Dipl.-Ing. Tanja Knaus

BIOCHEMICAL CHARACTERIZATION AND MECHANISTIC STUDIES ON TWO ZINC DEPENDENT PROTEINS

DISSERTATION

Zur Erlangung des akademischen Grades

einer Doktorin der technischen Wissenschaften (Dr. techn.)

erreicht an der

Technischen Universität Graz

Betreuer: Univ.-Prof. Mag. *rer.nat*. Dr. *rer. nat*. Peter Macheroux

Institut für Biochemie

Technische Universität Graz

Graz, 2012

"FLAVOPROTEINS ARE GREAT"

- P. Macheroux (2012)

Acknowledgements

First I would like to express my gratitude to my supervisor Peter Macheroux for giving me the opportunity to work on those interesting projects and for his constant support during the last four years. I would also like to thank him for giving me an insight regarding the interesting world of flavoproteins and sparking my interest in these yellow-coloured enzymes.

Furthermore I would like to thank Markus Schober for our great teamwork on the Pisa1 project and Michael Uhl who helped me whenever I had troubles with PyMol!

Moreover I would like to thank Susanne Horvath, Ute Stemmer and Martin Puhl for our nice conversations, motivating discussions and encouragements during our coffee breaks.

My thesis committee members Prof. Kurt Faber and Prof. Ulrike Wagner deserve special thanks for dedicating time and knowledge to advice and support me during the time of my PhD project.

Thanks to the highly motivated students: Elisabeth Eger, Martin Faccinelli, Julia Pitzer, Stefanie Mondschein and Julia Koop. It was a pleasure for me to supervise them during their bachelor and master projects.

Thank you, Francesco, for supporting me the whole last year. You picked me up when I was down and motivated me always to hang on. Thank you for correcting my sometimes very woozily English sentences and thank you very much for making my life as perfect as it is! I want to express my gratitude to my mum and my dad for their love, their great support and motivation during my whole life and for always believing in me! The strong relationship to all of you, affects my way of life in a very positive way!

Last but not least I would like to thank all people who have helped and inspired me during the time of my PhD thesis and who contributed to a pleasant atmosphere, enjoyable working hours in the lab and motivating discussions.

Abstract

Pisa1 from Pseudomonas sp. DSM6611: A highly enantio- and stereoselective secondary alkylsulfatase from *Pseudomonas* sp. DSM6611 (Pisa1) was heterologously expressed in *Escherichia coli* BL21 and purified to homogeneity for kinetic and structural studies. Structure determination of Pisa1 by X-ray crystallography showed that the protein belongs to the family of metallo-β-lactamases with a conserved binuclear Zn^{2+} cluster in the active site. In contrast to a closely related alkylsulfatase from *Pseudomonas aeruginosa* (SdsA1), Pisa1 showed preference for secondary rather than primary alkyl sulfates and enantioselectively hydrolyzed the *(R)* enantiomer of *rac*-2-octyl sulfate yielding (*S*)-2-octanol with inversion of absolute configuration as a result of C-O bond cleavage. In order to elucidate the mechanism of inverting sulfate ester hydrolysis, for which no counterpart in chemical catalysis exists, an invariant histidine residue (His317) near the sulfate binding site was identified as the general acid for crucial protonation of the sulfate leaving group. Additionally, amino acid replacements in the alkyl-chain binding pocket generated an enzyme variant, which lost its stereoselectivity towards *rac*-2-octyl sulfate. These findings are discussed in light of the potential use of this enzyme family for applications in biocatalysis.

ppBat from *Bacteroides thetaiotaomicron*: In a recent survey of flavin-dependent proteins, a putative protease from *Bacteroides thetaiotaomicron*, a microbe inhabiting the human gut, was identified. The structure of the protein (pdb code 3CNE) features a mono nuclear zinc site in addition to a flavin, which is sandwiched by two tryptophan residues provided by each of the two protomers in the dimeric protein. The zinc ion is coordinated with two cysteine-derived thiol groups and two water molecules forming a tetragonal ligation sphere. The large distance of ≈ 16 Å between the edge of the flavin cofactor and the zinc argues against a direct cooperation between these two cofactors. This intriguing combination of unusual cofactors sparked our interest in the protein. In this thesis, the biochemical characterization of the flavin as well as the zinc binding site of the protein is reported. Moreover it is demonstrated, that the recombinant protein is capable of binding not only naturally occurring flavin derivatives, such as lumichrome, riboflavin, FMN and FAD but also a variety of chemically modified "artificial" flavin analogs with K_D values in the nanomolar range. Additionally, it could be proved that the Trp164 is fundamental for binding the flavin cofactors but not for a dimerization of the protein and that the zinc ion has a structural role in ppBat.

Kurzfassung

Pisa1 aus Pseudomonas sp. DSM6611: Eine enantio- und stereoselektive Alkylsulfatase aus *Pseudomonas* sp. DSM6611 (Pisa1) wurde heterolog in *Escherichia coli* BL21 exprimiert und aufgereinigt, um kinetische und strukturelle Studien an diesem Enzym durchzuführen. Die Kristallstruktur von Pisa1 zeigte, dass dieses Protein zur Familie der Metallo-ß-Lactamasen gehört und einen konservierten binuklearen Zn^{2+}-Cluster im aktiven Zentrum besitzt. Im Gegensatz zu einer ähnlichen Alkylsulfatase aus *Pseudomonas aeruginosa* (SdsA1) bevorzugt Pisa1 sekundäre Alkylsulfate gegenüber primären als Substrate für die Hydrolyse. Das Enzym setzt enantioselektiv das (*R*)-Enantiomer von *rac*-2-Octylsulfat zu (*S*)-2-Octanol um. Dies geschieht unter Inversion der absoluten Konfiguration am chiralen Kohlenstoffatom als Resultat einer C-O Spaltung. Inverse Sulfatesterhydrolysen sind unter chemischen Bedingungen nicht möglich; im Falle von Pisa1 konnte gezeigt werden, dass ein Histidin Rest (His317) in der Nähe der Sulfatbindestelle als Säure dient, die für die Protonierung der Sulfat Abgangsgruppe benötigt wird. Des Weiteren führten Aminosäureaustausche in der Bindestelle der Alkylkette zur Herstellung einer Enzymvariante, welche die Stereoselektivität hinsichtlich *rac*-2-Octylsulfat verlor.

ppBat aus *Bacteroides thetaiotaomicron*: In einer Studie über flavinabhängige Enzyme wurde eine mögliche Protease aus dem menschlichen Darmorganismus *Bacteroides thetaiotaomicron* identifiziert. Die Kristallstruktur des Proteins (PDB Code: 3CNE) zeigt, dass es eine mononukleare Zinkbindestelle und zusätzlich ein Flavin aufweist. Dieses Flavin ist zwischen zwei Tryptophanresten aus den beiden Protomeren eingebettet. Das Zinkion ist durch die Thiolgruppen von zwei Cysteinresten und zwei Wassermolekülen tetragonal gebunden. Eine direkte Kooperation der zwei Kofaktoren ist durch den großen Abstand von etwa 16 Å zwischen dem Flavin und dem Zink unwahrscheinlich. Durch diese unübliche Kombination dieser zwei Kofaktoren wurde das Interesse an diesem Protein geweckt und es wurden sowohl die Flavinbindestelle, als auch die Zinkbindestelle biochemisch charakterisiert. Es konnte gezeigt werden, dass ppBat nicht nur natürlich vorkommende Flavine wie Lumichrom, Riboflavin, FMN und FAD, sondern auch chemisch modifizierte Flavinderivate mit K_D-Werten im nanomolaren Bereich bindet. Des Weiteren wurde nachgewiesen, dass die Aminosäure Trp164 fundamental für die Bindung des Flavinkofaktors, jedoch nicht für die Dimerisierung des Proteins zuständig ist und das Zinkion eine strukturstabilisierende Rolle in ppBat aufweist.

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommene Stellen als solche kenntlich gemacht habe.

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Graz, am …………………………   ……………………………………………………

Table of contents

Chapter 1

Zinc dependent proteins

1.1   Introduction to zinc in biology

Immediately after iron, the second transition metal for abundance and biological importance into mammals is zinc.[1] The importance of zinc for the growing of organisms and particularly its high concentration in the reproductive organs, infers about the participation of this metal in the catalysis of essential metabolic reactions. Nature has evolved various structural patterns whereby zinc is bound to protein scaffold. Zinc possesses a wide variety of different functions: it stabilizes the structure and the conformation of proteins, acts as a catalytic center or participates in the expression of the genetic code. Nowadays, more than 300 enzymes have been characterized to contain zinc. Interestingly it is the only metal known to have representatives in all enzyme classes.[2, 3]

According to the acid-base hard-soft theory[4], zinc is classified as a borderline metal thus neither hard nor soft. The hard-soft theory is based on the definition of acid and base according G. N. Lewis[5], which takes into account the electronic configuration of the species involved. A Lewis acid is a chemical species which possesses an unoccupied atomic or molecular orbital at low energy, whereas a Lewis base is a species which possesses a couple of unpaired electrons available to form a chemical bound. The feasibility of an acid-base reaction depends on the strength of the couple involved but also from another property: the hardness of the acid and the softness of the base. This property can only be described qualitatively:[6-8]

- Hard bases: the donor atoms have high electronegativity and low polarizability; hence they are difficult to oxidize since the electrons of the valence shell are tightly bound.

- Soft bases: the donor atoms have low electronegativity and high polarizability; hence they are easy to oxidize since the electrons of the valence shell are weakly bound.

- Hard acids: The acceptor atoms possess a small radio, a high positive charge and do not contain unpaired electrons in the valence shell; hence they have low polarizability and high electronegativity.

- Soft acids: The acceptor atoms possess a high radio, a small positive charge and contain unpaired electrons belonging to p or d orbitals of the valence shell; hence they have high polarizability and low electronegativity.

In a nutshell, the general rule for the hard-soft acid-base theory foresees that hard acids bind preferentially hard bases and soft acids bind preferentially soft bases.

The chemistry of zinc is dominated by the +2 oxidation state. As mentioned before, it is typically a borderline acid thus it binds preferentially borderline bases (i.e. nitrogen donors such as a histidine residue). However, this borderline feature allows zinc to bind also hard bases (i.e. oxygen donors such as a water molecule) and soft bases (i.e. sulphur donors such as a cysteine residue).[9, 10] In general zinc-complexes, hence zinc-proteins are particularly stable due to the favourable entropic factors coming from the elimination of a water molecule in the coordination sphere. Zinc has all the five 3d orbitals completely occupied, which leads to a coordination geometry not influenced by electronic factors. In fact, d^0 and d^{10} atoms are generally not influenced by stabilization of d orbitals. When a metal-protein interaction occurs, the d orbitals of the metal ion lying along the donor become destabilized, whereas the d orbitals pointing towards the contrary direction become stabilized. Thus, the latter orbitals of the metal ion are occupied first from the electrons coming from the donor (i.e. amino acid residues from the protein). In the case of Zn^{+2}, this stabilization does not exist since the d orbitals are already completely occupied. As a consequence, the only parameter which determines the geometry of the ligands around the Zn^{+2} center is the sterical hindrance of the ligands. These ligands are arranged around the metal center in order to minimize the repulsion among themselves. Therefore, the zinc center present in zinc proteins commonly adopts a slightly distorted tetrahedral geometry, coordinating three or four protein side chains. In fact, the repulsion among the ligands is minimized in a tetrahedral coordination. Additionally, the common rule of the 18 electrons for transition metals is simultaneously satisfied.[11, 12] However, also five-coordinated distorted trigonal bipyramidal geometry has been observed in the metal sites of various Zn dependent proteins (Scheme 1).[12]

tetrahedral   trigonal bipyramidal
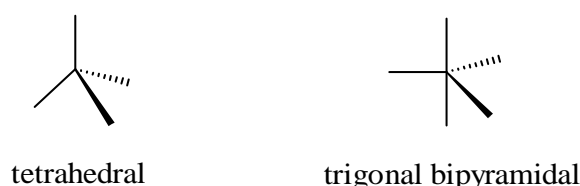
Scheme 1. Coordination geometry of Zn^{2+} in zinc dependent proteins.

As mentioned before, zinc is only present in the +2 oxidation state. Thus it cannot promote catalytic reactions whereby electrons are transferred via the metal center (redox reactions). Consequently, Zn^{+2} acts only as a strong Lewis acid hence activating the coordinated water molecule towards deprotonation to generate a hydroxide species (Scheme 2a) or polarizing the chemical bond of its ligands (i.e. Zn-OH + H^+, Scheme 2b) to generate a nucleophile for catalysis. Furthermore, zinc is appointed to preserve the structure in the active site of the enzymes and orient the substrates in the correct position (Scheme 2c).[13]
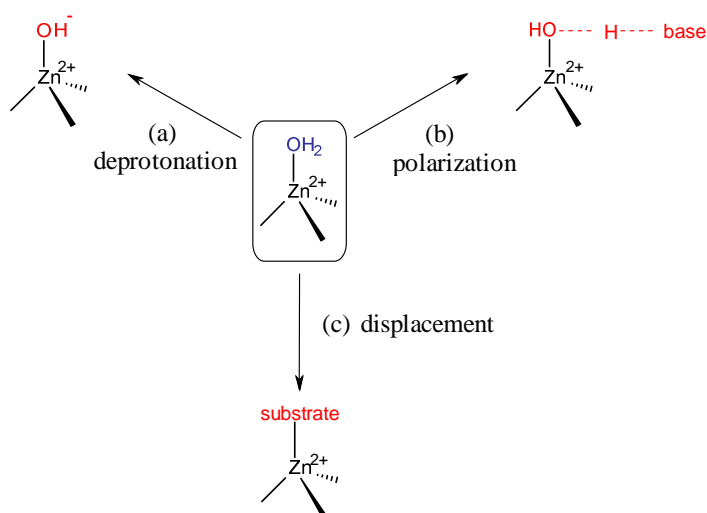
Scheme 2. Basic mechanisms of the activation of zinc bound water in the majority of zinc enzymes: (a) deprotonation of the ligand such as water to generate a nucleophilic zinc hydroxide ligand (b) additional polarization by a basic residue, and (c) displacement of the coordinated water molecule by the substrate.

1.2   Classification of zinc binding sites

So far, four distinct types of zinc binding sites in enzymes have been identified:

Catalytic zinc: Zn^{+2} is essential and directly involved in the enzymatic catalysis. The removal of the metal leads to loss of any catalytic activity although the enzyme usually maintains its tertiary structure. Those zinc sites consist of a bound water molecule and additional three amino acid residues, namely His, Glu, Asp or Cys in a binding frequency of His >> Glu > Asp = Cys. It is not surprisingly that histidine is the most frequent amino acid donor into zinc catalytic active sites, since the imidazole ring coordinates using a typically borderline nitrogen donor.[14]

Structural zinc: The metal is necessary to maintain the structural stability of the protein (often the quaternary structure in oligo-enzymes) and structural zinc sites possess four protein ligands without any additional water molecule. Structural zinc is present in the alcohol dehydrogenase (ADH)[15] or in the regulatory subunit of *Escherichia coli* aspartate carbamoyltransferase[16] whereby the metal is coordinated to four cysteine residues in a tetrahedral fashion. Such a coordination pattern hampers the access of water and other substrates to the ligands sphere. Thus, it seems that zinc-sulphur bounds are appointed to preserve the structure of the proximal catalytic site and that of the whole protein. In fact, the thermodynamic stabilization of a zinc-sulphur bound is comparable to the one of a disulphide bridge. Although sulphur is the preferred ligand in structural zinc sites, also combinations with His or Glu and Asp exist. Figure 1 shows the combination of structural zinc sites which have been observed in metalloenzymes so far.[17]
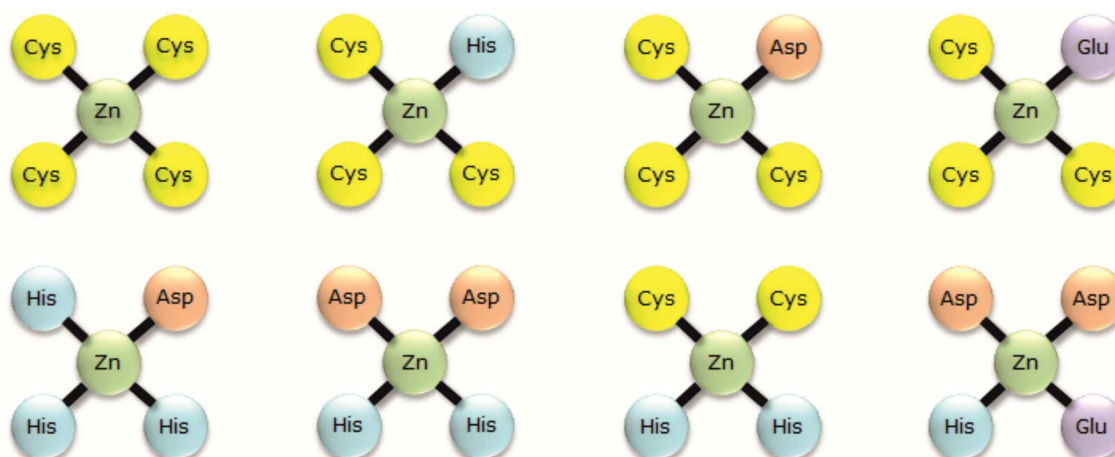
Figure 1. Schematic representation of the types of structural zinc sites observed in metalloenzymes until to date.

Cocatalytic zinc: Cocatalytic metal sites are important to stabilize the overall protein folding and to assist the catalytic activity of the enzyme. Cocatalytic zinc is present in enzymes containing two or more zinc atoms and/or other transition metals close each other (i.e. cluster). For instance a single amino acid residue can bind simultaneously two metal atoms and sometimes an additional bridge is formed by a water molecule. Asp and His are the most prevalent ligand in this type of zinc sites.[18,19] For instance, in the P1 nuclease from *Penicillium citrinum* (Pdb code: 1AK0), one catalytic zinc and other two additional zinc ions are found in the active site (Figure 2A).[20] Conversely, in *E. coli* alkaline phosphatase (Pdb code: 1ALK), two zinc ions and one magnesium ion are present (Figure

2B).[21] However, in all the cases, one zinc is the catalytic site and the remaining couple of metals, i.e. zinc or magnesium, is linked via a bridge amino acid residue (usually an aspartic acid). These additional metal ions are classified as cocatalytics.
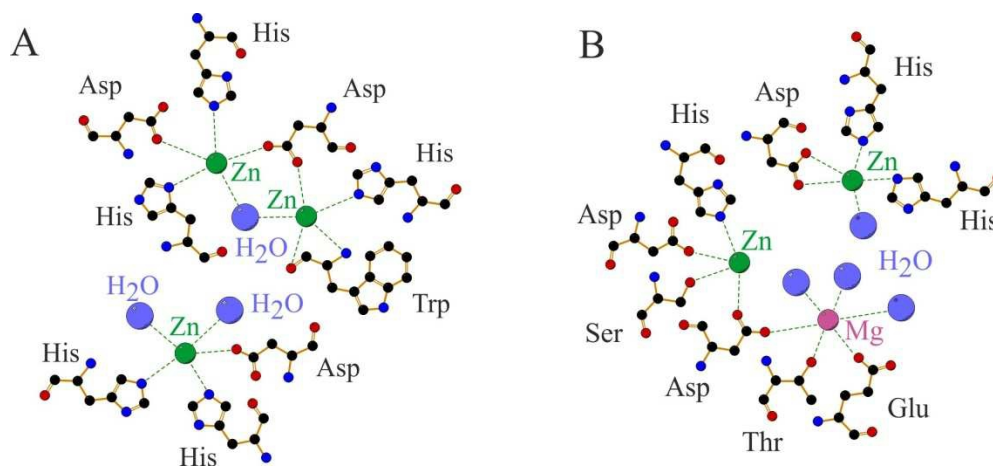
Figure 2. Schematic representation of three-metal cocatalytic sites: (A) P1 nuclease from *Penicillium citrinum* and (B) *E. coli* alkaline phosphatase, created with Ligplot.[22]

Protein interface zinc sites: Zinc is appointed to stabilize the quaternary structure of proteins. The ligands are located in the binding surface between the two interacting proteins or protein subunits and in principal protein interface zinc sites show general characteristics of catalytic or structural zinc binding sites. In principle four amino acid residues are bound to zinc (Glu, Asp, His or Cys) to achieve again a tetrahedral coordination geometry, albeit it can be extended to five and six ligands with donor atoms other than sulfur.[23, 24]

As mentioned before, His and Cys constitute the preferential ligands for catalytic and structural zinc, whereas Asp is predominant in cocatalytic zinc. It is noteworthy that cocatalytic zinc is also bound to other amino acid residues such as Ser, Thr, Lys or Trp. These ligands have lower affinity to the zinc center compared to His (N-imidazol), Glu, Asp (O-carboxylic) and Cys (SH-). Thus, the cocatalytic zinc is more prone to rearrangement when a substrate enters into the catalytic site. This rearrangement leads to the cleavage of the bond between the Asp and the other proximal catalytic zinc, which enhances its polarizability. The overall process ends up with the activation of the catalytic zinc which is then capable to act as a Lewis acid and interacts with the substrate. As an alternative, the formerly bridged Asp can act as a nucleophile or acid-base catalyst.[2]

## 1.3 Enzymatic reactions catalyzed by zinc enzymes

Although zinc proteins show representatives in all 6 enzyme classes, the main importance of zinc enzymes derives from their hydrolytic activity. The most common catalyzed reactions are: phosphoric ester hydrolysis and peptide bond hydrolysis. The latter also includes endopeptidase as well as exopeptidase (carboxypeptidases and aminopeptidases) activity. Additionally to the cleavage of amide and phosphate bonds two other important zinc enzymes should be mentioned: carbonic anhydrase, which hydrolyses CO_2 and alcohol dehydrogenase which catalysis oxidation reactions of alcohols to ketones or aldehydes.[13]

1.4 Prominent representatives of zinc dependent proteins

In the next chapters the zinc sites of the most prominent representatives of zinc dependent proteins and the mechanisms of selected ones are described.

1.4.1 Hydrolases acting on peptide bonds (Zinc peptidases)

Zinc peptidases catalyze the hydrolysis of peptide bonds. They are divided in two groups: endopeptidases (cleavage of peptide bonds within the protein chain) and exopeptidases (cleavage of a terminal amino acid from the protein chain). Exopeptidases can further be classified into carboxypeptidases (removal of the C-terminal amino acid) and aminopeptidases (removal of the N-terminal amino acid). Peptidase activity is illustrated in Figure 3.[25]
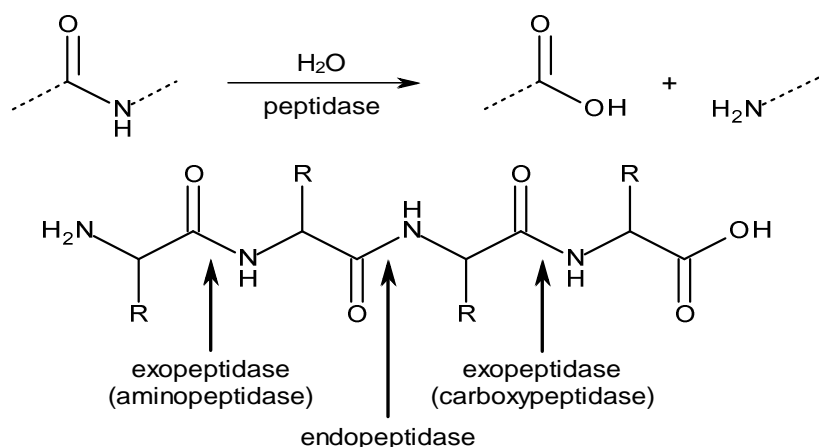
Figure 3. Peptidase activity of zinc dependent enzymes.

The zinc peptidases can be grouped in 5 superfamilies. In the first superfamily, also named superfamily of the thermolysin, the zinc center is coordinated by a Glu and two His residues. The thermolysin, the carboxypeptidase A, the D-Ala carboxypeptidase and many others belong to this superfamily. All these peptidases possess a similar active site, including the coordination to the zinc center and the activation of a water molecule. In contrast to many phosphatases, which preferentially use magnesium or manganese as active metal center, peptidases employ solely zinc to hydrolyze peptide bounds. The second superfamily, named as metzincin, is characterized by a zinc ion bound to three histidine residues. At this superfamily belong the peptidases from snake poison. In the third superfamily, both Glu and His bind the zinc center. Other amino acids are also involved in the coordination sphere although they have not been identified so far.

It is interesting to note that both thermolysins and metzincins conserve the same topology regarding to the catalytic domain; that is constituted by five β-strands and two α-helices. Thus, the two superfamilies have probably evolved from a common ancestor.
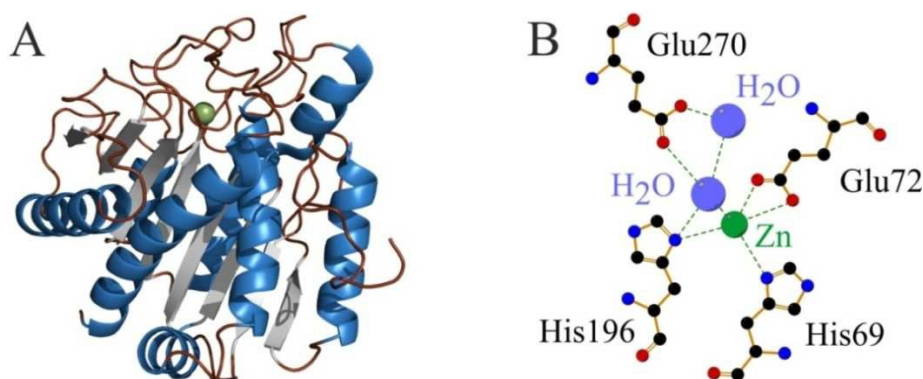
Metallocarboxypeptidases (Carboxypeptidase A – CPA)

Members of the metallocarboxypeptidase (CP) family catalyze the removal of C-terminal amino acids from proteins. According to their specificity, four classes may be distinguished: (i) A-like CPs for aromatic/hydrophobic residues, (ii) B-like CPs for basic residues, (iii) for acidic residues and (iv) inactive for standard CP substrates.[26]

Carboxypeptidase A (PDB code: 5CPA[27]), the prototype enzyme for metallocarboxypeptidases, is a digestive enzyme which has been isolated from bovine pancreas. It catalyzes the cleavage of C-terminal hydrophobic amino acid residues, preferentially possessing aromatic side chains such as Phe, as well as the hydrolysis of esters. CPA consists of ca. 300 amino acid residues with a MW of 34 kDa (Figure 4). The catalytic zinc center is located at the bottom of a cavity and it is coordinated to His69, His196, Glu72 and a water molecule. The water is additionally bound to Glu270 via a hydrogen bound.[27]

Structural investigations have led to two proposals concerning the mechanism of the hydrolysis. In the first proposal, an Arg145 places the peptide or the ester substrate in the

correct orientation via a strong interaction between the imine group of the amino acid and the carboxylic group of the substrate.

Figure 4. Structure representation of carboxypeptidase A (PDB code: 5CPA) from bovine pancreas. (A) Cartoon representation, showing the zinc ion in green. Picture was prepared with the program PyMol.[28] (B) Close up of the zinc binding site, prepared with LigPlot.[22]

This feature explains the specificity of the enzyme to cleave exclusively C-terminal amino acid residues. Then, the zinc coordinates the carbonylic oxygen of the peptide which polarizes the C-O bond. This favours the attack of the nucleophile Glu270 to form an anhydride Glu-C(O)-O-C(O)-R. The function of the zinc is fundamental since it polarizes the carbonylic bound and increases the acidity of the water molecule hence producing the more nucleophilic hydroxyl species. The final rearrangement of the penta-coordinated zinc gives the carboxylic acid and regenerates the native enzyme. The fact that some anhydrides have been found during the hydrolysis of a few substrates supports this proposal.[29]

The second hypothesis by Christiansen and Libscomb[30] (Figure 5) suggests that Arg145 serves to place the substrate in a specific orientation via hydrogen bonds between the amino groups of Arg145 and the carboxylic group of the peptide. Then, the synergic action of Glu270 and the Zn^{+2} allows the deprotonation of the coordinated water molecule hence enabling the direct attack of the hydroxyl moiety to the electrophilic carbon of the peptide. An additional Arg127 increases the nucleophilicity of the carbonylic oxygen of the peptide, allowing a direct attack of the zinc ion. The penta-coordinated intermediate evolves to the cleavage of the C-N bound and final substrate hydrolysis.[29]
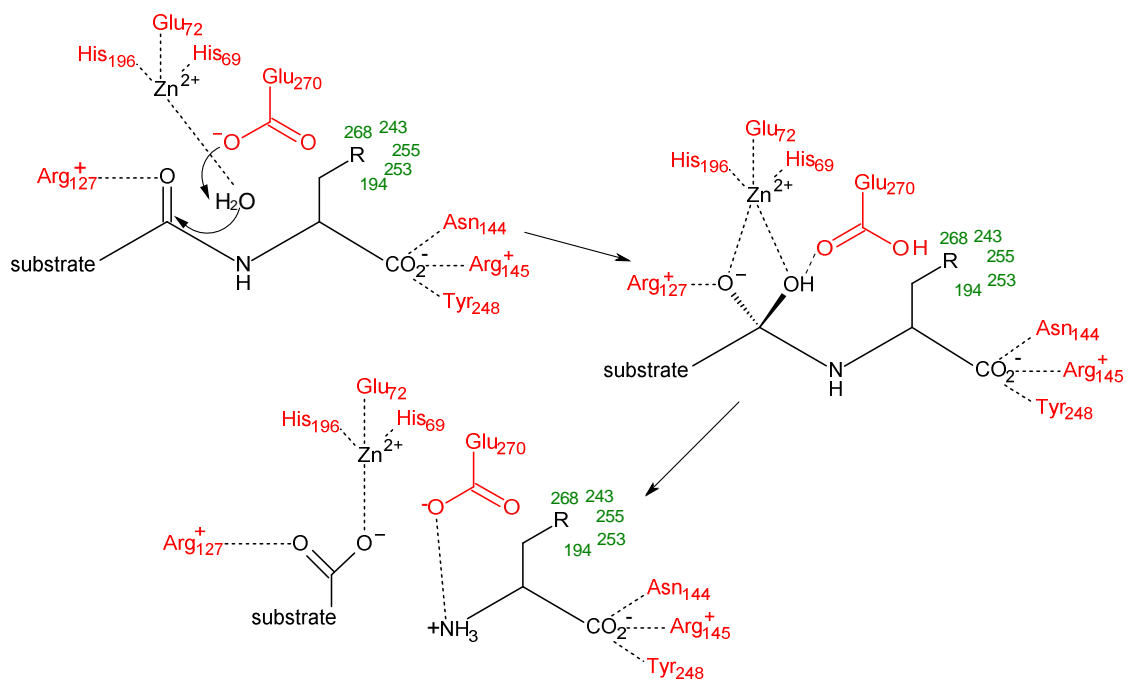
Figure 5. Catalytic mechanism of CPA as proposed by Christiansen and Libscomb (adapted from reference 29) with substrate atoms shown in black, and active sites residues and residues in the binding pocket colored in red and green, respectively.

Leucine aminopeptidase (LAP)

Leucine aminopeptidases are found in animals, plants and bacterica and play a key role during the modification and the degradation of proteins and in the metabolism of biologically active peptides by hydrolyzing the N-terminal amino acid of a polypeptide chain. The altered activity of the aminopeptidase was linked to diverse pathologies, for instance, cancer or eye cataract.[31, 32]

Until now more than 50 sequences for leucine aminopeptidase activity are found in the Swiss-Prot Database, whereby the two best studied enzymes were isolated from *Bos taurus* (LAP, PDB code: 1LAM[33]) and *Escherichia coli* (PepA, PDB code: 1GYT[34]).

LAP is a non-cooperative homohexamer, constituted by 487 amino acid residues per monomer; it contains two zinc centers which are spaced by 3 Å within the active site (Figure 6). Each zinc ion is penta-coordinated with an octahedral geometry lacking of one vertex. The two zinc ions are bridged by Glu334 (bidentate), Asp255 (monodentate) and a water molecule. Zn1 is additionally bound to Asp332, whereas Zn2 is bound to Asp273

and Lys250. Interestingly, the zinc ion, coordinated by the lysine residue also binds the terminal amino group of the substrate peptide.[35]
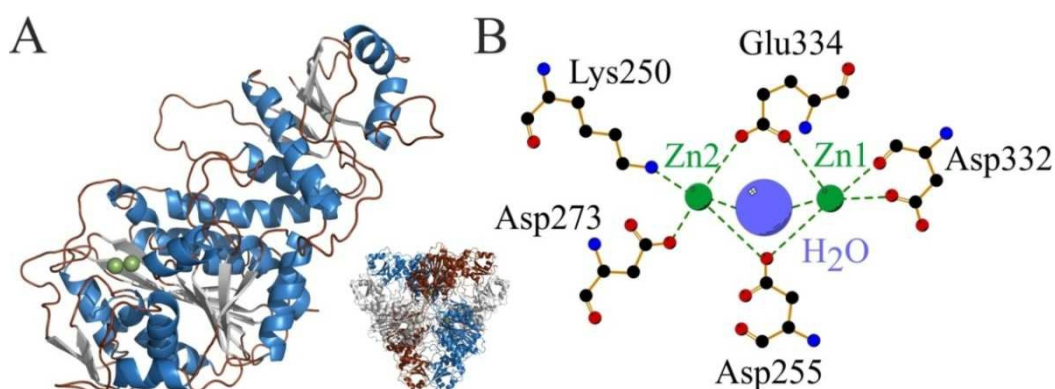
Figure 6. View of the structure of *Bos taurus* leucine aminopeptidase (PDB code: 1LAM). (A) Cartoon representation of the protomer, showing the zinc ions in green. The inset shows the naturally occurring hexameric structure. Pictures were prepared with the program PyMol.[28] (B) Close up of the zinc binding site, prepared with LigPlot.[22]

On the base of this and other structures, a catalytic mechanism has been proposed, whereby the substrate binds via its N-terminal amino group towards Zn2 whereas Zn1 polarizes the carbonylic oxygen. Subsequently, the nucleophilic attack of the activated water molecule (or the hydroxyl moiety) occurs. Additionally to the zinc ions, Lys262 and a bicarbonate ion bound to Arg336 and Leu360 stabilize the transition state. The bicarbonate ion shuttles the proton from the nucleophile to the leaving amino group.[35, 36]
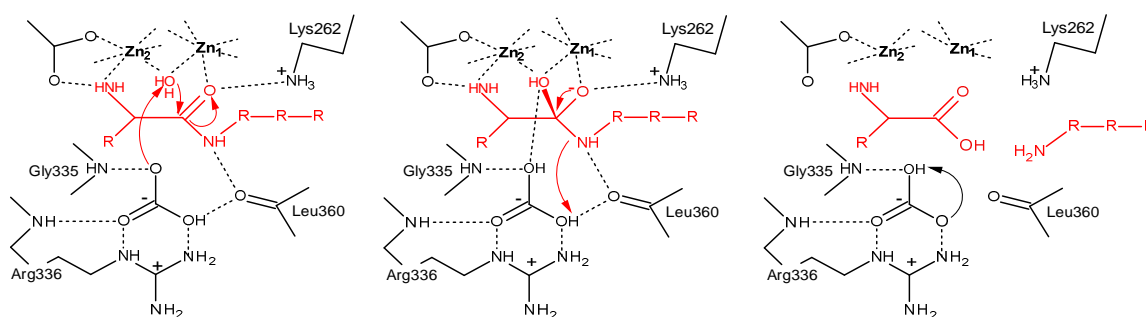
Figure 7. Proposed catalytic mechanism for the cleavage of the N-terminal amino acid by *B. taurus* leucine aminopeptidase (adapted from reference 36).

Methionine aminopeptidase

The methionine aminopeptidase obtained from *E. coli* is the prototypical member of this family and cleaves the N-terminal methionine from nascent polypeptide chains.[37] The structure has been established for the enzyme containing two cobalt ions as a replacement for the native zinc ions (PDB code 1MAT[38]). The two cobalt ions in the cocatalytic metal site are spaced by 3.2 Å and they are bound to the side chains of Asp97, Asp108, Glu204, Glu235 and His171 (Figure 8). The coordination of the metal ions is quasi octahedral. Furthermore a water molecule forms a bridge between Co1 and Co2 and a second water molecule binds to Co2. These two water molecules could form hydrogen bonds with His178 and Thr99, respectively. All these amino acid residues binding the metal centers are highly conserved in all the enzymes belonging to the homologues series, the only exception being the methionine aminopeptidase from *Bacillus subtilis*. In this homologue, the Glu235 is substituted with a Gln.[39]

Figure 8. Structure representation of *E. coli* methionine aminopeptidase (PDB code: 1MAT). (A) Cartoon representation, showing the metal ions in green. Picture was prepared with the program PyMol.[28] (B) Close up of the zinc binding site, prepared with LigPlot.[22]
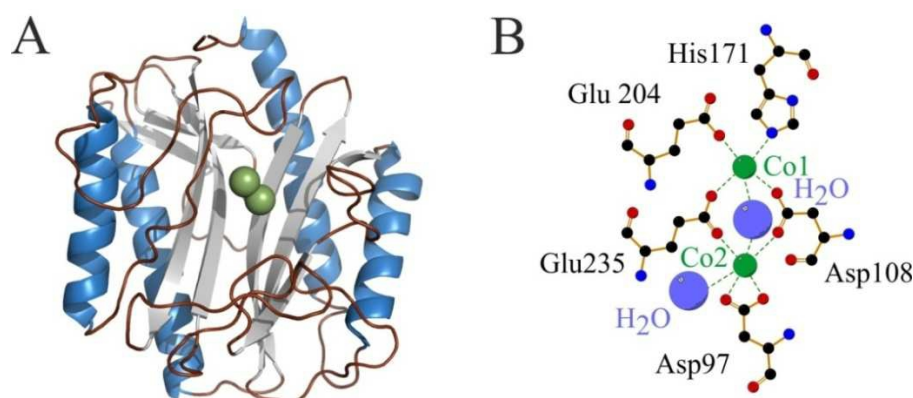
The main feature of the proposed mechanism for *E. coli* methionine aminopeptidase (displayed in Figure 9) is the nucleophilic attack of a metal-bound water or a hydroxide ion to the carbonyl carbon of the substrate which is assisted by Glu204 together with the metal center.

A noncovalently bound tetrahedral intermediate is formed and a proton transfer from Glu204 to the nitrogen atom of the scissile bond may facilitate breakdown of the indermediate to products.[40]
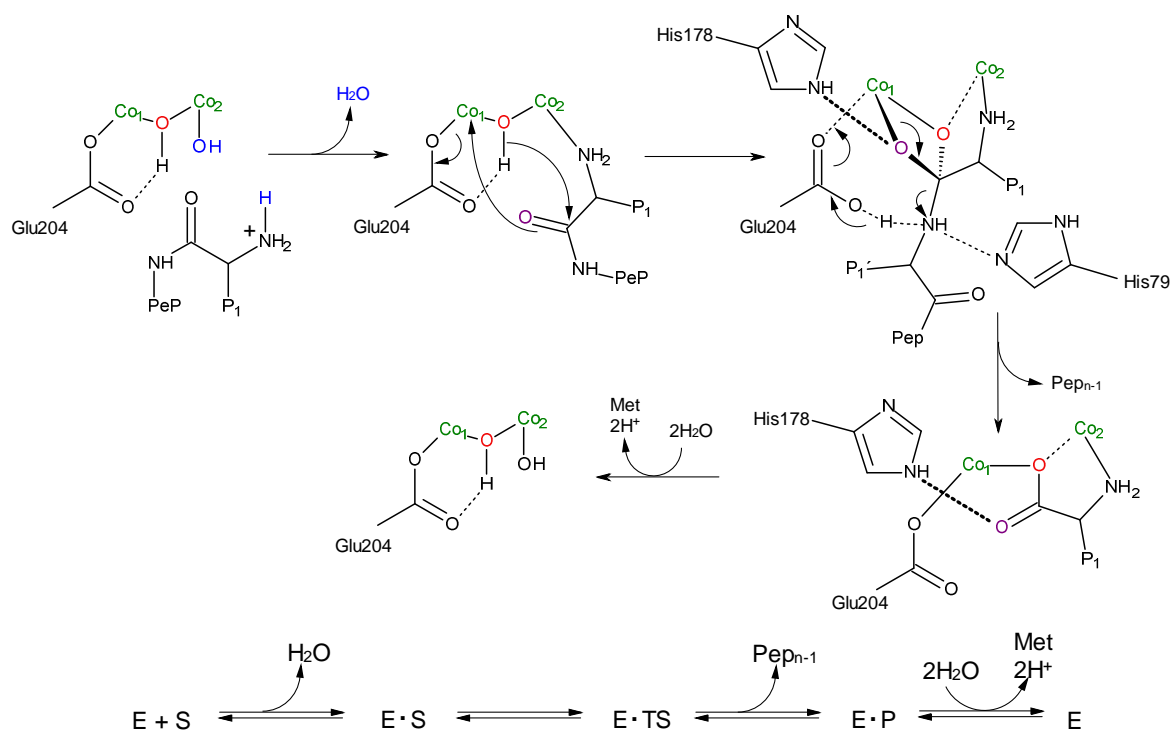
Figure 9. Proposed reaction mechanism for *E. coli* methionine aminopeptidase (adopted from reference 40).

Thermolysins

The thermolysins constitute a class of hydrolytic enzymes, whose structure is very similar to the one of carboxypeptidase A. The arrangement of the active site is ascribable to a convergent evolution. All members of that family contain a HEXXH sequence motif where the zinc ion is coordinated by two histidine residues and a glutamic acid which is necessary for catalysis.[41] Thermolysin from *Bacillus thermoproteolyticus* (PDB code: 3TMN[42]) is a thermostable zinc endopeptidase which preferentially cleaves peptides with a bulky, hydrophobic residue, for example phenylalanine or leucine in the R_1' position.[43] In addition to the zinc ion, there are four calcium ions necessary for the thermal stability of the enzyme. The zinc ion in the active site is coordinated to three amino acids, His142, His146, Glu166 and a water molecule (see Figure 10).[44]

The catalytic mechanism of thermolysin was described by Hangauer *et al.*[45] and it is similar to that of other metallopeptidases like carboxypeptidase A and methionine aminopeptidase (Figure 5 and Figure 9). When the substrate enters the active site, the water molecule bound to the zinc ion points towards a Glu143 residue that promotes the nucleophilic attack on the scissile peptide bond. A tetrahedral intermediate is further

stabilized by His231 and Tyr157. Again the proton which was accepted by Glu143 from the attacking water molecule is transferred to the leaving nitrogen. A stabilization of the doubly protonated tetrahedral nitrogen of the peptide bond takes place via the side chains of Asn112 and Ala113.
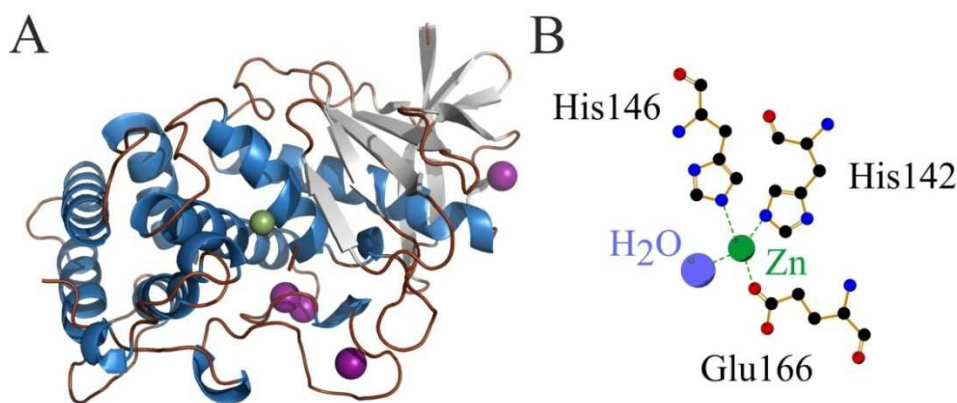
Figure 10. Overall structure of thermolysin from *Bacillus thermoproteolyticus* (PDB code: 3TMN). (A) Cartoon representation, showing the zinc ion in green and the calcium-ions in purple. Picture was prepared with the program PyMol.[28] (B) Close up of the zinc binding site, prepared with LigPlot.[22]

Superfamily of metzincins

A superfamily of zinc endopeptidases is called metzincins which occurs of a characteristic consensus motif **HEXXHXXGXXH**. This superfamily shows a significant relationship to the thermolysins (zinc binding motif **HEXXH**), although a histidine replaces a glutamate as a third ligand of the catalytically essential zinc. All members share a conserved methionine residue which is lacking in thermolysins. This Met is part of a fold into a hydrophobic pocket close to the active site and its importance presumably stems from the retention of the correct conformation and stabilization of the active site. Four distinct families of zinc peptidases, the astacins, matrix metalloproteases (matrixins, collagenases), adamalysins/reprolysins (snake venom proteinases/reproductive tract proteins) and the serralysins belong to the superfamily of metzincins as they show the above mentioned features. Moreover, their three-dimensional structures are topological similar (five-stranded β-sheet and three α-helices arranged in typical sequential).[46]

The metzincins are catalytically inactive in their native state (proenzyme). For most of the metzincins, the activation of the proenzyme involves a cleavage of the pro-peptide

resulting in an enzyme carrying two additional amino-terminal residues which are auto catalytically cleaved off in a second step. That allows the proenzyme to pass from the inactive to the active form.[47]

Astacin

Astacin from prawn of freshwater *Astacus astacus* is the first peptidase from the metzincins superfamily whose crystal structure has been solved (PDB code: 1AST[48], Figure 11). The physiological activity of this enzyme is the digestion of food proteins for example type I collagen or gelatin.[49] From its crystal structure, it is possible to infer the characteristic of the other family members. The terminal groups of all astacins are located in a proximal position to the active site. That implies an activation mechanism which involves the proteolytic cleavage of the N-terminal amino group of the proenzyme.[48]

Figure 11. Overall structure of astacin from *Astacus astacus* (PDB code: 1AST) (A) Cartoon representation, showing the zinc ion in green. Picture was prepared with the program PyMol.[28] (B) Close up of the zinc binding site, prepared with LigPlot.[22]

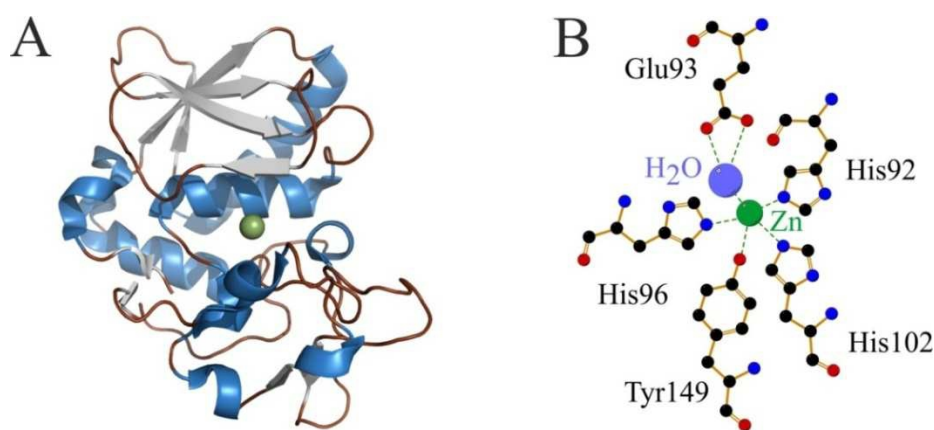The zinc center of the astacin is coordinated by five ligands in trigonal bipyramidal geometry. Tyr149 and His96 occupy the axial positions, whereas His92, His102 and a water molecule are placed in the equatorial plane. The carboxylic side chain of Glu93 is bound via hydrogen bonds to the nucleophilic water molecule coordinated to the zinc center.[47] This water molecule is additionally bound to Tyr149. This structural arrangement of the active site is analogous to the one of carboxypeptidase A and thermolysin; hence, the hydrolytic mechanism is the same (Figure 12).
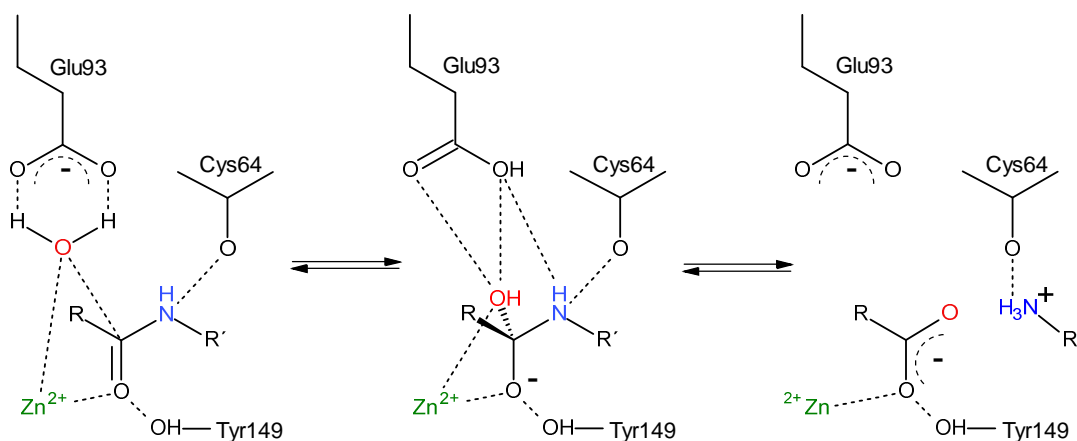
Figure 12. Catalytic mechanism of astacin (adopted from reference 47).

Despite the low sequence homology between astacin and thermolysin, they are superimposable on the base of a conserved helix in the active site. His102 and Tyr149 of astacin have a similar position to the ones of Glu166 and His231 of thermolysin. As a consequence, both enzymes employ an analogous mechanism, whereby the water molecule bound to the zinc is a nucleophile. Tyr149 of astacin could play as similar role as the one of His231 of thermolysin in the polarization of the carbonylic group of the substrate and Glu93 in astacin instead of Glu143 in thermolysins acts as a general base during catalysis.[47]

Snake venom protease family

Snakes of the family *Crotalidae* produce extremely aggressive proteolytic enzymes in their venom glands. These toxins belong to the superfamily of metzincins and are further divided in adamalysin, atrolysin and H_2-proteinase. Those proteins are appointed to the hydrolysis of the extracellular matrix. All of these enzymes characterized so far are inhibited from chelating metals and contain zinc. The pro-haemorrhage and proteolytic activity of that family is dependent on the presence of the zinc ion.[46]

The X-ray structures of adamalysin and atrolysin reveal the presence of a calcium ion, which probably has an additional catalytical role.

Adamalysin II, a 24 kDa protein, formerly known as proteinase II, was the first snake venom protein whose crystal structure has been solved. The zinc-endopeptidase was

isolated from the poison of the snake *Crotalus adamanteus* and the structure (PDB code 1IAG[50]) of its active site reveals several analogies to the one of astacin.
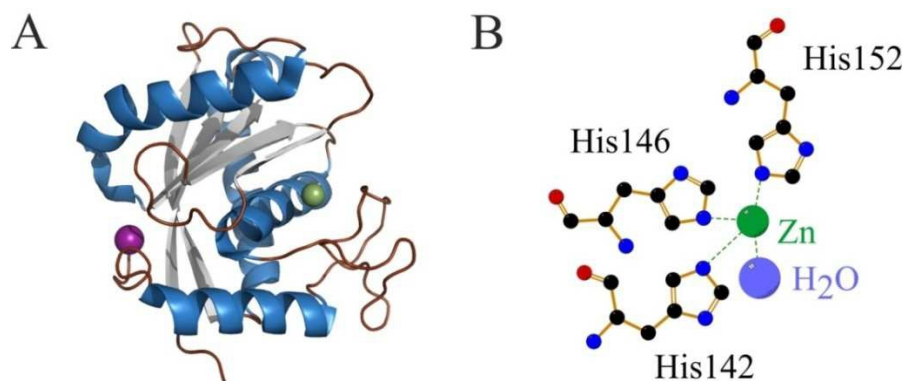
Figure 13. Overall structure of adamalysin II from *Crotalus adamanteus* (PDB code: 1IAG). (A) Cartoon representation, showing the zinc ion in green and the calcium ion in purple. Picture was prepared with the program PyMol.[28] (B) Close up of the zinc binding site, prepared with LigPlot.[22]

Thus, the zinc ion is tetrahedral coordinated by three histidine residues (His142, His146 and His152, respectively) and a water molecule (Figure 13). In contrast to astacin, adamalysin II lacks a fifth (tyrosine) residue.

It is worth to note that the oxygen atom closest to the zinc ion is the one belonging to the carboxylic group of Glu143. This oxygen may act as a base for the catalysis, albeit it is 4.1 Å far from the metal center. In general, the active site of the adamalysin is less complex that the one of the thermolysin although also these enzymes show a high catalytic efficiency.[50]

Atrolysin C (form d) is a zinc metalloprotease isolated from the venom of the western diamondback rattlesnake *Crotalus atrox* (PDB code 1ATL[51]). The zinc ion is essential for the haemorrhage and proteolytic activity whilst the replacement of Zn^{2+} with Co^{2+} does not determine a significant loss of the activity. According to the structure, a distorted tetrahedral coordination of the metal ion was found, whereby the zinc is bound to three histidues residues (His142, His146 and His152, respectively) and a water molecule. The water molecule bound to the zinc is placed 4.1 Å far from a glutamic acid residue (Glu143) as already mentioned in the case of adamalysin.

Beside the native structure, the enzyme was crystallized in presence of a dipeptide inhibitor. Figure 14 shows the binding of this inhibitor in the active site of atrolysin. The

dipeptide coordinates with both the oxygen atoms belonging to the same carboxylic moiety of a C-terminal side chain. Additionally, one of these oxygens is also bound to Glu143 via a hydrogen bond. In contrast to the proposed mechanism for the hydrolysis catalyzed by the thermolysin, it seems that the structures of adamalysin and atrolysin do not change significantly upon binding of a substrate or an inhibitor.[51]
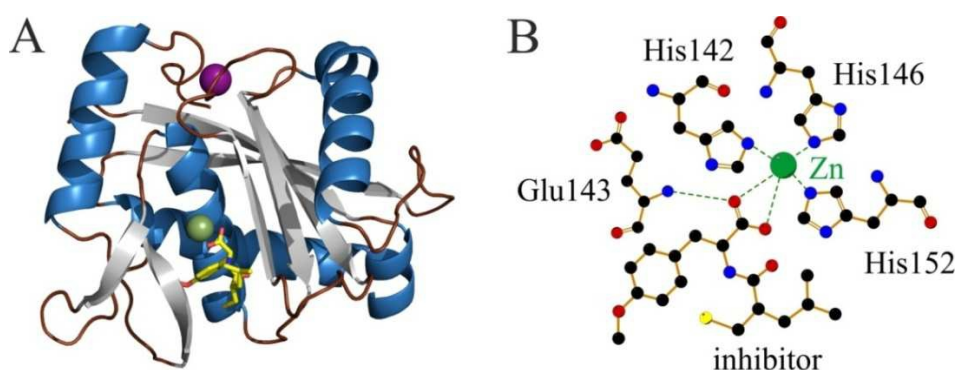
Figure 14. Overall structure of atrolysin C (form d) from *Crotalus atrox* (PDB code: 1ATL). (A) Cartoon representation, showing the zinc ion in green and the calcium ion in purple. Picture was prepared with the program PyMol.[28] (B) Close up of the zinc binding site with a dipeptide inhibitor bound, prepared with LigPlot.[22]

1.4.2 Hydrolases acting on ester bonds (Zinc phosphatases)

Alkaline phosphatase

Both prokaryotes and eukaryotes possess alkaline phosphatases, enzymes capable to catalyze the unspecific hydrolysis of a phosphate group from phospho-monoesters. These enzymes can additionally transfer phosphoryl groups to alcohols. The most studied alkaline phosphatase was isolated from *E. coli* (PDB code 1ALK[21]). In Figure 15 the native enzyme was complexed with phosphate (phosphate is a strong competitive inhibitor of alkaline phosphatase as well as a product of the enzyme). Alkaline phosphatase is a homodimer (MW 94 kDa), whereby the active sites are about 30 Å from each other. Each monomer contains two zinc ions (defined as Zn1 and Zn2) and a magnesium ion at a distance of about 5 Å from Zn2.[21] The crystal structure of the enzyme shows that the two catalytically essential zinc ions are separated by 4 Å. Although the magnesium ion is closer to the Zn2, it is not directly involved in the catalysis. Nonetheless, the Mg^{2+} contributes to the stabilization of the active site. Zn1 of alkaline phosphatase is classified

as catalytic zinc and it coordinates His331 and His412 and both oxygens of Asp327 (bidentate). Zn2 is bound to Asp369, His370 and Asp51, the last being a bridged ligand between Zn2 and the magnesium ion. Mg^{2+} is finally bound to Asp153, Thr155 and Glu322. Spectroscopic studies showed that Zn1 is likely penta-coordinated, whereas Zn2 and Mg might be hexa-coordinated by means of additional water molecules. Additional hydrogen bonding interactions with Asp153 hold the water molecules in place.
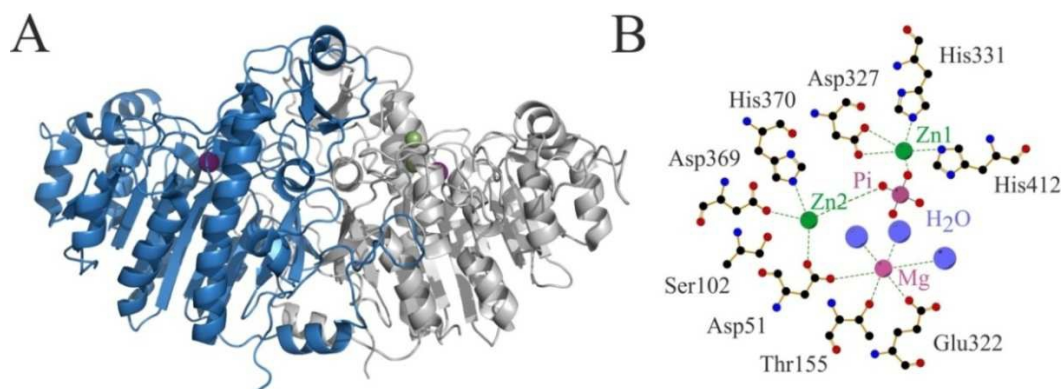
Figure 15. Overall structure of alkaline phosphatase from *E. coli* (PDB code: 1ALK). (A) Cartoon representation, showing the zinc ions in green and the magnesium ion in purple. Picture was prepared with the program PyMol.[28] (B) Close up of the metal binding site, prepared with LigPlot.[22]

The presence of a serine residue (Ser102) close to the couple Zn1, Zn2 is fundamental for the catalytic activity. In fact, there are strong evidences that this amino acid is reversibly phosphorylated during the catalysis and that Zn1 is the one able to coordinate the phosphate ion.[21, 52] In Scheme 3 the overall reaction mechanism concerning hydrolase and transphosphorylation activity of alkaline phosphatase as described by Wilson and Dayan[53] is depicted:
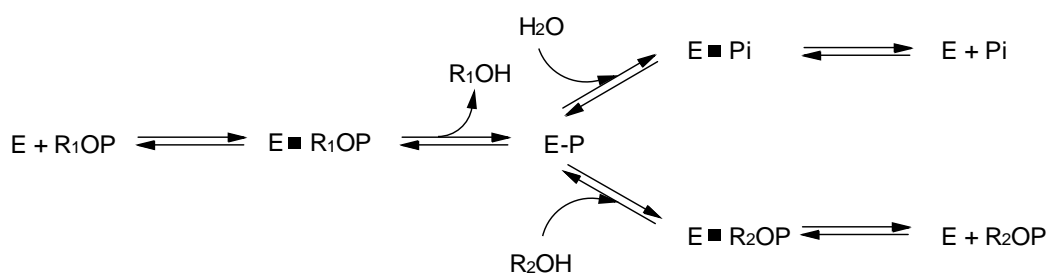
Scheme 3. Kinetic scheme of hydrolase and transphosphorylation reaction catalyzed by alkaline phosphatase (adapted from reference 52).

Ser102 is phosphorylated creating a covalently linked phosphoseryl intermediate E-P. This complex can either be hydrolyzed to a noncovalent enzyme-phosphate complex (E•Pi) or undergo a transphosphorylation reaction where the phosphate is transferred to an acceptor (R$_2$OH) like Tris or ethanolamine. The hydrolysis of the phosphoenzyme E-P is rate limiting under acidic conditions, whereas the phosphate dissociation from the E•Pi complex is rate limiting under basic pH.

The catalytic mechanism for the hydrolysis of phosphoesters is briefly summarized below (Figure 16):[21]
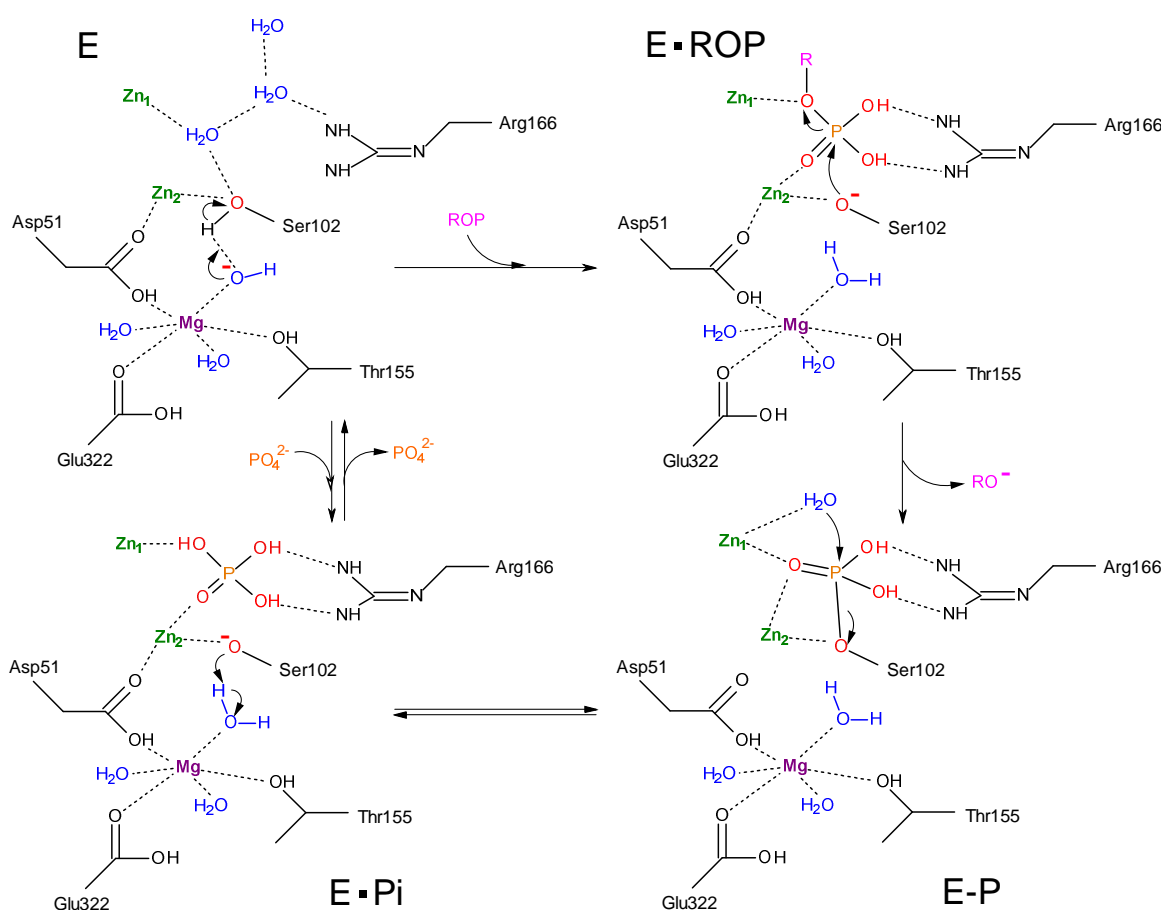
Figure 16. Catalytic mechanism for the hydrolysis of phosphoresters by alkaline phosphatase (adapted from reference 21).

Three water molecules fill the active site in the free enzyme and Ser102 is hydrogen bonded with the magnesium-coordinated hydroxide ion. The phosphomonoester enters in the active site to form the enzyme-substrate complex. The phosphate group coordinates the

Zn1 center employing a deprotonated oxygen atom after removal of a water molecule from the metal center. The result is the activation of the phosphorous atom to a subsequent nucleophilic attack. Furthermore, the binding of the substrate is synergetically increased by action of the positively charged Arg166 residue, proximal to the active site.

As a consequence of this process, a conformational rearrangement moves Ser102 towards Zn2, enabling the coordination of the serine after deprotonation of the hydroxyl moiety.

The Ser102 attacks the phosphorous atom which leads to the cleavage of the P-O ester bond; the alcohol is then set free from the active site. The release of the alcohol provokes the formation of a phosphoryl-serine via cleavage of the Zn1-phosphate bond and the concomitant deprotonation of the second water molecule present in the coordination sphere of the zinc.

Thus, the Zn1-OH group attacks the previously generated phosphoryl intermediate and the Zn2-Ser102 bond is subsequently restored. In other words, the phosphoryl group shifts from the serine residue to the zinc ion.

<u>Zn-dependent prokaryotic phospholipase C</u>

The phosphatidylcholin-hydrolyzing phospholipase C (PLC), a monomeric, 245 amino acids containing enzyme, was isolated from *Bacillus cereus* (PDB code 1AH7[54]). Two zinc ions are tightly bound to the protein scaffold. A more detailed structural survey revealed the presence of three zinc ions; however atomic absorption analysis showed the presence of 2.3 mol Zn per mol PLC, indicating that the third metal site is partially occupied even in solution (Figure 17). All three metal ions are oriented approximately trigonal bipyramidal. The relative distances among these zinc ions are: Zn1-Zn2 4.7 Å, Zn1-Zn3 3.3 Å (those zinc ions are bridged by Asp122), Zn2-Zn3 4.7 Å. The crystal structure of the enzyme bound to a substrate analogue showed that all the three zinc ions interact with the phosphonic group. The interaction of the three metal centers with the substrate is possible due to the displacement of the water molecules previously bound. However, the water molecule responsible for the hydrolytic attack may not come from the coordination sphere of a zinc ion but it may be bound to Glu146 via a hydrogen bond.

Figure 17. Overall structure of Zn-dependent prokaryotic phospholipase C from *Bacillus cereus* (PDB code: 1AH7). (A) Cartoon representation, showing the three zinc ions in green. Picture was prepared with the program PyMol.[28] (B) Close up of the metal binding site, prepared with LigPlot.[22]
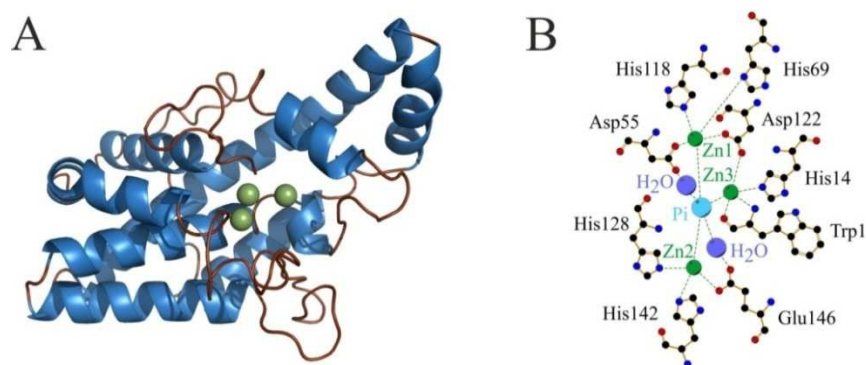
Nuclease P1

Nuclease P1, a glycoprotein, acts as a phosphomonoesterase as well as phosphodiesterase and cleaves preferentially single RNA and DNA filaments at the 3´-phosphate extremity to yield the 5´-mononucleotides. The structure of the enzyme obtained from *Penicillium citrinum* (PDB code: 1AK0[55]) is constituted by 270 amino acids and shows three zinc ions per molecule which are essential for catalytic activity.[56]
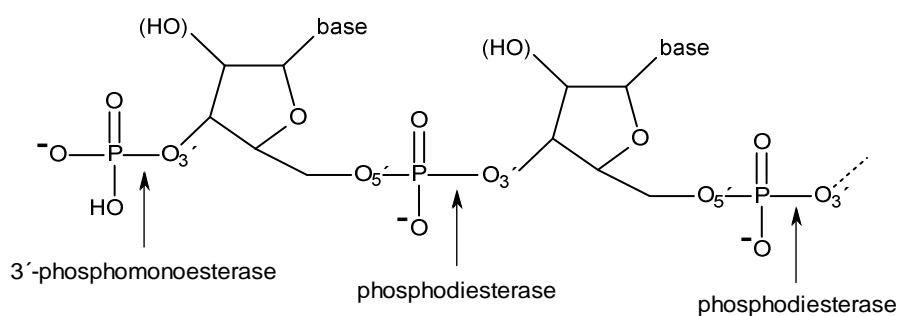
Figure 18. The two enzymatic activities of nuclease P1.

