ELSEVIER

Journal of Molecular Catalysis B: Enzymatic 10 (2000) 491–508

JOURNAL OF MOLECULAR CATALYSIS B: ENZYMATIC

www.elsevier.com/locate/molcatb

Lipase engineering database Understanding and exploiting sequence–structure–function relationships

Jürgen Pleiss, Markus Fischer, Marcus Peiker, Claudia Thiele, Rolf D. Schmid *

Institute of Technical Biochemistry, University of Stuttgart, Allmandring 31, D-70569 Stuttgart, Germany

Received 21 January 2000; received in revised form 21 January 2000; accepted 21 January 2000

Abstract

The Lipase Engineering Database is a WWW-accessible resource on sequence–structure–function relationships of microbial lipases. Sequences of 92 microbial lipases and homologous serine hydrolases were assigned to 32 homologous families and 15 superfamilies. Multisequence alignments of all homologous families and superfamilies are provided. Functionally relevant amino acids are annotated. The catalytic serine is part of the conserved nucleophilic elbow and was identified in all sequences by its conserved signature GxSxG. The complete catalytic machinery (catalytic triad and two oxyanion hole residues) could be annotated in 91% of LED sequence entries. Published mutants and their properties are provided. The X-ray structures of 22 lipases were superposed and consistently annotated. Sequence and structure data were applied to study the role of the first oxyanion hole residues. Although the backbone amides contribute to the oxyanion hole rather than side chains, the residues are well conserved. Two sequence signatures including the first oxyanion hole residue were identified: G**X** and GG**G**X. In the G**X** type, the position of the first oxyanion hole residue **X** is stabilized by one or several anchor residues. If **X** is hydrophilic, it is hydrogen bonded to hydrophilic anchor residues, while hydrophobic oxyanion hole residues bind to hydrophobic pockets. The GG**G**X type includes short chain length specific lipases and carboxylesterases. The first oxyanion hole residue **G** is stabilized by interaction of the dipeptide **G**X with the side chain of the second oxyanion hole residues, which is a conserved alanine as C-terminal neighbour of the catalytic serine. Thus, short chain specific lipases and carboxylesterase can be identified by combining the signatures GGGX and GxSAG. Consistently annotated aligned sequences and superimposed structures of microbial lipases help to understand the functional role of individual amino acids and thus the LED is a useful tool for protein engineering. The Lipase Engineering Database is available at http://www.led.uni-stuttgart.de/.

Keywords: Lipase; Database; Oxyanion hole; Alignment; Superposition; Mutant; Protein engineering

* Corresponding author. Tel.: +49-711-685-3193; fax: +49-711-685-3196.
E-mail address: rolf.d.schmid@rus.uni-stuttgart.de http://www.itb.uni-stuttgart.de/ (R.D. Schmid).

1381-1177/00/$ - see front matter
PII: S1381-1177(00)00092-8

Nomenclature:

The following abbreviations are used for organisms:

Ac. (*Acinetobacter*)
Al. (*Alternaria*)
An. (*Anas*)
As. (*Ascochyta*)
Asp. (*Aspergillus*)
B. (*Bacillus*)
Bo. (*Botrytis*)
C. (*Candida*)
Ca. (*Caenorhabditis*)
Co. (*Colletotrichum*)
Cu. (*Culex*)
D. (*Drosophila*)
Dic. (*Dictyostelium*)
E. (*Escherichia*)
F. (*Fusarium*)
Gal. (*Galactomyces*)
H. (Homo)
Haem. (*Haemophilus*)
Hel. (*Heliothis*)
M. (*Mammalian*)
Ma. (*Magnaporthe*)
Me. (*Mesocricetus*)
My. (*Mycus*)
Myc. (*Mycobacterium*)
Myp. (Mycoplasma)
O. (*Oryctolagus*)
P. (*Pseudomonas*)
Psy. (*Psycrobacter*)
R. (Rattus)
Rh. (*Rhizomucor* and *Rhizopus*)
S. (*Streptomyces*)
Sa. (*Saccharomyces*)
St. (*Staphylococcus*)
Tor. (*Torpedo*)
Vib. (*Vibrio*)
Y. (*Yarrowia*)

Abbreviations for enzymes with known structure:
BTL (*Bos taurus* lipase)
CALB (*Candida antarctica* lipase B)
CRL (*Candida rugosa* lipase)
FSC (*Fusarium solani* cutinase)
GCL (*Geotrichum candidum* lipase)
HLL (*Humicola lanuginosa* lipase)
PCL (*Pseudomonas cepacia* lipase)
PeCL (*Penicillium camembertii* lipase)
RDL (*Rhizopus delemar* lipase)
RML (*Rhizomucor miehei* lipase)
ROL (*Rhizopus oryzae* lipase)
SAB (*Streptomyces aureofaciens* bromoperoxidase)
SEL (*Streptomyces exfoliatus* lipase)
TCA (*Torpedo californica* acetylcholinesterase)

1. Introduction

Microbial lipases (triacylglycerol hydrolases, E.C. 3.1.1.3) are widely used to catalyse hydrolysis, alcoholysis, esterification and transesterification of triacylglycerols and other water insoluble esters, due to their broad substrate specificity, high regio- and stereoselectivity (for review: see Refs. [1,2]). Although all lipases catalyse the same reaction, their sequences vary widely, with the exception of a well-conserved pentapeptide, the GxSxG motif. Only when structures became available of the free enzyme [3–7] and in complexes with substrate analogous inhibitors [8–12], structural similarities became visible: all lipases belong to the same fold family, the α/β hydrolases [13], which consists of a central hydrophobic β-sheet, and amphiphilic helices packed to it; the catalytic machinery consists of a catalytic triad, three residues which are placed at highly conserved geometry in the loops on one side of the sheet (Ser in the *nucleophilic elbow* after strand 5, His after strand 8, Asp/Glu after strand 7), and an oxyanion hole formed by two backbone amides of a residue in the N-terminal region of the lipase and the C-terminal neighbour of the catalytic serine. In most lipases, a mobile element, the lid, covers the catalytic site in the inactive form of the lipase. It consists of one or two short α-helices linked to the body of the lipase by flexible structure elements. In the open, active form of the lipase, the lid moves away and makes the binding site accessible to the substrate.

Data have been accumulated on the role and the interplay of catalytic triad, oxyanion hole, lid and other residues, which are involved in catalytic activity and selectivity of lipases. The catalytic triad seems to be stabilized by a hydrogen-bonded layer of residues [14]; its geometry is not influenced by opening of the lid [7,8]. In lipases from filamentous fungi and *Candida rugosa* lipase, the oxyanion hole is already preformed in the closed form [15] and does not change upon opening of the lid, while in *Pseudomonas*, it is only formed upon opening of the lid [7], thus, indicating a tight interaction of lid and oxyanion hole residues. The first oxyanion hole residue in the N-terminal region is well

conserved inside homologous families. In the lipases from filamentous fungi, it is a serine or threonine and its side chain is postulated to add a third hydrogen to the stabilization of the oxyanion [8]. Mutagenesis experiments showed that exchanging this residue by a hydrophobic one results in loss of activity [14,16,17]. For lipases from *Pseudomonas*, the first oxyanion hole residue is hydrophobic (leucine, methionine or phenylalanine), therefore its side chain is not directly involved in oxyanion hole stabilization. However, this residue is also well conserved and thus seems to be functionally important, but its role is not yet fully understood. Other residues, which mediate catalytic activity or substrate binding have been identified by protein engineering [16,18,19].

