%) R (3%)	A (44%) E (32%) R (12%) S (8%) K (4%)
15	4.3	$5.7*10^{-67}$	40T	–	–	G	G	G	T
19	4.3	$7.3*10^{-84}$	39G	–	–	G	G	G	G
20	4.2	$2.6*10^{-87}$	106Q	Y	Y (89%) F (11%)	Y	Y (94%) F (4%) W (2%)	W (53%) M (47%)	Q (96%) A (4%)
23	4.2	$5.7*10^{-100}$	223D	A	D (79%) N (9%) R (7%) A (5%)	G	S (55%) G (45%)	G (78%) R (10%) A (6%) S (6%)	T (48%) D (28%) G (16%) S (4%) E (4%)
40	3.0	$3.4*10^{-112}$	103T	G	G	G	G	G	S (60%) T (40%)

Reference residue numbering is given as in CALB (1TCB). Positions are ranked by declined statistical significance (Z scores). *P* value for a position with rank *k* is a probability for the set of first *k* scores to be observed by chance. Amino acid content is shown for each subfamily at every position.
Number of sequences in each subfamily is given in parentheses.
DPP, dipeptidyl peptidases; CPD, carboxypeptidases; PIP, proline iminopeptidase; CALB, CALB-like proteins.

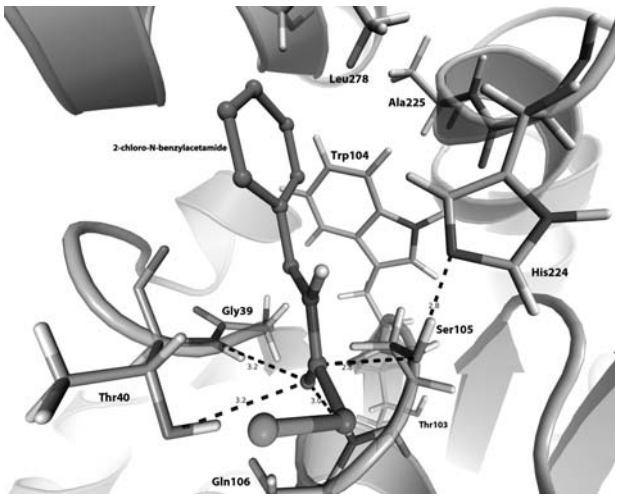

Fig. 2. Molecular model of CALB in complex with 2-chloro-*N*-benzylacetamide. Interactions of the substrate with catalytic Ser105 and oxyanion hole residues (Thr40 and Gln106) are shown. Subfamily-specific positions are shown.

However, observed orientation of the amide bond of the substrate might not be optimal for the nucleophilic attack.

The benzyl group of the substrate points towards a narrow binding cleft formed by hydrophobic sidechains of Trp104, Ile189, Leu278, and Ala281. Bioinformatic analysis has shown that position Leu278 in CALB-like group of proteins is occupied by branched aliphatic residues (68% of Leu, 24% of Ile and 4% of Val) while Ala281 preferably contains Ala (76%) and Gly (8%). Both residues are part of 277–285 loop that does not have a structural homology in peptidase enzymes reviewed in this study. Thus, positions 278 and 281 are occupied by residues with similar physicochemical properties within one subfamily and are absent from other subfamilies and thus should be also treated as subfamily-specific. A specific position 104 W in CALB is substituted in peptidases to residues with smaller side-chains. Interestingly, even though some dipeptidyl peptidases (like human DPP4) preserve a tryptophan in this position it has a completely different rotamer orientation that makes extra space available compared to CALB (χ_1/χ_2 angles are equal to $-157^\circ/124^\circ$ and $177^\circ/-132^\circ$, respectively). Specific position Ile189 is located above the catalytic His224 and isolates it from the solvent. Moreover, it creates hydrophobic environment for nitrogen and adjacent hydrogen atoms of the substrate amide bond. It can be substituted by both polar and hydrophobic residues in peptidases. Compared to narrow leaving group binding cleft CALB has wide acyl binding site. This observation is in agreement with structural classification of lipase substrate specificity (Pleiss *et al.*, 1998).