The zinc ions are grouped in a cluster, whereby Zn1 and Zn3 constitute a binuclear metal site (distance Zn1 to Zn3 ~ 3.4 Å) and a single metal site (Zn2) is found nearby. The distance from Zn2 to Zn1 and Zn3, respectively, are about 6 Å and 4.7 Å. Zn1 is coordinated to His60, His116, Asp45, Asp120 and a water molecule. Asp120 and this water form a bridge between Zn1 and Zn3 in the binuclear metal site. Zn3 is further bound to His6, to the N-terminal main chain nitrogen and the carbonyl oxygen to Trp1. Zn1 in

the single metal site binds to His126, His149, Asp153 and a water molecule (Figure 19). All the three zinc ions have approximately the same trigonal bipyramidal geometry with metal ligand distances between 1.9 and 2.5 Å.[55]

Figure 19. Overall structure of nuclease P1 from *Penicillium citrinum* (PDB code: 1AK0). (A) Cartoon representation, showing the zinc ions in green and the carbohydrate residues in yellow sticks. Picture was prepared with the program PyMol.[28] (B) Close up of the metal binding site, prepared with LigPlot.[22]

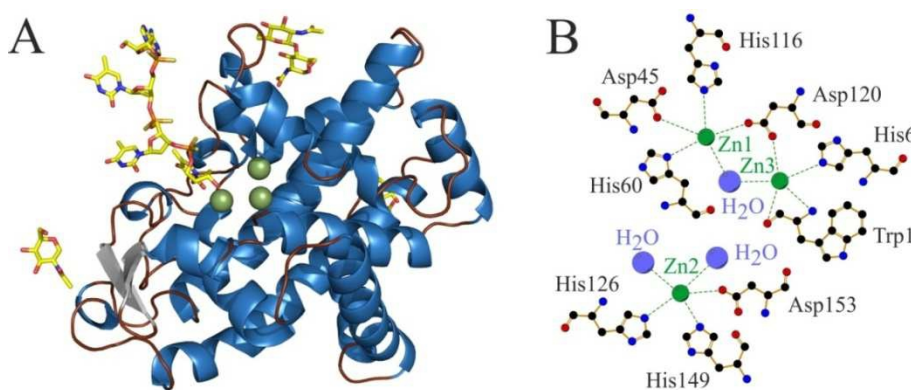The most plausible mechanism (Figure 20) for the nuclease P1 envisages the binding of the phosphate group (i.e. from the nucleotide) to the Zn2 via the free oxygen 3´.
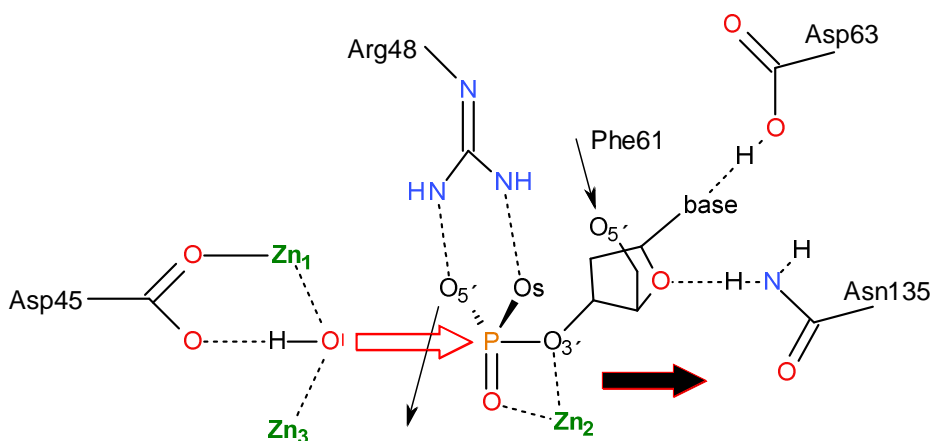
Figure 20. Schematic picture of the proposed three-metal ion mechanism (adapted from reference 57).

That displaces the two water molecules previously bound to the Zn2. Then, a residue of Phe61 allows the orientation of the nucleotide base in a way that it forms a hydrogen bond with Asp63. Simultaneously, a water molecule between Zn1 and Zn3 acts as a nucleophile and attacks the phosphate ion towards the $P-O_{3'}$ axis. In the transition state the hydroxyl

ion and the 3′ oxygen occupy the apical positions of a trigonal bipyramide. The function of Zn2 is to stabilize the oxygen $O_{3'}$, whereas Arg48 neutralizes the additional negative charge in the transition state. Finally, the cleavage of the phosphate group takes place forming the 5′-nucleoside.[57]

Phosphotriesterases

Phosphotriesterases, also named organophosphoros hydrolases, are bacterial enzymes, which catalyze the hydrolysis of organophosphates like paraoxon, parathion, sarin, soman and other inhibitors of mammalian acetylcholinesterases (Figure 21). They are useful for the detoxification of these classes of toxins and agriculture pesticides.[58]
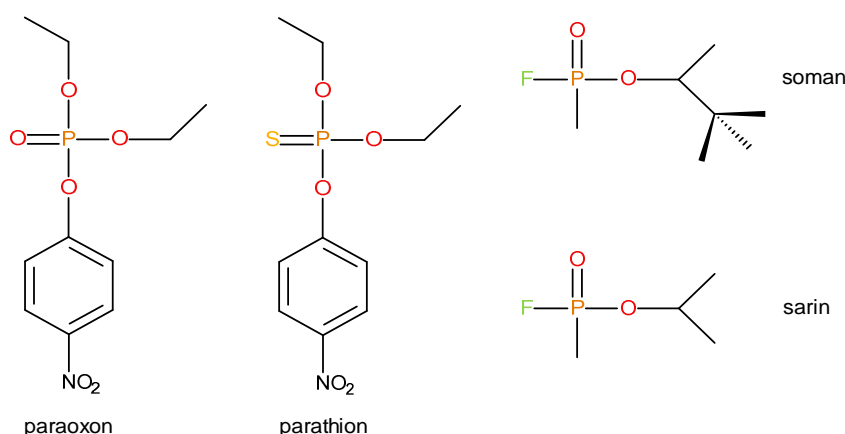
Figure 21. Structures of organophosphate compounds hydrolyzed by phosphotriesterases.

The native enzyme isolated from *Pseudomonas diminuta* (PDB code: 1DPM[59]) contains two zinc ions, although a comparable activity has been observed when the zinc ion was replaced by Co, Ni, Cd or Mn.[60]. The structure of the enzyme was characterized via replacement of the zinc ions with two cadmium ions. The two metal centers are spaced at 3.7 Å distance. A carbamoyl-lysine and a water molecule act as bridged ligands between the two cadium (zinc) ions (Figure 22). The other ligands are His55, His57 and Asp301 in the case of Zn1 and and His201 and His230 for Zn2, respectively.

Comparing the holo- to the apo-enzyme, one can observe a significant structural rearrangement upon binding of the metals. In the proposed mechanism, inversion of configuration at the phosphorous center occurs postulating the reaction to occur via a S_N2-

type mechanism. A metal ion probably polarizes the P-O bond via oxygen binding and the attack of a water molecule generates a bipyramidal trigonal intermediate.[59]
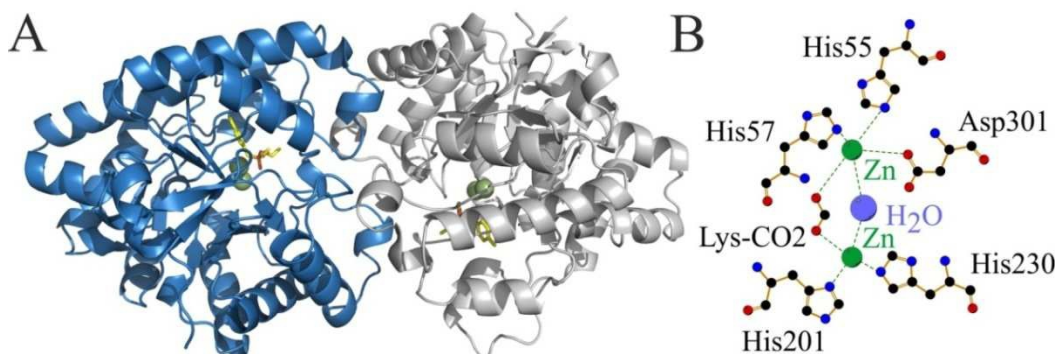
Figure 22. Overall structure of phosphotriesterase from *Pseudomonas diminuta* (PDB code: 1DPM). (A) Cartoon representation, showing the cadmium (zinc) ions in green and the bound substrate analog diethyl 4-methylbenzylphosphonate in yellow sticks. Picture was prepared with the program PyMol.[28] (B) Close up of the metal binding site, prepared with LigPlot.[22]

1.4.3 Hydrolases acting on carbon-nitrogen-bonds, other than peptide bonds

Metallo-ß-lactamases

Metallo-ß-lactamases are a group of enzymes widely distributed in both gram-positive and gram-negative bacteria that hydrolyze ß-lactam compounds such as penicillin, cephalosporin and carbapenem by cleavage of the ß-lactam C-N bond (Scheme 4). ß-Lactam antibiotics are the most used antibiotics which inhibit the synthesis of the cell wall in bacterial organism. The polysaccharide chains cannot be linked anymore, due to an inhibition of the step where the D-Ala-D-Ala peptide is cleaved, a prior step to the transpeptidation reaction that links polysaccharide chains. As a result bacteria often develop resistance against those antibiotics by synthesizing ß-lactamases.[61] So far four classes of ß-lactamase families are recognized: class A, B, C and D. Classes A, C and D act by a serine-based mechanism, whereas class B enzymes are zinc dependent. Class B metallo-ß-lactamases are further divided into three subclasses based on their sequence: class B1, B2 and B3.[62] A standard numbering scheme according on sequence alignment has been developed in 2001 by a metallo-ß-lactamase working group.[63] In this chapter the residues mentioned are numbered according to this scheme.
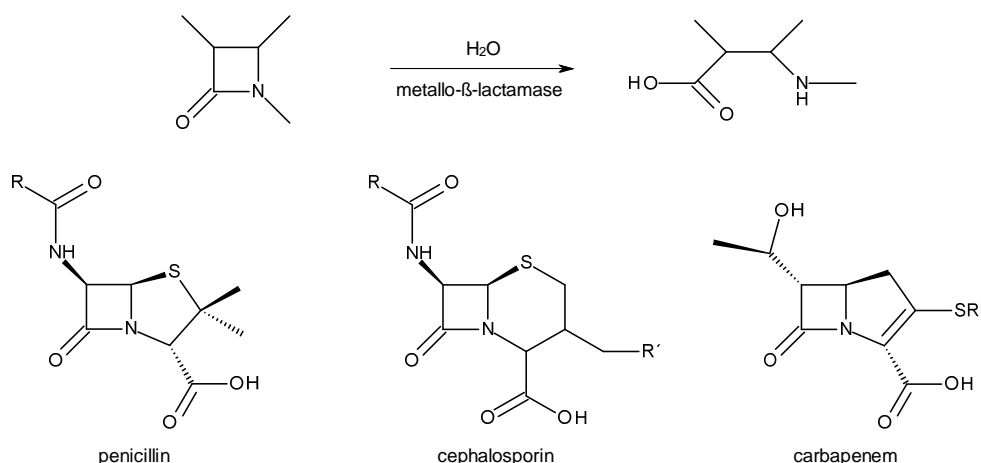
Scheme 4. Structures of ß-lactam antibiotics which are hydrolyzed by metallo-ß-lactamases.

The ligands to the active site zinc ions Zn1 and Zn2 are the key markers along the peptide chain and listed in Table 1. A common feature in all class B metallo-ß-lactamases is the coordination of Zn1 by two histidine residues, however the third ligand is a third histidine residue in the case of subclass B1, an asparagine residue in B2 and a histidine or glutamine residue in B3. Zn2 is ligated by an aspartic acid and a histidine in all three subclasses and by a cysteine in class B1 and B2 and a histidine in B3. As indicated in Table 1 all metallo-ß-lactamases bind two zinc ions forming a binuclear center. However, a few enzymes are known to be active also in the presence of just one zinc ion (enzymes from *Bacillus cereus*, *Bacteroides fragilis* and *Stenotrophomonas maltophilia* L1), although with lower catalytic activity.[64]

Table 1. Zinc ligands in metallo-ß-lactamases class B.

subclass	Zn1 ligands			Zn2 ligands		
B1	His116	His118	His196	Asp120	Cys221	His263
B2	Asn116	His118	His196	Asp120	Cys221	His263
B3	His/Gln116	His118	His196	Asp120	His121	His263

The overall fold of metallo-ß-lactamases is an αββα sandwich, whereby the core of the enzyme is formed by two β-sheets packed against one another. Two α-helices are on the surface of each sheet. Furthermore, a fifth α-helix is located between the last two β-strands of the N-terminal sheet forming a structural bridge to the C-terminal domain. Two αβ units are formed, the first one by the N-terminal half of the peptide chain, the second one by the C-terminal half. The active site is located in a channel at the top of the interface between

the two units.[64] In Figure 23 the metallo-ß-lactamase from *B. fragilis* (subclass B1) is shown (PDB code 1ZNB[65]). As mentioned before, Zn1 is coordinated to three histidine residues and one water molecule that forms a bridge between the two metal ions in a tetrahedral coordination sphere. Zn2 is further coordinated with histidine, aspartic acid and cysteine and a second water molecule trigonal bipyramidal.
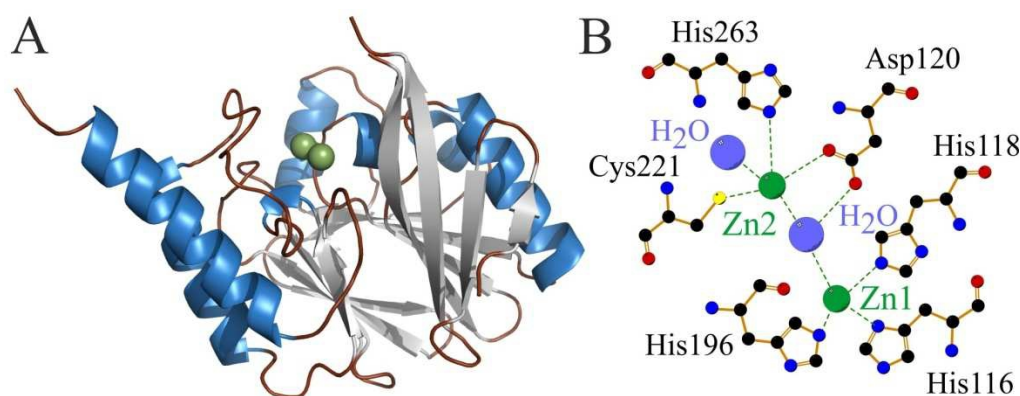
Figure 23. Overall structure of metallo-ß-lactamase from *Bacteroides fragilis* (subclass B1, PDB code:1ZNB). (A) Cartoon representation, showing the zinc ions in green. Picture was prepared with the program PyMol.[28] (B) Close up of the binuclear zinc binding site, prepared with LigPlot.[22]

As mentioned before, all metallo-ß-lactamases show a binuclear zinc binding site. However, there are also enzymes which are active in the mono-zinc state. As there are different electrostatic environments and different orientations and flexibilities of active site residues existing, the catalytic activities of both states have to be considered separately.[64]

As proposed by Carfi *et al*.[66], the structure-based mechanism of mono-zinc ß-lactamases resembles the mechanism as proposed for zinc peptidases (carboxylase and thermolysin) and is described in Figure 24. The water molecule, which is coordinated by the zinc ion and Asp120 (which serves as base during this reaction), is polarized and the resulting hydroxide attacks the ß-lactam carbonyl carbon atom. The resulting oxyanion is stabilized by the zinc ion and possibly by the amide group of Asn233. A change in the coordination of the zinc ion occurs from tetrahedral to trigonal bipyramidal, followed by the formation of a tetrahedral intermediate at the carbonyl carbon. As a result the C-N bond of the ß-lactam is cleaved and the proton which was accepted by Asp120 is transferred to the nitrogen atom.[66]
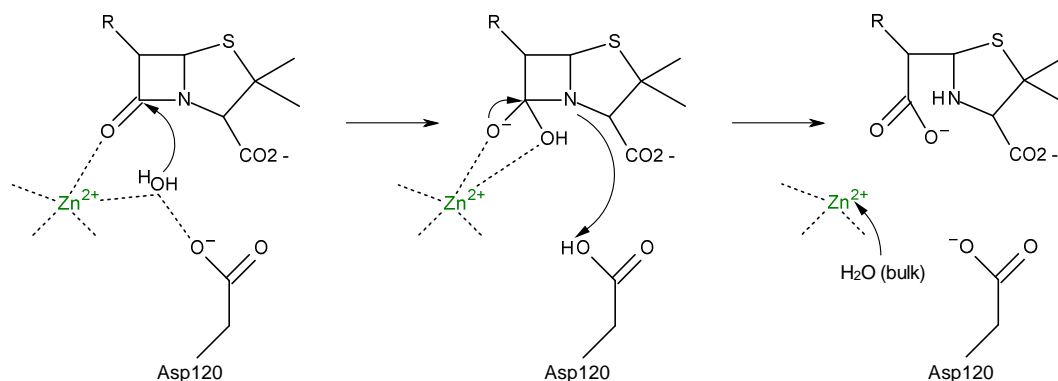
Figure 24. Proposed mono-zinc catalytic mechanism for the hydrolysis of penicillin by metallo-ß-lactamases (adapted from reference 66).

In the case of the binuclear zinc enzyme (Figure 25), which was first described for the *B. fragilis* metallo-ß-lactamase[65], the solvent molecule coordinated by the two zinc ions and Asp120 is likely existing as a hydroxide. The electron pair of the hydroxide is located appropriately for a nucleophilic attack onto the ß-lactam carbonyl carbon. Then an oxyanion tetrahedral intermediate is formed which is stabilized by one zinc ion and the amide group of Asn233. The coordination of Zn1 changes from tetrahedral to trigonal bipyramidal and from trigonal bipyramidal to octahedral in Zn2.[67] The ß-lactam C-N bond is cleaved and a proton is transferred from Asp120 to the nitrogen of the ß-lactam bond.
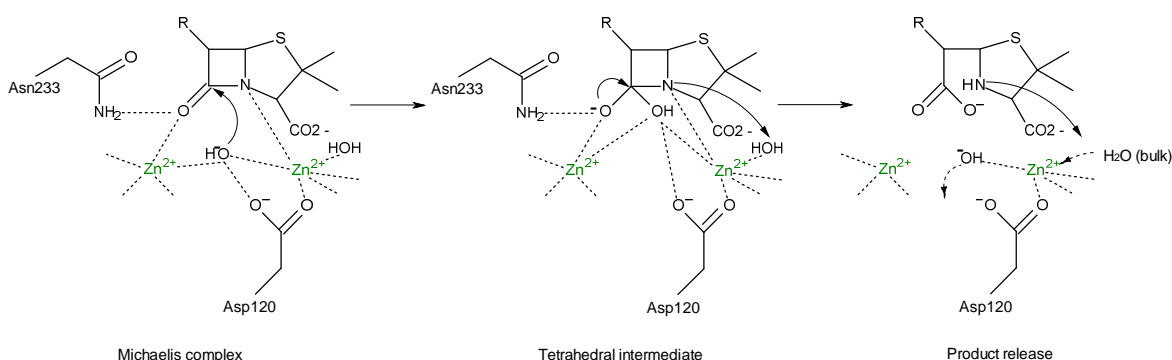
Figure 25. Proposed di-zinc catalytic mechanism for the hydrolysis of penicillin by metallo-ß-lactamases (adapted from reference 64).

The apical water molecule that coordinates to Zn2 is positioned to donate this proton. If this water is converted to a hydroxide, it moves to occupy the vacated nucleophilic hydroxide site, as the tetrahedral transition state decomposes and the degraded substrate diffuses away from the active site. The penta-coordination of Zn2 is then restored by a water molecule from bulk solvent.[64]

1.4.4 Lyases (Carbonic anhydrases)

Carbonic anhydrases (CA) are divided into three broad classes: α, β, and γ CAs, which are genetically unrelated families from mammalian, plants and bacteria without any amino acid homology between the classes.[68] Until now, 14 isoenzymes (CA I-XIV) of α-class CAs are known with varying tissue distributions and catalytic activities.[69]

The catalytic mechanism of the carbonic anhydrase is a typical example concerning the unique properties of enzymes containing zinc as a prosthetic group. The main features in the catalytic cycle of the carbonic anhydrase can be summarized as follows.

a) Taking into account the native form of the enzyme, the zinc is bound to three neutral histidine residues which permit to diminish the pKa of the further coordinated water molecule; in fact charged anionic ligands would transfer an excess of charge density on the zinc ion, hence reducing the acidity of the metal center.

b) During the catalytic cycle, the zinc ion changes its coordination number; as a consequence, two coordination sites must be quickly exchanged.

c) It is not actually known whether the CO_2 binds to the zinc center; nevertheless, the zinc ion activates the water molecule necessary for the hydrolysis and binds the product HCO_3^-.

The carbonic anhydrase catalyzes therefore the hydrolysis of CO_2 according to the following equilibrium:

$$CO_2 + H_2O \rightleftharpoons HCO_3^- + H^+$$
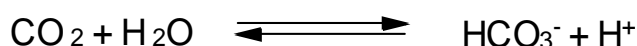
The chemical reaction normally occurs at an extremely low rate. In contrast, the zinc enzyme is capable to increase the hydrolysis reaction rate of a factor ca. 10^7. Due to this impressive acceleration rate and its important biological role in the conversion of CO_2 to HCO_3^- in red blood cells, the carbonic anhydrase has been widely investigated. The enzymatic hydrolytic reaction can be subdivided in two distinct half reactions:[70]

$$H_2O + E^*Zn\text{-}OH^- + CO_2 \rightleftharpoons E^*Zn\text{-}OH_2 + HCO_3^- \qquad (1)$$

$$E^*Zn\text{-}OH_2 + B \rightleftharpoons E^*Zn\text{-}OH^- + BH^+ \qquad (2)$$
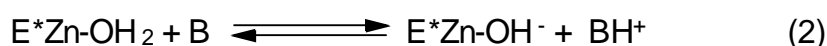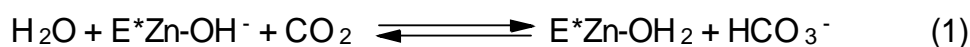
As mentioned before, 14 isoenzymes of the mammalian carbonic anhydrase are known, each of them containing one zinc ion. This variety of isoenzymes reflects the ubiquitous biological need to hydrate CO_2 to give HCO_3^- and H^+. All the isoenzymes found in mammals show a similar zinc coordination site; however, they differ in the way that the proton is released by the enzyme into the solvent during the final step of the catalytic mechanism.

The analysis of the crystal structure of human CA II (PDB code: 1MOO[71]) revealed that the monomeric enzyme is constituted by 259 amino acids and has a MW of approx. 30 kDa. As previously mentioned, the Zn^{2+} center is coordinated to three neutral histidine residues (His94, His96 and His119, respectively), whereas the fourth coordination position is filled by a water molecule which is further hydrogen bonded to Thr199. The final geometry resembles a distorted tetrahedral structure (Figure 26).
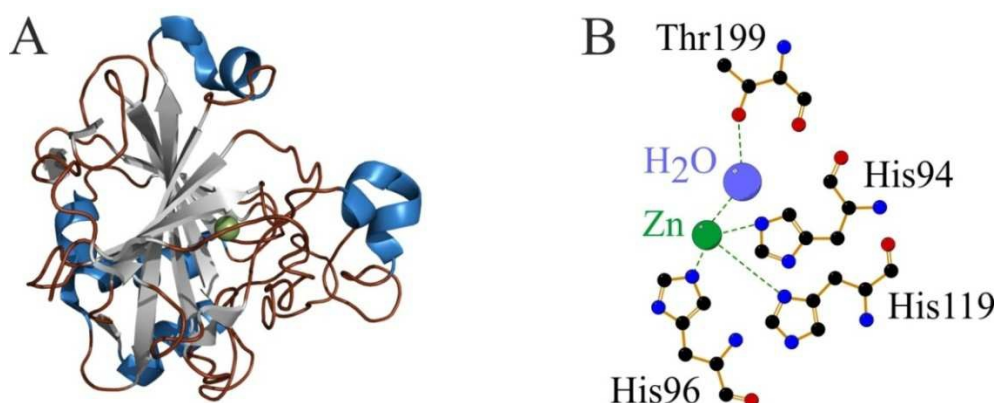
Figure 26. Overall structure of human CA II (PDB code: 1MOO). (A) Cartoon representation, showing the zinc ion in green. Picture was prepared with the program PyMol.[28] (B) Close up of the zinc binding site, prepared with LigPlot.[22]

The overall protein structure is stabilized by a network of hydrogen bonds. This stabilization is crucial for the catalysis, since the residues in the active site must be placed at specific intervals in order to optimize the coordination geometry during all the steps of the catalytic cycle, the electron density of the zinc ion and the correct orientation of the reagents.

The binding site of the carbonic anhydrase for the CO_2 is a hydrophobic pocket, adjacent to the active site (3 to 4 Å to the zinc ion). Val143 is located at the bottom of the pocket while Val121, Trp209 and Leu198 are located at its extremities. The hydrophobic pocket

possesses specific size and shape, which are fundamental for the reaction. Site direct mutagenesis has demonstrated that the replacement of a single amino acid residue in the hydrophobic pocket leads to variants having significantly lower or no activity.[72-74]

The presumable mechanism for the carbonic anhydrase can be summarized as follows (Figure 27):

Starting from the aqua-complex, a H^+ is transferred to the medium that the active enzyme is generated. The transfer of the H^+ does not occur directly from the Zn^{2+}-OH_2 to the medium but it is mediated by a "shuttle" amino acid residue. The efficiency of the transfer depends on the mobility of the "shuttle" amino acid residue and its exposure to the medium.
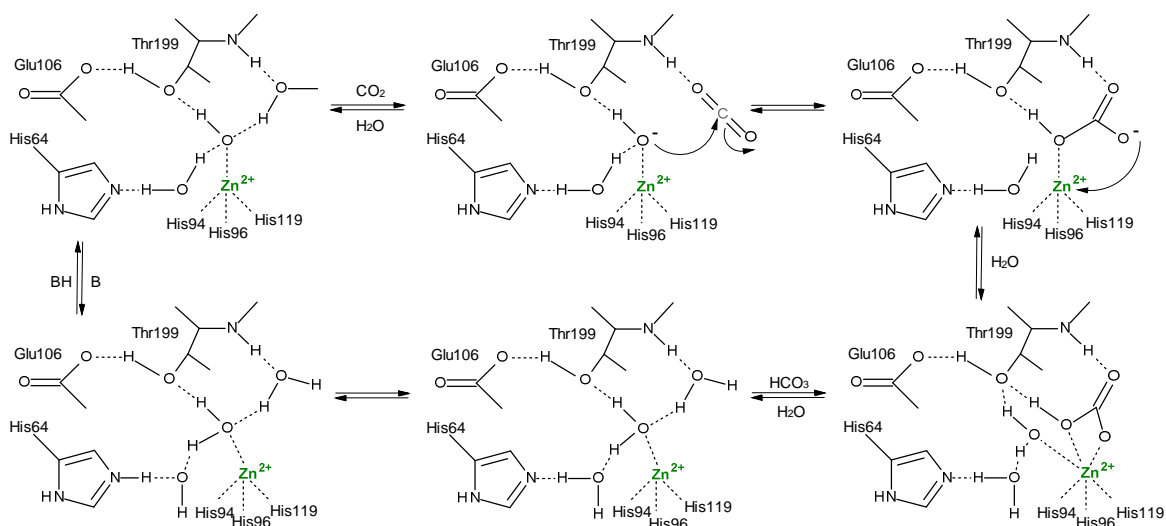
Figure 27. Catalytic mechanism of carbonic anhydrase (adopted from reference 69).

The isoenzymes II, IV, VI and VII transfer the proton in the following pathway: **(i)** the proton acceptor is His64, which points towards the active site; **(ii)** His64 accepts the proton and afterwards turns in order to release the proton into the medium; **(iii)** then, the pre-catalytic complex enzyme-CO_2 is formed. This adduct has such a short life that it is undetectable via X-ray spectroscopy. However, the structure of this complex can be inferred using the complex enzyme-phenol (i.e. the sole competitive inhibitor for the CO_2). The phenol enters into the hydrophobic pocket, forms Van der Waals interactions with the vicinal amino acid residues using its aromatic ring and finally generates a hydrogen bond between its hydroxyl group and the one coordinated to the Zn^{2+} center. **(iv)** The OH^- bound to the zinc center performs a nucleophilic attack to the CO_2 bound in the

hydrophobic pocket, forming HCO_3^-. The binding mode of HCO_3^- to the Zn^{2+} is matter of debate, mainly due to the role of Thr199 during the catalytic process. Studying the positioning of the azido ion (i.e. competitive inhibitor for HCO_3^-), it was deduced that HCO_3^- binds the zinc ion and establishes a Van der Waals interaction with the hydroxyl moiety of Thr199. Thus, the role of this residue is to maximize the catalytic efficiency by destabilizing the complex enzyme-product. This leads to a quick dissociation of the complex Zn^{2+}-HCO_3^-. **v)** In the final step, the HCO_3^- sets free from the active site due to the entry of another water molecule, hence allowing a new cycle to start again.[70]

The main role of the zinc is to act as electrostatic catalyst, since it stabilizes the transition state, having negative charge, which leads to the formation of HCO_3^-. Furthermore, as previously mentioned, the zinc ion decreases the pKa of the bound water molecules and acts therefore as a reservoir of nucleophilic OH^-.

In the human CA II, the zinc has four indirect ligands (also called second shell ligands) which stabilize the direct ligands (His94, His96 and His119, respectively) and help in the positioning for zinc ion coordination. Gnl92 forms a hydrogen bond with His94, Glu117 stabilizes His119, and Asp244 forms hydrogen bonds with His96 while Thr199 coordinates with the zinc-bound hydroxyl ion. The hydrogen bonds related to the direct ligands exercise a subtle but relevant influence on the zinc-protein affinity, on the pKa, on the reactivity of the solvent bound to the zinc ion and on the kinetic equilibria of the zinc-protein complex. Increasing the interaction between the zinc ion and the negative charges of the direct and indirect ligands diminishes the capability of the metal center to stabilize the hydroxyl group. The indirect interactions increase 100-fold the protein-metal affinity by pre-orienting the metal active site for the formation of the complex, minimize the increase in the conformational entropy required to the ligands coordination and increase the electrostatic interactions among the protein and the direct ligands of the zinc ion.[75]

1.4.5 Oxidoreductases

<u>Alcohol dehydrogenase (ADH)</u>

Alcohol dehydrogenases (ADHs) constitute a vast group of enzymes. Mammalian ADHs are divided into five classes, whereby classes I, II and IV (class IV is also called retinol dehydrogenases) have the ability to use ethanol as well as retinol as substrates, class III

lacks of ethanol activity.[76] In vivo, the zinc-dependent ADHs oxidize alcohols to aldehydes or ketones which are then further oxidized to carboxylic acids by the action of molybdenum-dependent aldehyde dehydrogenases. ADHs employ nicotinamide adenine nucleotide (NAD$^+$) as a cofactor. The oxidized form of the cofactor (NAD$^+$) accepts a hydride from the alcohol substrate to give the reduced form (NADH) and a carbonylic compound as a final product.[77]
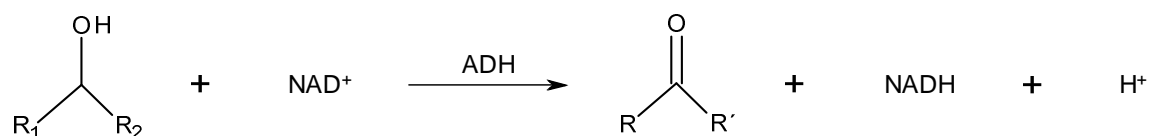
R$_1$ = aryl, alkyl
R$_2$= aryl, alkyl or H

The most studied zinc-dependent ADH was isolated from horse liver (HLADH, PDB code: 1HET[77]). The enzyme is a homodimer and each monomer has a MW of ca. 40 kDa. Each subunit contains two zinc ions and binds one molecule of NAD(H). In fact, only one of the two zinc ions (i.e. the catalytic zinc) is appointed to bind the coenzyme NAD$^+$.
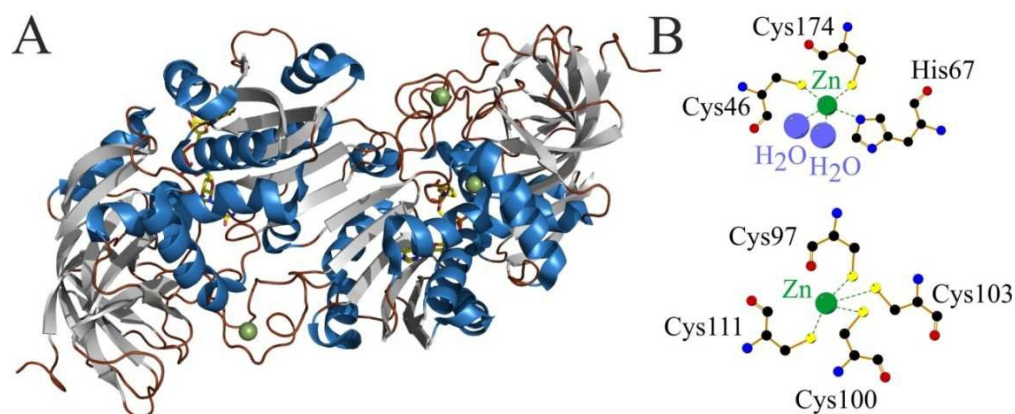
Figure 28. Overall structure of HLAHD (PDB code: 1HET). (A) Cartoon representation, showing the zinc ions in green and the nicotinamide residue in yellow sticks. Picture was prepared with the program PyMol.[28] (B) Close up of the zinc binding site, prepared with LigPlot.[22]

The catalytic zinc ion is anchored into the protein scaffold by a histidine residue (His67), two negatively charged cysteine residues (Cys46, Cys174) and a water molecule. The second zinc site is bound to four cysteine residues (Cys97, Cys100, Cys103 and Cys111) in almost perfect tetrahedral geometry and has a structural function (Figure 28). Particularly, the structural zinc ion impedes the access of exogenous water molecules into

the active site and permits to maintain the protein stability. Each subunit of the HLADH is divided in two domains; the substrate and the cofactor are located into a cavity between the two domains. The major domain includes the catalytic site with its two zinc ion and other protein groups which control the specificity of the substrate. The two zinc ions are placed at a distance of ca. 25 Å. Instead, the minor domain binds NAD^+. The two domains are linked each other to form a unique central nucleus in the protein structure. The native enzyme adopts a typical "open" conformation, whereas a "close" conformation is observed upon binding of NAD(H) in the active site.[78]

The nature of these conformations in the transition states are fundamental regarding the oxidation of alcohols. In fact, methanol is oxidized to formaldehyde at a very low rate, leading to several disorders and, in some cases, death (neurotoxicity of methanol). Longer chain alkylic alcohols such as ethanol, propanol and butanol are quickly oxidized to the related aldehydes. Furthermore, the oxidation of secondary alcohols by ADHs proceeds in a stereospecific fashion since only one of the two prochiral hydrogens is generally transferred as hydride (Scheme 5).[78, 79]
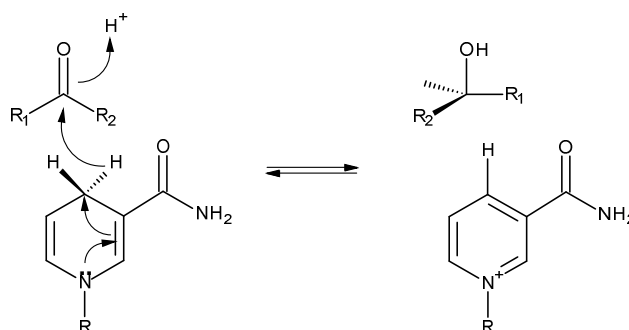
Scheme 5. Hydride transfer from NADH to an acceptor molecule.

There is much known about the kinetics of the assembly between the enzyme and the cofactor, however the actual mechanism regarding the hydride transfer and proton release is still under discussion.[80] In the standard textbook description of the HLADH mechanism[81], no water is involved in the reaction and can be described as follows: The alcohol substrate binds directly to the zinc ion site employing the oxygen atom of the hydroxyl moiety. This provokes the release of a water molecule, previously bound to the metal center and the deprotonation of the alcohol substrate. The binding of NAD^+ into the active site of the enzyme is accompanied by a conformational rearrangement of the protein

structure from the "open" to the "closed" conformation. This conformational change permits to exclude a further binding of other water molecules into the active site. Additionally, the "closed" conformation forces substrate and NAD^+ to adopt the correct reciprocal orientation, necessary for the hydride transfer. Hence the crucial step, that is the transfer of the hydride anion from the C-α carbon of the substrate to the nicotinamide ring of the cofactor, takes place. A ternary complex NADH-ADH-aldehyde is therefore generated, leading to a drastic decrease of the polarity of the active site. The aldehyde is then released from the active site and replaced by a water molecule; further water molecules can access into the active site, regenerating the "open" conformation. The "open" conformation favours the release of the NADH and another catalytic cycle can start again. The role of the zinc ion is to facilitate the deprotonation of the alcohol, increase the nucleophilicity of the hydride (i.e. due to an increased negative charge) and place the substrate in the correct position for the hydride transfer. The binding of the alcohol to the zinc ion occurs simultaneously to the release of the previously bound water molecule; thus the coordination number and the geometry of the zinc do not change, avoiding the binding of other potential ligands.

Another mechanism describes that a water molecule assists the zinc ion during the catalysis and goes through a cycle of protonation and deprotonation steps during catalysis.[77] This mechanism can be seen in Figure 29:
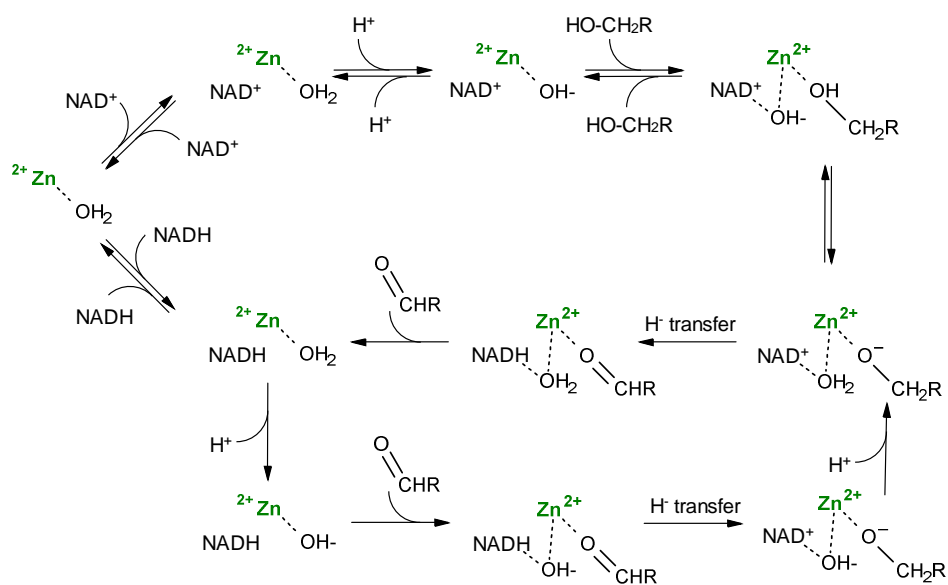
Figure 29. Reaction mechanism for HLADH where a water molecule is involved during catalysis (adapted from reference 78).

1.4.6   Structural zinc in biology

Besides the catalytic activity for hydrolysis and hydration of substrates, the zinc explicates an additional role in biology that is the stabilization of the protein scaffolds. In the previous paragraph, the structural role of the zinc ion has been already elucidated for the alcohol dehydrogenases. Among others, the most relevant examples of biological systems containing structural zinc ions which will be explained in this chapter are insulin, superoxodismutase (SOD) and zinc finger proteins.

Insulin

The zinc ion has a predominant structural function in peptide hormones, since it joins together two or more single polypeptide chains to form complex oligomers. For instance, zinc induces the oligomerization of the single units of insulin, a hormone employed for the treatment of diabetes. It is synthesized as proinsulin (single-chain precursor) in the pancreas and it is responsible for the regulation of the carbohydrate and fat metabolism in the body. After a removal of the N-terminal signal sequence, insulin is built by the formation of disulphide bonds.[82] The biologically active form of insulin is indeed the single monomer, possessing 51 amino acids and a MW of 5800 Da. Nevertheless, these monomers have a high tendency to generate aggregates. When zinc is present, the monomers assemble to the hexameric form (three dimers built up the hexamer) which shows reduced solubility and is stored in the granules as an insoluble precipitate. If the hexamer is secreted into the blood stream where only a low zinc concentration is present, the complex dissociates first into the dimers and then into the biologically active monomers.[83] That means that in the case of insulin zinc acts as a molecular "glue" (zinc binding constant 10^6 M^{-1}).[84]

The first structure of hexameric porcine insulin was reported in the year 1969 at 2.8 Å.[85] A more detailed structure was published in 1988 by Baker *et al*. (PDB code 4INS) at 1.5 Å (Figure 30).[86] The resting state of the insulin is a hexamer possessing a T_6 geometry. Each monomer is linked by two disulphide bonds. The hexamer assembly is held by two zinc ions, spaced at 15.9 Å. The zinc ions are placed along the ternary axis of symmetry of the protein aggregate. Thus, each Zn^{2+} is hexacoordinated by three histidine residues and three water molecules. The first eight amino acid residues of each B-chain are in an extended conformation. X-ray structure of the hexameric insuline revealed different conformations

depending on the reagent employed to induce the crystallization of the protein aggregate. A nomenclature has been adopted to distinguish those different hexameric conformations: T and R refer to an extended (T) or an α-helical (R) conformation in the first eight residues of the B-chain, respectively.[87]
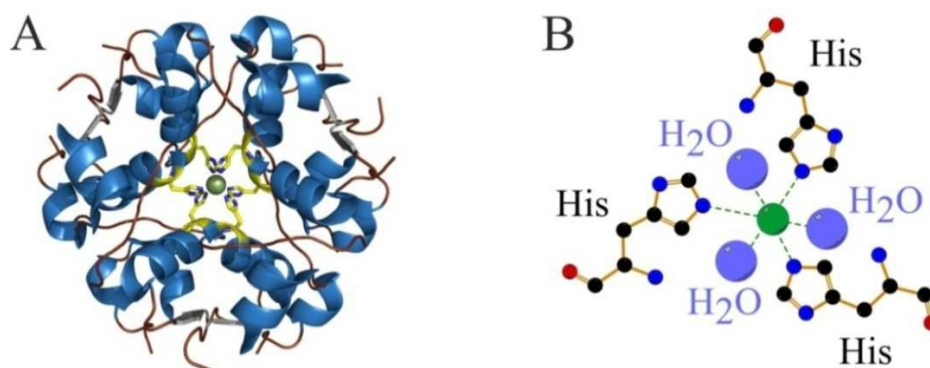
Figure 30. Overall structure of porcine insulin (PDB code: 4INS). (A) Cartoon representation, showing the T6 hexamer with zinc ions in green and the histidine residues in yellow sticks. Picture was prepared with the program PyMol.[28] (B) Close up of the zinc binding site, prepared with LigPlot.[22]

Crystallization in the presence of a high concentration of chloride ions (1 M sodium chloride) generates a T_3R_3 conformation with four instead of two potential zinc binding sites in one hexamer,[88] whereas the conformation R_6 is obtained upon addition of phenol.[89] The latter structure has a particular biological relevance since it is considered the active form of the insulin; thus pharmaceutical preparations of insulin contain phenolic compounds as stabilizers.

In the R conformation, both zinc ions have a pseudo-tetrahedral coordination and they are located into a cavity which is less accessible from exogenous molecules. In the structure T_3R_3, one zinc ion is pseudo-tetrahedral (one chlorid ion acts as the fourth ligand), while the latter is pseudo-octahedral (water molecules complete the octahedral structure).

As mentioned before, a molecule of phenol induces the transition from the T_6 to the R_6 forms, due to the conformational change of a peptide chain from a β-sheet to an α-helix. This modification is responsible for the closure of the whole protein structure, generating the biologically active form of the insulin. Various spectroscopic techniques have been employed to study the conformational transition from the T to the R forms in solution. An interesting approach is to use hexamers of insuline, whereby the zinc ion is replaced by

other bivalent metals such as Cu(II), Ni(II), Mn(II) and Co(II).[90] In particular, spectroscopic studies using Co(II)-insulin showed the strong correlation between the coordination geometry of the metal center and the protein structure. Studies in solution demonstrated that two allosteric processes lead to the stabilization of the R insuline: the binding of cyclic organic molecules to the six hydrophobic cavities of the hexamer and the coordination of anionic ligands to the metal center. The characteristic of the hexameric insuline to undergo structural modifications due to allosteric interactions with exogenous ligands is a strong insight that these events are also important during the formation of the complexes insuline-receptor.

Superoxide dismutase

Superoxide dismutases (SOD) are enzymes important in antioxidant defence in all organisms which are exposed to oxygen because they catalyze the dismutation of the toxic free radical superoxide ($O_2 \cdot^-$) into oxygen and hydrogen peroxide according to following chemical equation:[91]

$$2\,O_2 \cdot^- \; + \; 2\,H^+ \xrightarrow{\quad SOD \quad} H_2O_2 \; + \; O_2$$
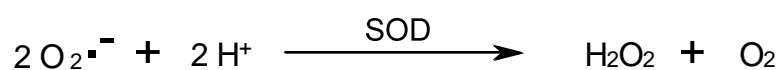
The product hydrogen peroxide is further converted to molecular oxygen and water by the catalase or the glutathione peroxidase.

SODs exist in several forms regarding their metal content, which can be copper and zinc or manganese or iron; therefore SODs are classified into three major families according to their metal cofactor. The copper-zinc-SOD (Cu-Zn-SOD) are mostly found in the cytosol of all eukaryotic cells. Conversely, superoxide dismutases from bacteria or chloroplasts are iron or manganese dependent. The choice of the copper in the eukaryotes came later during the evolution, probably due to the mediocre stability (i.e. kinetic and thermodynamic) of iron and manganese complexes in the bacterial SOD. In fact, loss of iron or manganese from SOD can damage the DNA. Instead, the Cu-Zn-SOD strongly binds the Cu^{2+} to avoid dissociation of the potentially toxic metal for the cell.[91] The Cu-Zn-SOD has an extremely high catalytic efficiency for the disproportionation of the superoxide ion with a k_{cat}/K_M for this reaction of $\sim 7 \times 10^9$ $M^{-1}s^{-1}$, thus very close to the diffusion velocity of molecules in aqueous solution.[92]

The Cu-Zn-SOD from bovine (PDB code: 2SOD[93]) is a dimeric enzyme with a MW of about 32 kDa, whereby each subunit contains one zinc and one copper ion (Figure 31).
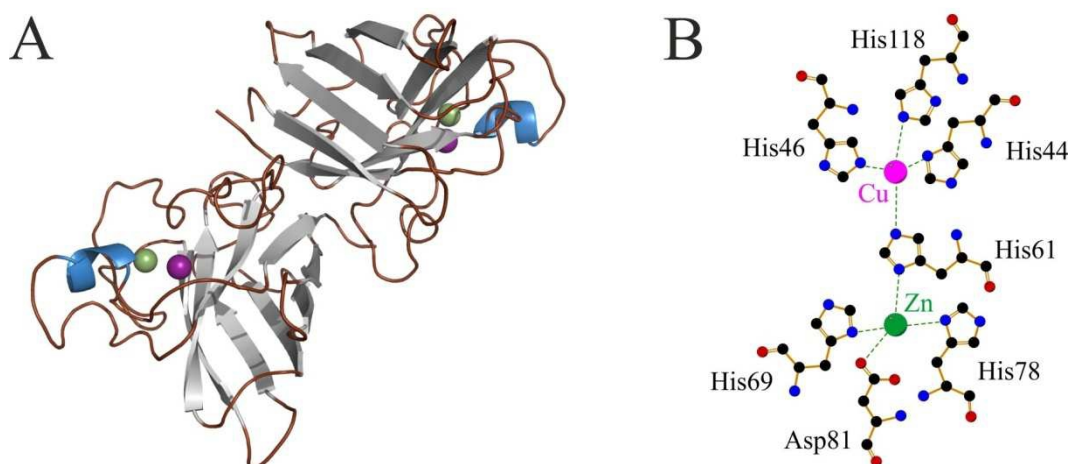
Figure 31. Overall structure of bovine Cu-Zn-SOD (PDB code: 2SOD). (A) Cartoon representation, showing the dimeric protein with the zinc ions in green and the copper ions in purple. Picture was prepared with the program PyMol.[28] (B) Close up of the copper-zinc binding site, prepared with LigPlot.[22]

The analysis of the X-ray structure of the bovine Cu-Zn superoxide dismutase shows a very compact structure (β-barrel). The zinc ion is tetrahedral coordinated with three histidue residues (His61, His69 and His78) and an aspartic acid (Asp81), whereas the copper ion is ligated to four histidine residues (His44, His46, His61 and His118). His61 bridges between the Cu and the Zn.

Concerning the zinc ion, its role is essentially structural due to the absence of any evidence regarding to the binding of water molecules, anions or other ligands to this metal center. Thus, it is unlikely that the superoxide ion can react at the Zn^{2+} site. In fact, the enzyme conserves almost completely its catalytic activity upon removal of the zinc ion from the protein scaffold.[93]

The mechanism of the Cu-Zn-SOD is summarized in Figure 32, whereby the zinc ion is not participating in the redox process but just stabilizing the intermediates during the catalytic cycle.[94] Hence, the mechanism of the disproportionation catalyzed by the Cu-Zn-SOD will not be discussed any further.
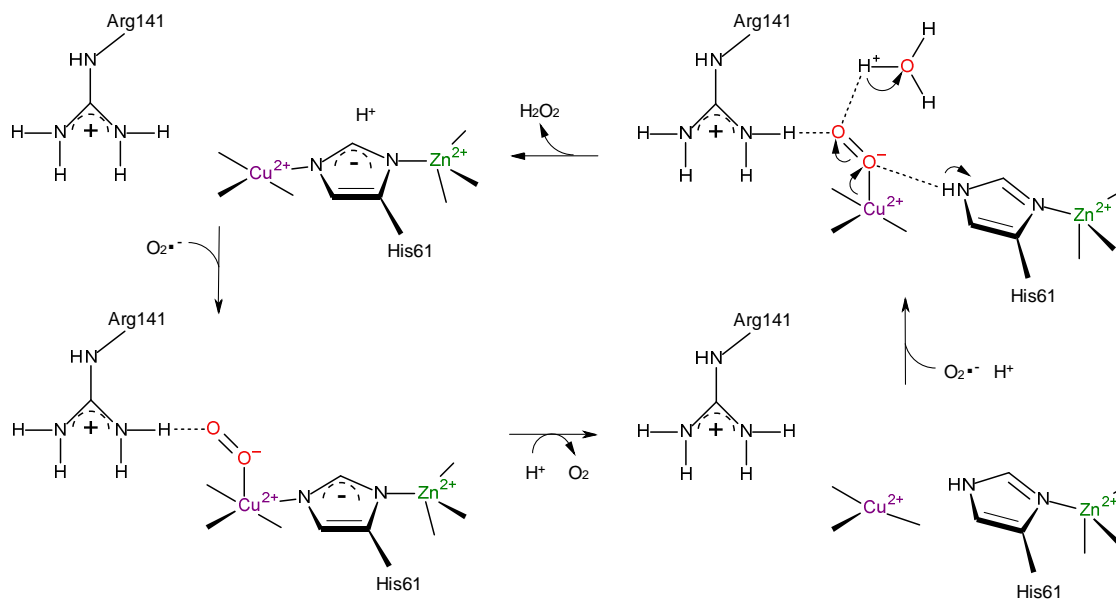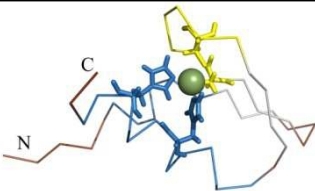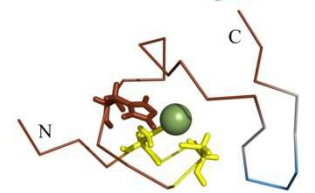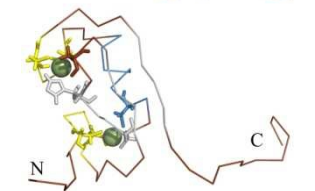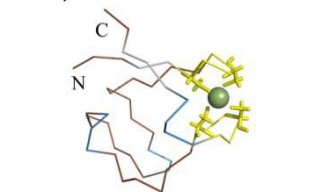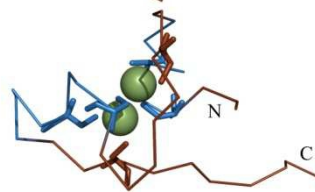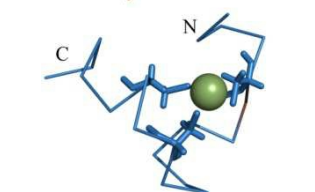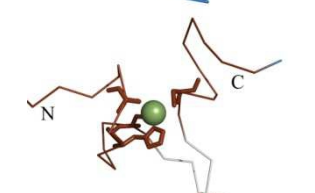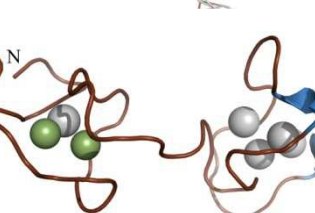
Figure 32. Schematic mechanism of the dismutation of superoxide by Cu-Zn-SOD (adapted from reference 94).

Zinc finger

The discovery of the zinc as a metal possessing a catalytic role in proteins is dated back to 1940 with the determination of the structure of the carbonic anhydrase. Only during the `80s, the structural role of the zinc was finally revealed. In 1983, it was proposed for the first time the structural role of the zinc in proteins containing steroids and nucleic acids. This hypothesis stemmed from the observation that the stability and biological activity of these proteins was reduced upon addition of EDTA. Further studies then revealed the presence of small domains based on Zn^{2+}. The first factor of genetic transcription finger has been found in the ovary from a South African frog *Xenopus laevis*; the protein was defined as zinc finger and named TFIIIA.[95] Nowadays the term "zinc finger" is used to identify any compact domain stabilized by a zinc ion.[96] Afterwards, zinc fingers have been found in a vast number of proteins involved in the genetic regulation. In spite of the large variety of these proteins, the main function of zinc fingers is the sequence-specific binding to DNA for regulation of gene expression.[97] For example more than 100 such proteins are present in the human brain.[98]

Eight different classes of fold groups based on structural properties in the vicinity of the zinc binding site were discovered during the last ten years. The most recurrent ones are depicted in Table 2.[99]

Table 2. Classification of zinc fingers according to their structure.

Fold group	Representative structure	Ligand placement
C_2H_2 like		Two ligands from a knuckle and two more from the C-terminus of a helix (PDB code: 1NCS[100])
Gag knuckle		Two ligands from a knuckle and two more from a short helix or loop (PDB code: 1DSV[101])
Treble clef		Two ligands from a knuckle and two more from the N-terminus of a helix (PDB code: 1CHC[102])
Zinc ribbon		Two ligands each from two knuckles (PDB code: 1IRN[103])
Zn_2/Cys_6		Two ligands from the N-terminus of a helix and two more from a loop (PDB code:1D66[104])
TAZ2 domain like		Two ligands each from the termini of two helices (PDB code: 1F81[105])
Zinc binding loops		Four ligands from a loop (PDB code: 1CW0[106])
Metallothionein		Cysteine rich metal binding loop (PDB code: 4MT2[107])

In the case of the zinc finger protein TFIIIA from the South African frog *Xenopus laevis*, nine tandem similar units were revealed from the structural analysis. These domains are fundamental for DNA binding and each of them is characterized from repetitive units of ca. 30 amino acids. The recurrent motif is the following: (Tyr-Phe)-X-Cys-X_{2-4}-Cys-X_3-Phe-X_5-Leu-X_2-His- X_{3-4}-His-X_{2-6}, whereby the term "X" indicates an amino acid that can vary from domain to domain. The zinc ion is tetra-coordinated to two cysteine residues and two histidine residues as could be seen in Figure 33A and B (PDB code: 1ZNF[108]) The three amino acid residues (-X_3-) that are placed within the two histidine residues form a hydrophobic core adjacent to the coordination site of the zinc ion. The biological function of these three residues may be the stabilization of the overall structure. The zinc finger TFIIIA protects 45 pairs of DNA bases.[95]

Figure 33. Overall structures of zinc finger proteins. (A) Cartoon representation of TFIIIA from *Xenopus laevis* (PDB code: 1ZNF) with the zinc ion in green. (B) Close up of the zinc binding site of TFIIIA. (C) Cartoon representation of Zif268 from *Mus musculus* in complex with DNA. Pictures were prepared with the programs PyMol[28] and LigPlot[22].

The first proposed structure regarding the interaction between a zinc finger and DNA was obtained from the mouse (*Mus musculus*, PDB code: 1AAY[109]) and is named Zif268. The complex is assembled by three zinc finger domains and the DNA double helix as depicted in Figure 33C.

The zinc fingers lie in the major loop of the double helix and each domain interacts with three contiguous DNA base pairs. These DNA bases are placed in such a way that only one of the DNA filaments forms covalent bonds with the zinc finger. The DNA filament involved in the binding is anti-parallel to the protein. In other words, considering the zinc finger from the amino to the carboxylic extremities, the DNA sequence involved is the one from the 5´ to the 3´ position. The interactions between DNA and protein involve the

amino acid residues 13, 16, 19 of the sequence X-(Thr-Phe)-X-Cys-X_2-Cys-X_3-Phe-X-X^{13}-X-X-X^{16}-Leu-X-X^{19}-His-X_3-His. In the first and third domains, X^{13}, X^{16} and X^{19} are an arginine, a glutamate and another arginine, respectively; these residues bind with 5`-GCG-3´. In the second domain, $X^{13}$, $X^{16}$ and $X^{19}$ are an arginine, a histidine and a threonine, respectively; these residues bind with 5`-TGG-3´. Five among the six direct interactions involve double hydrogen bonds between an arginine residue and a guanidine nucleotide, whereas the last one involves a single bond between a histidine residue and a guanidine nucleotide. The interaction of the zinc finger domains with the DNA bases occurs with the amino acid in position 2 (Asp2) belonging to the α-helix of each domain.[110]

Interestingly, the identification and the characterization of different classes of zinc finger domains revealed other proteins containing RNA or steroids, possessing different biological functions. For instance, the protein GAL4 is a transcription factor for the galactose oxidase which interacts with the DNA double helix in the same way as other steroid thyroid receptors.[111]

The zinc ion is fundamental in the zinc finger proteins. In fact, other metals such as iron and copper are capable to bind the DNA and the RNA but these metals also promote radical reactions which damage the nucleic acids. These radical reactions cannot occur in the presence of a zinc ion. The binding of the Zn^{2+} constrains the overall protein structure to adopt a more folded conformation, characterized by several bends and turns. The correlation between the binding of the metal and the protein folding is the main principle to elucidate the functions of these proteins. Finally, the binding of the Zn^{2+} to the protein does not generate any stabilization energy and the tetrahedral geometry is suitable to tolerate high torsional strengths. Yet, it is not clear whether the zinc accomplishes solely a structural function or it is involved in the transmission of the genetic information.[97]

1.4.7   Zinc storage proteins

Metallothioneins

Metallothioneins (MTs) are a family of small (6-7 kDa) cysteine-rich proteins that are localized in the membrane of the Golgi-apparatus. They are found in animals, higher plants, eukaryotes and some prokaryotes. About 33% of the amino acids are cysteins and

they have the highest metal content after ferritins.[112] MTs are able to bind physiologically (Cu^{2+} and Zn^{2+}) as well as nonessential (Cd^{2+}, Ag^{2+} and Hg^{2+}) metal ions through the thiol group of cysteins in a high thermodynamic stability but a low kinetic stability. In principle the metal binding is very tight, however in MTs a high metal exchange with other proteins is possible. The function of metallothioneins is still unclear; however it is suggested that they play a role in the intracellular distribution in the metals they bind. Furthermore, they may play a role in heavy metal detoxification, in the protection against oxidative stress and they additionally scavenge nitrogen oxide.[113]
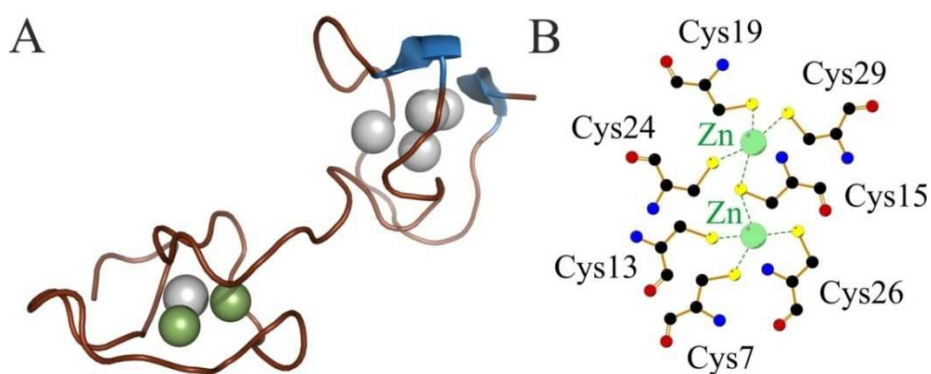
Figure 34. Overall structure of rat liver metallothionein 2 (PDB code: 4MT2). (A) Cartoon representation, showing the zinc ions in green and the cadmium ions in grey. Picture was prepared with the program PyMol.[28] (B) Close up of the zinc binding site, prepared with LigPlot.[22]

A classification of MTs was done according to their amino acid sequence and structural characteristics into (i) class I: they show homology to horse liver, (ii) class II: MTs show no homology to horse liver and (iii) class III: cysteine-rich enzymatically synthesized peptides. Figure 34 displays the overall structure and the metal binding sites in rat liver MT2 (PDB code: 4MT2[107]). In principle all mammalian MTs bind 7 metal ions in total in two separate domains: 3 metals are bound in the N-terminal ß-domain, whereas 4 metal ions are found in the C-terminal α-domain. Furthermore, each metal ion is ligated to the sulphur atoms of four cysteine residues.[114]

1.5 References

1. Broadley, M. R., White, P. J., Hammond, J. P., Zelko, I., and Lux, A. (2007) Zinc in plants., *New Phytologist 173*, 677-702.

2. Auld, D. S. (2001) Zinc coordination sphere in biochemical zinc sites., *BioMetals 14*, 271-313.

3. Vallee, B. L., and Auld, D. S. (1992) Active zinc binding sites of zinc metalloenzymes., *Matrix Suppl. 1*, 5-19.

4. Jolly, W. L. (1984) *Modern Inorganic Chemistry.*, New York: McGraw-Hill.

5. Lewis, G. N. (1923) *Valence and the Structure of Atoms and Molecules.*

6. Pearson, R. G. (1963) Hard and Soft Acids and Bases., *J. Am. Chem. Soc. 85*, 3533-3539

7. Pearson, R. G. (1968) Hard and soft acids and bases, HSAB, part 1: Fundamental principles., *J. Chem. Educ. 45*, 581.

8. Pearson, R. G. (1968) Hard and soft acids and bases, HSAB, part II: Underlying theories., *J. Chem. Educ. 45*, 643.

9. Auld, D. S. (2005) Zinc enzymes., in *King RB (ed) Encyclopedia of inorganic chemistry.*, pp 5885-5927, Wiley, Chichester.

10. Patel, K., Kumar, A., and Durani, S. (2007) Analysis of the structural consensus of the zinc coordination centers of metalloprotein structures., *Biochim Biophys Acta 1774*, 1247-1253.

11. Howard, J. (1996) Crystal Chemistry and Refractivity., p 31, Courier Dover Publications.

12. McCall, K. A., Huang, C., and Fierke, C., A. (2000) Function and Mechanism of Zinc Metalloenzymes., *J Nutr. 130*, 1437S-1446S.

13. Parkin, G. (2004) Synthetic Analogues Relevant to the Structure and Function of Zinc Enzymes., *Chem. Rev. 104*, 669-767.

14. Vallee, B. L., and Auld, D. S. (1990) Active-site zinc ligands and activated H2O of zinc enzymes., *Proc. Natl. Acad. Sci. U.S.A. 87*, 220-224.

15. Eklund, H., Nordström, B., Zeppezauer, E., Söderlund, G., Ohlsson, I., Boiwe, T., and Brändén, C. I. (1974) The structure of horse liver alcohol dehydrogenase., *FEBS Lett. 44*, 200-204.

16. Honzatko, R. B., Crawford, J. L., Monaco, H. L., Ladner, J. E., Ewards, B. F. P., Evans, D. R., Warren, S. G., Wiley, D. C., Ladner, R. C., and Lipscomb, W. N. (1982) Crystal and molecular structures of native and CTP-liganded aspartate carbamoyltransferase from *Escherichia coli.*, *J Mol Biol. 160*, 219-263.

17. Auld, D. S. (2004) Structural zinc sites., in *Handbook of Metalloproteins* (Messerschmidt, A., Bode, W., and Cygler, M., Eds.), pp 403-415, Wiley, Chichester.

18. Vallee, B. L. (1993) New perspective on zinc biochemistry: cocatalytic sites in multi-zinc enzymes., *Biochemistry 32*, 6493-6500.

19. Auld, D. S. (2004) Cocatalytic zinc sites., in *Handbook of Metalloproteins* (Messerschmidt, A., Bode, W., and Cygler, M., Eds.), pp 416-430, Wiley, Chichester.

20. Volbeda, A., Lahm, A., Sakiyama, F., and Suck, D. (1991) Crystal structure of *Penicillium citrinum* P1 nuclease at 2.8 A resolution., *EMBO J 10*, 1607-1618.

21. Kim, E. E., and Wyckoff, H. W. (1991) Reaction mechanism of alkaline phosphatase based on crystal structures. Two-metal ion catalysis., *J Mol Biol. 218*, 449-464.

22. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions., *Prot Engin 8*, 127-134.

23. Auld, D. S. (2001) Zinc sites in metalloenzymes and related proteins., in *Bertini I, Sigel A, Sigel H (eds) Handbook on metalloproteins*, pp 881-959, M. Dekker, New York.

24. Maret, W. (2004) Protein interface zinc sites: the role of zinc in the supramolecular assembly of proteins and in transient protein-protein interactions., in *Handbook of Metalloproteins* (Messerschmidt, A., Bode, W., and Cygler, M., Eds.), pp 432-440, Wiley, Chichester.

25. (2003) *The Handbook of Proteolytic Enzymes.*, Vol. **2**, Academic Press.

26. Stepanov, V. M. (1998) in *Handbook of Proteolytic Enzymes* (Barrett, A. J., Rawlings, N. D., and Woessner, J. F., Eds.), pp 1336-1228, Academic Press, San Diego.

27. Rees, D. C., Lewis, M., and Lipscomb, W. N. (1983) Refined crystal structure of carboxypeptidase A at 1.54 A resolution., *J.Mol.Biol. 168*, 367-387.

28. DeLano, W. L. (2002) The PyMOL Molecular Graphics System   http://www.pymol.org., (DeLano Scientific, S. C., CA, USA., Ed.).

29. Vendrell, J., Aviles, F. X., and Fricker, L. D. (2004) Metallocarboxypeptidases., in *Handbook of Metalloproteins* (Messerschmidt, A., Bode, W., and Cygler, M., Eds.), pp 176-189, Wiley, Chichester.