Published data on the role of individual residues is accessible in annotated sequence databases, the number of which is steadily increasing. Sequence databases like EMBL Nucleotide Sequence Database [20], Genbank [21], Swiss-Prot [22] and PIR [23] are archiving information on nucleotide and protein sequences. More specialized databases use these databases as primary data set, e.g. by classifying sequences into protein families [24,25]. Information on protein structure, as provided by the Protein Data Bank [26], is used for structural classification of proteins [27,28] and for superposing protein families [29]. Some databases assemble sequence information on a single protein family such as cholinesterases [30].

To integrate the different aspects of sequence and structure of microbial lipases, we have set up the Lipase Engineering Database, which integrates aligned sequences, superposed structures and annotated mutation information. It covers all microbial lipases and is consistently annotated. Comparison of aligned and classified sequences revealed new conserved patterns. By superposing the available lipases structures, these patterns could be associated with different architectures of oxyanion holes and revealed new aspects on the role and functionality of these oxyanion hole residues.

2. Methods

2.1. Obtaining protein sequences

Sequences of microbial lipases and homologous serine hydrolases were taken from the Swiss-Prot sequence database [22]. As first step, the Swiss-Prot sequence database was searched for the keywords , ‘triglyceride’ and ‘triacylglycerol’. Identical sequences, as determined by multisequence alignment with ClustalX 1.5b [31], and sequence fragments were not included. These sequences were used as templates for BLAST searches [32]. From the BLAST search results list, only sequences with a ‘smallest sum probability’ of more than 0.1 were selected. Additional BLAST searches were performed using the newly found sequences as templates.

2.2. Classification

In the first round, superfamilies were defined by all sequences found by iteration of the BLAST search for one template sequence. These were subdivided into homologous families, which were separated by a decrease of the ‘smallest sum probability’ of the BLAST search. This classification was verified by multisequence alignment of the homologous family. Sequences without a catalytic triad were removed, since they were supposed to have no esterase activity.

In the second round, a multisequence alignment for each superfamily was created with one representative sequence of each homologous family. This alignment was used to refine the classification of the superfamilies: all sequences which belong to one superfamily should align with the catalytic triad at the same position. A homologous family which did not fulfil this condition was assigned to a different superfamily.

2.3. Protein structure data

Protein structures were obtained from the Protein Data Bank, except for the structures of

lipases from *Penicillium camembertii* and *Rhizopus delemar* which were obtained from the EMBL-EBI MaxSprout server [33]. Secondary structure information were obtained from DSSP database [34]. Structures were superposed using SYBYL (Tripos, St. Louis, MO). In a first round, the following atoms were used as reference: C_α atoms of a pentapeptide with the catalytic serine in the center (e.g. G142–G146 for catalytic S144 in *Rh. miehei* * lipase), of the catalytic histidine and its nearest neighbours (e.g. D256–L258 in *Rh. miehei* lipase for H257 as catalytic histidine), and of the catalytic aspartic or glutamic acid (e.g. D203 in *Rh. miehei* lipase), as well as the amide hydrogen atom of the first oxyanion hole residue (S82 in *Rh. miehei* lipase). Based on this superposition, the most structurally conserved four residues of the central β-strand were identified (e.g.Y172-G175 in *Rh. miehei* lipase). In a second round, the superposition was refined taking as reference the C_α atoms of these four residues, two additional residues of the catalytic serine-containing β-strand (e.g. V140, T141 in *Rh. miehei* lipase), and the set of atoms of the first round. For the investigation of the oxyanion hole, the superposed structures were analyzed using SYBYL.

2.4. Design of LED-online

To integrate information on sequence, structure and function, four properties are stored and tightly linked: (1) Sequence alignment identifies residues which are conserved among a homologous family or a superfamily. (2) Superposed structures allow to study sites in the lipases which have no sequence similarity, but are structurally conserved. Sequence and structure are consistently annotated and linked. (3) A table of published mutations and their effect on function helps to understand the role of individual residues. (4) Links to other WWW resources, like homologous sequences and protein topology, integrate LED into the ensemble of WWW accessible biological resources.

All information on lipases provided by the LED is filed in static HTML pages and thus can be accessed using a WWW browser. The user can enter the LED at three different tables: sequence, structure or list of mutations. The sequence table includes information on homologous families and on enzyme function for all members of each family, and links structures to the respective Swiss-Prot entry. Sequences of each homologous family and representative sequences of each superfamily were aligned and annotated. Annotations in sequence alignments were derived from Swiss-Prot or, if available, PDB entries. If similarity was sufficient, information was transferred to all members of a superfamily.

The structure table includes superposed, annotated and downloadable structure entries, links to the respective PDB entries, references, and annotations of functionally important residues. The following sets of amino acids were annotated: (1) BML: binding region for medium sized substituents of secondary alcohols [35], (2) CAT: residues of the catalytically active Ser–Asp/Glu–His triad, (3) LID: residues forming the lid were determined by superposition of open and closed structures. (4) M## (## = 01,...,99): mutations described in literature, (5) OXY: oxyanion hole residues, (6) S## (## = 01,...,99): disulfide bridges obtained from Swiss-Prot and PDB. LED-online is accessible at http://www.led.uni-stuttgart.de.

3. Results

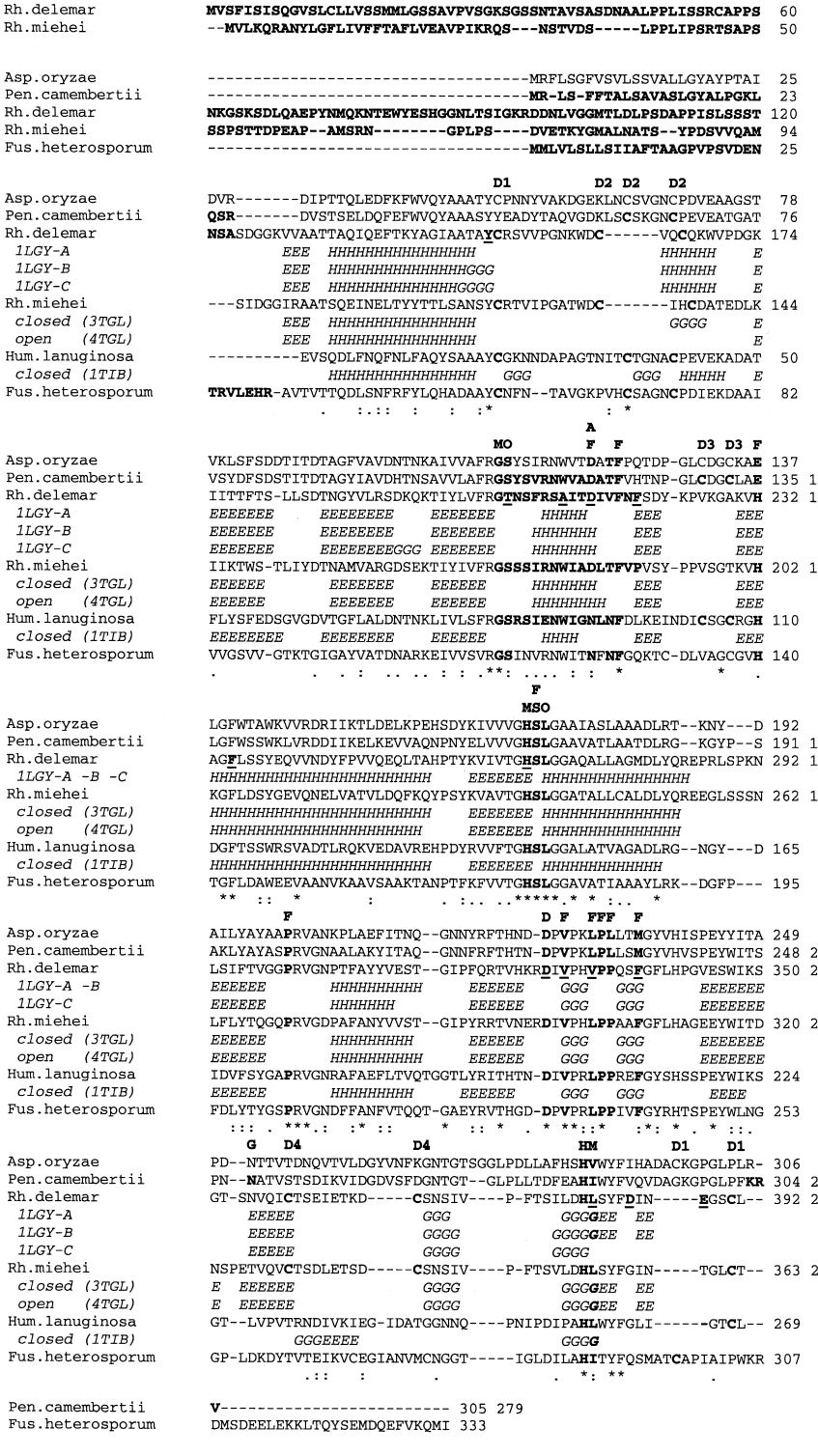
3.1. LED: sequences and structures of microbial lipases

