Structure-Function Studies of Epoxide Hydrolases

Agata Naworyta
Faculty of Natural Resources and Agricultural Sciences
Department of Molecular Biology
Uppsala

Doctoral Thesis
Swedish University of Agricultural Sciences
Uppsala 2010

Acta Universitatis Agriculturae Sueciae
2010:5

ISSN 1652-6880
ISBN 978-91-576-7482-1
© 2010 Agata Naworyta, Uppsala
Print: SLU Service/Repro, Uppsala 2010

”Gdy ruszyć chcesz w najdalszą z dróg
Tam gdzie nie sięga wzrok
Musisz wziąć oddech i
Zrobić pierwszy krok

Choć w sercu lęk, nie cofaj się
Tylko tak warto żyć
W chmurach gdzieś lśni Twój szczyt
Odważ się i idź..."

A.M. Jopek

Structure-Function Studies of Epoxide Hydrolases.

Abstract

Epoxides are three-membered cyclic ethers formed in cells via several metabolic pathways. Epoxide hydrolases (EHs) are enzymes that hydrolyse epoxides to the corresponding diols.

The main goal of this thesis was to investigate the structures of EHs from the α/β-hydrolase family.

The first part concerns the structural and functional analysis of a protein-water channel found in EHs in many plants. Thermostability studies, sequence analysis and determination of the x-ray structure of a mutated EH enzyme from *Solanum tuberosum* led to the conclusions that the water channel in plants participates in stabilization of the protein structure and furthermore, it forms an efficient system to enable transfer of protons that are required for enzymatic catalysis.

The second part describes how computational methods together with structural and kinetic information identified factors that are responsible for the enhanced enantioselectivity of an improved variant of EH from *Aspergillus niger* obtained during a directed evolution process. The x-ray structure of the mutant showed that dramatic changes in the active site explain why the preferred (*S*)-substrate binds more easily in the active site than the disfavored (*R*)-enantiomer. The study underscores the importance of obtaining structural data when attempting to understand the results of directed evolution.

The last part presents the structures of two novel microbial EHs that have been shown to produce chemically valuable 1,2-diols and exhibit high enantioselectivity. Their similarity to the mammalian microsomal EH, a key enzyme in detoxification, provided new information about its possible structure. The improved sequence alignment based on the structural work gives new insights on the connections between sequences/structures and the broad scope of selectivities among EHs.

Keywords: chirality, crystallography, diols, enantioselectivity, enzyme, epoxide, epoxide hydrolases, regioselectivity, substrate specificity, X-ray structure,

Author's address: Agata Naworyta, SLU, Department of Molecular Biology, S-751 24 Uppsala, Sweden
E-mail: agata@xray.bmc.uu.se

Contents

List of Publications 7
Abbreviations 9

1 Background 11
1.1 Introduction to the chiral world 11
1.2 Epoxides 12
1.3 Epoxide hydrolases 13
    1.3.1 Mammalian EHs – microsomal and soluble 14
    1.3.2 Plant EHs 15
    1.3.3 Microbial EHs 16
1.4 Structures and mechanism of EH 17

2 Current investigation 23
2.1 Aim and scope of the thesis 23
Proton wires in plant epoxide hydrolases (paper I) 24
    2.2.1 Protein crystallization and structure solution 26
    2.2.2 Structural results 26
    2.2.3 Sequence comparisons 28
    2.2.4 Functional studies 30
    2.2.5 Conclusions 31
2.3 Directed evolution of EH from *Aspergillus niger* (paper II) 32
    2.3.1 Protein crystallization and structure solution 33
    2.3.2 Structure of wild-type *An*EH in complex with inhibitor 35
    2.3.3 Structure of mutated enzyme 37
    2.3.4 Molecular dynamics calculations 40
    2.3.5 Docking experiments 42
    2.3.6 Inhibition studies on mutated enzyme 43
    2.3.7 Conclusions 44
2.4 Introducing novel microbial epoxide hydrolases (paper III) 45
    2.4.1 Functional studies 45
    2.4.2 Protein purification, crystallization, and solving the structures 47
    2.4.3 Structural results 49
    2.4.4 Structure vs. selectivity of EHs 52
    2.4.5 Sequence comparisons 57
    2.4.6 Shedding light on possible structure of human microsomal EH 60
    2.4.7 Conclusions 61

3 Future perspectives 63

References 65

Acknowlegments **69**

List of Publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I Thomaeus A., **Naworyta A.,** Mowbray S.L., Widersten M. (2008). Removal of distal protein-water hydrogen bonds in a plant epoxide hydrolase increases catalytic turnover but decreases thermostability. *Protein Science* 17(7), 1275-84.

II Reetz M.T., Bocola M., Wang L.W., Sanchis J., Cronin A., Arand M., Zou J., Archelas A., Bottalla A.L., **Naworyta A.**, Mowbray S.L. (2009). Directed evolution of an enantioselective epoxide hydrolase: uncovering the source of enantioselectivity at each evolutionary stage. *J Am Chem Soc.* 131(21), 7334-43.

III **Naworyta A.**, Zhao L., Weiner D.P., Mowbray S.L. Structures of two novel microbial epoxide hydrolases showing high and broad enantioselectivity. (manuscript)

Additional paper

Stern A.L.*, **Naworyta A.***, Cazzulo J.J., Mowbray S.L. (2010). Structure of type B ribose 5-phosphate isomerase from *Trypanosoma cruzi* reveals the basis of sugar specificity in the structural family. Submitted to *Biochemical Journal*

Papers I-II are reproduced with the permission of the publishers.
* First authorship shared

Abbreviations

A. niger/An	*Aspergillus niger*
A. radiobacter/Ar	*Agrobacterium radiobacter*
AU	Asymmetric unit
Da	Dalton – atomic mass unit
E. coli/Ec	*Escherichia coli*
EH(s)	Epoxide hydrolase(s)
EET	Epoxyeicosatrienoic acids
ESRF	European Synchrotron Radiation Facility
GPE	Phenyl glycidyl ether
*Hs*sEH	Soluble EH from *Homo sapiens*
HEPES	4-(2-hydroxyethyl)-1-piperazineethane-sulfonate
MD	Molecular dynamics
mEH	Microsomal EH
M. tuberculosis/Mt	*Mycobacterium tuberculosis*
*Mm*sEH	EH from *Mus musculus*
PDB	Protein Data Bank
PEG	Polyethylene glycol
r.m.s	Root mean square
sEH	Soluble EH
*St*EH1	EH from *Solanum tuberosum*
Tris-HCl	Tris(hydroxymethyl)aminomethane
TSO	*Trans*-stilbene oxide

1 Background

1.1 Introduction to the chiral world

The French microbiologist Louis Pasteur (1822-1895) is most famous for his remarkable breakthroughs in the prevention of infectious diseases, but in his early work as a chemist, he also wrote a doctoral thesis on crystallography. He described there a phenomenon of the asymmetry of chemical compounds. Pasteur was the first one who found that some molecules exist in isoforms that are exact mirror images of each other. We call them *chiral* (*χειρ, Greek*), which means “hand” because like hands, they lack symmetry and therefore are not identical or superimposable (Fig. 1).

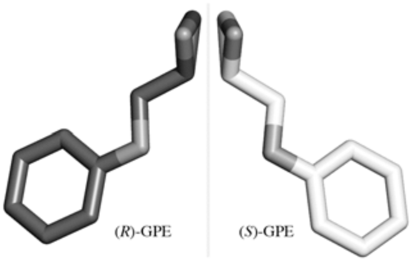


Figure 1. Chirality of phenyl glycidyl ether (GPE).

Since Pasteur’s discovery many other chemicals, particularly biologically active compounds, have been described as chiral. The essential building blocks of life, like amino acids or sugars, have been found to have isoforms called *enantiomers*. This means that the resulting proteins, including cell receptors, hormones and enzymes, are also chiral, and what is even more important, they are enantioselective, i.e. they can distinguish between enantiomers of other chiral molecules. This in turn, has significant impact in the pharmacological industry because various drugs in their enantiomeric forms may be absorbed, activated or degraded differently. The enantiomers may have equal pharmacological activities, but more often only one is active whereas the others can be toxic or have negative side effects. Thence, it is of high importance to be able to distinguish and purify particular enantiomers from racemic mixtures. This means that pharmaceutical companies are always looking for easy and cheap ways of developing enantiomerically pure drugs.

An enormous part of the chiral intermediate market relies on production of enantiopure **epoxides** and diols, which are key building blocks for the synthesis of a number of pharmaceutical compounds. For instance, various β-adrenergic receptor agonists, anti-cancer agents, and anti-HIV drugs are synthesized from epoxide-containing substrates. Several chemical methods have been developed to prepare chiral epoxides and diols, but due to many obstacles such as low efficiency and enantioselectivity, they are not suitable for scaled up synthesis. That is why the general trend is moving towards biocatalysis that involves **epoxide hydrolases** - enzymes transforming epoxides to their corresponding diols.

1.2 Epoxides

Epoxides are organic three-membered heterocycles that contain polarized oxygen-carbon bonds. Epoxide (oxirane) rings are highly strained and unstable in a water environment, and exhibit nucleophilic activity that can cause irreversible toxic effects in reactions with such critical cell components as DNA, amino acids or purines. This in turn can lead to mutagenesis and carcinogenesis (Szeliga and Dipple 1998).

Epoxides can be formed in cells via several metabolic pathways. Frequently they are generated by enzymatic oxidation of xenobiotic compounds, like during transformation of inhaled toxicants by cytochrome P-450 monooxygenase in human lungs (Hukkanen, Pelkonen et al. 2002). Another common route of epoxide formation is via endogenous metabolic pathways. For instance in plants, many epoxide-containing lipids take part in response to stress as wound-healing agents (Howe and Schilmiller 2002). They are also messengers in plant-pathogen interactions and precursors of cutin, which in turn serves as a protective barrier for plants (Pinot, Benveniste et al. 1999). In mammals, fatty acid epoxides are involved in blood vessel relaxation (Carroll, Schwartzman et al. 1987; Carroll and McGiff 2000) and inflammatory processes (Node, Huo et al. 1999; Kozak, Kluger et al. 2000).

Very often epoxide-containing compounds play roles as signaling and regulatory molecules, but when not controlled, they may cause unnecessary chemical changes when reacting with other cell components. This makes the regulation of epoxide levels very critical, and so multiple systems have evolved in order to transform them into less active compounds.

Two types of enzymes are responsible for the detoxification, of which one group is glutathione transferases, known as general detoxifying enzymes (Mannervik and Danielson 1988). By contrast, epoxide hydrolases (EHs) have been described as more specific for particular epoxide elimination (Seidegard and Ekstrom 1997). The latter enzymes, by opening the ring of epoxides, produce less active diols that show increased water solubility and therefore can be easily excreted (Fig. 2).

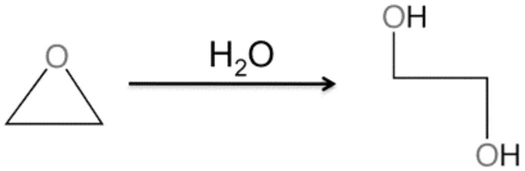


Figure 2. Hydrolysis of an epoxide to the corresponding diol.

1.3 Epoxide hydrolases

EHs are ubiquitously found in nature and have been identified in many organisms including mammals, plants, insects and various microorganisms such as yeasts, filamentous fungi and bacteria. The functions of EHs vary in the different organisms, but the four main roles can be described as: detoxification (mostly in mammals), catabolism of specific carbon sources (in bacteria), response to stress (in plants) and regulation of signaling molecules (in mammals)

Till now, several distinct sub-types of EHs have been described in pro- and eukaryotes. These include mammalian soluble epoxide hydrolase (sEH) and microsomal epoxide hydrolase (mEH), insect juvenile hormone EH, limonene-1,2-epoxide hydrolase (LEH), cholesterol 5,6-oxide hydrolase (ChEH), hepoxilin A_3 hydrolase and zinc–metalloprotein leukotriene A_4 hydrolase (LTA_4). With the exceptions of last three, most of the epoxide hydrolases can be grouped into two main families. One is populated by EHs from microorganisms, with the LEH from *Rhodococcus erythropolis* as the archetype (Barbirato, Verdoes et al. 1998) and a recently discovered EH from *Mycobacterium tuberculosis* as a further example (Johansson, Unge et al. 2005). The other family, formed by enzymes with an α/β-hydrolase fold, is larger and more diverse (Ollis, Cheah et al. 1992). Based on their subcellular localization and biochemical properties, enzymes within this family have been divided into two major groups, sEH and mEH, and both have their representatives in humans.

All EHs that have been studied and are presented in this thesis are members of the α/β-hydrolase family and show sequence similarity to the human enzymes, sEH and mEH. This introduction will therefore concentrate on the α/β-hydrolase family members and will not discuss the other EHs that are not related in sequence and structure and thus have different catalytic mechanisms.

1.3.1 Mammalian EHs – microsomal and soluble

Of all known EHs, the first characterized was mammalian microsomal epoxide hydrolase (mEH), originally found in liver. It is an ~50kDa strongly hydrophobic protein, with a transmembrane anchor of 20 residues at the N-terminal end. It has now been found in all evaluated mammalian tissues, with the highest activity in liver, testis, lungs and heart. Intracellularly, mEH is mostly attached to the smooth endoplasmic reticulum. The enzyme is very well known by pharmacologists and toxicologists because of its role in the metabolism of xenobiotic epoxide-containing compounds that might cause mutagenic and carcinogenic effects. It hydrolyses a number of environmental contaminants, like polyaromatic hydrocarbons, and drugs, e.g. 1,3-butadiene, phenytoin and benzopyrene. Despite the fact that mEH has been extensively investigated for decades, the x-ray structure remains unsolved, and because of the medical importance it is still urgently needed.

The soluble epoxide hydrolase in mammals has been described as a homodimer with 62kDa subunits, where each subunit is build by two distinct functional domains having completely different structures (Gomez, Morisseau et al. 2004). The 35kDa C-terminal domain shows the epoxide hydrolase activity and has the α/β-hydrolase fold. The N-terminal domain is smaller, ~ 25kDa, shows phosphatase activity and has the topology typical for the haloacid dehalogenase superfamily. The activity of sEH has been found in all vertebrates inspected so far, including the teleost fish rainbow trout (*Salmo gairdneri*), mouse (*Mus musculus*), rat (*Rattus norvegicus*), rabbit

(*Oryctolagus cuniculus*), domestic horse (*Equus caballus*), rhesus monkey (*Macaca mulatto*), and human (*Homo sapiens*). The distribution of sEH in tissues is very broad – from liver, kidney, lungs, heart, brain, spleen, adrenals, urinary bladder, to vascular endothelium, smooth muscles, skin, and leukocytes, but the highest specific activity has been reported in liver and kidneys (Pacifici, Temellini et al. 1988). Its critical role is in regulation of metabolic cascades influenced by epoxide-containing lipids. The best studied of the sEH substrates are epoxides of arachidonic acid (epoxyeicosatrienoic acids, EETs), that being a major component of the endothelium-derived hyperpolarizing factor, function as regulatory molecules in vascular, pulmonary and renal physiology.

These two types of mammalian EH enzymes are distinguished not only by their subcellular localization, but also by different pH optima, physical properties and substrate specificity. In general, mEH hydrolyses mono- and *cis*-1,2-disubstituted epoxides, whereas the soluble enzyme prefers *trans*- over *cis*-epoxides (Ota and Hammock 1980) (Fig. 3).

1.3.2 Plant EHs

Soluble EHs have been isolated from different plants, such as soybean (*Glycine max),* mouse eared cress (*Arabidopsis thaliana*), potato (*Solanum tuberosum*), common tobacco (*Nicotiana tabacum*), oilseed rape (*Brassica napus*), spurge (*Euphorbia lagascae*), and lemon (*Citrus jambhiri*). The activity has been reported also in spinach (*Spinacia olerecea*) and apple (*Malus pumila*). EHs have been localized in many cell types including seedlings, germinated seeds, roots, fruit, tubers and leaves. They are located mostly in the cytosol, with a minor fraction associated with isolated microsomes (Blee and Schuber 1992).

The substrate specificity and regulation of transcription shows that the primary functions of this enzyme in plants lie in host defence and growth. Apart from the constitutive soluble EHs, there are also multiple inducible isoforms. The expression of these increases after exposure to hormones involved in growth, differentiation of meristematic tissues, germination, development, fruit ripening, and host-defence (Edqvist and Farbos 2003). Physical trauma in potato leaves (Stapleton, Beetham et al. 1994) or viral infection in the common tobacco have also been reported to increase the transcription of sEH (Guo, Durner et al. 1998). Moreover, plants produce massive amounts of epoxide-containing lipids that are used in cutin synthesis (Kolattukudy 2001) and the host defence response (Blee 2002) (Fig. 3).

These kinds of lipids are the preferred endogenous substrates for plant soluble EHs, which by hydrolyzing them, play defensive roles in both active (production of anti-fungal acids) and passive (cutin biosynthesis) ways.

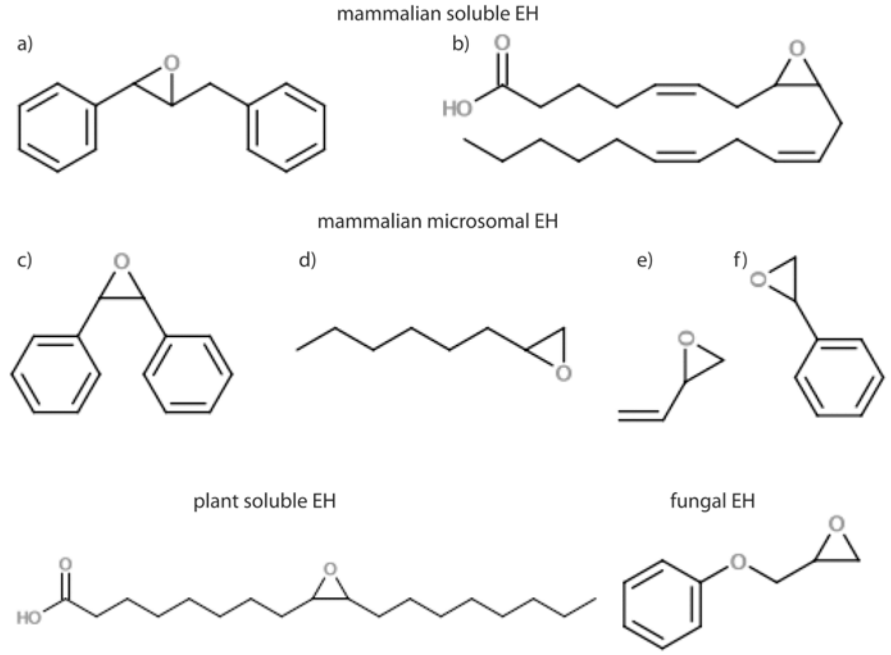


Figure 3. Epoxides as substrates for different EHs a). *trans*-stilbene oxide, b). 8(9)-epoxyeicosatrienoic acid, c). *cis*-stilbene oxide, d). octane 1,2-epoxide, e). butadiene monoxide, f). styrene 7,8-oxide, g). 9(10)-epoxy octadecanoid acid, h). glycidyl phenyl ether.