30. Christianson, D. W., and Lipscomb, W. N. (1989) Carboxypeptidase A., *Acc. Chem. Res. 22*, 62–69.

31. Buffone, G. J., Spence, J. E., Fernbach, S. D., Curry, M. R., O´Brien, W. E., and Beaudet, A. L. (1988) Prenatal diagnosis of cystic fibrosis: microvillar enzymes and DNA analysis compared., *Clin Chem. 34*, 933-937.

32. Taylor, A., Surgenor, T., Thomson, D. K., Graham, R. J., and Oettgen, H. (1984) Comparison of leucine aminopeptidase from human lens, beef lens and kidney, and hog lens and kidney., *Exp Eye Res. 38*, 217-229.

33. Burley, S. K., David, P. R., and Lipscomp, W. N. (1990) Molecular structure of leucine aminopeptidase at 2.7-A resolution., *Proc.Natl.Acad.Sci.USA 87*, 6878-6882.

34. Sträter, N., Sherratt, D. J., and Colloms, S. D. (1999) X-ray structure of aminopeptidase A from *Escherichia coli* and a model for the nucleoprotein complex in Xer site-specific recombination., *EMBO J 18*, 4513-4522.

35. Sträter, N., and Lipscomb, W. N. (2004) Leucine aminopeptidase., in *Handbook on Metalloproteins* (Messerschmidt, A., Bode, W., and Cygler, M., Eds.), pp 199-207, Wiley, Chichester.

36. Sträter, N., Sun, L., Kantrowitz, E., and Lipscomb, W. N. (1999) A bicarbonate ion as a general base in the mechanism of peptide hydrolysis by dizinc leucine aminopeptidase., *Proc Natl Acad Sci U S A. 96*, 11151-11155.

37. Ben-Bassat, A., Bauer, K., Chang, S. Y., Myambo, K., Boosman, A., and Chang, S. (1987) Processing of the initiation methionine from proteins: properties of the *Escherichia coli* methionine aminopeptidase and its gene structure., *J Bacteriol. 169*, 751-757.

38. Roderick, S. L., and Matthews, B. W. (1993) Structure of the cobalt-dependent methionine aminopeptidase from *Escherichia coli*: a new type of proteolytic enzyme., *Biochemistry 32*, 3907-3912.

39. Matthews, B. W. (2004) Methionine aminopeptidase., in *Handbook of Metalloproteins* (Messerschmidt, A., Bode, W., and Cygler, M., Eds.), pp 95-103, Wiley, Chichester.

40. Lowther, T. W., Zhang, Y., Sampson, P. B., Honek, J. F., and Matthews, B. W. (1999) Insights into the mechanism of *Escherichia coli* methionine aminopeptidase from the

structural analysis of reaction products and phosphorus-based transition-state analogues., *Biochemistry* **38**, 14810-14819.

41. Holland, D. R., Hausrath, A. C., Juers, D., and Matthews, B. W. (1995) Structural analysis of zinc substitutions in the active site of thermolysin., *Protein Sci.* **4**, 1955-1965.

42. Holden, H. M., and Matthews, B. W. (1988) The binding of L-valyl-L-tryptophan to crystalline thermolysin illustrates the mode of interaction of a product of peptide hydrolysis., *J.Biol.Chem.* **263**, 3256-3260.

43. Matsubara, H., Singer, A., Sasaki, R., and Jukes, T. H. (1965) Observations on the specificity of a thermostable bacterial protease "thermolysin". *Biochem Biophys Res Commun.* **21**, 242-247.

44. Matthews, B. W. (2004) Thermolysin., in *Handbook of Metalloproteins* (Messerschmidt, A., Bode, W., and Cygler, M., Eds.), pp 85-93, Wiley, Chichester.

45. Hangauer, D. G., Monzingo, A. F., and Matthews, B. W. (1984) An interactive computer graphics study of thermolysin-catalyzed peptide cleavage and inhibition by N-carboxymethyl dipeptides., *Biochemistry* **23**, 5730-5741.

46. Stöcker, W., Grams, F., Baumann, U., Reinemer, P., Gomis-Ruth, F. X., McKay, D. B., and Bode, W. (1995) The metzincins--topological and sequential relations between the astacins, adamalysins, serralysins, and matrixins (collagenases) define a superfamily of zinc-peptidases., *Protein Sci.* **4**, 823-840.

47. Stöcker, W., and Yiallouros, I. (2004) Astacin., in *Handbook of Metalloproteins* (Messerschmidt, A., Bode, W., and Cygler, M., Eds.), pp 116-129, Wiley, Chichester.

48. Bode, W., Gomis-Ruth, F. X., Huber, R., Zwilling, R., and Stöcker, W. (1992) Structure of astacin and implications for activation of astacins and zinc-ligation of collagenases., *Nature* **358**, 164-167.

49. Vogt, G., Stöcker, W., Storch, V., and Zwilling, R. (1989) Biosynthesis of *Astacus* protease, a digestive enzyme from crayfish., *Histochemistry.* **91**, 373-381.

50. Gomis-Ruth, F. X., Kress, L. F., and Bode, W. (1993) First structure of a snake venom metalloproteinase: a prototype for matrix metalloproteinases/collagenases., *EMBO J.* **12**, 4151-4157.

51. Zhang, D., Botos, I., Gomis-Ruth, F. X., Doll, R., Blood, C., Njoroge, F. G., Fox, J. W., Bode, W., and Meyer, E. F. (1994) Structural interaction of natural and synthetic inhibitors with the VENOM METALLOPROTEINASE, ATROLYSIN C (FORM-D). *Proc.Natl.Acad.Sci.USA* **91**, 8447-8451.

52. Kantrowitz, E. (2004) *E. coli* alkaline phosphatase., in *Handbook of Metalloproteins* (Messerschmidt, A., Bode, W., and Cygler, M., Eds.), pp 71-82, Wiley, Chichester.

53. Wilson, I. B., and Dayan, J. (1965) THE FREE ENERGY OF HYDROLYSIS OF PHOSPHORYL-PHOSPHATASE., *Biochemistry* **4**, 645-645.

54. Hough, E., Hansen, L. K., Birknes, B., Jynge, K., Hansen, S., Hordvik, A., Little, C., Dodson, E., and Derewenda, Z. (1989) High-resolution (1.5 A) crystal structure of phospholipase C from *Bacillus cereus.*, *Nature* **338**, 357-360.

55. Romier, C., Dominquez, R., Lahm, A., Dahl, O., and Suck, D. (1998) Recognition of single-stranded DNA by nuclease P1: high resolution crystal structures of complexes with substrate analogs., *Proteins* **32**, 414-424.

56. Shishido, K., and Habuka, N. (1986) Purification of S1 nuclease to homogeneity and its chemical, physical and catalytic properties., *Biochim Biophys Acta* **884**, 215-218.

57. Volbeda, A., Romier, C., and Suck, D. (2004) Nuclease P1., in *Handbook of Metalloproteins* (Messerschmidt, A., Bode, W., and Cygler, M., Eds.), pp 51-61, Wiley, Chichester.

58. Benning, M. M., Kuo, J. M., Raushel, F. M., and Holden, H. M. (1994) Three-Dimensional Structure of Phosphotriesterase: An Enzyme Capable of Detoxifying Organophosphate Nerve Agents., *Biochemistry 33*, 15001-15007.

59. Vanhooke, J. L., Benning, M. M., Raushel, F. M., and Holden, H. M. (1996) Three-dimensional structure of the zinc-containing phosphotriesterase with the bound substrate analog diethyl 4-methylbenzylphosphonate., *Biochemistry 35*, 6020-6025.

60. Omburo, G. A., Kuo, J. M., Mullins, L. S., and Raushel, F. M. (1992) Characterization of the zinc binding site of bacterial phosphotriesterase., *J Biol Chem. 267*, 13278-13283.

61. Bush, K. (1999) beta-Lactamases of increasing clinical importance., *Curr Pharm Des. 5*, 839-845.

62. Ambler, R. P. (1980) The structure of beta-lactamases., *Philos Trans R Soc Lond B Biol Sci. 289*, 321-331.

63. Galleni, M., Lamotte-Brasseur, J., Rossolini, G. M., Spencer, J., Dideberg, O., and Frère, J. M. (2001) Standard numbering scheme for class B beta-lactamases., *Antimicrob Agents Chemother. 45*, 660-663.

64. Herzberg, O., and Fitzgerald, P. M. D. (2004) Metallo-ß-Lactamases., in *Handbook of Metalloproteins* (Messerschmidt, A., Bode, W., and Cygler, M., Eds.), pp 217-234, Wiley, Chichester.

65. Concha, N. O., Rasmussen, B. A., Bush, K., and Herzberg, O. (1996) Crystal structure of the wide-spectrum binuclear zinc beta-lactamase from *Bacteroides fragilis.*, *Structure. 4*, 823-836.

66. Carfi, A., Pares, S., Duée, E., Galleni, M., Duez, C., Frère, J. M., and Dideberg, O. (1995) The 3-D structure of a zinc metallo-beta-lactamase from *Bacillus cereus* reveals a new type of protein fold., *EMBO J. 14*, 4914-4921.

67. Ullah, J. H., Walsh, T. R., Taylor, I. A., Emery, D. C., Verma, C. S., Gamblin, S. J., and Spencer, J. (1998) The crystal structure of the L1 metallo-beta-lactamase from *Stenotrophomonas maltophilia* at 1.7 A resolution., *J Mol Biol. 284*, 125-136.

68. Hewett-Emmet, D., and Tashian, R. E. (1996) Functional diversity, conservation, and convergence in the evolution of the alpha-, beta-, and gamma-carbonic anhydrase gene families., *Mol Phylogenet Evol. 5*, 50-77.

69. Duda, D., and McKenna, R. (2004) Carbonic anhydrases (α-class). in *Handbook of Metalloproteins* (Messerschmidt, A., Bode, W., and Cygler, M., Eds.), pp 249-263, Wiley, Chichester.

70. Christianson, D. W., and Fierke, C., A. (1996) Carbonic Anhydrase: Evolution of the Zinc Binding Site by Nature and by Design., *Acc. Chem. Res. 29*, 331-339.

71. Duda, D., Govindasamy, L., Agbandje-McKenna, M., Tu, C., Silverman, D. N., and McKenna, R. (2003) The refined atomic structure of carbonic anhydrase II at 1.05 A resolution: implications of chemical rescue of proton transfer., *Acta Crystallogr.,Sect.D 59*, 93-104.

72. Liang, J. Y., and Lipscomb, W. N. (1990) Binding of substrate CO_2 to the active site of human carbonic anhydrase II: a molecular dynamics study., *Proc Natl Acad Sci U S A. 87*, 3675-3679.

73. Merz, K. M. (1991) Carbon dioxide binding to human carbonic anhydrase II., *J Am Chem Soc. 113*, 406-411.

74. Alexander, R. S., Nair, S. K., and Christianson, D. W. (1991) Engineering the hydrophobic pocket of carbonic anhydrase II., *Biochemistry 30*, 11064-11072.

75. Cox, D. J., Hunt, J. A., Compher, K. M., Fierke, C, A, and Christianson, D. W. (2000) Structural Influence of Hydrophobic Core Residues on Metal Binding and Specificity in Carbonic Anhydrase II., *Biochemistry 39*, 13687-13694.

76. Jörnvall, H., Höög, J. O., Persson, B., and Parés, X. (2000) Pharmacogenetics of the alcohol dehydrogenase system., *Pharmacology 61*, 184-191.

77. Meijers, R., Morris, R. J., Adolph, H. W., Merli, A., Lamzin, V. S., and Cedergen-Zeppezauer, E. S. (2001) On the enzymatic activation of NADH., *J.Biol.Chem. 276*, 9316.

78. Meijers, R., and Cedergen-Zeppezauer, E. S. (2004) Zn-dependent medium-chain dehydrogenases/reductases., in *Handbook of Metalloproteins* (Messerschmidt, A., Bode, W., and Cygler, M., Eds.), pp 5-33, Wiley, Chichester.

79. Cedergen-Zeppezauer, E. S. (1983) Crystal-structure determination of reduced nicotinamide adenine dinucleotide complex with horse liver alcohol dehydrogenase maintained in its apo conformation by zinc-bound imidazole., *Biochemistry 22*, 5761-5772.

80. Northrop, D. B., and Cho, Y. K. (2000) Effects of High Pressure on Solvent Isotope Effects of Yeast Alcohol Dehydrogenase., *Biophys. J. 79*, 1621-1628.

81. Fersht, A. (1999) *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding.*, W. H. Freeman and Company, New York.

82. Chan, S. J., Keim, P., and Steiner, D. F. (1976) Cell-free synthesis of rat preproinsulins: characterization and partial amino acid sequence determination., *Proc Natl Acad Sci U S A. 73*, 1964-1968.

83. Smith, D. G. (2004) Insulin., in *Handbook of Metalloproteins* (Messerschmidt, A., Bode, W., and Cygler, M., Eds.), pp 367-377, Wiley, Chichester.

84. Goldman, J., and Carpenter, F. H. (1974) Zinc binding, circular dichroism, and equilibrium sedimentation studies on insulin (bovine) and several of its derivatives., *Biochemistry 13*, 4566-4574.

85. Adams, M. J., Blundell, T. L., Dodson, E. J., Dodson, G. G., Vijayan, M., Baker, E. N., Harding, M. M., Hodgkin, D. C., Rimmer, B., and Sheat, S. (1969) Structure of Rhombohedral 2 Zinc Insulin Crystals., *Nature 224*, 491-495.

86. Baker, E. N., Blundell, T. L., Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Hodgkin, D. M., Hubbard, R. E., Isaacs, N. W., Reynolds, C. D., Sakabe, K., Sakabe, N., and Vijayan, N. M. (1988) The structure of 2Zn pig insulin crystals at 1.5 A resolution., *Philos Trans R Soc Lond B Biol Sci. 319*, 369-456.

87. Kaarsholm, N. C., Chong Ko, H., and Dunn, M. F. (1989) Comparison of solution structural flexibility and zinc binding domains for insulin, proinsulin, and miniproinsulin., *Biochemistry 28*, 4427–4435.

88. Bentley, G., Dodson, E., Dodson, G., Hodgkin, D. C., and Mercola, D. (1976) Structure of insulin., *Nature 261*, 166-168.

89. Derewenda, U., Derewenda, Z., Dodson, E. J., Dodson, G. G., Reynolds, C. D., Smith, D. G., Sparks, C., and Swenson, D. (1989) Phenol stabilizes more helix in a new symmetrical zinc insulin hexamer., *Nature 338*, 594-596.

90. Kurtzhals, P., Kiehr, B., and Sørensen, A. R. (1995) The cobalt(III)-insulin hexamer is a prolonged-acting insulin prodrug., *J Pharm Sci. 84*, 1164-1168.

91. McCord, J. M., and Fridovich, I. (1988) Superoxide dismutase: the first twenty years (1968-1988). *Free Radic Biol Med. 5*, 363-369.

92. Peter, H., Löffler, G., and Petrides, P. (2006) *Biochemie und Pathobiochemie (Springer-Lehrbuch) (German Edition)*, Berlin: Springer.

93. Tainer, J. A., Getzoff, E. D., Beem, K. M., Richardson, J. S., and Richardson, D. C. (1982) Determination and analysis of the 2 A-structure of copper, zinc superoxide dismutase., *J Mol Biol. 160*, 181-217.

94. Tainer, J. A., Getzoff, E. D., Richardson, J. S., and Richardson, D. C. (1983) Structure and mechanism of copper, zinc superoxide dismutase., *Nature 306*, 284-287.

95. Miller, J., McLachlan, A. D., and Klug, A. (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus oocytes.*, *EMBO J. 4*, 1609-1614.

96. Klug, A., and Schwabe, J. W. (1995) Protein motifs 5. Zinc fingers., *FASEB J. 9*, 597-604.

97. Laity, J. H. (2004) Cys2His2 zinc finger proteins., in *Handbook of Metalloproteins* (Messerschmidt, A., Bode, W., and Cygler, M., Eds.), pp 307-323, Wiley, Chichester.

98. Becker, K. G., Nagle, J. W., Canning, R. D., Biddison, W. E., Ozato, K., and Drew, P. D. (1995) Rapid isolation and characterization of 118 novel C2H2-type zinc finger cDNAs expressed in human brain., *Hum Mol Genet. 4*, 685-691.

99. Krishna, S. S., Majumdar, I., and Grishin, N. V. (2003) SURVEY AND SUMMARY: Structural classification of zinc fingers., *Nucleic Acids Res. 31*, 532-550.

100. Dutnall, R. N., Neuhaus, D., and Rhodes, D. (1996) The solution structure of the first zinc finger domain of SWI5: a novel structural extension to a common fold., *Structure. 4*, 599-611.

101. Klein, D. J., Johnson, P. E., Zollars, E. S., De Guzman, R. N., and Summers, M. F. (2000) The NMR structure of the nucleocapsid protein from the mouse mammary tumor virus reveals unusual folding of the C-terminal zinc knuckle., *Biochemistry 39*, 1604-1612.

102. Barlow, P. N., Luisi, B., Milner, A., Elliott, M., and Everett, R. (1994) Structure of the C3HC4 domain by 1H-nuclear magnetic resonance spectroscopy. A new structural class of zinc-finger., *J Mol Biol. 237*, 201-211.

103. Dauter, Z., Wilson, K. S., Sieker, L. C., Moulis, J. M., and Meyer, J. (1996) Zinc- and iron-rubredoxins from *Clostridium pasteurianum* at atomic resolution: a high-precision model of a ZnS4 coordination unit in a protein., *Proc Natl Acad Sci U S A. 93*, 8836-8840.

104. Marmorstein, R., Carey, M., Ptashne, M., and Harrison, S. C. (1992) DNA recognition by GAL4: structure of a protein-DNA complex., *Nature 356*, 408-414.

105. De Guzman, R. N., Liu, H. Y., Martinez-Yamount, M., Dyson, H. J., and Wright, P. E. (2000) Solution structure of the TAZ2 (CH3) domain of the transcriptional adaptor protein CBP., *J Mol Biol. 303*, 243-253.

106. Tsutakawa, S. E., Jingami, H., and Morikawa, K. (1999) Recognition of a TG mismatch: the crystal structure of very short patch repair endonuclease in complex with a DNA duplex., *Cell 99*, 615-623.

107. Braun, W., Vasak, M., Robbins, A. H., Stout, C. D., Wagner, G., Kagi, J. H., and Wuthrich, K. (1992) Comparison of the NMR solution structure and the x-ray crystal structure of rat metallothionein-2., *Proc.Natl.Acad.Sci.USA 89*, 10124-10128.

108. Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A., and Wright, P. E. (1989) Three-dimensional solution structure of a single zinc finger DNA-binding domain., *Science 245*, 635-637.

109. Elrod-Erickson, M., Rould, M. A., Nekludova, L., and Pabo, C. O. (1996) Zif268 protein-DNA complex refined at 1.6 A: a model system for understanding zinc finger-DNA interactions., *Structure. 4*, 1171-1180.

110. Pavletich, N. P., and Pabo, C. O. (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A., *Science 252*, 809-817.

111. Fischer, J. A., Gininger, E., Maniatis, T., and Ptashne, M. (1988) GAL4 activates transcription in *Drosophila*., *Nature 332*, 853-856.

112. Kaegi, J. H. R., and Schaeffer, A. (1988) Biochemistry of metallothionein., *Biochemistry 27*, 8509-8515.

113. Zangger, K., and Armitage, I. M. (2004) Metallothioneins., in *Handbook of Metalloproteins* (Messerschmidt, A., Bode, W., and Cygler, M., Eds.), pp 353-364, Wiley, Chichester.

114. Otvos, J. D., and Armitage, I. M. (1980) Structure of the metal clusters in rabbit liver metallothionein., *Proc Natl Acad Sci U S A. 77*, 7094-7098.

Chapter 2

Sulfatases

2.1   Introduction in Sulfatases

Sulfur is the seventh most abundant element in the human body by weight and it is therefore essential for all living cells. About 0.5-1% of the cell dry weight in bacteria is sulfur, which is needed especially for the amino acids cysteine and methionine. Furthermore, sulfur is needed in a variety of enzyme cofactors, for example coenzyme A, biotin and thiamine and forms a part of iron-sulfur clusters where its role is essential in performing redox processes. Moreover sulfur is important in protein assembly and protein structure by forming disulfide bridges in peptide chains.[1]

Although inorganic sulfate is the preferred source of bacteria for growing, for example in soil environments, sulfur is predominantly present in organically bound form, as sulfate esters or sulfonates (> 90%).[2] For the mobilization of alternative sulfur sources, microorganisms induce the expression of a heterogeneous class of enzymes, called sulfatases.[1, 3] They catalyze the cleavage of a sulfate ester bond yielding the corresponding alcohol and inorganic sulfate.[4-6]

Whilst the majority of hydrolases, like lipases, esterases and proteases do not alter the stereochemistry of the substrate during catalysis, sulfate ester hydrolysis can occur - depending on the mechanism of the enzyme - either via a cleavage of the S-O bond or of the C-O bond of a *sec*-alkyl sulfate.[7] The corresponding alcohol shows then either retention or inversion of the configuration at its stereogenic carbon atom, respectively (Scheme 1).
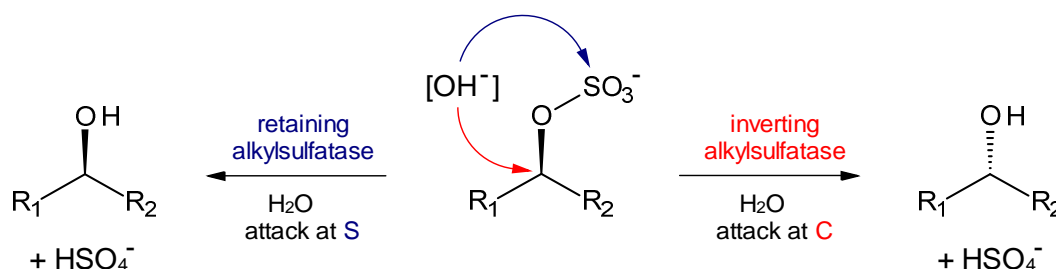

Scheme 1. Sulfate ester hydrolysis of sec-alkyl sulfates depending on the mechanism of the enzyme.

Based on mechanistic aspects, sulfatases can be divided[8] into three different subgroups: (i) aryl- and carbohydrate sulfatases which act on sulfated carbohydrates and steroids,[9] (ii)

Fe(II) sulfatases which belong to the α-ketoglutarate-dependent dioxygenase superfamily[10] and (iii) metallo-ß-lactamase related sulfatases.[8]

2.1.1   Aryl- and Carbohydrate Sulfatases

Members of this group share a highly conserved C/S-X-P-X-R-X_4-T-G amino acid consensus motif which encodes for a cysteine or serine residue in the active site. The oxidation of this cysteine (in prokaryotes and most eukaryotes) or this serine (in some bacterial sulfatases) generates a C_α-formylglycine hydrate residue which is essential for the catalytic activity of this group of sulfatases and generated by various formylglycine-generating enzymes. [4] The hydrated C_α-formylglycine nucleophilic attacks the sulfur atom of the sulfate ester substrate. The alcohol with a retained configuration at the chiral carbon atom is released, and a sulfated enzyme intermediate is formed. A subsequent elimination of the sulfate group regenerates the aldehyde (Figure 1).[9]
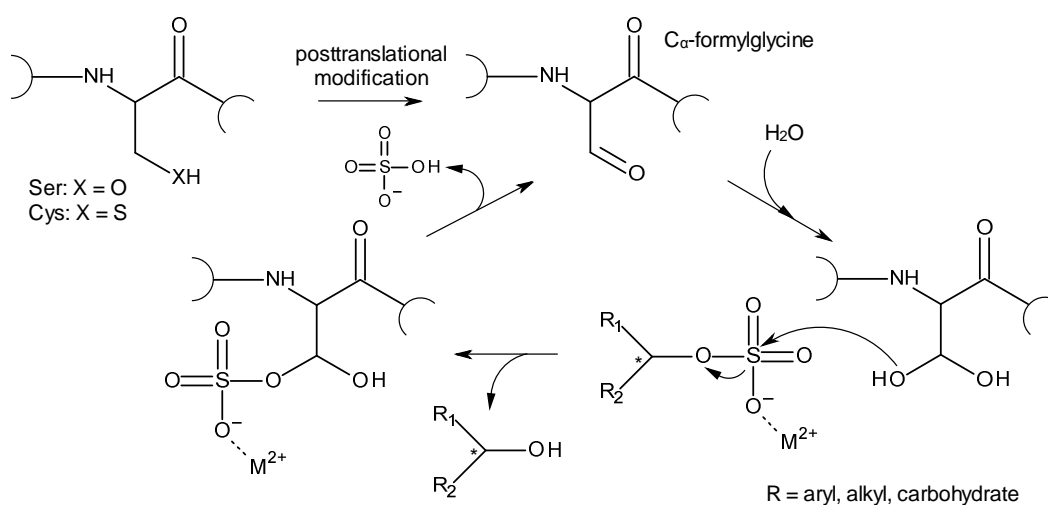
Figure 1. Catalytic mechanism of aryl- and carbohydrate sulfatases (adapted from reference 5).

2.1.2   α-Ketoglutarate-dependent Dioxygenases

Members of the α-ketoglutarate-dependent dioxygenase family catalyze the oxygenolytic cleavage of a variety of different alkyl sulfate esters into the corresponding aldehyde and inorganic sulfate using α-ketoglutarate as electron acceptor.[3] Two histidine residues and one carboxylate are binding the iron in this class of enzymes which is widespread in prokaryotes as well as in eukaryotes.[11] A proposed mechanism for this enzyme family was

described for the alkylsulfatase AtsK from *Pseudomonas putida* S-313 by Müller *et al.* which cleaves a variety of aliphatic alkyl sulfate esters to provide the strain with the sulfur required for growth under sulfate-limited conditions (Figure 2).[10]
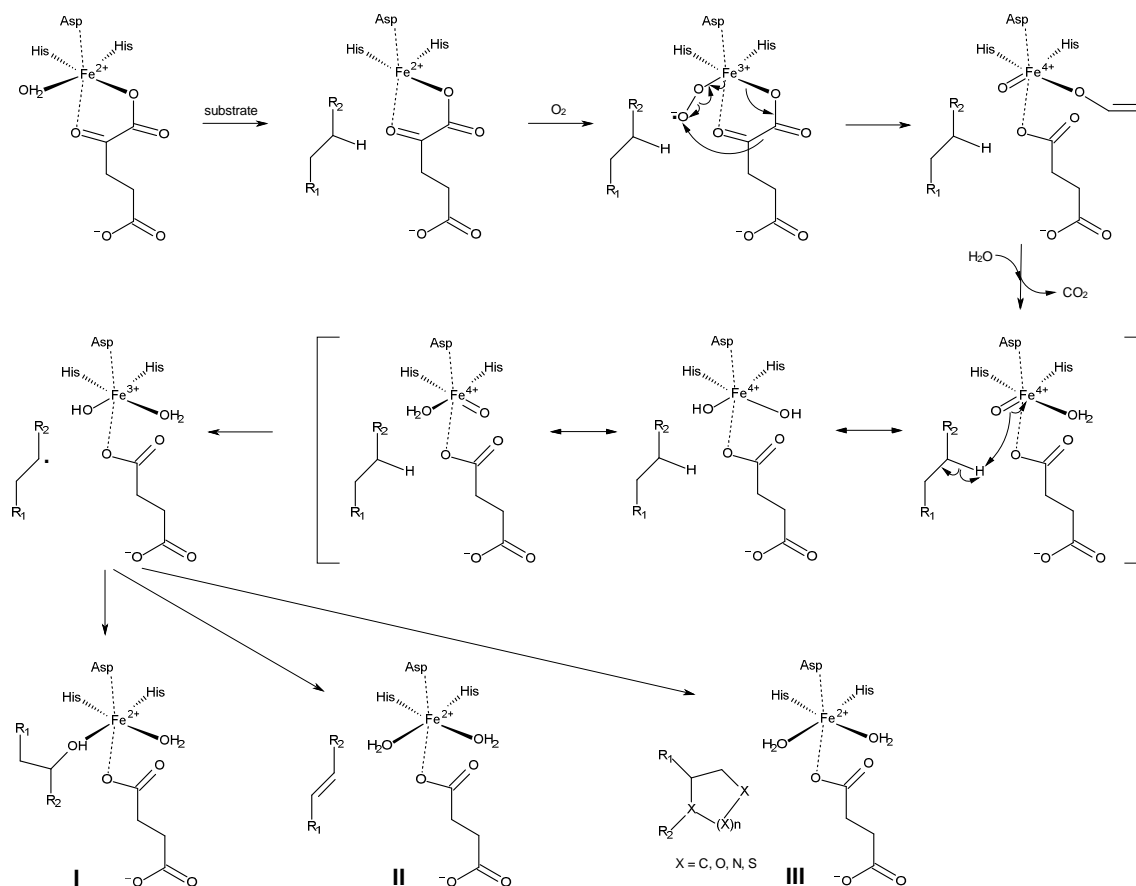
Figure 2. Proposed mechanism for enzymatic reactions catalyzed by α-ketoglutarate-dependent dioxygenases (adapted from reference 10).

In the first step a water molecule coordinated to the iron center is replaced by the sulfate ester substrate which enters the active site of the enzyme and may activate the enzyme for catalysis. In the second step, a dioxygen molecule coordinates to the iron center which induces the decarboxylation of the cosubstrate α-ketoglutarate, thereby forming succinate and carbon dioxide. At the same time a highly reactive ferryl intermediate Fe(IV)=O with the remaining oxygen atom is built and carbon dioxide is released from the iron coordination sphere. Concomitantly, a water molecule fills the sixth coordination site of the iron again, or succinate coordinates in a bidenate manner. A rearrangement of the iron ligands occurs to bring one oxygen atom close to the substrate to facilitate the highly positively charged ferryl intermediate to abstract a hydrogen atom from the sulfate ester.

In that step Fe(IV) is reduced to Fe(III) and a radical species of the substrate is formed. A hydroxyl transfer between the radical species and the iron complex hydroxylates the substrate (I) which is the most common reaction in this class of enzymes. In some cases another hydrogen atom is subtracted from the substrate radical forming the desaturated substrate species (II) or it leads to a cyclization or ring expansion (III). In the last step the Fe(II) center is regenerated by oxidation of the substrate, succinate is replaced by a new cosubstrate and the enzyme is ready for the next turnover.[10]

Reactions catalyzed by sulfatases belonging to α-ketoglutarate-dependent dioxygenases destruct the chiral center at the carbon atom; hence this class of enzymes is of limited interest for a stereoselective biotransformation of sulfate esters.[5]

2.1.3   Metallo-β-lactamase related Sulfatases

SdsA1: The first member of metallo-β-lactamase related sulfatases was identified in 2006 by a group in Germany.[8] It was isolated from the soil and water bacterium *Pseudomonas aeruginosa*. The enzyme is able to hydrolyze biocidic sodium dodecyl sulfate (SDS) - a detergent of most commercial hygiene products - into insoluble 1-dodecanol and soluble sulfate and was therefore named SdsA1. Hydrolysis of SDS makes *P. aeruginosa* able to grow under 0.1% SDS as sole carbon and sulfur source. Furthermore, SdsA1 accepts medium- and short chain alkyl sulfates as substrates, for example decyl, octyl and hexyl sulfate.
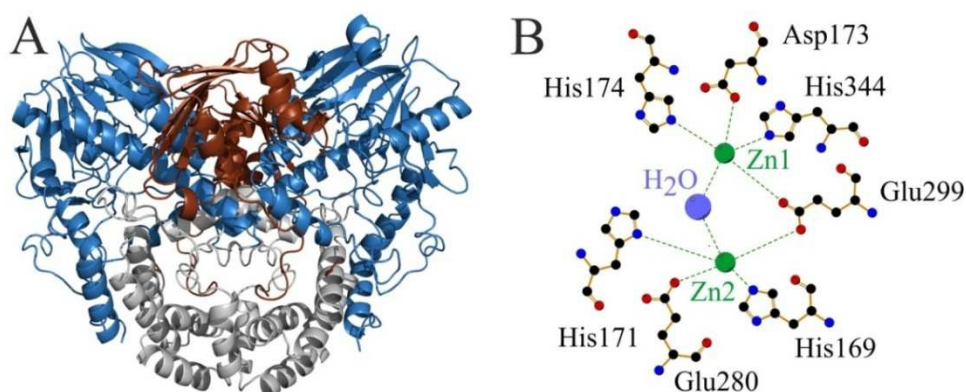
Figure 3. Structure representation of SdsA1 (PDB code: 2CGC) from *Pseudomonas aeruginosa*. (A) Cartoon representation showing the N-terminal catalytic domain in blue, the α-helical dimerization domain in grey and the C-terminal domain in brown. Picture was prepared with the program PyMol.[12] (B) Close up of the zinc binding site, prepared with LigPlot.[13]

It features a Zn^{2+} binding motif **THxHxDHxGG**-102-**E**-18-**AE**-44-**H** and was classified as a member of the metallo-β-lactamase family.[14] Figure 3 displays the overall structure of the symmetric dimeric enzyme and the Zn binding site in SdsA1 (PDB code: 2CG2).[8]

Each monomer is divided into three domains, an N-terminal αββα-sandwich domain which performs the catalytic activity of SdsA1 and harbors the zinc binding motif (shown in blue). This domain is typical for metallo-β-lactamases. The second domain is exclusively α-helical (shown in grey) and is responsible for the interlock of the two monomers. It ensures resistance to high concentrations of SDS. The brown-colored part of the enzyme shows the third, C-terminal domain of SdsA1. It consists of a five-stranded mixed β-sheet covered by short α-helices on one side.

In the active site of SdsA1 two zinc ions are present at a distance of 3.3 Å. They are bridged by a water molecule and a glutamic acid (Glu299). Zn1 shows distorted trigonal bipyramidal coordination sphere and is further coordinated to two histidine residues (His174 and His344) and aspartic acid (Asp173). In contrast to Zn1, Zn2 shows tetrahedral coordination and is additionally ligated with His169, His171 and Glu280.

As mentioned before, SdsA1 cleaves short and medium chain alkyl sulfates, however only primary alkyl sulfates have been tested; hence the stereo selectivity of the enzyme was not investigated in detail.

Pisa1: The first alkylsulfatase specific for the hydrolysis of *sec*-alkyl sulfates could be isolated from the *Pseudomonas* sp. DSM 6611 strain and was termed Pisa1 (*Pseudomonas inverting sec*-alkylsulfatase one).[15]

Interestingly, the enzyme shows a high overall structural similarity with SdsA1, the *prim*-alkyl sulfatases from *Pseudomonas aeruginosa* (Figure 3A),[8] including an N-terminal catalytic αββα-sandwich domain, typical for metallo-ß-lactamases, an exclusively α-helical dimerization domain and a C-terminal mixed domain (Figure 4A).[16]

According to SdsA1, Pisa1 features a binuclear zinc-cluster (Figure 4B), whereby the two zinc-ions are coordinated by histidine (His184 and His355 for Zn1 and His179 and His181 for Zn2), aspartic acid (Asp183 for Zn1 and Asp310 for Zn1 and Zn2) and glutamic acid (Glu291 for Zn1) residues.
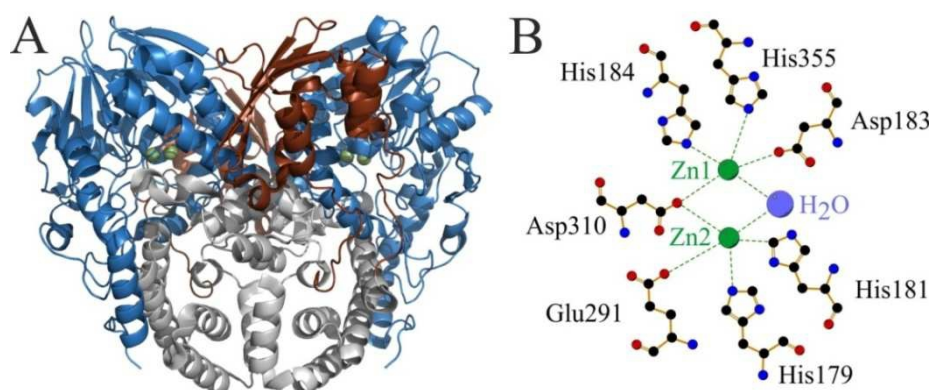
Figure 4. Structure representation of Pisa1 (PDB code: 2YHE) from *Pseudomonas* sp. DSM6611. (A) Cartoon representation showing the N-terminal catalytic domain in blue, the α-helical dimerization domain in grey and the C-terminal domain in brown. Picture was prepared with the program PyMol.[12] (B) Close up of the binuclear zinc binding site, prepared with LigPlot.[13]

Pisa1 hydrolyzes preferentially the *(R)*-form of a variety of *rac-sec*-alkyl sulfate esters, yielding a strictly homochiral mixture of *(S)*-configurated alcohols and non-reacted *(S)*-sulfate esters. That means that the enzyme works under strict inversion of configuration at the chiral carbon atom, a catalytic event which does not have an analog in chemo catalysis, giving access to valuable enantiopure building blocks for pharmaceutical and other industrial applications.[17]

2.2   References

1. Kertesz, M. A. (2000) Riding the sulfur cycle - metabolism of sulfonates and sulfate esters in Gram-negative bacteria, *FEMS Microbiol Rev. 24*, 135-175.

2. Scherer, H. W. (2001) Sulphur in crop production, *Eur J Agron 14*, 81-111.

3. Kahnert, A., and Kertesz, M. A. (2000) Characterization of a sulfur-regulated oxygenative alkylsulfatase from *Pseudomonas putida* S-313, *J Biol Chem. 275*, 31661-31667.

4. Bojarova, P., and Williams, S. J. (2008) Sulfotransferases, sulfatases and formylglycine-generating enzymes: a sulfation fascination, *Curr Opin Chem Biol. 12*, 573-581.

5. Gadler, P., and Faber, K. (2007) New enzymes for biotransformations: microbial alkyl sulfatases displaying stereo- and enantioselectivity, *Trends Biotechnol. 25*, 83-88.

6. Hanson, S. R., Best, M. D., and Wong, C. H. (2004) Sulfatases: Structure, mechanism, biological activity, inhibition, and synthetic utility, *Angew Chem Int Ed Engl. 43*, 5736-5763.

7. Faber, K., and Kroutil, W. (2005) New enzymes for biotransformations, *Curr Opin Chem Biol. 9*, 181-187.

8. Hagelueken, G., Adams, T. M., Wiehlmann, L., Widow, L., Kolmar, H., Tummler, B., Heinz, D. W., and Schubert, W. D. (2006) The crystal structure of SdsA1, an alkylsulfatase from *Pseudomonas aeruginosa*, defines a third class of sulfatases, *Proc Natl Acad Sci U S A. 103*, 7631-7636.

9. Boltes, I., Czapinska, H., Kahnert, A., von Bulow, R., Dierks, T., Schmidt, B., von Figura, K., Kertesz, M. A., and Uson, I. (2001) 1.3 angstrom structure of arylsulfatase from *Pseudomonas aeruginosa* establishes the catalytic mechanism of sulfate ester cleavage in the sulfatase family, *Structure 9*, 483-491.

10. Mueller, I., Kahnert, A., Pape, T., Sheldrick, G. M., Meyer-Klaucke, W., Dierks, T., Kertesz, M., and Uson, I. (2004) Crystal structure of the alkylsulfatase AtsK: Insights into the catalytic mechanism of the Fe(II) alpha-ketoglutarate-dependent dioxygenase superfamily, *Biochemistry 43*, 3075-3088.

11. Ryle, M. J., and Hausinger, R. P. (2002) Non-heme iron oxygenases., *Curr Opin Chem Biol. 6*, 193-201.

12. DeLano, W. L. (2002) The PyMOL Molecular Graphics System *http://www.pymol.org.*, (DeLano Scientific, S. C., CA, USA., Ed.).

13. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions., *Prot Engin. 8*, 127-134.

14. Daiyasu, H., Osaka, K., Ishino, Y., and Toh, H. (2001) Expansion of the zinc metallo-hydrolase family of the beta-lactamase fold, *FEBS Lett. 503*, 1-6.

15. Schober, M., Gadler, P., Knaus, T., Kayer, H., Birner-Gruenberger, R., Guelly, C., Macheroux, P., Wagner, U., and Faber, K. (2011) A Stereoselective Inverting sec-Alkylsulfatase for the Deracemization of sec-Alcohols, *Org Lett. 13*, 4296-4299.

16. Knaus, T., Schober, M., Kepplinger, B., Faccinelli, M., Pitzer, J., Faber, K., Macheroux, P., and Wagner, U. (2012) Structure and mechanism of an inverting alkylsulfatase from *Pseudomonas sp.* DSM6611 specific for secondary alkylsulfates, *FEBS Journal. **under revision***.

17. Schober, M., Knaus, T., Toesch, M., Macheroux, P., Wagner, U., and Faber, K. (2012) The Substrate Spectrum of the Inverting sec-Alkylsulfatase Pisa1, *Adv Synth Cat. 354*, 1737-1742.

Chapter 3

Aims of the project Pisa1 from *Pseudomonas* sp. DSM6611

Aims of the project Pisa1 from *Pseudomonas* sp. DSM6611

SdsA1 is a member of the alkylsulfatase family, which was isolated from the waste water organism *Pseudomonas aeruginosa.* This enzyme catalyzes the hydrolysis of the sulphate group belonging to various primary medium- and short-chain alkyl sulfates as substrates. Since SdsA1 hydrolyzes principally primary alkyl sulfates, the enzyme was not considered of interest for biocatalytic applications. However, a more accurate study regarding the mechanism of the hydrolysis using deuterated water revealed that SdsA1 cleaves the sulfate moiety in a perfect stereoselective fashion. More astonishing was the observation, that the only product obtained was the alcohol possessing a deutered hydroxyl group. To the best of our knowledge, a similar mechanism of sulfate hydrolysis does not occur in chemo-catalysis.

More recently, another alkylsulfatase (Pisa1) was identified from *Pseudomonas* sp. DSM6611. Molecular modeling of Pisa1 indicated that the novel enzyme might possess high structural similarity with SdsA1. Interestingly Pisa1 was capable to hydrolyze secondary rather than primary alkyl sulfates, although the stereochemistry of the reaction was still the same as for SdsA1. The possibility to hydrolyze secondary alkyl sulfates would allow employing Pisa1 in biocatalysis for the production of chiral secondary alcohols, which are key intermediates for the synthesis of pharmaceuticals and agrochemicals.

The aim of the project was to elucidate the reaction mechanism for the hydrolysis catalyzed by Pisa1. We were particularly interested to identify the amino acid residues responsible for the nucleophilic attack and the binding mode of the substrates into the active site of the enzyme. Structural and computational studies revealed which might have been the amino acid residues fundamental for the catalysis. Various variants were designed, prepared and evaluated during catalysis to compare their catalytic properties with our assumptions based on computational studies.

Chapter 4

A stereoselective inverting *sec*-alkylsulfatase for the deracemization of *sec*-alcohols

AUTHOR CONTRIBUTIONS

The manuscript has been published on ORGANIC LETTERS (2011), VOL. 13, NO. 16, PAGES 4296-4299. The research was carried out in cooperation with the Department of Chemistry, Organic & Bioorganic Chemistry (University of Graz, Austria), the Center for Medical Research (Medical University of Graz, Austria) and the Institute of Molecular Biosciences, Structural Biology (University of Graz, Austria). The main work was part of the PhD thesis of MARKUS SCHOBER, Department of Chemistry, Organic & Bioorganic Chemistry (University of Graz, Austria), who performed most of the experiments. My contribution consisted in the expression and purification of the proteins and the determination of the kinetic parameters for the conversion of alkyl sulfate esters by Pisa1 and SdsA1.

A Stereoselective Inverting *sec*-Alkylsulfatase for the Deracemisation of *sec*-Alcohols

Markus Schober[†], Petra Gadler[†,^], Tanja Knaus[‡], Heidemarie Kayer[‡], Ruth Birner-Grünberger[§], Christian Gülly[§], Peter Macheroux[‡], Ulrike Wagner[#] and Kurt Faber[†]*

Department of Chemistry, Organic & Bioorganic Chemistry, University of Graz, Heinrichstrasse 28, A-8010 Graz, Austria; Institute of Biochemistry, Graz University of Technology, Petersgasse 12, A-8010 Graz, Austria; Center for Medical Research, Medical University of Graz, Stiftingtalstrasse 24, A-8010 Graz, Austria; Institute of Molecular Biosciences, Structural Biology, University of Graz, Humboldtstrasse 50/III, A-8010 Graz, Austria

kurt.faber@uni-graz.at

ABSTRACT
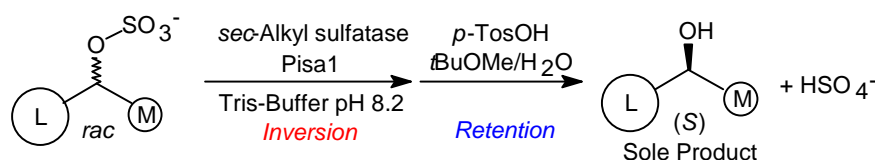
A metallo-β-lactamase-type alkylsulfatase was found to catalyse the enantioselective hydrolysis of *sec*-alkylsulfates with strict inversion of configuration. This catalytic event, which does not have an analog in chemo-catalysis, yields homochiral (*S*)-configurated alcohols and nonreacted sulfate esters. The latter could be converted into (*S*)-*sec*-alcohols as the sole product in up to >99% e.e via a chemo-enzymatic deracemization protocol on a preparative scale.

Sulfatases are a heterogenic group of enzymes which catalyse the cleavage of the sulfate ester bond yielding the corresponding alcohol and hydrogen sulfate.[1,2,3,4] In contrast to the majority of hydrolases,[5] which do not alter the stereochemistry of the substrate during catalysis, the stereochemical course of sulfate ester hydrolysis may proceed via cleavage of the S-O or the C-O bond of a *sec*-alkyl sulfate going in hand with retention or inversion of the stereogenic carbon atom, respectively (Scheme 1).[1] Recently, sulfatases were re-classified based on

their mode of catalysis:[6] Aryl- and carbohydrate-sulfatases act on sulfated carbohydrates and steroids.[4] These enzymes possess a highly conserved -C/S-X-P-X-A-X$_4$-T-G- consensus motif,[2] which codes for a Cys or Ser residue within the active site. The latter is posttranslationally modified into a hydrated α-formylglycine moiety, which attacks the S-atom of the sulfate ester going in hand with S-O bond cleavage while the absolute configuration at C is retained.[7] On the other hand, sulfatases belonging to the Fe^{2+}-dependent group of dioxygenases oxidatively cleave a sulfate ester at the expense of α-ketoglutarate as electron acceptor to yield an aldehyde and inorganic sulfate, leading

[†] University of Graz – Department of Chemistry
[‡] Graz University of Technology
[§] Medical University of Graz
[#] University of Graz – Institute of Molecular Biosciences
[^] Present Adress: Département de Chimie, Université de Montréal, 2900, Boulevard Édouard-Montpetit, Canada

(1) Gadler, P.; Faber, K. *Trends Biotechnol.* **2006**, *25*, 83-88.
(2) Kertesz, M. A. *FEMS Microbiol. Rev.* **2000**, *24*, 135-175.
(3) Bojarova, P.; Williams, S. J. *Curr. Opin. Chem. Biol.* **2008**, *12*, 573-581.
(4) Hanson, S. R.; Best, M. D.; Wong, C.-H. *Angew. Chem. Int. Ed.* **2004**, *43*, 5736-5763.
(5) Bornscheuer, U. T.; Kazlauskas, R. J. *Hydrolases in Organic Synthesis,* Wiley-VCH, Weinheim, **2005**.

(6) Hagelueken, G.; Adams, T. M.; Wiehlmann, L.; Widow, U.; Kolmar, H.; Tümmler, B.; Heinz, D. W.; Schubert, W.-D. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 7631-7636.
(7) Boltes, I; Czapinska, H.; Kahnert, A.; von Bülow, R.; Dierks, T.; Schmidt, B.; von Figura, K.; Kertesz, M. A.; Uson, I. *Structure* **2001**, *9*, 483-491.

Scheme 1. Stereochemical course of retaining and inverting alkylsulfatases
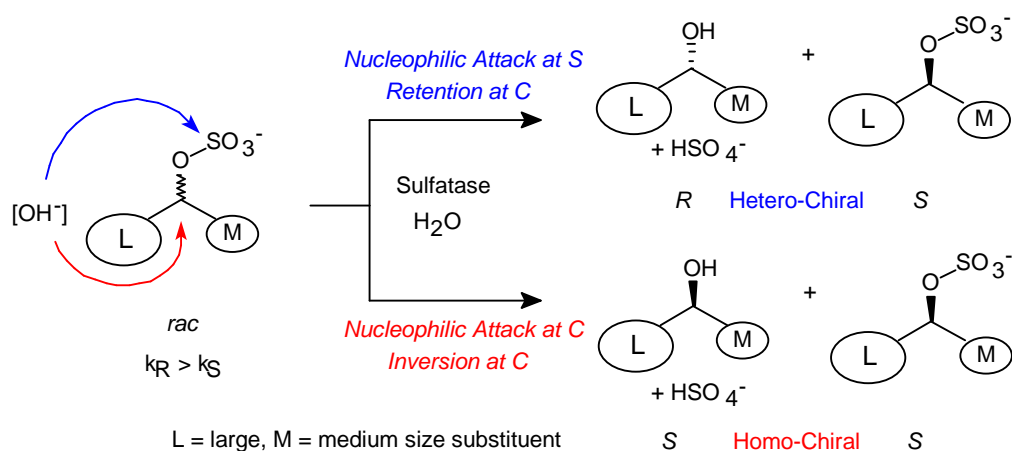
to destruction of the stereogenic centre at C.[8] The third class of sulfatases is related to metallo-β-lactamases and (to date) harbours only a single member: Sodium dodecyl sulfatase (SdsA1).[6] Since SDS is an achiral *prim*-sulfate ester, the stereochemical course of SdsA1-hydrolysis is not 'visible'. However, since the nucleophilic water molecule (W2 in the crystal structure 2cfu) strongly interacts with the sulfur atom of the substrate surrogate inhibitor 1-decanesulfonate, retention at carbon was assumed.[6] Inverting alkylsulfatases[9] were previously studied in *Pseudomonas* C12B,[10] *Comamonas terrigena*[11] and *Rhodococcus ruber* DSM 44541[12] but due to the lack of biochemical and structural data the mechanism of inverting alkylsulfatases remained unknown. In search for a well-characterised stable *sec*-alkylsulfatase, which would allow to design a deracemisation process for *sec*-alcohols via enantio-convergent hydrolysis of the corresponding sulfate esters, an extended whole-cell screening for *sec*-alkylsulfatase activity was

conducted, which revealed *Pseudomonas* sp. DSM 6611 as the most promising candidate.[13] Chromatographic protein purification followed by tryptic digestion and peptide mass fingerprinting allowed to assign the obtained peptide masses to the predicted open reading frames using the full genomic sequence of the strain.[14] The protein band with the lowest mobility on SDS-PAGE (termed Pisa1, identified by *de novo* sequencing and peptide matching) surprisingly turned out to be a homologue of SdsA1.[15] The Pisa1 gene was amplified by PCR, cloned and expressed in an *E. coli* BL21 strain with a C-terminal hexa-histidine tag. Pisa1 displayed the desired catalytic properties: (*R*)-2-Octyl sulfate (>99% ee) was quantitatively hydrolysed to yield (*S*)-2-octanol (>99% ee) through strict inversion of configuration, whereas the (*S*)-enantiomer was completely unreactive. Hydrolysis of *rac*-2-octyl sulfate ceased at 50% conversion to furnish a homochiral product mixture of (*S*)-2-octanol and unreacted (*S*)-2-octyl sulfate, indicating perfect enantioselectivity (E >200).

(8) Müller, I.; Kahnert, A.; Pape, T.; Sheldrick, G. M.; Meyer-Klaucke, W.; Dierks, T.; Kertesz, M. A.; Uson, I. *Biochemistry* **2004**, *43*, 3075-3088.

(9) Bartholomew, B.; Dodgson, K. S.; Matcham, G. W. J.; Shaw, D. J.; White, G. F. *Biochem. J.* **1977**, *165*, 575-580.

(10) Shaw, D. J.; Dodgson, K. S.; White, G. F. *Biochem. J.* **1980**, *187*, 181-196; White, G. F. *Appl. Microbiol. Biotechnol.* **1991**, *35*, 312-316.

(11) Fitzgerald, J. W.; Dodgson, K. S.; Matcham, G. W. J. *Biochem. J.* **1975**, *149*, 477-480; Matcham, G. W. J.; Dodgson, K. S.; Fitzgerald, J. W. *Biochem. J.* **1977**, *167*, 723-729; Barrett, C. H.; Dodgson, K. S.; White, G. F. *Biochem. J.* **1980**, *191*, 467-473.

(12) Pogorevc, M.; Faber, K. *Appl. Environ. Microbiol.* **2003**, *69*, 2810-2815.

(13) Gadler, P.; Faber, K. *Eur. J. Org. Chem.* **2007**, 5527-5530.

(14) Angelova, M.; Gadler, P.; Kayer, H.; Krempl, P.; Gülly, C.; Faber, K.; Macheroux P.; Thallinger, G. manuscript in preparation.

(15) The sequence was submitted to the European Nucleotide Archive (accession #: FR850678 – confidential until 01.07.2011). No measurable sequence-similarities of Pisa1 with other sulfatases, such as the α-ketoglutarate-dependent alkylsulfatase AtsK from *Pseudomonas putida*, the arylsulfatase PAS from *P. aeruginosa* and the human steroid sulfatases HArsA, N-acetylgalactosamin 4-sulfatase or cerebroside-3-sulfate sulfatase, could be detected (≤10%).

Although Pisa1 shares a 44% sequence identity with SdsA1 the substrate preference of both proteins differs significantly: Based on k_{cat} and K_M-values, SdsA1 has a 150-fold affinity for the *prim*-sulfate ester **1a**, whereas Pisa1 has a pronounced (190-fold) opposite preference for the *sec*-sulfate ester analog **2a**. Hence, Pisa1 is the first inverting *sec*-alkylsulfatase that is characterized on a molecular level.[15,21]

The stereochemical course of hydrolysis for both enzymes was investigated in detail using unlabelled 1-octyl- (**1a**) and *rac*-2-octyl sulfate (**2a**) in ^{18}O-enriched buffer (label >98%, Scheme 2, part A). GC-MS analysis of the formed 1-octanol (**1b**) and (*S*)-2-octanol (**2b**, ee >99%) showed complete incorporation of the ^{18}O-label in the product within analytical limits, proving C-O bond cleavage in both cases.

Scheme 2. Proof of inversion for Pisa1 and SdsA1 via ^{18}O-labeling (A), substrate spectrum of Pisa1 (B) and deracemisation of *rac*-**2a** (C).
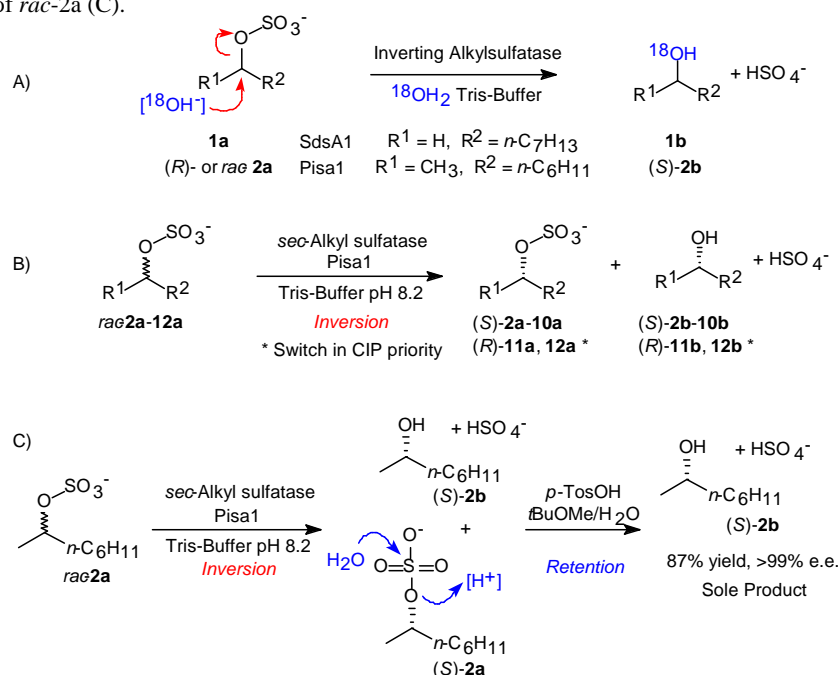
The nucleophilic attack of (formal) [OH[-]] at the chiral carbon atom of an alkyl sulfate ester as exerted by alkylsulfatases SdsA1 and Pisa1 leading to inversion of configuration at C is a remarkable catalytic event, which does not have a direct counterpart in chemical catalysis. The latter event would generate SO_4^{2-}, which is a very poor leaving group.[16] Hence, inverting nucleophilic hydrolysis of sulfate esters by hydroxide, acetate or methoxide proceeds extremely slowly and is not feasible for preparative purposes.[17,18] In order to facilitate the departure of the sulfate moiety, it has to be converted into a good leaving group, i.e. HSO_4^-

rather than SO_4^{2-}.[19] Consequently, acid-catalysed *sec*-sulfate ester hydrolysis is a fast process, which proceeds through retention at C.[20,17] Based on the crystal structure of Pisa1,[21] an acid-base-type mechanism for the inverting enzymatic alkyl sulfate hydrolysis can be proposed (Fig. 1).

(16) Being the anion of the weak acid HSO_4^- (pKa = +1.9 to +2.7).

(17) Batts, B. D. *J. Chem. Soc. (B)* **1966**, 551-555; Burwell Jr., R. L. *J. Am. Chem. Soc.* **1952**, *74*, 1462-1466.

(18) Wallner, S. R.; Nestl, B.; Faber, K. *Tetrahedron* **2005**, *61*, 1517-1521.

(19) Based on the pKa of methyl monosulfate, which was calculated/estimated as pKa -8.4 or pKa -3.4, thus protonation of a sulfate ester requires a strong acid of pKa ca. <2; see: a) Klages, F.; Jung, H. A.; Hegenberg, P. *Chem. Ber.* **1966**, *99*, 1704-1711; b) Fulford, J. E.; Dupuis, J. W.; March, R. E. *Can. J. Chem.* **1978**, *56*, 2324-2330.

(20) Depending on the conditions, acid-catalysed sulfate ester hydrolysis may be plagued by racemisation,[17b] which can be suppressed in presence of 1,4-dioxane as mediator, see: Batts, B. D. *J. Chem. Soc. (B)* **1966**, 547-551.

(21) The coordinates of the crystal structure of Pisa1 (2.7 Å resolution) are available from the PDB (code 2YHE); A detailed study on the structural and mechanistic aspects of Pisa1 will be published: Knaus, T.; Schober, M.; Kepplinger, B.; Gadler, P.; Macheroux, P.; Faber, K.; Wagner, U. G. manuscript in preparation.

(i) The anionic substrate is positioned by a sulfate binding site consisting of positively charged (Arg328/323) and H-bonding residues (Asn318, Thr321); (ii) a highly conserved 'nucleophile site' consisting of a binuclear Zn^{2+}-binding cluster (Asp310/183, His179/181, Glu291) activates a water molecule to provide the nucleophile [OH⁻], which attacks the carbon atom bearing the sulfate ester moiety.

The latter is facilitated through (iii) concomitant protonation of the liberated inorganic sulfate (presumably by Tyr417), to yield HSO_4^- as a good leaving group. The hydrophobic binding sites of Pisa1 and SdsA1 differ significantly with respect to the relative size of residues, which most likely accounts for the opposite substrate specificity concerning *prim*- versus *sec*-alkyl sulfate esters.
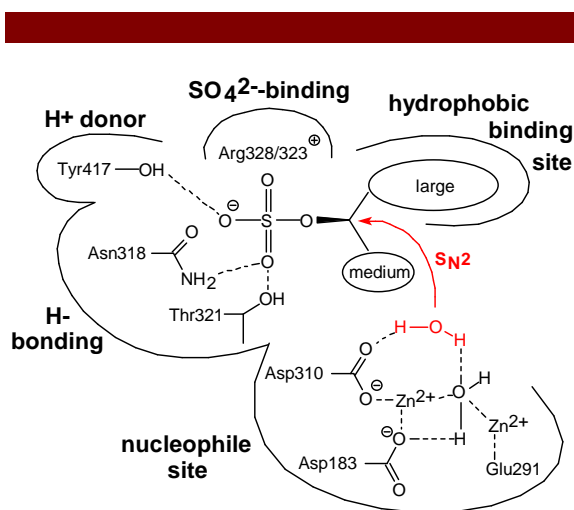
Figure 1. Schematic proposal for an acid-base mechanism for alkyl sulfate ester hydrolysis catalysed by Pisa1 (His179/181/184/355 coordinating the Zn^{2+} ions were omitted for clarity).

The substrate spectrum of Pisa1 proved to encompass a range of linear, branched or cyclic ω-1 to ω-3 *sec*-alkyl sulfates (Table 1, Scheme 2, part B). Straight-chain (ω-1)- and (ω-2)-sulfate esters (**2a-5a**) were resolved with perfect enantioselectivity, the (ω-3)-analog **6a** bearing two C_3/C_4-chains of similar size gave a respectable E-value of E = 10. Substrates bearing branched (**7a**), aromatic (**8a, 9a**) or cyclic side chains (**10a**) gave excellent results.

In contrast to the previously employed resting whole cell preparation of *Pseudomonas* sp. DSM 6611[13] the reaction rates were far superior and conversion values usually reached the theoretical limit of 50%, which is required for a deracemisation process. Most remarkably, substrates *rac*-**11a** and *rac*-**12a**, derived from propargylic alcohols bearing a synthetically useful terminal acetylene unit, were transformed with the same perfect stereoselectivity.

With the highly active and stereoselective inverting *sec*-alkylsulfatase Pisa1 in hand, the feasibility of a deracemisation protocol for *sec*-alcohols could be demonstrated as follows (Scheme 2, part C): Treatment of *rac*-2-octyl sulfate (**2a**, 1g) with Pisa1 in aqueous buffer gave equimolar amounts of (*S*)-**2b** and (*S*)-**2a**. The latter was hydrolyzed under acidic conditions (*p*-TosOH in *t*BuOMe/H₂O/1,4-dioxane)[18] to yield enantiopure (*S*)-**2b** (>99 % e.e) in 87 % isolated yield from the racemate (0.49 g).

Table 1. Substrate spectrum of Pisa1 compared to results with whole cells of *Pseudomonas* sp. DSM 6611.

Substrate	R^1	R^2	*Ps*. sp. DSM 6611[a]		Pisa1	
			c.	E[b]	c.	E[b]
rac-**2a**	CH_3	n-C_6H_{13}	21	>200	50	>200
rac-**3a**	CH_3	n-C_5H_{11}	17	>200	50	>200
rac-**4a**	CH_3	n-C_7H_{15}	7	>200	50	>200
rac-**5a**	C_2H_5	n-C_5H_{11}	18	>200	50	>200
rac-**6a**	n-	n-C_4H_9	20	6	57	10
rac-**7a**	CH_3	$(CH_2)_2CH=CMe_2$	9	>200	50	>200
rac-**8a**	CH_3	CH_2Ph	<1	n.d.	30	>200
rac-**9a**	CH_3	$(CH_2)_2Ph$	15	>200	50	>200
rac-**10a**	CH_3	c-C_6H_{11}	5	>200	10	>200
rac-**11a**	$HC{\equiv}C$	n-C_4H_9	n. i.	—	50	>200
rac-**12a**	$HC{\equiv}C$	n-C_5H_{11}	n. i.	—	50	>200

[a] Whole resting cells of *Pseudomonas* sp. DSM 6611, data from ref. [13]; [b] Enantioselectivity expressed as Enantiomeric Ratio (*E*-value); n. d. not determined due to exceedingly low conversion; n. i. = not investigated.

Supporting Information Available: Synthesis of sulfate esters, NMR-spectra, cloning, expression and purification of Pisa1 and SdsA1, H₂¹⁸O labeling experiments, determination of enantiomeric excess and absolute configuration, and preparative scale deracemization. This material is available free of charge via the Internet at http://pubs.acs.org.

Chapter 5

Supplementary Information
A stereoselective inverting *sec*-alkylsulfatase for the deracemization of *sec*-alcohols

A Stereoselective Inverting *sec*-Alkylsulfatase for the Deracemisation of *sec*-Alcohols

Markus Schober[1], Petra Gadler[1], Tanja Knaus[2], Heidemarie Kayer[2], Ruth Birner-Grünberger[3], Christian Gülly[3], Peter Macheroux[2], Ulrike Wagner[4] and Kurt Faber[1]*[1]

[1]Department of Chemistry, Organic & Bioorganic Chemistry, University of Graz, Heinrichstrasse 28, A-8010 Graz, Austria;

[2]Institute of Biochemistry, Graz University of Technology, Petersgasse 12, A-8010 Graz,

[3]Austria; Center for Medical Research, Medical University of Graz, Stiftingtalstrasse 24, A-8010 Graz, Austria;

[4]Institute of Molecular Biosciences, Structural Biology, University of Graz, Humboldtstrasse 50/III, A-8010 Graz, Austria

Electronic Supporting Information

* to whom correspondence should be addressed: Tel.: +43-316-380-5332; fax: +43-316-380-9840; email: Kurt.Faber@Uni-Graz.at

General

Competent cells One Shot® TOP10 and One Shot® BL21 Star™ (DE3) were purchased from Invitrogen and transformed according to the manufacturer´s protocol. 1-Octyl sulfate (**1a**), 1-octanol (**1b**) and alcohols **2b-10b** were purchased from Sigma Aldrich and Alfa Aesar. $H_2^{18}O$ for the preparation of ^{18}O labelled buffer was purchased from Rotem (label >98%). NMR spectra were recorded on a Bruker spectrometer at 300 (^1H) and 75 (^{13}C) MHz. Shifts ( ) are given in ppm and coupling constants (*J*) are given in Hz.

Synthesis of alkyl sulfate esters

Racemic sulfate esters **2a-10a**, (*R*)- and (*S*)-(**2a**) were prepared from the corresponding alcohols **2b-10b**, (*R*)- and (*S*)-(**2b**) by using NEt_3*SO_3 following a known procedure[11] with the following modifications: Triethylamine-SO_3 complex was added in 0.9 eq. to avoid residual complex in the lyophilisate. Hence the re-crystallisation step in methanol was not required.