In total, 92 sequences of microbial lipases and homologous serine hydrolases were assembled and aligned. They were grouped in 15 superfamilies and 32 homologous families (Table 1 and http://www.led.uni-stuttgart.de/

Table 1
Classification of lipases in superfamilies and homologous families; catalytic triad in bold letters, oxyanion hole residues underlined; number of Swiss-Prot entries in parentheses
(1)–(9): Structures available; PDB entries are (1) 1LBS, 1LBT, 1TCA, 1TCB, 1TCC; (2) 1CRL, 1CLE, 1LPO, 1LPP, 1LPM, 1LPN, 1LPS, 1TRH; (3) 1AKN, 1AQL, 2BCE; (4) 1ACJ, 1ACK, 1ACL, 1AMN, 2ACE, 1VOT; (5) 1AGY, 1CEX, 1CUS, 1OXM, 1XZK, 1XZM, 2CUT; (6) 1TGL, 3TGL, 4TGL, 5TGL, 1LGY, 1TIA, 1TIB, 1TIC; (7) 1JFR; (8) 1A88, 1A8S, 1A8Q, 1BRT, 1BRO, 1A8U, 1A7U; (9) 1OIL, 2LIP, 3LIP, 4LIP, 5LIP, 1CVL, 1QGE, 1TAH; (10) sequence taken from PDB entry 1JFR.

Superfamilies	Homologous families	Source	Oxyanion hole	Ser	Asp/Glu	His	Swiss-Prot
Acinetobacter calcoaceticus	Ac. esterase	*Ac. calcoaceticus*		147 GD**S**CG			EST_ACICA
Bacillus subtilis	B. lipase	*B. subtilis*		106 AH**S**MG			LIPA_BACSU
Candida antarctica	C. antarctica lipase B[1]	*C. antarctica*	61 LVPGTGT	128 TW**S**QG	210 AT**D**EI	247 ID**H**AG	LIPB_CANAR
Candida rugosa	C. rugosa lipase[2]	*C. rugosa* (5), *Gal. geotrichum* (2)	134 WIFGGGF	222 GE**S**AG	354 ND**E**GT	462 TF**H**SN	LIP1_CANRU
Carboxylesterases	A. oxidans hydrolase	*A. oxidans*	106 WIHGGGL	186 GQ**S**GG	305 RD**E**GT	400 AV**H**CI	PCD_ARTOX
	B. subtilis esterase	*B. subtilis*	102 WIHGGAF	187 GE**S**AG	308 RD**E**GY	397 AF**H**AL	PNBA_BACSU
	Ca. elegans cytoplasmic esterase	*Ca. elegans*	121 YIHGGGY	205 GQ**S**AG	328 EY**E**GL	443 AV**H**CT	EST2_CAEEL
	Ca. elegans gut esterase	*Ca. briggsae*, *Ca. elegans*	113 WVHGGGY	197 GY**S**AG	318 NK**E**GS	449 SP**H**AN	EST1_CAEBR
	Cu. pipiens esterase	*Cu. pipiens*	105 YIYGGGF	189 GH**S**AG	322 SE**E**GL	440 TA**H**AD	EST1_CULPI
	Dic. esterase	*Dic. discoideum* (2)	133 FIPGGAF	213 GE**S**AG	338 QD**E**AI	441 VC**H**GT	CRYS_DICDI
	Hel. virescens esterase	*Hel. virescens*	136 FIHGGGF	218 GQ**S**AG	349 SS**E**CE	463 VG**H**IE	ESTJ_HELVI
	M. bile salt activated lipase[3]	*R. norvegicus*, *M. musculus*, H. sapiens, Bos taurus	123 WIYGGAF	209 GE**S**AG	338 DM**D**GH	453 AD**H**AD	BAL_RAT
	M. carboxylesterase	*R. norvegicus* (5), *M. musculus* (3), *O. cuniculus* (2), *Me. auratus*, *H. sapiens*, *An. platyrhynchos*	138 WIHGGGL	219 GE**S**AG	338 KQ**E**FG	451 GD**H**GD	EST1_RAT
	My. persicae esterase	*My. persicae*	131 HIHGGGY	212 GM**S**AG	337 QD**E**GL	461 PT**H**GD	ESTE_MYZPE
	Pseudoobscura esterase	*D. pseudoobscura* (3), *D. melanogaster* (2), *D. simulans*, *D. mauritiana*, *D. virilis*	124 QIHGGAF	205 GH**S**AG	336 TE**D**GG	465 TV**H**GD	ESTB_DROPS
	Cholinesterases[4]	*Tor. californica*	135 WIYGGGF	219 GE**S**AG	346 KD**E**GS	459 VI**H**GY	ACES_TORCA