Both oxyanion hole residues—Thr40 and Gln106—showed subfamily-specific variations of amino acid content. Identification of Gln106 as subfamily-specific is especially interesting owing to the fact that only main-chain NH group is catalytically important. It was previously discussed that orientation of catalytically important backbone atoms of the oxyanion hole residues can be stabilized by anchoring interactions with surrounding residues (Pleiss *et al.*, 2000). Sidechain of Gln106 in CALB participates in hydrogen bonds with Met72, Thr76, Asn79, and also γ1-oxygen of Thr40. Thus, sidechain of Thr40 in CALB seems to have dual function—as a hydrogen bond donor to stabilize oxyanion and as a hydrogen bond acceptor to support orientation of Gln106 (Fig. 3). Peptidases have hydrophobic residues (Tyr, Trp, Phe) at position Gln106 that are associated with appropriate hydrophobic pockets in the structure. For example, Tyr147 in CPDW is part of a hydrophobic pocket formed by Pro54, Tyr151, Leu178, Ile216, and Trp308. Consequently, Thr40 of CALB is replaced by Gly in peptidases. Thus, sidechains of oxyanion hole residues may have no direct role in catalytic process but could support functional orientation of backbone atoms. Subfamily-specific

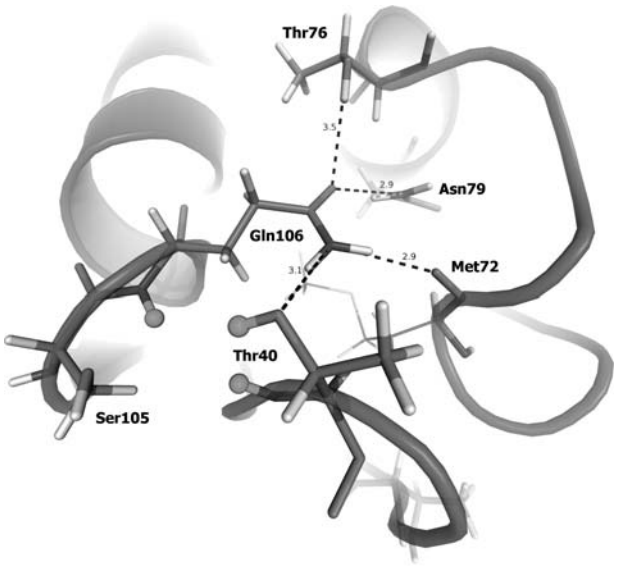

Fig. 3. Dense hydrogen bond network formed by side-chains of oxyanion hole residues Thr40 and Gln106 in CALB that stabilize functional orientation of the backbone. Hydrogen atoms that can participate in hydrogen-bond stabilization of oxyanion during catalysis are shown as spheres.

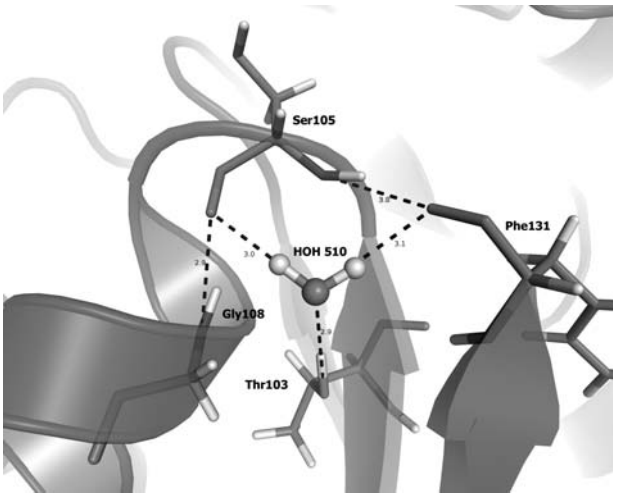

Fig. 4. Thr103 participates in hydrogen bond network with the catalytic Ser105 via coordination of a water molecule.

variations of oxyanion hole residues seem to have structural role accommodated by surrounding environment.

Together with Thr40 two more subfamily-specific residues—Gly39 and Thr42—are located in the 3A-loop. Gly39 is close to both catalytic Ser105 and the substrate. Thr42 is located on the distal side of the loop but its side-chain oxygen could play a structural role of a hydrogen bond donor to the main-chain oxygen of Gly39.