NMR Data and yields for alkyl sulfates:

rac-2-Octyl sulfate (2a):

^1H NMR (300 MHz, D_2O): δ = 4.46-4.36 (m, 1H), 1.62-1.43 (m, 2H), 1.33-1.15 (m, 11H), 0.78 (t, 5.4 Hz, 3H); ^{13}C NMR (75 MHz, D_2O): δ = 78.8, 36.0, 31.0, 28.2, 24.4, 21.9, 20.0, 13.4; 82% yield.

rac-2-Heptyl sulfate (3a):

^1H NMR (300 MHz, D_2O): δ = 4.45- 4.35 (m, 1H), 1.57-1.41 (m, 2H), 1.32-1.15 (m, 9H), 0.78 (t, 5.4 Hz, 3H). ^{13}C NMR (75 MHz, D_2O): δ = 78.8, 36.0, 30.8, 24.1, 21.8, 20.0, 13.3; 76% yield.

rac-2-Nonyl sulfate (4a):

^1H NMR (300 MHz, D_2O): δ = 4.44-4.34 (m, 1H), 1.59-1.39 (m, 2H), 1.35-1.14 (m, 13H), 0.76 (t, 5.0 Hz, 3H); ^{13}C NMR (75 MHz, D_2O): δ = 78.8, 35.9, 31.0, 28.4, 28.3, 23.3, 22.0, 19.9, 13.4; 70% yield.

rac-3-Octyl sulfate (5a):

^1H NMR (300 MHz, D_2O): δ = 4.29-4.21 (m, 1H), 1.70-1.47 (m, 4H), 1.28-1.13 (m, 6H), 0.85-0.75 (m, 6H); ^{13}C NMR (75 MHz, D_2O): δ = 83.6, 32.9, 30.9, 26.5, 23.9, 21.8, 13.3, 8.5; 83% yield.

rac-4-Octyl sulfate (6a):

^1H NMR (300 MHz, D_2O): δ = 4.35-4.27 (m, 1H), 1.60-1.45 (m, 4H), 1.35-1.15 (m, 6H), 0.84-0.76 (m, 6H); ^{13}C NMR (75 MHz, D_2O): δ = 82.3, 35.7, 33.1, 26.4, 21.9, 17.6, 13.2; 70% yield.

rac-6-Methyl-5-hepten-2-yl sulfate (7a):

^{1}H NMR (300 MHz, D$_2$O): δ = 5.14 (t, 35.5 Hz, 1H), 4.45-4.34 (m, 1H), 2.09-1.90 (m, 2H), 1.64-1.46 (m, 8H), 1.23 (d, 7.8 Hz, 3H); ^{13}C NMR (75 MHz, D$_2$O): δ = 133.9, 123.6, 78.2, 36.1, 24.8, 23.2, 19.9, 16.9; 74% yield.

rac-1-Phenylprop-2-yl sulfate (8a):

^{1}H NMR (300 MHz, D$_2$O): δ = 7.33-7.20 (m, 5H), 4.66-4.60 (m, 1H), 2.94-2.81 (m, 2H), 1.20 (d, 7.6 Hz, 3H); the signal at 4.65-4.59 was occluded by the H$_2$O-signal, but could be identified and annotated via 2D-NMR spectroscopy; ^{13}C NMR (75 MHz, D$_2$O): δ = 137.5, 129.9, 128.4, 126.6, 78.3, 41.9, 19.3; 80% yield.

rac-4-Phenylbut-2-yl sulfate (9a):

^{1}H NMR (300 MHz, D$_2$O): δ = 7.30-7.12 (m, 5H), 4.45-4.35 (m, 1H), 2.72-2.53 (m, 2H), 1.86-1.71 (m, 2H), 1.25 (d, 7.5 Hz, 3H); ^{13}C NMR (75 MHz, D$_2$O): δ = 142.2, 128.6, 128.6, 126.0, 77.9, 37.9, 30.8, 19.9; 71% yield.

rac-1-Cyclohexylethyl sulfate (10a):

^{1}H NMR (300 MHz, D$_2$O): δ = 4.23-4.15 (m, 1H), 1.72-1.35 (m, 6H), 1.24-0.83 (m, 8H); ^{13}C NMR (75 MHz, D$_2$O): δ = 82.6, 42.8, 28.0, 27.8, 26.0, 25.6, 16.9; 85% yield.

rac-1-Heptyn-3-yl sulfate (11a):

^{1}H NMR (300 MHz, DMSO-d$_6$): δ = 4.67-4.62 (m, 1H), 3.28 (d, 7.9 Hz, 1H), 1.70-1.53 (m, 2H), 1.42-1.21 (m, 4H), 0.86 (t, 5.9 Hz, 3H); ^{13}C NMR (75 MHz, DMSO-d$_6$): δ = 84.2, 75.6, 66.0, 35.7, 27.0, 22.3, 14.4; 82% yield.

rac-1-Octyn-3-yl sulfate (12a):

^{1}H NMR (300 MHz, DMSO-d$_6$): δ = 4.67-4.62 (m, 1H), 3.24 (d, 6.8 Hz, 1H), 1.68-1.51 (m, 2H), 1.42-1.16 (m, 6H), 0.83 (t, 5.6 Hz, 3H); ^{13}C NMR (75 MHz, DMSO-d$_6$): δ = 83.8, 75.7, 66.4, 35.8, 31.2, 24.3, 22.4, 14.3; 79% yield.

Cloning and expression of alkylsulfatase Pisa1 from *Pseudomonas sp.* DSM6611

Genomic DNA from *Pseudomonas sp.* DSM6611 was prepared using the PureLink^{TM} Genomic DNA Mini Kit (Invitrogen). The open reading frame of the putative *Pseudomonas* alkylsulfatase was amplified by PCR using CTAGCTAGCATGTCCCGCTTCATTCGCGCCAG as forward and CCGCTCGAGGGGGTTCGACGATATTGAACTTCGGGCT as reverse primer. The PCR product was digested with *Nhe*I and *Xho*I (Fermentas) and cloned into the corresponding restriction sites of a modified pET-21a(+) vector (Novagen) using a T4 DNA ligase (Fermentas). The start codon present in the vectors *Nde*I restriction site had previously been removed by site-directed mutagenesis (ATG to CTG). The stop codon was omitted in order to obtain a protein with a C-

terminal hexa-histidine tag. The ligated plasmid was first transformed into *E. coli* TOP10 cells. After plasmid preparation with a PureLink™ HiPure Plasmid Midiprep Kit (Invitrogen), the vector was further transformed into *E. coli* BL21 STAR cells according to the manual of the supplier. Ligation success was verified by restriction digest of the plasmid and subsequent agarose gel (1%) analysis. Cells were grown in LB- medium containing ampicillin (100 µg/mL) and $ZnSO_4$ (30 µg/mL) at 37° C and 120 rpm until the culture reached an OD_{600} of 0.7. After cooling to 20° C IPTG was added to a final concentration of 0.5 mM and Pisa1 was expressed at 20 ˚C and 120 rpm overnight. Cells were harvested at 4° C and 8000 rpm for 15 min. The cell pellet was washed once with sodium chloride (0.9%) and stored at -20 ˚C.

Cloning and expression of SdsA1 from *Pseudomonas aeruginosa*[2]

The gene *sdsa1* was synthesized by DNA 2.0 in a pJ201 vector with a kanamycin resistence. The lyophilized vector was treated as suggested by the supplier. Afterwards it was transferred into a pET-21a(+) vector (Novagen) using *Nde*I and *Xho*I restriction enzymes (Fermentas) following standard protocols. After transformation into competent *E. coli* BL21 (DE3) cells (as described for Pisa1), the protein was expressed with a C-terminal hexa-histidine tag. Expression was carried out as described for Pisa1.

Purification of His-tagged Pisa1 and SdsA1

After resuspension in phosphate buffer (50 mM, 300 mM NaCl, 10 mM Imidazol, pH 8.2) cells were disrupted by sonification using a Sonics & Materials Vibra Cell CV26, 13 mm tip, 30% amplitude, pulse 1s on, 2s off. SDS-PAGE analysis of the soluble and insoluble fraction showed that >80% of the protein was in the soluble fraction (Fig. S2, right). After centrifugation (4° C, 18000 rpm, 20 min) the supernatant was filtered through a MN 615 filter paper (Machery Nagel) and subjected to Ni-NTA affinity chromatography on a Bio-Rad DuoFlow FPLC system equipped with a Ni Sepharose 6 FF column employing standard procedures as described by the manufacturer (GE Healthcare). After SDS-PAGE selected fractions were exposed to dialysis in Tris-Cl (100 mM, pH 8.2) overnight (Fig. S2, left). The yield of this fraction is 18.5 mg protein/L of culture, it is >95% pure and can be used for biocatalytic transformations.

For protein crystallisation, the protein solution was concentrated and further purified with a MonoQ anion exchange 5/50 GL column (5 x 50 mm, GE Healthcare) previously equilibrated with Tris-Cl (20 mM, pH 8.2). The enzymes were eluted with a 10 mL linear gradient of 0-0.4 M NaCl in the same buffer at a flow rate of 1 mL/min and protein of >99% purity was dialysed in Tris-Cl (100 mM, pH 8.2) overnight. Final concentration was determined by the method of Bradford (as described) and purified proteins were flash-frozen and stored at -20° C. No significant loss of activity <10% was detected during several months.

Figure S2. SDS-PAGE of purified Pisa1 (left) and protein composition after cell disruption (right).
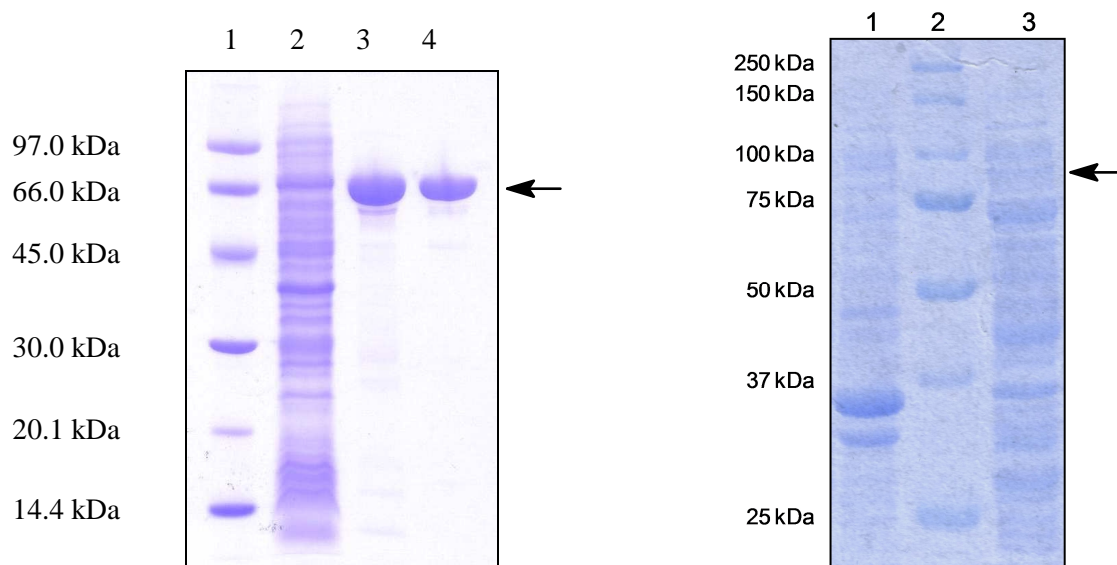
Left: Lane 1: LMW standard (GE Healthcare); lane 2: crude extract; lane 3: pool after Ni-NTA affinity chromatography; lane 4: pool after anion exchange chromatography.

Right: Lane 1: insoluble protein fraction (pellet); lane 2: MW standard (Bio-Rad Precision Plus Standard All Blue); lane 3: soluble protein fraction (supernatant).

Determination of protein concentrations

For determination of protein concentrations, Bradfords method was used: Protein solution (20 µL, diluted if necessary) was added to 980 µL of 1-fold Bio-Rad Bradford solution (diluted from a 5-fold concentrated stock solution) in a plastic cuvette. The reaction mixture was incubated at RT for 10 min and the extinction was measured at 595 nm against a blank containing Bradford solution and water. Protein concentrations were calculated with a calibration curve using bovine serum albumin.

$H_2^{18}O$ **labelling experiments**

Enzymatic assays of alkyl sulfates (**1a**, **2a**, 4 mg, 17 µmol) were conducted in unlabeled and ^{18}O-labeled Tris-HCl (250 µL, 100 mM, pH 8.2 for Pisa1, pH 7.5 for SdsA1). The reaction mixture was shaken at 30° C and 120 rpm for 24 h. After extraction with ethyl acetate (1 mL) the organic phase was dried over anhydrous sodium sulfate. Product **1b** was analyzed on a Agilent 7890A system equipped with an achiral HP-5 column and subsequently derivatized as acetate as described above. Product **2b** was directly derivatized. GC/MS spectra from ^{18}O-labelling experiments were recorded on an Agilent 7890A GC system equipped with an Agilent 5975C mass selective detector (electron impact, 70 eV). Measurements were carried out on an a) Agilent HP-5-MS column (30 m

x 0.25 mm x 0.25 µm film) or b) a Varian Chirasil Dex CB column (25 m x 0.32 mm x 0.25 µm film) using helium as carrier gas.

The following methods were used: a) injector temperature 200° C, flow: 0.62 mL/min; temperature program: 80° C, hold for 1.0 min, 15° C/min, to 110° C, 4° C/min, to 130° C, 10° C/min, to 180° C. b) injector temperature 250° C, flow: 0.5 mL/min; temperature program: 40° C, hold for 2.0 min, 20° C/min, to 180° C; c) injector temperature: 250° C, flow: 0.55 mL/min; temperature program: 100° C, hold for 0.5 min, 10° C/min, to 300° C.

Table S3. GC/MS-measurements

Compund	Column	Method	Retention time [min]
1-octanol (**1b**)	HP-5	c)	4.0
1-octyl acetate (**1c**)	HP-5	b)	8.8
1-octyl acetate (**1c**)	Dex-CB	a)	5.5
(*R*)-2-octyl acetate [(*R*)-**2c**]	Dex-CB	a)	4.5
(*S*)-2-octyl acetate [(*S*)-**2c**])	Dex-CB	a)	4.0

Activity assays for Pisa1 and SdsA1

Alkyl sulfate [**1a**, *rac*-, (*R*)-**,** or (*S*)-**2a, 3a-10a**] (5 mg) was dissolved in Tris-HCl (0.98 mL, 100 mM, pH 8.2). An aliquot of Pisa1 enzyme solution (20 µL, 130 µg, 1.8 nmol) was added. The reaction mixture was shaken at 30° C and 120 rpm for 24 h. After extraction with ethyl acetate (1 mL) the organic phase was dried over anhydrous sodium sulfate, derivatized as acetate ester (see below) and analyzed by chiral GC. Biotransformations of alkyl sulfates (**1a**, **2a**, 5 mg) with SdsA1 were conducted with purified SdsA1 (20 µL, 130 µg, 1.8 nmol) as shown for Pisa1 but at pH 7.5. The identity and absolute configuration of samples was confirmed by comparison with reference samples.

Stereoselectivity assays for Pisa1

Step1, enzymatic hydrolysis:

Alkyl sulfate [**2a-10a**] (50 mg) was dissolved in Tris-HCl (9.8 mL, 100 mM, pH 8.2). Purified enzyme solution (200 µL, 1.3 mg, 17.7 nmol) was added. The reaction mixture was shaken at 30° C and 120 rpm for 6 h. Afterwards the aqueous solution was extracted twice with ethyl acetate (3 mL). An aliquot of the organic layer (1 mL) was dried over anhydrous sodium sulfate, derivatized to the corresponding acetate (**2c-10c**) as described below and analyzed by chiral GC. The remaining organic phase was discarded. The aqueous phase was lyophilized overnight.

Step 2, acidic hydrolysis:

The lyophilisate, *p*-toluenesulfonic acid monohydrate (350 mg, 1.8 mmol) and 1,4-dioxane (10 µL, 0.12 mmol) were dissolved in methyl *tert*-butyl ether/deionized water (20 mL, 97:3). The reaction was stirred under reflux at 40° C for 2 h. After cooling to room temperature saturated $NaHCO_3$ (5 mL) was added to stop the reaction. After extraction with ethyl acetate the organic phase was dried over anhydrous sodium sulfate. The solvent was evaporated under reduced pressure, the residual alcohol was redissolved in ethyl acetate (1 mL), derivatized as described below and analyzed by chiral GC.

Derivatisation of alcohols

Acetic anhydride (100 µL) and DMAP (4-dimethylaminopyridine, cat.) were added to the dried organic phase containing the product alcohol (**1b**-**10b**). The reaction mixture was shaken at 30° C and 120 rpm for 18 h. The reaction was quenched with deionized water (300 µL) and the organic phase was dried over anhydrous sodium sulfate. The derivatives were analyzed on a Varian Chirasil Dex CB column (25 m x 0.32 mm x 0.25 µm film).

Determination of enantiomeric excess

Enantiomeric excess of alcohols (**2b**-**10b**) derivatised as acetates (**2c**-**10c**) was determined using an Agilent Technologies 7890A GC-FID system equipped with an Agilent Technologies 7683B autosampler and a Varian Chirasil Dex CB column (25 m x 0.32 mm x 0.25 µm film). Injector temperature: 200° C, flow: 2.0 mL/min, temperature program: 80° C, hold for 1.0 min, 15° C/min, to 110° C, 4° C/min, to 130° C, 10° C/min, to 180° C. Blank tests were performed in the absence of enzyme.

Method A: Injector temperature 200° C, flow 2.0 mL/min, temperature program: 80° C, hold for 1.0 min, 15° C/min to 110° C, 4° C/min to 130° C, 10° C/min to 180° C.

Method B: Injector temperature: 200° C, flow 2.0 mL/min, temperature program: 80° C, hold for 1.0 min, 3° C/min to 100° C, 15° C/min to 150° C.

Method C: Injector temperature: 200° C, flow 2.0 mL/min, temperature program: 80° C, hold for 1.0 min, 15° C/min to 140° C, 4° C/min to 160° C, 10° C/min to 180° C.

Table S2. GC-measurements.

Compound	Method	Retention time [min]	
1-octanol (**1b**)	A	6.3	
1-octyl acetate (**1c**)	A	7.0	
2-octyl acetate (**2c**)	A	5.3 (*S*)	5.9 (*R*)
2-heptyl acetate (**3c**)	B	5.8 (*S*)	7.1 (*R*)
2-nonyl acetate (**4c**)	A	7.1 (*S*)	7.8 (*R*)
3-octyl acetate (**5c**)	A	5.1 (*S*)	5.9 (*R*)
4-octyl acetate (**6c**)	A	4.6 (*S*)	4.8 (*R*)
6-methylhept-5-en-2-yl acetate (**7c**)	A	5.3 (*S*)	5.7 (*R*)
1-phenylprop-2-yl acetate (**8c**)	A	8.7 (*S*)	9.1 (*R*)
4-phenylbut-2-yl acetate (**9c**)	C	8.4 (*S*)	8.7 (*R*)
cyclohexylethyl acetate (**10c**)	A	6.9 (*S*)	7.5 (*R*)
1-heptyn-3-yl acetate (**11c**)	B	6.3 (*R*)*	7.3 (*S*)*
1-octyn-3-yl acetate (**12c**)	C	5.5 (*R*)*	5.9 (*S*)*

* Switch in CIP priority.

## Determination of absolute configuration

Absolute configuration was proven by comparison of the GC retention times with commercial reference samples of enantiopure alcohols (*R*)- and (*S*)-2b, (*R*)-3b, (*R*)-4b, (*S*)-5b, (*R*)- and (*S*)-7b, (*S*)-8b, (*S*)-9b, (*R*)- and (*S*)-12b after derivatisation. The absolute configuration of 10b was determined to be (*S*) by comparison of GC retention times with a sample prepared according to literature.[3] A sample of (*S*)-11c was prepared according to literature.[4] Compound 6b was determined to be (*S*) by optical rotation: $[\alpha]_D^{22}$ = +0.5° (c = 2, CHCl$_3$); ref.[5]:(*R*)-6b $[\alpha]_D^{22}$ = -0.65° (neat); ref.[6]: (*S*)-6b $[\alpha]_D^{25}$ = +0.75° (neat).

Preparative scale deracemisation

2-Octyl sulfate (*rac*-2a, 1 g, 4.3 mmol) was dissolved in Tris-HCl (99 mL, 100 mM, pH 8.2). Purified Pisa1 (1 mL, 6.5 mg, 88.2 nmol) was added. The reaction mixture was shaken at 30° C and 120 rpm for 24 h. The aqueous solution was then extracted with *t*BuOMe (3 x 100 mL). The combined organic phases were extracted with deionized water (100 mL) and saturated sodium chloride solution (100 mL). The organic layer was then dried with anhydrous sodium sulfate and filtered through a sintered glass filter. The solvent was evaporated under reduced pressure (250 mbar, RT). The product obtained was a clear yellow oil (248 mg, 1.9 mmol). Its purity was

confirmed by NMR and GC/MS analysis. The aqueous phase was lyophilized overnight. The lyophilisate, *p*-toluenesulfonic acid monohydrate (1.75 g, 9.2 mmol) and 1,4-dioxane (100 µL, 1.2 mmol) were dissolved in MTBE /deionized water (200 mL, 97:3). The reaction was stirred under reflux at 40° C for 5 h. After cooling to room temperature saturated sodium hydrogen carbonate (50 mL) was added to stop the reaction. The aqueous phase was extracted with MTBE (2x 50 mL). The combined organic phases were extracted with deionized water (100 mL) and saturated sodium chloride solution (100 mL) and dried with anhydrous sodium sulfate. After filtering through a sintered glass filter, the solvent was evaporated under reduced pressure (250 mbar, RT). The product was obtained as a clear yellow oil (243 mg, 1.9 mmol). (*S*)-2b was obtained in a total isolated yield of 87% (491 mg, 3.8 mmol) with an ee of >99%; $[\alpha]_D^{20}$ = +7.21, (CHCl$_3$, c = 2), lit.[7] (*S*)-2b $[\alpha]_D^{20}$ = 9.12. The purity of the product was proven by ^{1}H- and ^{13}C-NMR: ^{1}H NMR (300 MHz, CDCl$_3$): δ = 3.86-3.75 (m, 1H), 1.55-1.17 (m, 13H), 0.93-0.86 (m, 3H); ^{13}C NMR (75 MHz, CDCl$_3$): δ = 68.2, 39.4, 31.8, 29.3, 25.7, 23.5, 22.6, 14.1.

^{1}H-NMR (2a) *rac*-2-Octyl sulfate; D$_2$O
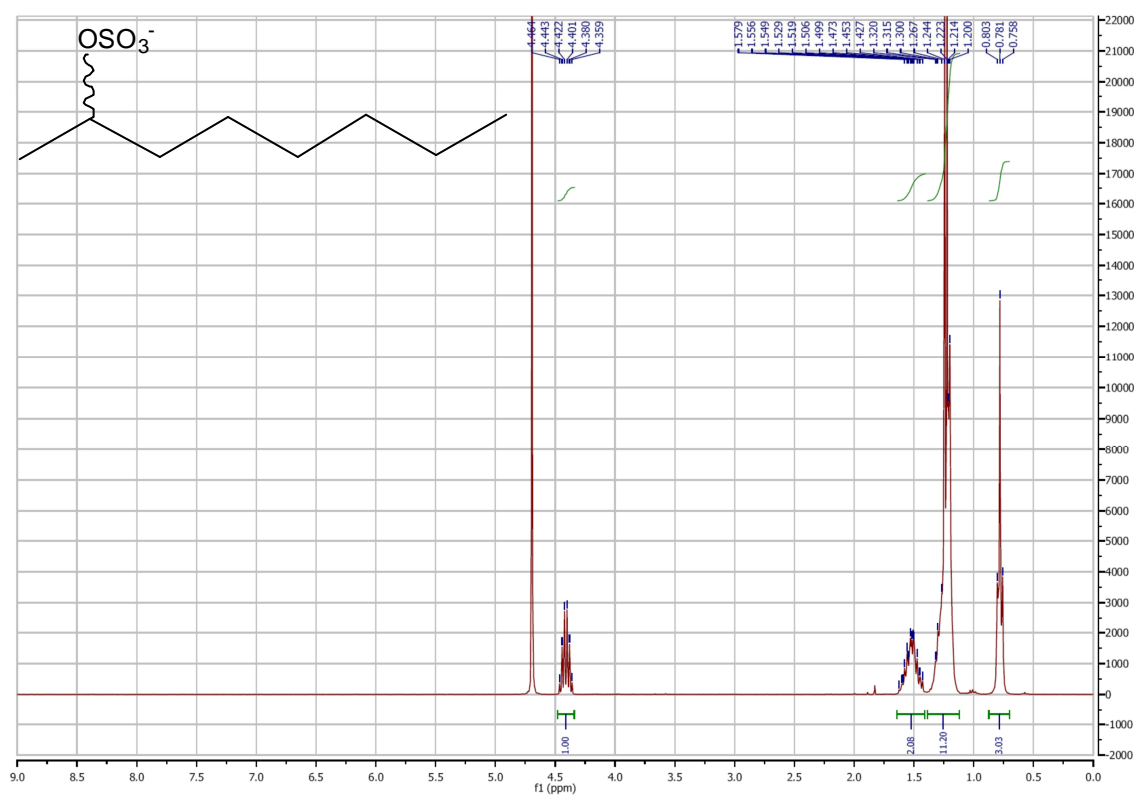
^{13}C-NMR (**2a**) *rac*-2-Octyl sulfate; D$_2$O
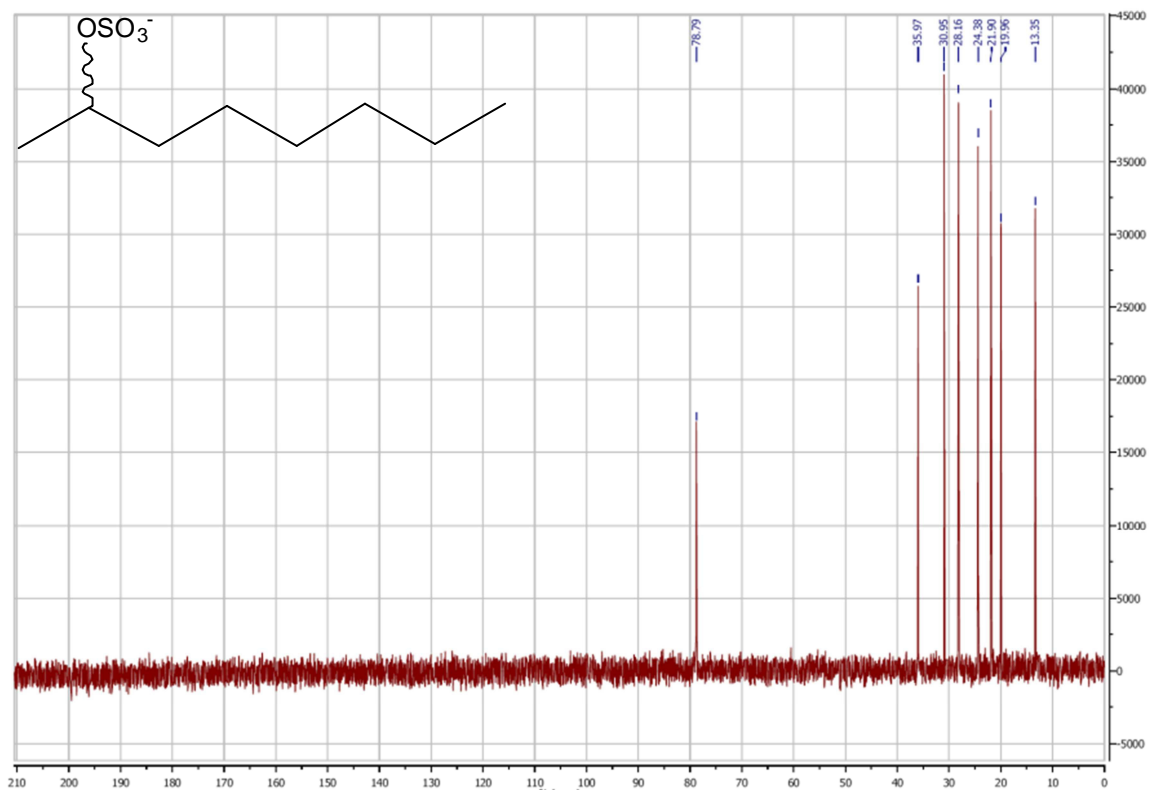
^{1}H-NMR (**3a**) *rac*-2-Heptyl sulfate; D$_2$O
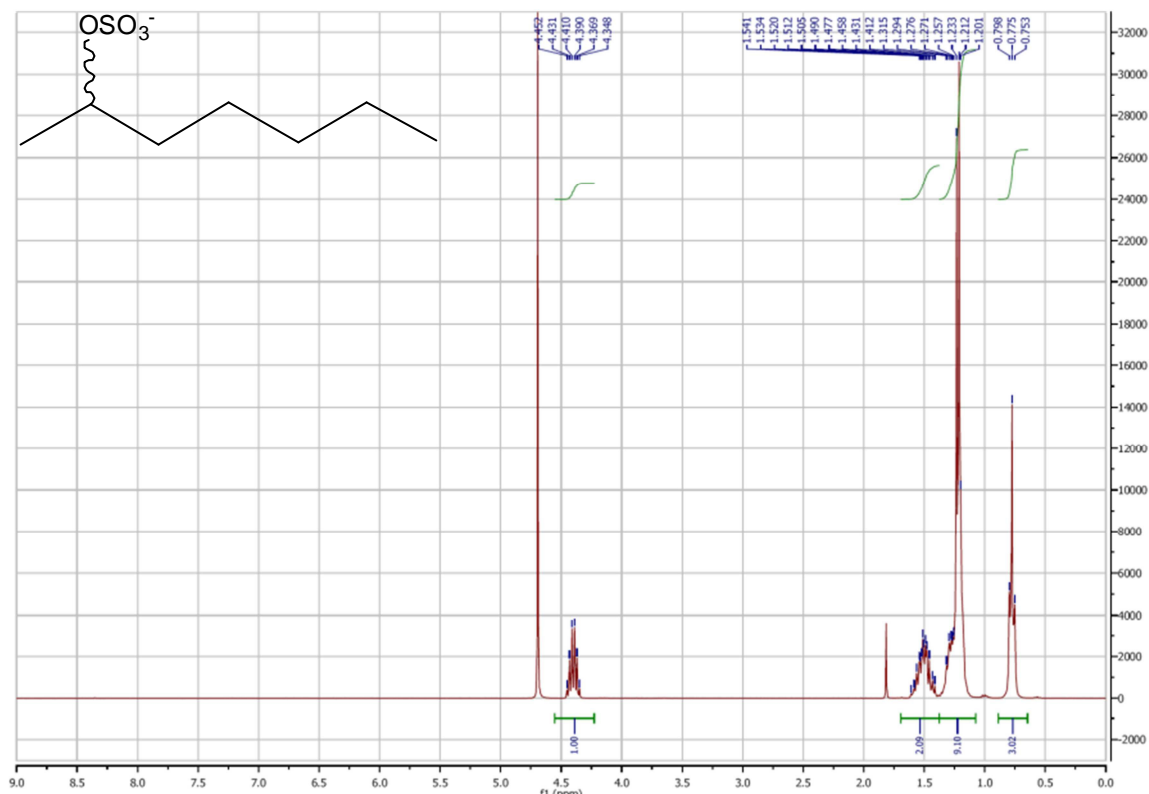
^{13}C-NMR (**3a**) *rac*-2-Heptyl sulfate; D$_2$O
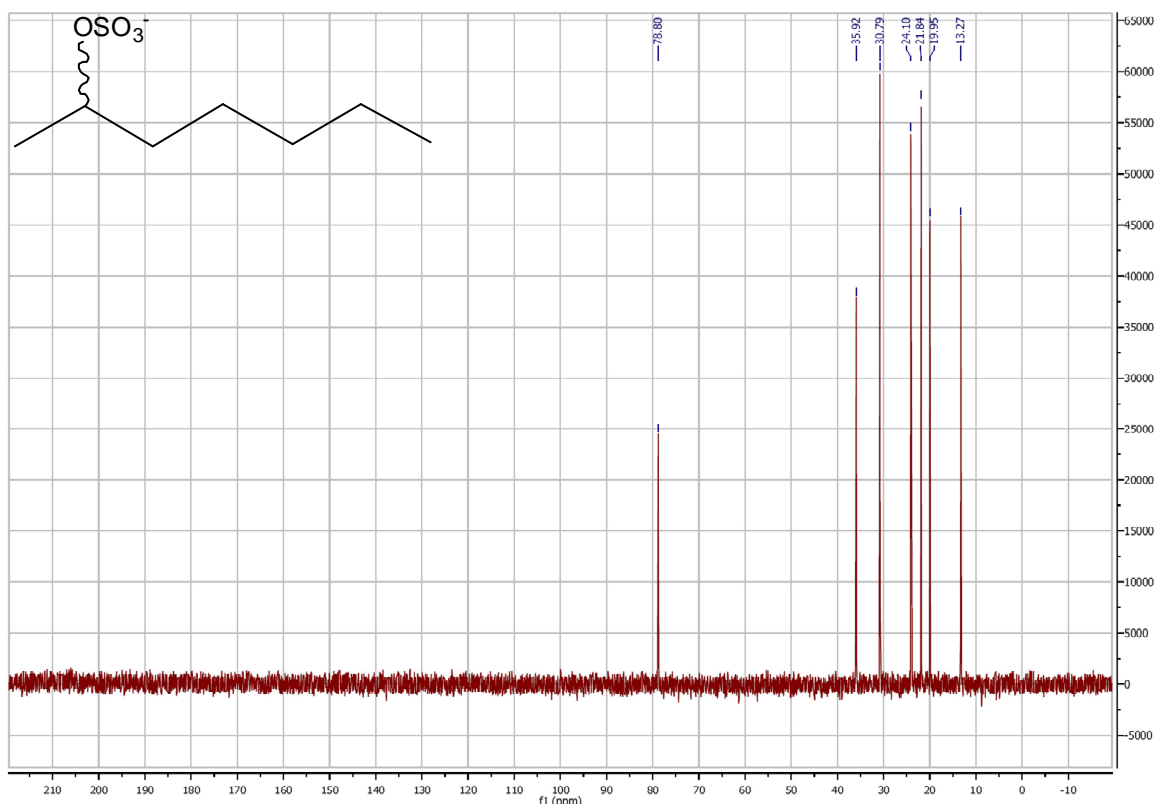
^{1}H-NMR (**4a**) *rac*-2-Nonyl sulfate; D$_2$O
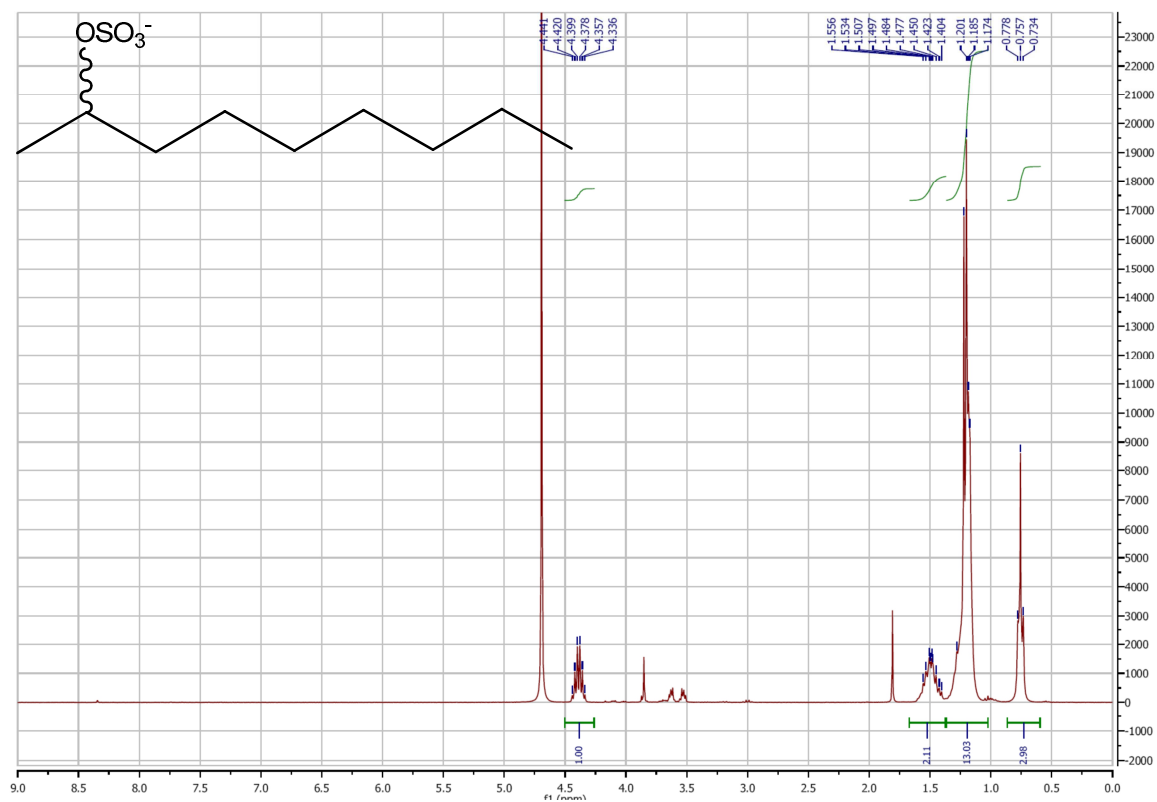
^{13}C-NMR (**4a**) *rac*-2-Nonyl sulfate; D_2O
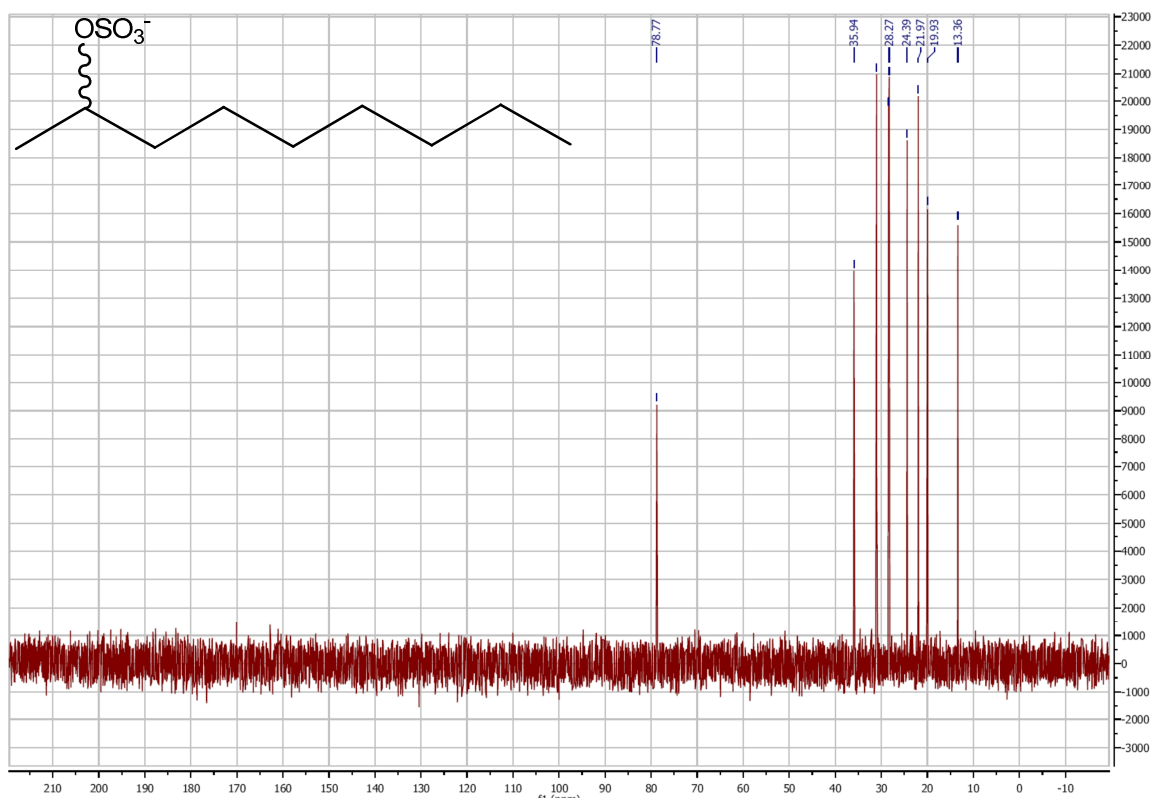
^{1}H-NMR (**5a**) *rac*-3-Octyl sulfate; D_2O
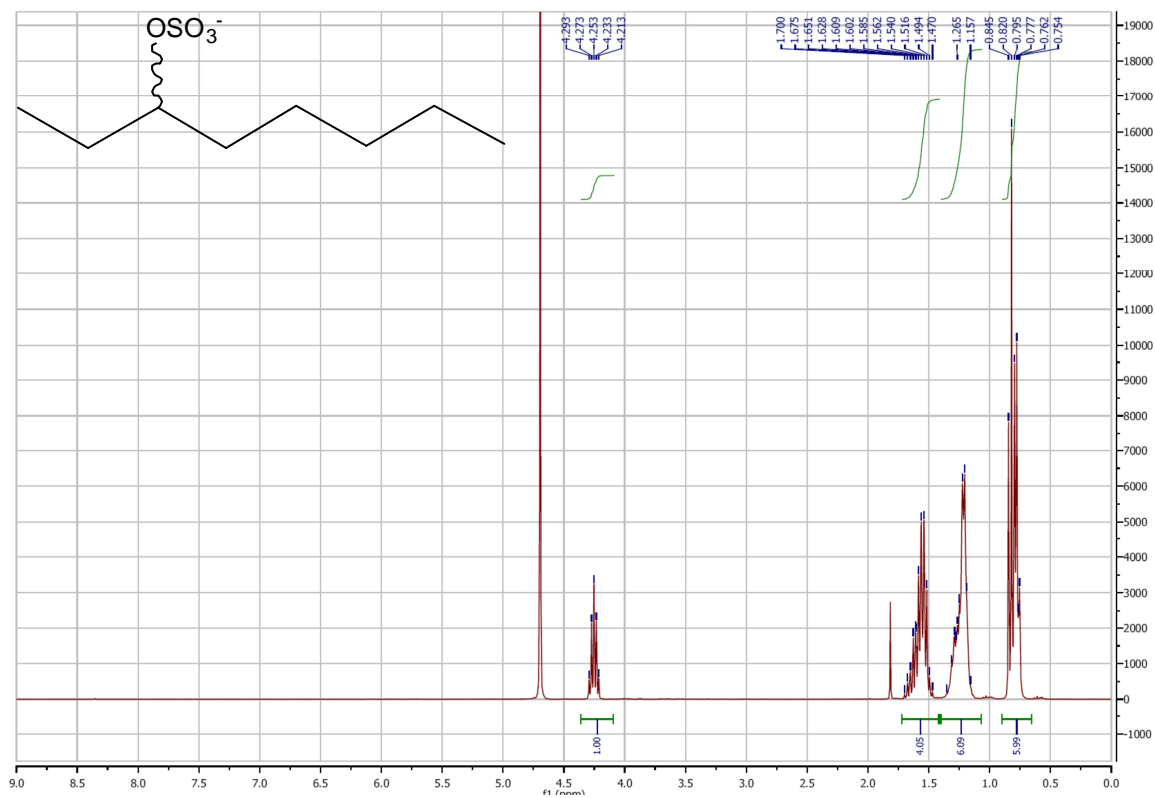
^{13}C-NMR (**5a**) *rac*-3-Octyl sulfate; D$_2$O
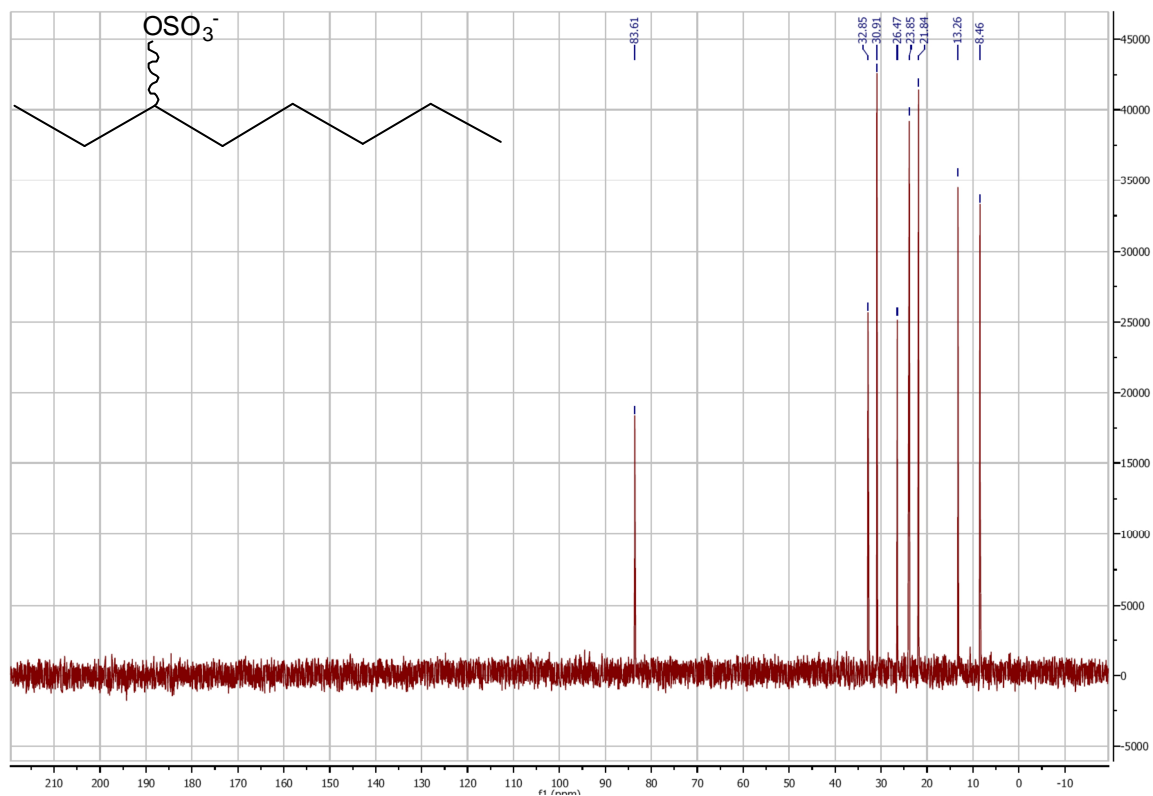
^{1}H-NMR (**6a**) *rac*-4-Octyl sulfate; D$_2$O
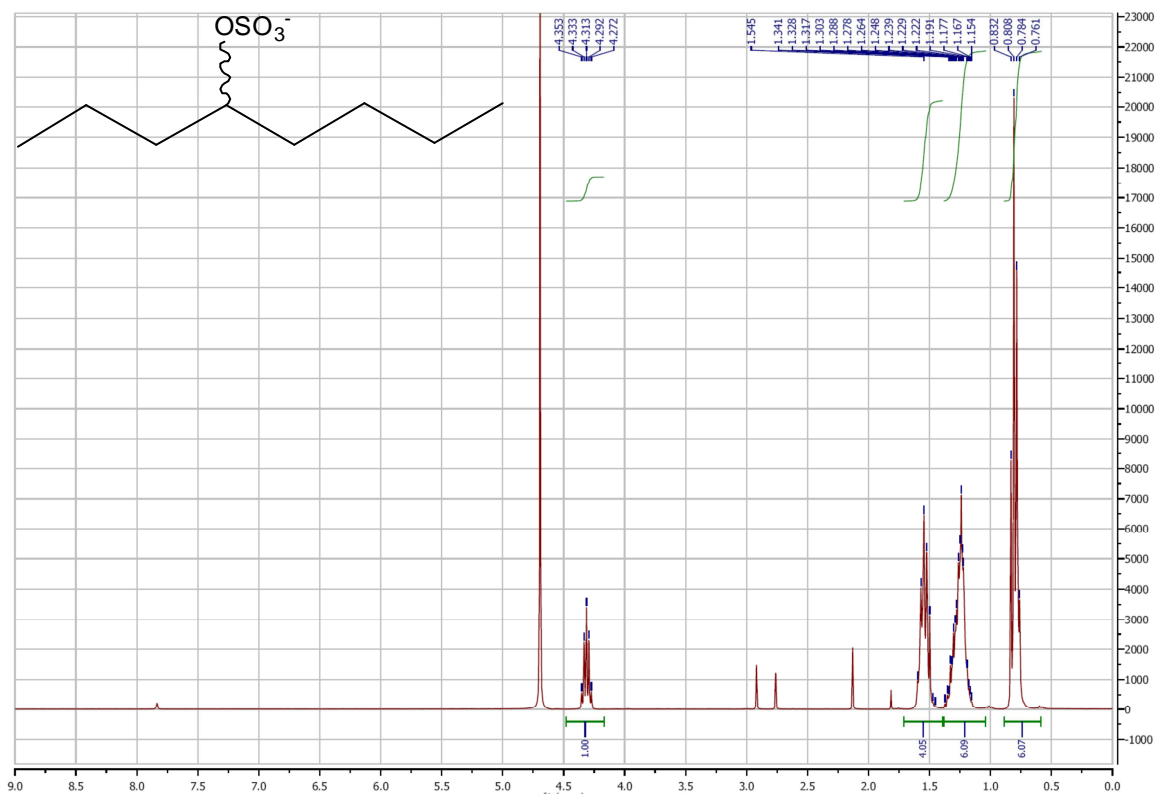
^{13}C-NMR (**6a**) *rac*-4-Octyl sulfate; D$_2$O
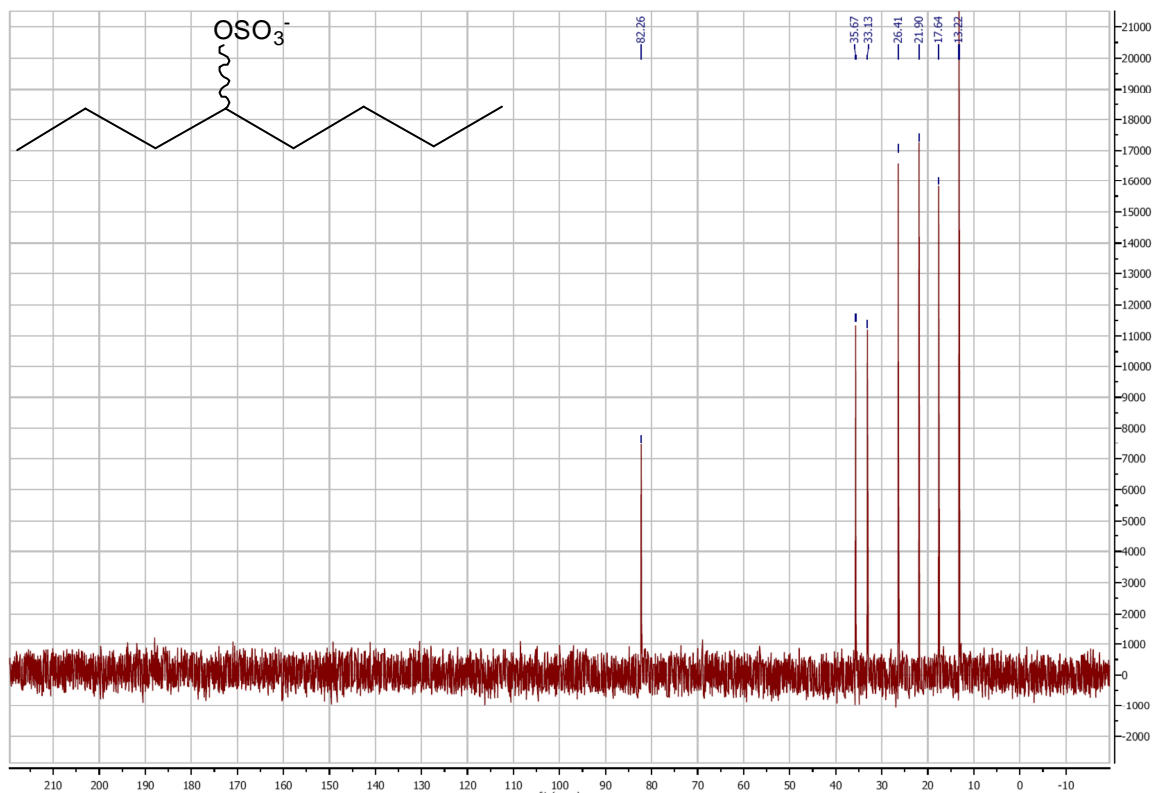
^{1}H-NMR (**7a**) *rac*-6-Methyl-5-hepten-2-yl sulfate; D$_2$O
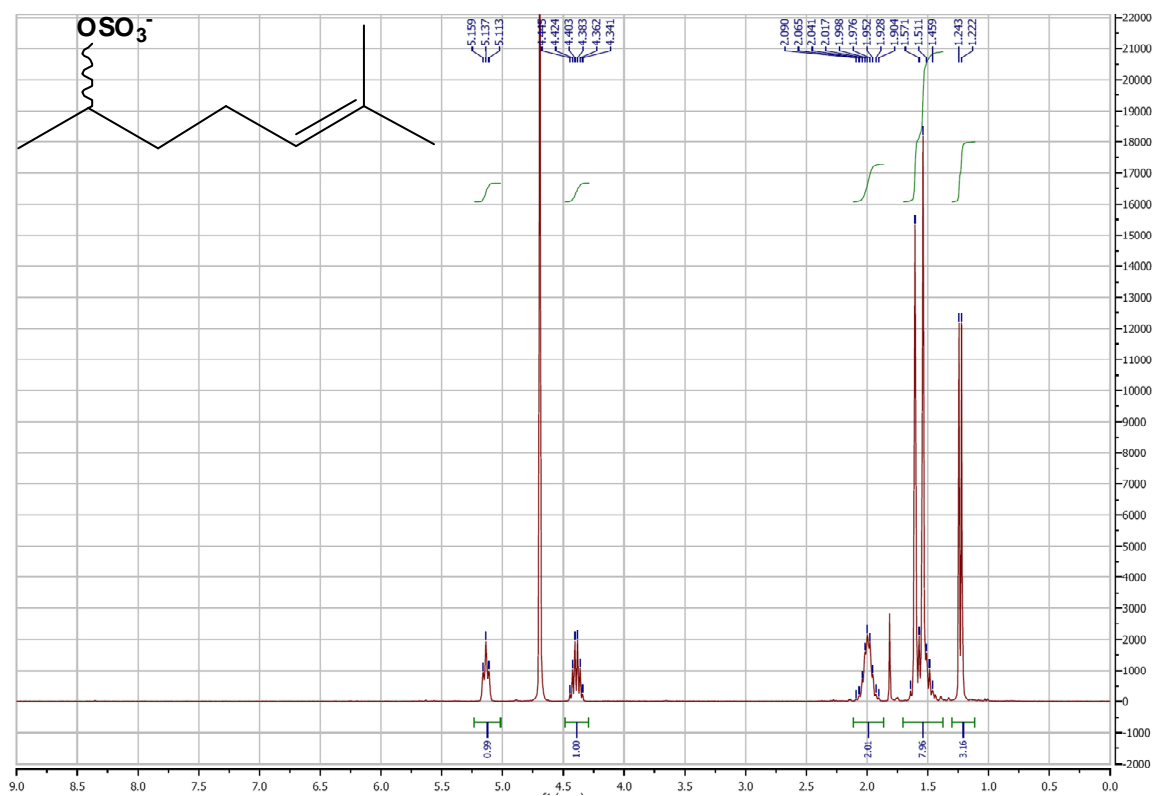
¹³C-NMR (**7a**) *rac*-6-Methyl-5-hepten-2-yl sulfate; D$_2$O
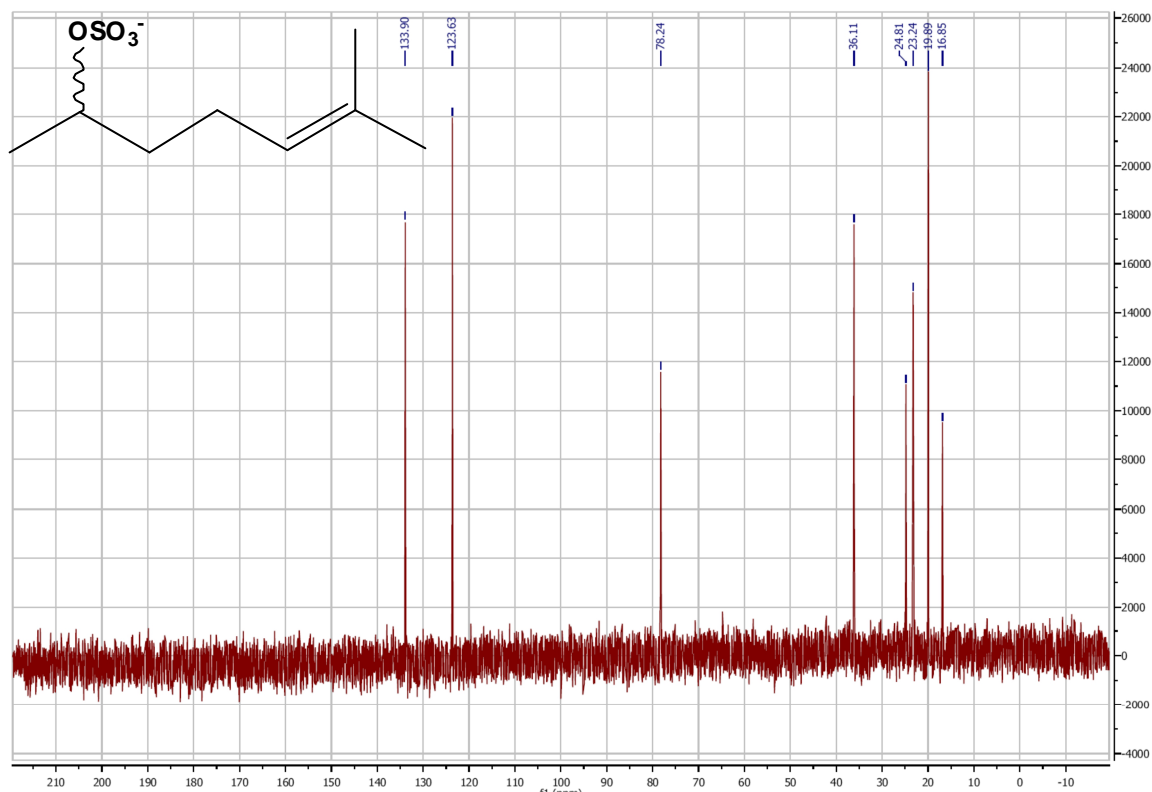
¹H-NMR (**8a**) *rac*-1-Phenylprop-2-yl sulfate; D$_2$O
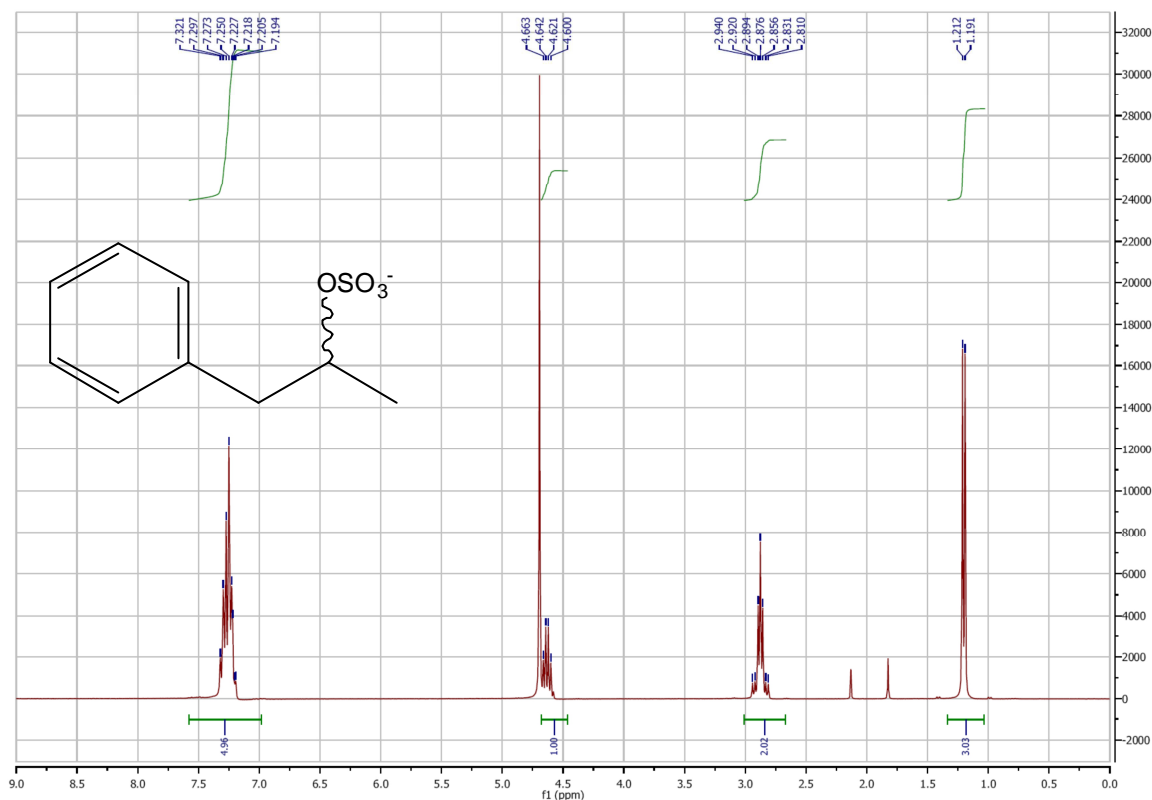
¹³C-NMR (**8a**) *rac*-1-Phenylprop-2-yl sulfate; D₂O
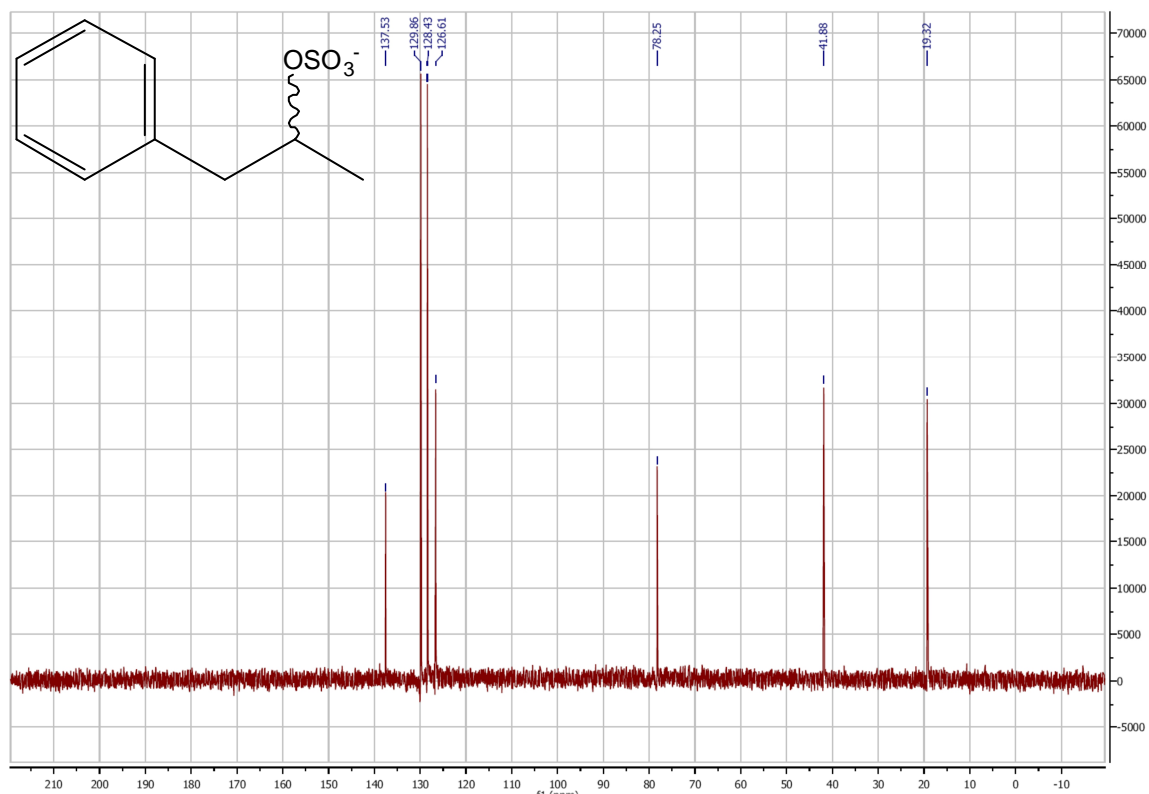
¹H-NMR (**9a**) *rac*-4-Phenylbut-2-yl sulfate; D₂O
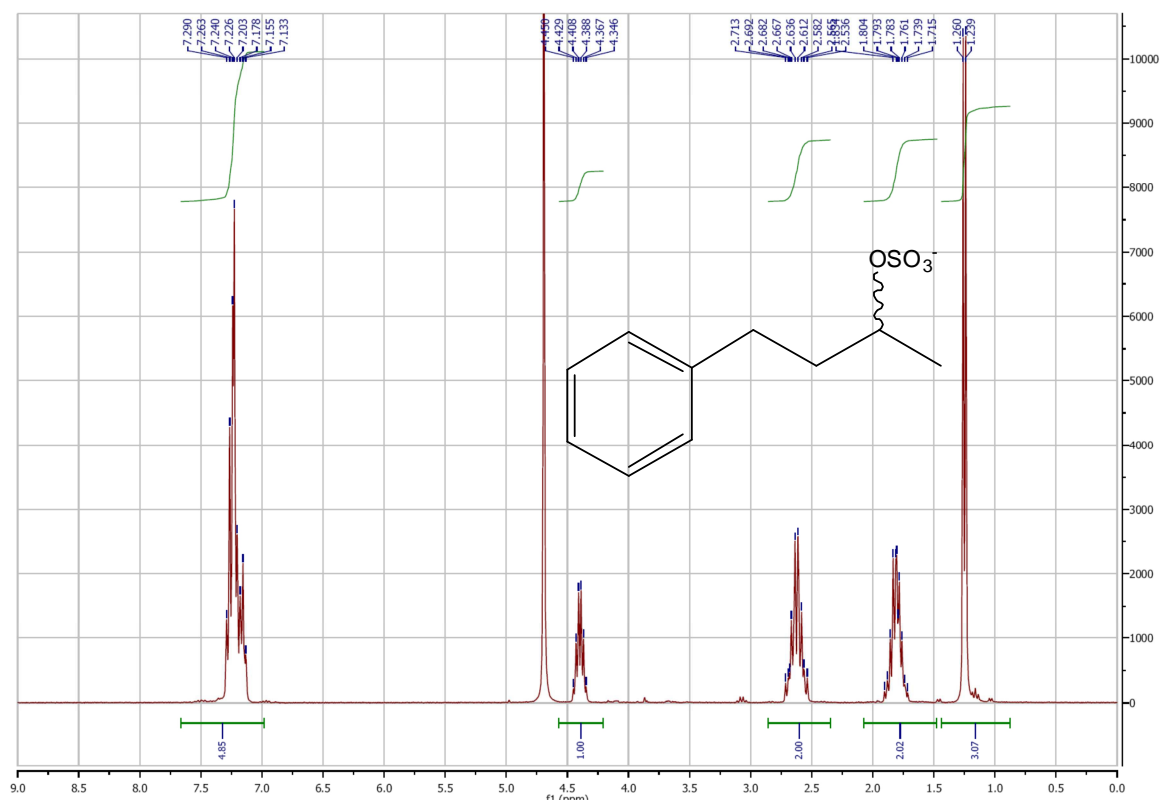
^{13}C-NMR (**9a**) *rac*-4-Phenylbut-2-yl sulfate; D_2O
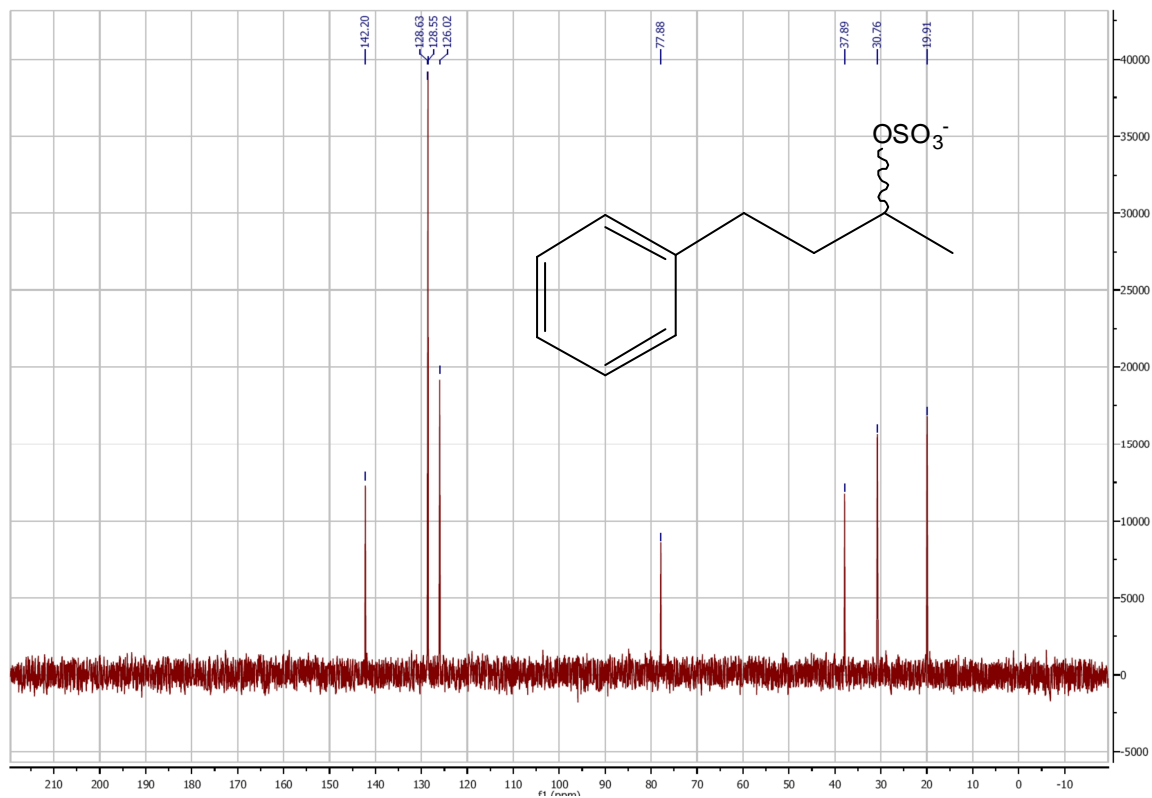
^{1}H-NMR (**10a**) *rac*-1-Cyclohexylethyl sulfate; D_2O
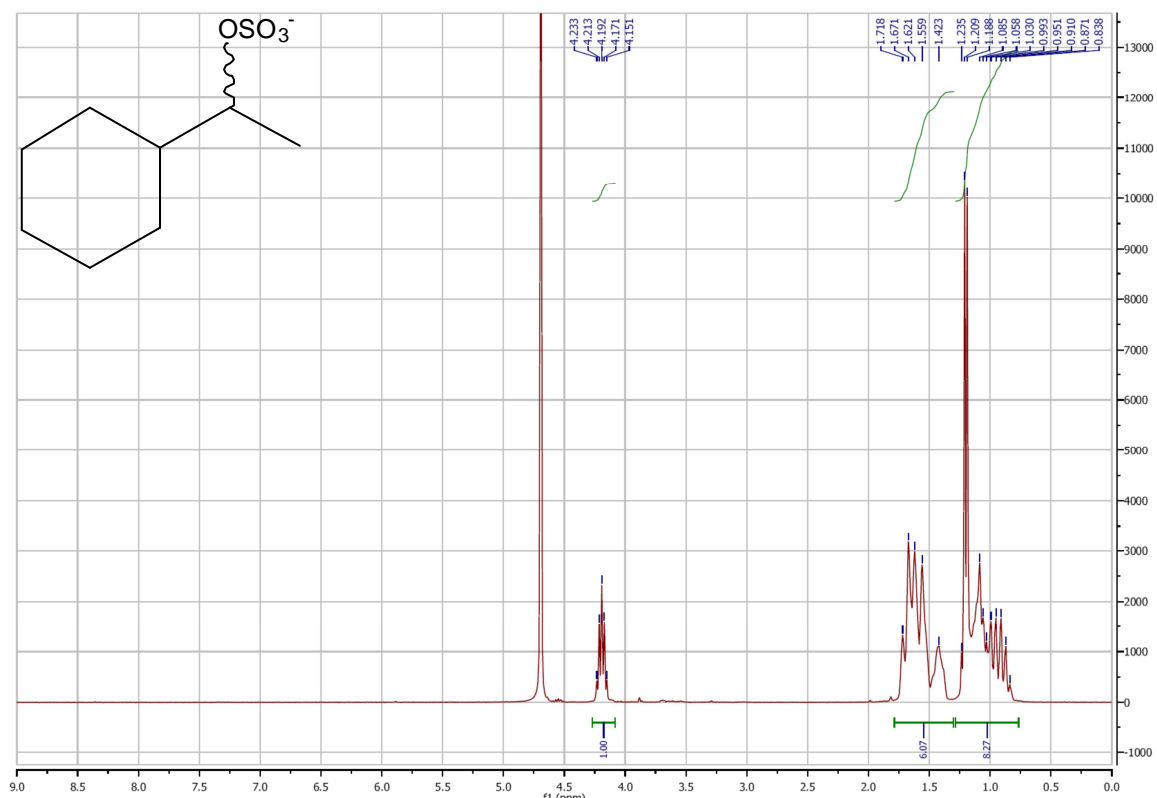
¹³C-NMR (10a) *rac*-1-Cyclohexylethyl sulfate; D₂O
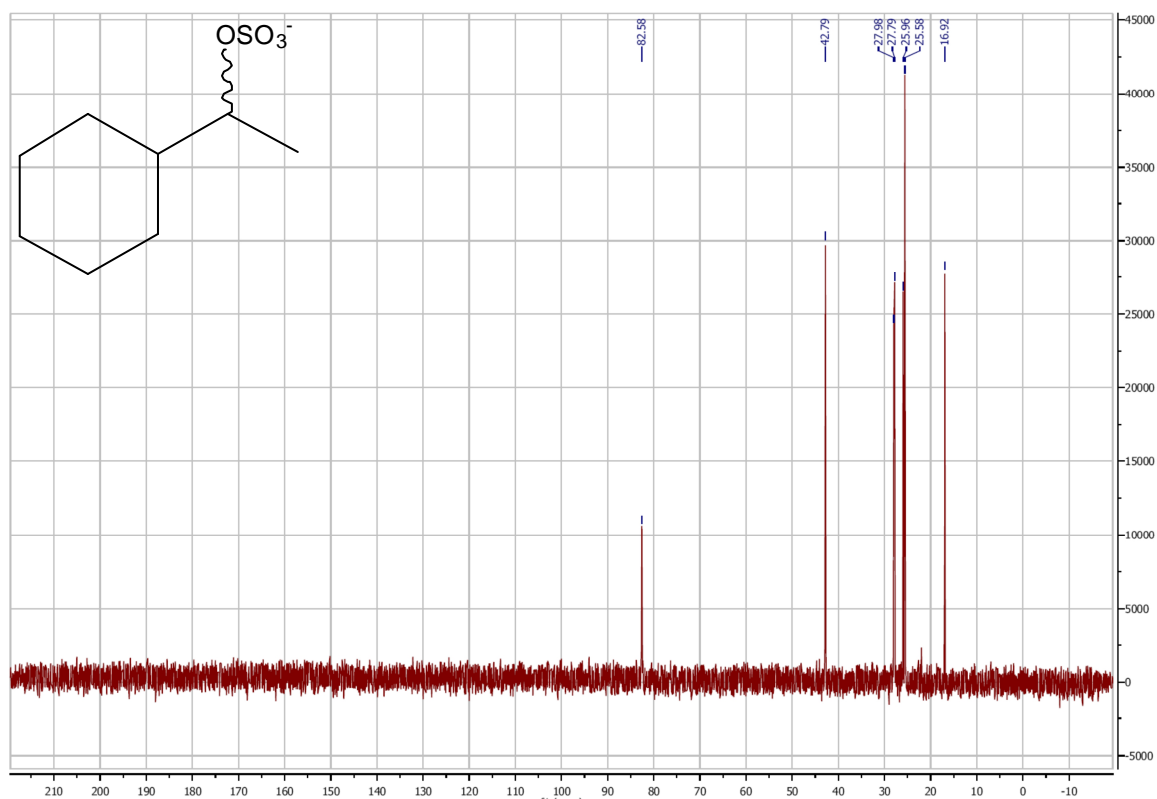
¹H-NMR (11a) *rac*-1-Heptyn-3-yl sulfate; DMSO-d₆
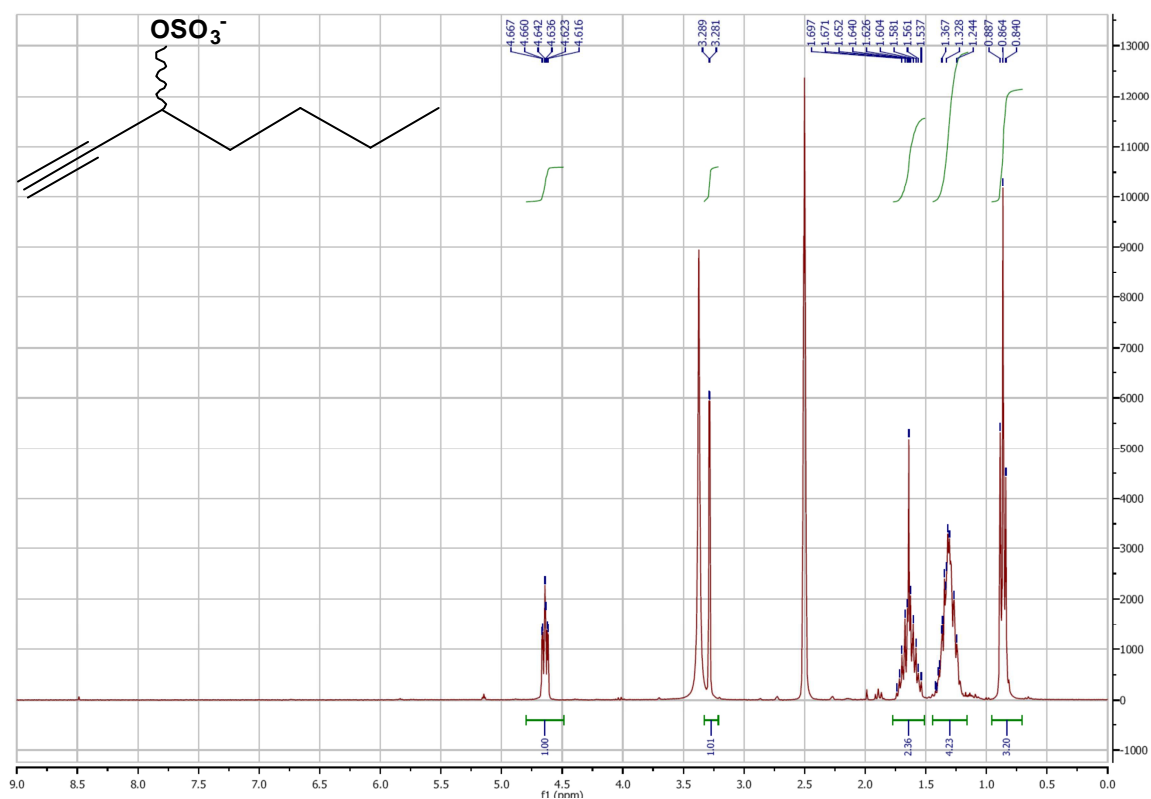
^{13}C-NMR (**11a**) *rac*-1-Heptyn-3-yl sulfate; DMSO-d$_6$