Cutinases	Fusarium solani[5]	*F. solani*	54 YARGSTE	134 GY**S**QG	189 VG**D**LV	202 AP**H**LA	CUTI_FUSSO
	Colletotrichum gloeosporioides	*Co. gloeosporioides, Co. capsici, Ma. grisea, Al. brassicicola, As. rabiei, Asp. oryzae, Bo. cinerea*	53 FARASTE	134 GY**S**QG	189 IA**D**AV	202 PA**H**FL	CUTI_COLGL
	Mycobacterium tuberculosis	*Myc. tuberculosis*	36 FARGTGE	108 GY**S**QG	175 PT**D**PI	189 SG**H**ID	CUT3_MYCTU
Filamentous fungi	Rhizomucor lipase[6]	*Rh. miehei Rh. delemar, Pen. camembertii, Hum. lanuginosa*	172 VFRGSSS	236 GH**S**LG	295 ER**D**IV	249 LD**H**LS	LIP
	Saccharomyces YJ77 lipase	*Sa. cerevisiae*	98 AFRGSTT	179 GH**S**LG	251 TG**D**YI	313 YE**H**RA	YJ77_YEAST
Haemophilus influenzae	Haemophilus lipase	*Haem. influenzae, E. coli*		117 GH**S**MG		264 SG**H**WV	Y193_HAEIN
Moraxella1	Moraxella lipase 1[7]	*Moraxella* sp., *S. exfoliatus*	59 ISPGFTA	131 GH**S**MG	175 DG**D**TV	207 AS**H**FT	-[10]
Moraxella2	Moraxella lipase 2	*Moraxella* sp.		237 GD**S**AG			LIP2_MORSP
Mycoplasma	Mycoplasma lipase	*Myp. pneumoniae* (3), *Myp. genitalium* (3)	26 FLHGFGS	95 GH**S**MG	210 SN**D**IV	239 VG**H**SP	ESL2_MYCPN
	Moraxella lipase 3	*Moraxella* sp., *Psy. immobilis*	72 LIHGFGG	140 GN**S**MG	262 DK**D**QV	290 VG**H**VP	LIP3_MORSP
	Non-haem peroxidase[8]	*P. pyrrocinia, S. lividans, P. fluorescens, S. aureofaciens* (2)	26 FHHGWPL	94 GH**S**TG	226 ED**D**QI	255 YS**H**GM	PRXC_PSEPY
Pseudomonas	Cepacia lipase[9]	*P. aeruginosa, Vib. cholerae, P. fragi, P. cepacia, P. glumae*	*57 LVHGLSG*	*129 GH**S**QG*	*306 QN**D**GL*	*328 WN**H**LD*	*LIP _ BURCE*
	Staphylococcus lipase	*St. epidermidis, St. hyicus, St. aureus*	315 LVHGFNG	416 GH**S**MG	607 EN**D**GL	646 WD**H**VD	LIP_STAEP
Pseudomonas fluorescens	Pseudomonas fluorescens lipase	*P. fluorescens* (2)		204 GH**S**LG			LIPA_PSEFL
Saccharomyces cerevisiae	Saccharomyces TGL2 lipase	*S. cerevisiae*		142 AH**S**MG			TGL2_YEAST
Yarrowia lipolytica	Yarrowia lipolytica lipase	*Y. lipolytica*	101 WIHGGGN	191 GE**S**AG	301 IV**D**GT	390 CH**H**AV	LIP1_YARLI


Rh.delemar
Rh.miehei
Asp.oryzae
Pen.camembertii
Rh.delemar
Rh.miehei
Fus.heterosporum
D1
D2 D2
D2
Asp.oryzae
Pen.camembertii
Rh.delemar
1LGY-A
1LGY-B
1LGY-C
Rh.miehei
closed (3TGL)
open (4TGL)
Hum.lanuginosa
closed (1TIB)
Fus.heterosporum
A
MO
F F
D3 D3 F
Asp.oryzae
Pen.camembertii
Rh.delemar
1LGY-A -B -C
Rh.miehei
closed (3TGL)
open (4TGL)
Hum.lanuginosa
closed (1TIB)
Fus.heterosporum
F
MSO
Asp.oryzae
Pen.camembertii
Rh.delemar
1LGY-A -B
1LGY-C
Rh.miehei
closed (3TGL)
open (4TGL)
Hum.lanuginosa
closed (1TIB)
Fus.heterosporum
F
D F FFF F
G D4
D4
HM
D1 D1
Asp.oryzae
Pen.camembertii
Rh.delemar
1LGY-A
1LGY-B
1LGY-C
Rh.miehei
closed (3TGL)
open (4TGL)
Hum.lanuginosa
closed (1TIB)
Fus.heterosporum
Pen.camembertii
Fus.heterosporum

seq/seq.html). Nine homologous families contain members with known X-ray structure (free or in complex with an inhibitor molecule). A total of 53 structure data sets for 20 microbial lipases and homologous serine hydrolases have been superposed and annotated http://www.led.uni-stuttgart.de/struc/struc.html).

Multisequence alignments (Fig. 1 for the homologous family Rhizomucor lipase) for superfamilies and homologous families are annotated by information on catalytically relevant residues (catalytic triad, oxyanion hole), disulfide bridges, scissile fatty acid and alcohol binding residues and secondary structure information. Data on published mutants and their properties are provided (Table 2 and http://www.led.uni-stuttgart.de/mut/mucorales.html for homologous family Rhizomucor lipase). Mutations are consistently annotated in superposed structures.

In nearly all lipases, there is a common pattern G–X_1–S–X_2–G for the nucleophile serine S and the oxyanion hole residue X_2; exceptions are: T–W–S–Q–G in *C. antarctica* lipase B and A–H–S–M–G in lipase from *Bacillus subtilis* and *Saccharomyces cerevisiae*. In all homologous families, the catalytically active serine could be assigned to at least one member using annotation by Swiss-Prot, and transferred to all lipases of the homologous family. Annotation of the catalytically active serine was verified for eight superfamilies (Candida antarctica, Candida rugosa, Carboxylesterases, Cutinases, Filamentous fungi, Moraxella 1, Mycoplasma, Pseudomonas) by known X-ray structures, and for the remaining seven superfamilies by local homology to superfamilies with known structure.

The catalytic histidine and aspartic or glutamic acid, however, were more difficult to find, since they are not part of conserved patterns, but could only be identified by mutation experiments, by structure determination or by homology to hydrolases with known catalytic triad. Similarly, the first of the oxyanion hole residue and the lid residues could only be identified by structure determination or by homology to a lipase with known structure. For the eight superfamilies with known structures, these residues were identified and annotated consistently by multisequence alignments. Additionally for the superfamily Yarrowia lipolytica, the catalytic triad and the first oxyanion hole residue were annotated by local similarity and verified by a structure model obtained from the 3DCrunch database [36]. Thus the catalytic triad and the oxyanion hole residues were annotated for nine superfamilies, containing 26 homologous families and 84 sequences, which corresponds to 81% of homologous families and 91% of LED sequence entries.

3.2. LED in action: understanding the role of the side chains of the oxyanion hole residues

*3.2.1. Two types of oxyanion holes: G**X** and GG**G**X*

The oxyanion hole consists of two residues, which donate their backbone amide protons to stabilize the substrate in the transition state. One residue is located in the structurally conserved nucleophilic elbow. As a consequence, its backbone amide is positioned identically in all lipases. In contrast, the other oxyanion hole residue is not located in a region with conserved sequence and structure, but in a loop on top of the central β-strand 3, according to the numbering scheme introduced by Ollis et al. [13]. In all

Fig. 1. Annotated LED multisequence alignment for the Rhizomucor lipase homologous family (http://www.led.uni-stuttgart.de/seq/filamentous_fungi/rhizomucor/align.html). Relevant residues are annotated (**bold**): catalytic triad (S: serine, D: aspartic acid, H: histidine), oxyanion hole (O), anchor residue (A), disulfide bridges (D1,D2,D3,D4), scissile fatty acid binding residues (F), alcohol binding residues (M), mutations (underlined), signal, propeptide, ER retention signal and lid; secondary structure information (*italic*) is calculated from PDB entries using DSSP [34] (H: helix, E: strand); conserved amino acids are marked (.:*).

Table 2
Mutations of lipases from *Rh. delemar* and *Rh. oryzae*, and their effect on chain length specificity and catalytic activity (http://www.led.uni-stuttgart.de/mut/mucorales.html)
s.f.a.: Scissile fatty acid.