SSP Thr103 in CALB is inaccessible to the solvent and located in the [TS].X.Nu.X.G pattern of the nucleophilic elbow in CALB-like subfamily of proteins while G.X.Nu.X.G pattern is observed in other α/β-hydrolases (Table I). In general, presence of hydrophilic residues in the protein core is thermodynamically unfavorable unless stabilizing polar interaction are formed (Bolon and Mayo, 2001). Thr103 seems to be involved in coordination of a burried water molecule HOH510 (in crystallographic structure 1TCB) that in turn is a hydrogen bond donor to main-chain oxygens of Phe131 and catalytic Ser105. Buried solvent molecules usually form structurally important hydrogen bonds to main-chain atoms that are not involved in intramolecular hydrogen bonds (Park and Saven, 2005). In case of CALB, however, the main-chain oxygen of Ser105 is additionally a hydrogen bond acceptor from Gly108, and main-chain nitrogen is a hydrogen bond donor to Phe131 (distances 2.9–3.1 Å and angles 150°–180°). Thus, a dense hydrogen bond network is formed that can support orientation of the catalytic elbow in CALB (Fig. 4). Molecular dynamics (MD) was used to study hydrogen bond pattern of the nucleophilic elbow in peptidases as they all contain Gly at position Thr103. A similar network was observed but its exact topology was not conserved in different structures. One hydrogen bond between main-chain nitrogen of the catalytic serine and main-chain oxygen of C-terminal residue of the β6 strand (Phe131 in CALB) was present in all observed peptidases. Contrary, number and origin of hydrogen bonds with main-chain oxygen of the catalytic serine can vary. In both DPP4 and DPP7 main-chain oxygen of the catalytic serine participated in only one hydrogen bond with main-chain nitrogen of Tyr634 and Met166, respectively (both residues are homologs of Leu109 in CALB). In CPDW main-chain oxygen of catalytic Ser146 can accept two hydrogen bonds from main-chain nitrogens of Gly149 and Leu178 (homologs of Gly108 and Asp134 in CALB). Similarly, two hydrogen bonds can be formed by main-chain oxygen of the catalytic Ser110 in XcPIP—with side-chain oxygen of Ser113 and main-chain nitrogen of Thr114 (homologs of Gly108 and Leu109 in CALB). Additionally, the main-chain nitrogen of Ser113 is within hydrogen bond distance. MD simulation of CALB mutant T103G indicated a hydrogen bond between main-chain oxygen of Ser105 and main-chain nitrogen of Gly108. Additionally, second compensatory hydrogen bond with main-chain nitrogen of Leu109 was observed, as in peptidases. Thus, substitution T103G does not disrupt the intramolecular stabilizing network but leads to different orientation of the catalytic Ser105 by introducing hydrogen bond pattern of nucleophilic elbow similar to peptidases.

To sum up, SSPs can be used as hot-spots to stabilize near-to-attack conformation of the amide substrate in the active site, introduce more space and decrease hydrophobicity of the leaving group binding subsite in order to rationally design promiscuous catalytic activity in CALB.

Screening and evaluation of mutant libraries

In silico library of 24 CALB variants was constructed by introducing single mutations into SSPs. Molecular docking was used to predict binding states of 2-chloro-*N*-benzylacetamide in active sites of enzyme mutants and to estimate free energy of binding. Resulting enzyme–substrate complexes were analyzed using knowledge-based structure filtration. This post-docking tool was used to select productive binding modes of the substrate by applying simple geometric constraints to crucial catalytic interactions. Four criteria were implemented to describe near-to-attack conformation of the substrate in the active site—distance between γ-oxygen of the catalytic

Table II. Amidase activity of single CALB mutants evaluated *in silico* and by experimental assay

Mutant	Docking evaluation		Experimentally evaluated amidase activity	Mutant	Docking evaluation		Experimentally evaluated amidase activity
	Frequency, %	Binding energy, kcal/mol			Frequency, %	Binding energy, kcal/mol	
G39A	90	−5.9	2.8	A225E	74	−5.7	0.3
A225M	83	−5.8	1.1	WT	73	−5.7	1.0
A225V	80	−5.8	1.2	T42S	73	−5.7	1.2
I189H	79	−5.7	1.1	I189Q	73	−5.6	0.4
D223N	78	−5.8	1.2	A225F	72	−5.7	0.8
D223G	77	−5.8	1.3	T40G	67	−5.0	0.1
A225K	77	−5.8	2.0	T42G	66	−5.7	1.9
W104F	76	−5.7	2.0	A281S	65	−5.8	0.5
I189N	76	−5.4	1.0	T40A	65	−5.2	0.1
T103G	75	−5.7	3.0	L278A	63	−5.5	2.5
A225L	74	−5.8	1.2	I189T	59	−5.5	0.1
A225I	74	−5.7	1.6	I189S	53	−5.2	0.1

Experimentally measured amidase activity values are given as a ratio of those for mutant and wild-type enzymes. Entries are sorted by declining 'frequency' parameter.