1.3.3 Microbial EHs

About 20% of the microorganisms with the whole genome sequence known possess putative EH genes (van Loo, Kingma et al. 2006). They have been identified in various bacteria, yeasts, and fungi (Archelas and Furstoss 2001), including *A. radiobacter, A. niger, M. tuberculosis, B. megaterium, S. antibioticus, Pseudomonas spp., Corynobacterium spp., Norcardia spp., Arthrobacter spp.,* etc.

Although the EHs have been intensively studied over the last few decades, the best characterized in terms of structure-activity properties are still those of mammalian origin. A relatively small number of EHs from microorganisms have been explored in detail and very little is known about their functions. However interest in microbial EHs has increased very recently, since several studies had shown that they catalyze the enantioselective conversion of industrially important epoxides and chiral diols, often showing high enantioselectivity and broad substrate scope.

1.4 Structures and mechanism of EH

The α/β-hydrolase superfamily has one of the most frequently observed folds and includes not only EHs but also other hydrolases like lipases, esterases, haloalkane dehalogenases, carboxypeptidases, iminopeptidases and haloperoxidases. It is a diverse group of enzymes with very low sequence identities (often below 20%) and very different enzymatic functions. This superfamily is an example of divergent evolution that creates enzymes that come from a common ancestor and have the same fold, but different catalytic properties.

To date, six structures of EHs belonging to the α/β-hydrolase family have been solved, including an enzyme from *Agrobacterium radiobacter (An*EH) (Nardini, Ridder et al. 1999), sEH from *Mus musculus* (*Mm*sEH) (Argiriadi, Morisseau et al. 2000), a fungal enzyme from *Aspergillus niger (An*EH) (Zou, Hallberg et al. 2000), sEH from human *(Hs*sEH) (Gomez, Morisseau et al. 2004), a potato (*Solanum tuberosum*) EH (*St*EH1) (Mowbray, Elfstrom et al. 2006), and bacterial *Mt*EH from *Mycobacterium tuberculosi*s (Biswal, Morisseau et al. 2008).

Although these enzymes are different in sequence, their overall structure is highly conserved. They all consist of the α/β-hydrolase core domain, and a smaller lid domain (Fig. 4). The main domain is composed of a central twisted eight-stranded β-sheet surrounded by α-helices. The superposition of available EH structures shows that both β-strands and α-helices are well conserved within the main domain. The lid domain is built mostly by α-helices arranged in two layers. The first helix usually leads to the active site that lies in the cleft between the α/β-hydrolase domain and the lid. The next three helices of the lid form a nearly triangular arrangement over the active site, and are linked by a cap-loop to the last two helices, which in turn lie between the triangle and the α/β-hydrolase domain. The two domains are connected by loops I will call NL- and CL-loops (where L stands for the lid domain, while N and C stand for the N- and C-terminal part of the main domain, respectively). The length, fold and character of those three loops vary between different EHs and are thought to be connected with substrate specificity of particular enzymes.

As mentioned before, EHs from the α/β-hydrolase family have been separated into two major groups, soluble (also referred to as cytosolic) and microsomal EHs. Most of the enzymes (mammalian and insect) from the latter one contain additionally an N-terminal membrane anchor. The microbial EHs from the same group lack the anchor and therefore are soluble, but still are ~70 residues longer than EHs from the cytosolic group. That is caused by the presence of an N-terminal meander that caps the lid domain (Fig. 4). This meander forms a long hairpin-like structure with several α-helices and interacts with both of the other domains, stabilizing dimer interactions. Most of reported EHs are in fact homodimers, with the exception of the plant enzymes, which have an additional loop in the lid domain, preventing dimerization.

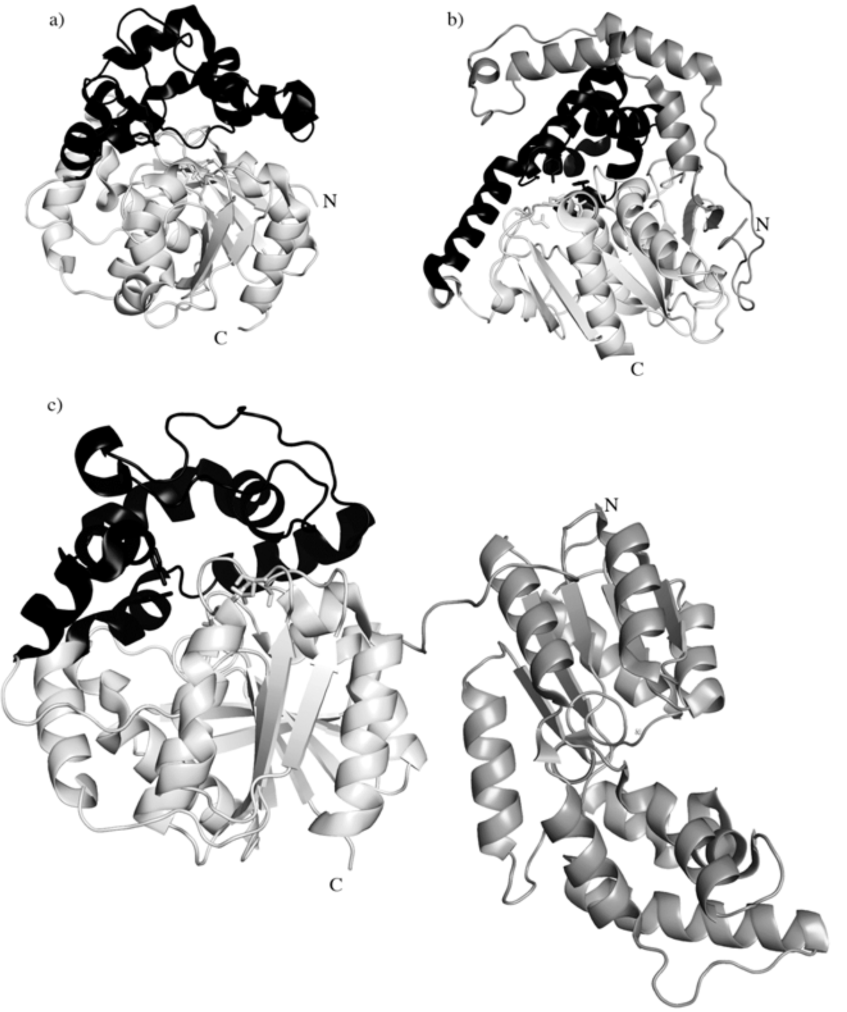


Figure 4. Overall structures of EHs from the α/β-hydrolase family. a) *St*EH1, b) *An*EH, c). *Hs*sEH. For clarity, all enzymes are shown as monomers, but *An*EH and *Hs*sEH are actually dimers. The core domains are colored in white and the lid domains in black. For *An*EH, the N-terminal meander is shown in gray, while for *Hs*sEH this color indicates the additional N-terminal domain displaying phosphatase activity. Catalytic residues as well as two tyrosines from the lid are shown in stick representations and colored as the domains from which they are originate.

Despite a generally low sequence similarity, EHs possess a high conservation of the catalytic triad, which is composed of nucleophile-histidine-acid (Asp-His-Asp/Glu). These residues are located in the main domain while two tyrosines, which are also involved in substrate binding, come from the lid.

There are two main steps in the EH mechanism (Fig. 5). The first one is associated with the opening of the epoxide ring. The two tyrosines from the lid make hydrogen bonds with the oxygen of the substrate, while nucleophilic aspartate is attacking one of the epoxide-ring carbons. This results in the formation of a covalent alkyl-enzyme intermediate, which is subsequently hydrolysed by a water molecule in the second step of reaction. The catalytic histidine together with a second acidic residue (Asp or Glu) function as a charge-relay pair, which is responsible for water activation.

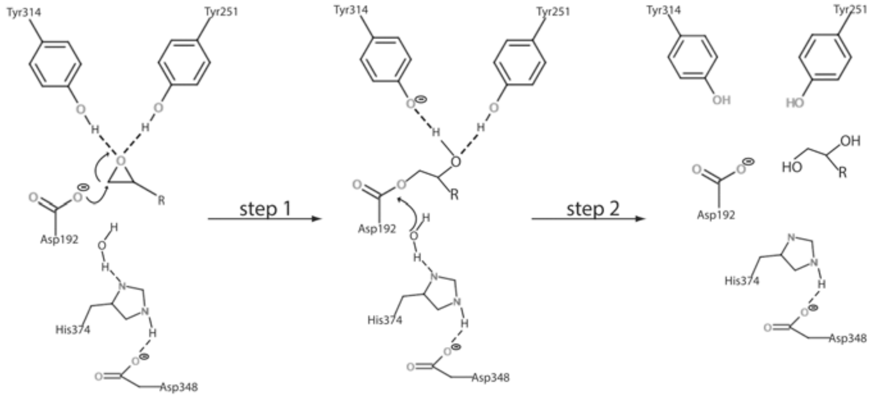


Figure 5. Schematic view of epoxide hydrolysis catalyzed by epoxide hydrolase (in this case *An*EH). Tyr314 and Tyr251 are responsible for initial binding and positioning of the substrate in the active site. In the first step of the enzymatic reaction, the unsubstituted carbon atom of the epoxide is attacked by the carboxylate sidechain of the nucleophilic Asp192. This opens the epoxide ring and the covalent alkylenzyme intermediate is formed. In the second step, a water molecule, previously activated by the His374/Asp348 charge relay, hydrolyses the intermediate and the diol is released as a product. Catalytic residues as well as two tyrosines from the lid are shown and labeled (*An*EH numbering).

The kinetic mechanism of EH in its simplest form is shown on Scheme 1, where K_s is the equilibrium dissociation constant of the substrate-enzyme complex (Michaelis complex), k_2 is the rate of alkylenzyme intermediate formation, and k_{-2} and k_3 are the rates of its hydrolysis, which in turn can lead either towards diol production or back into enzyme-substrate complex. The second step of the reaction in the majority of cases limits the rate of diol production (Widersten, Gurell et al. 2009).

$$E + S \overset{Ks}{\rightleftarrows} E\text{-}ES \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} E\text{-}alkyl \xrightarrow{k_3} E + diol$$

Scheme 1. Kinetic mechanism of EH.

This mechanism applies to many reactions of epoxide-containing substrates (Tzeng, Laughlin et al. 1998) (Thomaeus, Carlsson et al. 2007). However, it does not take into account regiospecificity of the enzymes, i.e. which of the carbon atoms of the epoxide ring is attacked by the nucleophile, which in turn has important consequences for the stereochemistry of the final products of the reaction. It is noteworthy that since only one oxygen atom from a water molecule is incorporated into the product, the opening of the epoxide ring occurs in a *trans*-specific manner. Also, the configuration of the carbon atom that is attacked by the nucleophile can be either retained or inverted during the reaction, depending on the substitution pattern of the attacked carbon and the regioselectivity of the enzyme (Fig. 6).

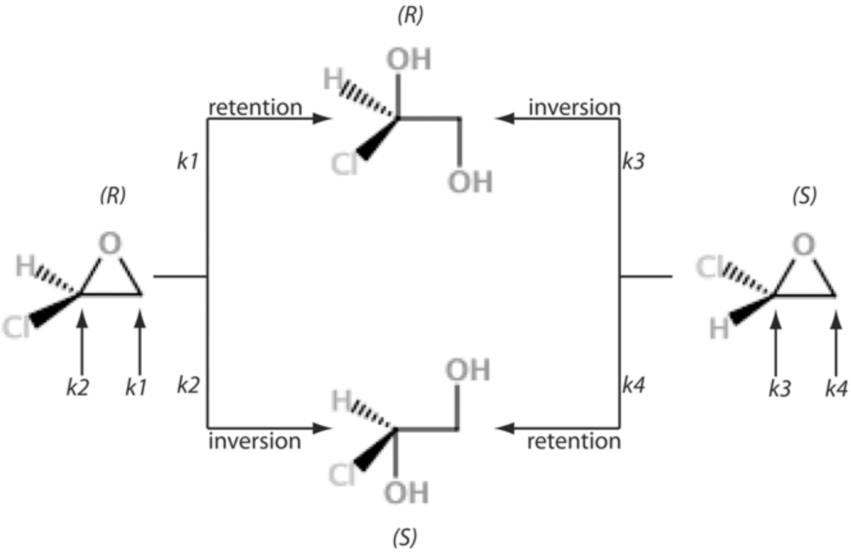


Figure 6. Stereochemistry of the reaction catalyzed by EH, illustrated for a chloro-substituted ethylene oxide. In the hydrolysis of a racemic mixture of epoxides, four hydrolytic pathways (k_1-k_4) are possible. Depending on the enantioselectivity of the enzyme, different enantiomers are hydrolyzed with different rates, whereas its regioselectivity determines which carbon will be attacked. In the majority of cases, the aspartate nucleophile attacks the less hindered carbon of the epoxide ring ($k_1 > k_2$ and $k_4 > k_3$).

2 Current investigation

2.1 Aim and scope of the thesis

The work presented here describes the functional and structural studies of epoxide hydrolases from different organisms.

The first part discusses the x-ray structure of an EH from a plant (*S. tuberosum*), analysis of which has led to the discovery of a protein-water channel leading from bulk solvent to the active site of the enzyme. Kinetics studies together with structural information from the mutated protein are presented here, giving new insights on how the enzyme works. Moreover, sequence comparisons showed that this system is present also in other species, suggesting that plants might have evolved it for the reasons that are further discussed in **paper I**.

The next chapter of the thesis describes results obtained in the directed evolution of EH from *A. niger* (**paper II**). Several computational methods were combined with enzyme kinetics and analysis of x-ray structures, which have led to the identification of factors that are responsible for the improved enantioselectivity of the enzyme.

The last part of the thesis addresses the problem that inadequate structural information is available on epoxide hydrolases, particularly from the mEH branch. **Paper III** presents two structures of recently discovered proteins that are the most closely related to date to the human mEH, an enzyme that has been difficult to over-express and study. Its structure remains unknown, but because of mEH's medical importance, there is an urgent need. The analysis of these structures provides new insights on the fold of mEH. Also, since the sequence identity within the EH superfamily is often very low, and at the same time this group of enzymes has a very wide spectrum of substrate specificities, the new structural results present exciting potential in terms of possible prediction of enzymatic preferences.

2.2 Proton wires in plant epoxide hydrolases (paper I)

The soluble epoxide hydrolase from potato (*St*EH1) was first characterized in 2000 (Morisseau, Beetham et al. 2000). The x-ray crystal structure of wild-type enzyme was solved in 2006 (Mowbray, Elfstrom et al. 2006). The enzyme is a typical sEH with the common overall structure consisting of the core α/β-domain and the α-helical lid domain. The active site is located between those two domains and five catalytic residues were easily identified by analogy with other EH structures.

It has been shown that the pair of tyrosines from the lid (in this case Tyr154/Tyr235) is responsible for proper orientation of the epoxide in the active site by interacting with the epoxide oxygen. The hydrogen bonds with the phenol groups also stabilize the oxyanion that is formed during the first step of the enzymatic reaction, when the epoxide is attacked by the nucleophile Asp105. In the second step of the catalysis, the previously formed alkylenzyme intermediate is hydrolyzed by a water molecule activated by the general relay pair His300/Asp265. The first x-ray structure of *St*EH1 thus helped to understand the catalytic mechanism of the *St*EH1.

Another interesting residue, Glu35, was later identified and proposed to be a linker in a chain of residues connecting the catalytic His300 with the solvent through hydrogen bonding (Thomaeus, Carlsson et al. 2007). While His300 activates the hydrolytic water molecule, Glu35 is responsible for its proper orientation before interacting with the alkylenzyme intermediate. It has also been proposed that the attack of the base-activated water molecule is simultaneous with protonation of the negatively charged intermediate, where the necessary proton is transferred from solvent through a chain of water molecules to the catalytic tyrosine, Tyr154. These water molecules are coordinated by several residues that are located in a tunnel leading from the protein surface to the interior of enzyme. This hydrogen bond network goes through the series Pro186-His269-Tyr249-Leu266-His153-Tyr154-alkylenzyme-His300-Glu35-Ser39-Tyr219-Arg41-Glu215, where Pro186 and Glu215 are situated at the protein/solvent interface, on two opposite entries of the tunnel (Fig. 7).

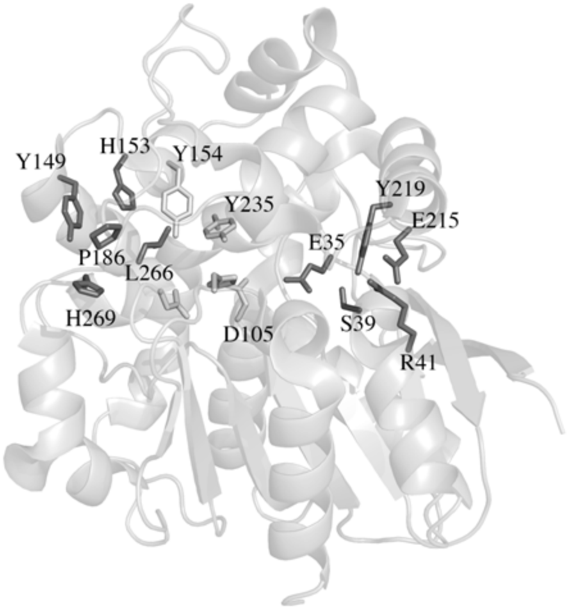


Figure 7. Overall structure of *St*EH1, showing the proposed proton wire. The catalytic residues are shown in white, whereas residues involved in forming the putative proton wire are shown in black.

Both the protein channel and the chain of water molecules were identified from the x-ray structure of wild-type *St*EH1. To understand the role of this putative network in catalysis, two single mutants, Y149F and H153F, and one double mutant, Y149F/H153F, were constructed, expressed and characterized by steady-state kinetic measurements and temperature inactivation studies. These particular mutations were chosen to disrupt the hydrogen bond network in the chain of waters and to understand the importance of the putative proton wire in the protonation step of the enzymatic catalysis. For better interpretation of the functional data, the x-ray structure of Y149F mutant was determined and compared to the wild-type structure.

The protein was studied in collaboration with the Department of Biochemistry and Organic Chemistry at Uppsala University. The mutagenesis, protein expression, purification and characterization were performed by Ann Gurell and Mikael Widersten.

2.2.1 Protein crystallization and structure solution

The first crystallization trials of the Y149F mutant were carried out directly after the last step of purification, performed on a HiPrep Sephacryl S-200 (16/60) column equilibrated with 0.1 M sodium phosphate (pH 7.4). Many different crystallization conditions were tested, together with the previously used 25% PEG 10,000, 0.1 M HEPES, pH 7.5 (Mowbray, Elfstrom et al. 2006) but no crystals ever appeared. Since the wild-type enzyme was successfully crystallized when the protein solution contained 30 mM Tris-HCl, an additional step consisting of a quick buffer exchange on a PD10 column was performed before setting up new crystallization experiments. With the new protein solution, crystals were easily obtained in the same conditions as for the wild-type *St*EH1 (Fig. 8).

Figure 8. Crystals of the Y149F mutant of *St*EH1.

The rod-like crystals (~0.1 x 0.05 x 1 mm^3) appeared within a couple of days, and a data set of 99.6% completeness to 2 Å resolution could be collected. The crystals possessed the symmetry of the space group $P2_12_12_1$ and the cell dimensions a=55.9; b=96.7; c=122.2 Å.

2.2.2 Structural results

The point mutation of Y149F caused only a few, very small changes in the structure. The overall fold remained the same, and superposition with the wild-type enzyme resulted in an r.m.s. difference of 0.2 Å when all Cα atoms were compared.