Source	Mutation	Structural effect of the mutation	Experimental observation	Reference
Rh. oryzae	Y28F	open conformation destabilized	0.02% activity	[14]
Rh. oryzae	T82A	stabilization of oxyanion hole by anchor residue gets lost	0.04% activity	[14]
Rh. oryzae	T82V	stabilization of oxyanion hole by anchor residue gets lost	0.14% activity	[14]
Rh. oryzae	T82S	stabilization of oxyanion hole by anchor residue gets lost	12% activity	[14]
Rh. oryzae	A88W	geometry of tetrahedral intermediate less favourable	56% activity	[14]
Rh. oryzae	D91N	stabilization of oxyanion hole by anchor residue gets lost	7% activity	[14]
Rh. oryzae	F95Y	blocks binding at C12 of s.f.a.	60% (30%) increase of caproic acid methyl ester relative to oleic (stearic) acid	[43]
Rh. delemar	F95D	hydrophilic at C12 position of s.f.a.	2-fold decrease of hydrolysis	[16]
Rh. delemar	F95D F214R	right wall of hydrophobic crevice becomes hydrophilic (C10–C16 of s.f.a.)	3-fold increase of tricaprylin relative to triolein	[44]
Rh. delemar	F112W	blocks binding at C6 of s.f.a.	50% increase of tributyrin relative to triolein	[16]
Rh. delemar	F112Q	hydrophilic at C6 position of s.f.a.	no activity	[44]
Rh. oryzae	H143F	breakdown of active site geometry	0.04% activity	[14]
Rh. oryzae	H143S	active site geometry less stabilized	2.2% activity	[14]
Rh. oryzae	D203A	breakdown of active site geometry	0.06% activity	[14]

Rh. delemar	V206T	hydrophilic at C4 position of s.f.a.	10–20% activity	[16]
Rh. delemar	V209W	blocks binding at C10 of s.f.a.	2-fold increase of tributyrin relative to triolein	[16]
Rh. delemar	V209W F112W	blocks binding at C10 and C6 of s.f.a.	80-fold increase of tributyrin relative to tricaprylin; no triolein hydrolysis	[44]
Rh. oryzae	F214Y	blocks binding at C16 of s.f.a.	20% increase of caproic acid methyl ester relative to oleic and stearic acid	[43]
Rh. oryzae	L258A	increased size of His gap	Shifts stereoselectivity for *sn*-2 substituted triacylglycerols towards *sn*-1	[45]
Rh. oryzae	L258S	increased size of His gap	Shifts stereoselectivity for *sn*-2 substituted triacylglycerols towards *sn*-1	[45]
Rh. oryzae	L258F	decreased size of His gap	Shifts stereoselectivity for *sn*-2 substituted triacylglycerols towards *sn*-3	[45]
Rh. oryzae	L258F L254A	decreased size of His gap	Shifts stereoselectivity for *sn*-2 substituted triacylglycerols towards *sn*-3	[45]
Rh. oryzae	L258F L254V	decreased size of His gap	Shifts stereoselectivity for *sn*-2 substituted triacylglycerols towards *sn*-3	[45]
Rh. oryzae	L258F L254F	decreased size of His gap	Shifts stereoselectivity for *sn*-2 substituted triacylglycerols towards *sn*-3	[45]
Rh. oryzae	D262G	not essential for function	96% activity	[14]
Rh. oryzae	E265Q	no influence on active site geometry	98% activity	[14]
Rh. oryzae	E265D	breakdown of active site geometry	0.14% activity	[14]

lipases, a glycine is conserved which contacts the nucleophilic elbow. Representative structures from six of the nine superfamilies with annotated oxyanion hole (Candida antarctica (CALB), Cutinases (FSC), Filamentous fungi (RML), Moraxella 1 (SEL), Mycoplasma (SAB), Pseudomonas (PCL)) showed that the oxyanion hole is formed by the amide backbone of the C-terminal neighbour **X** of this conserved glycine. We termed this type of oxyanion hole 'G**X** type', with **X** being the oxyanion hole residue. While the position of the backbone amide is conserved in all six lipase structures, the side chain of **X** differs. Multisequence alignments of these superfamilies and the respective homologous families revealed that the type of amino acid **X** is conserved inside the superfamilies: hydrophilic in Candida antarctica (T), Cutinases (S, T, R), Filamentous fungi (S, T), and hydrophobic in Moraxella 1 (F), Mycoplasma (F, W), Pseudomonas (L, F, M).

The superfamilies Candida rugosa, Carboxylesterases and Yarrowia lipolytica differ from these lipases. Structure superposition of representative structures (CRL, GCL, BTL, TCA) revealed that the oxyanion hole residue is shifted by one position toward the C-terminus as compared to the G**X** type (Fig. 2), and is glycine and alanine in CRL/GCL/TCA and BTL, respectively. The oxyanion hole forming residue is followed by a conserved hydrophobic residue X. We termed this type the 'GG**G**X type'. Investigation of the multisequence alignments of these superfamilies and the respective homologous families revealed that only in the case of the mammalian carboxylesterase from *Mus musculus* the oxyanion hole residue is formed by a serine instead of a glycine or

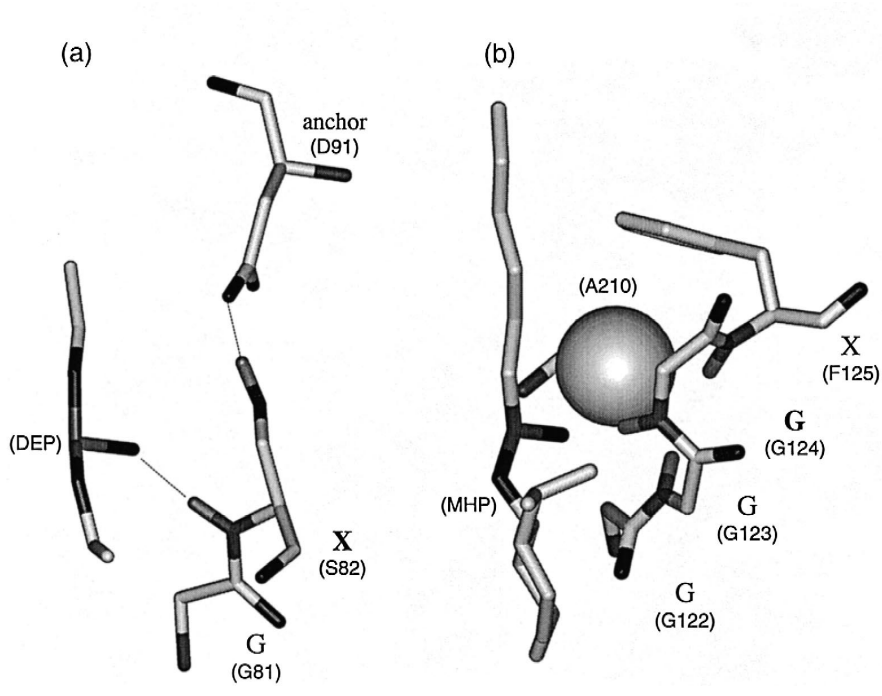


Fig. 2. Illustration of the two types of oxyanion holes. (a) G**X** type in RML (PDB entry 4TGL): stabilization of substrate analogous inhibitor diethylphosophate (DEP) by hydrogen bonding to first oxyanion hole residue **X** (S82); stabilization of S82 by hydrogen bond to the anchor residue D91. (b) GG**G**X type in CRL (PDB entry 1LPM): stabilization of substrate analogous inhibitor (1*R*)-menthyl hexyl phosphonate by the first oxyanion hole residue **G** (G124); stabilization of the oxyanion by locking side chain of A210 between G124 and the side chain of X (F125).

alanine. In most cases, the hydrophobic residue X is F, L or Y, in two cases M or W. The exception to this rule is the lipase from *Yarrowia lipolytica*, where X is N. Although the structure of the oxyanion hole residue is different for the G**X** and the GG**G**X type, the position of the hydrogen atom is similar, and therefore its function is conserved (Fig. 2).