^{1}H-NMR (**12a**) *rac*-1-Octyn-3-yl sulfate; DMSO-d$_6$
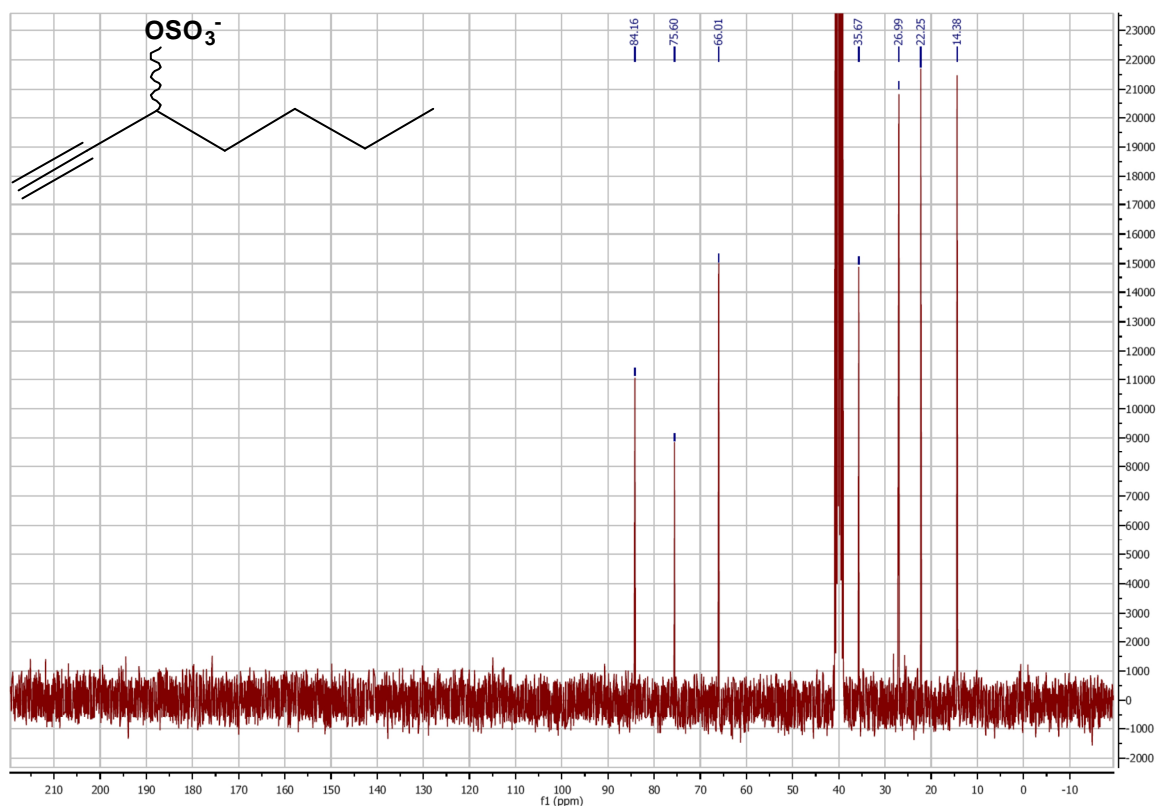
13C-NMR (**12a**) *rac*-1-Octyn-3-yl sulfate; DMSO-d₆
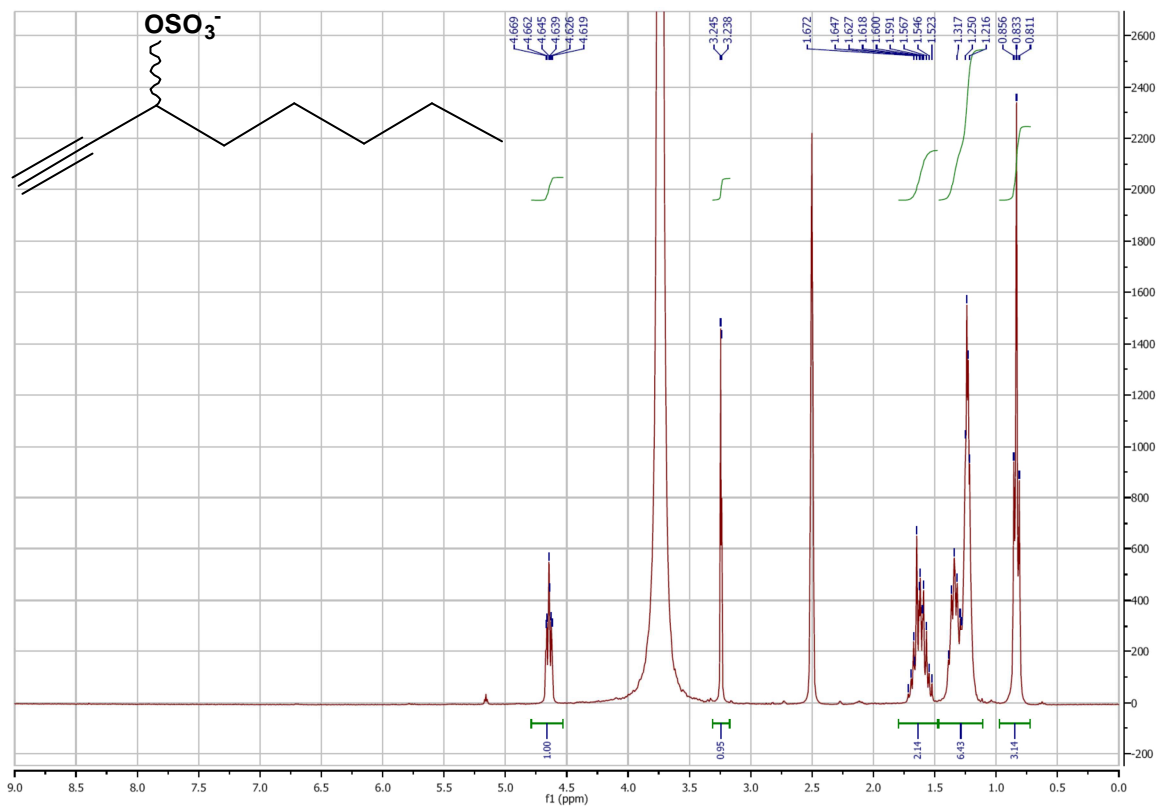
13C-NMR (12a) rac-1-Octyn-3-yl sulfate; DMSO-d6
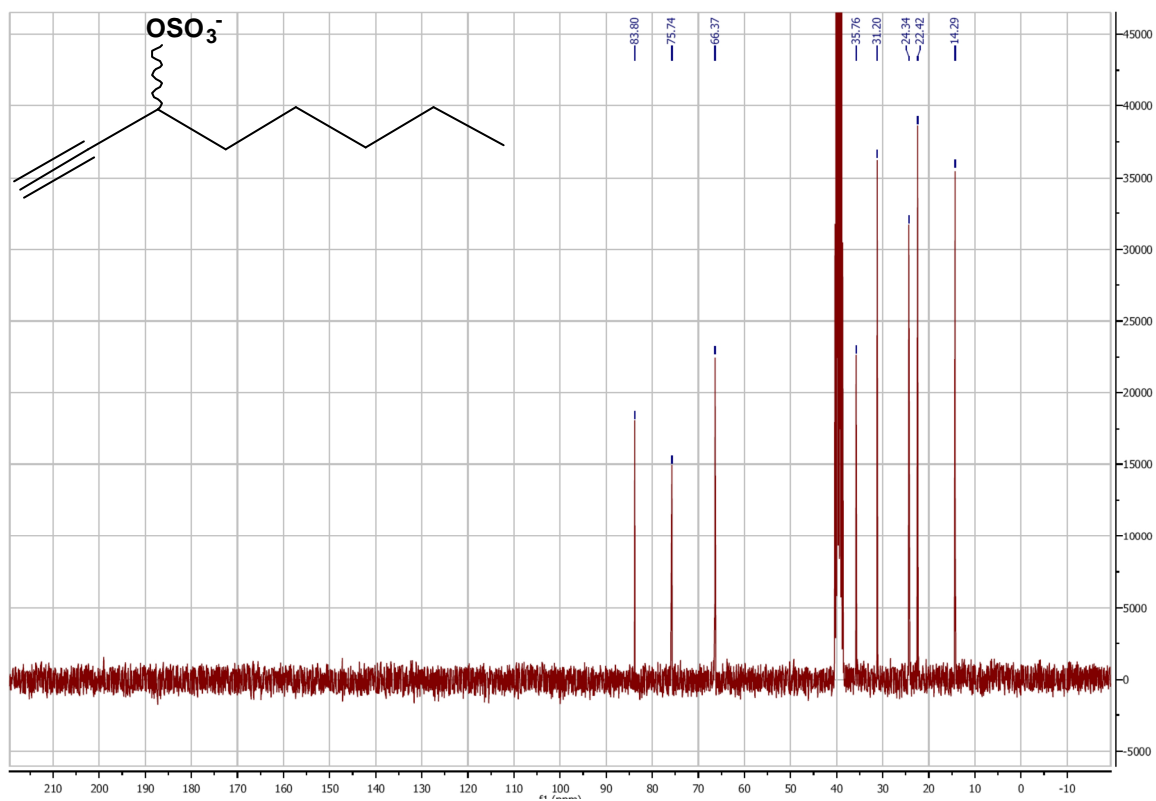
1H-NMR ((*S*)-2b) (*S*)-2-Octanol; CDCl3
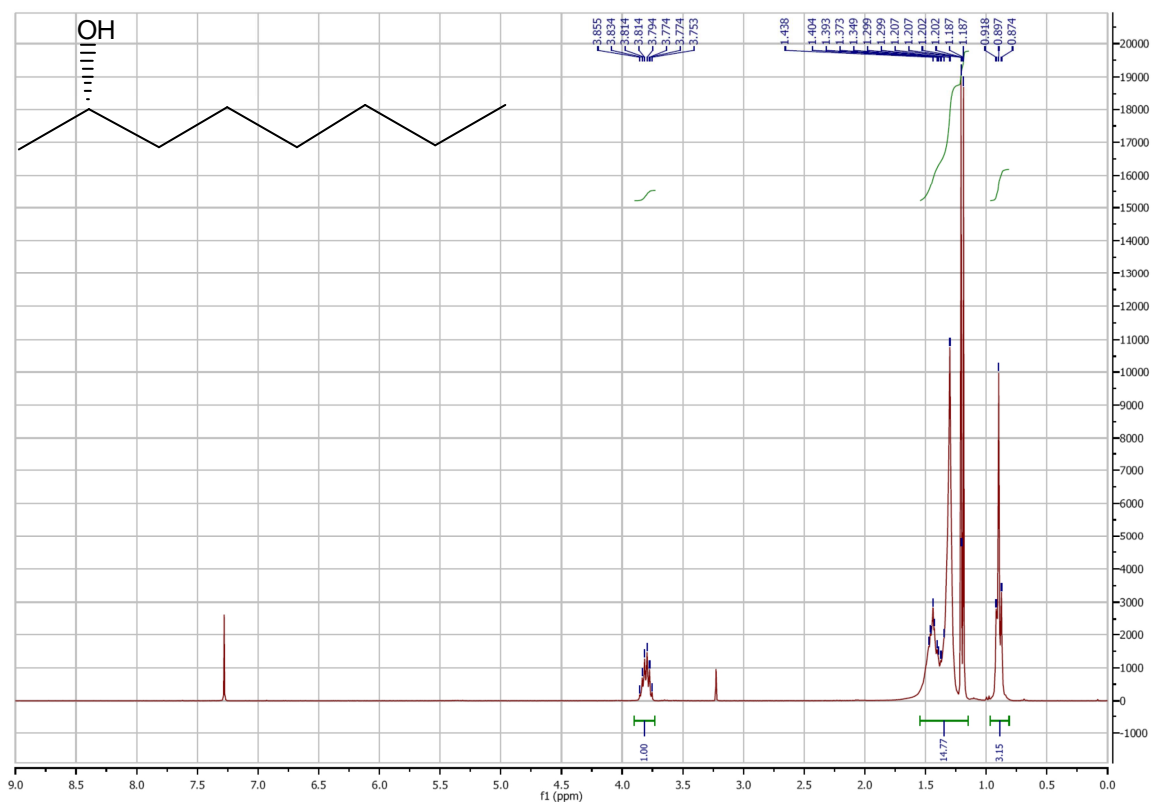
¹³C-NMR (**(*S*)-2b**) (*S*)-2-Octanol; CDCl₃
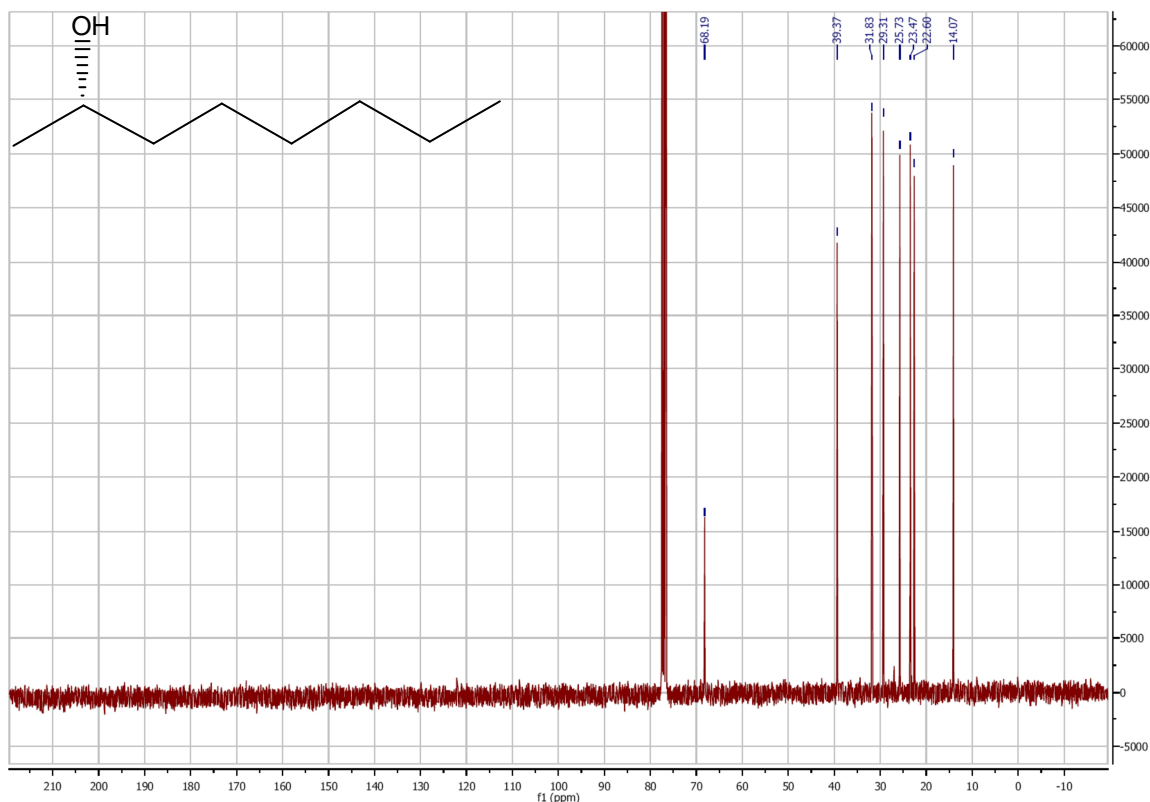
References and Notes

[1]   Pogorevc M.; Kroutil W.; Wallner S. R.; Faber K. *Angew. Chem. Int. Ed.* **2002**, *41*, 4052-4054; Pogorevc M.; Faber K. *Tetrahedron: Asymmetry* **2002**, *13*, 1435-1441.

[2]   Hagelueken G.; Adams T. M.; Wiehlmann L.; Widow U., Kolmar H.; Tümmler B.; Heinz D. W.; Schubert W.-D. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 7631-7636.

[3]   Ema T.; Ura N.; Yoshii M.; Korenaga T.; Sakai T. *Tetrahedron* **2009**, 65, 9583-9591.

[4]   Abad J.-L.; Villorbina G.; Fabrias G.; Camps F. *J. Org. Chem.* **2004**, *69*, 7108-7113.

[5]   Canales E.; Gonzalez A.; Soderquist J. *Angew. Chem. Int. Ed.* **2007**, *46*, 397-399.

[6]   Lodogan J. G.; Ley S. V.; Pattenden G. (eds.) *Dictionary of Organic Compounds*, Chapman & Hall, London, 6th edn., **1995**, p. 1003.

[7]   Ochiai M.; Yoshimura A.; Miyamoto K.; Hayashi S.; Nakanishi W.; *J. Am. Chem. Soc.* **2010**, *132*, 9236-9239.

Chapter 6

The substrate spectrum of the inverting *sec*-alkylsulfatase Pisa1

AUTHOR CONTRIBUTIONS

FULL PAPERS

Advanced Synthesis & Catalysis

The Substrate Spectrum of the Inverting *sec*-Alkylsulfatase Pisa1

Markus Schober,[a] Tanja Knaus,[b] Michael Toesch,[a] Peter Macheroux,[b]
Ulrike Wagner[c] and Kurt Faber[a]

[a] Department of Chemistry, Organic & Bioorganic Chemistry, University of Graz, Heinrichstrasse 28, A-8010 Graz, Austria.
Fax: +43-316-380-9840; phone+43-316-380-5332; email: Kurt.Faber@Uni-Graz.at
[b] Institute of Biochemistry, Graz University of Technology, Petersgasse 12, A-8010 Graz, Austria.
[c] Institute of Molecular Biosciences, Structural Biology, University of Graz, Humboldtstrasse 50/III, A-8010 Graz, Austria.

Abstract: The substrate spectrum of the inverting alkylsulfatase Pisa1 was investigated using a range of *sec*-alkylsulfate esters bearing aromatic, olefinic and acetylenic moieties. Perfect enantioselectivities were obtained for substrates bearing groups of different size adjacent to the sulfate ester moiety. Insufficient selectivities could be doubled by using DMSO as co-solvent. Hydrolytically unstable benzylic sulfate esters could be sufficiently stabilised by introduction of electron-withdrawing substituents. Overall, Pisa1 appears to be a very useful inverting alkylsulfatase for the deracemisation of *rac-sec*-alcohols via enzymatic hydrolysis of their corresponding sulfate esters, which furnishes homochiral products possessing 'anti-Kazlauskas' configuration.

Keywords: alkylsulfatase, inversion, substrate spectrum, deracemization.

Introduction

Driven by the demand to enhance the economic balance of chemical processes, the design of asymmetric catalytic protocols, which allow the transformation of a racemate into a single stereoisomeric product without the occurrence of an 'unwanted' isomer is an important topic.[1] Depending on the type of starting material, these so-called deracemisation methods are based on dynamic kinetic resolution,[2] stereo-inversion[3] or enantio-convergent processes.[4] In the latter, both starting enantiomers are transformed via independent pathways via retention or inversion of configuration to furnish a single stereoisomer. In order to fulfil this prerequisite, catalysts must not only be enantioselective (by preferring one enantiomer over the other), but also stereoselective (by inverting one enantiomer, but retaining the other). Enzymes that elicit the complex potential to invert the stereochemistry of their substrate during catalysis are rather rare and encompass haloalkane dehalogenases,[5] epoxide hydrolases[6] and (alkyl) sulfatases.[7] Depending on the subtype of enzyme, enzymatic sulfate ester hydrolysis may proceed via cleavage of the S-O or the C-O bond, which causes retention or inversion of configuration at the stereogenic carbon centre.[8] Whereas the mechanism of retaining sulfatases is well understood,[9] much less is known about inverting (alkyl) sulfatases.[10] It was only recently, that they have been identified as members of the metallo-β-lactamase family, which act through nucleophilic displacement of HSO_4^- via (formal) $[OH^-]$, which is provided from water by a binuclear Zn^{2+}-cluster.[11] We have recently characterised the first inverting *sec*-alkylsulfatase (Pisa1) from *Pseudomonas* sp. DSM 6611 on a molecular level.[12,13] In order to

evaluate the applicability of this useful enzyme, which transforms a *rac-sec*-alkyl sulfate into a homochiral mixture of a *sec*-alcohol and non-

reacted sulfate ester, the substrate tolerance and the operational stability of this enzyme is reported here.

Scheme 1. Kinetic resolution of *sec*-alkylsulfate esters *rac*-**1a-24a** using the inverting alkylsulfatase Pisa1 yields homochiral products.
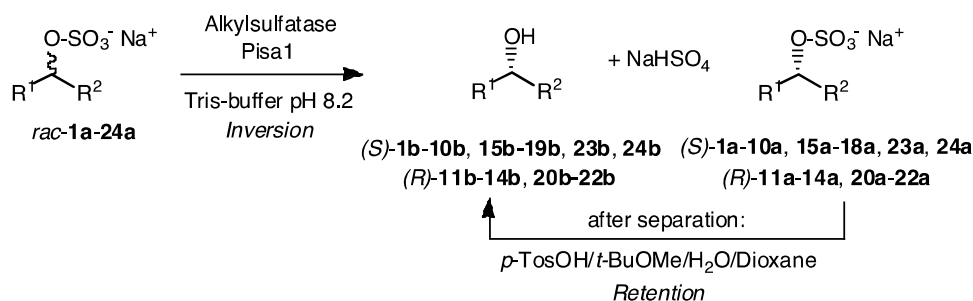
Compound	R^1	R^2	Compound	R^1	R^2
1a,b	Me	*n*-Pr	**13a,b**	$CH=CH_2$	*n*-Hexyl
2a,b	Me	*n*-Bu	**14a,b**	$CH=CH_2$	*n*-Heptyl
3a,b	Me	$(CH_2)_2CHMe_2$	**15a,b**	Me	C☐C-Et
4a,b	Me	*n*-Hexyl	**16a,b**	Me	CH_2-C☐C-Me
5a,b	Me	*n*-Octyl	**17a,b**	Me	C☐C-Ph
6a,b	Et	*n*-Pentyl	**18a,b**	Et	C☐C-Me
7a,b	Et	*n*-Bu	**19a,b**	Et	C☐C-Et
8a,b	*n*-Pr	*n*-Bu	**20a,b**	C☐CH	*n*-Pr
9a,b	Me	CH_2-CH=CH_2	**21a,b**	C☐CH	CH_2-CHMe_2
10a,b	Me	$(CH_2)_2CH=CH_2$	**22a,b**	C☐CH	*n*-Pentyl
11a,b	$CH=CH_2$	*n*-Bu	**23a,b**	Me	Ph
12a,b	$CH=CH_2$	*n*-Pentyl	**24a,b**	Me	m,m-$(CF_3)_2C_6H_3$-

Results and Discussion

In order to provide a broad coverage of substrate structures, *sec*-alkyl sulfates of varying chain length (*rac*-**1a-8a**) were chosen with special emphasis on the relative position of the sulfate ester moiety ranging from (ω-1) to (ω-3) (Scheme 1). The latter substrates bearing an 'internal' (ω-3)-functionality are difficult to tackle with traditional lipase-[14] or carbonyl reductase-technology.[15] From a synthetic viewpoint, substrates bearing lipophilic alkene- or alkyne-functionalities (*rac*-**9a-20a**) allow access to valuable building blocks via functional group

manipulation via (stereo)selective redox-reactions. Finally, two stereochemically sensitive benzylic sulfate esters were selected (*rac*-**23a**, *rac*-**24a**). Since Pisa1 acts via a single pathway through strict inversion of configuration, as proven by ^{18}O-labeling experiments,[12] *E*-values can be applied for the description of the enantioselectivity.[16] Since the formed product (a lipophilic alcohol) and the non-reacted (water-soluble) sulfate ester enantiomers have extremely different characteristics with respect to their relative solubilities, *E*-values were calculated from the enantiomeric excess of substrate and product (ee$_S$, ee$_P$), because the latter are relative

values and are thus (almost) independent of sample manipulations.[17,18] The overall procedure was as follows (Scheme 1): In a first step, racemic substrates were subjected to enzymatic hydrolysis in Tris-buffer at the pH optimum of 8.2. After the conversion came close to mid-point (depending on the reactivity of the substrate 2-72h), the formed alcohol was separated from the nonreacted sulfate ester by extraction. The latter was hydrolysed under acidic conditions, which proceeds with strict retention of configuration.[19] Deracemization experiments were performed by linking both hydrolysis steps in a sequential fashion (without

separation of intermediates) in a one-pot procedure using substrate *rac*-**4a**. The absolute configuration of products was elucidated by co-injection of reference samples of alcohols **1b**, **2b**, **5b**, **7b**,[20] **9b**, **10b**, **11b**,[21] **12b**, **13b**,[21] **17b**,[22] **20b**,[23] **21b**,[24] **23b** and **24b**, which were obtained commercially or were prepared via independent synthesis according to literature. The absolute configuration of compounds **3b**,**14b** and **15b** was determined via optical rotation. Alcohols **16b**, **18b** and **19b** were correlated via catalytic hydrogenation yielding **2b**, 3-hexanol and **7b**, respectively. Compounds **4b**, **6b**, **8b** and **22b** were elucidated as recently reported.[12]

Table 1. Kinetic resolution of alkyl sulfates *rac*-**1a**-**24a** using the inverting *sec*-alkylsulfatase Pisa1

Substrate	t [h]	Cosolvent	Conv.[%][a]	Ee$_P$	Ee$_S$	Config.[b]	E-Value[a]
rac-**1a**	24	none	34	83	42	(*S*)	16
rac-**2a**	24	none	39	98	62	(*S*)	155
rac-**3a**	24	none	50	>99	>99	(*S*)	>200
rac-**4c**[c]	2	none	50	>99	>99	(*S*)	>200
rac-**5a**	24	none	50	>99	>99	(*S*)	>200
rac-**6a**[c]	6	none	50	>99	>99	(*S*)	>200
rac-**7a**	6	none	34	96	49	(*S*)	87
rac-**8a**[c]	24	none	57	60	80	(*S*)	10
rac-**9a**	24	none	7	94	7	(*S*)	34
rac-**10a**	24	none	50	>99	>99	(*S*)	>200
rac-**11a**	6	none	35	83	46	(*R*)[d]	17
rac-**11a**	8	DMSO	45	88	74	(*R*)[d]	36
rac-**12a**	6	none	48	90	87	(*R*)[d]	102
rac-**12a**	8	DMSO	47	99	87	(*R*)[d]	>200
rac-**13a**	6	none	50	94	94	(*R*)[d]	125
rac-**13a**	8	DMSO	49	99	95	(*R*)[d]	>200
rac-**14a**	6	none	50	95	96	(*R*)[d]	149
rac-**14a**	8	DMSO	46	99	83	(*R*)[d]	>200
rac-**15a**	24	none	44	74	58	(*S*)	12
rac-**16a**	72	none	4	>99	4	(*S*)	>200
rac-**17a**	6	none	49	93	89	(*S*)	80
rac-**17a**	6	DMSO	47	98	89	(*S*)	>200
rac-**18a**	4	none	57	55	80	(*S*)	8
rac-**19a**	6	none	56	32	40	(*S*)	3
rac-**20a**	24	none	23	91	27	(*R*)[d]	28
rac-**21a**	24	none	24	>99	31	(*R*)[d]	>200
rac-**22a**[c]	24	none	50	>99	>99	(*R*)[d]	>200
rac-**23a**[e]	6	none	n.d	33	n.d	(*S*)	n.d
rac-**23a**[e]	6	DMSO	n.d	81	n.d.	(*S*)	n.d
rac-**24a**	72	none	10	>99	11	(*S*)	>200
rac-**24a**	72	DMSO	13	>99	14	(*S*)	>200

[a] Calculated from ee$_S$ and ee$_P$;[18]

[b] since the enzymatic hydrolysis of the preferred (*R*)-enantiomer proceeds with strict inversion of configuration, the alcohol formed and the non-reacted sulfate ester possess the same absolute configuration;

[c] data taken from ref. [12];

[d] switch in CIP priority;

[e] not determined due to competing (non-enzymatic) autohydrolysis.

FULL PAPERS

The substrate-selectivity pattern revealed several clear trends (Table 1). Overall, substrates bearing the sulfate ester grup in the (ω-1)-position show that the enantioselectivity strongly depends on the relative size of both alkyl substituents: With increasing chain length or branching, E-values steadily climbed from **1a** ($E = 16$ via **2a** ($E = 155$) to **3a-5a** ($E > 200$). Analogous observations were made when the functional group was gradually moved towards the center of the molecule: Whereas the (ω-2)-substrate **6a** was nicely resolved, compounds **7a** and **8a** showed reduced stereoselectivities due to the decreasing size-difference of both alkyl substituents. Similar trends can be deduced from substrates bearing a terminal olefinic group (**9a**, **10a**).

When the position of the alkene unit was switched to yield allylic sulfate esters **11a-14a**, E-values were surprisingly modest ($E = 17$ to 149). In order to enhance the stereoselectivity of the reaction by suppressing the background of slow (non-enzymatic) autohydrolysis of the substrate,[25] a range of organic co-solvents were tested using rac-**12a** as test substrate (Fig. 1). Whereas water-immiscible co-solvents, such as toluene, and (cyclo)hexane had little effect on the catalytic performance of Pisa1, t-BuOMe, ethyl acetate and chlorinated hydrocarbons caused a drop in activity and selectivity. In contrast, some water-miscible solvents, such as ethylene glycol, DMSO, 1,2-dimethoxyethane and methanol led to an increase of selectivity (ee_P from 90% to >99%) at a small cost of relative activity (~10%). Other solvents caused a dramatic loss of activity. Based on these data, DMSO (20% v:v) turned out to be the best selectivity-enhancer by roughly doubling E-values of substrates **11a-14a**. The exact molecular reason for the selectivity-enhancing effects exerted by DMSO can be attributed to suppression of spontaneous (non-enzymatic) hydrolysis and/or alteration of the catalytic properties of the enzyme, as observed for an alkyl sulfatase from *Rhodococcus ruber* DSM 44541.[26]
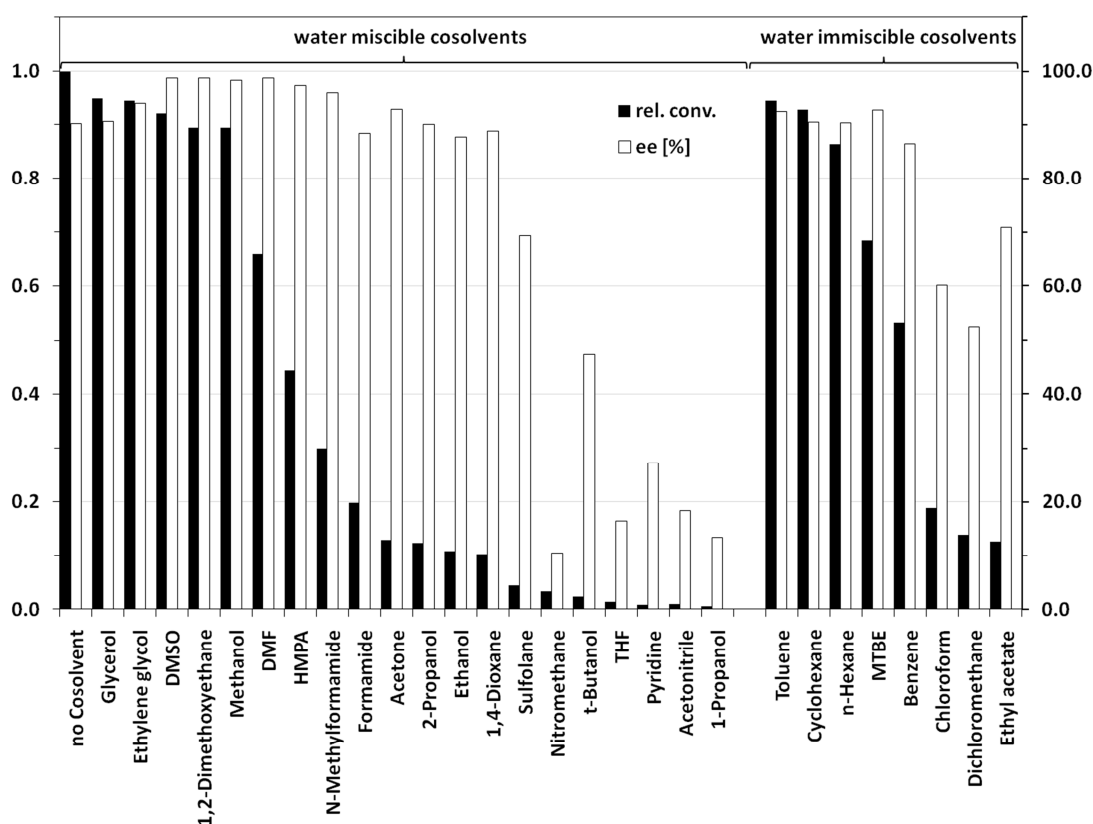
Figure 1. Relative conversion and enantioselectivity of Pisa1 for rac-**12a** in presence of organic cosolvents. Reaction conditions: rac-**12a** (5 mg, 22 mM), Pisa1 (20 µL, 130 µg, 1.8 µM), Tris-HCl buffer (100 mM, pH 8.0, 20% v/v organic cosolvent), reaction volume: 1000 µL, 30 °C, 120 rpm, 6 h.

FULL PAPERS

With (ω-1)-substrates bearing an internal alkyne unit (**15a**-**17a**) the relative position of the acetylene moiety had a strong impact on the selectivity (**15a**, **16a**) as well as the size-difference of substituents (**18a**, **19a**). The selectivity of **17a** could be drastically increased from E = 80 to >200. In contrast to more flexible allylic substrates, rigid propargylic sulfate esters (**21a**, **22a**) were resolved with excellent selectivity, as long as both substituents differed in size. Unfortunately, the benzylic sulfate **23a** proved to be hydrolytically unstable and showed spontaneous (non-enzymatic) hydrolysis, which was assumed to proceed via an S_N1-mechanism involving a benzylic carbenium ion derived by departure of the good leaving group HSO_4^-. In order to prove this hypothesis, electron-withdrawing CF_3-substituents (which impede resonance-stabilization) were added. Indeed, substrate **24a** turned out to be perfectly stable and thus could be resolved with excellent enantioselectivity.

Temperature-stability tests using substrate *rac*-**4a** with pre-incubated Pisa1 revealed that the enzyme was very stable at 20-30 °C (24 h incubation), which caused only a small loss of activity (≤10%). However, incubation at 40 °C caused an activity loss of ~20% and at 50 °C significant enzyme deactivation was observed. The latter was underlined by detailed kinetic studies performed at 20, 30 and 40 ℃, which revealed a deterioration of catalytic parameters (rising K_M, decreasing k_{cat}) at elevated temperatures. A melting point of 57 ℃ was determined for Pisa1 by CD-measurements.

Conclusions

Investigation of the substrate spectrum of the inverting alkylsulfatase Pisa1 revealed that a wide variety of *sec*-alkylsulfate esters were accepted. Aromatic, olefinic and acetylenic moieties, which allow further functional group manipulation, were nicely tolerated. Perfect enantioselectivities were obtained for substrates bearing groups of different size adjacent to the sulfate ester moiety. For stereochemically 'difficult' substrates possessing

substituents of similar size, insufficient selectivities could be significantly enhanced by using DMSO as co-solvent. Hydrolytically labile benzylic sulfate esters, which are impeded by background-hydrolysis, could be sufficiently stabilised by introduction of electron-withdrawing groups. Overall, Pisa1 proved to be a very useful inverting alkylsulfatase for the deracemisation of *rac-sec*-alcohols via enzymatic hydrolysis of their corresponding sulfate esters, which furnishes homochiral products. Although Pisa1 exhibits the same stereochemical preference for *sec*-alkyl sulfate esters as lipases for *sec*-alcohol carboxylic esters (cf. Kazlauskas-rule [27]), anti-Kazlauskas products, which are accessible by using subtilising-based dynamic kinetic resolution,[28] are formed via inversion of configuration.

Experimental

General Remarks

Competent cells One Shot® TOP10 and One Shot® BL21 Star™ (DE3) were obtained from Invitrogen and transformed according to the protocol of the supplier. Alcohols *rac*-**1b**-**24b**, the respective enantiopure alcohol references and all other reagents were purchased from Sigma Aldrich, Acros, Carl Roth and Alfa Aesar. NMR spectra were recorded on a Bruker spectrometer at 300 (^{1}H) and 75 (^{13}C) MHz. Shifts ($\square$) are given in ppm and coupling constants (*J*) are given in Hz.

Synthesis of alkyl sulfate esters.

Racemic sulfate esters **1a-3a**, **5a**, **7a**, **9a-21a**, **23a** and **24a** were prepared from the corresponding alcohols **1b-3b**, **5b**, **7b**, **9b-21b**, **23b** and **24b** by using NEt_3*SO_3 following a known procedure[12] with the following modification: NaH (60% suspension in paraffin oil) was added directly in equimolar amounts.

FULL PAPERS *Advanced* Synthesis & Catalysis

Cloning, expression and purification of alkylsulfatase Pisa1 from *Pseudomonas sp.* DSM6611.

Pisa1 was obtained according to a known procedure.[12]

Screening for Activity and Determination of Stereoselectivity

The activity of Pisa1 was tested using alkyl sulfate esters *rac*-**1a-3a**, *rac*-**5a**, *rac*-**7a**, *rac*-**9a-21a**, *rac*-**23a** and *rac*-**24a** (5 mg) as previously described.[12] Depending on the relative activity, reaction times of 2-72 h were applied.

For the calculation of the conversion and E-values from e.e.$_P$ and e.e.$_S$ (after acid-catalysed hydrolysis of the non-reacted sulfate ester with retention of configuration yielding the corresponding alcohol) a previously described two step protocol[12] was employed with the following modifications:

Step1 (enzymatic hydrolysis): Alkyl sulfate (**1a-24a**, 25 mg) was dissolved in Tris-HCl (4.9 mL, 100 mM, pH 8.2) and a solution of purified enzyme (0.65 mg, 8.85 nmol protein in 0.1 mL) was added. The reaction mixture was shaken at 30 °C and 120 rpm for 2-72 h. After extraction with ethyl acetate (2 mL), an aliquot of the organic layer (1 mL) was dried over anhydrous sodium sulfate, derivatized to the corresponding acetate **1c-24c**, which was analyzed by GC using a chiral stationary phase. The aqueous phase was washed twice with ethyl acetate (3 mL) to remove traces of alcohol and was lyophilized overnight. When DMSO was used as a cosolvent, the extraction was repeated 3 times with ethyl acetate (5 mL) to remove residual traces of DMSO before lyophilisation.

Step 2 (acidic hydrolysis):[19] The lyophilisate was dissolved in a mixture of methyl *tert*-butyl ether/deionized water (10 mL, 97:3) and 1,4-dioxane (5 µL, 0.06 mmol) and *p*-toluenesulfonic acid monohydrate (170 mg, 0.9 mmol) were added. The mixture was stirred under reflux at 40 °C for 2 h. After cooling to room temperature, saturated NaHCO$_3$ solution (5 mL) was added to quench the reaction. After phase separation the organic phase

was dried over anhydrous sodium sulfate. The solvent was evaporated under reduced pressure, the remaining alcohol was redissolved in ethyl acetate (1 mL), derivatized as acetate[12] and analyzed by GC using a chiral stationary phase.

Enantio-convergent one-pot two-step hydrolysis of *rac*-2-octyl sulfate (**4a**)

rac-2-Octyl sulfate (**4a**, 10 mg) was hydrolyzed using Pisa1 (0.19 mg) in Tris/HCl buffer (0.1 M, pH 8.0, 2 mL) by shaking at 30 °C and 120 rpm for 6 h in a closed glass vial. Methyl *tert*-butyl ether (2 mL), *p*-TosOH (60 mg) and 1-octanol (2 mg, as internal standard) were added and the mixture was incubated for 24h with shaking at 40 °C or 60 °C, resp. After cooling to 0 °C, sat. NaHCO$_3$ solution was added (1mL) to neutralize the acid, the mixture was centrifuged, the organic layer was separated, dried (Na$_2$SO$_4$) and derivatized as acetate as described above. GC-analysis showed that the relative accuracy of the conversion from double experiments was ±7% using a calibration curve. Alcohol (*S*)-**4b** was obtained as the sole product (average values from double experiments): (i) T = 40 °C, c = 78%, e.e. 97.5%; (ii) T = 60 °C, c = 93%, e.e. 98.7%.

The use of microwave heating for the acid-catalyzed hydrolysis step could not be employed due to significant racemisation: (*R*)-2-octyl sulfate (**4a**) in Tris/HCl buffer (0.1 M, pH 8.0, 1 mL) containing *p*-TosOH (30 mg) was placed in a sealed vial and heated for 15 min to 100 °C, 110 °C, 120 °C and 180 °C, respectively. After cooling to 0 °C, EtOAc (1 mL) containing 1-octanol (1 mg, as internal standard) was added. The organic layer was separated, dried (Na$_2$SO$_4$) and derivatized as acetate as described above. GC-analysis revealed that the enantiomeric purity of (*R*)-**4b** was significantly diminished. 100 °C: e.e. 44%; 110 °C: e.e. 36%; 120 °C: e.e. 39%; 180 °C: e.e. 49%, respectively.

Electronic Supporting Information

NMR data and yields of substrates, determination of enantiomeric excess and absolute configuration of products, GC-analytics, optical rotation values,

FULL PAPERS Advanced Synthesis & Catalysis

detailed procedures for cosolvent-, temperature-stability- and pH-studies and determination of protein concentration are given in the electronic supporting information.

Acknowledgements

References

[1] a) N. J. Turner, *Curr. Opin. Chem. Biol.* **2010**, *14*, 115-121; b) D. Tessaro, G. Molla, L. Pollegioni, S. Servi, in: *Modern Biocatalysis: Stereoselective and Environmentally Friendly Reactions*, W.-D. Fessner, T. Anthonsen (eds.), Wiley-VCH, Weinheim, **2009**, pp. 195-228; c) J. Steinreiber, K. Faber, H. Griengl, *Chem. Eur. J.* **2008**, *14*, 8060-8072; d) B. Martin-Matute, J.-E. Bäckvall, in: *Organic Synthesis with Enzymes in Non-Aqueous Media*, G. Carrea, S. Riva (eds.), Wiley-VCH, Weinheim, **2008**, pp. 113-144; e) K. Faber, *Chem. Eur. J.* **2001**, *7*, 5004-5010.

[2] a) M.-J. Kim, Y. Ahn, J. Park, *Curr. Opinion Biotechnol.* **2002**, *13*, 578-587; b) B. Martin-Matute, J.-E. Bäckvall, *Curr. Opin. Chem. Biol.* **2007**, *11*, 226-232; c) J. H. Lee, K. Han, M.-J. Kim, J. Park, *Eur. J. Org. Chem.* **2010**, 999-1015; d) Y. Kim, J. Park, M.-J. Kim, *ChemCatChem* **2011**, *3*, 271-277; e) H. Pellissier, *Adv. Synth. Catal.* **2011**, *353*, 659-676.

[3] a) C. C. Gruber, I. Lavandera, K. Faber, W. Kroutil, *Adv. Synth. Catal.* **2006**, *348*, 1789-1805; b) I. Fotheringham, I. Archer, R. Carr, R. Speight, N. J. Turner, *Biochem. Soc. Trans* **2006**, *34*, 287-290; c) A. J. Carnell, *Adv. Biochem. Eng. Biotechnol.* **1999**, *63*, 57-72.

[4] a) P. Gadler, S. M. Glueck, W. Kroutil, B. M. Nestl, B. Larissegger-Schnell, B. T. Ueberbacher, S. R. Wallner, K. Faber, *Biochem. Soc. Trans* **2006**, *34*, 296–300; b) S. R. Wallner, M. Pogorevc, H. Trauthwein, K. Faber, *Eng. Life Sci.* **2004**, *4*, 512-516.

[5] a) D. B. Janssen, in: *Biocatalysis in the Pharmaceutical and Biotechnology Industry,* R. N. Patel (ed.), **2007**, CRC Press, Boca Raton, pp. 441-462; b) T. Kurihara, N. Esaki, K. Soda, *J. Mol. Catal. B: Enzym.* **2000**, *10*, 57-65.

[6] a) E. J. de Vries, D. B. Janssen, *Curr. Opin. Biotechnol.* **2003**, *14*, 414-420; b) A. Steinreiber, K. Faber, *Curr. Opin. Biotechnol.* **2001**, *12*, 552-558.

[7] a) P. Gadler, K. Faber, *Trends Biotechnol.* **2007**, *25*, 83-88; b) S. R. Hanson, M. D. Best, C.-H. Wong, *Angew. Chem. Int. Ed.* **2004**, *43*, 5736-5763.

[8] a) M. A. Kertesz, *FEMS Microbiol. Rev.* **2000**, *24*, 135-175; b) P. Bojarova, S. J. Williams, *Curr. Opin. Chem. Biol.* **2008**, *12*, 573-581.

[9] I. Boltes, H. Czapinska, A. Kahnert, R. von Bülow, T. Dierks, B. Schmidt, K. von Figura, M. A. Kertesz, I. Uson, *Structure* **2001**, *9*, 483-491.

[10] a) K. S. Dodgson, G. F. White, J. W. Fitzgerald, *Sulfatases of Microbial Origin*, **1982**, CRC Press, Boca Raton; b) B. Bartholomew, K. S. Dodgson, G. W. J. Matcham, D. J. Shaw, G. F. White, *Biochem. J.* **1977**, *165*, 575-580; c) G. F. White, *Appl. Microbiol. Biotechnol.* **1991**, *35*, 312-316; d) M. Pogorevc, K. Faber, *Appl. Environ. Microbiol.* **2003**, *69*, 2810-2815; e) P. Gadler, T. C. Reiter, K. Hoelsch, D. Weuster-Botz, K. Faber, *Tetrahedron: Asymmetry* **2009**, *20*, 115-118; f) P. Gadler, K. Faber, *Eur. J. Org. Chem.* **2007**, 5527-5530.

[11] G. Hagelueken, T. M. Adams, L. Wiehlmann, U. Widow, H. Kolmar, B. Tümmler, D. W. Heinz, W.-D. Schubert, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 7631-7636.

[12] a) M. Schober, P. Gadler, T. Knaus, H. Kayer, R. Birner-Grünberger, C. Gülly, P. Macheroux, U. Wagner, K. Faber, *Org. Lett.* **2011**, *13*, 4296-4299.

[13] European Nucleotide Archive sequence accession code : FR850678; PDB code of the crystal structure (2.7 Å resolution): 2YHE.

[14] U. T. Bornscheuer, R. J. Kazlauskas, *Hydrolases for Organic Synthesis*, 2nd ed., Wiley-VCH, **2006**, Weinheim.

[15] K. Faber, S. M. Glueck, B. Seisser, W. Kroutil, in: *Enzyme Technologies: Metagenomics, Evolution, Biocatalysis and Biosynthesis*, W.-K. Yeh, H.-C. Yang, J. R. McCarthy (eds.), Wiley, **2010**, pp. 199-248.

[16] a) A. J. J. Straathof, J. A. Jongejan, *Enzyme Microb. Technol.* **1997**, *21*, 559-571; b) C.-S. Chen, Y. Fujimoto, G. Girdaukas, C. J. Sih, *J. Am. Chem. Soc.* **1982**, *104*, 7294-7299.

[17] In contrast, conversion (c) is an absolute value, which is much more sensitive to errors derived from sample manipulation, such as extraction, etc.; see:K. Faber, *Biotransformations in Organic Synthesis*, 6th edn., Springer, Heidelberg, **2011**, pp. 38-43.

[18] J. L. L. Rakels, A. J. J. Straathof, J. J. Heijnen, *Enzyme Microb. Technol.* **1993**, *15*, 1051-1056.

[19] S. R. Wallner, B. Nestl, K. Faber, *Tetrahedron* **2005**, *61*, 1517-1521.

[20] H. Hirata, M. Ikushima, M. Watanabe, K. Kawauchi, M. Miyagishi, Y.-W. Chen, H. Yanagishita, *J. Oleo. Sci.*, **2003**, *52*, 99-108.

[21] F. D. Özdemirhan, M. Celik, S. Atli, C. Tanyeli, *Tetrahedron: Asymmetry*, **2006**, *17*, 287-291.

[22] X. Zhang, Z. Lu, C. Fu, S. Ma, *Org. Biomol. Chem.*, **2009**, *7*, 3258–3263.

[23] J. L. Abad, F. Camps, G. Fabriàs, *J. Am. Chem. Soc.* **2007**, *129*, 15007-15012.

FULL PAPERS

Advanced
Synthesis &
Catalysis
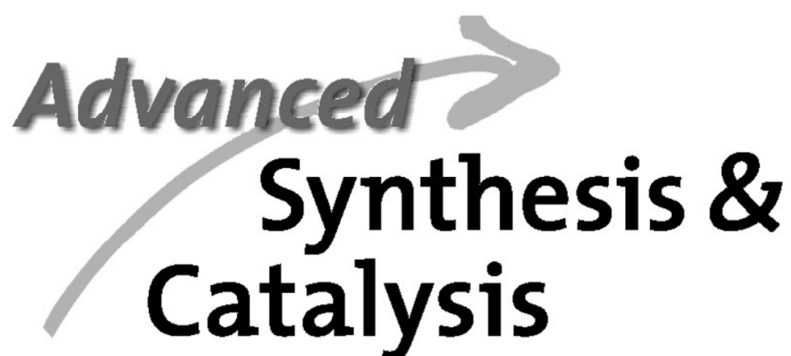
[24]   F. A. Marques, A. Zimmermann, C. E. Delay, B.H.L.N. Sales Maia, *Lett.Org. Chem.*, **2008**, *5*, 69-72.

[25]   K. Faber, M. Mischitz and A. Kleewein, In: *Preparative Biotransformations*, S. M. Roberts (ed.), Wiley, New York, 1997, pp. 0:079-0:084.

[26]   M. Pogorevc, U. T. Strauss, T. Riermeier, K. Faber, *Tetrahedron: Asymmetry* **2002**, *13*, 1443-1447.

[27]   R. J. Kazlauskas, A. N. E. Weissfloch, A. T. Rappaport, L. A. Cuccia, *J. Org. Chem.* **1991**, *56*, 2656-2665; Q. Jing, R. J. Kazlauskas, *Chirality* **2008**, *20*, 724-735.

[28]   a) M.-J. Kim, Y. I. Chung, Y. K. Choi, H. K. Lee, D. Kim, J. Park, *J. Am. Chem. Soc.* **2003**, *125*, 11494-11495; b) L. Boren, B. Martin-Matute, Y. Xu, A. Cordova, J.-E. Bäckvall, *Chem. Eur. J.* **2006**, *12*, 225-232; c) M.-J. Kim, H. K. Lee, J. Park, *Bull. Korean Chem. Soc.* **2007**, *28*, 2096-2098.

Chapter 7

Supplementary Information

The substrate spectrum of the inverting *sec*-alkylsulfatase Pisa1

The Substrate Spectrum of the Inverting *sec*-Alkylsulfatase Pisa1

Markus Schober[1], Tanja Knaus[2], Michael Toesch[1], Peter Macheroux[2], Ulrike Wagner[3] and Kurt Faber[1]*[1]

[1]Department of Chemistry, Organic & Bioorganic Chemistry, University of Graz, Heinrichstrasse 28, A-8010 Graz, Austria;

[2]Institute of Biochemistry, Graz University of Technology, Petersgasse 12, A-8010 Graz,

[4]Institute of Molecular Biosciences, Structural Biology, University of Graz, Humboldtstrasse 50/III, A-8010 Graz, Austria

Supporting Information

* to whom correspondence should be addressed: Tel.: +43-316-380-5332; fax: +43-316-380-9840; email: Kurt.Faber@Uni-Graz.at

NMR Data and yields for alkyl sulfates

Synthesis of substrates **4a**, **6a**, **8a** and **22a** has been described elsewhere.[1]

***rac*-2-Pentyl sulfate (1a):** ^{1}H NMR (300 MHz, D$_2$O): δ = 4.43-4.32 (m, 1H), 1.55-1.13 (m, 7H), 0.79 (t, *J* = 5.7 Hz, 3H); ^{13}C NMR (75 MHz, D$_2$O): δ = 78.5, 38.2, 19.9, 17.9, 13.1; 83% yield.

***rac*-2-Hexyl sulfate (2a):** ^{1}H NMR (300 MHz, D$_2$O): δ = 4.45-4.34 (m, 1H), 1.62-1.41 (m, 2H), 1.33-1.15 (m, 7H), 0.78 (t, *J* = 5.6 Hz, 3H); ^{13}C NMR (75 MHz, D$_2$O): δ = 78.8, 35.7, 26.7, 21.8, 19.9, 13.2; 78% yield.

***rac*-5-Methyl-2-hexyl sulfate (3a):** ^{1}H NMR (300 MHz, D$_2$O): δ = 4.42-4.30 (m, 1H), 1.60-1.36 (m, 3H), 1.24-1.07 (m, 5H), 0.78-0.73 (m, 6H); ^{13}C NMR (75 MHz, D$_2$O): δ = 79.1, 33.8, 33.6, 27.1, 21.8, 21.7, 19.9; 80% yield.

***rac*-2-Decyl sulfate (5a):** ^{1}H NMR (300 MHz, D$_2$O): δ = 4.47-4.36 (m, 1H), 1.67-1.41 (m, 2H), 1.36-1.14 (m, 15H), 0.85-0.78 (m, 3H); ^{13}C NMR (75 MHz, D$_2$O): δ = 78.4, 36.2, 31.5, 29.0, 29.0, 28.8, 24.7, 22.3, 20.0, 13.6; 66% yield.

***rac*-3-Heptyl sulfate (7a):** ^{1}H NMR (300 MHz, D$_2$O): δ = 4.28-4.19 (m, 1H), 1.66-1.47 (m, 4H), 1.29-1.16 (m, 4H), 0.84-0.73 (m, 6H); ^{13}C NMR (75 MHz, D$_2$O): δ = 83.6, 32.6, 26.5, 26.4, 21.9, 13.2, 8.4; 72% yield.

***rac*-4-Penten-2-yl sulfate (9a):** ^{1}H NMR (300 MHz, D$_2$O): δ = 5.89-5.73 (m, 1H), 5.15-5.05 (m, 2H), 4.55-4.43 (m, 1H), 2.37-2.31 (m, 2H), 1.26 (d, *J* = 8.0 Hz, 3H); ^{13}C NMR (75 MHz, D$_2$O): δ = 133.9, 117.9, 77.4, 40.2, 19.5; 67% yield.

***rac*-5-Hexen-2-yl sulfate (10a):** ^{1}H NMR (300 MHz, D$_2$O): δ = 5.85-5.72 (m, 1H), 5.03-4.84 (m, 2H), 4.44-4.34 (m, 1H), 2.10-1.98 (m, 2H), 1.69-1.50 (m, 2H), 1.22 (d, *J* = 7.3 Hz, 3H); ^{13}C NMR (75 MHz, D$_2$O): δ = 138.6, 114.7, 78.0, 35.2, 28.8, 19.9; 82% yield.

***rac*-1-Hepten-3-yl sulfate (11a):** ^{1}H NMR (300 MHz, DMSO-d6): δ = 5.84-5.72 (m, 1H), 5.14-4.98 (m, 2H), 4.47 (q, *J* = 27.2 Hz, 1H), 1.60-1.44 (m, 2H), 1.31-1.19 (m, 4H), 0.85 (t, *J* = 5.9 Hz, 3H); ^{13}C NMR (75 MHz, DMSO-d6): δ = 139.8, 114.8, 76.8, 34.7, 27.1, 22.6, 14.5; 77% yield.

***rac*-1-Octen-3-yl sulfate (12a):** ^{1}H NMR (300 MHz, DMSO-d6): δ = 5.82-5.69 (m, 1H), 5.14-4.98 (m, 2H), 4.47 (q, *J* = 26.8 Hz, 1H), 1.61-1.39 (m, 2H), 1.31-1.14 (m, 6H), 0.82 (t, *J* = 5.4 Hz, 3H); ^{13}C NMR (75 MHz, DMSO-d6): δ = 139.3, 115.2, 77.3, 34.8, 31.6, 24.4, 22.5, 14.3; 76% yield.

***rac*-1-Nonen-3-yl sulfate (13a):** ^{1}H NMR (300 MHz, DMSO-d6): δ = 5.84-5.72 (m, 1H), 5.15-4.98 (m, 2H), 4.47 (q, 27.2, 1H), 1.60-1.40 (m, 2H), 1.32-1.15 (m, 8H), 0.85 (t, *J* = 5.8 Hz, 3H); ^{13}C NMR (75 MHz, DMSO-d6): δ = 139.8, 114.8, 76.8, 35.0, 31.8, 29.2, 24.8, 22 .5, 14.4; 74% yield.