Two molecules were found in the asymmetric unit, with an r.m.s. difference of 0.3 Å when all C^α atoms compared. In the active site of the A molecule, clear electron density compatible with a molecule of PEG was found. PEG is a known competitive inhibitor of mEH (Kerr, Rettie et al. 1989) and in this case its binding is reasonable since it was present at 25% (w/v) in the crystallization conditions. The terminal part of the PEG lies close to the lid tyrosines, while the remainder of the visible density meanders out to the surface of the protein through a binding cavity. The inhibitor binds to the nucleophile Asp105 and to a water molecule that interacts with Tyr154 and Tyr235 (Fig. 9a). In molecule B, no additional electron density was observed; instead a water molecule was found at the same position as the terminal oxygen of PEG in molecule A (Fig. 9b). The binding modes of the PEG and water molecules are similar to those previously seen in the wild-type structure, and indicate the expected position of the oxygen in the epoxide ring of a potential substrate.

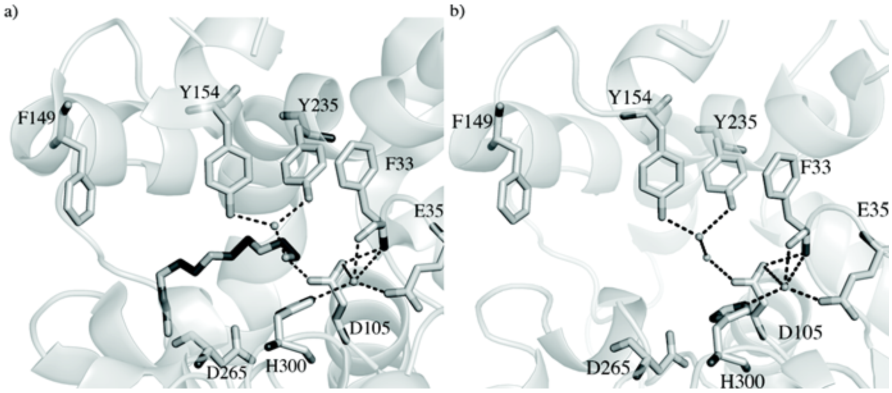


Figure 9. Active site of *St*EH1. The Y149F mutant is shown in the A (a) and B (b) molecules of asymmetric unit. Catalytic residues are labeled as well as the mutated residue 149 that faces out into the solvent (at left on both panels). The water molecules together with PEG bound to the tyrosines are shown. The hydrogen bonds are dashed.

Replacing Tyr149 with a phenylalanine (i.e. removing an OH moiety) generated a space that in the mutated protein is filled by the closest water molecule coming from the solvent. The other water molecules of the putative wire remained in the same positions, coordinated by the same residues as in the wild-type enzyme (Fig. 10a-b).

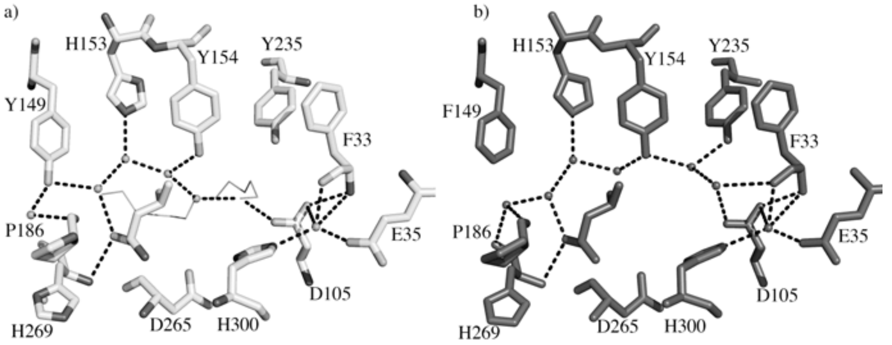


Figure 10. Comparison of proton wires in the active site of (a) wild-type *St*EH1 and (b) Y149F mutant. The catalytic residues as well as the ones involved in water coordination are labeled. The PEG molecule present in the wild-type structure is shown in a stick representation.

2.2.3 Sequence comparisons

A structure-based sequence alignment of epoxide hydrolases including sequences from other plants shows that the Tyr149 residue is present in eight out of the ten EHs that were compared (Fig. 11). This suggests that such highly conserved residue might be involved in coordination of water molecules in other plants. In contrast, the His153 residue found in *St*EH1 is otherwise present only in the EH of rice; other enzymes have a tyrosyl residue in that position. When modeled in *St*EH1, the hydroxyl group of this tyrosine replaces the histidine imidazole moiety in the hydrogen-bonding chain. The other residues from the putative wire, like Pro186, Leu266 and His269, are also present in other sequences but their conservation is suggested reflect the need to maintain the correct local structure.

It is noteworthy that the exit route of the enzyme channel formed by Glu35-Tyr219-Ser39-Arg41-Glu215 is structurally conserved in mammalian sEHs (Thomaeus, Carlsson et al. 2007), while the entry pathway built by Pro186-His269-Tyr149-Leu266-His153 could not be identified in any of the other available structures.

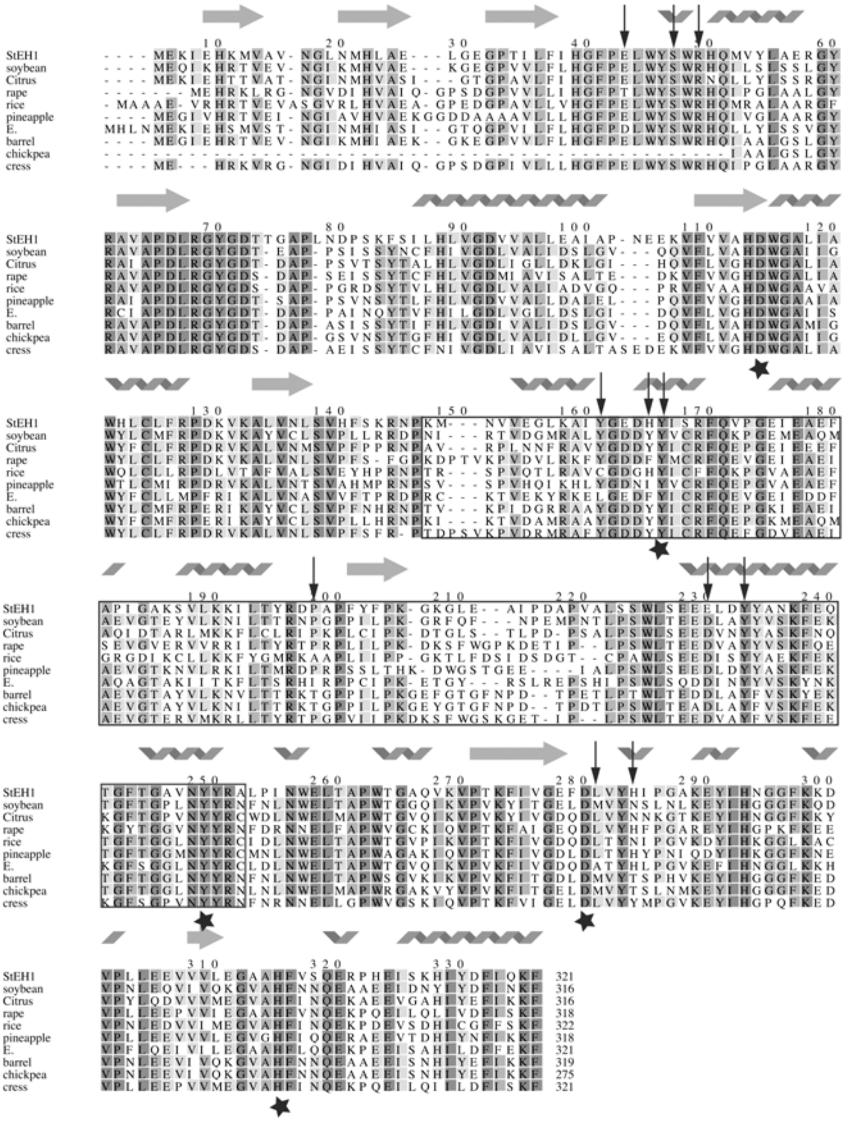

StEH1
soybean
Citrus
rape
rice
pineapple
E.
barrel
chickpea
cress

Figure 11. Sequence alignment of soluble epoxide hydrolases in plants. Arrows indicate residues participating in the putative proton wire; the catalytic residues are marked with stars. The lid domain is boxed and the secondary structural elements of *St*EH1 are shown.

2.2.4 Functional studies

Enzyme kinetics

The enzyme activities with both enantiomers of *trans*-stilbene oxide (TSO) were determined by steady state kinetics studies for all three mutants. The results are listed in Table 1.

Table 1. *Steady-state kinetic parameters of TSO hydrolysis by wild-type and mutated StEH1.*

	R,R-TSO			*S,S*-TSO		
Enzyme	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M ($s^{-1}μM^{-1}$)	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M ($s^{-1}μM^{-1}$)
Wild-type	23±2	10±1	2.4±0.2	3.8±0.1	4.7±0.5	0.80±0.07
Y149F	44±2	36±3	1.2±0.06	6.8±0.4	6.8±0.9	1.04±0.09
H153F	15±2	94±20	0.016±0.01	12±1	11±1	1.1±0.1
Y149F/H153F	n.d.	n.d.	0.20±0.01	12±1	5.2±0.5	2.2±0.2

In the reaction of Y149F mutant with *R,R*-TSO, both k_{cat} and K_m were higher when compared to the wild-type enzyme, which resulted in lower catalytic efficiency (k_{cat}/K_M). For the reaction with *S,S*-TSO, k_{cat} and K_m were only slightly increased, however, this did not affect k_{cat}/K_m.

In the case of the H153F mutant, for the hydrolysis of *R,R*-TSO, k_{cat} was not changed, while K_m increased (9-fold), which resulted in a 15-fold drop in catalytic efficiency. With the *S,S*-enantiomer, both k_{cat} and K_m were increased (2- and 3-fold), resulting in an unchanged k_{cat}/K_m.

The difficulties in reaching enzyme saturation within the solubility range of the substrate meant that k_{cat} and K_m parameters could not be measured individually for the reaction of the double mutant (Y149F/H153F) with *R,R*-TSO as a substrate. However the catalytic efficiency could be measured, and decreased more than 10-fold compared to wild-type enzyme. When the *S,S*-enantiomer was used as a substrate, k_{cat} increased, while K_m remained unaffected, resulting in an increase of k_{cat}/K_m (2-fold).

We interpret this to mean that, in the Y149F mutant, the additional water molecule filled the available space and joined the water chain, and that therefore the enzyme did not lose its catalytic activity. Nevertheless, the increase in k_{cat} of the mutant when compared to the wild-type enzyme can not be caused by improved proton transport within the chain for two reasons: a) the rates of proton transfer between atoms are much faster than the catalytic turnover numbers, and b) it is not protonation of alkylenzyme that is the rate-limiting step but instead its hydrolysis (Thomaeus, Carlsson et al. 2007).

Thermostability studies

The catalytic activities of the wild-type enzyme and all three mutants (Y149F, H153F, and Y149F/H153F) were determined after incubation at 55 °C. The wild-type enzyme exhibited the highest degree of resistance to the higher temperature, with a half-life of approximately 2.25 hours. The single mutants showed stabilities with intermediate half-lives, of one hour, whereas the double mutant exhibited the lowest stability with the shortest half-life (twenty minutes). The wild-type enzyme was also more resistant to thermal inactivation in higher temperatures. It retained 65% of its activity after 5 minutes incubation at 60 °C, whereas the Y149F mutant showed only 30% of its initial activity at 30 °C.

The structure of Y149F mutant shows that even though the water chain is maintained, this mutation removes two hydrogen bond interactions involving protein residues, which might destabilize the structure. Removing the maximal number of interactions in the double mutant drastically lowers its half-life at higher temperature. The drop in stability should be interpreted as a time-dependent loss in catalytic activity rather than as a direct analysis of structural unfolding. The thermostability tests show clearly that all three mutations are responsible for loss of protein-water hydrogen bonding, which in turn destabilizes the mutant proteins.

2.2.5 Conclusions

Based on the thermostability data and conservation of most of the residues coordinating water molecules in the chain, two functions have been proposed for the water channel in *St*EH1. First of all, it can stabilize the enzyme structure via hydrogen bonds between water molecules and protein side chains. The second suggested role of this water-protein network is a function as a proton wire for efficient protonation of the alkylenzyme intermediate during catalysis. In the active site, the His300 imidazole must be unprotonated in order to function as a catalytic base, which results in an urgent need for an efficient proton transfer. On the other hand, the alkylenzyme intermediate requires protonation prior to its hydrolysis to form the final product. The analysis of the *St*EH1 structure suggests that plant epoxide hydrolases might have evolved an efficient system to supply and also remove protons from the active site by their continuous transport from the solvent to the buried active site and then again back to the solvent.

2.3 Directed evolution of EH from *Aspergillus niger* (paper II)

The recombinant epoxide hydrolase from *A. niger* (*An*EH) was first described in 1999 (Arand, Hemmer et al. 1999) and the x-ray structure of the wild-type *An*EH at 1.8 Å resolution was published a year later (Zou, Hallberg et al. 2000). Among other substrates, the enzyme is known as a catalyst in the hydrolysis of glycidyl phenyl ether (GPE). This compound has an aryl oxypropanolamine structure containing one chiral centre, characteristic for all glycidyl ethers that are important intermediates for the synthesis of β-blockers, commonly used pharmaceutical drugs in hypertension treatment (Iakovou, Kazanis et al. 1999).

The hydrolysis of GPE by *An*EH is an example of a kinetic resolution where, one of the enantiomers from a racemic mixture is selectively degraded, leaving the other one behind (Fig. 12). In this type of reaction, the relative rate of the hydrolysis of two enantiomers is reflected by the enantioselectivity factor, *E*.

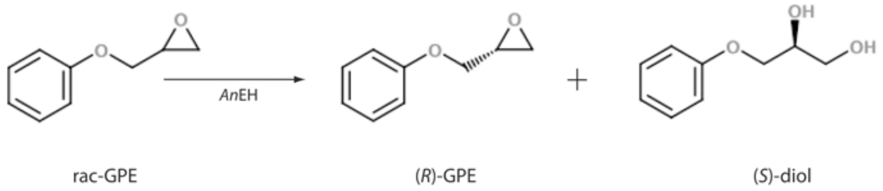


Figure 12. Schematic representation of GPE hydrolysis catalysed by *An*EH.

The wild-type *An*EH catalyzes this reaction with low enantioselectivity (*E* = 4.6), in slight favor of the (*S*)-diol. It has been shown previously that random mutagenesis of the enzyme could slightly increase the enantioselectivity factor (Reetz, Torre et al. 2004). That publication describes several dozen mutants that were created by using error-prone polymerase chain reaction (epPCR) and screened on the basis of their enantioselectivity. In this way, a new, improved variant of the *An*EH had been discovered, displaying an *E*-value of 10.8, in favor of the (*S*)-diol.

Based on previous experiments, the combinatorial active-site saturation test (CAST) was developed to perform mutagenesis of sites around the binding pocket of an enzyme. In this method a gene of a previously improved mutant is used as a template to perform random mutations at another site. The process can be repeated until the desired degree of catalyst improvement is achieved (Reetz, Wang et al. 2006).

In our work, the previous x-ray structure was used to choose six different sites around the active site for the mutagenesis experiments. Screening of the mutants created led to the discovery of a highly enantioselective enzyme (with an *E* value of 193). Using kinetics, MD calculations, molecular modeling and X-ray structures we tried to analyze the factors responsible for the enhanced enantioselectivity of the mutant generated by directed evolution. In addition, inhibition studies were performed on both mutated and wild-type enzyme, and the structure of the latter one in complex with an inhibitor were solved.

Both wild-type and mutated enzyme were studied in collaboration with the Max-Planck Institute in Mulheim, and the Institute of Pharmacology and Toxicology in Zurich. Protein expression, purification and biochemical characterization were performed by A. Cronin and LW. Wang, while molecular dynamics simulations were done by M. Bocola and J. Sanchis.

2.3.1 Protein crystallization and structure solution

Crystals of wild-type *An*EH were obtained by sitting drop vapor diffusion at room temperature, as follows: 2 µl of protein solution (17 mg/ml in 10 mM Tris-HCl, pH 7.4, 20 mM NaCl, 1 mM EDTA, and 0.02% sodium azide) were mixed with 2 µl of the reservoir (20% PEG6000, 0.1 M MES, pH 6.0, and 0.1 M unbuffered sodium acetate). Crystals (~0.3 x 0.4 x 0.02 mm^3) grew within one week and before flash-cooling, were soaked for 4 days in a cryoprotectant solution containing 20% PEG6000, 20% glycerol, 0.1 M MES, pH 6.0, and 0.1 M unbuffered sodium acetate, and 10 mM valpromide.

The LW202 mutant protein was crystallized by vapour diffusion in hanging drops of protein solution (10 mg/ml in 10 mM Tris-HCl, pH 7.4, 20 mM NaCl, 1 mM EDTA, and 0.02% sodium azide) mixed with reservoir solution (15% PEG 3350, 0.25 M ammonium formate) in 1:1 ratio. Crystals (~0.3 x 0.4 x 0.8 mm3) appeared within 5 days. Before being flash-cooled in liquid nitrogen, crystals were transferred to cryoprotectant composed of 20% glycerol, 15% PEG 3350, and 0.25 M ammonium formate.

For both structures, data were collected at the European Synchrotron Radiation Facility (ESRF), Grenoble. Diffraction data were indexed using MOSFLM (Leslie 1999) and processed with SCALA (Evans 1993) as implemented in the CCP4 interface (Potterton et al. 2003). The crystals possessed the symmetry of the space group $P2_1$. The protein model from the previously reported structure in the same space group (PDB (Berman et al. 2000) entry code 1QO7 (Zou et al. 2000)) was used as the starting point for alternating cycles of rebuilding (Jones et al. 1991) and refinement (Murshudov et al. 1997). Two molecules were found in the asymmetric unit. The structures were refined by alternating cycles of refinement in REFMAC (Murshudov, Vagin et al. 1997) and model rebuilding in O (Jones, Zou et al. 1991). Statistics for data processing and the final model are reported in Table 2.

Table 2. Data collection and refinement statistics. The space group in each case is $P2_1$. Information in parentheses refers to the highest resolution shell.

Data collection statistics	LW202 mutant	Wild-type in complex with valpromide
Cell axial lengths (Å)	61.8, 89.7, 75.3 (β= 104.8°)	63.0, 89.7, 75.8 (β= 105.3°)
Resolution range (Å)	30.54 – 1.5 (1.58 – 1.5)	40.0 - 2.1 (2.14 – 2.1)
Average multiplicity	1.7 (1.7)	3.6 (2.7)
Completeness (%)	95.5 (98.2)	99.5 (92.8)
R_{meas} (%)	2.7 (12.8)	6.6 (21.3)
$< \langle I \rangle / \sigma \langle I \rangle >$	15.8 (7.2)	13.7 (4.3)
Refinement statistics		
Resolution range (Å)	30.0 - 1.5	38.2 – 2.1
R-value, R_{free} (%)	18.3, 20.9	17.5, 22.1
Mean B-factor, protein atoms, A and B molecules ($Å^2$)	13.7, 13.9	13.6, 13.9
Mean B-factor, solvent atoms ($Å^2$)	28.0	18.6
Mean B-factor, valpromide atoms ($Å^2$)	-	17.6
Mean B-factor, formic acid ($Å^2$)	11.4	-
Ramachandran plot outliers (%)[a]	1.6	1.5
r.m.s. deviation from ideal bond length (Å)[b]	0.01	0.01
r.m.s. deviation from ideal bond angle (°)	1.01	1.19

---

[a] Calculated using a strict-boundary Ramachandran plot

[b] Using the parameters of Engh and Huber.