*3.2.2. Role of the oxyanion hole side chain in the G**X** type*

What is the role of the oxyanion hole side chain **X** in the G**X** type? While the side chain is generally not involved in interaction with the substrate, it is well conserved inside a superfamily. In four G**X** superfamilies with known structure of the open form of the lipase (CALB, FSC, RML, PCL), it interacts with at least one side chain, which we termed the 'anchor residue' (Fig. 2a). In PCL, the hydrophobic anchor residue L167 is at 3.6 Å distance from the hydrophobic oxyanion hole residue L17. In CALB, FSC and RML, where the oxyanion hole residue is hydrophilic (T40, S42, S82, respectively), the closest anchor residue (Q157, N84, D91, respectively) is in hydrogen bonding distance (Table 3). Oxyanion hole residue and anchor residue line the bottom of the deep hydrophobic cleft to which the scissile fatty acid binds [37]. In the non-haem bromoperoxidase from *Streptomyces aureofaciens*, the anchor is formed by a group of residues. The side chain of the oxyanion hole residue F32 is fixed by a hydrophobic binding pocket formed by the residues P33, F61, F163, A198 and A201. An exception of this concept is the lipase from *S. exfoliatus*. This lipase is lacking the typical lid and the hydrophobic oxyanion hole residue F63 is exposed to the protein surface. There is no hydrophobic binding pocket that is fixing the side chain and only a weak contact to L93 can be observed. Instead, the side chain of F63 is pointing toward the hydrophobic substrate interface, and thus would be stabilized.

Like the oxyanion hole residues, the anchor residues of FSC, RML and PCL are conserved in the respective superfamilies: hydrophobic for the Pseudomonas superfamily (L,V, I; N as one exception), hydrophilic for the Filamentous fungi (D, N) and Cutinase superfamily (N, F, S, D).

For RML and PGL, the latter being highly homologous to PCL, the structure of the closed form of the lipase has been experimentally determined. In this inactive form, the contact between the oxyanion hole and the anchor residue does not exist. In the closed form of RML, the side chain of the oxyanion hole residue S82 is fixed by a hydrogen bond to the backbone amide of S84, while the distance between the oxyanion hole residue and the anchor residue is 9.3 Å. Compared to the open form, the backbone amide hydrogen of the oxyanion hole residue moves away from the substrate binding site by 0.6 Å. Thus, this pre-formed oxyanion hole is no longer optimal for oxyanion stabiliza-

Table 3
Oxyanion hole residue and anchor residue in the G**X** type oxyanion holes. Possible variations for oxyanion hole and anchor residue in superfamily in parentheses

Enzyme	Oxyanion hole	Anchor	Distance (Å)	Contact
RML	S82 (S,T)	D91 (D,N)	2.5	H-bond
PCL	L17 (L,M,F)	L167 (L,I,V,N)	3.6	hydrophobic
CALB	T40	Q157, Q106, L73	3.0, 3.1, 3.8	H-bond H-bond hydrophobic
FSC	S42 (S,T,R)	N84, Q121 (N,S,D,F)	3.0, 3.0	H-bond H-bond
SAB	F32 (F,W)	hydrophobic binding pocket (P33, F61, F163, A198, A201)		hydrophobic
SEL	F63 (F,Y)	only weak contact to L93		

tion, even though the S82 backbone amide is orientated toward the substrate binding site.

In the closed form of PGL, which is an example for oxyanion holes formed only upon opening of the lid, the distance between the oxyanion hole residue side chain L17 and the anchor L167 increases to 6 Å, compared to the open form of PCL. In the closed form, L17 interacts with L149 (distance 3.6 Å), which is located on the lid. This interaction causes a rotation of the backbone amide of L17 away from the oxyanion hole and results in a distance of 2.1 Å between the amide hydrogen atoms in closed and open form.

Of the six lipases of the G**X** type, five (PCL, CALB, FSC, SAB, SEL) have the same topology (central strand 4–3–5–6) [13], while RML has a different topology (2–3–4–5). The position of the closest anchor residue differs between these lipases. In PCL and CALB it is located at the C-terminus of helix D'_3 and helix D'_2, respectively. Both helices are located before the crossover helix D. In RML and FSC, it is located at the C-terminus of helix B'_1. In RML, this helix forms the lid and is followed by B'_2, the helix before the crossover helix B. In FSC helix B'_1 is structurally conserved to helix B'_1 in RML, but helix B'_2 is missing.

*3.2.3. Stabilizing the oxyanion hole in the GG**G**X type*

The stabilization of the oxyanion hole structure in the GG**G**X type differs from the G**X** type. In addition to the well-conserved oxyanion hole residue **G**, its C-terminal hydrophobic neighbour X is also highly conserved. Comparison of representative structures for the GG**G**X type superfamilies Candida rugosa and Carboxylesterase showed that the second oxyanion hole residue, the C-terminal neighbour of the catalytic serine, is a conserved alanine, which stabilizes the oxyanion hole. Like a spring lock, the side chain of this alanine is embraced by the backbone amide groups of the oxyanion hole forming glycine **G** and its hydrophobic neighbour X, and the side chain of X (Fig. 2b). Sequence comparison of these superfamilies shows that for all but three carboxylesterases the second oxyanion hole residue is a conserved alanine (exceptions: G, S, T for lipases from *Arthrobacter oxidans* (PCD_ARTOX), *M. musculus* (ESTN_MOUSE), *Drosophila virilis* (ESTS_DROVI), respectively).

Comparison of the open structures of CRL, BTL and TCA also revealed that the dipeptide **G**X is placed in a β-loop, which is structurally similar in open and closed conformation. The conformation of the dipeptide GG, however, changes upon activation of the lipases. In the open form, it exists either of two conformations: the backbone amide of the N-terminal neighbour (G123 in CRL) of the oxyanion hole glycine **G** (G124 in CRL) stabilizing the oxyanion by hydrogen bonding [38,39], or being orientated away from the substrate toward a bound water molecule [10]. In the closed form, it is in a stabile β-strand conformation, with the amide hydrogen of G123 pointing away from the oxyanion hole.

4. Discussion

4.1. Annotation and classification

Information on protein sequences as well as number and size of sequence databases is rapidly growing. In addition to pure sequence information, annotated sequence databases like Swiss-Prot include information like similarities to other proteins, annotation of catalytic residues and published mutants. However, for each new entry, annotation has to be performed manually by the database curators, thus becoming the bottleneck in maintaining annotated databases: while the automatically translated EMBL database consists of 225,165 entries (Nov. 1999), the fully annotated Swiss-Prot database includes only 82,229 entries (Nov. 1999) [22].

The strength of databases like Swiss-Prot are their broad spectrum of provided biological in-

formation and their standardized nomenclature. On the other hand, the availability of Web-based resources led to a multitude of biological databases, which provide additional information on proteins like databases on classification of sequences (SYSTERS [25] and Pfam [24]). Query systems like SRS [40] or Entrez [41] have proved to be of great help in integrating heterogenous databases in a common query system.