Ser105 and substrate carbonyl carbon ('Ser105 γO ··· C'), distances from substrate carbonyl oxygen to backbone nitrogen atoms of Thr40 and Gln106 ('Thr40 N ··· O' and 'Gln106 N ··· O') and to γ1-oxygen of Thr40 ('Thr40 γ1O ··· O', if applicable)—all limited to at most 3.5 Å. Enzyme–substrate complexes were ranked by frequency of substrate binding in near-to-attack conformation out of 100 independent docking runs. Selected mutants were produced experimentally and their amidase activity was measured (Table II). Five single CALB mutants—T103G, G39A, L278A, A225K and W104F—showed 3.0-, 2.8-, 2.5-, 2.0- and 2.0-fold increase of the amidase activity, respectively, compared with the wild type. Four of them were among the top 10 docking predictions. Furthermore, in agreement with computational predictions both T40G and T40A variants showed decreased amidase activity measured experimentally. These substitutions can lead to disruption of described above stabilizing hydrogen bond network in the oxyanion hole and, consequently, loss of activity.

In silico library of nine multiple CALB mutants was constructed for the second round of mutagenesis by combining the best single variants obtained from the first generation. Near-to-attack substrate-binding modes were predicted using molecular docking and resulting enzyme–substrate complexes were used as starting points for MD simulations. Geometry parameters describing the nucleophilic attack trajectory—distances and angles between catalytic residues and substrate—were retrieved from unconstrained runs of MD trajectories. Three MD simulations were calculated for each mutant with different starting velocities and the one representing most stable enzyme–substrate complex (with minimal intermolecular distances) was further considered. Obtained data were used to rank the CALB mutants by their ability to stabilize near-to-attack conformation of the amide (Table III). Variants G39A.T103G.W104[FYQ].L278A and G39A.W104F.L278A were the most efficient in accommodating the substrate while G39A.T103G and G39A.W104F. A225K.L278A showed the worse results comparable to the wild-type lipase. Selected second generation mutants were produced and characterized experimentally. Experimental assay showed good correlation with *in silico* predicted catalytic properties. Amidase activity was improved in mutants that stabilized the near-to-attack conformation during MD runs while no improvement was observed in those mutants where near-to-attack conformation was not observed. Eventually, the best variants G39A.T103G.W104F.L278A and G39A.W104F.L278A showed 11.2- and 6.3-fold improvement of amidase activity compared with the wild-type enzyme.

It has been observed that two multiple mutants showed experimentally measured amidase activity comparable to the wild type while corresponding single mutations provided activity increase. For example, individual mutants G39A and T103G showed 2.8- and 3.0-fold improvement, respectively, but the double mutant G39A.T103G had decreased activity (0.8-fold) what shows that positive effect of one mutation can be destroyed by introduction of other positive mutation. It is yet to be understood how to introduce concerted structural changes to get synergistic effect of multiple positive mutations and exclude negative influence of SSPs on each other. Despite certain progress in understanding influence of SSPs on enzyme function it is still poorly explored area which needs further development. We hope that systematic analysis of SSPs and their role in enzyme catalytic mechanisms will help to establish structure–function relationship.

Discussion

Opportunity to design functional promiscuity of enzymes is one of the most attractive and challenging trials in protein engineering. There is, however, no clear methodology how to change structure of any protein in order to modify its properties.