2.3.2 Structure of wild-type *An*EH in complex with inhibitor

The wild-type enzyme was soaked with its inhibitor 2-propylpentanamide (valpromide). This compound is commonly used as a drug in epilepsy treatment but is also known as a competitive inhibitor for the human liver mEH (Kerr, Rettie et al. 1989). It is been reported to inhibit other EHs, like the soluble EH from *S. tuberosum* (Mowbray, Elfstrom et al. 2006) and the unrelated limonene-1,2-epoxide hydrolase from *R. erythropolis* (Arand, Hallberg et al. 2003). In the *An*EH case, the binding studies gave a measured K_i of 250 μM.

The complex structure of *An*EH with valpromide is very similar to that of the apo enzyme. The r.m.s. difference between these structures is 0.2 Å when all Cα atoms are included. The active site residues have the same positions, and the only residue within the binding-pocket that is different, is Leu349. Here, both main and side chains are moved, due to the accommodation of the inhibitor. The amide oxygen of valpromide forms strong, short hydrogen bonds to the lid tyrosines (Tyr251 and Tyr314), whereas the amide nitrogen is bound to the nucleophilic Asp192 and the main chain of the oxyanion hole residue Trp117 (Fig 13).

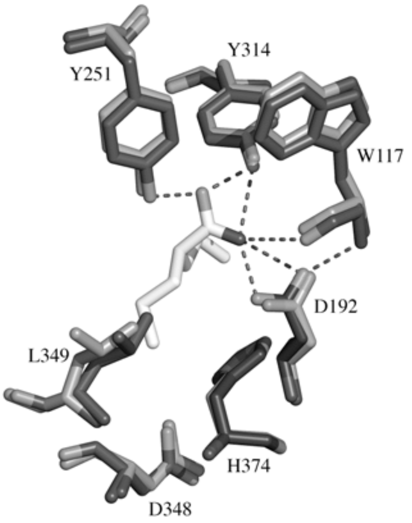


Figure 13. Comparison of the active sites in the wild-type structure of *An*EH with and without valpromide. The ligand-free enzyme is shown in gray, while the valpromide containing structure is colored black. Residues expected to participate in binding are labeled, together with the Leu349 that shows the largest movements in the active site. The hydrogen bonds between inhibitor and active-site residues are dashed. The ligand is shown in white.

Previously, *St*EH1 was the only member of the α/β-hydrolase family that had been reported in complex with valpromide (Mowbray, Elfstrom et al. 2006). this enzyme shares 19% amino acid sequence identity with *An*EH, with an r.m.s. difference of 1.7 Å when all C^{α} atoms are compared. In both structures, the amide moiety of valpromide interacts with the nucleophilic aspartate, but the rest of the ligand is located in different positions in the binding pocket. The superposition based on the Cα positions on the five catalytic residues shows clearly differences in shapes of the active sites that explain the different positioning of bound inhibitor (Fig. 14a-b). The active site of *St*EH1 is very broad on both sides of the catalytic residues, which gives an extra cavity deeper inside the protein, in this case filled by the two aliphatic tails of valpromide (at left in panel a). In *An*EH the binding pocket is much smaller and includes only the "outer" side of the catalytic triad (at right in both panels). This gives limited space for the inhibitor to bind in *An*EH, especially due to the fact that the oxygen of the ligand amide binds strongly to the lid tyrosines. In *St*EH1, those two residues bind to a water molecule instead and valpromide is moved further into the "inner" cavity of the active site. In this structure the "outer" part of the pocket that leads to the solvent is occupied by a small fragment of PEG (Fig. 14a-b).

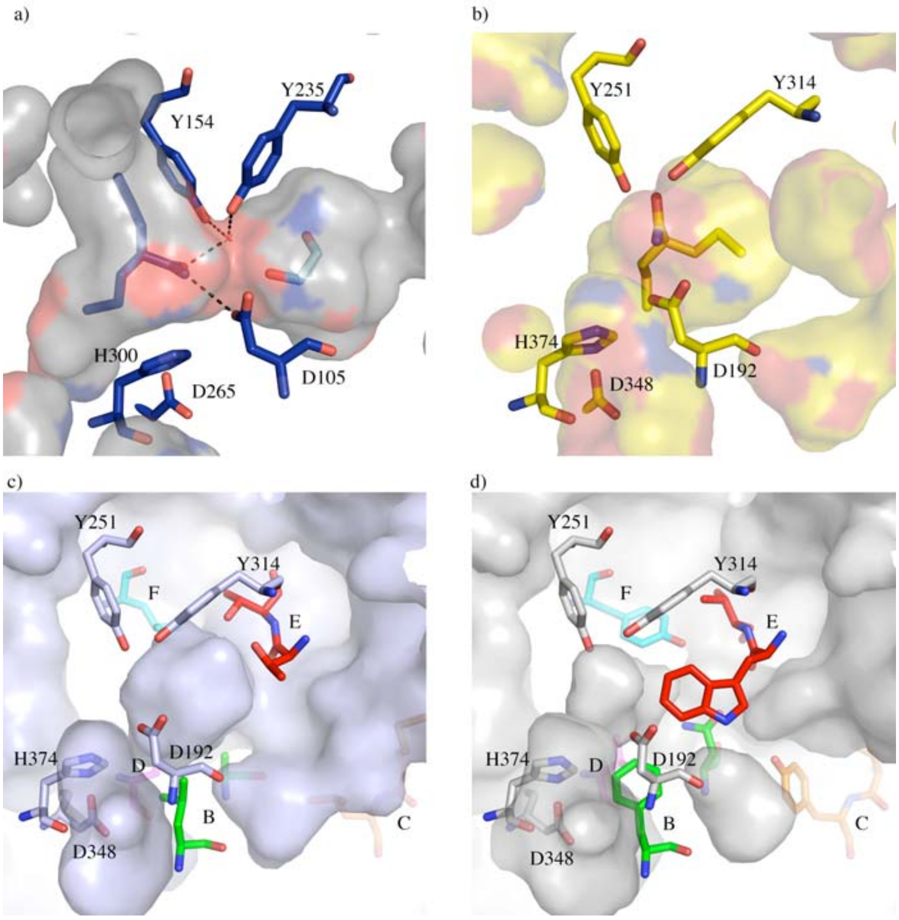

a)
Y235
Y154
H300
D265
D105
b)
Y251
Y314
H374
D348
D192
c)
Y251
Y314
F
E
H374
D
D192
B
C
D348
d)
Y251
Y314
F
E
H374
D
D192
B
C
D348

Figure 14. Comparison of the active site cavities. a) *St*EH1 with valpromide and a PEG molecule, b) *An*EH wild-type in the complex with valpromide, c). apo *An*EH wild-type, d). apo *An*EH LW202 mutant. The catalytic triad (D192/105, H374/300 and D348/265) and the lid tyrosines (Y251/154 and Y314/235) are shown for *An*EH/*St*EH1 and labeled. For the wild-type *An*EH, the sites chosen for mutations are colored as follows: B (green), C (orange), D (magenta), E (red), and F (cyan) and correspond to the mutated residues in LW202.

2.3.3 Structure of mutated enzyme

The active site of *An*EH lies between the α/β-hydrolase domain and the lid domain. A narrow hydrophobic tunnel leads to the active site, and the catalytic triad, as well as the two tyrosines sticking from the lid, are located in and around the tunnel. By modeling substrate (GPE) into that tunnel, six sites were chosen for mutagenesis experiments, named as A (composed by residues 193/195/196), B (215/217/219), C (329/330), D (349/350), E

(317/319), and F (244/245/249) (see Fig. 1 in the paper). Several cycles of mutated protein were created and tested for their enantioselectivity, among which the highest *E* value (of 193) was found in the LW202 mutant, with nine mutations in total (in site B: L215F/A217N/R219S; site C: M329P/L330Y; site D: L249Y/C350V; site E: T317W; and site F: T318V) (Fig. 2 and 8a in the paper).

The overall structure of the apo LW202 mutant is very similar to that of the apo wild-type *An*EH (Fig. 15). The main domain has the classic α/β-hydrolase fold, built two regions of residues 86-229 and 321-396. The lid domain is composed by six α-helices, corresponding to residues 230-320. The first helix leads to the active site, which is located in the cleft between two domains. Helices 2-4 form a triangle over the active site, and are connected to the last two by a relatively short cap-loop (residues 284-291). Like in the wild-type *An*EH, the last helix is disordered and the density for the following loop is also not visible (residues 320-327). The N-terminal region (residues 2-85) forms a curved meander that caps the lid domain; this type of feature is not found in enzymes like *St*EH1, and is more characteristic of the mEH family. The meander's interactions with the helices of the lid domains from both molecules present in the asymmetric unit are thought to play a role in stabilizing the formation of dimers.

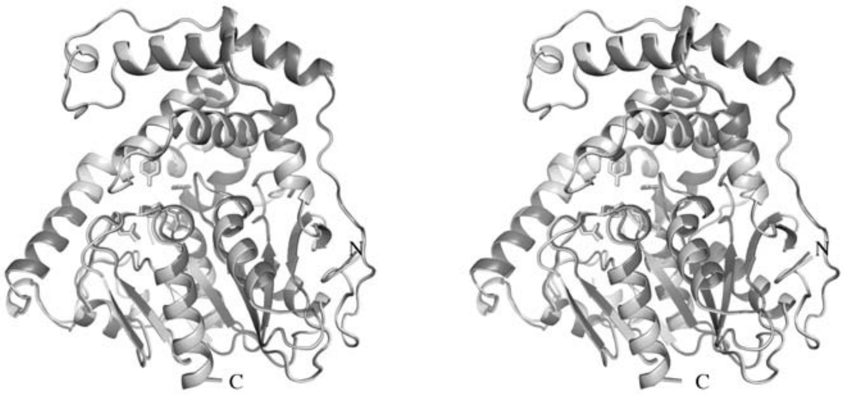


Figure 15. Stereo view of overall structure of LW202 mutant (white) superimposed onto the wild-type *An*EH (black). The C- and N-termini are labeled, and the catalytic residues are shown in stick representation.

Comparison of the two molecules in the asymmetric unit gives an r.m.s. difference of 0.3 Å, when they are superimposed with all Cα atoms matching within a 2 Å cut-off. Due to crystal packing, two regions in molecule B (217-222 and around residue 356) are different when compared to molecule A. Molecule B has also an extra disordered loop (residues 222-229) (Fig. 16).

The 1.5 Å resolution of the data collected let us model multiple conformations for several residues in both molecules. Most of them are located in the N-terminal meander (residues 32, 36, 38, and 62) and in the loops pointing out to the solvent (e.g. 236, 253, 368, 383), but some of them were also found within the binding pocket.

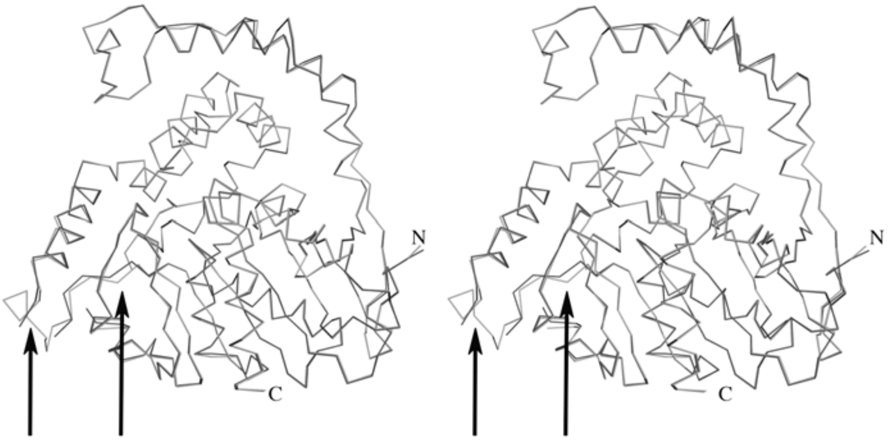


Figure 16. Superposition of two molecules present in the asymmetric unit of LW202 mutant (stereo view). The arrows point the the most striking differences that occur in the NL-loop region, residues 217-222 (at bottom left). Its C-terminal part is disordered in this molecule, but visible in molecule A (white). N- and C-termini are labeled.

There are two active sites per dimer, each composed by residues from a single subunit. Asp192 is the catalytic nucleophile and His374/Asp348 form a general base-charge relay pair. This catalytic triad is located on the main α/β-hydrolase domain, whereas the two tyrosine residues Tyr251 and Tyr314 that are expected to assist the epoxide ring opening come from the lid. In both molecules, clear electron density for formic acid was observed. Since ammonium formate was present at 0.25 M in the mother liquor, its binding in the active site is reasonable. The ligand forms strong interactions with the nucleophile Asp192, as well as the two catalytic tyrosines (Fig. 17a).

Despite the fact that all catalytic residues are in the structurally conserved positions, the binding pocket is significantly different when compared to the wild-type *An*EH (Fig. 14c-d). The changes are mostly caused in a very direct way by the mutations that have been introduced. The "new" residues are not only different in size and character but also have a secondary effect on other residues. For instance, the side chain of Phe196 in the LW202 mutant has a different conformation, pointing out from the active site. This conformational change is caused by the substitution at residue 317 of threonine with much bigger tryptophan. Within the active site, two residues show mobility and were modeled in multiple conformations (Ser195 and Val318) (Fig. 17a).

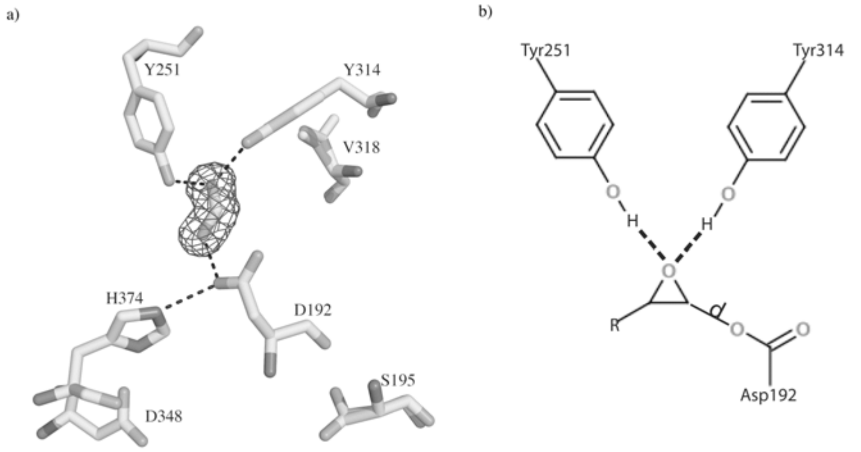


Figure 17. The active site of *An*EH. a) X-ray structure of the LW202 mutant. Electron density is shown for bound formic acid in the A molecule of the asymmetric unit, using a SIGMAA-weighted $2|F_o|-|F_c|$ map contoured at $1\sigma = 0.4$ e/Å^3. The catalytic residues are labeled and the hydrogen interactions are dashed. Two residues with multiple conformations (S195 and V318) are also visible. b) Schematic view of the first step of hydrolysis. The nucleophilic Asp192 attacks one of the epoxide carbons, while the oxygen from the substrate is bound to two tyrosines from the lid. The distance d is marked between the oxygen of Asp192 and one carbon of an epoxide.

2.3.4 Molecular dynamics calculations

As for *St*EH1, the hydrolysis of epoxides catalyzed by *An*EH is believed to occur in two steps. First, two tyrosines (Tyr251 and Tyr314) bind and activate the substrate by forming hydrogen bonds to the oxygen atom of the epoxide, which is followed by an attack of Asp192 on the oxirane ring resulting in the formation of a covalent enzyme-substrate ester intermediate

(Fig. 17b). In the second step, the intermediate is hydrolysed by the attack of a water molecule, previously activated through proton extraction using the His374/Asp348 charge relay.

To study how the *(R)*- and *(S)*-substrates are positioned in the narrow binding pocket of *An*EH, MD simulations were performed. Based on the apo wild-type structure of *An*EH, a model of the enzyme harboring either *(R)*- or *(S)*-GPE was constructed with the additional assumption of two tyrosines binding to the oxygen of the substrate. Similar models were built for all intermediate *An*EH mutants until the final LW202 mutant was reached. The same procedure was repeated for both enantiomers of the substrate as covalently bound to Asp192 in the form of an ester intermediate. The most significant finding is the correlation between the measured *E* values for the particular mutants and the differences in the calculated distance from the MD experiments (**Δd),** described as a distance between the attacking oxygen atom of Asp192 and the epoxide carbon atom (Fig. 17b). For the wild-type enzyme, the **Δd** is expected to be 0.8 Å, with the preferred *(S)*-enantiomer slightly closer to the attacking aspartic acid. In the high enantioselective mutant LW202, the **Δd** increases to 1.6 Å (Fig. 18). For the preferred *(S)*-substrate, the calculated distance does not change substantially, but the *(R)*-GPE is predicted to be positioned much further from Asp192 (d = 5.4 Å), which disfavors its ring-opening by nucleophilic attack (Table 3).

In summary, the structural changes in the active site created by the 9 mutations tend to disfavor binding of the *(R)*-enantiomer, making it more difficult for this enantiomer to be positioned within hydrogen bond-distance to the two tyrosines and at the same time to be close enough for the nucleophilic attack by Asp192.

Table 3. Results of MD calculations.

Mutations	d_R	d_S	Δd_{R-S}	*E*
Wild-type	4.3	3.5	0.8	4.6
L215F/A217N/R219S	4.8	4.0	0.8	14
As above plus M329P/L330Y	4.9	4.0	0.9	21
As above plus C350V	5.1	4.0	1.1	24
As above plus L249Y	5.1	3.9	1.2	35
LW202	5.4	3.8	1.6	115

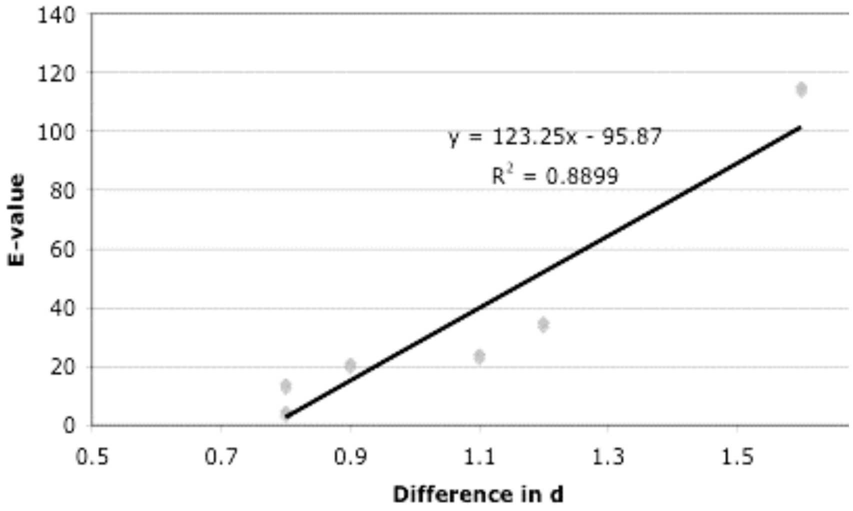


Figure 18. Relation between differences in the calculated distance d from Asp192 to the C-atom of epoxide and enantioselectivity (*E*) of the particular mutant.