***rac*-1-Decen-3-yl sulfate (14a):** ^{1}H NMR (300 MHz, DMSO-d6): δ = 5.84-5.72 (m, 1H), 5.15-4.98 (m, 2H), 4.47 (q, *J* = 27.4 Hz, 1H), 1.59-1.41 (m, 2H), 1.32-1.16 (m, 10H), 0.85 (t, *J* = 5.8

Hz, 3H); ^{13}C NMR (75 MHz, DMSO-d6): δ = 139.8, 114.7, 76.8, 35.0, 31.7, 29.5, 29.2, 24.8, 22.6, 14.4; 79% yield.

rac-3-Hexyn-2-yl sulfate (15a): ^{1}H NMR (300 MHz, D$_2$O): δ = 5.00-4.92 (m, 1H), 2.19-2.09 (m, 2H), 1.41 (d, *J* = 9.4 Hz, 3H), 1.00 (t, *J* = 7.5 Hz, 3H); ^{13}C NMR (75 MHz, D$_2$O): δ = 89.4, 77.9, 66.6, 22.1, 12.7, 11.7; 65% yield.

rac-4-Heptyn-2-yl sulfate (16a): ^{1}H NMR (300 MHz, D$_2$O): δ = 4.50-4.39 (m, 1H), 2.50-2.33 (m, 2H), 2.13-2.03 (m, 2H), 1.27 (d, *J* = 8.0 Hz, 3H), 0.98 (t, *J* = 7.4 Hz, 3H); ^{13}C NMR (75 MHz, D$_2$O): δ = 85.5, 75.5, 75.3, 25.8, 19.1, 13.2, 11.6; 65% yield.

rac-4-Phenyl-3-butyn-2-yl sulfate (17a): ^{1}H NMR (300 MHz, D$_2$O): δ = 7.46-7.22 (m, 5H), 5.19 (q, *J* = 38.8 Hz, 1H), 1.51 (d, *J* = 10.0 Hz, 3H); ^{13}C NMR (75 MHz, D$_2$O): δ = 131.6, 129.1, 128.5, 121.3, 87.2, 85.8, 66.4, 21.7; 62 % yield.

rac-4-Hexyn-3-yl sulfate (18a): ^{1}H NMR (300 MHz, D$_2$O): δ = 4.85-4.77 (m, 1H), 1.82-1.68 (m, 5H), 0.92 (t, *J* = 6.9 Hz, 3H); ^{13}C NMR (75 MHz, D$_2$O): δ = 84.5, 76.6, 71.5, 28.9, 8.5, 2.6; 65% yield.

rac-4-Heptyn-3-yl sulfate (19a): ^{1}H NMR (300 MHz, D$_2$O): δ = 4.88-4.80 (m, 1H), 2.25-2.14 (m, 2H), 1.79-1.72 (m, 2H), 1.06 (t, *J* = 8.0 Hz, 3H), 0.94 (t, *J* = 7.2 Hz, 3H); ^{13}C NMR (75 MHz, D$_2$O): δ = 90.2, 76.9, 71.4, 29.0, 12.9, 11.7, 8.6; 65% yield.

rac-1-Hexyn-3-yl sulfate (20a): ^{1}H NMR (300 MHz, D$_2$O): δ = 4.86 (dt, *J* = 32.1 and 8.3 Hz, 1H), 2.88 (d, *J* = 5.2 Hz, 1H), 1.76-1.67 (m, 2H), 1.45-1.32 (m, 2H), 0.83 (t, *J* = 6.2 Hz, 3H); ^{13}C NMR (75 MHz, D$_2$O): δ = 81.5, 75.6, 69.1, 37.2, 17.7, 12.7; 86% yield.

rac-5-Methyl-1-hexyn-3-yl sulfate (21a): ^{1}H NMR (300 MHz, D$_2$O): δ = 4.68 (dt, *J* = 33.7 and 9.8 Hz, 1H), 3.27 (d, *J* = 6.9 Hz, 1H), 1.84-1.55 (m, 2H), 1.49-1.39 (m, 1H), 0.92-0.82 (m, 6H); ^{13}C NMR (75 MHz, D$_2$O): δ = 84.4, 75.5, 64.8, 45.1, 24.5, 23.1, 22.6; 86% yield.

rac-1-Phenylethyl sulfate (23a): ^{1}H NMR (300 MHz, DMSO-d6): δ = 7.38-7.17 (m, 5H), 5.24-5.16 (q, *J* = 34.0 Hz, 1H), 1.43 (d, *J* = 9.4 Hz, 3H); ^{13}C NMR (75 MHz, DMSO-d6): δ = 144.1, 128.3, 127.2, 126.3, 73.9, 23.8; 74% yield.

rac-1-[3,5-Bis(trifluoromethyl)phenyl]ethyl sulfate (24a): ^{1}H NMR (300 MHz, DMSO-d6): δ = 8.02 (s, 2H), 7.96 (s, 1H), 5.45-5.37 (q, *J* = 35.7 Hz, 1H), 1.46 (d, *J* = 9.6 Hz, 3H); ^{13}C NMR (75 MHz, DMSO-d6): δ = 148.3, 131.0, 130.6, 130.1, 129.7, 127.1, 125.7, 122.1, 72.6, 23.9; 43% yield.

Determination of enantiomeric excess of products

The enantiomeric excess of alcohols (**1b-3b**, **5b**, **7b**, **9b-21b**, **23b** and **24b**) was determined via GC after derivatization to the corresponding acetates (**1c-3c**, **5c**, **7c**, **9c-21c**, **23c** and **24c**) using an Agilent Technologies 7890A GC-FID system equipped with an Agilent Technologies 7683B autosampler and a Varian Chirasil Dex CB column (25 m x 0.32 mm x 0.25 µm film). The following methods were used:

Method A: Injector temperature 200° C, flow 2.0 mL/min He; temperature program: 80° C, hold for 1.0 min, 15° C/min to 110° C, 4° C/min to 130° C, 10° C/min to 180° C.

Method B: Injector temperature: 200° C, flow 2.0 mL/min He; temperature program: 80° C, hold for 1.0 min, 3° C/min to 100° C, 15° C/min to 150° C.

Method C: Injector temperature: 200° C, flow 2.0 mL/min He; temperature program: 80° C, hold for 1.0 min, 15° C/min to 140° C, 4° C/min to 160° C, 10° C/min to 180° C.

Method D: Injector temperature: 200° C, flow 2.0 mL/min He; temperature program: 80° C, hold for 1.0 min, 3° C/min to 95° C, 15° C/min to 150° C.

Method E: Injector temperature: 200° C, flow 2.0 mL/min He; temperature program: 70° C, hold for 1.0 min, 3° C/min to 85° C, 15° C/min to 150° C.

Table S1. GC-measurements.

Compound	Method	Retention time [min]	
2-pentyl acetate (**1c**)	E	3.5 (*S*)	4.5 (*R*)
2-hexyl acetate (**2c**)	D	3.8 (*S*)	4.7 (*R*)
5-methyl-2-hexyl acetate (**3c**)	B	5.0 (*S*)	5.9 (*R*)
2-octyl acetate (**4c**)[b]	A	5.3 (*S*)	5.9 (*R*)
2-decyl acetate (**5c**)	C	6.4 (*S*)	6.6 (*R*)
3-octyl acetate (**6c**)[b]	A	5.1 (*S*)	5.9 (*R*)
3-heptyl acetate (**7c**)	B	5.3 (*S*)	5.8 (*R*)
4-octyl acetate (**8c**)[b]	A	4.6 (*S*)	4.8 (*R*)
4-penten-2-yl acetate (**9c**)	E	3.6 (*S*)	4.3 (*R*)
5-hexen-2-yl acetate (**10c**)	D	4.0 (*S*)	4.9 (*R*)
1-hepten-3-yl acetate (**11c**)	B	5.2 (*R*)[a]	5.9 (*S*)[a]
1-octen-3-yl acetate (**12c**)	A	4.9 (*R*)[a]	5.2 (*S*)[a]
1-nonen-3-yl acetate (**13c**)	A	6.5 (*R*)[a]	6.8 (*S*)[a]
1-decen-3-yl acetate (**14c**)	C	6.1 (*R*)[a]	6.3 (*S*)[a]
3-hexyn-2-yl acetate (**15c**)	D	4.2 (*S*)	4.5 (*R*)
4-hexyn-2-yl acetate (**16c**)	D	6.9 (*S*)	7.3 (*R*)
4-phenyl-3-butyn-2-yl acetate (**17c**)	C	8.3 (*S*)	8.4 (*R*)
4-hexyn-3-yl acetate (**18c**)	D	4.9 (*S*)	5.2 (*R*)
4-heptyn-3-yl acetate (**19c**)	B	5.9 (*S*)	6.4 (*R*)
1-hexyn-3-yl acetate (**20c**)	D	4.3 (*R*)[a]	4.9 (*S*)[a]
5-methyl-1-hexyn-3-yl acetate (**21c**)	B	5.0 (*R*)[a]	5.5 (*S*)[a]
1-octyn-3-yl acetate (**22c**)[b]	A	5.5 (*R*)[a]	5.9 (*S*)[a]
1-phenylethyl acetate (**23c**)	A	6.6 (*S*)	7.2 (*R*)
1-[3,5-bis(trifluoromethyl) phenyl]ethyl acetate (**24c**)	A	4.5 (*S*)	4.8 (*R*)

[a] Switch in CIP priority; [b] data taken from ref. [1].

Determination of absolute configuration

The absolute configuration of products was elucidated by co-injection of commercial reference samples of alcohols (*S*)-**1b**, (*S*)-**2b**, (*R*)-**5b**, (*S*)-**9b**, (*S*)-**10b**, (*R*)- and (*S*)-**12b**, (*S*)-**23b** and (*R*)-**24b** after derivatization as acetates on GC. Reference samples of **7b**[2], **11b**[3], **13b**[3], **17b**[4], **20b**[5] and **21b**[6] were prepared according to literature. For the determination of optical rotation values, biotransformations were upscaled as previously described.[1] Compounds **3b** and **15b** were determined to be (*S*): (*S*)-**3b**: $[\alpha]_D^{22}$ = +11.1° (c = 1, CHCl$_3$); ref.[7]: (*R*)-**3b** was described as (+), (*S*)-**3b** was described as (-); (*S*)-**15b**: $[\alpha]_D^{22}$ = -22.1° (c = 1, CHCl$_3$); ref.[8]: (*R*)-**15b**: $[\alpha]_D^{25}$ = +27.2° (c = 0.58, Et$_2$O). Compound **14b** was determined to be (*R*) by optical rotation: $[\alpha]_D^{22}$ = -5.2° (c = 1, CHCl$_3$); ref.[9]: (*R*)-**14b**: $[\alpha]_D^{20}$ = -8.2° (c = 1.1, CHCl$_3$). Alcohol **16b** was reduced to the respective saturated alcohol **2b** with Pd/C under H$_2$ atmosphere and subsequently derivatized as acetate, GC-analysis revealed (*S*)-2c. Alcohol **18b** was reduced to yield 3-hexanol with Pd/C under an H$_2$ atmosphere after biotransformation. After subsequent derivatization as acetate, GC-analysis revealed (*S*)-configuration according to literature.[10] Alcohol **19b** was reduced to the respective saturated alcohol 7b with Pd/C under an H$_2$ atmosphere and subsequently derivatized as acetate. GC-analysis revealed (*S*)-configuration.[2]

Catalytic reduction of alcohols: Substrates **16a**, **18a** and **19a** were subjected to the chemo-enzymatic two-step protocol as described above. After extraction of the product alcohol with ethyl acetate (3 mL), 1 mL of the organic phase was processed as usual for the determination of the e.e. as described above. A second sample (1 mL) was hydrogenated using Pd/C (5 mg, 10% Pd) at ambient pressure. After 16 h the catalyst was removed by centrifugation. The product was derivatized as described before[1] and analyzed by chiral GC.

Cosolvent studies

Substrate *rac*-**12a** (5 mg, 22 µmol) was dissolved in Tris-HCl (780 µL, 100 mM, pH 8.0). Cosolvent [200 µL, 20% (v/v)] and an aliquot of Pisa1 enzyme solution (20 µL, 130 µg, 1.8 nmol) were added. The reaction mixture was shaken at 30 °C and 120 rpm for 6 h. Afterwards 1-octanol was added as an internal standard (2.5 mg, 19 µmol; dissolved in 1000 µL ethyl acetate) and the organic phase was dried over anhydrous sodium sulfate, after derivatization, the acetate ester **12c** was analyzed by chiral GC.[1]

Temperature stability studies

An aliquot of Pisa1 enzyme solution (20 µL, 130 µg, 1.8 nmol) was preincubated in Tris-HCl (880 µL, 100 mM, pH 8.0) at a fixed temperature for 1-24 h (Table S3). After the mixture reached room temperature, substrate *rac*-**4a** [5 mg, 22 µmol, dissolved in Tris-HCl (100 µL, 100 mM, pH 8.0)] was added. The reaction was shaken at 30 °C and 120 rpm for 60 min. Afterwards 1-octanol was added as internal standard (2.5 mg, 19 µmol; dissolved in 1000 µL ethyl acetate) and the organic phase was dried over anhydrous sodium sulfate, derivatized as acetate ester[1] and analyzed by chiral GC.
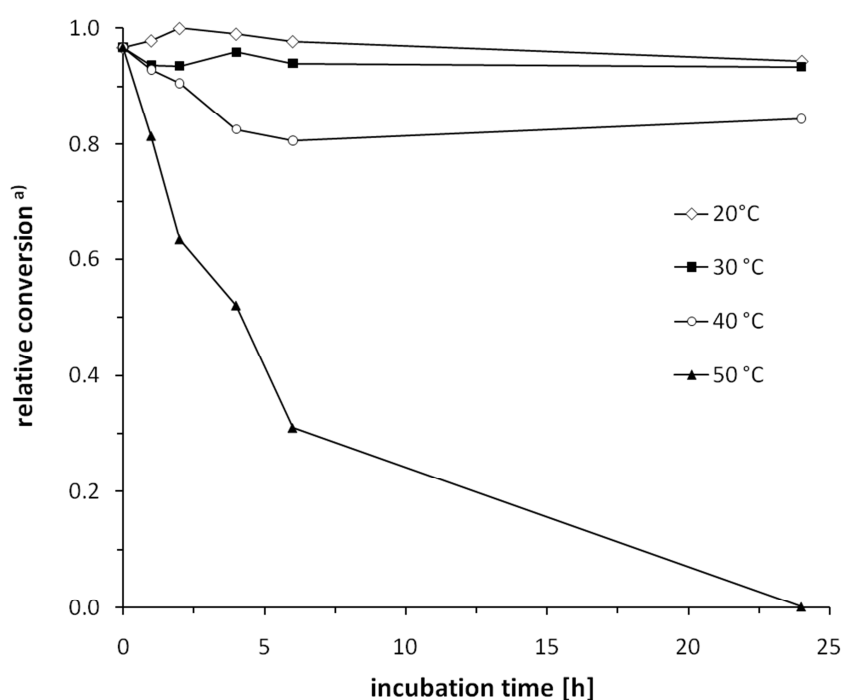
Fig. S1. Conversion of *rac*-**4a** with Pisa1 after heat-treatment of the enzyme. [a] Corresponding to that of the untreated enzyme.

Kinetic parameters for the hydrolysis of (*R*)-**4a** were determined using a VP-ITC system (MicroCal) according to the manufacturer's instruction.[11] All experiments were performed at 20 °C, 30 °C and 40 °C, respectively, in Tris-HCl (100 mM, pH 8.2) and solutions were degassed immediately before measurements were taken. The enthalpy of reaction was determined by single injection method, by adding substrate (*R*)-**4a** (1.5 mM, 1 x 20 µL, duration time 39.9 sec) to Pisa1 [125 nM for 20 and 30 °C and 6 µM for 40 °C]. For the kinetic parameters multiple injection method was used, titrating Pisa1 [20 nM for 20 and 30 °C and 200 nM for 40 °C] with (*R*)-**4a** (7 mM, 24x 10 µL, duration time 20 sec, spacing time 300 sec). Nonlinear least-quares fitting with Origin version 7.0 (MicroCal) for ITC data analysis was used to obtain K_M, and k_{cat}.

Table S2. Kinetic parameters of Pisa1 for the conversion of (*R*)-**4a** at different temperatures.

T [°C]	K_M [µM]	k_{cat} [min^{-1}]	k_{cat}/K_M [M^{-1}·sec^{-1}]
20	144±31	428±52	50.0*10^3±4.8*10^3
30	114±18	351±48	51.9*10^3±8.9*10^3
40	352±98	40±14	2.0*10^3±0.8*10^3

Determination of Pisa1 melting point.

The melting point (T_m) was determined by circular dichroism (CD) measurements at 15-95 °C using a Jasco J-715 spectrometer equipped with a Xe-lamp. Spectra were recorded in buffer (Tris-HCl, 100 mM, pH 8.0) at 208 nm. The melting point (57 °C) was calculated from a sigmoid fit of the data points.

pH Studies

Substrate *rac*-**4a** (5 mg, 22 µmol) was dissolved in buffer (980 µL, 100 mM; pH 5-6: acetic acid/sodium acetate, pH 7-9: Tris-HCl, pH 10: glycine-NaOH). An aliquot of Pisa1 enzyme solution (20 µL, 130 µg, 1.8 nmol) was added and the mixture was shaken at 30 °C and 120 rpm for 15-180 min. Afterwards 1-octanol was added as an internal standard (2.5 mg, 19 µmol; dissolved in 1000 µL ethyl acetate) and the organic phase was dried over anhydrous sodium sulfate, derivatized as acetate ester[1] and analyzed by chiral GC.
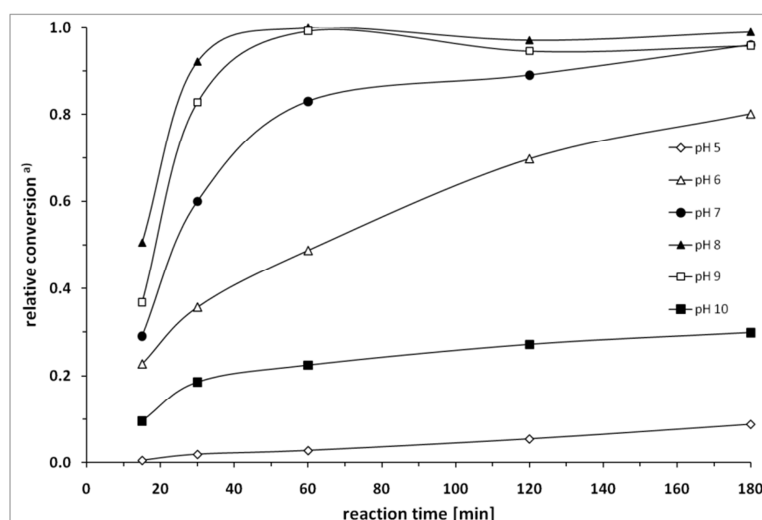
Fig. S2: Conversion of *rac*-**4a** with Pisa1 at pH values of 5-10. Reaction conditions: *rac*-**4a** (5 mg, 22 mM), Pisa1 (20 µL, 130 µg, 1.8 µM), buffer (100 mM; pH 5-6: acetic acid/sodium acetate, pH 7-9: Tris-HCl, pH 10: glycine-NaOH), reaction volume: 1000 µL, 30°C, 120 rpm, 15-180 min; [a] determined by the internal standard.

Determination of protein concentrations

For the determination of protein concentrations, UV absorption at 280 nm according to Bradfords method[1] were measured using a Nanodrop ND-1000 spectrophotometer: Purified protein solution (60 μL, diluted if necessary) was directly measured in an UV permeable cuvette. The concentration was determined at 280 nm according to Lambert-Beers law. The extinction coefficient was calculated with the ProtParam tool from ExPASy.[12]

^{1}H-NMR (1a) *rac*-2-Pentyl sulfate; D2O
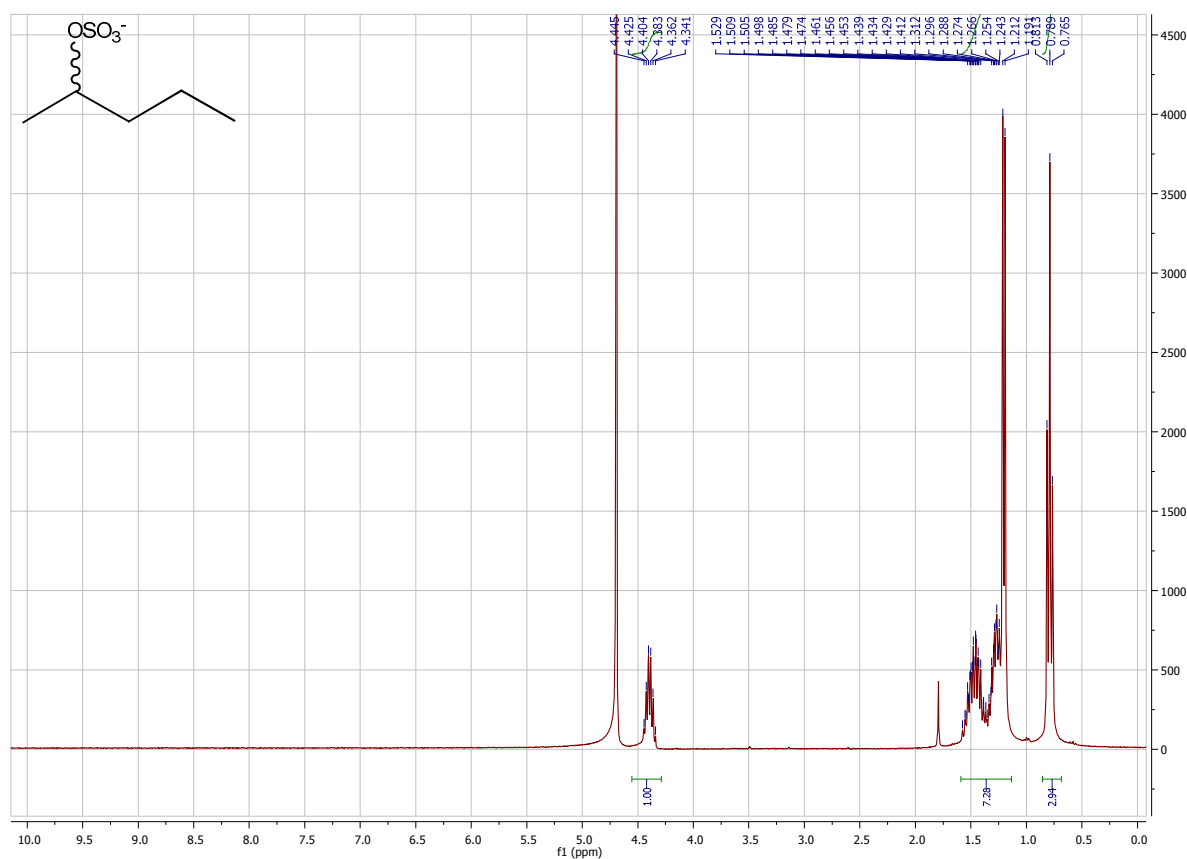
^{13}C-NMR (1a) *rac*-2-Pentyl sulfate; D$_2$O
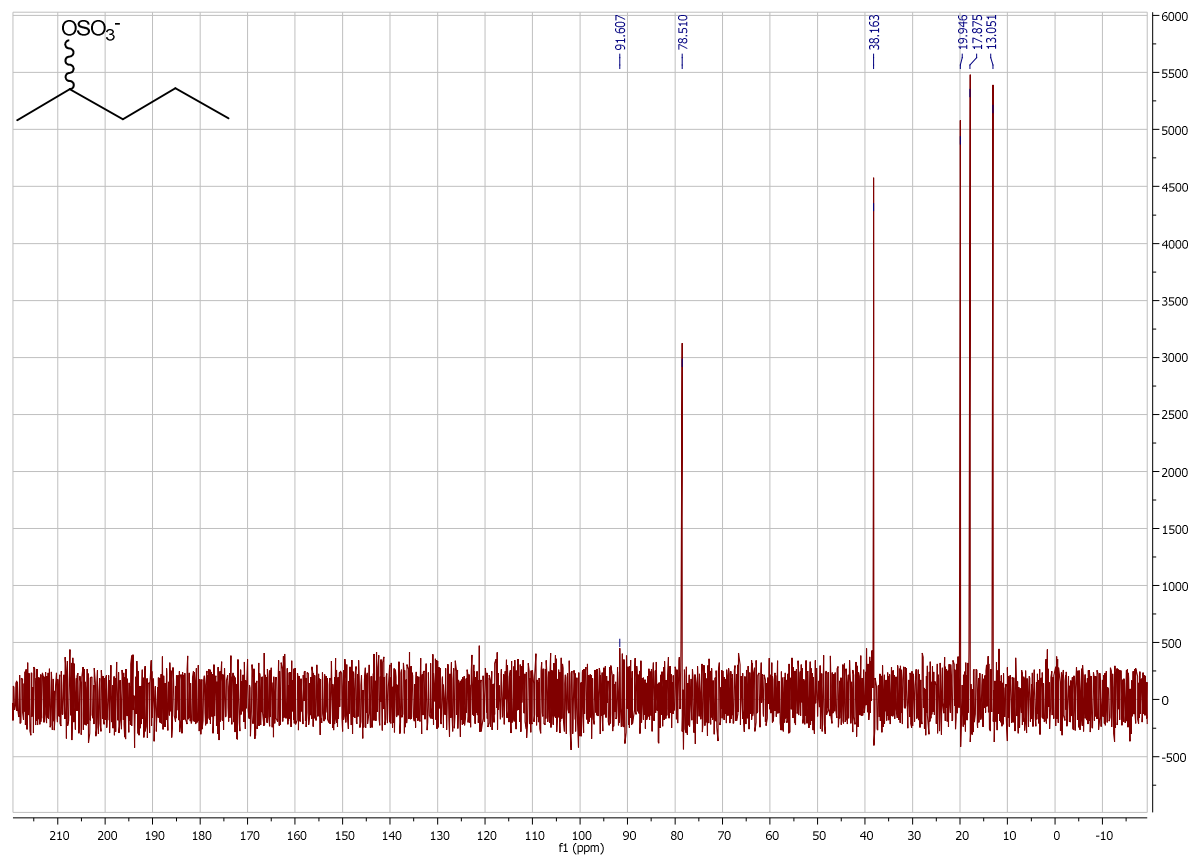
^{1}H-NMR (2a) *rac*-2-Hexyl sulfate; D$_2$O
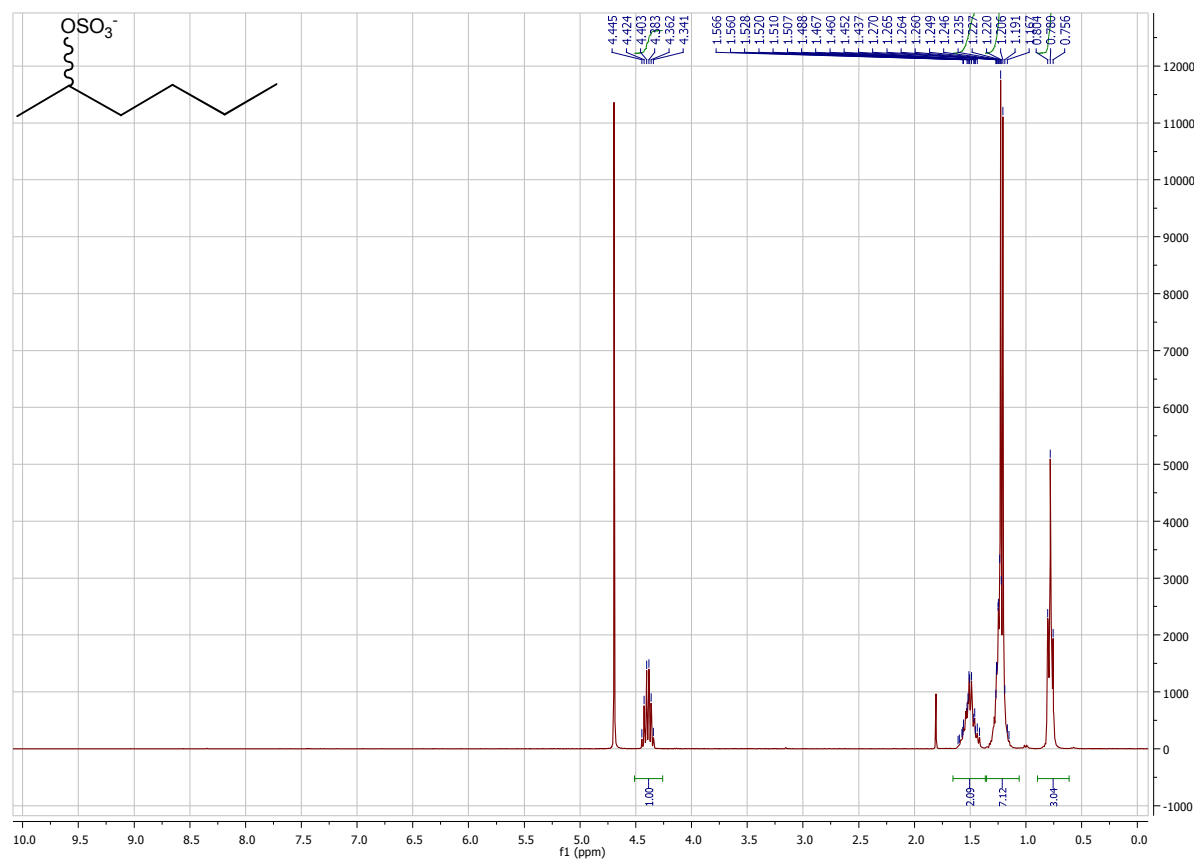
¹³C-NMR (2a) *rac*-2-Hexyl sulfate; D₂O
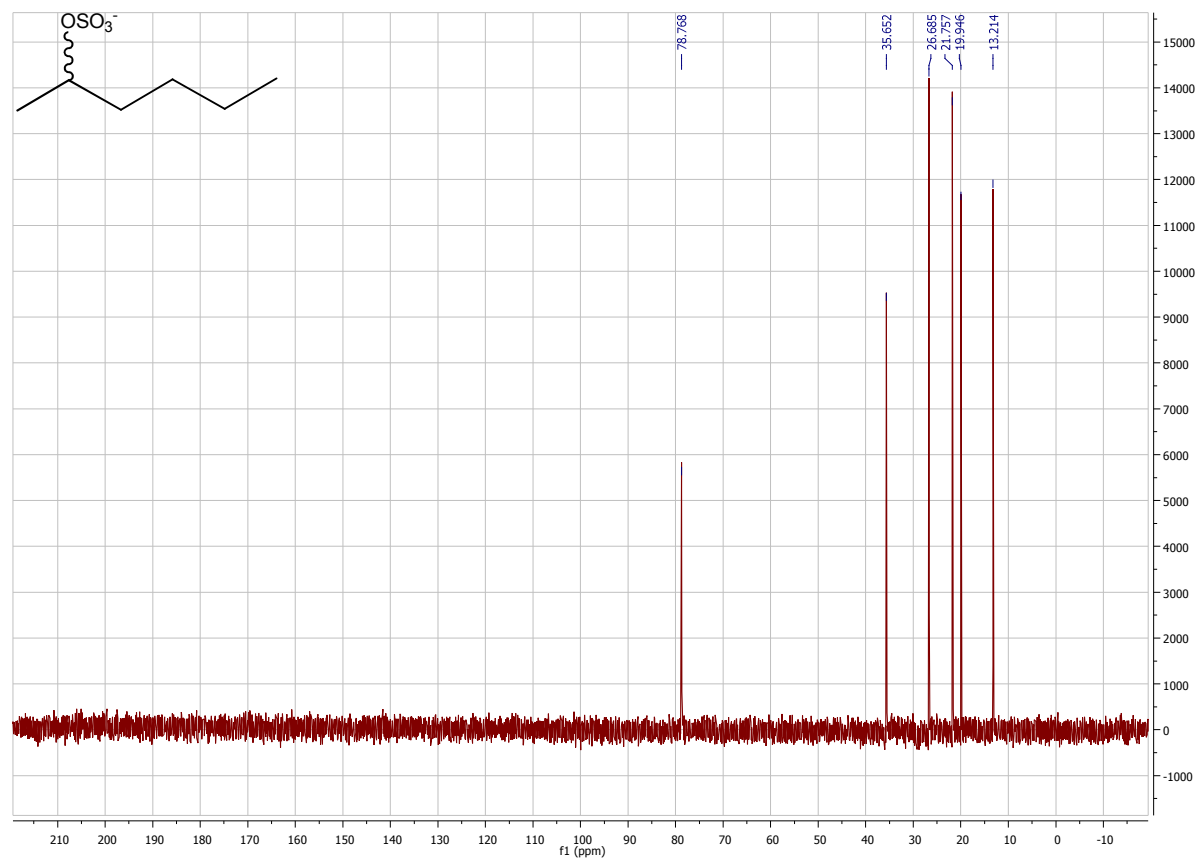
¹H-NMR (3a) *rac*-5-Methyl-2-hexyl sulfate; D₂O
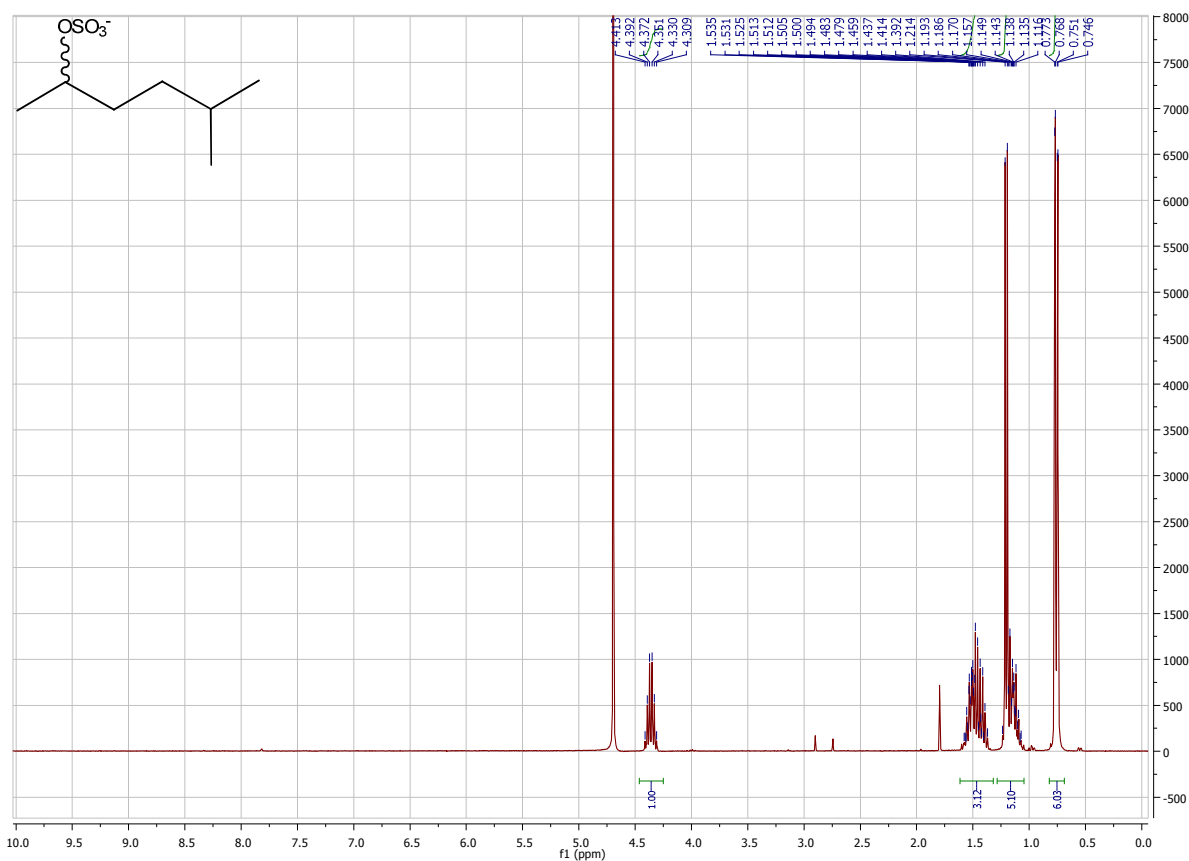
¹³C-NMR (3a) *rac*-5-Methyl-2-hexyl sulfate; D₂O
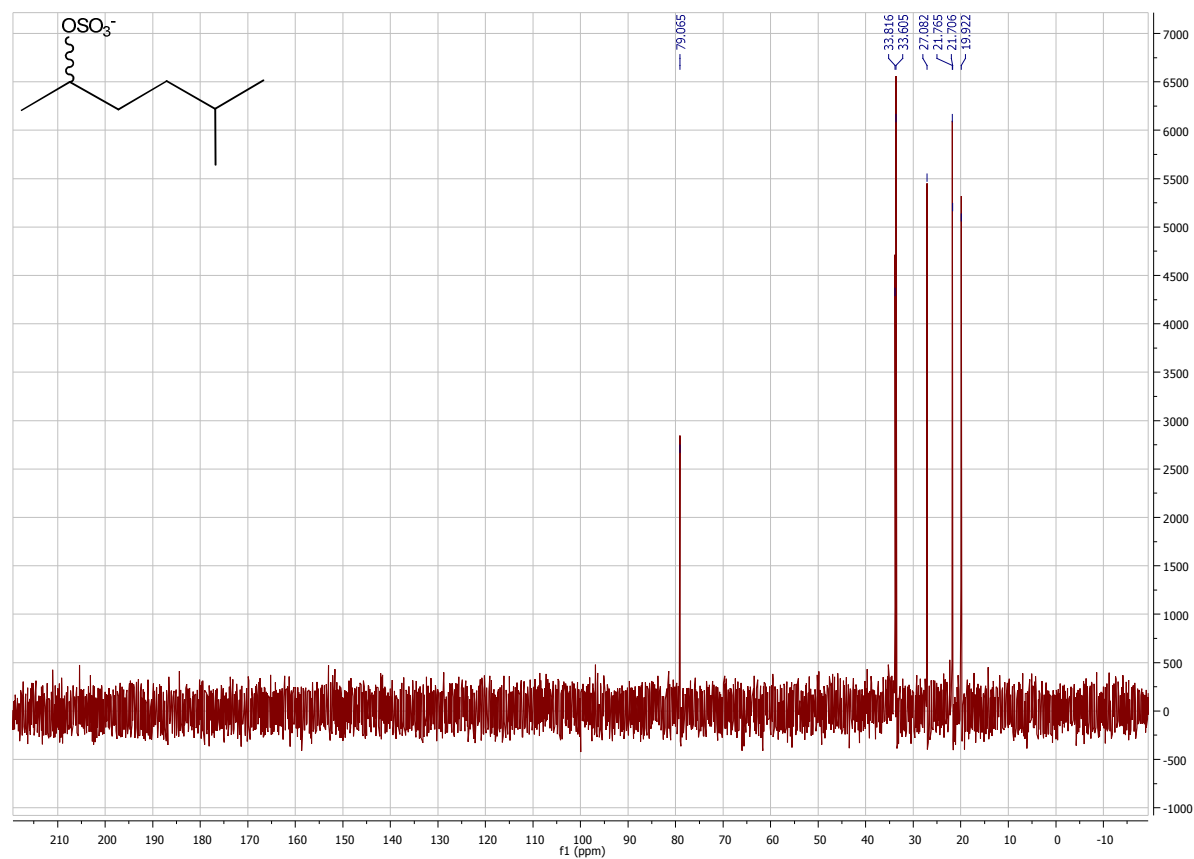
¹H-NMR (5a) *rac*-2-Decyl sulfate; D₂O
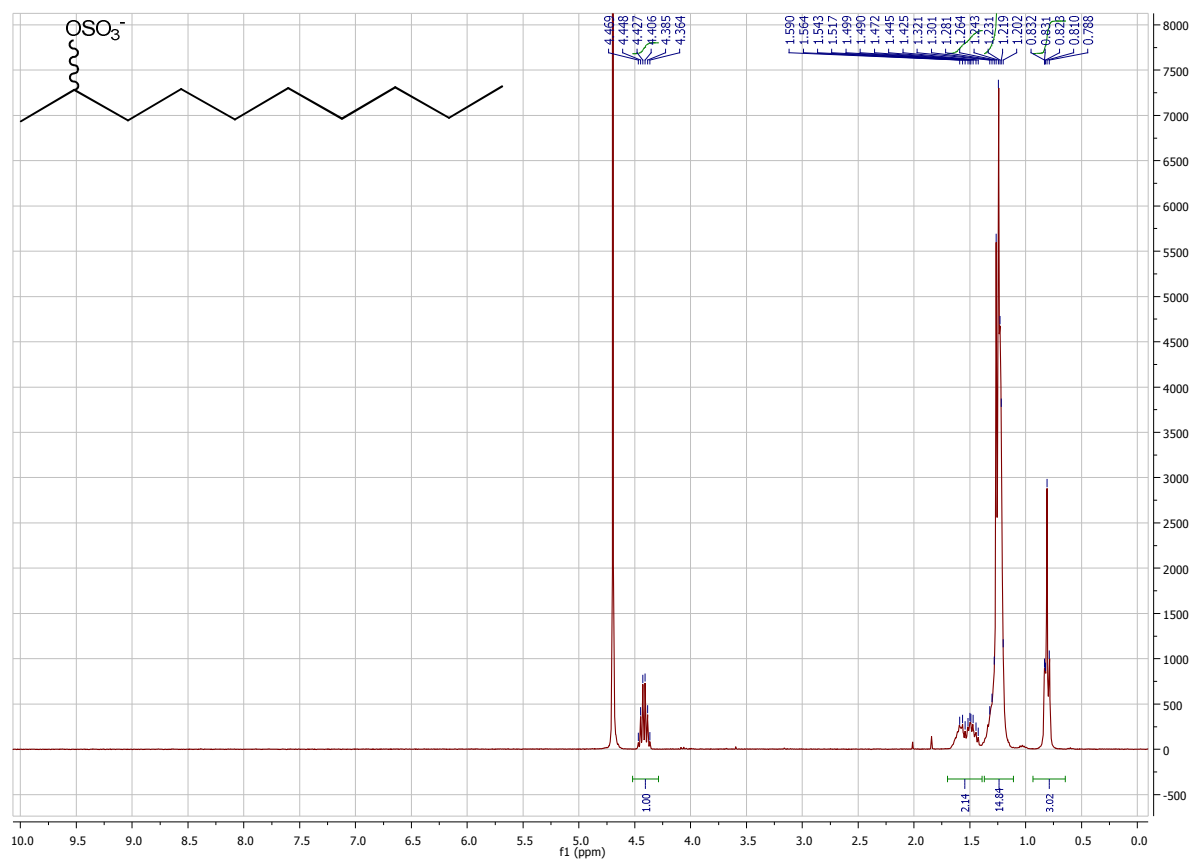
^{13}C-NMR (5a) *rac*-2-Decyl sulfate; D$_2$O
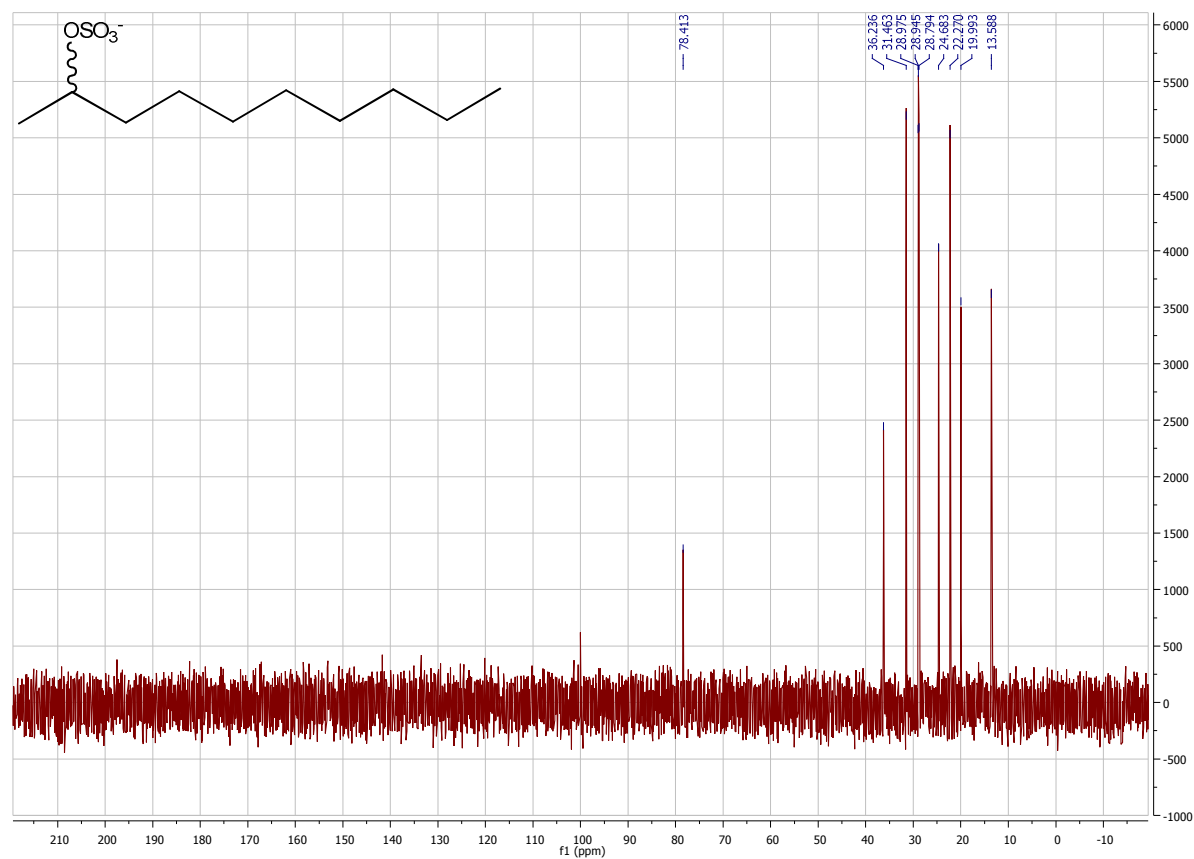
^{1}H-NMR (7a) *rac*-3-Heptyl sulfate; D$_2$O
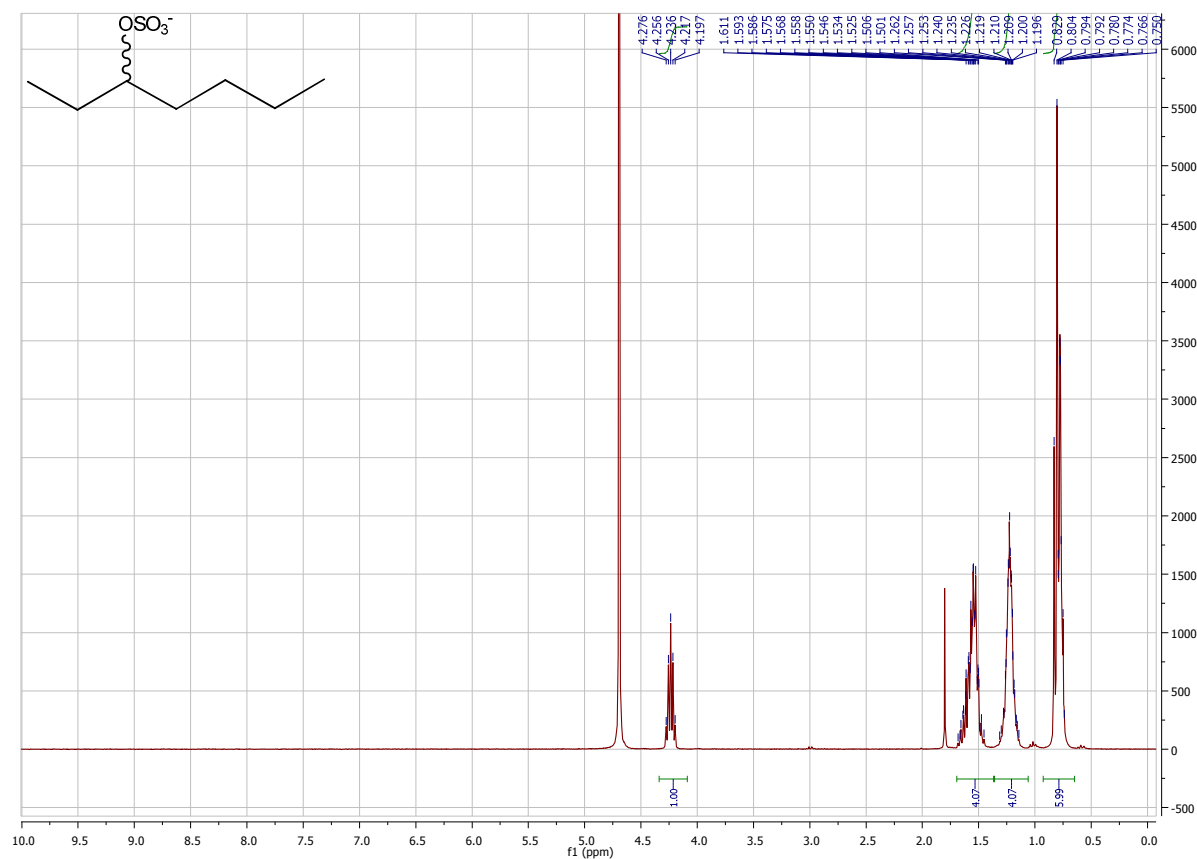
^{13}C-NMR (7a) *rac*-3-Heptyl sulfate; D$_2$O
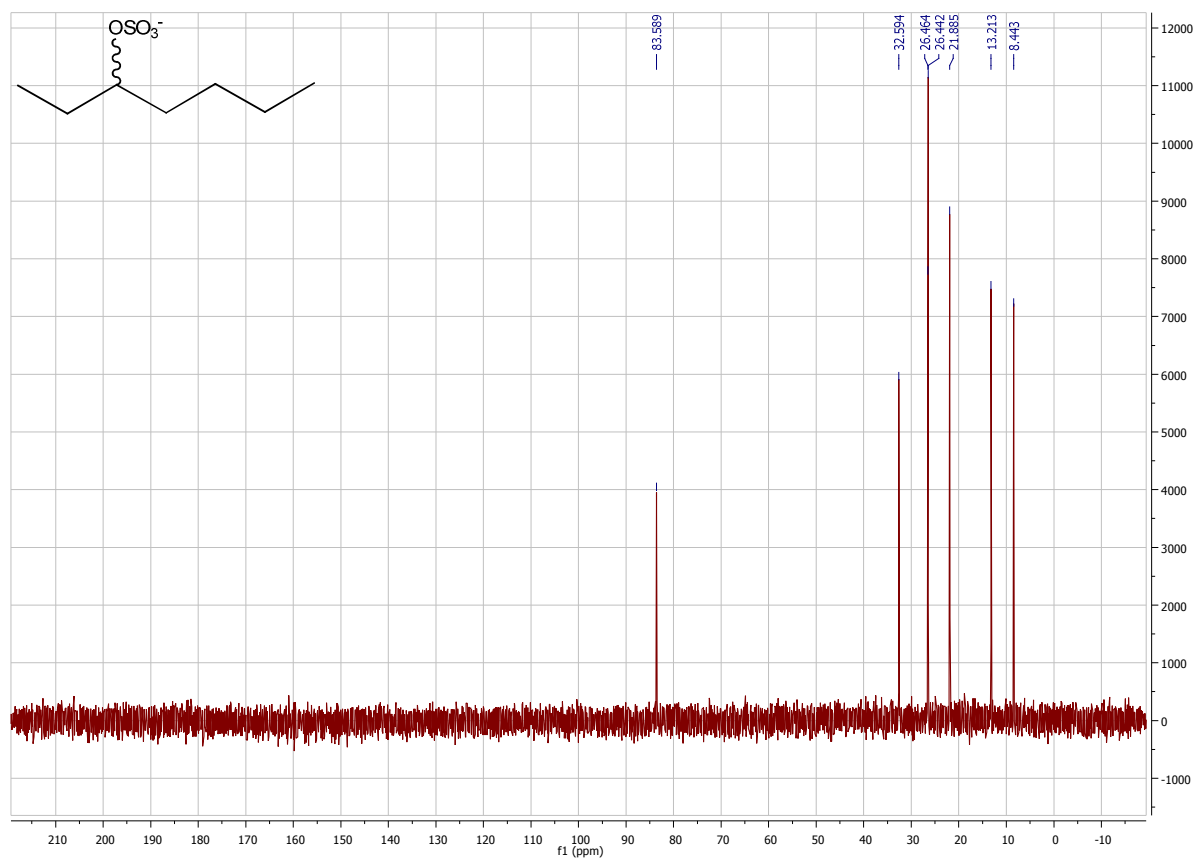
^{1}H-NMR (9a) *rac*-4-Penten-2-yl sulfate; D$_2$O
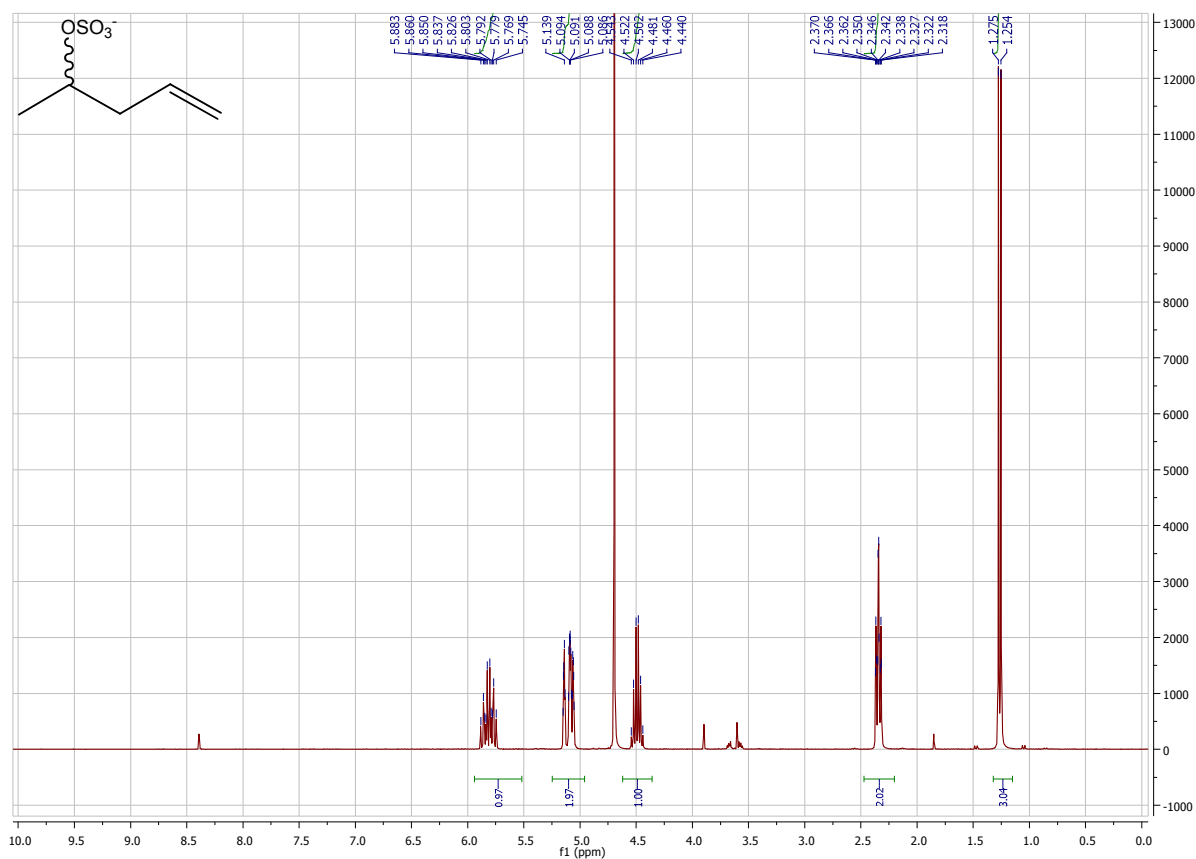
^{13}C-NMR (9a) *rac*-4-Penten-2-yl sulfate; D_2O
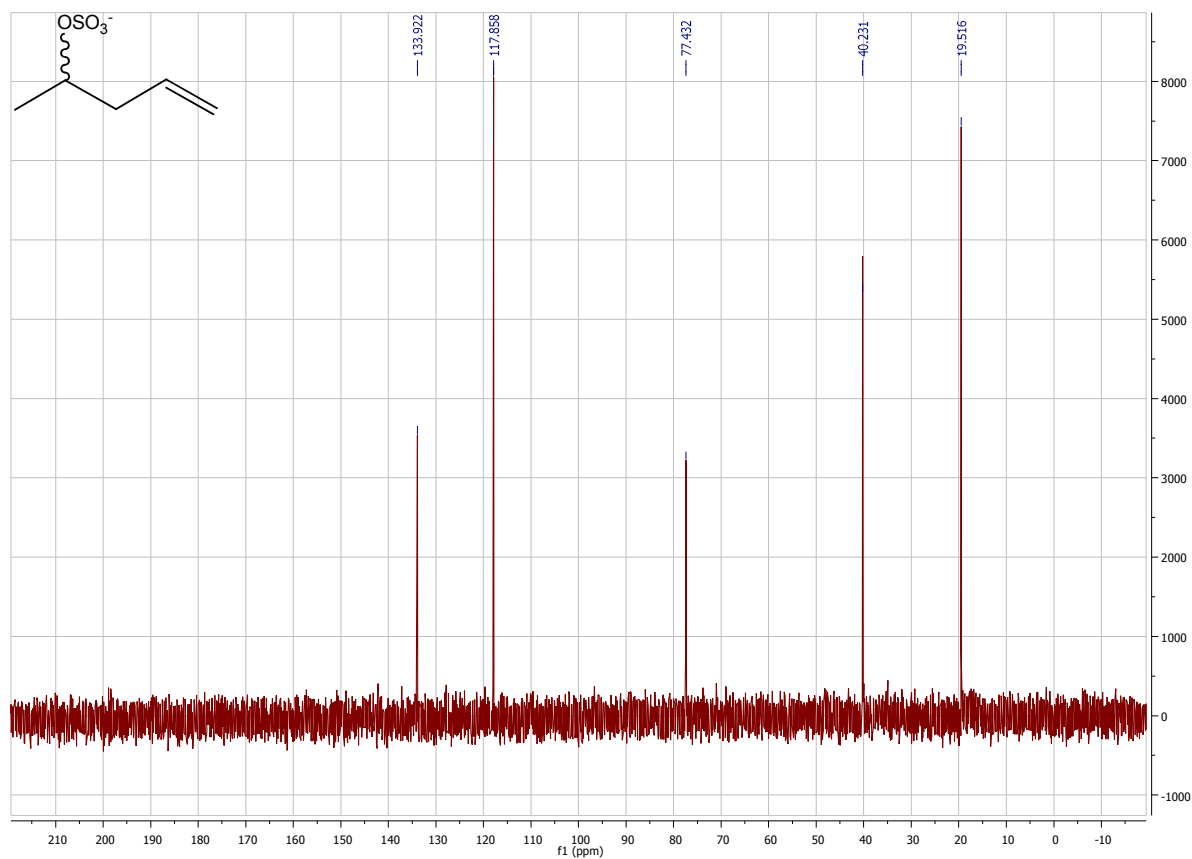
^{1}H-NMR (10a) *rac*-5-Hexen-2-yl sulfate; D_2O
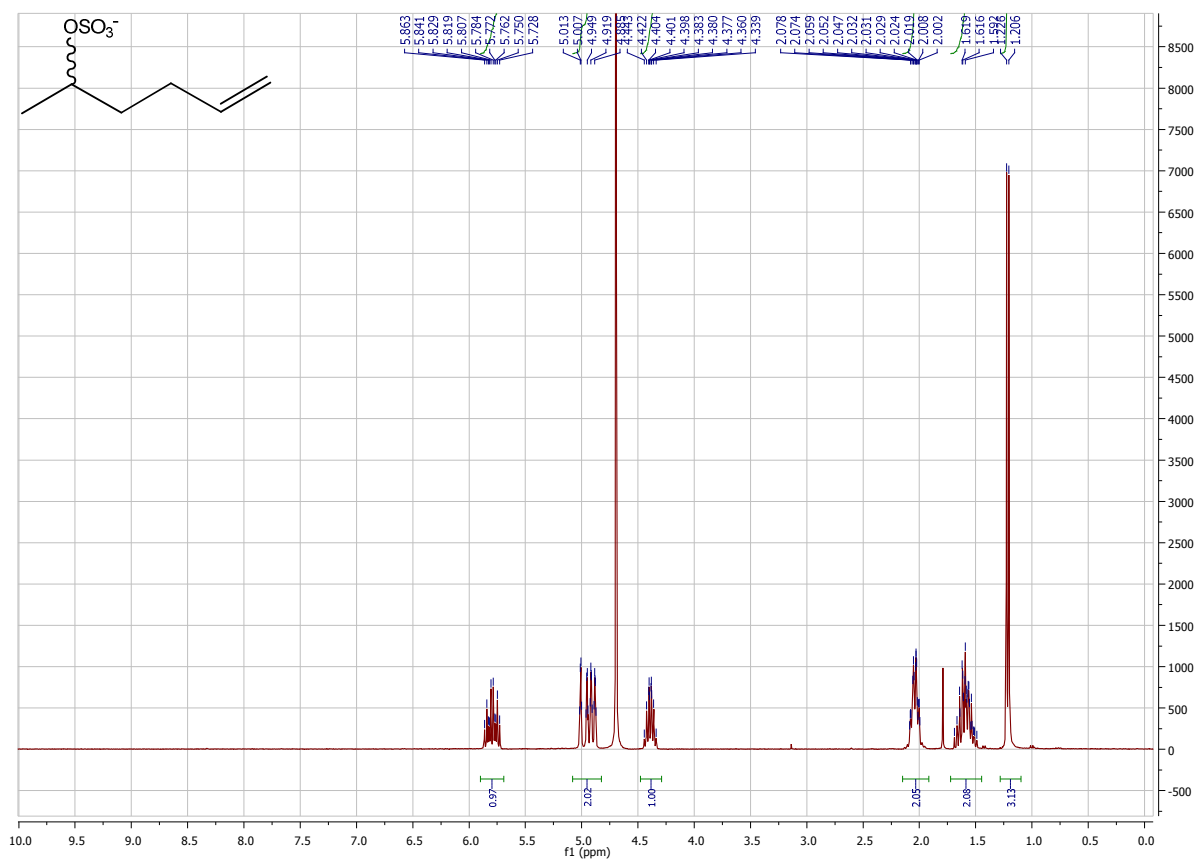
¹³C-NMR (10a) *rac*-5-Hexen-2-yl sulfate; D₂O
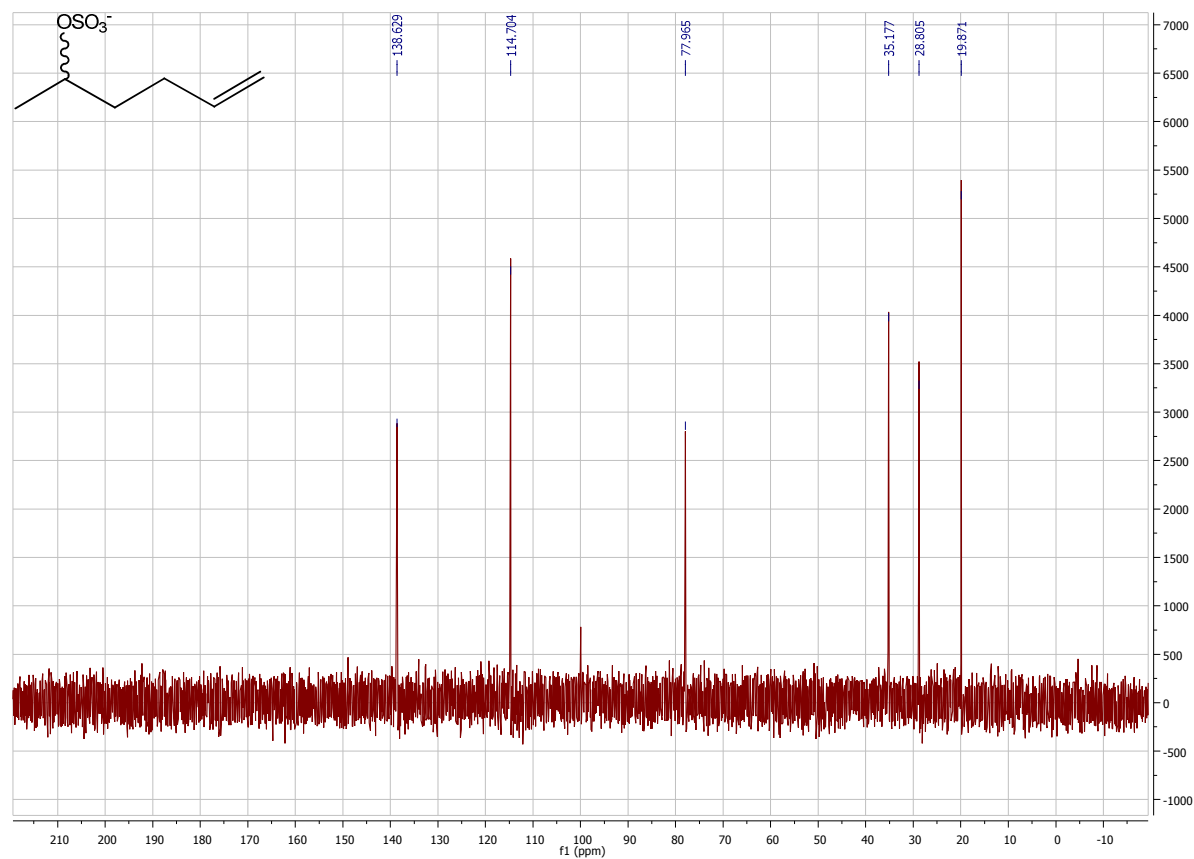
¹H-NMR (11a) *rac*-1-Hepten-3-yl sulfate; DMSO-d₆
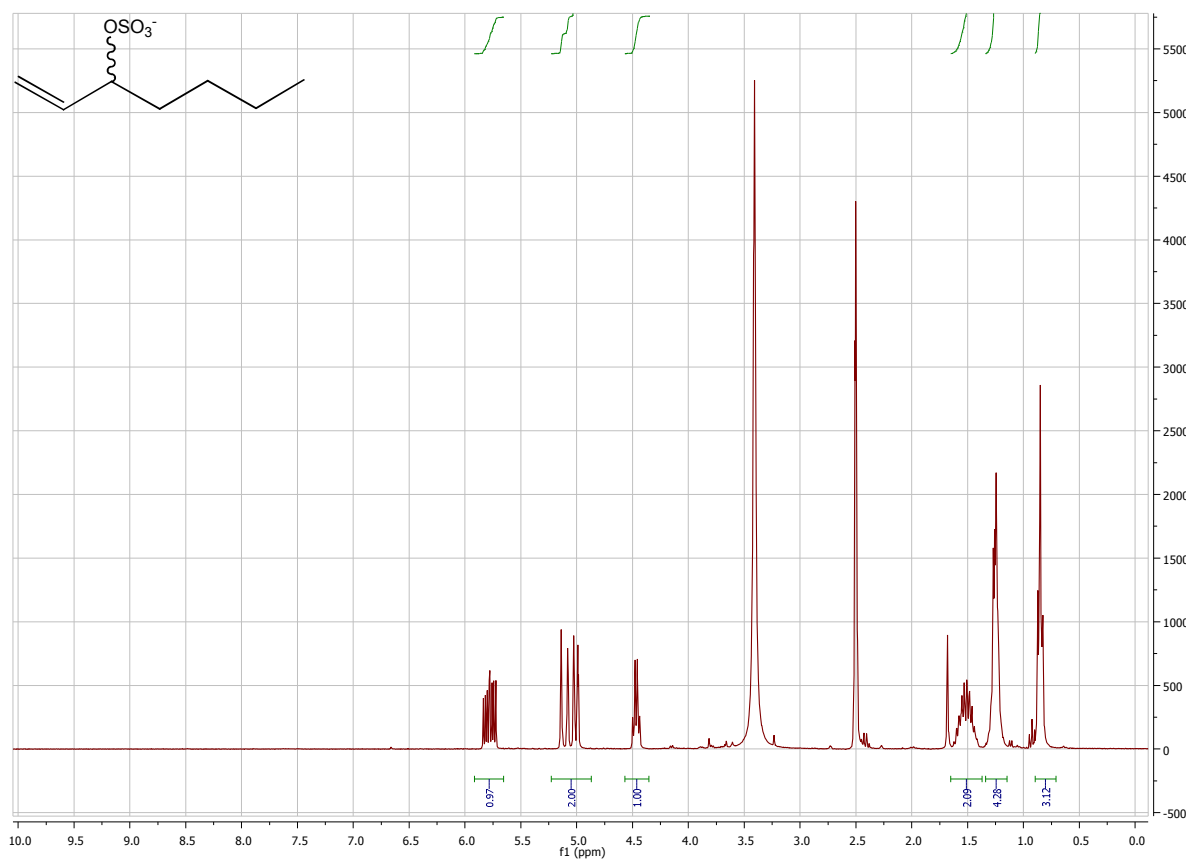
^{13}C-NMR (11a) *rac*-1-Hepten-3-yl sulfate; DMSO-d$_6$