The α/β-hydrolase fold superfamily is one of the largest groups of functionally diverse enzymes whose homologous members catalyze different reactions but share a conserved structure and common catalytic machinery. Members of this group include dienelactone hydrolases, lipases, thioesterases, serine carboxypeptidases, proline iminopeptidases, haloalkane dehalogenases, haloperoxidases, epoxide hydrolases and others. The fact that lipases are very poor catalysts for hydrolysis of amides whereas proteases with high amidase

Table III. Amidase activity of multiple CALB mutants evaluated *in silico* and by experimental assay

Mutant	Molecular dynamics evaluation							E
	Ligand C(N) = O RMSD, Å	Distance, Å				Angle, °		
		Ser105 γO ··· C	Thr40 N ··· O	Thr40 γ1O ··· O	Gln106 N ··· O	$A_{O\cdots C=O}$	$D_{O\cdots C(N)=O}$	
G39A.T103G.W104F.L278A	0.85 ± 0.33	3.11 ± 0.23	3.34 ± 0.61	2.79 ± 0.22	3.79 ± 0.76	94.99 ± 10.75	94.84 ± 14.73	11.2
G39A.T103G.W104Y.L278A	0.89 ± 0.26	3.28 ± 0.23	3.07 ± 0.22	2.78 ± 0.20	3.57 ± 0.32	91.07 ± 8.06	111.73 ± 11.00	2.8
G39A.T103G.W104Q.L278A	0.93 ± 0.37	3.19 ± 0.25	3.26 ± 0.35	3.49 ± 0.63	3.90 ± 0.44	79.40 ± 10.53	106.16 ± 12.14	1.9
G39A.W104F.L278A	1.01 ± 0.30	3.27 ± 0.22	3.01 ± 0.19	2.88 ± 0.24	3.46 ± 0.29	95.72 ± 8.70	91.40 ± 28.18	6.3
G39A.W104F	1.19 ± 0.49	3.26 ± 0.27	4.73 ± 1.53	5.72 ± 1.00	5.18 ± 1.22	108.53 ± 27.84	91.76 ± 29.41	4.2
G39A.T103G.L278A	1.88 ± 0.55	3.48 ± 0.34	5.39 ± 0.82	4.12 ± 0.70	6.39 ± 0.89	133.55 ± 22.79	89.10 ± 22.43	3.8
G39A.L278A	1.91 ± 0.44	3.34 ± 0.28	5.15 ± 0.64	3.91 ± 0.60	6.11 ± 0.66	126.06 ± 20.27	88.87 ± 13.79	3.3
G39A.T103G	2.29 ± 0.79	3.45 ± 0.56	5.46 ± 1.09	4.10 ± 0.92	6.19 ± 1.01	123.60 ± 25.19	78.94 ± 46.04	0.8
WT	2.67 ± 1.20	4.52 ± 0.96	6.92 ± 1.47	5.18 ± 1.19	7.40 ± 1.29	137.23 ± 23.82	68.67 ± 136.28	1.0
G39A.W104F.A225K.L278A	3.13 ± 0.96	5.34 ± 1.04	6.46 ± 1.03	4.79 ± 1.01	8.10 ± 1.36	148.14 ± 21.18	70.15 ± 61.19	0.8

Experimentally measured amidase activity values are given as a ratio of those for mutant and wild type enzymes. Entries are sorted by declining root mean square deviation (RMSD) of atoms that belong to amide bond of the substrate.
E, Experimentally evaluated amidase activity.

activity have the same α/β-hydrolase fold has been attracting increasing attention in recent years. In this work comparative bioinformatic analysis of α/β-hydrolases was used to identify differences in organization of corresponding active sites with lipase and peptidase activities. Subfamily-specific positions—conserved within subfamilies of lipases and peptidases but different between them—are evolutionary flexible and variable in nature and were supposed to be responsible for functional discrimination between homologs. Therefore, identified positions were used as hot-spots to improve promiscuous amidase activity of CALB. *In silico* library of suggested mutants was constructed and molecular modeling was used to evaluate binding of 2-chloro-*N*-benzylacetamide. Experimental evaluation showed that amidase activity was improved in those mutants that were capable of stabilizing near-to-attack conformation of the substrate in enzyme–substrate complex. The G39A.T103G.W104F.L278A mutant demonstrated productive substrate-binding *in silico* and showed 11-fold increase of experimentally measured amidase activity compared with the wild-type CALB.