2.3.5 Docking experiments

The predictions based on MD calculations agree with the x-ray structure obtained of the highly enantioselective mutant LW202. The automated docking studies on this particular protein were difficult due to the relatively high flexibility and several multiple conformations that were modeled during the structure refinement. Still some interpretations could be made based on a comparison of available structures of the wild-type and LW202. It becomes clear that upon the mutations, the active site shape and cavity change dramatically, especially around sites B and E (Fig 14c-d). Some of the differences are also due to the movement of Phe196 (part of site A, but not mutagenized in LW202).

When the substrate is manually docked into the active site of LW202 in a manner that provides the optimal positioning with respect to the Asp192 and two tyrosines, only the *(S)*-enantiomer fits easily, as it does in the wild-type structure. In the case of the *(R)*-enantiomer, this binding mode places the epoxide ring in an unproductive orientation, away from Asp192. In addition, its phenyl moiety lies in an "aromatic box" formed by Phe215, Tyr249 and Trp317 that causes severe steric clashes (Fig. 19).

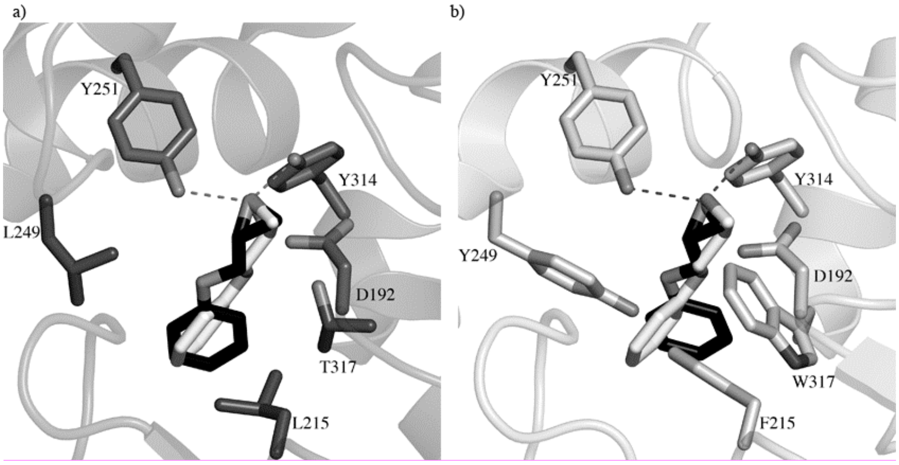


Figure 19. Active sites. a) The x-ray structure of the wild-type enzyme of *An*EH. b) LW202 mutant. The enantiomeric forms of GPE are manually docked and colored in white (*S*-GPE) and black (*R*-GPE). The catalytic Asp192, as well as the two tyrosines from the lid and three mutated residues, are labeled and shown in stick representation; interactions between tyrosines and epoxide oxygen atom are dashed.

2.3.6 Inhibition studies on mutated enzyme

Several efforts to soak crystals of mutated enzyme with valpromide were not successful; no electron density for the compound could be ever obtained. Later, the analysis of the apo structure of LW202 combined with inhibition studies revealed why this inhibitor could not be bound. Valpromide shows 10-fold weaker binding to the mutated protein than to the wild-type enzyme (K$_i$s of 2.5 mM and 0.25 mM, respectively). Also, the compound has two *n*-propyl groups ("arms"), which in a chiral environment like the active site of protein turn into diastereotopic groups (Fig 20a).

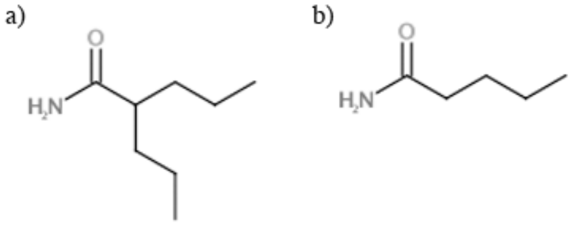


Figure 20. Structure of EH inhibitors a) valpromide and b) valeramide.

When compared to the substrate (GPE) used in the MD studies, the valpromide combines stereochemical features of both enantiomers in the sense that one arm corresponds to the CH_2OPh moiety of the favored *(S)*-GPE, and the other one to the placement of the CH_2OPh in the disfavored (*R*)-GPE. Since the latter one is not accepted by the mutant in a manner that ensures hydrogen bonds to the lid tyrosines and nucleophile at the same time, binding one of the valpromide arms is also disfavored, which makes this compound a poor inhibitor.

To confirm these ideas, another possible inhibitor was tested – valeramide (Fig 20b). This compound lacks one of the "arms" present in valpromide, which leaves the alkyl group free to rotate at the α-carbon. There is no structure of *An*EH with valeramide available, but we suspect that this kind of compound could easily suit the steric requirements in the chiral pocket of mutant enzyme. Based on the available structural data, valeramide is predicted to bind to the tyrosines (Tyr251 and Tyr314) and nucleophile Asp192 in analogous way to the favored substrate *(S)*-GPE. Kinetic studies in presence of *(S)*-GPE showed that valeramide is indeed a very efficient inhibitor, with K_i= 0.075 mM.

2.3.7 Conclusions

Based on kinetic studies, MD calculations, molecular modeling and x-ray structures, we analyzed the factors that are responsible for the increased enantioselectivity of a mutated variant of the *An*EH enzyme previously generated by the process of directed evolution. The enhanced enantioselectivity in the production of *(S)*-diol was based on the model reaction of racemic glycidyl phenyl ether hydrolysis, and the selectivity factor of the mutant LW202 described here was increased over 25-fold in comparison to the wild-type enzyme (*E*=4.6).

The molecular dynamics studies, together with analysis of structural data for the LW202 mutant, helped in the identification of the main reasons underlying this dramatic change in enzymatic behavior, showing that the preferred (*S*)-substrate binds easily in the binding pocket, whereas the (*R*)-GPE encounters many steric clashes arising from the ways that the mutations change the shape of the active site. Some of the changes involve residues that were not mutated, and so would be very difficult to predict.

This project shows the importance of obtaining x-ray structures in order to validate molecular modeling and computational calculations and help in the interpretations on a molecular level. The work described is also the first showing the structure of an enzyme with activity evolved by directed evolution methods.

2.4 Introducing novel microbial epoxide hydrolases (paper III)

This project was carried out in collaboration with the Verenium Corporation (former Diversa), which is located in San Diego, California. The company constructed libraries of DNA isolated from a wide variety of microenvironments around the world. By using activity-based high-throughput assays they discovered novel microbial EHs (about 50 in total). At the time when the project started, fewer than fifteen microbial epoxide hydrolases had been reported and very few of them carefully studied. All new proteins discovered by Verenium were found to be unique at the sequence level and have diverse substrate scope and enantioselectivity (Zhao, Han et al. 2004).

The main goal of the project, from the company's point of view, was to apply these enzymes to the synthesis of commercially important chiral diols/epoxides, since these are key intermediates for the biosynthesis of active pharmaceutical ingredients. On the other hand, the phylogenetic analysis showed that the new enzymes were highly diverse and represented a wide range of sequence space. Possible new structures could provide deeper insights into the general knowledge about EHs and help to understand the connection between their sequence/structure and substrate specificity. Therefore selected proteins were further characterized including x-ray structures determination and sequence analysis.

2.4.1 Functional studies

The newly discovered EHs have been screened with regard to their potential biocatalytic properties. In order to probe the functional diversity of these enzymes, several substrates representing both the terminal and internal, alkyl and aryl epoxides were used in screening. The studies showed that the set of enzymes exhibit broad substrate scope and novel activities. Active hits were found for every type of tested substrate, including those for which other known microbial epoxide hydrolases were ineffective (e. g. *cis*-stilbene oxide). In addition, many of these enzymes showed excellent enantioselectivity. For instance, desymmetrization of *meso*-epoxides was not effective by other known microbial EHs due to lack of activity or poor enantioselectivity (Fig. 21). Moreover, all previously reported EHs had been observed to predominantly form *R,R*-diols, whereas within the newly discovered EHs, the first enzymes providing access to *S,S*-diols were found (Zhao, Han et al. 2004). Table 5 shows a few examples of the enzymes exhibiting the highest enantioselectivity using particular substrates.

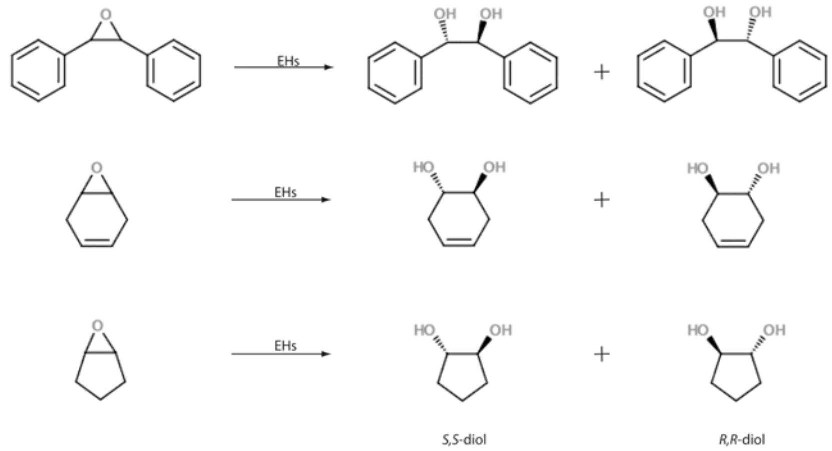


*Figure 21.*Desymmetrization of *meso*-epoxides carried out by the new microbial EHs.

Table 5. Broad substrate scope of the newly discovered EHs.

Epoxide-containing substrate	Best hits for *R*-isomer	Best hits for *S*-isomer
Cis-stilbene oxide	BD9300	BD9126
	BD9883	
	BD9884	
Cyclohexene oxide	BD9883	BD10159
	BD10332	BD10090
	BD10721	
3-chlorostyrene oxide	BD8876	BD8877
Epichlorohydrin	BD7520	BD8877

The results indicated the functional diversity of the discovered EHs and highlighted their potential use in many synthetic applications, such as the synthesis of key intermediates for the manufacture of important pharmaceuticals, including antibiotics, HIV protease inhibitors, anti-obesity and anti-depression drugs.

2.4.2 Protein purification, crystallization, and solving the structures

Some of the best enzymes identified above (Table 5) were characterized further. The selected proteins (12 in total) were overexpressed in *E. coli* as constructs fused with a poly-His tag at C-terminus for purification on nickel-immobilized metal affinity columns. The purified enzymes were subjected to size exclusion chromatography. Because of very fast precipitation between subsequent steps of purification, some further optimization was needed, which is described in detail in Table 6.

Table 6. *Different approaches were tested in order to avoid protein precipitation during purification. n/e means "no effect".*

Experiment	Description	Effect
Use of different buffers	Tris-HCl; sodium acetate; HEPES; CAPS	n/e
Try different pH of buffers	4.5; 7.5; 8.0; 9.0; 10.0	n/e
Change salt concentration	0; 350; 500; 1000; 2000 [mM]	No precipitation when >1000mM
Reduce protein concentration	1-5 [mg/ml]	n/e
Use of additives	L-arginine 50-400 [mM]	n/e
	Glycerol (10-30%)	n/e
Use of detergents	Triton X-100 (0.5%)	n/e
	Beta-octyl glucoside (0.5%)	n/e
Reduce concentration of imidazole in elution buffer	500 -> 200 [mM]	n/e
Change buffer immediately after affinity column purification	Use PD10 column between affinity and gel filtration steps	n/e
	Dialyse sample against gel filtration buffer without imidazole	n/e
	Quick buffer exchange by several dilutions and concentrations on Centricon/Vivaspin	No precipitation
Work at 4 °C	From the cell harvesting to setting up crystallization drops	No precipitation
Avoid freezing/thawing proteins	Work with fresh batch of protein, or one stored at 4 °C	No precipitation

To summarize, the proteins tended to precipitate when purified at room temperature, or with a high concentration of imidazole present in the buffers. Another limitation was the fact, that the proteins could not be stored at -20 °C and again thawed. Taking this under consideration, to avoid precipitation all steps of purification (including cells harvesting, affinity chromatography, and gel filtration) in further experiments, were performed at 4 °C or directly on ice, when needed. In addition, in order to get rid of imidazole, the protein buffer after elution from affinity column was immediately exchanged by several dilutions and concentrations on Centricon/Vivaspin columns with appropriate size cut-off. Concentrated protein samples in 50 mM Tris-HCl pH 8.0, 200 mM NaCl, and 10 mM β-ME were then applied to a size-exclusion chromatography column (Hiload^{TM} 16/60 Superdex^{TM} 200, Amersham Pharmacia Biotech). Eluted fractions containing significant amounts of the desired proteins as judged by SDS/PAGE analysis were concentrated to 20 mg/ml (A_{280}, $\varepsilon = 2.0$ ml mg^{-1}·cm^{-1}) by using a Centriprep 10K concentrator (Millipore) and stored at 4 °C.

Six proteins of interest (out of 12) were purified and obtained in the soluble fraction. Consequently they were entered into the crystallization trials, by using vapor diffusion method (mostly sitting drops, consisting of fresh protein and reservoir solution in 1:1 ratio). Four of the proteins were found to crystallize easily in different kinds of crystallization solutions, but always containing PEG. They tended to form crystal clusters rather than single forms. Therefore in order to increase the size and quality of the crystals, microseeding was carried out (Bergfors 2003) (Fig. 22), where the initial crystal clusters were crushed and used as a stock to seed several drops that had been left to equilibrate for 1-5 hours. During the seeding experiments, the concentration of precipitant was gradually lowered to finally reach 18-20% PEG, which was about 5-7% less than in the original drops where the crystal-clusters were found. After seeding, single crystals grew within 3-7 days and were big enough for data collection, diffracting to approximately 1.7 – 2 Å (for BD9883 and BD9300, respectively) and 2.5-3 Å (for BD8877 and BD9884).

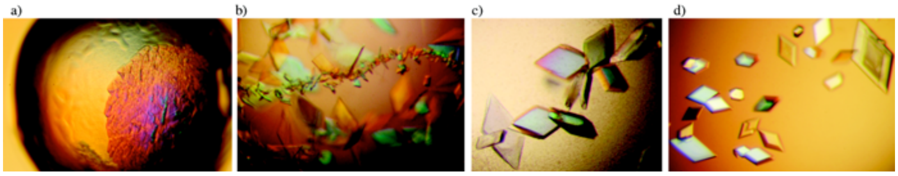


Figure 22. Streak-seeding method. The initial crystal-cluster (at panel a) was used to streak-seed the previously equilibrated drops (panels b-d). The characteristic streak-line, formed by micro-crystals is visible on panel b.

The structures of two proteins were solved (BD9300 and BD9883), both by molecular replacement using either AMoRe (Navaza 2001) or Phaser (McCoy, Grosse-Kunstleve et al. 2005). Since the EH structure from *A. niger* (Zou, Hallberg et al. 2000) had the highest sequence similarity (35% in equivalent regions) to the BD9300, it was used as a search model. The sequence identity between BD9300 and BD9883 was 38%, which let us use the functional dimer of the former to solve the structure of the latter one.

2.4.3 Structural results

The overall structures of BD9300 and BD9883 are compared in Fig. 23a-b. As for other EHs of the microsomal class, they consist of three major parts: an N-terminal meander (residues 1-68/2-71 in BD9300/BD9883, respectively), a main α/β-hydrolase domain (two segments built by residues 69-201/72–203 and 319-384/321-387), and a smaller lid domain (residues 216-310/215-310).

The long curved N-terminal meander has a hairpin-like structure built by several α-helices. It interacts with the other two domains and is also partly responsible for dimer stabilization. The main domain has a classical α/β-hydrolase fold. The lid domain is mainly α-helical and forms an insert in the main domain. These two domains are connected by the NL-loop (residues 202-215/204-214) and the CL-loop (residues 311-318/311-320) (Fig. 23c).

The two molecules found in each asymmetric unit form a homodimer, stabilized by the interactions between the α-helices of the lids and the N-terminal meanders from both subunits. The helices of the lid domain are completely buried in the dimer interface, and these regions have the lowest temperature factors (~16 and ~10 Å^2, in BD9300 and BD9883). The central parts of the both domains are the most ordered while the loops show higher flexibility. Also, when the two structures are compared, the r.m.s differences are ~1.2 Å (with 3.5 Å cut-off); the C$^\alpha$ atoms that do not match are located in the NL- and CL-loops, and in the cap-loop. The NL-loop has the highest crystallographic temperature factors (~40 Å^2), and this region has also the weakest (although visible) electron density in both structures (Fig. 23d-e).

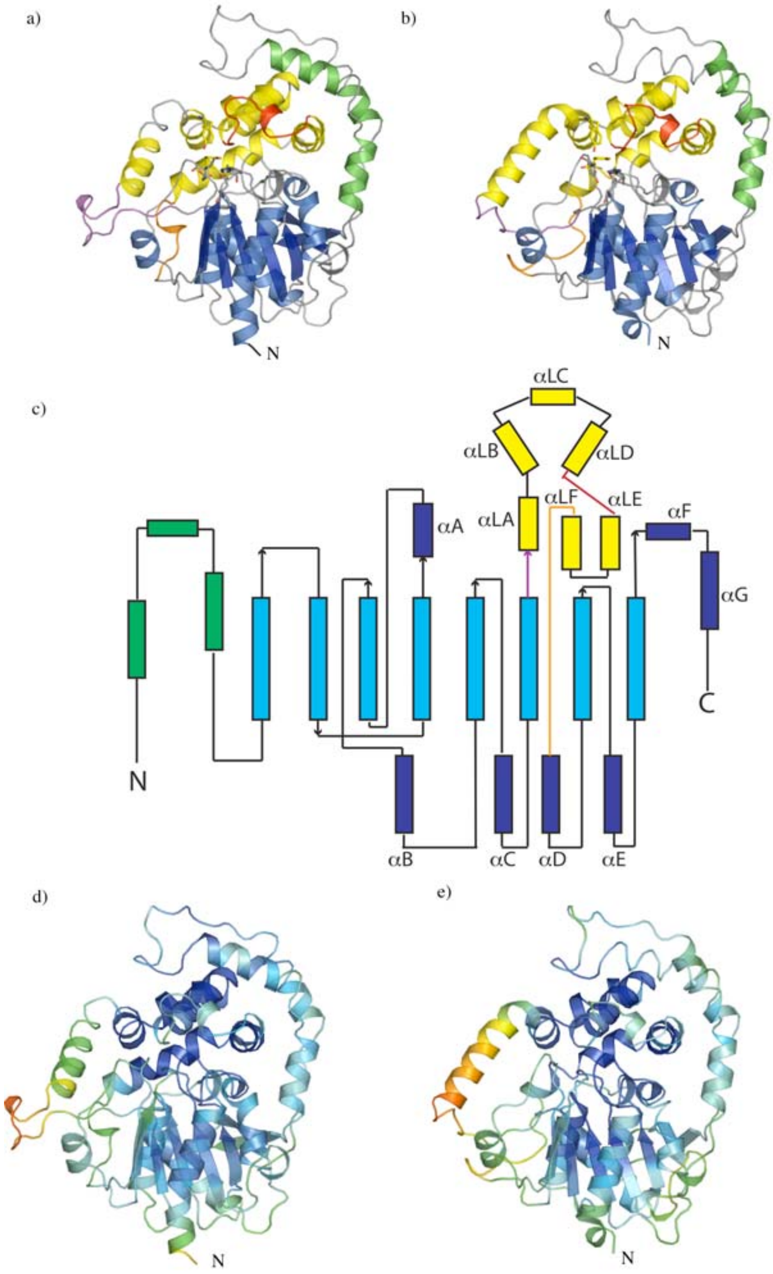

a)
N
b)
N
c)
αLC
αLB
αLD
αLF
αLE
αA
αLA
αF
αG
N
C
αB
αC
αD
αE
d)
N
e)
N

Figure 23. Overall structures. Comparison of a) BD9300 and b) BD9883. The N-terminal meander is shown in green, the main domain in blue and the lid domain in yellow. N- and C- termini are labeled. The NL-, CL- and cap-loops are colored in magenta, orange, and red, respectively Active site residues are shown in a stick representation. c) A topology diagram of BD9300 with the same color convention. d) BD9300 and e) BD9883 colored by crystallographic temperature factors.