^{1}H-NMR (12a) *rac*-1-Octen-3-yl sulfate; DMSO-d$_6$
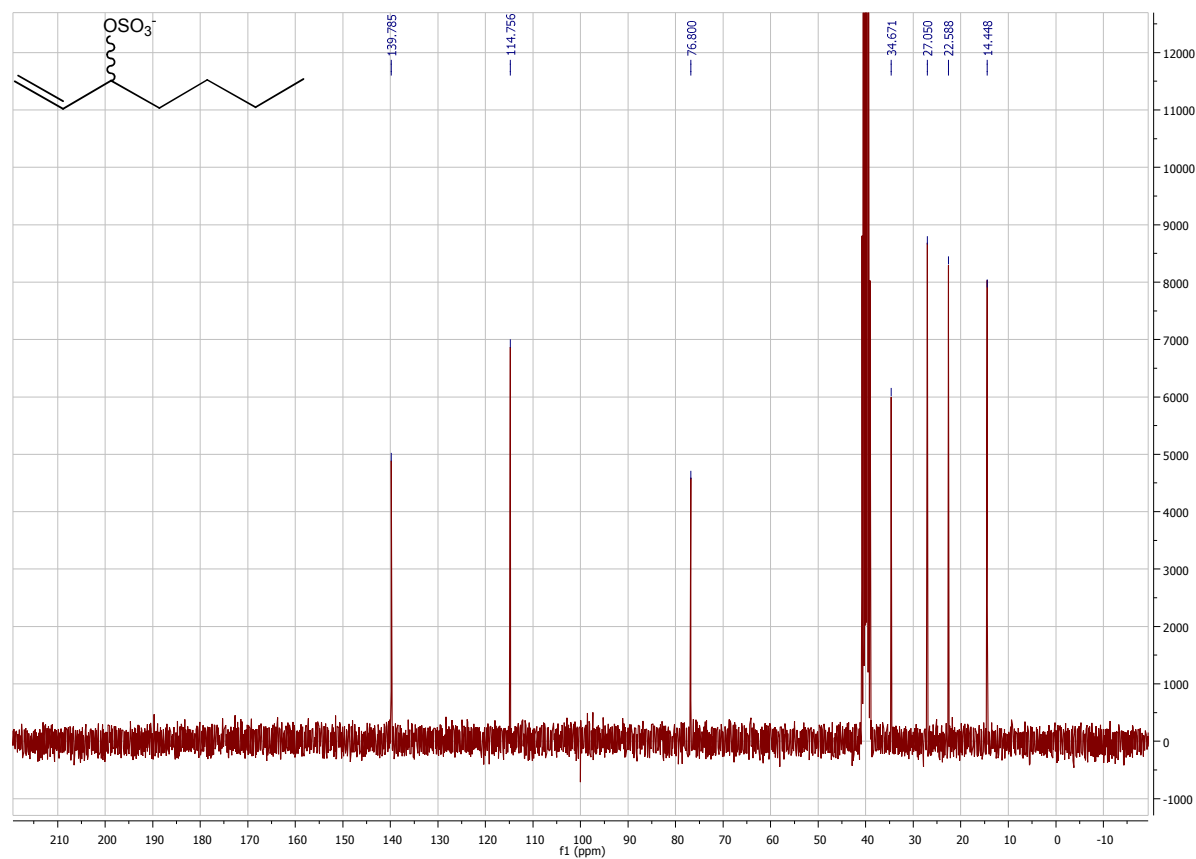
13C-NMR (12a) *rac*-**1-Octen-3-yl sulfate; DMSO-d₆**
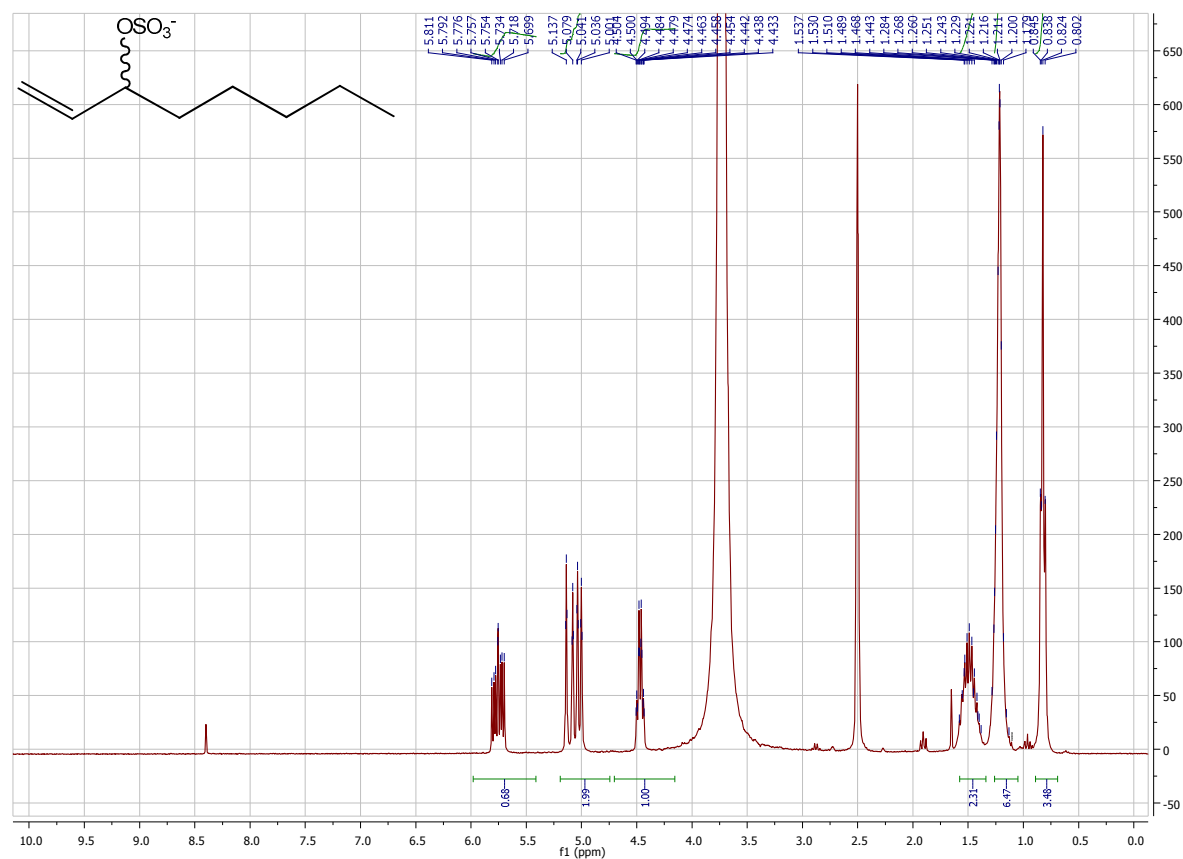
^{13}**C-NMR (12a)** *rac*-**1-Octen-3-yl sulfate; DMSO-d$_6$**
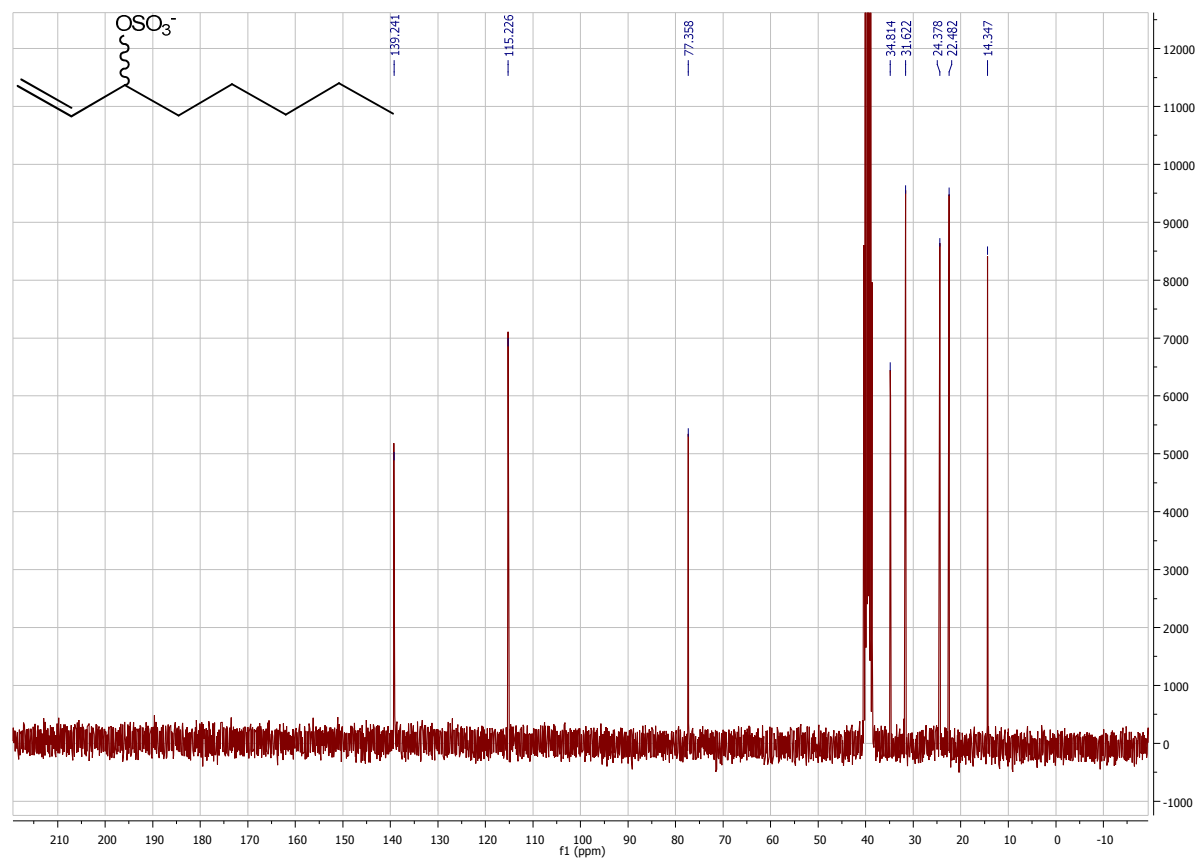
^{1}**H-NMR (13a)** *rac*-**1-Nonen-3-yl sulfate; DMSO-d$_6$**
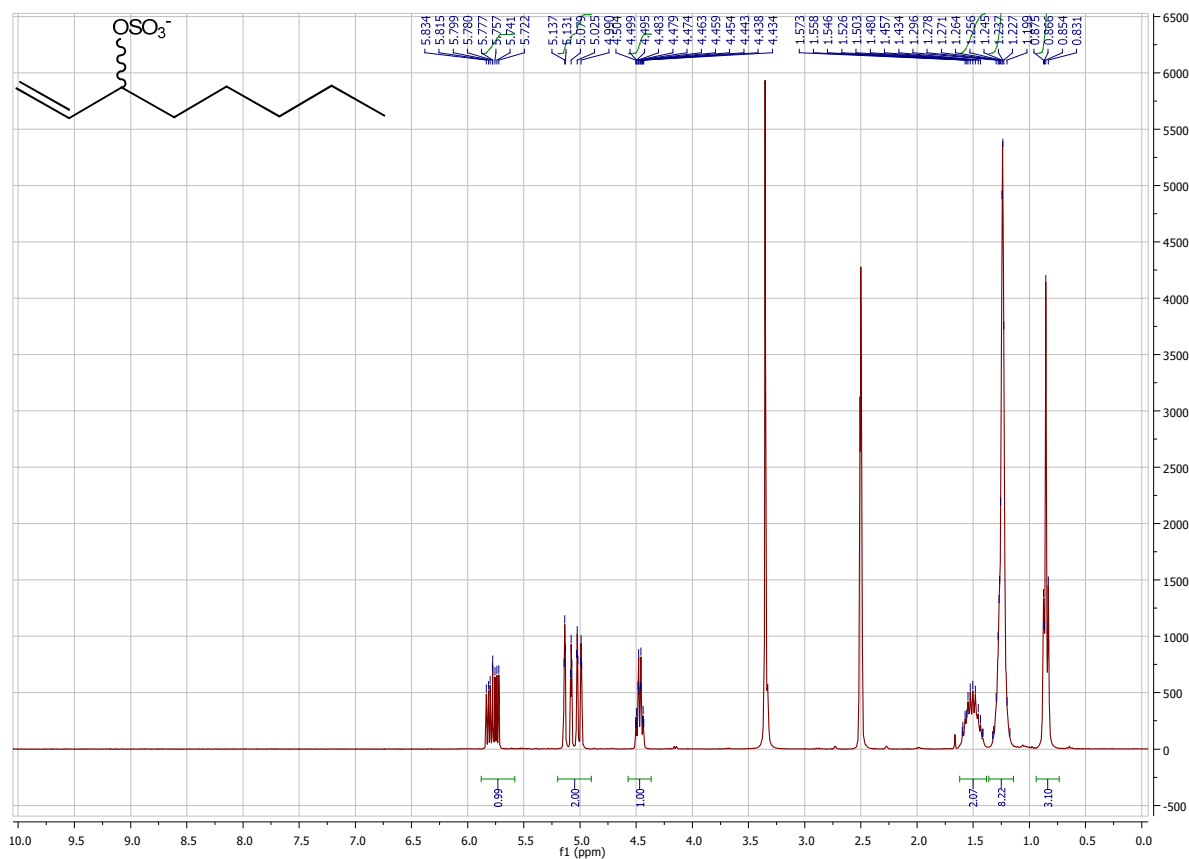
¹³C-NMR (13a) *rac*-1-Nonen-3-yl sulfate; DMSO-d₆
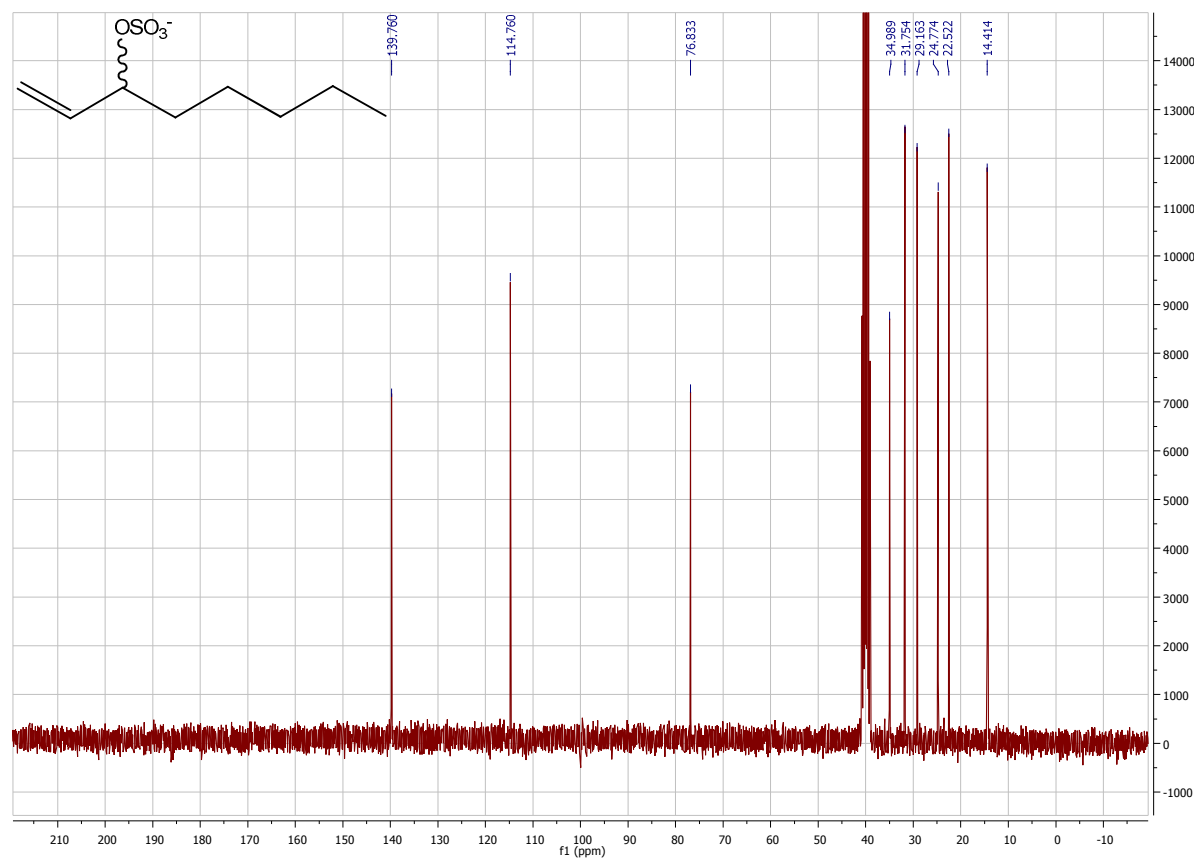
¹H-NMR (14a) *rac*-1-Decen-3-yl sulfate; DMSO-d₆
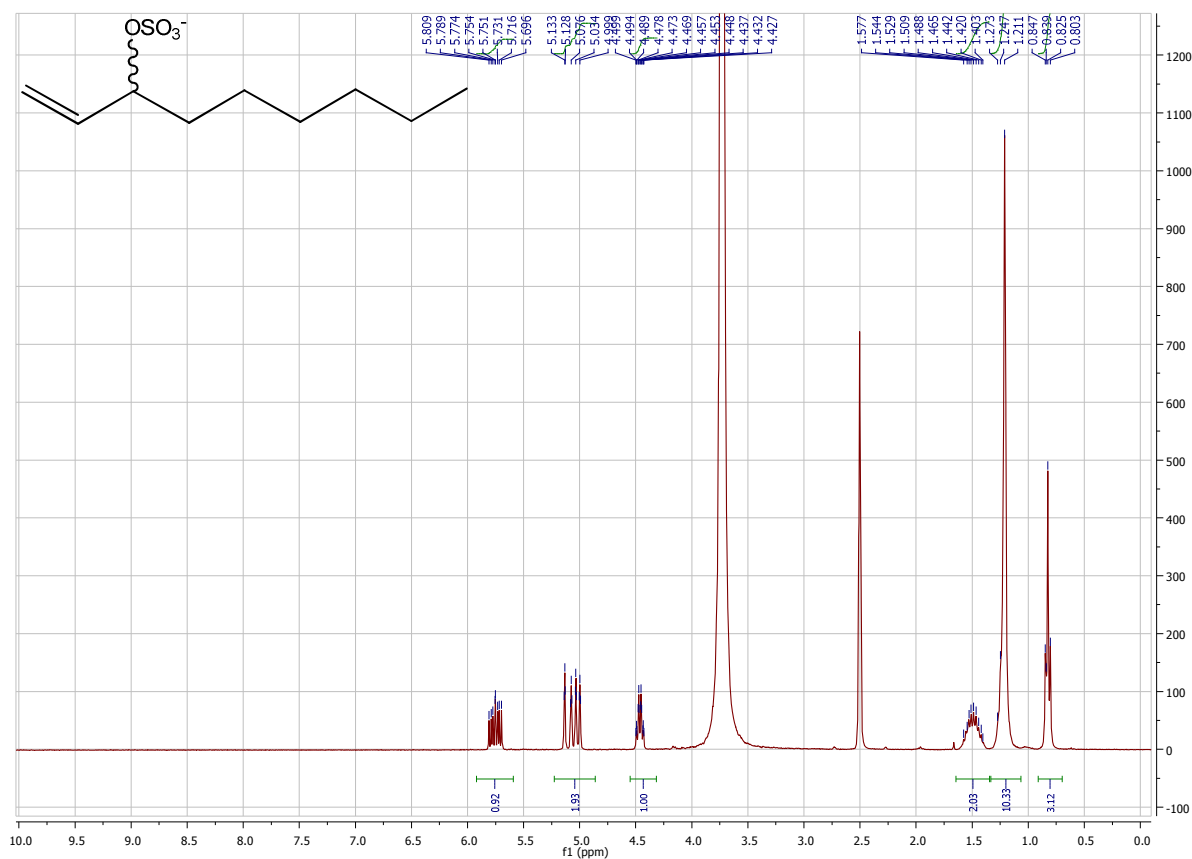
^{13}C-NMR (14a) *rac*-1-Decen-3-yl sulfate; DMSO-d$_6$
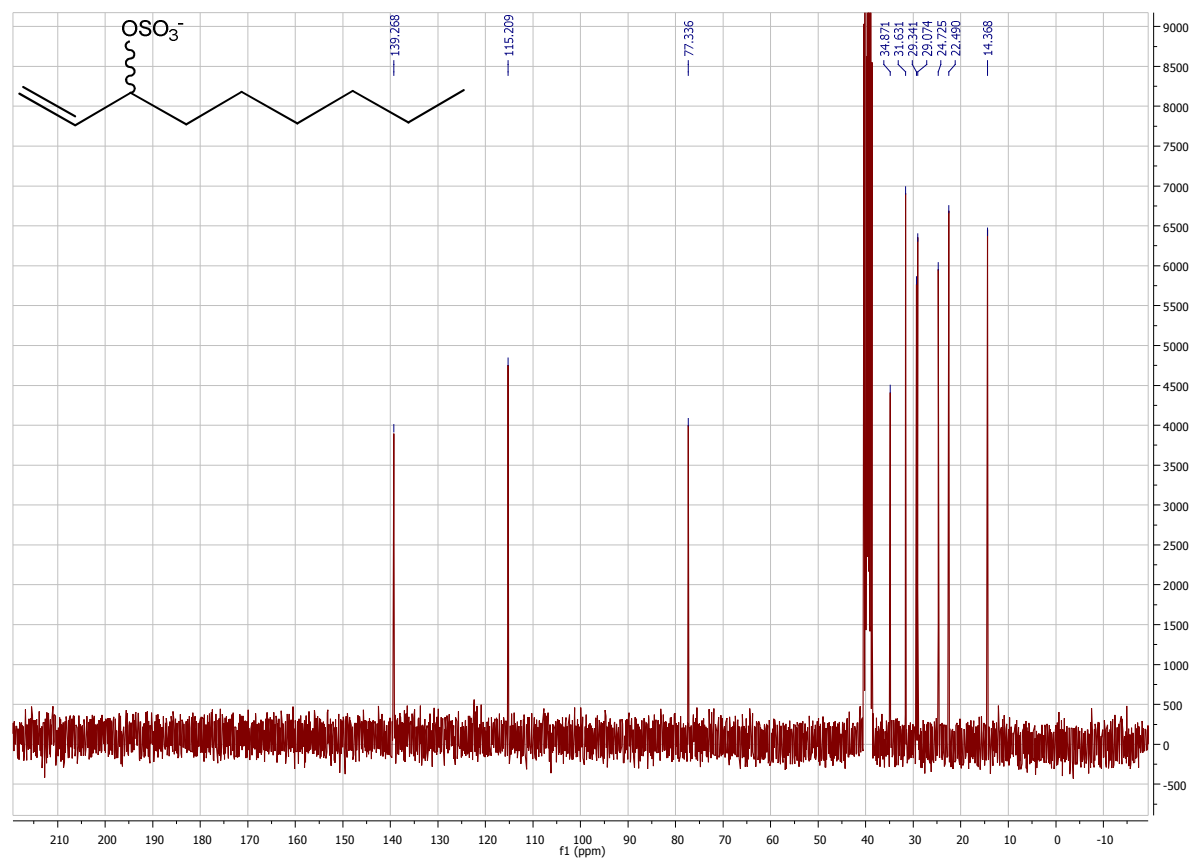
^{1}H-NMR (15a) *rac*-3-Hexyn-2-yl sulfate; D$_2$O
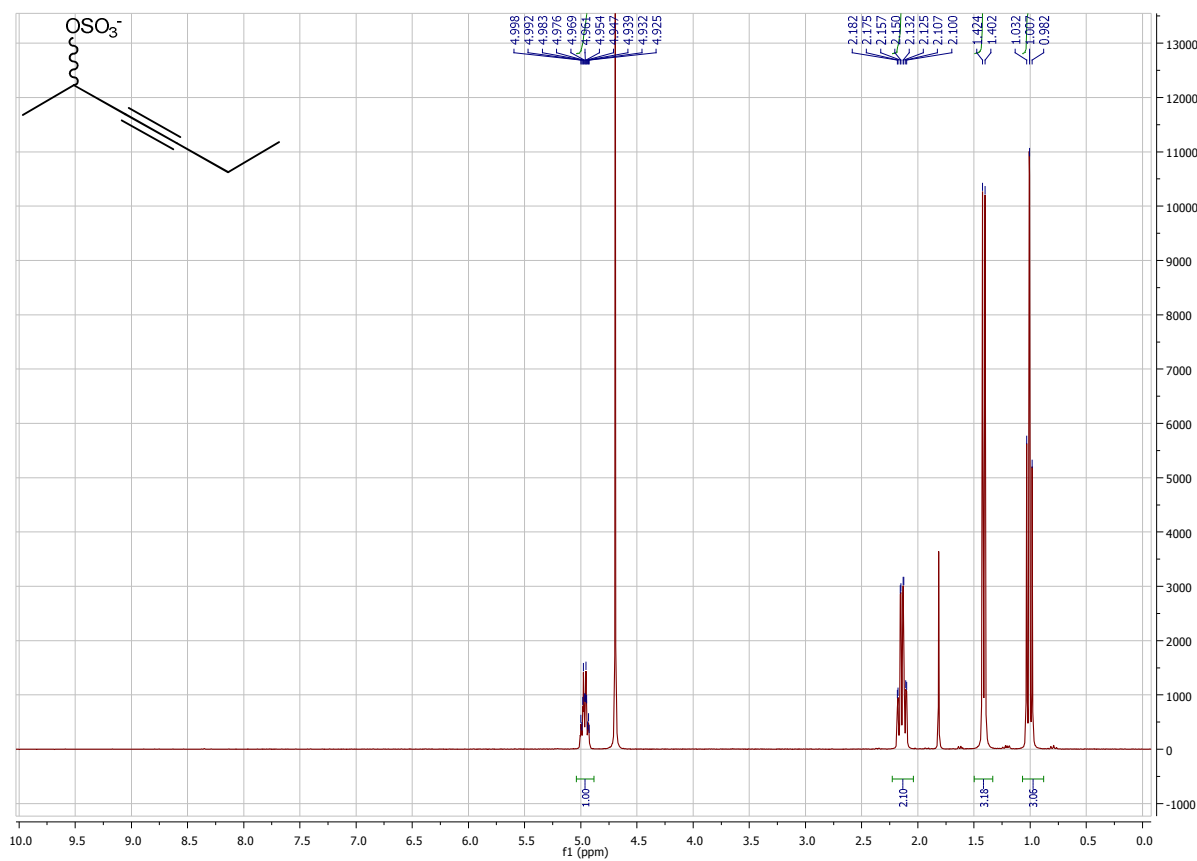
¹³C-NMR (15a) *rac*-3-Hexyn-2-yl sulfate; D₂O
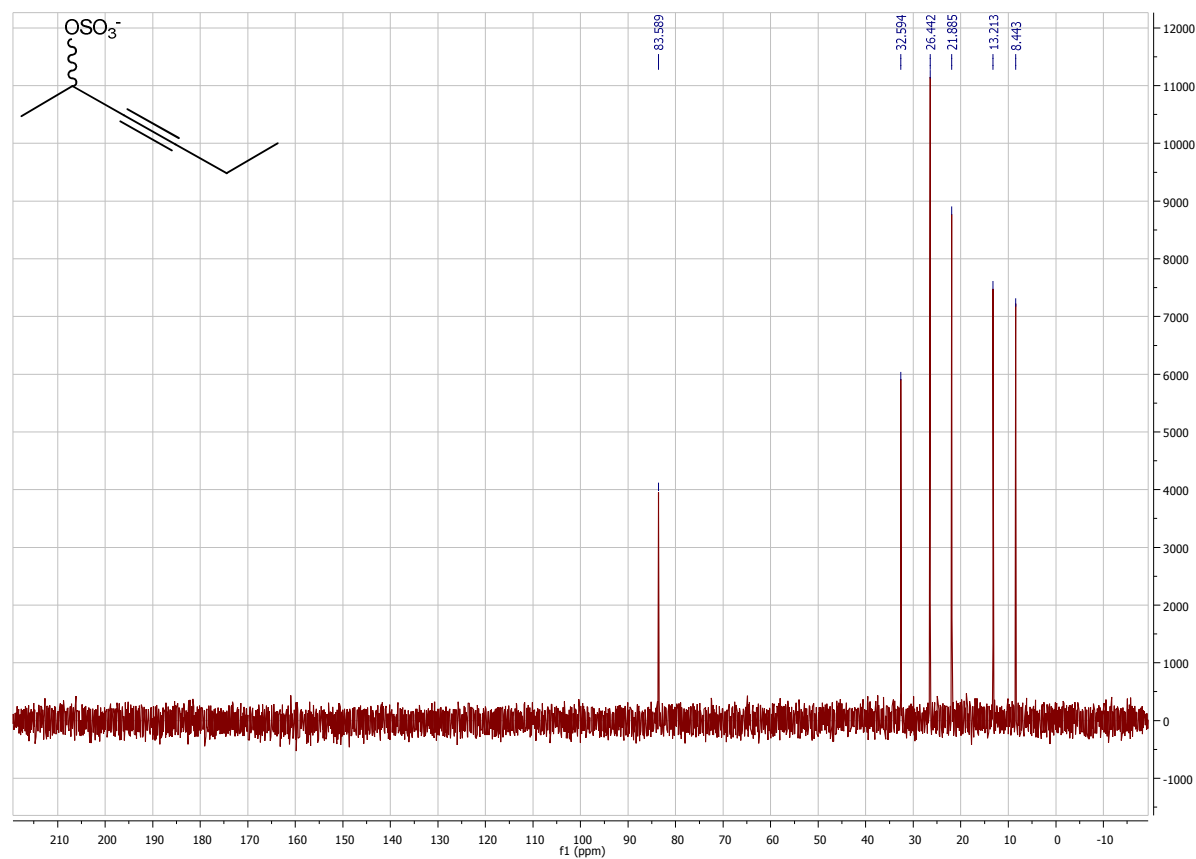
¹H-NMR (16a) *rac*-4-Heptyn-2-yl sulfate; D₂O
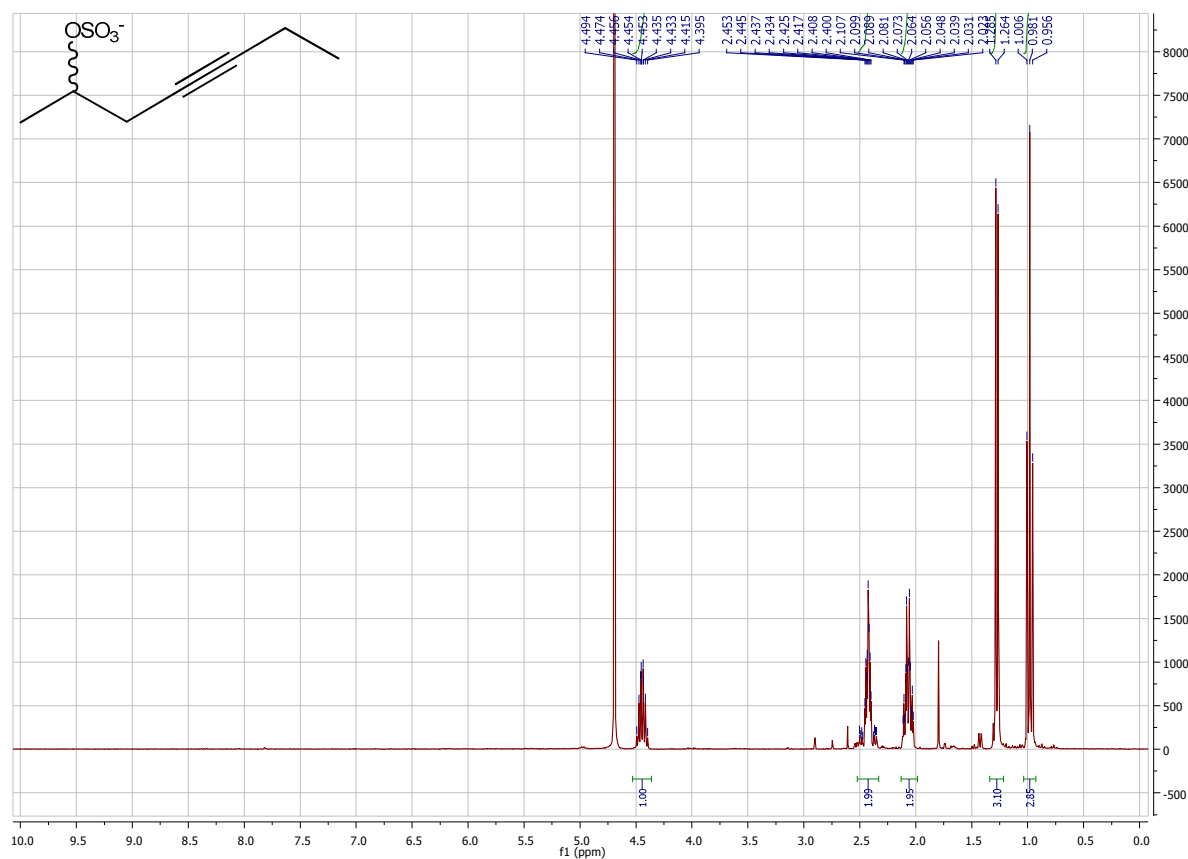
¹³C-NMR (16a) *rac*-4-Heptyn-2-yl sulfate; D₂O
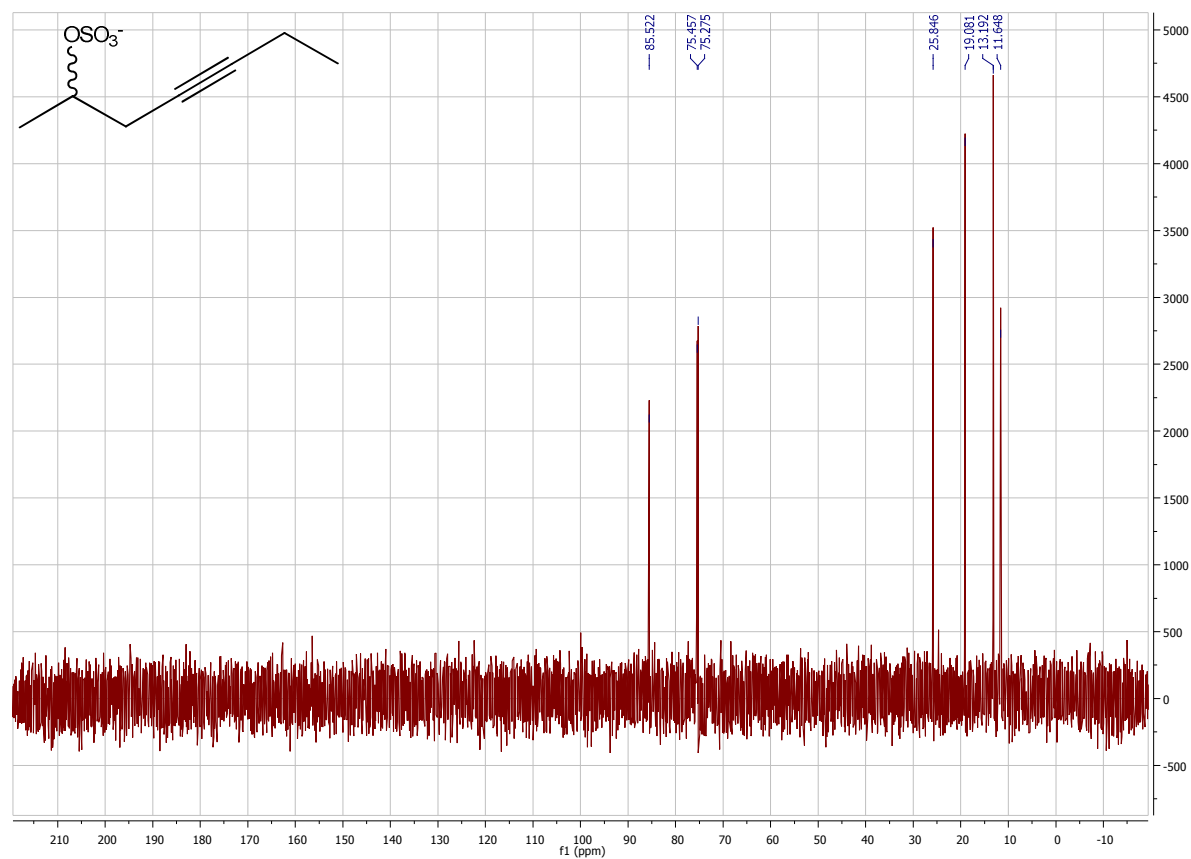
¹H-NMR (17a) *rac*-4-Phenyl-3-heyptyn-2-yl sulfate; D₂O
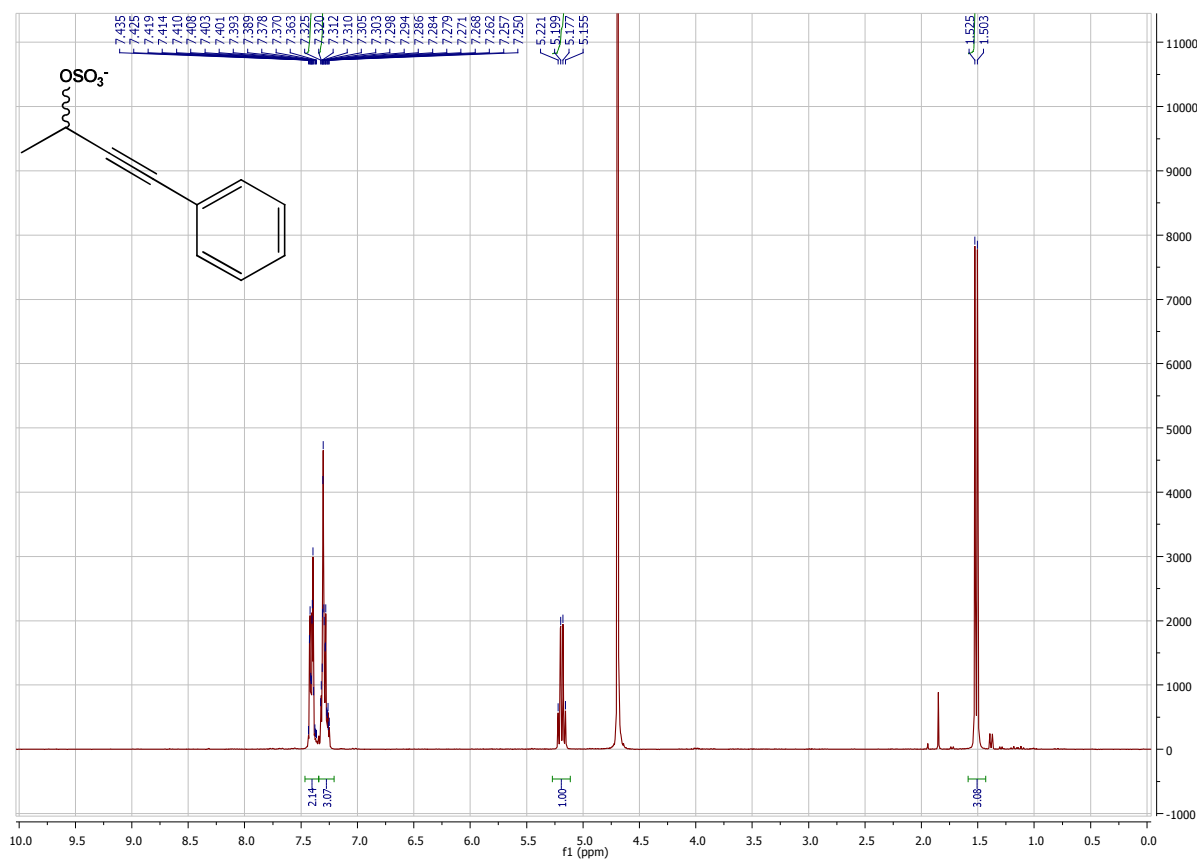
^{13}C-NMR (17a) *rac*-4-Phenyl-3-heyptyn-2-yl sulfate; D$_2$O
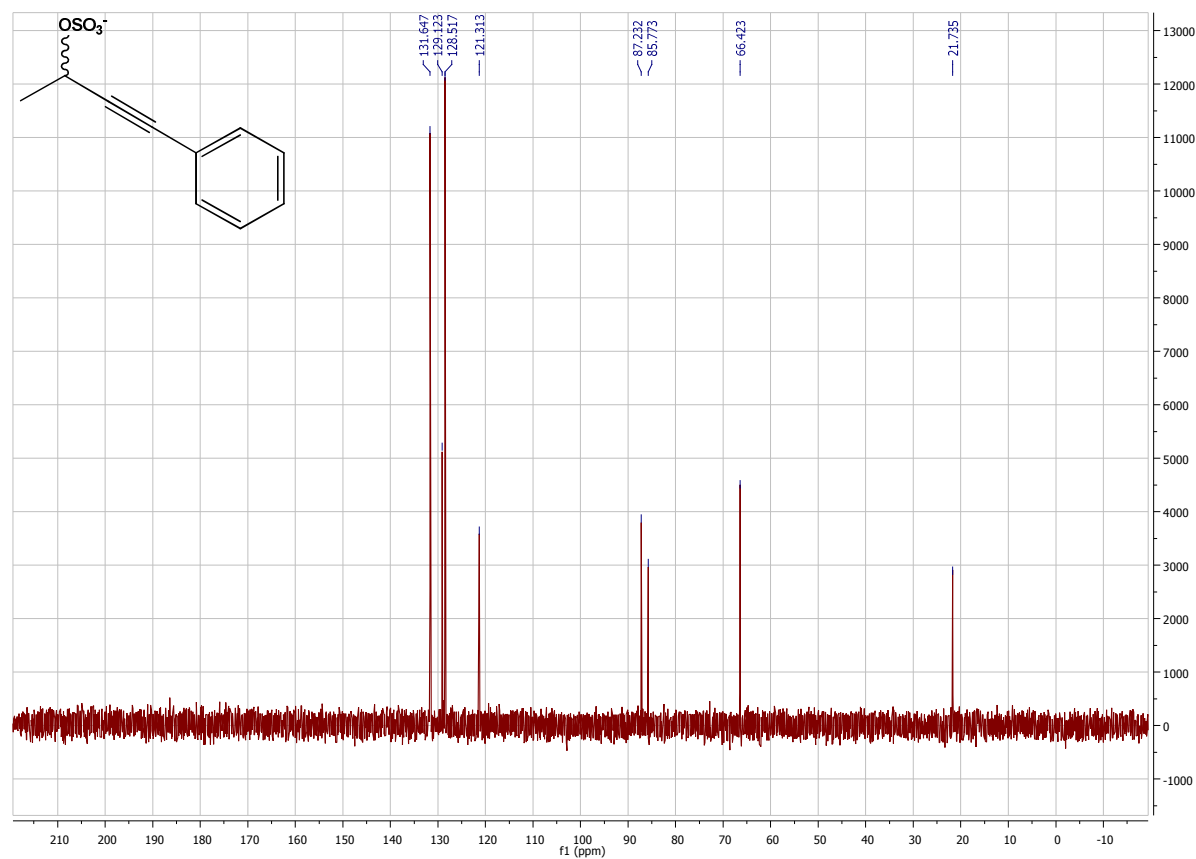
^{1}H-NMR (18a) *rac*-4-Hexyn-3-yl sulfate; D$_2$O
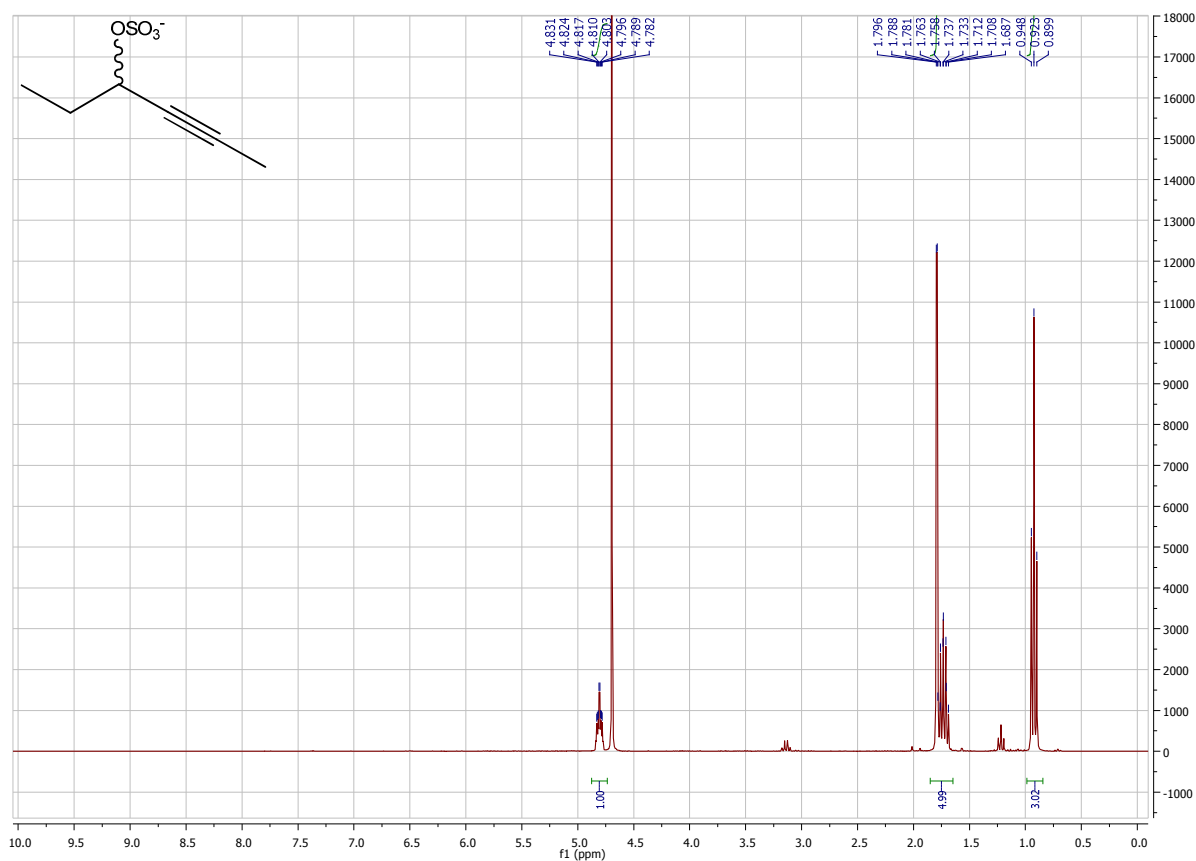
^{13}C-NMR (18a) *rac*-4-Hexyn-3-yl sulfate; D_2O
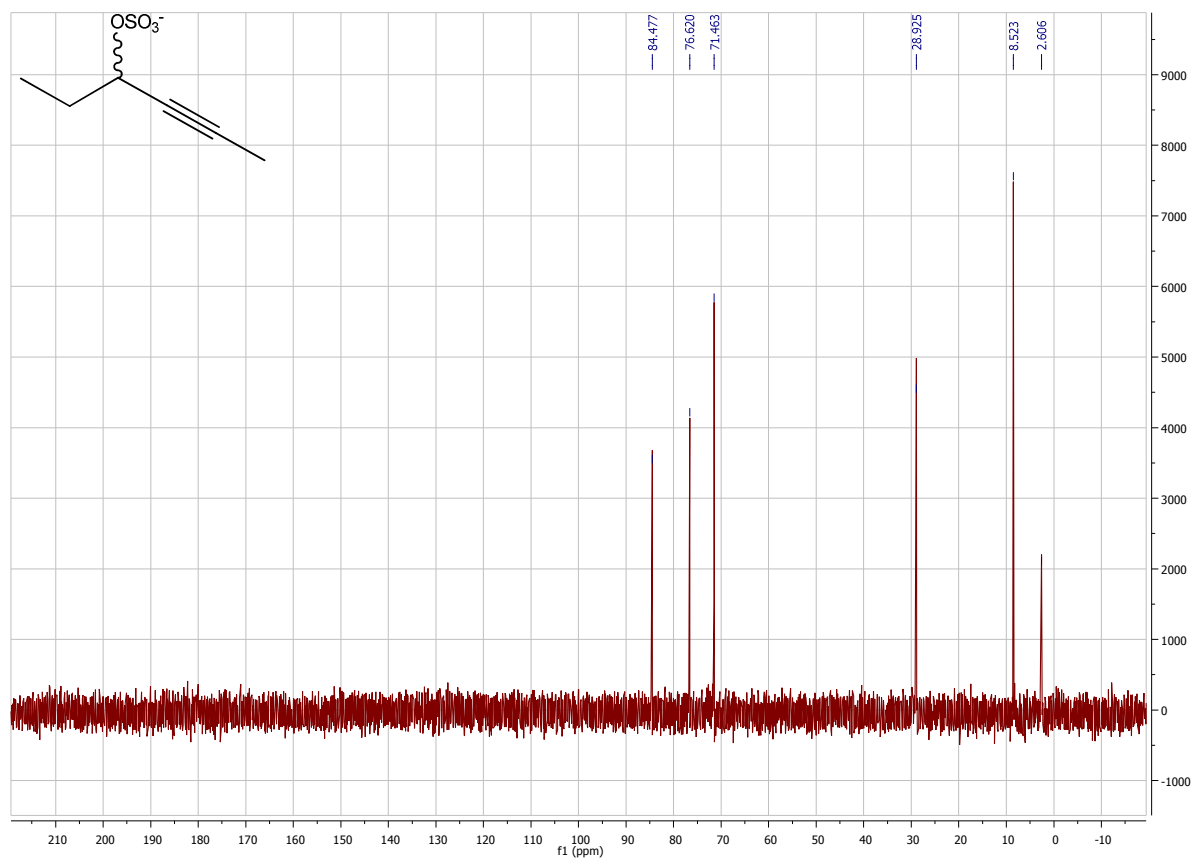
^{1}H-NMR (19a) *rac*-4-Heptyn-3-yl sulfate; D_2O
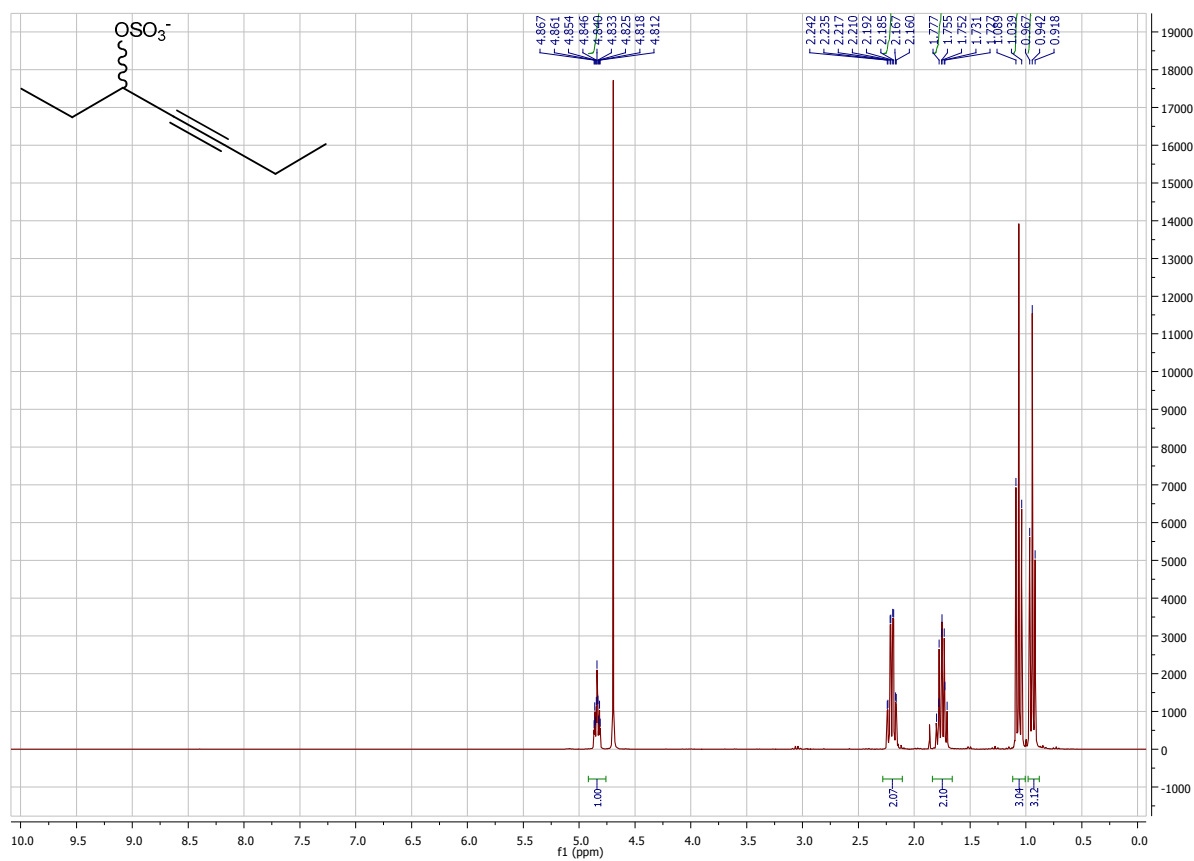
¹³C-NMR (19a) *rac*-4-Heptyn-3-yl sulfate; D₂O
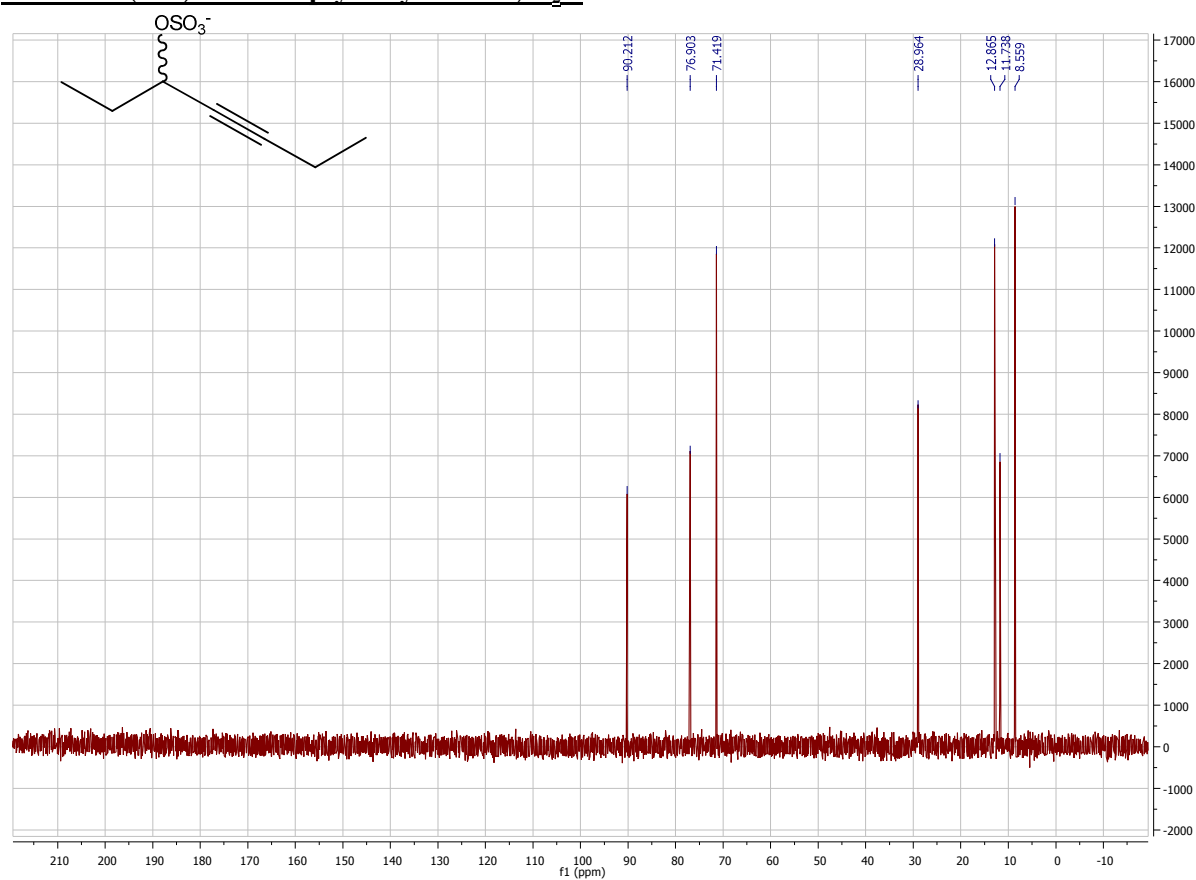
¹H-NMR (20a) *rac*-1-Hexyn-3-yl sulfate; D₂O
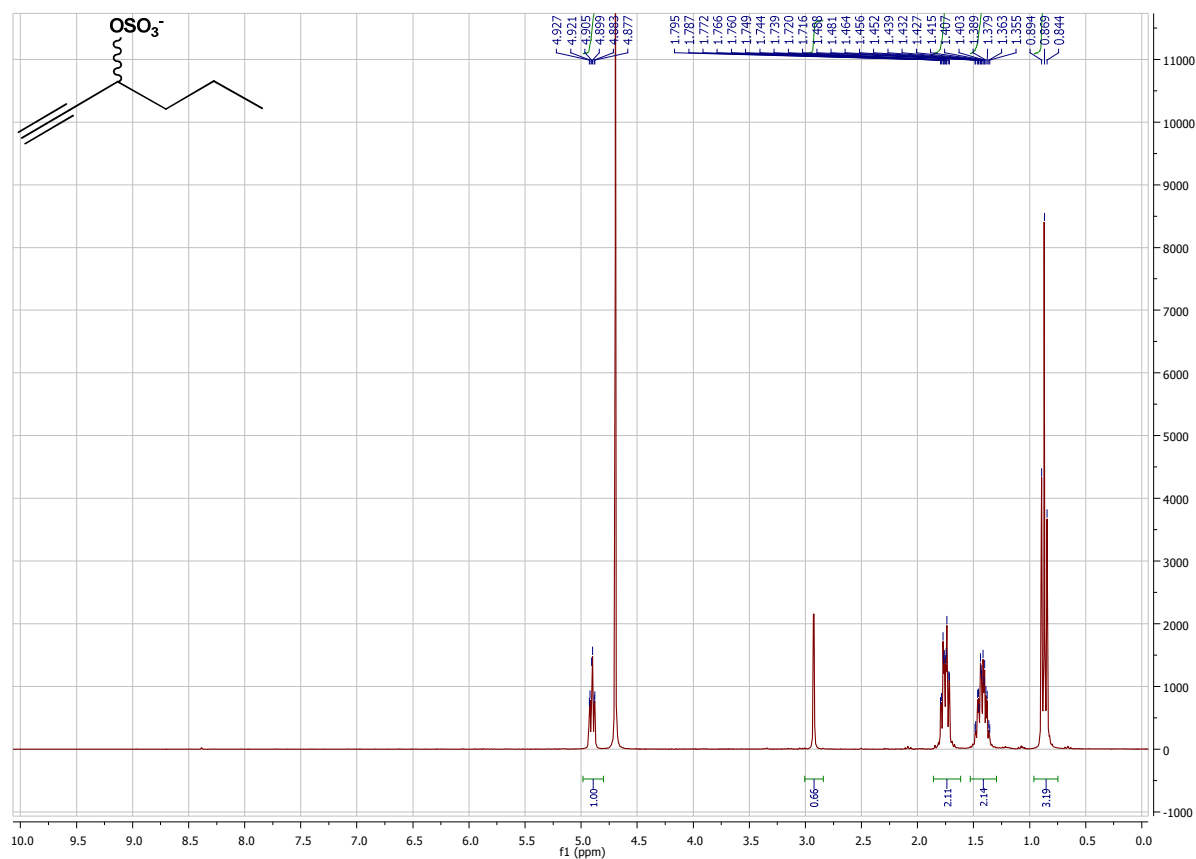
^{13}C-NMR (20a) *rac*-1-Hexyn-3-yl sulfate; D$_2$O
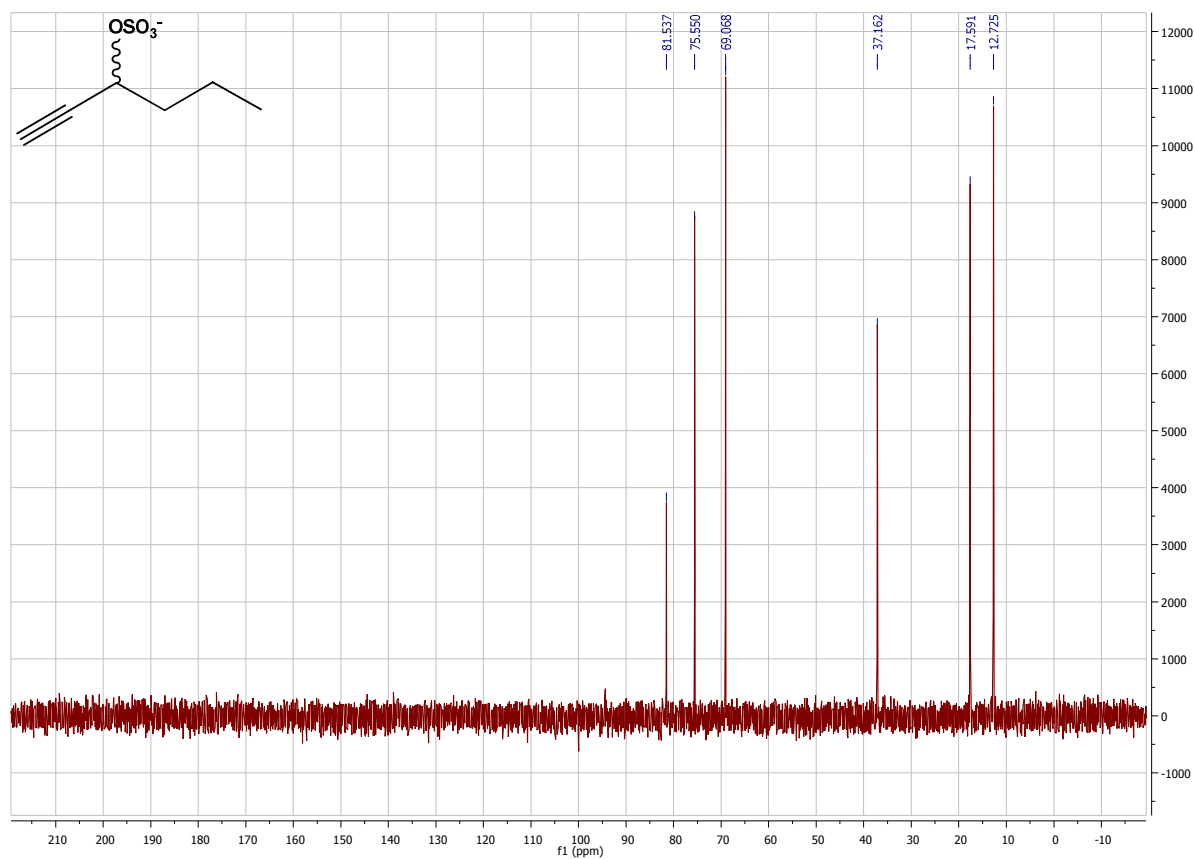
^{1}H-NMR (21a) *rac*-5-Methyl-1-hexyn-3-yl sulfate; DMSO-d$_6$
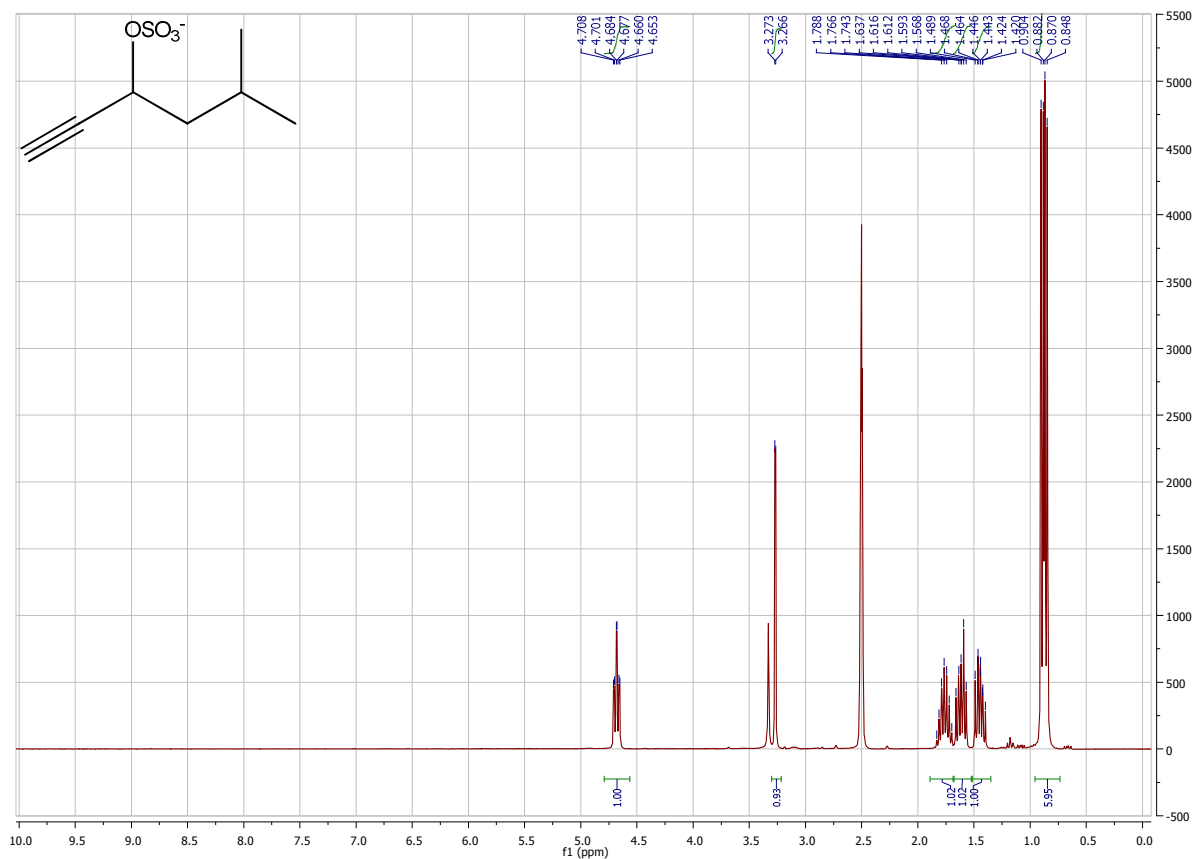
¹³C-NMR (21a) *rac*-5-Methyl-1-hexyn-3-yl sulfate; DMSO-d₆
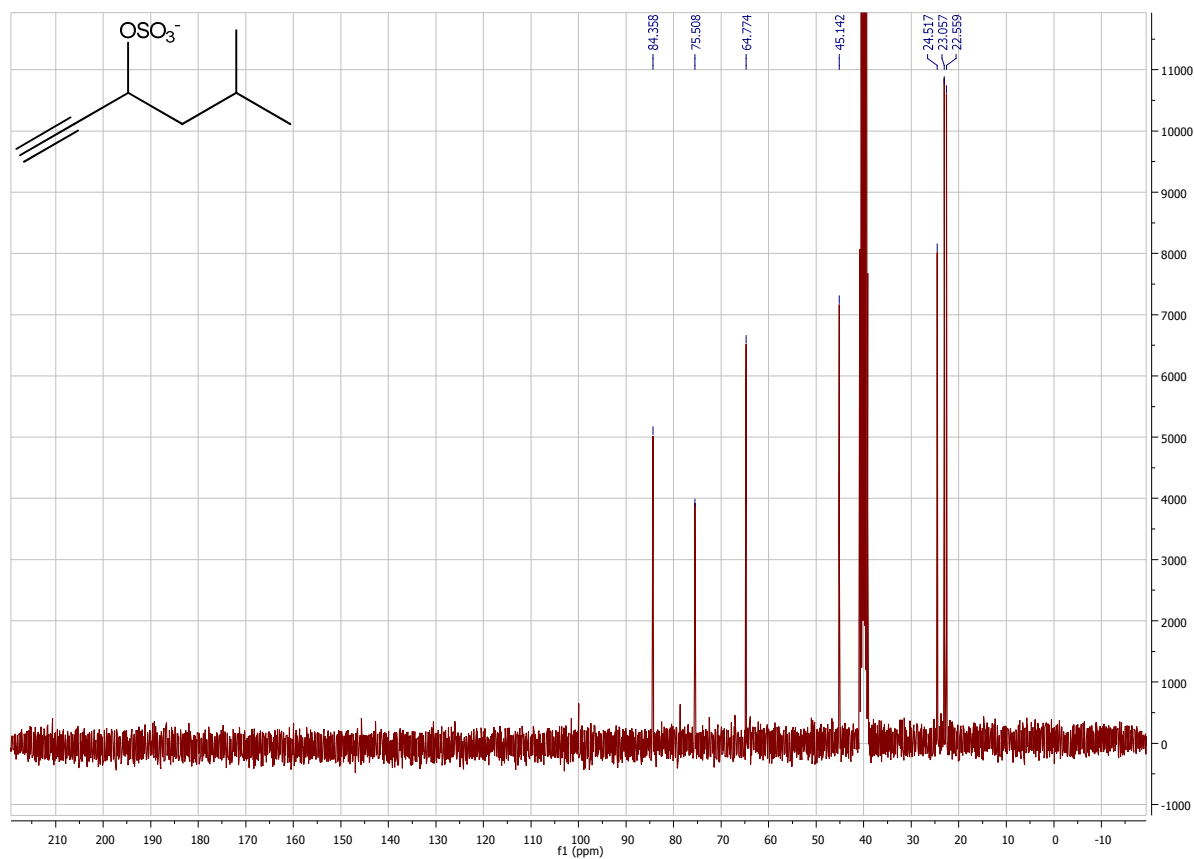
¹H-NMR (23a) *rac*-1-Phenylethyl sulfate; DMSO-d₆
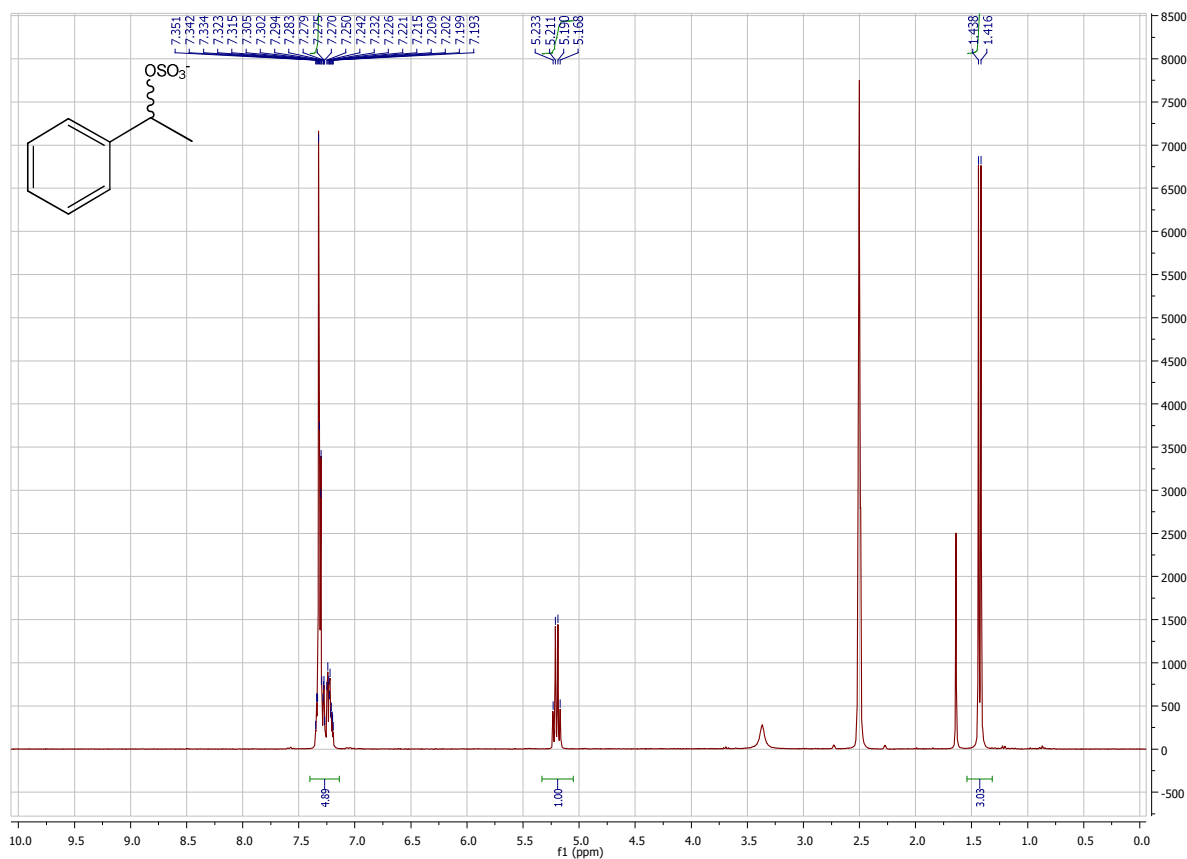
^{13}C-NMR (23a) *rac*-1-Phenylethyl sulfate; DMSO-d$_6$
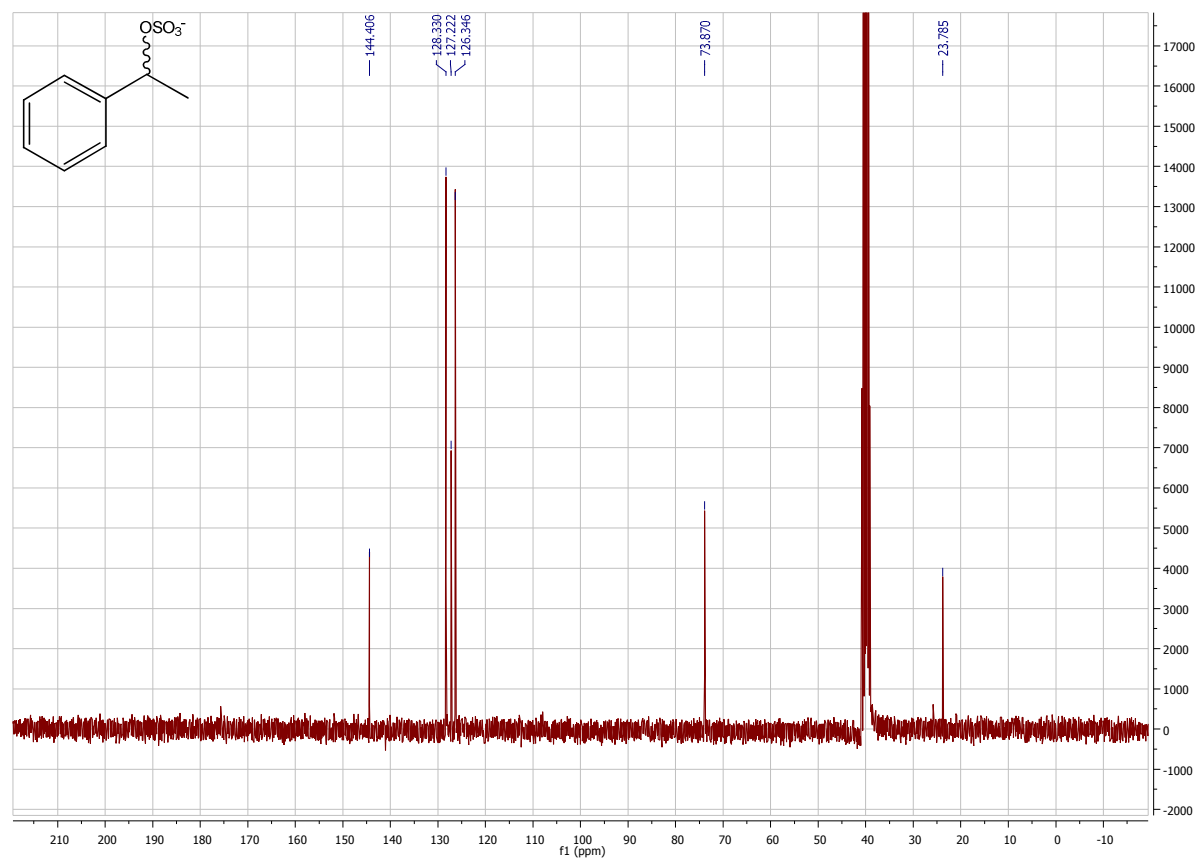
^{1}H-NMR (24a) *rac*-1-[3,5-bis(trifluoromethyl)phenyl]ethyl sulfate; DMSO-d$_6$
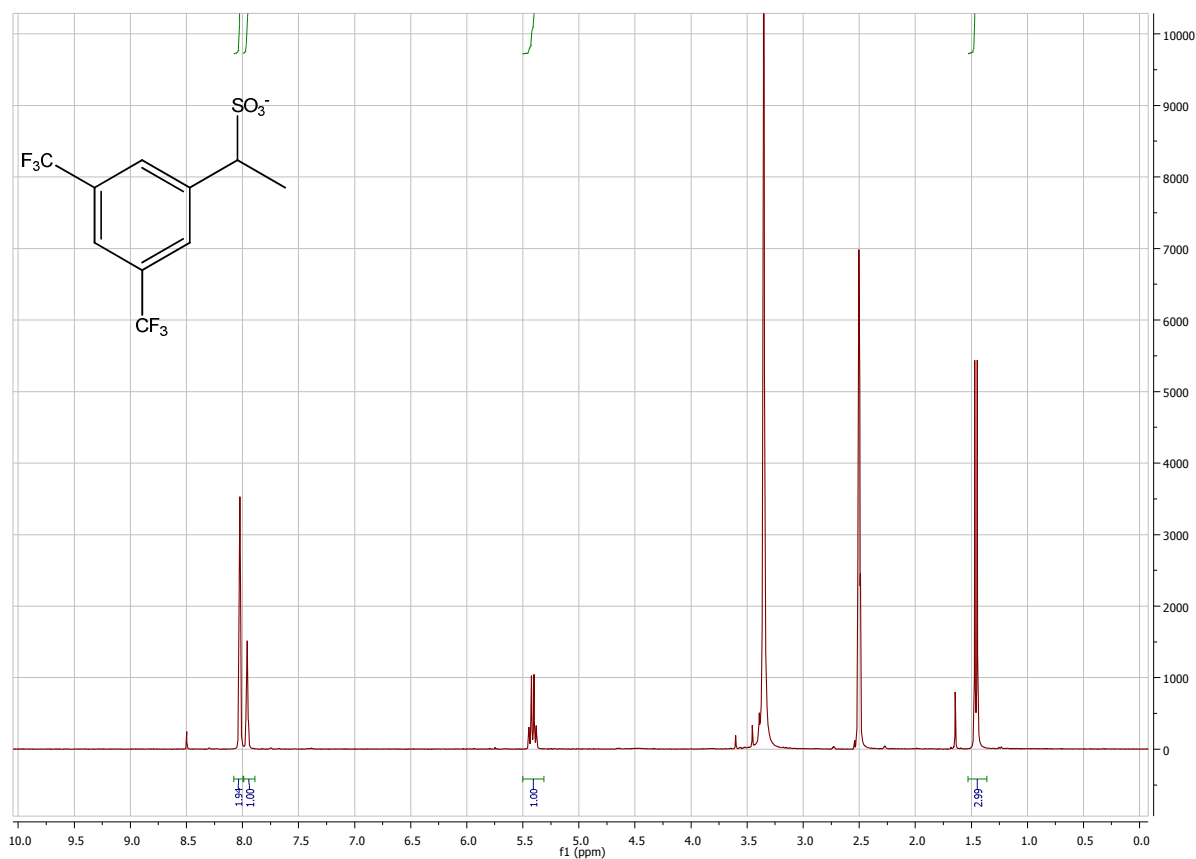
¹³C-NMR (24a) *rac*-1-[3,5-bis(trifluoromethyl)phenyl]ethyl sulfate; DMSO-d₆
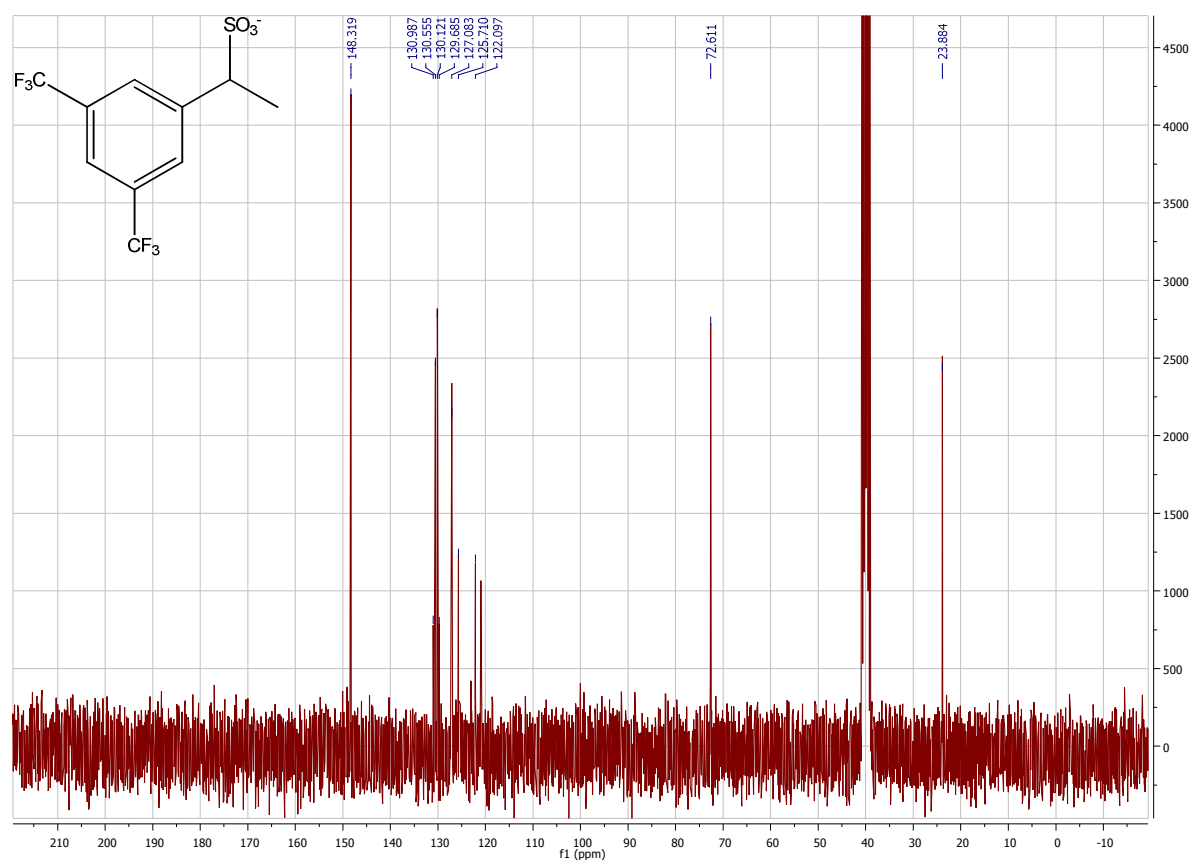
References and Notes