While the goal of all these databases is to cover all sequences, LED focuses on one family, the microbial lipases, with the goal to provide all information which is relevant to understand sequence–structure–function relationship. A similar database, ESTHER [30], focuses on sequences of cholinesterases and provides alignments and mutant information on this class of enzymes.

In LED, based on global multisequence alignments, all lipase sequences were classified into homologous families and superfamilies, to detect similarities and variations. This classification was compared with SYSTERS [25], generating sequence clusters by iterated BLAST searches, and the protein family database Pfam [24], classifying proteins by Hidden Markow Models based on manually improved multisequence alignments for various protein families. The SYSTERS clusters or cluster subfamilies generally correspond to homologous families in LED. Differences were found for two homologous families (Mycoplasma lipase and Rhizomucor lipase), where LED merges different SYSTERS clusters, and the superfamily Candida rugosa, which in SYSTERS is a subfamily of the cluster corresponding to the LED superfamily Carboxylesterases. This is caused by the fact that our classification was based on multisequence alignments of the complete sequences rather than on local similarity and observed the additional criterion that the position of the catalytic triad must be conserved inside one superfamily. The Pfam classification is based on protein domains, thus a single enzyme can be a member of several domain families, which is very useful for comparing proteins with only distant similarity. Most of the homologous families in LED are members of two Pfam A-families: the ‘abhydrolases’ and the ‘Coesterases’. The ‘Coesterase’ family only contains the GG**G**X type superfamilies Candida rugosa, Carboxylesterses and Yarrowia lipolytica. The ‘abhydrolase’ family contains five homologous families from three superfamilies. The remaining families are classified in different Pfam B-families.

Comparing and analysing families assists to a consistent and fast annotation of conserved residues, as demonstrated for the families of lipases. For all superfamilies, for which structure information was available, the catalytic triad and the oxyanion hole could be identified. By sequence comparison, these functional residues were assigned in 81% of homologous families and 91% of all LED sequences entries, compared to 48% of lipases entries in Swiss-Prot. Sequence alignment and structure superposition of the homologous family Cepacia lipases revealed that the catalytic aspartic acid was incorrectly assigned by Swiss-Prot for four of five entries. For the homologous families Mycoplasma lipase, Moraxella lipase 3, Moraxella lipase 2 and Acinetobacter esterase, a wrong histidine was annotated as catalytically active in Swiss-Prot. The homologous families Mycoplasma lipase and Moraxella lipase 3 are members of the superfamily Mycoplasma, including also the homologous family of non-haem peroxidases, which contains entries with known structure. Sequence comparison of these families indicated that the catalytically active histidine was confused with a well-conserved histidine in the N-terminal neighbourhood of the first oxyanion hole residue. Although there is no lipase with structure information in the superfamilies Moraxella 2 and Acinetobacter calcoaceticus, the catalytic histidine seems also to be confused with this conserved oxyanion hole histidine. The incorrectly annotated histidine is localized in the N-terminal part of the sequence and not C-terminally to the active serine as

expected. Also, the patterns *H*GG**A**F and *H*GG**G**F near this histidine indicate that the histidine is localized in the GG**G**X oxyanion hole region. These results show that specialized databases like LED can help to enhance annotation coverage and assist general sequence databases like Swiss-Prot.

The multisequence alignments for the superfamilies and the homologous families allow to transfer information from one member to all members of the corresponding family. Thus, secondary structure information of lipases with known structure can be mapped on sequences with unknown structure, functional relevant residues like the catalytic triad and the oxyanion hole can be identified, and effects of mutations can be correlated to sequence and structure. For RDL and ROL, several mutations and their experimentally observed effects have been published (Table 3). By analyzing the pattern of conservation among the family of homologous lipases, an understanding of the experimental observations was achieved, like their effect on catalytic activity [14] or chain length specificity [37].

Since the α/β hydrolase fold is the only structural feature which is common to lipases from different superfamilies, superposition of all lipases is not unique. Our strategy of superposition was focused on the substrate and the active site, to compare the position and orientation of the tetrahedral C-atom of the substrate intermediate and its interaction with the lipase. For experimental protein–inhibitor-complexes of RML and CRL (phosphate and phosphonate as inhibitors, respectively), the structure superposition suggested by LED was compared to superposition by the FSSP database [29], the latter focusing on the superposition of the overall fold. The deviation of the phosphorus atoms and its four ligands was 1 Å in the LED superposition compared to 1.5 Å as superposed by FSSP. Thus, superposition by LED is optimized for the active site, although the geometry and physico-chemical properties differ greatly in both lipases.

4.2. Role of the oxyanion hole side chain

The oxyanion hole side chain has been suggested to be involved in oxyanion stabilization, donating a third hydrogen bond, as concluded from crystal structure of RML [8]. For the homologous family Rhizomucor lipase, experimental data on oxyanion hole mutants are available. The exchange of the oxyanion hole forming threonine for a serine in RDL [14,16] resulted only in a moderate decrease in activity, while the exchange for hydrophobic amino acids like alanine and valine [14,16] inactivated the mutant. A similar effect also was observed for PeCL [17]. Results from LED suggest that the oxyanion hole side chain of the G**X** type is interacting with at least one amino acid, the anchor residue, which stabilizes this oxyanion hole residue. This concept predicts that lipases of the G**X** type are only active if geometry and physico-chemical properties of oxyanion hole residue and anchor residue fit. This concept is supported by investigations for ROL: exchanging the anchor residue from aspartic acid to the less polar asparagine has shown to decrease activity to 7.2% [14]. In Pseudomonas lipases, the hydrophobic oxyanion hole residue (leucine, methionine, phenylalanine) fits well into a hydrophobic pocket and has a geometry similar to ROL. Recently, it has been replaced by threonine without changing the hydrophobic anchor (Quyen Dinh Thi, personal communication). As a result, the mutant had only 2% of activity compared to wild type, which is in accordance with the anchor–oxyanion hole residue interaction. The additional hydroxyl group did not contribute to the stabilization of the oxyanion.

The anchor–oxyanion hole residue interaction seems to be a sensitive system, optimized for each pair of residues. While in ROL, replacement of the anchor residue for an asparagine resulted in low activity, in HLL, with sequence similarity to ROL, the anchor residue is N92 (determined by sequence similarity). Since only a half-opened X-ray structure is available, the geometry in the activated form is

not known. In the open form, additional hydrogen bonds might assist the anchor residue, like in FSC, where the anchor residue N84 stabilizes the oxyanion hole by a hydrogen bond. In addition, Q121, the second oxyanion hole residue, seems to stabilize the correct orientation of the first oxyanion hole residue by hydrogen bonding.

While architecture of lipases and geometry of the catalytic triad are structurally conserved in all lipases, the other element of the catalytic machinery, the oxyanion hole, could be assigned to either of two groups, based on sequence and structure: the G**X** and the GG**G**X type. Thus the oxyanion hole sequence is a signature which allows to predict the structure of the oxyanion hole. Classification into two types of oxyanion holes seems to correlate to differences in substrate specificity: the GG**G**X pattern occurs in carboxylesterases and short chain specific lipases, the G**X** pattern in medium and long chain specific lipases. For lipases and carboxylesterases the protein family and domain database PROSITE [42] provides two patterns including the functionally relevant and conserved GxSxG motif. Additionally, for carboxylesterases a second, structurally relevant pattern is described, including β-strand 2 of the α/β hydrolase fold. The GG**G**X pattern has not yet been identified as carboxylesterase specific consensus sequence, since it is highly repetitive in a vast number of protein families. However, a pattern combining GG**G**X with GxSxG, is specific for carboxylesterases. It can even be improved by specifying the second oxyanion hole residue in GxS*x*G, which is a well-conserved alanine. In addition, X in GG**G**X is F, L or Y. The resulting pattern GGG[F,L,Y]–$x_{n(n=70\ldots100)}$–GxSAG is highly specific for functional sites of carboxylesterases.