The active sites of both proteins are located in the cleft between the two domains. The catalytic triad comes from the α/β-hydrolase domain. Of these, Asp176/179 (BD9300/BD9883) is the nucleophile, and His358 together with Glu332 (His361/Asp334 in BD9883) function as the general base-charge relay. From the lid domain, a pair of tyrosines Tyr235/Tyr304 in BD9300, and Tyr234/Tyr304 in BD9883 hydrogen bonds to the oxygen of the epoxide ring and are thought to assists in its opening.

The active sites of the proteins are compared on Fig. 24. Electron density for PEG was observed in the active sites of both subunits in BD9883, modeled as tri- and di-ethylene glycol. The oxygen of its terminal hydroxyl group binds to the lid tyrosines, and one carboxylate oxygen of the nucleophile (Asp179) (Fig. 24a). The other carboxylate oxygen of Asp179 interacts with the main-chain amide nitrogens of two nearby tryptophan residues (Trp103 and Trp180). The remainder of the visible PEG density extends out towards the solvent.

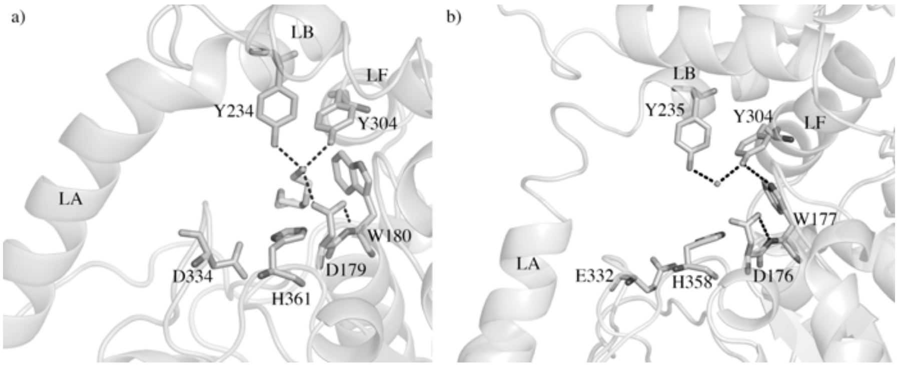


Figure 24. Active sites. The active sites are shown for BD9883 and BD9300 in panels a and b, respectively. Side chains of the catalytic residues are shown and labeled. The active site water and PEG molecules are shown and helixes of the lid domain labeled as LA, LB and LF.

The position of the terminal hydroxyl group of the PEG in BD9883 is occupied by a water molecule in BD9300 (Fig. 24b). Equivalent interactions are formed with the lid tyrosines, but the distance to the nucleophile Asp176 is larger (~4 Å). As in BD9883, Asp176 also binds to two tryptophan residues (Trp100 and Trp177), and lies near His358, which in turn hydrogen-bonds to its partner in the charge relay, Glu332 (which are equivalent to His361 and Asp334 in BD9883).

2.4.4 Structure vs. selectivity of EHs

Structural comparisons with other enzymes from the α/β-hydrolase superfamily are summarized in Table 7. Only six epoxide hydrolases within this family are structurally available till date. Among those, the most closely related to either BD9300 or BD9883 is the fungal EH from *A. niger* (*An*EH). The next structure is that of a bacterial EH from *A. radiobacter* (*Ar*EH), with 22% sequence identity, although due to a possible error in part of this structure (involving the active site), it will not be discussed further. More distant are the EH from *M. tuberculosis (Mt*EH), as well as the plant (*St*EH1) and mammalian soluble enzymes (*Hs*sEH and *Mms*EH).

Among all available structures, the most similar part is within the α/β-hydrolase domain. Both helices and strands are of more or less equal size and fold. The largest differences in this domain involve loops that fold out to the solvent. The lid domains, when compared, have only two residues truly conserved, both pointing into the active site and assisting the opening of a ring in the epoxide-containing substrate.

The most striking differences in the lid are seen in the length and fold of the first (αLA) and the last helix (αLF). The first one is rather long in BD9300 and BD9883, as it is in *An*EH, whereas other structures like mammalian EHs and *St*EH1 have very short αLAs (6-8 residues comparing to 20 of BD9300). Usually this helix leads into the active site and is directly connected with αLB, where one of the catalytic tyrosines is located. In the *Mt*EH, the first helix of the lid is short, as is seen in *Hs*sEH, but here is followed by a long loop forming an extra β–sheet (not present in other structures) and it folds out toward the solvent rather than interacting with the second helix of the lid. The next three helices (αLB – αLE) are placed similarly and have approximately the same lengths in all structures. They form a triangle over the active site and are connected to the last two helices (αLE and αLF) by the cap-loop, which varies in length, and in some structures like BD9300/BD9883 and *St*EH1, also forms a short helix (~4 residues) in the middle part. *Mt*EH has the longest cap-loop (~67 residues), and although partially disordered, it forms an additional α-helix and a small β-sheet. The last helix of the lid (αLF) is also very variable. In some structures it is relatively long (~20 residues), as seen in *St*EH1, *Hs*sEH or *Mt*EH, but it also may be very short, like in *Ar*EH (10) and *Mms*EH (7 residues). In both our proteins, this helix is medium sized (13-14 residues).

The lid domain forms an insertion in the main α/β-hydrolase domain and is connected to the latter by two loops (NL- and CL-loops). The manner in which the NL-loop folds, depends greatly on the nature of the first helix in the lid (αLA). Since this helix is of similar size in both our proteins, as well as in *An*EN, the intrinsic nature of the preceding NL-loop is analogous. Nevertheless, the positioning of the equivalent helices and loops in these structures differs a lot, which affects the volume of their active-sites (Fig 25). Since in BD9883 the NL-loop and following helix (αLA) are positioned much closer to the last helix of the lid (αLF), the tunnel leading to the active site pocket is much more narrow than in BD9300 and *An*EH (Fig. 26a-c). Both *St*EH1 and *Hs*sEH have relatively short NL-loops, which place αLA on the opposite side of the active-site cavity compared to the other enzymes (much closer to αLB and αLF); the result is in each case a large difference in the entrance to the active site. In *Ar*EH, the NL-loop and αLA are disordered (residues 137-149), whereas in *Mt*EH this loop has a completely different structure and topology, which results in a smaller opening leading from the solvent to the catalytic residues.

The second loop, connecting the two domains, also varies a lot among the different EH structures. For instance, in *St*EH1, *Mt*EH *Hs*sEH, the CL-loop folds into a helix that is connected to αLF by a kink (near Pro239 of *St*EH1). In *An*EH this part of the structure is disordered (residues 319-329), and cannot be compared. In BD9883, the shift in the NL-loop (in comparison to corresponding fragment of BD9300) moves the CL-loop that brings it further outside the active-site cavity than its equivalent in BD9300 (Fig. 25).

Table 7. Comparison of the BD9300 and BD9883 to the available structures within the α/β-hydrolase family.

Protein/Source	Number of residues	No of atoms within 3.5Å cut-off (% id of matched)	r.m.s.d.	No of atoms within 3.5Å cut-off (% id of matched)	r.m.s.d.	Catalytic triad	PDB code
		Comparing to *BD9300*		Comparing to *BD9883*			
*An*EH	385	267 (33)	1.27	312 (36)	1.24	D192/H374/D348	1QO7
*Ar*EH	282	238 (21)	1.5	230 (23)	1.44	D107/H275/D246[a]	1EHY
*Mt*EH	352	208 (19)	1.74	200 (18)	1.61	D104/H333/D302	2E3J
*St*EH1	321	229 (15)	1.78	225 (21)	1.67	D105/H300/D265	2CJP
*Hs*sEH	302[b]	196 (18)	1.8	208 (19)	1.78	D333/H523/D495	1VJ5
*Mm*sEH	310[c]	193 (18)	1.8	206 (18)	1.74	D333/H523/D495	1EK2
Bromoperoxidase A1 *(Streptomyces aureofaciens)*	274	178 (16)	1.7	187 (15)[d]	1.84	S94/H252/D223	1A8Q
Hydroxynitrile lyase *(Hevea brasiliensis)*	257	160 (9)	1.75	160 (8)	1.88	S80/H235/D207	1YAS
Haloalkane dehalogenase (Xanthobacter autotrophicus)	310	165 (17)	1.79	190 (14)	2.0	D124/H252/D223	2HAD
Lipase *(Candidia antarctica)*	317	102 (14)	1.8	111 (13)	2.0	S105/H224/D187	1TCB
Proline iminopeptidase *(Xanthomona campestris)*	313	146 (23)	1.86	146 (19)	1.85	S110/H294/D266	1AZW
Chloroperoxidase L *(Streptomyces lividans)*	275	195 (17)	1.86	200 (17)	1.9	S94/H253/D224	1A8S

---

[a] The third member of the active site triad is assigned by comparison with other sequences/structures.

[b] Using only domain 2, residues 245-547.

[c] Using only domain 2, residues 245-254.

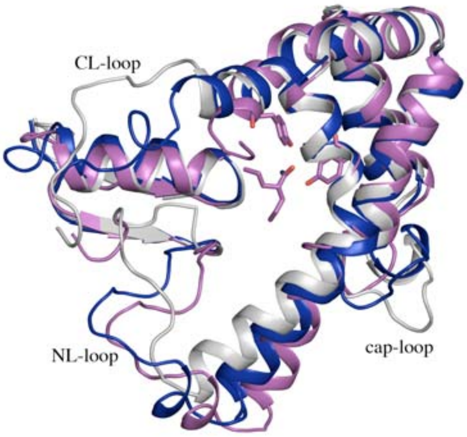

CL-loop
NL-loop
cap-loop

Figure 25. The entry to the active site of superimposed BD9300, BD9883 and *An*EH with valpromide bound (PDB entry code 3G0I). BD9300 is colored blue, BD9883 by white, and *An*EH magenta. The active site tyrosines and valpromide molecule are shown in stick representations.

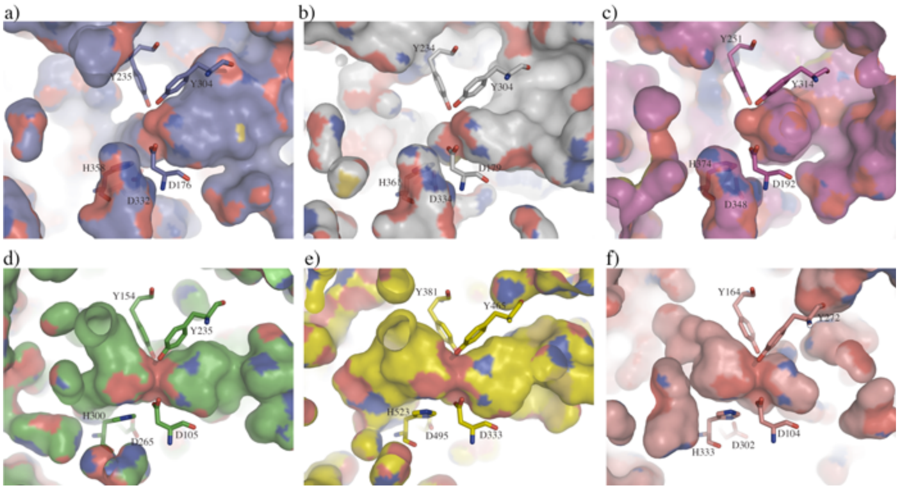

a)
Y235
Y304
H358
D176
D332
b)
Y234
Y304
D179
H365
D334
c)
Y251
Y314
H374
D192
D348
d)
Y154
Y235
H300
D105
D265
e)
Y381
Y465
H523
D495
D333
f)
Y164
Y272
D104
H333
D302

Figure 26. Comparison of the active site cavities. The active sites of BD9300, BD9883, *An*EH, *St*EH1, *Hs*sEH, and *Mt*EH are shown in the panels a-f, respectively. The views are the same, with the "outside" of the active sites (the one facing the solvent) to the right in each case. The catalytic residues are shown in stick representation and labeled.

Structure comparisons among reported EHs show that the differences in the lid domain, as well as connected loops, definitely affect not only the shape, but also size and accessibility of the active site (Fig. 25 and 26). The question that remains unanswered is how those differences in size/shape of the cavities underlie the substrate specificities of particular enzymes. One could suspect that the enzymatic regioselectivity (i.e. preference for which of the carbons in the epoxide ring to attack) depends on the substrate's placement in the active site. This in turn, dictates which particular diol is produced (see Fig. 6).

Also, if we compare the space in the active sites that is available for incoming substrates, it becomes clear that among available EH structures only in BD9300, BD9883 and *An*EH is the "inside" part of the cavity not available (at left in Fig. 26a-c). This is mostly caused by a tryptophan residue that is located at the end of αLD in these proteins and points into the active site (Trp268 in BD9300). In the rest of the enzymes, residues with smaller side chains replace the "big" tryptophan, at the same time leaving more space in the active site. This observation is correlated with the fact that highly substituted substrates have been shown to bind only to enzymes like *St*EH1. We predict that those EHs that lack tryptophan at this position will also have broader cavities in the binding pocket, and be able to accept more highly substituted substrates.

Since only a few structures of EHs have been solved till now, the systematic analysis that could provide some "general" architecture allowing predictions of substrate specificity for new sequences was not possible. However some attempts of clustering EHs by their sequence alignments were done and published in 2004 (Barth, Fischer et al. 2004). At that time the only available structures were those of *Ar*EH, *An*EH and *Mms*EH. Based on this limited structure information and sequence alignments of other EHs, the authors divided EH sequences into three different clusters. The main criteria that were used here were based on the length and secondary structures of two loops, of which one is the defined earlier as the cap-loop, whereas the second, described by Barth *et al.* as NC-loop, covers the NL-loop together with the first helix of the lid (αLA). The authors have claimed that the clustering by loop length is highly correlated with substrate specificity and could help in structure modelling of new epoxide hydrolases.

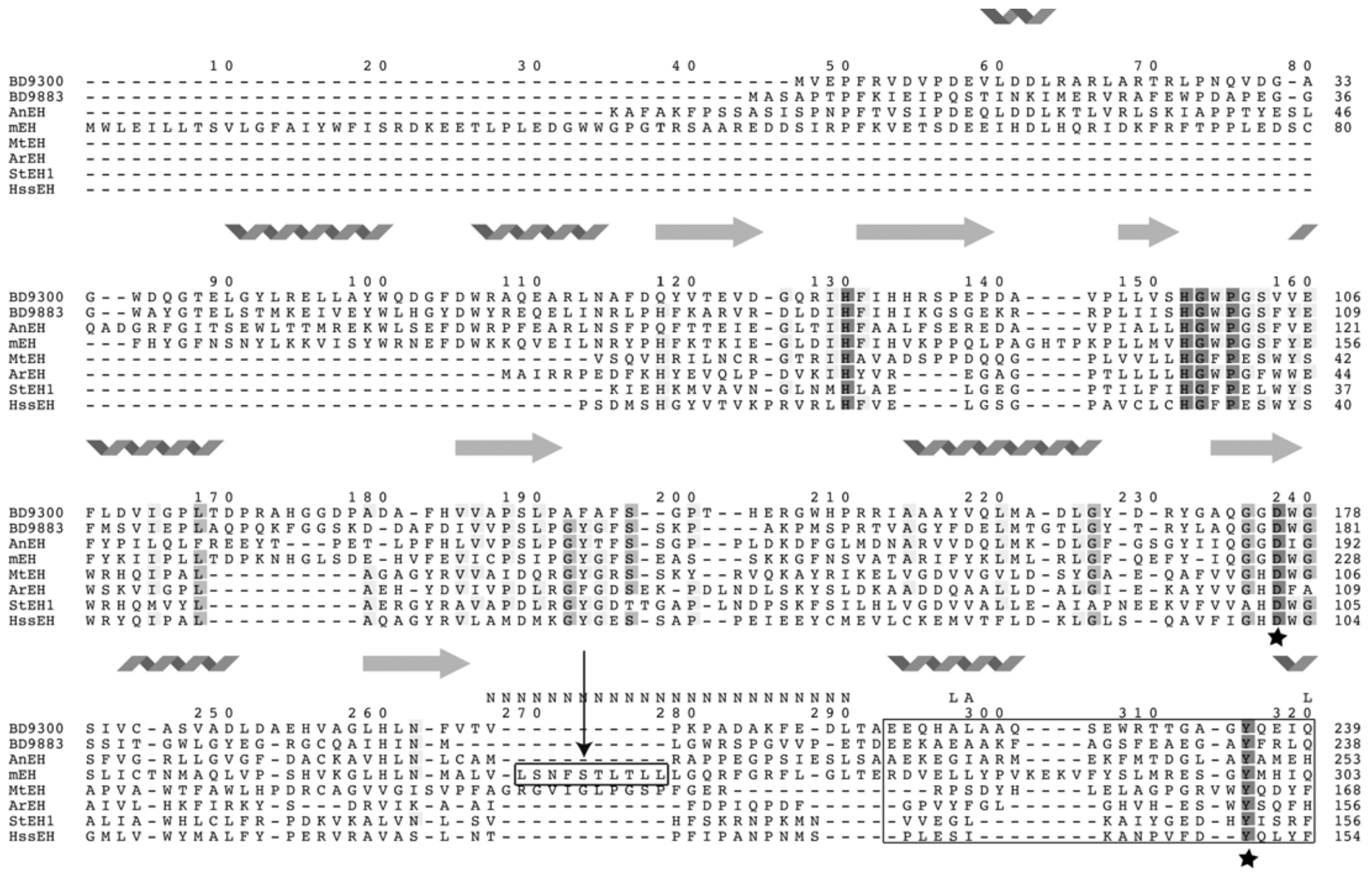
Since then, the number of available structures has doubled and comparison of only BD9300, BD9883 and *An*EH (Fig. 25) shows clearly that even if the length of NC-loops is similar, they can fold differently within the active site and its surroundings. Moreover, none of the cap-loops in either structure lies closely enough to the active site to affect its shape or determine the substrate specificity of the enzyme.