Subfamily-specific positions in α/β-hydrolase family can play very different roles. Positions Gly39, Trp104 and Leu278 are part of narrow leaving group-binding cleft in the active site. Mutations of these positions can regulate interactions with the leaving group of amide substrate. Furthermore, G39A was the best single mutation identified by molecular docking experiments which led to better accommodation of benzyl group of the substrate by introducing more favorable hydrophobic contacts. Thr103 is located in the hydrophobic core of the protein between β-strands 3, 5 and 6 of the central β-sheet and α-helix C. Introducing a polar Thr group into the protein core was previously experimentally evaluated to incur an energetic penalty of about 1–3 kcal/mol (Blaber *et al.*, 1993). Despite the stability tradeoff, buried polar residues may play key role in ensuring unique protein topology and fold specificity (Bryson *et al.*, 1998; Bolon and Mayo, 2001). Thr103 does not directly interact with surrounding residues but participates in a hydrogen bond network with the catalytic Ser105 via coordination of a buried water molecule. This network, however, is different in peptidases that have substitution to Gly at position Thr103 as part of a conserved G.X.Nu.X.G nucleophile pattern. Thus, mutation T103G preserves catalytic function and supports orientation of the nucleophilic elbow favorable for amide hydrolysis by introducing hydrogen bond pattern similar to peptidases.

It is important to note that some of subfamily-specific positions identified in this work were previously involved in preparation of CALB mutants. The CALB variant T103G was shown to possess increased thermostability of the mutant (Patkar *et al.*, 1997, 1998). Mutation W104H was previously shown to have increased enantioselectivity in ester hydrolysis (Patkar *et al.*, 1998; Naik *et al.*, 2010).

Mutation of Leu252 in *Pseudomonas aeruginosa* lipase (Ala225 in CALB) to Met increased amidase activity towards *N*-(2-naphthyl)oleamide, whereas the esterase activity was not affected (Nakagawa *et al.*, 2007). In our work, however, mutation A225M showed a very modest 1.1-fold effect. The best mutation at this position A225K showed 2-fold increase. Finally, the role of position Ile189 in CALB has been recently explored using molecular modeling and experimental techniques (Syrén *et al.*, 2012). It has been suggested that substitution at position Ile189 can introduce important hydrogen bond with the substrate to support amide hydrolysis. In particular, mutants I189N and I189Q were shown to have increased ratio of amidase over esterase activity. In our work, however, these variants did not show increased amidase activity to the selected substrate. In general, it can be concluded that important role of some subfamily-specific positions identified by comparative bioinformatic analysis was previously shown to be functionally important for different enzymes.

We have used bioinformatic analysis to select hot spots for rational engineering of promiscuous catalytic activity in

CALB. However, value of subfamily-specific positions should be further explored in order to develop a systematic tool to study structure–function relationship in enzymes and to use this information for rational protein engineering.

Acknowledgements

Molecular docking and dynamics simulations were run on Lomonosov Moscow State University supercomputer clusters 'Chebyshev' and 'Lomonosov'.

Funding

This work was supported as a joint EC-Russian FP7-KBBE-2008-2B project 'IRENE' by the European Commission (Grant Agreement 227279) and the Russian Ministry for Science and Education (Contract 02.740.11.0866).

References

Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) *Nucleic Acids Res.*, **25**, 3389–3402.

Blaber,M., Lindstrom,J.D., Gassner,N., Xu,J., Heinz,D.W. and Matthews,B.W. (1993) *Biochemistry*, **32**, 11363–11373.

Bolon,D.N. and Mayo,S.L. (2001) *Biochemistry*, **40**, 10047–10053.

Bryson,J., Desjarlais,J., Handel,T. and DeGrado,W. (1998) *Protein Sci.*, **7**, 1404–1414.

Carr,P.D. and Ollis,D.L. (2009) *Protein Pept. Lett.*, **16**, 1137–1148.

Darden,T., York,D. and Pedersen,L. (1993) *J. Chem. Phys.*, **98**, 10089–10093.

DePaul,A.J., Thompson,E.J., Patel,S.S., Haldeman,K. and Sorin,E.J. (2010) *Nucleic Acids Res.*, **38**, 4856–4867.

Dolinsky,T.J., Czodrowski,P., Li,H., Nielsen,J.E., Jensen,J.H., Klebe,G. and Baker,N.A. (2007) *Nucleic Acids Res.*, **35**, W522–W525.

Finkelstein,A.V. and Ptitsyn,O. (2002) *Protein Physics: A Course of Lectures*. Boston: Academic Press, pp. 354.

Henke,E. and Bornscheuer,U.T. (2003) *Anal Chem.*, **75**, 255–260.

Holmquist,M. (2000) *Curr. Protein Pept. Sci.*, **1**, 209–235.

Huey,R., Morris,G.M., Olson,A.J. and Goodsell,D.S. (2007) *J. Comput. Chem.*, **28**, 1145–1152.