Acknowledgements

We thank BASF for financial support.

References

[1] R.D. Schmid, R. Verger, Angew. Chem. Int. Ed. Engl. 37 (1998) 1608.
[2] R.J. Kazlauskas, U.T. Bornscheuer, Biotransformations with Lipases vol. 8a Wiley-VCH, Weinheim, NY, 1998.
[3] L. Brady, A.M. Brzozowski, Z.S. Derewenda, E. Dodson, G. Dodson, S. Tolley, J.P. Turkenburg, L. Christiansen, B. Huge-Jensen, L. Norskov et al., Nature 343 (1990) 767.
[4] C. Martinez, P.d. Geus, M. Lauwerexs, G. Matthyssens, C. Cambillau, Nature 356 (1992) 615.
[5] P. Grochulski, Y. Li, J.D. Schrag, F. Bouthillier, P. Smith, D. Harrison, B. Rubin, M. Cygler, J. Biol. Chem. 268 (1993) 12843.
[6] J. Uppenberg, M.T. Hansen, S. Patkar, J.T. Alwyn, Curr. Biol. 2 (1994) 293.
[7] J.D. Schrag, Y. Li, M. Cygler, D.A. Lang, T. Burgdorf, H.-J. Schmid, R. Hecht, D. Schomburg, T. Rydel, J. Oliver, L. Strickland, M. Dunaway, S. Larson, A. McPherson, Structure 5 (1997) 187.
[8] U. Derewenda, A.M. Brzozowski, D.M. Lawson, Z.S. Derewenda, Biochemistry 31 (1992) 1532.
[9] C. Martinez, A. Nicolas, H.v. Tilbeurgh, M.P. Egloff, C. Cudrey, R. Verger, C. Cambillau, Biochemistry 33 (1994) 83.
[10] M. Cygler, P. Grochulski, R.J. Kazlauskas, J.D. Schrag, F. Bouthillier, B. Rubin, A.N. Serreqi, A.K. Gupta, J. Am. Chem. Soc. 116 (1994) 3180.
[11] J. Uppenberg, N. Ohrner, M. Norin, K. Hult, G.J. Kleywegt, S. Patkar, V. Waagen, T. Anthonsen, T.A. Jones, Biochemistry 34 (1995) 16838.
[12] D.A. Lang, M.L.M. Mannesse, G.H. DeHaas, H.M. Verheij, B.W. Dijkstra, Eur. J. Biochem. 254 (1998) 333.
[13] D.L. Ollis, E. Cheah, M. Cygler, B. Dijkstra, F. Frolow, S.M. Franken, M. Harel, S.J. Remington, I. Silman, J. Schrag, J.L. Sussman, K.H.G. Verschueren, A. Goldman, Protein Eng. 5 (1992) 197.
[14] H.D. Beer, G. Wohlfahrt, J.E.G. McCarthy, D. Schomburg, R.D. Schmid, Protein Eng. 9 (1996) 507.
[15] Z.S. Derewenda, U. Derewenda, G.G. Dodson, J. Mol. Biol. 227 (1992) 818.
[16] R.D. Joerger, M.J. Haas, Lipids 29 (1994) 377.
[17] S. Yamaguchi, T. Mase, K. Takeuchi, Biosci., Biotechnol., Biochem. 56 (1992) 315.
[18] H. Scheib, J. Pleiss, A. Kovac, F. Paltauf, R.D. Schmid, Protein Sci. 8 (1999) 215.
[19] M. Martinelle, H. Holmquist, I.G. Clausen, S. Patkar, A. Svendsen, K. Hult, Protein Eng. 9 (1996) 519.
[20] G. Stoesser, P. Sterk, M.A. Tuli, P.J. Stoehr, G.N. Cameron, Nucleic Acids Res. 25 (1997) 7.
[21] D.A. Benson, M.S. Boguski, D.J. Lipman, J. Ostell, Nucleic Acids Res. 25 (1997) 1.
[22] A. Bairoch, R. Apweiler, Nucleic Acids Res. 25 (1997) 31.
[23] D.G. George, W.C. Barker, H.W. Mewes, F. Pfeiffer, A. Tsugita, Nucleic Acids Res. 24 (1996) 17.
[24] A. Bateman, E. Birney, R. Durbin, S.R. Eddy, R.D. Finn, E.L. Sonnhammer, Nucleic Acids Res. 27 (1999) 260.
[25] A. Krause, P. Nicodeme, E. Bornberg-Bauer, M. Rehmsmeier, M. Vingron, Bioinformatics 15 (1999) 262.

[26] F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyer Jr., M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, J. Mol. Biol. 112 (1977) 525.
[27] T.J.P. Hubbard, A.G. Murzin, S.E. Brenner, C. Chothia, Nucleic Acids Res. 25 (1997) 236.
[28] C.A. Orengo, A.D. Michie, S. Jones, D.T. Jones, M.B. Swindells, J.M. Thornton, Structure 5 (1997) 1093.
[29] L. Holm, C. Sander, Science 273 (1996) 595.
[30] X. Cousin, T. Hotelier, P. Lievin, J.P. Toutant, A. Chatonnet, Nucleic Acids Res. 24 (1996) 132.
[31] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G. Higgins, Nucleic Acids Res. 25 (1997) 4876.
[32] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Nucleic Acids Res. 25 (1997) 3389.
[33] L. Holm, C. Sander, J. Mol. Biol. 218 (1991) 183.
[34] W. Kabsch, C. Sander, Biopolymers 22 (1983) 2577.
[35] R.J. Kazlauskas, Trends Biotechnol. 12 (1994) 464.
[36] M. Peitsch, Bio/Technology 13 (1995) 685.
[37] J. Pleiss, M. Fischer, R.D. Schmid, Chem. Phys. Lipids 93 (1998) 67.
[38] P. Grochulski, F. Bouthillier, R.J. Kazlauskas, A.N. Serreqi, J.D. Schrag, E. Ziomek, M. Cygler, Biochemistry 33 (1994) 3494.
[39] J.L. Sussman, M. Harel, F. Frolow, C. Oefner, A. Goldman, L. Toker, I. Silman, Science 253 (1991) 872.
[40] T. Etzold, A. Ulyanov, P. Argos, Methods Enzymol. 266 (1996) 114.
[41] G.D. Schuler, J.A. Epstein, H. Ohkawa, J.A. Kans, Methods Enzymol. 266 (1996) 141.
[42] K. Hofmann, P. Bucher, L. Falquet, A. Bairoch, Nucleic Acids Res. 27 (1999) 215.
[43] H. Atomi, U. Bornscheuer, M.M. Soumanou, H.D. Beer, G. Wohlfahrt, R.D. Schmid, Oils–Fats–Lipids, 21st World Congress Int. Soc. Fat. Res., Bridgewater, Microbial lipases — from screening to design (1996) 49.
[44] R.R. Klein, G. King, R.A. Moreau, M.J. Haas, Lipids 32 (1997) 123.
[45] H. Scheib, J. Pleiss, P. Stadler, A. Kovac, A.P. Potthoff, L. Haalck, F. Spener, F. Paltauf, R.D. Schmid, Protein Eng. 11 (1998) 675.