When comparing the kinetic data that are available for these three proteins, it becomes clear that even quite similar enzymes (in terms of sequence, size of loops and overall fold) can have very different activities. For instance, the enzymes presented here are highly enantioselective towards the formation of *R,R*-diols from *cis*-stilbene oxide and cyclopentene oxide (Zhao, Han et al. 2004). In contrast, the most similar to them, *An*EH as a wild-type, catalyzes the conversion of GPE with only slight favor towards its *S*-enantiomer, while the LW202 mutant (described in chapter 2.3) has this preference increased more than 100-fold (Reetz, Bocola et al. 2009).

2.4.5 Sequence comparisons

Sequence identity between available EH sequences is generally very low. The most conserved regions are within the main domain, which is reflected in the structure comparison. The variable size and fold of the lid domains as well as connected NL- and NC-loops are correlated with the fact that, within this region, only two active-site tyrosines are conserved.

A structure-based sequence alignment, including EHs of known structures as well as the mEH, is shown in Fig. 27. The analysis of those sequences combined with the kinetic data could provide some suggestions in terms of predicting their activity. However, very often direct comparison of available data is impossible, since many studies published till now were designed differently and the enzymes were tested on various different compounds, often in the form of a racemic mixture. Based on the sequence alignment, similar to one shown in Fig. 27, but containing also other sequences of the most interesting enzymes presented here, a phylogenetic tree was constructed (Fig. 28). It reveals interesting trends regarding enzymatic activity where *R,R*- and *S,S*-diol producing enzymes seem to cluster together in the sequence relationships.


10 20 30 40 50 60 70 80
BD9300 --MVEPFRVDVPDEVLDDLRARLARTRLPNQVDG-A 33
BD9883 ---MASAPTPFKIEIPQSTINKIMERVRAFEWPDAPEG-G 36
AnEH ----------------------------------KAFAKFPSSASISPNPFTVSIPDEQLDDLKTLVRLSKIAPPTYESL 46
mEH MWLEILLTSVLGFAIYWFISRDKEETLPLEDGWWGPGTRSAAREDDSIRPFKVETSDEEIHDLHQRIDKFRFTPPLEDSC 80
MtEH --
ArEH --
StEH1 --
HssEH --
90 100 110 120 130 140 150 160
BD9300 G--WDQGTELGYLRELLAYWQDGFDWRAQEARLNAFDQYVTEVD-GQRIHFIHHRSPEPDA----VPLLVSHGWPGSVVE 106
BD9883 G--WAYGTELSTMKEIVEYWLHGYDWYREQELINRLPHFKARVR-DLDIHFIHIKGSGEKR----RPLIISHGWPGSFYE 109
AnEH QADGRFGITSEWLTTMREKWLSEFDWRPFEARLNSFPQFTTEIE-GLTIHFAALFSEREDA----VPIALLHGWPGSFVE 121
mEH ---FHYGFNSNYLKKVISYWRNEFDWKKQVEILNRYPHFKTKIE-GLDIHFIHVKPPQLPAGHTPKPLLMVHGWPGSFYE 156
MtEH ---------------------------------VSQVHRILNCR-GTRIHAVADSPPDQQG----PLVVLLHGFPESWYS 42
ArEH ---------------------------MAIRRPEDFKHYEVQLP-DVKIHYVR----EGAG----PTLLLLHGWPGFWWE 44
StEH1 ----------------------------------KIEHKMVAVN-GLNMHLAE----LGEG----PTILFIHGFPELWYS 37
HssEH --------------------------------PSDMSHGYVTVKPRVRLHFVE----LGSG----PAVCLCHGFPESWYS 40
170 180 190 200 210 220 230 240
BD9300 FLDVIGPLTDPRAHGGDPADA-FHVVAPSLPAFAFS--GPT--HERGWHPRRIAAAYVQLMA-DLGY-D-RYGAQGGDWG 178
BD9883 FMSVIEPLAQPQKFGGSKD-DAFDIVVPSLPGYGFS-SKP----AKPMSPRTVAGYFDELMTGTLGY-T-RYLAQGGDWG 181
AnEH FYPILQLFREEYT---PET-LPFHLVVPSLPGYTFS-SGP--PLDKDFGLMDNARVVDQLMK-DLGF-GSGYIIQGGDIG 192
mEH FYKIIPLLTDPKNHGLSDE-HVFEVICPSIPGYGFS-EAS---SKKGFNSVATARIFYKLML-RLGF-QEFY-IQGGDWG 228
MtEH WRHQIPAL----------AGAGYRVVAIDQRGYGRS-SKY--RVQKAYRIKELVGDVVGVLD-SYGA-E-QAFVVGHDWG 106
ArEH WSKVIGPL----------AEH-YDVIVPDLRGFGDSEK-PDLNDLSKYSLDKAADDQAALLD-ALGI-E-KAYVVGHDFA 109
StEH1 WRHQMVYL----------AERGYRAVAPDLRGYGDTTGAP-LNDPSKFSILHLVGDVVALLE-AIAPNEEKVFVVAHDWG 105
HssEH WRYQIPAL----------AQAGYRVLAMDMKGYGES-SAP--PEIEEYCMEVLCKEMVTFLD-KLGLS--QAVFIGHDWG 104
NNNNNNNNNNNNNNNNNNNNNNNN
LA
L
250 260 270 280 290 300 310 320
BD9300 SIVC-ASVADLDAEHVAGLHLN-FVTV-----------PKPADAKFE-DLTAEEQHALAAQ----SEWRTTGA-GYQEIQ 239
BD9883 SSIT-GWLGYEG-RGCQAIHIN-M--------------LGWRSPGVVP-ETDEEKAEAAKF----AGSFEAEG-AYFRLQ 238
AnEH SFVG-RLLGVGF-DACKAVHLN-LCAM-----------RAPPEGPSIESLSAAEKEGIARM----EKFMTDGL-AYAMEH 253
mEH SLICTNMAQLVP-SHVKGLHLN-MALV-LSNFSTLTLLLGQRFGRFL-GLTERDVELLYPVKEKVFYSLMRES-GYMHIQ 303
MtEH APVA-WTFAWLHPDRCAGVVGISVPFAGRGVIGLPGSPFGER-------------RPSDYH----LELAGPGRVWYQDYF 168
ArEH AIVL-HKFIRKY-S---DRVIK-A-AI------------FDPIQPDF------GPVYFGL------GHVH-ES-WYSQFH 156
StEH1 ALIA-WHLCLFR-PDKVKALVN-L-SV-----------HFSKRNPKMN-----VVEGL--------KAIYGED-HYISRF 156
HssEH GMLV-WYMALFY-PERVRAVAS-L-NT-----------PFIPANPNMS-----PLESI--------KANPVFD--YQLYF 154
★
★

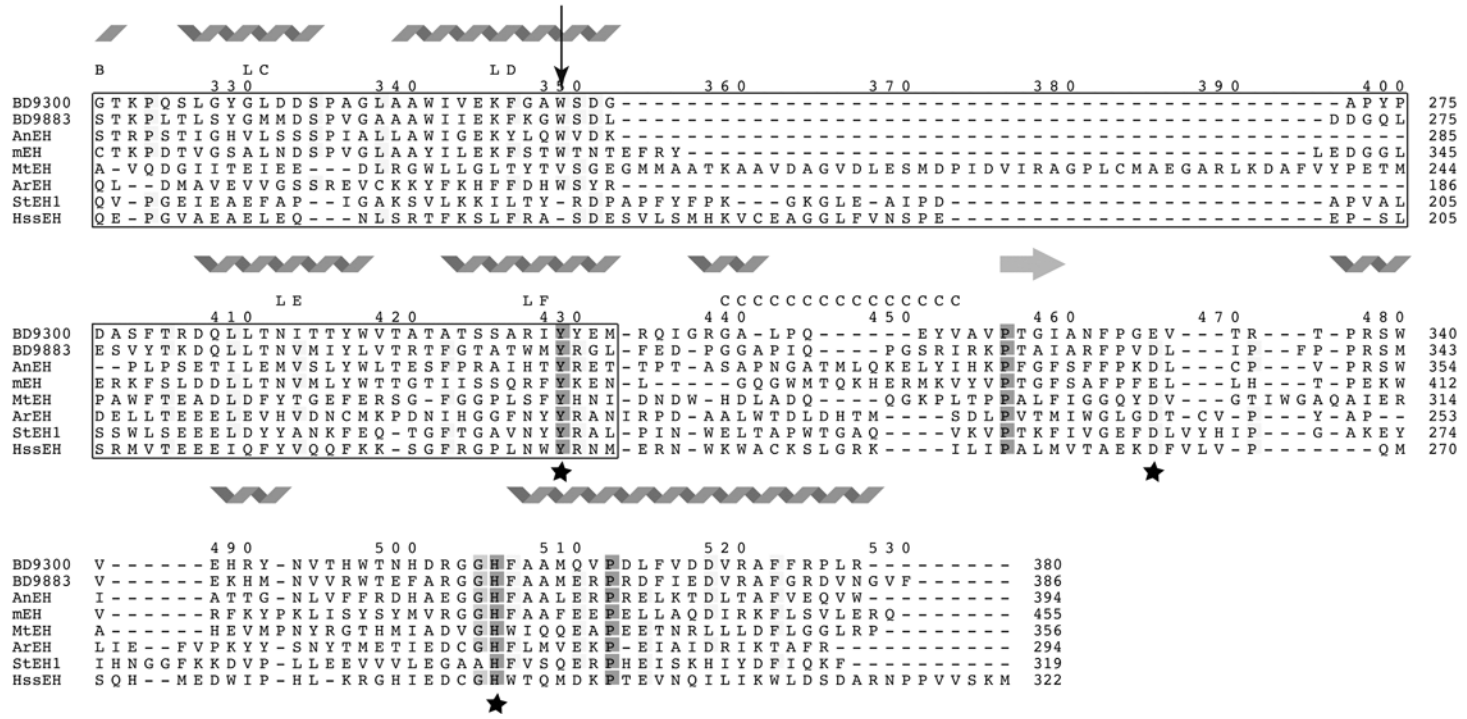

B
LC
LD
330
340
350
360
370
380
390
400
BD9300 GTKPQSLGYGLDDSPAGLAAWIVEKFGAWSDG--APYP 275
BD9883 STKPLTLSYGMMDSPVGAAAWIIEKFKGWSDL---DDGQL 275
AnEH STRPSTIGHVLSSSPIALLAWIGEKYLQWVDK-- 285
mEH CTKPDTVGSALNDSPVGLAAYILEKFSTWTNTEFRY--------------------------------------LEDGGL 345
MtEH A-VQDGIITEIEE---DLRGWLLGLTYTVSGEGMMAATKAAVDAGVDLESMDPIDVIRAGPLCMAEGARLKDAFVYPETM 244
ArEH QL--DMAVEVVGSSREVCKKYFKHFFDHWSYR-- 186
StEH1 QV-PGEIEAEFAP--IGAKSVLKKILTY-RDPAPFYFPK---GKGLE-AIPD-----------------------APVAL 205
HssEH QE-PGVAEAELEQ---NLSRTFKSLFRA-SDESVLSMHKVCEAGGLFVNSPE-----------------------EP-SL 205
LE
LF
CCCCCCCCCCCCCCC
410
420
430
440
450
460
470
480
BD9300 DASFTRDQLLTNITTYWVTATATSSARIYYEM-RQIGRGA-LPQ------EYVAVPTGIANFPGEV---TR---T-PRSW 340
BD9883 ESVYTKDQLLTNVMIYLVTRTFGTATWMYRGL-FED-PGGAPIQ----PGSRIRKPTAIARFPVDL---IP--FP-PRSM 343
AnEH --PLPSETILEMVSLYWLTESFPRAIHTYRET-TPT-ASAPNGATMLQKELYIHKPFGFSFFPKDL---CP---V-PRSW 354
mEH ERKFSLDDLLTNVMLYWTTGTIISSQRFYKEN-L-----GQGWMTQKHERMKVYVPTGFSAFPFEL---LH---T-PEKW 412
MtEH PAWFTEADLDFYTGEFERSG-FGGPLSFYHNI-DNDW-HDLADQ----QGKPLTPPALFIGGQYDV---GTIWGAQAIER 314
ArEH DELLTEEELEVHVDNCMKPDNIHGGFNYYRANIRPD-AALWTDLDHTM----SDLPVTMIWGLGDT-CV-P---Y-AP-- 253
StEH1 SSWLSEEELDYYANKFEQ-TGFTGAVNYYRAL-PIN-WELTAPWTGAQ----VKVPTKFIVGEFDLVYHIP---G-AKEY 274
HssEH SRMVTEEEIQFYVQQFKK-SGFRGPLNWYRNM-ERN-WKWACKSLGRK----ILIPALMVTAEKDFVLV-P-------QM 270
490
500
510
520
530
BD9300 V------EHRY-NVTHWTNHDRGGHFAAMQVPDLFVDDVRAFFRPLR--------- 380
BD9883 V------EKHM-NVVRWTEFARGGHFAAMERPRDFIEDVRAFGRDVNGVF------ 386
AnEH I------ATTG-NLVFFRDHAEGGHFAALERPRELKTDLTAFVEQVW--------- 394
mEH V------RFKYPKLISYSYMVRGGHFAAFEEPELLAQDIRKFLSVLERQ------- 455
MtEH A------HEVMPNYRGTHMIADVGHWIQQEAPEETNRLLLDFLGGLRP-------- 356
ArEH LIE--FVPKYY-SNYTMETIEDCGHFLMVEKP-EIAIDRIKTAFR----------- 294
StEH1 IHNGGFKKDVP-LLEEVVVLEGAAHFVSQERPHEISKHIYDFIQKF---------- 319
HssEH SQH--MEDWIP-HL-KRGHIEDCGHWTQMDKPTEVNQILIKWLDSDARNPPVVSKM 322

Figure 27. Structure-based sequence alignment. The sequences were obtained from GenBank {Benson, 2003 #248} with accession numbers as follows: *St*EH1 (75102548), *Hs*sEH (27597073), *Ar*EH (2292731), *An*EH (6165234), *Mt*EH (161760896), mEH (4503583). The first 233 residues of *Hs*sEH (the phosphatase domain) are omitted. Secondary structural elements of BD9300 were assigned with the structure in hand, while the various loops discussed in the text are marked as N and C, respectively, for the NL- and CL-loops. The lid domain is boxed, the catalytic residues are indicated by stars, and shading shows the sequence conservation. The tryptophan at the end of αLD discussed in the text is indicated by an arrow, as well as the insertion within the NL-loop in mEH.

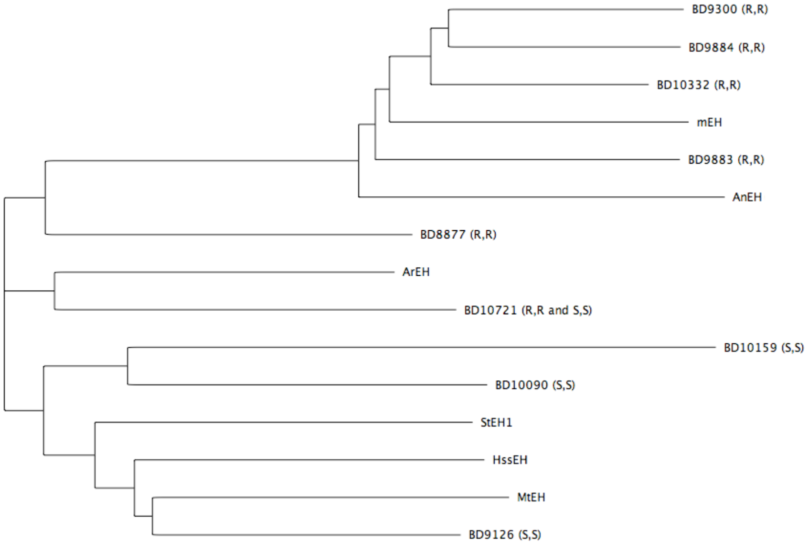


Figure 28. Phylogenetic tree. The analysis is based on the sequence alignment described in text. The reported enantioselectivities in the production of *R,R*/*S,S*-diol product by the hydrolysis of *meso*-epoxides are shown in the parentheses.

2.4.6 Shedding light on possible structure of human microsomal EH

The organisms from which these two proteins originate are not known, but the sequence comparisons showed that they have the greatest similarity among all EH structures to the poorly understood microsomal branch of the EH family. This family includes mammalian microsomal EH (mEH), the structure of which still remains unavailable. Previously, *An*EH has been the only structure representing this class of EHs. The sequences of all members of the mEH subfamily are on average 100-120 amino acids longer than most other EHs, mainly due to a hydrophobic N-terminal extension (shown in Fig. 27). The mammalian mEH was discovered decades ago, and since then many efforts have been made to purify and crystallize the enzyme, so far without success. Even when the N-terminal extension is removed form the constructs, the protein remains insoluble.

Comparison of the sequences between our two proteins and human mEH provide deeper insights into a possible structure of the latter one. As seen in Fig. 27, mEH has an additional hydrophobic insertion, which in the sequence alignment is located in the NL-loop region (of BD9300 or BD9883). Figure 24a shows clearly that this loop folds out from the main domain to the solvent and lies on the surface of the protein, at the entry to the active site. Considering the fact that mEH is associated mainly with the membranes, such an insertion could place this enzyme in an arrangement that makes it more effective in metabolizing the hydrophobic epoxides, among which many are the xenobiotic derived compounds.

2.4.7 Conclusions

The main goal of this project was a structural characterization of several epoxide hydrolases that had been recently discovered from microbial sources. Within this set of enzymes, many were found to exhibit high enantioselectivity levels and broad substrate scopes. The structures of two proteins have been solved and refined to high resolution. Both of them were previously shown to produce chemically valuable 1,2-diols and exhibit high enantioselectivity in reactions for which other known microbial EHs were not effective. Sequence analysis showed that these enzymes are also the most similar structures available to the mammalian microsomal EH, a key enzyme in degradation of xenobiotics. The structural investigation provided new ideas about this poorly understood but important enzyme. Furthermore, structural comparisons between these new structures and other EHs from the α/β-hydrolase family provided deeper insights on the connection between their sequence/structure and substrate specificity. Based on the structural work, we proposed an improved sequence alignment, showing that the differences in the lid, and the loops connecting it to the main domain, are the main factors that underlie the very wide spectrum of EH's substrates.

3 Future perspectives

All projects described in the thesis provided many new insights in the analysis of the epoxide hydrolases from the α/β-hydrolase family. Although these enzymes have been known for decades now, and much has been published about their general catalytic mechanism, the structural information still remains insufficient to allow detailed predictions of any kind concerning their enzymatic selectivity. Sequence comparisons can provide some suggestions, but because of generally very low levels of sequence identity (particularly in the lid domain and the loops connecting it to the α/β-hydrolase domain), there are still important gaps in the EH story that remain and should be addressed in future. The EHs presented here represent a collection of enzymes with various types of substrate scopes, activity and selectivity. The potential of such a library is enormous but more consistent kinetic studies together with systematic analysis on the sequence and structural level are still needed.

To date, the EH from *S. tuberosum (St*EH1) is the only example of a structure of a predominantly *S*-selective enzyme. The sequence comparisons to other EHs from plants imply that they might prefer *S*-enantiomers as well, which suggests new candidates for further work.