Humphrey,W., Dalke,A. and Schulten,K. (1996) *J. Mol. Graph.*, **14**, 33–38.

Jorgensen,W.L., Chandrasekhar,J., Madura,J.D., Impey,R.W. and Klein,M.L. (1983) *J. Chem. Phys.*, **79**, 926–935.

Kazlauskas,R.J. and Bornscheuer,U.T. (1998) In Rehm,H.J. and Reed,G. (eds), *Biotechnology: Biotransformation I*, **Vol. 8a**. Weinheim: Wiley-VCH, pp. 37–191.

Krissinel,E. and Henrick,K. (2004) *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2256–2268.

Lee,M.M., Chan,M.K. and Bundschuh,R. (2008) *Bioinformatics*, **24**, 1339–1343.

Naik,S., Basu,A., Saikia,R., Madan,B., Paul,P., Chaterjee,R., Brask,J. and Svendsen,A. (2010) *J. Mol. Catal. B, Enzym.*, **65**, 18–23.

Nakagawa,Y., Hasegawa,A., Hiratake,J. and Sakata,K. (2007) *Protein Eng. Des. Sel.*, **20**, 339–346.

Notredame,C., Higgins,D.G. and Heringa,J. (2000) *J. Mol. Biol.*, **302**, 205–217.

Ollis,D.L., Cheah,E., Cygler,M., *et al*. (1992) *Protein Eng.*, **5**, 197–211.

Olsson,M.H.M., Søndergard,C.R., Rostkowski,M. and Jensen,J.H. (2011) *J. Chem. Theory Comput.*, **7**, 525–537.

Park,S. and Saven,J.G. (2005) *Proteins*, **60**, 450–463.

Patkar,S.A., Svendsen,A., Kirk,O., Clausen,I.G. and Borch,K. (1997) *J. Mol. Catal. B: Enzym.*, **3**, 51–54.

Patkar,S., Vind,J., Kelstrup,E., Christensen,M.W., Svendsen,A., Borch,K. and Kirk,O. (1998) *Chem. Phys. Lipids.*, **93**, 95–101.

Phillips,J.C., Braun,R., Wang,W., *et al*. (2005) *J. Comput. Chem.*, **26**, 1781–1802.

Pleiss,J., Fischer,M., Peiker,M., Thiele,C. and Schmid,R.D. (2000) *J. Mol. Catal. B, Enzym.*, **10**, 491–508.

Pleiss,J., Fischer,M. and Schmid,R.D. (1998) *Chem. Phys. Lipids*, **93**, 67–80.

Radisky,E.S. and Koshland,D.E., Jr. (2002) *Proc. Natl. Acad. Sci. USA*, **99**, 10316–10321.

Rubin,B. and Dennis,E.A., (eds), (1997) *Lipases, Part B: Enzyme Characterization and Utilization. Methods in Enzymol*, **Vol. 286**. San Diego: Academic Press, pp. 3–563.

Schmid,R.D. and Verger,R. (1998) *Angew. Chem. Int. Ed.*, **37**, 1609–1633.

Stroganov,O.V., Novikov,F.N., Stroylov,V.S., Kulkov,V. and Chilov,G.G. (2008) *J. Chem. Inf. Model.*, **48**, 2371–2385.

Suplatov,D.A., Arzhanik,V.K. and Švedas,V.K. (2011) *Acta Naturae*, **3**, 93–98.

Suplatov,D.A., Shalaeva,D.N., Kirilin,E.M., Arzhanik,,V.K. and Švedas,V.K., Bioinformatic analysis of protein families for identification of variable amino acid residues responsible for functional diversity (submitted, 2012).

Syrén,P.O., Hendil-Forssell,P., Aumailley,L., Besenmatter,W., Gounine,F., Svendsen,A., Martinelle,M. and Hult,K. (2012) *Chembiochem*, **13**, 645–648.

Tina,K.G., Bhadra,R. and Srinivasan,N. (2007) *Nucleic Acids Res.*, **35**, W473–W476.

Uppenberg,J., Hansen,M.T., Patkar,S. and Jones,T.A. (1994) *Structure*, **2**, 293–308.

Van Rantwijk,F. and Sheldon,R.A. (2004) *Tetrahedron*, **60**, 501–519.

Zaks,A. and Klibanov,A.M. (1984) *Science*, **224**, 1249–1251.