The mutated EH from *A. niger* (LW202) described here was the first reported structure of a mutated enzyme after successful directed evolution that had enhanced its enzymatic properties. The analysis of the x-ray structure was accompanied by MD calculations at all stages of evolutionary pathway. This project shows that combining computational methods with structural studies could help in the future in obtaining a better understanding of the catalytic mechanism of the proteins. The process of directed evolution could be productively applied in further experiments on some of newly discovered EHs isolated from environmental samples.

Within this set of enzymes, the two new proteins that have their structure determined might help in solving the insolubility problem of microsomal EH from humans, which in turn could lead to solving the most desired structure among all EHs. This branch of the enzymes has the lowest sequence similarity, and therefore more structures of the microbial EH type are almost certainly needed.

Furthermore, some of this family of enzymes show very interesting and novel enantio- and regioselectivities, like for instance BD9126, BD10721 and BD10159, which preferentially produce *S*,*S*-diols in the desymmetrization of *meso*-epoxides. These proteins have been purified but never crystallized; solving those structures would be highly interesting and could give useful ideas on how the substrate is positioned in the binding pocket of EH. So far, the simple interpretations based on the shape of active site are not sufficient for a deeper understanding of detailed mechanism of EHs.

Another interesting enzyme, which deserves further investigation, is BD8877, the most active in formation of *R,R*-diols from *meso*-epoxides (among reported EHs). The sequence comparisons suggest that the cavity of its active site might be broad enough to accept highly substituted epoxides, which to our knowledge was never tested but could give many interesting results.

Finally, some of these proteins are possible candidates for trapping an alkylenzyme intermediate, which would provide many additional insights into the catalytic mechanism of those enzymes. Since it is the hydrolysis step that is generally thought to be rate limiting, trapping such an intermediate is in theory possible. To date, I have made many unsuccessful attempts at cocrystallizing and soaking the EHs crystals with epoxide-containing substrates, but there is no x-ray structure available of such complex yet.

References

Arand, M., B. M. Hallberg, et al. (2003). "Structure of Rhodococcus erythropolis limonene-1,2-epoxide hydrolase reveals a novel active site." Embo J **22**(11): 2583-92.

Arand, M., H. Hemmer, et al. (1999). "Cloning and molecular characterization of a soluble epoxide hydrolase from Aspergillus niger that is related to mammalian microsomal epoxide hydrolase." Biochem J **344 Pt 1**: 273-80.

Archelas, A. and R. Furstoss (2001). "Synthetic applications of epoxide hydrolases." Curr Opin Chem Biol **5**(2): 112-9.

Argiriadi, M. A., C. Morisseau, et al. (2000). "Binding of alkylurea inhibitors to epoxide hydrolase implicates active site tyrosines in substrate activation." J Biol Chem **275**(20): 15265-70.

Barbirato, F., J. C. Verdoes, et al. (1998). "The Rhodococcus erythropolis DCL14 limonene-1,2-epoxide hydrolase gene encodes an enzyme belonging to a novel class of epoxide hydrolases." FEBS Lett **438**(3): 293-6.

Barth, S., M. Fischer, et al. (2004). "Sequence and structure of epoxide hydrolases: A systematic analysis." Proteins-Structure Function and Bioinformatics **55**(4): 846-855.

Bergfors, T. (2003). "See to crystals." J Struct Biol **142**: 66-76.

Biswal, B. K., C. Morisseau, et al. (2008). "The molecular structure of epoxide hydrolase B from Mycobacterium tuberculosis and its complex with a urea-based inhibitor." J Mol Biol **381**(4): 897-912.

Blee, E. (2002). "Impact of phyto-oxylipins in plant defense." Trends Plant Sci **7**(7): 315-22.

Blee, E. and F. Schuber (1992). "Occurrence of fatty acid epoxide hydrolases in soybean (Glycine max). Purification and characterization of the soluble form." Biochem J **282 (Pt 3)**: 711-4.

Carroll, M. A. and J. C. McGiff (2000). "A new class of lipid mediators: cytochrome P450 arachidonate metabolites." Thorax **55 Suppl 2**: S13-6.

Carroll, M. A., M. Schwartzman, et al. (1987). "Vasoactivity of arachidonic acid epoxides." Eur J Pharmacol **138**: 281-3.

Edqvist, J. and I. Farbos (2003). "A germination-specific epoxide hydrolase from Euphorbia lagascae." Planta **216**(3): 403-12.

Gomez, G. A., C. Morisseau, et al. (2004). "Structure of human epoxide hydrolase reveals mechanistic inferences on bifunctional catalysis in epoxide and phosphate ester hydrolysis." Biochemistry **43**(16): 4716-23.

Guo, A., J. Durner, et al. (1998). "Characterization of a tobacco epoxide hydrolase gene induced during the resistance response to TMV." Plant J **15**(5): 647-56.

Howe, G. A. and A. L. Schilmiller (2002). "Oxylipin metabolism in response to stress." Curr Opin Plant Biol **5**(3): 230-6.

Hukkanen, J., O. Pelkonen, et al. (2002). "Expression and regulation of xenobiotic-metabolizing cytochrome P450 (CYP) enzymes in human lung." Crit Rev Toxicol **32**(5): 391-411.

Iakovou, K., M. Kazanis, et al. (1999). "Synthesis of oxypropanolamine derivatives of 3,4-dihydro-2H-1,4-benzoxazine, beta-adrenergic affinity, inotropic, chronotropic and coronary vasodilating activities." Eur J Med Chem **34**(11): 903-917.

Johansson, P., T. Unge, et al. (2005). "Structure of an atypical epoxide hydrolase from mycobacterium tuberculosis gives insights into its function." Journal of Molecular Biology **351**(5): 1048-1056.

Jones, T. A., J. Y. Zou, et al. (1991). "Improved methods for building protein models in electron density maps and the location of errors in these models." Acta Crystallogr A **47 (Pt 2)**: 110-9.

Kerr, B. M., A. E. Rettie, et al. (1989). "Inhibition of human liver microsomal epoxide hydrolase by valproate and valpromide: in vitro/in vivo correlation." Clin Pharmacol Ther **46**(1): 82-93.

Kolattukudy, P. E. (2001). "Polyesters in higher plants." Adv Biochem Eng Biotechnol **71**: 1-49.

Kozak, W., M. J. Kluger, et al. (2000). "Molecular mechanisms of fever and endogenous antipyresis." Ann N Y Acad Sci **917**: 121-34.

Mannervik, B. and U. H. Danielson (1988). "Glutathione transferases--structure and catalytic activity." CRC Crit Rev Biochem **23**(3): 283-337.

McCoy, A. J., R. W. Grosse-Kunstleve, et al. (2005). "Likelihood-enhanced fast translation functions." Acta Crystallogr D Biol Crystallogr **61**(Pt 4): 458-64.

Morisseau, C., J. K. Beetham, et al. (2000). "Cress and potato soluble epoxide hydrolases: Purification, biochemical characterization, and comparison to mammalian enzymes." Archives of Biochemistry and Biophysics **378**(2): 321-332.

Mowbray, S. L., L. T. Elfstrom, et al. (2006). "X-ray structure of potato epoxide hydrolase sheds light on substrate specificity in plant enzymes." Protein Sci **15**(7): 1628-37.

Murshudov, G. N., A. A. Vagin, et al. (1997). "Refinement of macromolecular structures by the maximum-likelihood method." Acta Crystallographica Section D-Biological Crystallography **53**: 240-255.

Nardini, M., I. S. Ridder, et al. (1999). "The x-ray structure of epoxide hydrolase from Agrobacterium radiobacter AD1. An enzyme to detoxify harmful epoxides." J Biol Chem **274**(21): 14579-86.

Navaza, J. (2001). "Implementation of molecular replacement in AMoRe." Acta Crystallogr D Biol Crystallogr **57**(Pt 10): 1367-72.

Node, K., Y. Huo, et al. (1999). "Anti-inflammatory properties of cytochrome P450 epoxygenase-derived eicosanoids." Science **285**(5431): 1276-9.

Ollis, D. L., E. Cheah, et al. (1992). "The alpha/beta hydrolase fold." Protein Eng **5**(3): 197-211.

Ota, K. and B. D. Hammock (1980). "Cytosolic and microsomal epoxide hydrolases: differential properties in mammalian liver." Science **207**(4438): 1479-81.

Pacifici, G. M., A. Temellini, et al. (1988). "Cytosolic epoxide hydrolase in humans: development and tissue distribution." Arch Toxicol **62**(4): 254-7.

Pinot, F., I. Benveniste, et al. (1999). "Production in vitro by the cytochrome P450 CYP94A1 of major C18 cutin monomers and potential messengers in plant-pathogen interactions: enantioselectivity studies." Biochem J **342 (Pt 1)**: 27-32.

Reetz, M. T., M. Bocola, et al. (2009). "Directed evolution of an enantioselective epoxide hydrolase: uncovering the source of enantioselectivity at each evolutionary stage." J Am Chem Soc **131**(21): 7334-43.

Reetz, M. T., C. Torre, et al. (2004). "Enhancing the enantioselectivity of an epoxide hydrolase by directed evolution." Org Lett **6**(2): 177-80.

Reetz, M. T., L. W. Wang, et al. (2006). "Directed evolution of enantioselective enzymes: iterative cycles of CASTing for probing protein-sequence space." Angew Chem Int Ed Engl **45**(8): 1236-41.

Seidegard, J. and G. Ekstrom (1997). "The role of human glutathione transferases and epoxide hydrolases in the metabolism of xenobiotics." Environ Health Perspect **105 Suppl 4**: 791-9.

Stapleton, A., J. K. Beetham, et al. (1994). "Cloning and expression of soluble epoxide hydrolase from potato." Plant J **6**(2): 251-8.

Szeliga, J. and A. Dipple (1998). "DNA adduct formation by polycyclic aromatic hydrocarbon dihydrodiol epoxides." Chem Res Toxicol **11**(1): 1-11.

Thomaeus, A., J. Carlsson, et al. (2007). "Active site of epoxide hydrolases revisited: a noncanonical residue in potato StEH1 promotes both formation and breakdown of the alkylenzyme intermediate." Biochemistry **46**(9): 2466-79.

Tzeng, H. F., L. T. Laughlin, et al. (1998). "Semifunctional site-specific mutants affecting the hydrolytic half-reaction of microsomal epoxide hydrolase." Biochemistry **37**(9): 2905-11.

van Loo, B., J. Kingma, et al. (2006). "Diversity and biocatalytic potential of epoxide hydrolases identified by genome analysis." Applied and Environmental Microbiology **72**(4): 2905-2917.

Widersten, M., A. Gurell, et al. (2009). "Structure-function relationships of epoxide hydrolases and their potential use in biocatalysis." Biochim Biophys Acta.

Zhao, L., B. Han, et al. (2004). "Epoxide hydrolase-catalyzed enantioselective synthesis of chiral 1,2-diols via desymmetrization of meso-epoxides. lzhao@diversa.com." J Am Chem Soc **126**(36): 11156-7.

Zou, J., B. M. Hallberg, et al. (2000). "Structure of Aspergillus niger epoxide hydrolase at 1.8 A resolution: implications for the structure and function of the mammalian microsomal class of epoxide hydrolases." Structure **8**(2): 111-22.

Acknowlegments

The last three chapters were about *Science*. This one will be about *Life*. My life, that wouldn't be the same without people around me. Something tells me this chapter will be the longest one...

First of all, I would like to thank my supervisor **Sherry** for excellent support and guidance. The last couple of years have been extremely educational and inspiring. Thank you for letting me work so independently, for your optimism and never ending patience.

My co-supervisor **Alwyn,** for everyday smile, your „*how many pages?*" and „*how is it going?*" over last couple of months...

I would also like to thank my collaborators: **Mikael Widersten, Ann Gurell** and **Diana Lindberg –** it was a pleasure to work with you, thanks for never ending supply of fresh protein, inspiring discussions and meetings. I am sorry I wasn't able to get the intermediate structure!

Ana Laura – you brought fresh air into our corridor, full of enthusiasm, ideas and always in good mood. Thank you for sharing with me the RpiB project, I have learnt a lot during last year! I hope to continue my Spanish course so one day I could understand your South American soul!

Emma – my first mentor, the most patient girl I've ever met. You guided me not only through the world of proteins and crystals, but what's even more important you took care of me outside the lab. *Tack so mycket Emma, jag lovar att en dag jag ska prata svenska!*

Urszula - thank you for being such a good friend, for your never ending support, even now when you're not in Uppsala anymore; for checking my Swedish homeworks; for arranging for us a place (or even two) to live; for all our talks and emails. I own you so much – my coach and real friend!

Glareh – I bless the day when you moved to Uppsala with your PhD! Thank you for your smile when I'm happy, tears when I cry, for your support, jokes, talks, advices, sharing rooms is Como, Tallberg, Denmark, even those late nights and early mornings in ESRF when we should've been separated from each other for our own sake. It is so good to have you here!

Wojtek – for the famous daily „breakthroughs"! For your way of working and setting priorities. I think I have learnt a lot just by watching you. Thanks for our talks; Polish sense of humor; your distance to everything, and for all emails asking about my thesis.

Michiel – for sharing tango fascination, all our talks, for being a good friend always when needed (e.g. killing the sheriff when his wife wanted to go home), for all the evenings with wine, movies, pictures...

Cha San – My Angel -thank you for taking care of everyone! I probably gained couple of kilos since you have moved to Uppsala, but it was the sweetest time of my whole PhD! Thanks for all cookies, chocolates, fruits, lunches... You, Glareh and probably my Mom are the only ones in my life who care when and what I actually eat...

Anatoly – for talks, ESRF trips and tips, your Eastern spirit and sense of humor. And of course for your cat, even if I didn't manage to convince Artur to have one...; **Daniel** – for always sneaking to my office with cup of coffee or piece of cake, for small chats&longer talks. It was good to share with you „the_thesis_writing_stress". I promise one day we will go climbing outdoors!; **Lotta** – for sharing with me the Kingdom of the Beer Club, and being the best party organizer ever; **PavelP** – for being good neighbor, having everything I needed to borrow (climbing gear, bottle of Absinth, projector... you name it). Thanks for organizing all Flogsta activities and for sharing with me the true love to „Palermo"; **Annette+Nisse** for being such a friendly couple; **Magnus** – for all your effort you put to understand Eastern European people, for good advices about apartments and cats and for your fantastic sense of humor!

Sanjee and Wimal for being so kind and nice, always having a good word and smile; **Christofer** – for Ice-cream Club every summer; for good travel tips about Istambul and Iceland; **AnnaJ** – for daily portion of sarcastic humor; **Adrian** - for everyday smile, **Frederik** – for scientific enthusiasm and for starting our collaboration with Diversa; **Jonas** – for great talks and for jumping in to my office in the last couple weeks, asking „*how is it going*" and repeating that „*the thesis shouldn't be too long*"; **Gerard** for accepting me as an exchange student in your group long time ago. Your very kind emails helped me to choose cold, dark Sweden, instead of sunny Italy or even better... Finland. Thanks for all nice group dinners, pizzas, volleyball during those days. **Lars** for xray-course; **Torsten** for „fika reminders"; **Tex** for everyday smile and good sense of humor, book recommendations, great dinner, talks about food and wine;

Karin, Evalena, Nina and Tex – you Ladies are like a huge Lab-Wikipedia, you simply know everything what one could ever ask! Thanks for all answers, tips and advices! I would be lost in the lab without you!

Erling, Christer and **Elleonor** – for always helping with practical matters.

The Monday Pottery Club (**Karin, Nina, Urszula, Annette, Lotta, Glareh, Monika, Yang, Lena)** – for all the girl-talks and pottery inspiration.

The youngest team in our corridors: **Yang, Cha San, Avinash, Dee, Majid**, **Miao** for reminding me about coffee breaks, keeping my sugar level as high as possible, all discussions during lunches and fikas. Many thanks for the volleyball games - I loved them!

All former PhD students at the time when I came to Sweden for the first time: **Jenny, Al, Talal, Patrik, Marian, Faribourz, Andrea, Rosie, Malin, Sara, Martin and Louise** you all made my first few months in Sweden so nice that the decision of coming back was very easy.

All@xray including **Alina, AnnaL., Anton, David, Filipe, Frida, Hasse, Henke, HenrikB., HenrikI., Gunnar, Gösta, , Inger, Janos, Jerry, Kaspars, Lena, Margareta, Maria, Mark, Mats, Marvin, Mikael, Nic, Nils, Nina, Saeid, Stefan, Torleif, Ulla, Yafei, Åsa...** - thank you all for being here and making xray corridors such a nice place to work!

But outside the labs there was also another Life…

Ania & Grzes Mysliwy (with little Gustaw and even smaller Maja) – for a great friendship from the very first day! For being fantastic family! I am so grateful for all your support during these years! All dinners, evenings with „Taniec z gwiazdami", parties, lunches, climbing trips, barbeques in Tursborg and many others. Hope you will never ever think of moving back to Poland! We need you here!

Grzesiek & Agata with little Sonia – for parties, beers, talks, movies, climbing...

Gabrysia & Bartek – for all the games until early mornings, for sailing, campings, skiing, for „ferry_trips_for_Christmas", for being exactly like us but the other way around...

„Sunday Climbing School" (**Andzia, Asia, Michal i Jackie, Lukasz, Ania+Grzes, Agata+Grzes, Piotrki, Bartek, Darek, Karolina, Ewa+Krzysiek +Franek , Agnieszka, Kuba**...) – for great friendships, support, fantastic weekends and sports – not only climbing but skiing, ice-skating, kayaking, gokarting...!

The main team of www.poloniainfo.se: **Kajetan, Magda, Ewa, Michal, Ulka, Asia, Iza, Pawel, Kasia, Dana, Bogus, Magda, Monika...**– for giving me an opportunity to do something that is so much different than science. I had a chance to meet interesting people and be involved in fantastic events. You all together make very supportive and friendly team. It is so much fun to work with you guys! And who knows maybe one day when I grow up, I will be a journalist...

not_only_Flogsta_gang **(Pavel P., Pavel K., Michiel, Anne, Juliane, Rafal, Wojtek, Monika, Leo, Maria, Michael, Martin, Maria, Zuzka**...) - for playing cards (and letting me be „a decent farmer's wife"), barbecues, wine& movie-sessions, Absinth parties, cube...

My friends far far away, oceans apart... **(Magda S., Magda S., Ksenia**) – you're SIMPLY THE BEST! I am so glad our friendship survives even now when we are on different continents. I am so grateful for all your support in emails, and facebook chats during all these years. I love you girls!

Moim kochanym Mamie i Tacie – bez Was nie dałabym sobie rady. Wspieraliście mnie od samego początku, od pierwszej chwili gdy narodził się pomysł wyjechania do Szwecji. Byliście ze mną myślami codziennie przez ostatnie lata i to sie dało odczuć! Dziękuję za Wasze wsparcie, wyrozumiałość i słodkie smsy każdego dnia. Kocham Was bardzo!

Krynia & Jola, Wojtek & Agnieszka - the best brothers and sisters one could ever have! Thanks for your support, encouragement and accepting Artur as a member of our family :)

Haneczka i Misio – for being the best kids in the world! It was always so much fun to have you here in Sweden! I am so proud of who you are! Thanks for understanding that the only fairy tales I know are about viruses, DNA and this kind of stuff. I hope it won't affect your future choices and one day you understand that working with bacteria is not as cool as it looks like…

Artur – I simply don't know what to say... You must be an Angel because only angels have so much of patience and understanding. Honey, I hope our life will get back to normal now. CIEM :*

Agata