[1]   M. Schober, P. Gadler, T. Knaus, H. Kayer, R. Birner-Grünberger, C. Gülly, P. Macheroux, U. Wagner, K. Faber, *Org. Lett.* **2011**, *13*, 4296-4299.

[2]   H. Hirata, M. Ikushima, M. Watanabe, K. Kawauchi, M. Miyagishi, Y.-W. Chen, H. Yanagishita, *J. Oleo. Sci.*, **2003**, *52*, 99-108.

[3]   F. D. Özdemirhan, M. Celik, S. Atli, C. Tanyeli, *Tetrahedron: Asymmetry*, **2006**, *17*, 287-291.

[4]   X. Zhang, Z. Lu, C. Fu, S. Ma, *Org. Biomol. Chem.*, **2009**, *7*, 3258–3263.

[5]   J. L. Abad, F. Camps, G. Fabriàs, *J. Am. Chem. Soc.* **2007**, *129*,15007-15012.

[6]   F. A. Marques, A. Zimmermann, C. E. Delay, B. H. L. N. Sales Maia, *Lett. Org. Chem.*, **2008**, *5*, 69-72.

[7]   M. Takakiyo, Y. Tomoyuki, Mitsubishi Gas Chem. Co. Inc., EP893422 A1, **1999**; *Chem. Abstr.* **1999**: 90411, CAN *130*:138398.

[8]   K. Nakamura, K. Takenaka, *Tetrahedron: Asymmetry*, **2002**, *13*, 415-422.

[9]   R. Fu, J. Chen, L.-C. Guo, J.-L. Ye, Y.-P. Ruan, P.-Q. Huang, *Org. Lett.*, **2009**, *11*, 5242–5245.

[10]   E. Keinan, E. K. Hafeli, K. K. Seth, R. Lamed, *J. Am. Chem. Soc.* **1986**, *108*, 162-169.

[11]   *Use of Isothermal Titration Calorimetry to Measure Enzyme Kinetics Parameters*, ITC Application Note, MicroCal, 22 Industrial Drive East, Nothhampton, MA 01060, USA.

[12]   http://web.expasy.org/protparam/.

Chapter 8

Structure and mechanism of an inverting alkylsulfatase from *Pseudomonas* sp. DSM6611 specific for secondary alkyl sulfates

AUTHOR CONTRIBUTIONS

The manuscript has been sent to FEBS Journal and is now under minor revision. The research was carried out in cooperation with the Department of Chemistry, Organic & Bioorganic Chemistry (University of Graz, Austria), and the Institute of Molecular Biosciences, Structural Biology (University of Graz, Austria). Crystallization experiments and structure elucidations were performed at the Institute of Molecular Biosciences, Structural Biology (University of Graz, Austria) by ULRIKE WAGNER. My contribution consisted in the expression and purification of SdsA1 and the cloning, expression and purification of Pisa1 and its variants. Moreover, I performed the activity assays and determined the kinetic parameters of Pisa1 and its variants for the hydrolysis of alkyl sulfates.

Structure and mechanism of an inverting alkylsulfatase from *Pseudomonas* sp. DSM6611 specific for secondary alkylsulfates

Tanja Knaus[1], Markus Schober[2], Bernhard Kepplinger[3], Martin Faccinelli[1], Julia Pitzer[1], Kurt Faber[2], Peter Macheroux[1*] and Ulrike Wagner[3*]

[1]Institute of Biochemistry, Graz University of Technology, A-8010 Graz, Austria

[2]Institute of Chemistry, University of Graz, A-8010 Graz, Austria

[3]Institute of Molecular Biosciences, University of Graz, A-8010 Graz, Austria

Running title: *Secondary alkylsulfatase Pisa1*

Corresponding authors: Prof. Dr. Ulrike Wagner, University of Graz, Institute of Molekulare Biowissenschaften / Strukturbiologie, Humboldtstrasse 50/III, A-8010 Graz, Austria; Tel.: +43-316-3805431; Fax: +43-316-3805431; Email: ulrike.wagner@uni-graz.at and

Prof. Dr. Peter Macheroux, Graz University of Technology, Institute of Biochemistry, Petersgasse 12, A-8010 Graz, Austria; Tel.:+43-316-8736450; Fax: +43-316-8736952; Email: peter.macheroux@tugraz.at

Abbreviations: Pisa1, *Pseudomonas* inverting *sec*-alkylsulfatase 1; SDS, sodium dodecyl sulfate; SdsA1, alkylsulfatase from *Pseudomonas aeruginosa*;

Key words: Enzyme mechanism, Hydrolases, Metalloenzymes, X-ray crystallography, Zinc

_____

SUMMARY

A highly enantio- and stereoselective secondary alkylsulfatase from *Pseudomonas* sp. DSM6611 (Pisa1) was heterologously expressed in *Escherichia coli* BL21 and purified to homogeneity for kinetic and structural studies. Structure determination of Pisa1 by X-ray crystallography showed that the protein belongs to the family of metallo-β-lactamases with a conserved binuclear Zn^{2+} cluster in the active site. In contrast to a closely related alkylsulfatase from *Pseudomonas aeruginosa* (SdsA1), Pisa1 showed preference for secondary rather than primary alkyl sulfates and enantioselectively hydrolyzed the *(R)* enantiomer of *rac*-2-octyl sulfate yielding (*S*)-2-octanol with inversion of absolute configuration as a result of C-O bond cleavage. In order to elucidate the mechanism of inverting sulfate ester hydrolysis, for which no counterpart in chemical catalysis exists, we designed variants of Pisa1 guided by the three-dimensional structure and docking experiments. In the course of these studies we identified an invariant histidine residue (His317) near the sulfate binding site as the general acid for crucial protonation of the sulfate leaving group. Additionally, amino acid replacements in the alkyl-chain binding pocket generated an enzyme variant, which lost its stereoselectivity towards *rac*-2-octyl sulfate. These findings are discussed in light of the potential use of this enzyme family for applications in biocatalysis.

Introduction

Sulfatases are a heterogeneous group of enzymes, which catalyze the cleavage of the sulfate ester bond yielding the corresponding alcohol and hydrogen sulfate[1-4]. In contrast to the majority of hydrolases, such as proteases and carboxyl ester hydrolases, which do not alter the stereochemistry of the substrate during catalysis[5], the stereochemical course of sulfate ester hydrolysis can be controlled by choice of the type of enzyme: Depending on the mechanism of the enzyme, cleavage of the S-O or the C-O bond of a *sec*-alkyl sulfate causes retention or inversion of the stereogenic carbon atom, respectively (Scheme 1).
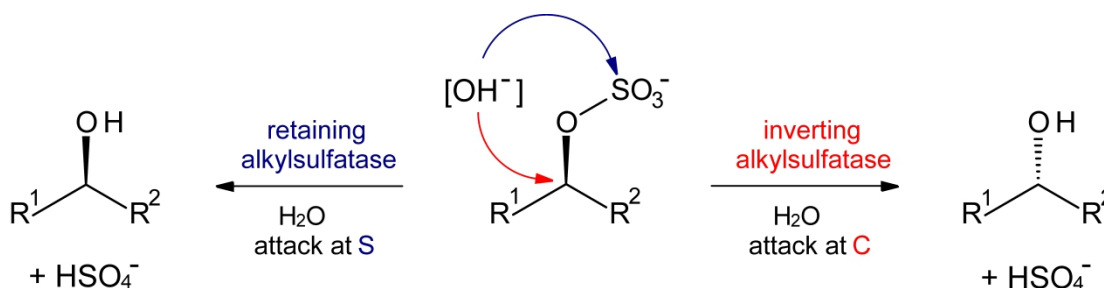
Scheme 1 Stereochemical course of retaining and inverting alkylsulfatases**.**

Previously, sulfatases were classified according to their preferred substrate-type into aryl-, carbohydrate- and alkyl-sulfatases[3]. More recently, sulfatases were re-classified based on mechanistic considerations[6] into (i) aryl- and carbohydrate sulfatases acting on sulfated carbohydrates and on sulfated steroid hormones[7], (ii) sulfatases that oxidatively cleave a sulfate ester at the expense of α-ketoglutarate as electron acceptor to yield an aldehyde and inorganic sulfate[8] and (iii) sulfatases belonging to the family of metallo-β-lactamases[6]. In the case of aryl- and carbohydrate sulfatases a hydrated α-formylglycine active site nucleophile attacks the S-atom of the sulfate ester resulting in retention of the absolute configuration at the carbon. In the course of the reaction of α-ketoglutarate dependent sulfatases the stereogenic carbon center is destroyed. The third class of sulfatases is represented by sodium dodecyl sulfatase (SdsA1) produced by *Pseudomonas aeruginosa* and enables the bacterium to survive under bacteriocidal conditions by hydrolysing SDS.

Based on the crystal structure of SdsA1 it was proposed that sulfate ester cleavage occurs by nucleophilic attack of a water molecule, which is activated by a highly conserved binuclear Zn^{2+} cluster. Since the substrate is an achiral *prim*-sulfate ester, the stereochemical course of SdsA1-hydrolysis is not 'visible' and thus was not investigated in detail. However, based on the short distance of the presumed nucleophile (Water2) to the S-atom of 1-decyl sulfonate (used as substrate-surrogate inhibitor), S-O bond cleavage was assumed implying retention at the carbon. SdsA1 was unreactive on sulfated sugars and aryl sulfates. *Sec*-alkyl sulfate esters were not investigated as potential substrates[6].

The existence of alkylsulfatases acting on *sec*-sulfate esters was derived from the chance observation of a bacterial contamination in a commercial shampoo preparation[9]. During subsequent studies, a range of bacterial strains possessing *sec*-alkyl sulfatase activity were isolated from sewage sludge, two of which were investigated in detail[10]: Depending on the culture conditions, *Pseudomonas* C12B (NCIMB 11753 = ATCC 43648) produced up to three, and *Comamonas terrigena* (NCIMB 8193) formed two *sec*-alkyl sulfatases. First hints on the enantiopreference of these enzymes came from the observation that *rac-sec*-alkyl sulfates were rapidly hydrolyzed until 50% conversion, while further reaction took place at a much slower rate[11-13]. The stereochemical course of these enzymes with respect to retention versus inversion was studied using enantioenriched substrates[12]. Subsequent studies revealed *Rhodococcus ruber* DSM 44541 as a promising source of an inverting *sec*-alkylsulfatase[14, 15]. Although the enzyme responsible for the stereoselective hydrolysis of *sec*-sulfate esters could be purified to a certain extent, no information concerning its amino acid sequence could be obtained due to the instability of the protein. In search for a more stable alternative, an extended screening for *sec*-alkyl sulfatase activity was performed among *Pseudomonads* and close relatives thereof, which revealed *Pseudomonas* sp. DSM 6611 as the most promising candidate[16]. This strain was isolated from soil for its ability to degrade halogenated aromatic compounds, such as 4-fluorobenzoate[17]. A secondary alkylsulfatase was isolated and sequenced by peptide mass fingerprinting. In combination with the full genomic sequence of *Pseudomonas* sp. DSM 6611, the predicted open reading frame was determined and the gene was cloned and heterologously expressed in *E. coli* BL21[18, 19]. The protein termed Pisa1 (*Pseudomonas inverting secondary alkylsulfatase 1*) displayed the desired catalytic properties: Hydrolysis of (*R*)-2-octyl sulfate proceeded quantitatively through complete inversion of configuration yielding (*S*)-2-octanol. The stereochemical course of inversion was independently verified by hydrolysis of unlabelled ^{16}O-(*R*)-2-octyl sulfate in ^{18}O-labelled buffer (label >97%), which showed complete incorporation of the ^{18}O-label in the product alcohol. The hydrolysis of *rac*-2-octyl sulfate ceased at 50% conversion to furnish (*S*)-2-octanol and unreacted (*S*)-2-octyl sulfate, indicating perfect enantioselectivity (enantiomeric ratio, E >200).

In a previous study we investigated the properties of Pisa1 and SdsA1 with regard to the stereochemical course of the sulfate ester cleavage and substrate specificity[18]. This study revealed that both alkylsulfatases cleave the C-O bond resulting in inversion at the carbon. On the other hand, it was demonstrated that the two enzymes have different substrate specificities towards primary and secondary alkyl sulfates: while Pisa1 prefers *sec*-octyl sulfate, SdsA1 shows a 140-fold higher catalytic efficiency towards 1-octyl sulfate. These findings ignited our interest in the structure of Pisa1 as a basis to rationalize the

conservation of the cleavage mechanism while at the same time the enzymes have acquired different substrate specificities. Here, we report the elucidation of the Pisa1 structure by X-ray

crystallography combined with a site-directed mutagenesis study to identify residues in the active site of the enzyme crucial for catalysis.

Results

Structure description

The overall structures of SdsA1 and Pisa1 exhibit a high degree of similarity. The rms distances between the C_a atoms of SdsA1 with sulfate in the active site (PDB code: 2cg2) and orthorhombic crystals of wt Pisa1 (hitherto called wt form-I) calculated using program LSQMAN[20, 21] showed a maximum and an average distance of 15.01 Å and 1.5 Å, respectively. Figure 1 shows a comparison between the structure of Pisa1 (PDB code: 2yhe) and SdsA1 (PDB code: 2cg2).
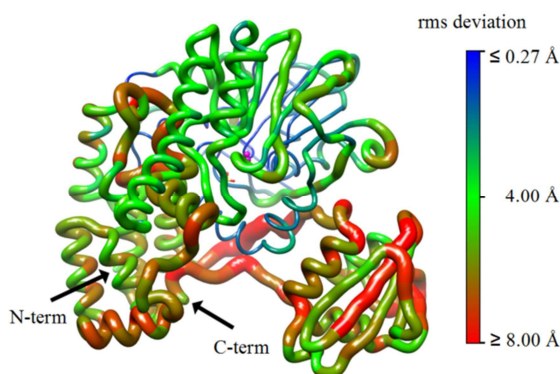
Figure 1. Comparison of overall structures of Pisa1 and SdsA1. C_α-rms differences calculated with lsqman are represented by the color as well as the thickness of the worm. The overall rms is 1.5 Å. The picture was created with CHIMERA [22].

As described for the SdsA1 structure, Pisa1 also forms a dimer. Each protomer consists of the same three domains as SdsA1, the N-terminal, catalytic, αββα-sandwich domain, an α-helical dimerization domain and an α,β-

mixed C-terminal domain. In contrast to SdsA1 where the dimer is formed via a crystallographic 2-fold axis, the active Pisa1 unit consists of two molecules related by non-crystallographic symmetry. In wt-form-I and the double variant Ser233Tyr/Phe250Gly structures these dimers form an imperfect $6_{1.2}$-helix parallel to the crystallographic c-axis, with rotation angles between 65° and 68° and a translation of approximately 0.2 fractional coordinates. While all six molecules of the wt form-I structure are adequately defined, both molecules of wt from-II (trigonal crystals) have poorly defined regions: while molecule A has a very well defined N-terminal part ranging from residue 30-531, the density for residues 532-661 is not very well defined but nevertheless interpretable at least for the main chain atoms. The density for molecule B is almost not interpretable for residues 31-207, 237-306, 402-417 and the loop region linking the N and the C-terminal domain (527-545). In contrast to wt form-I structure no sulfate was found in the ligand binding site of wt form-II molecule A. The electron density for the double variant structure Ser233Tyr/Phe250Gly is weak but especially around the active site residues the density was well defined, thus making it useful for docking experiments. Figure 2 shows a B-factor representation of the dimer for the three crystal structures. Only those parts forming the

interface between the dimers are well ordered in both molecules. Despite the lower overall quality of the structure obtained from wt form-II crystals, the loop regions are well defined in contrast to those of the structure derived from wt form-I. Figure 3 shows a representative section of the well-defined parts of the electron density of Pisa1 wt form-I, form-II and the double mutant Ser233Tyr/Phe250Gly, respectively. A summary of the crystallographic data and refinement statistics is provided in Table 1 (see also Materials & Methods).
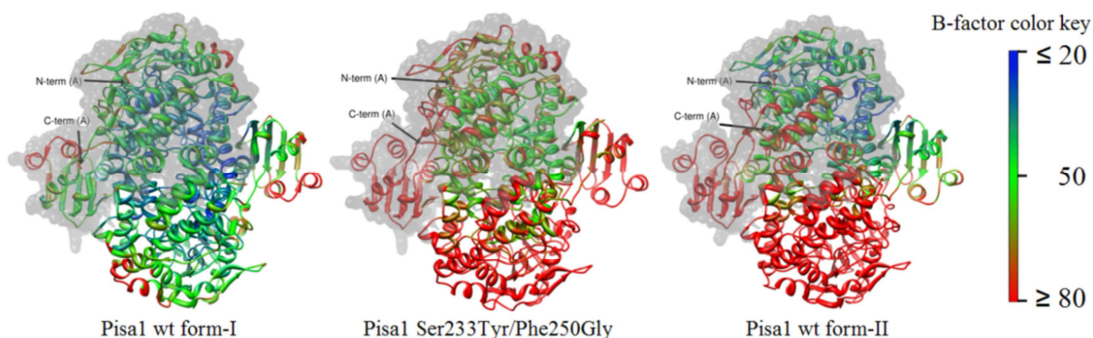
Figure 2. Worm-style plots of the active dimers created with CHIMERA [22]. The residual B-factor (defined as $B_i = 8\pi^2 U_i^2$ where U_i^2 is the mean square displacement of atom i) reveals the quality of the structures. For a better understanding molecule A of each structure is marked by a transparent surface and the N- and C-terminus are denoted. The view is approximately down the 2-fold axis.
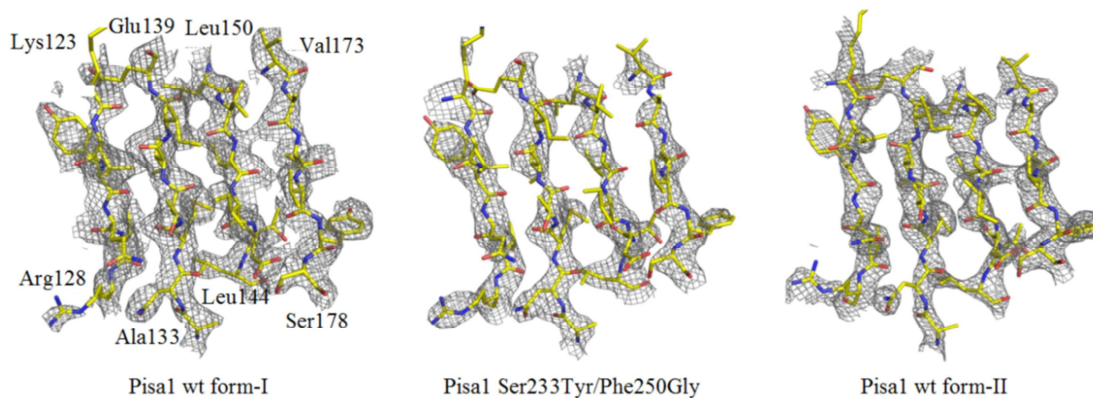
Figure 3. Electron density (drawn with a σ-level of 1.6) representation of the β-strands A123-128, A133-139, A144-150, A173-178 for all three crystal structures. Pictures were created with PyMol.

Comparison of the zinc binding sites

Both alkylsulfatases, SdsA1 and Pisa1, have conserved catalytic machineries consisting of a binuclear zinc center. Comparison of the zinc binding sites of SdsA1 and Pisa1 (PDB codes: 2cg2 and 2yhe, respectively) reveals a high degree of similarity despite the different substrate- and regio-specificities of the enzymes: all of the interaction partners of the zinc centers are identical with the exception of the bridging acidic amino acid. This position is occupied by aspartic acid in Pisa1 instead of a glutamic acid in SdsA1 (Figure 4).

Table 1. Data collection and refinement statistics

	Pisa1 wt form-I	Pisa1 Ser233Tyr/Phe250Gly	Pisa1 wt form-II		
Space group	$P2_12_12_1$	$P2_12_12_1$	$P3_121$		
Cell dimensions (Å)	a=147.69	a=75.2	a=142.0		
	b=54.53	b=202.0	c=119.7		
c=94.90	c=248.5				
Dataset			remote	inflection	peak
Wavelength (Å)	1.27030	1.0044	1.18076	1.28099	1.28031
Resolution (Å)	19.9-2.7	45.7-3.0	19.6-2.7	10.0-2.7	10.0-2.7
Unique reflections	73894 (10637)[a]	62063 (9310)[a]	36072 (5553)[a]	33729 (5172)[a]	33799 (5225)[a]
Overall completeness (%)	99.8 (100)[a]	78.4 (79.4)[a]	93.3 (100)[a]	92.6 (98.1)[a]	92.7 (98.9)[a]
R_{sym}[1] (%)	17.0 (80.1)[a]	7.4 (54.4)[a]	8.5 (66.1)[a]	9.6 (57.8)[a]	10.0 (60.8)[a]
Multiplicity	10.1 (9.3)[a]	3.0 (2.7)[a]	7.4 (7.5)[a]	7.3 (6.9)[a]	7.3 (6.9)[a]
$I/(\sigma I)$	9.1 (1.7)[a]	10.8 (2.0)[a]	18.8 (3.1)[a]	15.4 (3.2)[a]	15.9 (3.2)[a]
Wilson plot B-factor (Å²)	53.43	84.30	65.36	38.12	84.12
R_{cryst}[2]	18.3 (26.0)[a]	18.7 (32.5)[a]	22.1 (26.6)[a]		
R_{free}[3] (%)	23.0 (33.1)[a]	25.6 (38.8)[a]	28.1 (32.2)[a]		
Protein atoms	29825	29647	9900		
Water molecules	1166	318	136		
Heterogen atoms	42	42	9		
RMSD bond lengths (Å)	0.014	0.010	0.010		
RMSD bond angles (°)	1.834	1.316	1.23		
Mean B-factor[4] (Å²)	47.74	86.83	77.70		

[a] Numbers in parenthesis refer to high resolution shells.
[1] $R_{sym} = \Sigma|(I - \langle I \rangle)/\Sigma(I)$ where I is the observed intensity
[2] $R_{cryst} = (\Sigma \|F_{obs}| - |F_{calc}\|) / (\Sigma|F_{obs}|)$

Comparison of the substrate binding sites

The sulfate group in SdsA1 is bound through seven hydrogen bonds; however, two of these entail the guanidinium nitrogens of Arg323 with approximately 50% occupancy. In the case of Pisa1, the sulfate group appears to be more tightly bound by eight hydrogen bonds (Figure 5). In both structures an arginine claw is involved in sulfate binding (Figure 5). The wt form-II of Pisa1 shows no density at the sulfate position. Interestingly, the arginine claw is in an "open" conformation in this structure. In SdsA1 this arginine (Arg312) adopts alternate conformations with the side chain oriented towards the sulfate or is pointing away from the sulfate group with approximately equal occupancy. In SdsA1, this alternate conformation of Arg312 is found in the sulfate-free as well as in the sulfate-bound form.
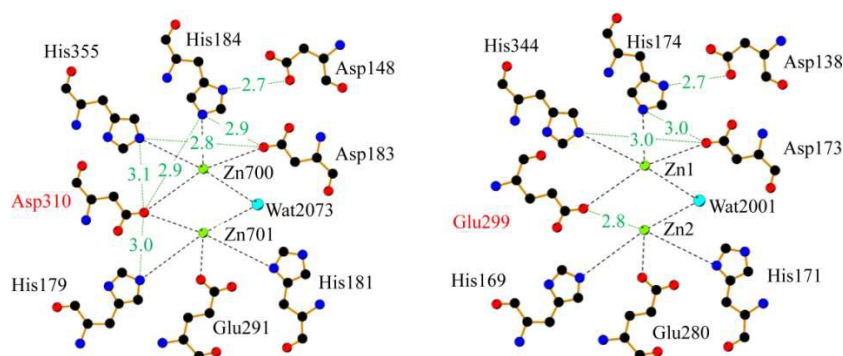
Figure 4. Schematic comparison of the zinc binding sites of Pisa1 (left) and SdsA1 (right) containing bound sulfate made with ligplot [23]**.**

A water molecule near the sulfate group was postulated as the nucleophile for the hydrolysis in SdsA1. This assumption does not hold any longer since SdsA1 also works under
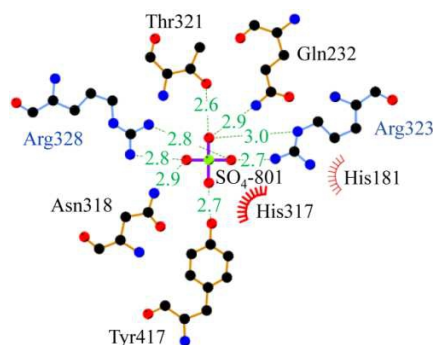
inversion. A water molecule in the right position to perform a nucleophile attack at the chiral carbon could not be found in any of the Pisa1 structures.
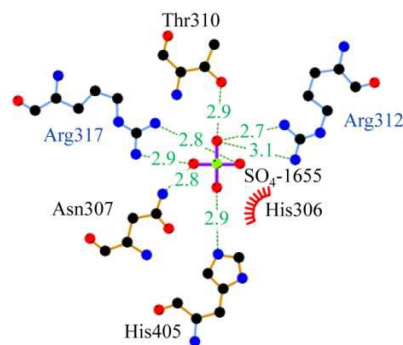
Figure 5. Schematic comparison of the sulfate binding sites of Pisa1 (left) and SdsA1 (right) generated with ligplot [23]. The arginine residues involved in the Arginine Claw are displayed in blue.

For a better understanding of the substrate binding pocket, we performed soaking and co-crystallization experiments with *R* and *S* substrate and inhibitor (sodium 1-nonanesulfonate) with wild type Pisa1 and the double variant crystals. Unfortunately, the resulting density showed no incorporation of substrate/inhibitor. A possible reason might be that a sulfate, presumable already incorporated during expression, binds very tightly and hence might be difficult to be displaced by substrates or inhibitors under soaking conditions. We therefore crystallized Pisa1 expressed without addition of $ZnSO_4$ but crystals obtained from this expression batch did not diffract at all. Therefore we resorted to docking experiments. The results could not fully rationalize the preference of wild-type Pisa1 for *R* enantiomers. However, they highlight an interesting difference between wild-type enzyme and the Ser233Tyr/Phe250Gly double variant: As shown in Figure 6, the hydrophobic side chain of the substrates points towards the

entrance of the reaction cavity in the wild-type, whereas in the double variant the hydrophobic tail of the substrate also points in the hydrophobic hole created by the Phe250Gly exchange (Figure 6).
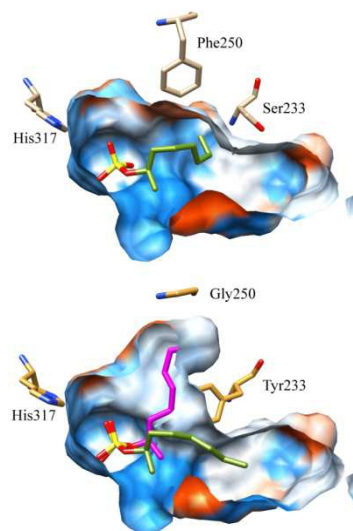
Figure 6. The figure shows a cut through the active site of Pisa1 wild-type (left) with a docked (*R*)-2-octyl sulfate (green) and Pisa1 Ser233Tyr/Phe250Gly double variant (right) with both docked enantiomers, *R* colored in green and *S* in magenta. The mutated residues are displayed as well as His317 identified as the general acid. The Phe250Gly mutation opens the possibility for an alternative binding of the side chain thereby allowing both enantiomers to react.

Due to the narrow entrance into this binding pocket it will not be equally populated as the "normal" one. It is conceivable that the *S* substrate can utilize this additional binding site and thus make it accessible for nucleophile attack.

Kinetic characterization of wild-type Pisa1 and variants

Kinetic analysis of the alkyl sulfate ester cleavage eludes simple spectrophotometric assays and hence we have employed microcalorimetry to obtain kinetic parameters for SdsA1 and Pisa1. Toward that end, we first determined the reaction enthalpy for sulfate ester cleavage (Figure 7) and then measured the rate of enzymatic substrate turnover as a function of substrate concentration (Figure 8). SdsA1 was previously described as a primary alkylsulfatase with a preference towards sodium dodecylsulfate (SDS), however, no kinetic parameters were reported for this or other substrates. Because Pisa1 has a preference for shorter alkyl-chains, we have compared the two alkylsulfatases using 1- and *(R)*-2-octyl sulfate. As presented in Table 2, SdsA1 is clearly more specific for 1-octyl sulfate as it has a 6.5 times higher k_{cat} and a 22-fold lower K_M than Pisa1 resulting in a 141 times higher k_{cat}/K_M for the primary model substrate. In contrast, Pisa1 exhibited a 194 times higher k_{cat}/K_M for the secondary model substrate leading to the classification of the two enzymes as primary (SdsA1) and secondary (Pisa1) alkylsulfatases.
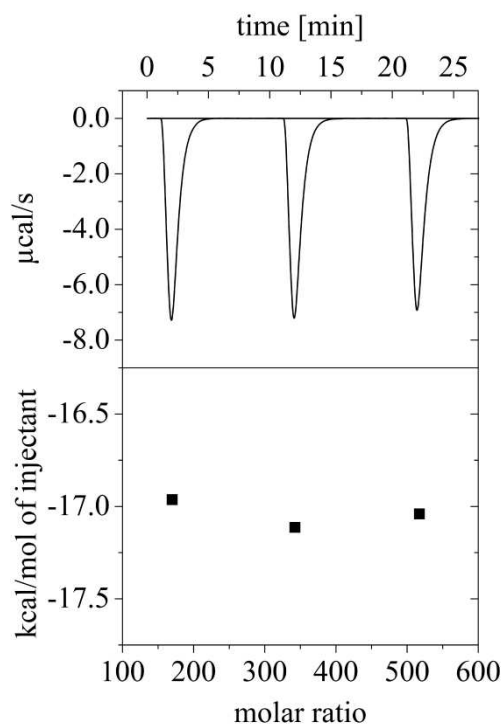
Figure 7. Determination of ΔH for the hydrolysis of *(R)*-2-octyl sulfate by isothermal titration calorimetry. Pisa1 (125 nM) was titrated with *(R)*-2-octyl sulfate (1.5 mM, 3 x 20 µl, duration time 39.9 s, spacing time 1,000 s) in 100 mM Tris/HCl, pH 8.2 buffer at 25 °C yielding a ΔH of -16.76 kcal/mol.

We have recently demonstrated that alkyl sulfate ester cleavage by Pisa1 results in the inverted absolute stereoconfiguration at the chiral carbon center, *i.e.* (*R*)-2-octyl sulfate is hydrolyzed to (*S*)-2-octanol. This stereochemical course of the reaction implies that the hydroxide attacks the chiral carbon rather than the sulfur atom of the sulfate group leading to C-O bond cleavage. Such a reaction mechanism generates a negative charge on the oxygen and requires synchronous protonation by an active site amino acid residue. In order to identify this putative general acid, we have replaced histidine 317 (His317Ala) and tyrosine 417 (Tyr417His, Phe and Asp). The His317Ala variant showed a 1300-fold

decrease in k_{cat} while the K_M was only slightly increased from 151 to 240 µM. All of the tyrosine 417 variants exhibited severely impaired catalytic parameters: The Tyr417Asp variant was found to be inactive whereas the Tyr417His and Tyr417Phe variants showed 238 and 114-fold decreased k_{cat} values, respectively, and 13- to 65-fold increased Michaelis-Menten parameters (Table 2).
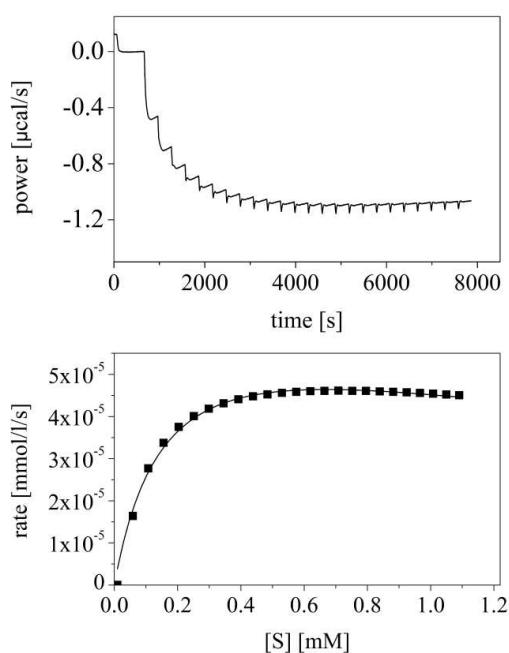
Figure 8. Representative example for the determination of the kinetic parameters by isothermal titration calorimetry. The measurement shows the titration of 20 nM Pisa1 with *(R)*-2-octyl sulfate (7 mM, 24 x 10 µl, duration time 20 s, spacing time 300 s) in 100 mM Tris/HCl, pH 8.2 buffer at 25 °C. Panel A: Raw data which were obtained by ITC measurement, Panel B: Determination of k_{cat}, K_M and K_I using nonlinear least-squares fitting.

These results indicate that the imidazole side chain of His317 is mainly involved in protonation of the sulfate leaving group and not in substrate binding. On the other hand, it appears that the tyrosine side chain is critically involved in the correct positioning of the substrate's sulfate group as suggested by the large increase in the K_M value observed for the Tyr417Phe and Tyr417His variants.

Inspection of the substrate binding pocket of Pisa1 and comparison to the previously reported structure of SdsA1 prompted us to generate several single and double variants with the aim to enable hydrolysis of the (*S*)-2-octyl sulfate stereoisomer. Towards this aim, we have generated the Ser233Tyr, Phe250Gly single and the Ser233Tyr/Phe250Gly and Ser233Phe/Phe250Ser double variants. The two single variants Ser233Tyr and Phe250Gly showed 2.5- and 190-fold reduced catalytic activity towards (*R*)-2-octyl sulfate and the latter variant had a 28-fold higher K_M. Neither one of these variants had a measurable activity towards (*S*)-2-octyl sulfate. The two double variants had 3-4 times lower k_{cat} values and the double variant carrying the Phe250Gly replacement also exhibited a ten-fold higer K_M value. This double variant also showed activity towards (*S*)-2-octyl sulfate. As summarized in Table 2, the Ser233Tyr/Phe250Gly variant still processes the *R* stereoisomer with higher catalytic efficiency ($k_{cat}/K_M = 919$ M^{-1} s^{-1}) than the *S* stereoisomer ($k_{cat}/K_M = 46$ M^{-1} s^{-1}), which translates into an E-value of 20.

DISCUSSION

Despite the overall structural similarity of SdsA1 and Pisa1, the two proteins show only 43% identity and 61% similarity in amino acid sequence (see Supplementary Figure S2). In this context it is worth noting that *Pseudomonas* sp. DSM6611 was originally

selected for its ability to degrade halogenated aromatic compounds, *e.g.* 4-fluorobenzoate [17]. The preferred source for soil bacteria appears to be inorganic sulfur[3], however, in this environment sulfur is predominantly (>90%) bound to organic compounds as sulfonates and sulfate esters[24]. Since *Pseudomonas* sp.

DSM6611 was isolated from uncontaminated soil (Prof. Bernard Hauer, personal communication), Pisa1 could be part of the enzymatic repertoire of this strain to release sulfate from organic sulfate esters and hence make it available for bacterial growth in the soil.

Table 2. Kinetic parameters of Pisa1, its variants and SdsA1 for the hydrolysis of alkyl sulfates.

Enzyme	Substrate	k_{cat} [min^{-1}]	K_M [µM]	K_I [mM]	k_{cat}/K_M [M^{-1}·s^{-1}]
Pisa1	*(R)*-2-octylsulfate	**262 ± 62**	**151 ± 24**	**3.8 ± 1.1**	**28918**
SdsA1	*(R)*-2-octylsulfate	5.4 ± 0.5	605 ± 147	1.4 ± 0.2	149
Ser233Tyr	*(R)*-2-octylsulfate	104 ± 3.5	99 ± 4	19 ± 6	17508
Phe250Gly	*(R)*-2-octylsulfate	1.4 ± 0.2	4282 ± 2890	-	5
Ser233Tyr/Phe250Gly	*(R)*-2-octylsulfate	84 ± 1	1523 ± 30	-	919
Ser233Phe/Phe250Ser	*(R)*-2-octylsulfate	68 ± 7	115 ± 11	2.7 ± 0.5	9855
His317Ala	*(R)*-2-octylsulfate	0.2 ± 0.02	240 ± 111	-	14
Tyr417His	*(R)*-2-octylsulfate	1.1 ± 0.8	*	n.d.	-
Tyr417Phe	*(R)*-2-octylsulfate	2.3 ± 0.5	**	n.d.	-
Tyr417Asp	*(R)*-2-octylsulfate	no conversion (GC)			
Ser233Tyr/Phe250Gly	*(S)*-2-octylsulfate	23 ± 6	8420 ± 450	18.6 ± 4.2	46
SdsA1	1-octylsulfate	337 ± 106	30 ± 14	0.33 ± 0.07	187222
Pisa1	1-octylsulfate	52 ± 5	651 ± 25	0.5 ± 0.08	1331

* no saturation observed up to [S] =10 mM

** at least ≈ 2 mM

From a mechanistic point of view, chemical (non-enzymatic) sulfate ester hydrolysis is not a trivial task. In order to facilitate the departure of the sulfate moiety, it has to be converted into a good leaving group. This can be achieved by protonation of the sulfate ester (the pK_a of methyl monosulfonate was calculated/estimated to -8.4 and -3.4, respectively) by a strong acid of a pK_a ca. <2[25, 26] to yield HSO$_4^-$ as a good leaving group (the

pK_a of H$_2$SO$_4$ is -9 to -3 and the pK_a of HSO$_4^-$ is 1.9 to 2.7). On the other hand, basic conditions generate SO$_4^{2-}$, which - being the anion of a weak acid - is a very poor leaving group. Although base-catalyzed hydrolysis proceeds under inversion of configuration as expected, the reaction rates are exceedingly low and alkaline hydrolysis is not feasible for practical purposes[27, 28]. The same holds for O-nucleophiles with enhanced nucleophilicity,

such as acetate or methoxide[29]. Bearing the difficulties of chemical (non-enzymatic) sulfate monoester hydrolysis in mind, we have addressed the issue of protonation of the sulfate leaving group based on the active site architecture revealed by X-ray crystallographic analysis of Pisa1. As shown in Figure 5, SdsA1 and Pisa1 share an active site histidine (His317 in Pisa1) as a potential source for the protonation of the leaving group. In addition Tyr417 and His405 in Pisa 1 and SdsA1, respectively, may act as an active site acid to fulfill the envisaged protonation. To investigate the potential role of these two amino acids we have generated amino acid replacements by active site directed mutagenesis: His317Ala, as well as Tyr417Phe, Tyr417His and Tyr417Asp.

The characterization of the purified variant proteins revealed that enzymatic hydrolysis of the model substrate is three orders of magnitude slower for the His317Ala variant compared to the wild-type enzyme. This variant showed only a marginal effect on K_M indicating that His317 does not play a major role in substrate binding. Therefore, we suggest that the imidazole side chain of His317 participates in the protonation of the sulfate group, which is the crucial step in preparing the substrate for hydrolysis.

In contrast to the His317Ala variant, the Tyr417Phe and Tyr417His variants showed considerably higher enzymatic activities (5- and 10-times, respectively). On the other hand, K_M values were strongly affected. In fact, substrate saturation was not achieved rendering K_M values inaccessible in our experimental set-

up. In the case of the Tyr417Asp variant k_{cat} could not be determined indicating that a negatively charged amino acid residue is prohibitive for substrate binding and enzymatic hydrolysis. These data support a role in substrate binding and positioning rather than protonation of the substrate's leaving group.

In addition to addressing issues revolving around the enzymatic reaction mechanism, we have also generated a small set of single and double variants (see Table 2) to investigate the high stereopreference towards the *(R)* configuration of 2-octyl sulfate. In this set of variants, the Ser233Tyr/Phe250Gly double variant accepted (*S*)-2-octyl sulfate as a substrate although the catalytic efficiency was 20-fold lower as with (*R*)-2-octyl sulfate. In comparison with wild type enzyme the catalytic efficiency was even further reduced by a factor of 630 (see Table 2). This indicates that the amino acid replacements have led to an alternative lipophilic binding channel enabling productive binding of the *S* configured substrate (see Figure 6). These results demonstrate that the amino acids lining the hydrophobic substrate binding pocket govern the stereopreference of the reaction and strongly affect the catalytic efficiency of substrate hydrolysis. Further studies in our laboratories will focus on the generation of variants that show inverted stereopreference (*i.e. S* specificity) and display altered chain-length specificity in order to improve the accessible substrate range of this enzyme family for biocatalytic applications [18].

Materials and Methods

Expression and purification of recombinant protein in E. coli *host cells*

The expression and purification of SdsA1, Pisa1 and variants thereof was performed as described previously [6, 18].

Activity assays for Pisa1, its variants and SdsA1

Enzymatic activity towards the hydrolysis of 1- and 2-octyl sulfate, respectively, was analyzed by chiral GC measurements as described in a previous study[6, 18, 19].

Crystallization, data collection and structure elucidation of Pisa1

Pisa1 and Ser233Tyr/Phe250Gly double variant crystals were grown at 20 °C by sitting-drop vapor-diffusion in silica hydrogel with equal volume (5 μl) of protein (10 mg/ml) and reservoir solution (acetate buffer pH 4.6, and either 30% PEG 2000 or PEG 4000). Fragile leaflet shaped orthorhombic crystals with six molecules in the asymmetric unit were obtained using PEG 2000 (wt form-I) whereas PEG 4000 produced rodlike trigonal crystals hosting a dimer in the asymmetric unit (wt form-II). These crystals diffracted to 3.0 Å (double variant Ser233Tyr/Phe250Gly) and 2.7 Å (wt form-I and II) resolution, respectively, and were used for structure determination.

Data collection for both crystal forms obtained for wt Pisa1 were performed at the European Molecular Biology Laboratory in Hamburg, Germany. The experiments were carried out at 100 K using crystals flash frozen in a mixture of 20% glycerol and mother liquor using liquid nitrogen. A single wavelength dataset from wt form-I crystals was collected at beam line X11. A MAD experiment was performed for wt form-II crystals employing the four Zn^{2+} atoms as anomalous scatter (beamline X12). Both beam lines were equipped with MAR CCD detectors. Data from the double variant crystals were collected at the ESRF Grenoble at beam line ID14-4. Indexing and integration of the collected data were done with XDS[30], scaling with SCALA[31].

Phasing of Pisa1 wt form-II dataset was achieved through multi wavelength anomalous dispersion techniques, solve [32] found three of the four Zn^{2+} ions, a density modification using resolve[32] gave an interpretable density for most parts of the structure. First trace was built using buccaneer[33] and coot[34]. It turned out that both molecules related by a non-crystallographic dyad are partly disordered.

The structures of wt form-I as well as double variant were solved by molecular replacement using program phaser of the ccp4 package[35] and a starting model built with the program modeller [36] based on the SdsA1 coordinates. Whereas the wt form-I structure was well defined in most parts, the structure of the double variant was highly disordered. Both, wt form-II as well as the double variant structures were included in this report only because they show features that are not evident in the wt form-I structure. Table 1 gives an overview of data collection and refinement.

Atomic coordinates and structure factors were deposited in the PDBe under accession codes 2yhe (wt form-I), 4av7 (double mutant) and 4axh (wt form-II).

Docking studies

Molecules A from the Pisa1 wt from-I and from the double variant were subjects of docking experiments using the program GLIDE [37-40]. *(R)*-2-Octyl sulfate and *(S)*-2-octyl sulfate were used as possible substrates. All three structures were prepared using the protein preparation wizard. The substrate was built inside Maestro and both configurations (*R* and *S*) were calculated with ligprep. The position of the ligand's sulfate group was restrained to the position of the sulfate ion in the crystal structures.

Site-directed mutagenesis

The Pisa1 variants were prepared according to the protocol described in the QuikChange® XL-site directed mutagenesis kit (Stratagene) with a slight modification. A two-step PCR reaction was carried out: in the first step, separate reactions for forward und reverse primer were performed for 4 cycles, in the second step, equal amounts of the forward and the reverse primer were mixed for another 18 cycles. The primers used for the insertion of the mutations are summarized in Table S1 in the supplementary information available for this manuscript.

Isothermal titration calorimetry (ITC)

Kinetic parameters for the hydrolysis of *(R)*-2-octyl sulfate, *(S)*-2-octyl sulfate and 1-octyl sulfate were determined using a VP-ITC system (MicroCal) according to the instruction of the manufacturer. All experiments were performed at 25 °C in 100 mM Tris/HCl, pH 8.2 buffer and solutions were degassed immediately before measurements were taken. The enthalpy of the reaction was determined by the single injection method: The substrate *(R)*-2-octyl sulfate was added (1.5 mM, 3 x 20 µl, duration time 39.9 s, spacing time 1,000 s) to Pisa1 (125 nM) and 1-octyl sulfate (150 µM, 3 x 20 µl, duration time 39.9 s, spacing time 600 s) to SdsA1 (125 nM), respectively. For the kinetic parameters the multiple injection method was used for the titration of Pisa1, its variants and SdsA1 using *(R)*-2-octyl sulfate, *(S)*-2-octyl sulfate and 1-octyl sulfate (the enzyme and the substrate concentration was varied from 8 nM to 200 µM or from 800 µM to 55 mM, depending on the turnover numbers and K_M values of the proteins, respectively). Nonlinear least square fitting with Origin version 7.0 (MicroCal) was used for ITC data analysis to obtain k_{cat}, K_M and K_I.

REFERENCES

1. Bojarova, P., and Williams, S. J. (2008) Sulfotransferases, sulfatases and formylglycine-generating enzymes: a sulfation fascination, *Current opinion in chemical biology* **12**, 573-581.

2. Gadler, P., and Faber, K. (2007) New enzymes for biotransformations: microbial alkyl sulfatases displaying stereo- and enantioselectivity, *Trends in biotechnology* **25**, 83-88.

3. Kertesz, M. A. (2000) Riding the sulfur cycle--metabolism of sulfonates and sulfate esters in gram-negative bacteria, *FEMS microbiology reviews* **24**, 135-175.

4. Hanson, S. R., Best, M. D., and Wong, C. H. (2004) Sulfatases: structure, mechanism, biological activity, inhibition, and synthetic utility, *Angewandte Chemie (International ed* **43**, 5736-5763.

5. Bornscheuer, U., and Kazlauskas, R. (2005) Hydrolases in organic synthesis, Wiley-VCH, Weinheim, Germany.

6. Hagelueken, G., Adams, T. M., Wiehlmann, L., Widow, U., Kolmar, H., Tummler, B., Heinz, D. W., and Schubert, W. D. (2006) The crystal structure of SdsA1, an alkylsulfatase from *Pseudomonas aeruginosa*, defines a third class of sulfatases, *Proceedings of the National Academy of Sciences of the United States of America* **103**, 7631-7636.

7. Boltes, I., Czapinska, H., Kahnert, A., von Bulow, R., Dierks, T., Schmidt, B., von Figura, K., Kertesz, M. A., and Uson, I. (2001) 1.3 A structure of arylsulfatase from *Pseudomonas aeruginosa* establishes the catalytic mechanism of sulfate ester cleavage in the sulfatase family, *Structure* **9**, 483-491.

8. Muller, I., Kahnert, A., Pape, T., Sheldrick, G. M., Meyer-Klaucke, W., Dierks, T., Kertesz, M., and Uson, I. (2004) Crystal structure of the alkylsulfatase AtsK: insights into the catalytic mechanism of the Fe(II) alpha-ketoglutarate-dependent dioxygenase superfamily, *Biochemistry* **43**, 3075-3088.

9. Williams, O. B., and Rees, H. B., Jr. (1949) Bacterial utilization of anionic surface-active agents, *Journal of bacteriology* **58**, 823.

10. Dodgson, K. S., White, G. F., and Fitzgerald, J. (1982) Sulfatases of microbial origin, CRC Press, Boca Raton.

11. Bartholomew, B., Dodgson, K. S., and Gorham, S. D. (1978) Purification and properties of the S1 secondary alkylsulphohydrolase of the detergent-degrading micro-organism, *Pseudomonas* C12B, *The Biochemical journal* **169**, 659-667.

12. Bartholomew, B., Dodgson, K. S., Matcham, G. W., Shaw, D. J., and White, G. F. (1977) A novel mechanism of enzymic ester hydrolysis. Inversion of configuration and carbon-oxygen bond cleavage by secondary alkylsulphohydrolases from detergent-degrading micro-organisms, *The Biochemical journal* **165**, 575-580.

13. Shaw, D. J., Dodgson, K. S., and White, G. F. (1980) Substrate specificity and other properties of the inducible S3 secondary alkylsulphohydrolase purified from the detergent-degrading bacterium Pseudomonas C12B, *The Biochemical journal* **187**, 181-190.

14. Pogorevc, M., and Faber, K. (2002) Enantioselective stereoinversion of sec-aklyl sulfates by an alkylsulfatase from *Rhodococcus ruber* DSM44541, *Tetrahedron:Asymmetry* **13**, 1435-1441.

15. Pogorevc, M., Kroutil, W., Wallner, S. R., and Faber, K. (2002) Enantioselective stereoinversion in the kinetic resolution of rac-sec-alkyl sulfate esters by hydrolysis with an alkylsulfatase from *Rhodococcus ruber* DSM 44541 furnishes homochiral products, *Angewandte Chemie (International ed* **41**, 4052-4054.

16. Gadler, P., and Faber, K. (2007) Highly enantioselective biohydrolysis of sec-alkyl sulfate esters with inversion of configuration catalysed by Pseudomonas spp., *Eur. J. Org. Chem.*, 5527-5530.

17. Oltmanns, R. H., Muller, R., Otto, M. K., and Lingens, F. (1989) Evidence for a new pathway in the bacterial degradation of 4-fluorobenzoate, *Applied and environmental microbiology* **55**, 2499-2504.

18. Schober, M., Gadler, P., Knaus, T., Kayer, H., Birner-Grunberger, R., Gully, C., Macheroux, P., Wagner, U., and Faber, K. (2011) A stereoselective inverting *sec*-alkylsulfatase for the deracemization of sec-alcohols, *Org Lett* **13**, 4296-4299.

19. Schober, M., Knaus, T., Toesch, M., Macheroux, P., Wagner, U., and Faber, K. (2012) The substrate spectrum of the inverting *sec*-alkylsulfatase Pisa1, *Adv. Synth. Catal.* **354**, 1737-1742.

20. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (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)**, 110-119.

21. Kleywegt, G. J. (1996) Use of non-crystallographic symmetry in protein structure refinement, *Acta crystallographica* **52**, 842-857.

22. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis, *J Comput Chem* **25**, 1605-1612.

23. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions, *Protein Engineering* **8**, 127-134.

24. Scherer, H. W. (2001) Sulphur in crop production, *Eur. J. Agron.* **14**, 81-111.

25. Fulford, J. E., Dupuis, J. W., and March, R. E. (1978) Gas phase ion-molecule reactions of dimethyl sulfoxide, *Can. J. Chem.* **56**, 2324-2330.

26. Klages, F., Jung, H. A., and Hegenberg, P. (1966) Determination of acidity by means of aliphatic diazo compounds. III. Dynamic acidity of very strong acids, *Chem. Ber.* **9**, 1704-1711.

27. Batts, B. D. (1966) Alkyl hydrogen sulfates. II. The hydrolysis in aqueous acid solution, *J. Chem. Soc. (B)*, 551-555.

28. Burwell, R. J. J. (1952) The hydrolysis of optically active *secondary* butyl hydrogen sulfate, *J. Am. Chem. Soc.* **74**, 1462-1466.

29. Wallner, S. R., Nestl, B., and Faber, K. (2005) Stereoselective hydrolysis of *sec*-monoalkyl sulfate esters with retention of configuration, *Tetrahedron* **61**, 1517-1521.

30. Kabsch, W. (2010) Integration, scaling, space-group assignment and post-refinement, *Acta crystallographica* **66**, 133-144.

31. Evans, P. (2006) Scaling and assessment of data quality, *Acta crystallographica* **62**, 72-82.

32. Terwilliger, T. C., and Berendzen, J. (1999) Automated MAD and MIR structure solution, *Acta crystallographica* **55**, 849-861.

33. Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains, *Acta crystallographica* **62**, 1002-1011.

34. Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics, *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126-2132.

35. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *J. Appl. Cryst.* **40**, 658-674.

36. Eswar, N., Eramian, D., Webb, B., Shen, M. Y., and Sali, A. (2008) Protein structure modeling with MODELLER, *Methods in molecular biology (Clifton, N.J* **426**, 145-159.

37. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., and Shenkin, P. S. (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, *Journal of medicinal chemistry* **47**, 1739-1749.

38. Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C., and Mainz, D. T. (2006) Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes, *Journal of medicinal chemistry* **49**, 6177-6196.

39. Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., and Banks, J. L. (2004) Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening, *Journal of medicinal chemistry* **47**, 1750-1759.

40. Park, M. S., Gao, C., and Stern, H. A. (2011) Estimating binding affinities by docking/scoring methods using variable protonation states, *Proteins* **79**, 304-314.

Chapter 9

Supplementary Information
Structure and mechanism of an inverting alkylsulfatase from *Pseudomonas* sp. DSM6611 specific for secondary alkyl sulfates

Supplemental Data

Structure and mechanism of the first inverting alkylsulfatase specific for secondary alkylsulfates

Tanja Knaus, Markus Schober, Bernhard Kepplinger, Martin Faccinelli, Julia Pitzer, Kurt Faber, Peter Macheroux and Ulrike Wagner

Table of Contents

Supplementary Table 1

Primer used for the insertion of the mutations in Pisa1.

Variant		Sequence (5´→ 3´)
Ser233Tyr	for	GTCGCGGCCAGTACCAGTATGGCGTCATGGTGCCG
	rev	CGGCACCATGACGCCATACTGGTACTGGCCGCGAC
Ser233Phe	for	CGCGGCCAGTACCAGTTTGGCGTCATGGTGCC
	rev	GGCACCATGACGCCAAACTGGTACTGGCCGCG
Phe250Gly	for	GTCGACAGCGGCCTGGGCAAGACCACGGC
	rev	GCCGTGGTCTTGCCCAGGCCGCTGTCGAC
Phe250Ser	for	GTCGACAGCGGCCTGAGCAAGACCACGGCG
	rev	CGCCGTGGTCTTGCTCAGGCCGCTGTCGAC
His317Ala	for	GCCGCCGGCGATGGCCAACCTGCTCACC
	rev	GGTGAGCAGGTTGGCCATCGCCGGCGGC
Tyr417His	for	GCGCAGCTACCACGGCGCGCTGTCG
	rev	CGACAGCGCGCCGTGGTAGCTGCGC
Tyr417Phe	for	CTGCGCAGCTACTTCGGCGCGCTGTC
	rev	GACAGCGCGCCGAAGTAGCTGCGCAG
Tyr417Asp	for	CCTGCGCAGCTACGATGGCGCGCTGTCGAC
	rev	GTCGACAGCGCGCCATCGTAGCTGCGCAGG

Supplementary Figure 1

Sequence alignment

The initial alignment was performed using the program BLAST [1] and refined by manual adjustments. Amino acid residues coordinating to the two zinc centers in the active sites are highlighted in red (invariant) and pink (conservative exchange). Residues involved in substrate recognition and positioning are shown in yellow (invariant) and green (not conserved).

```
          1                                              48
Pisa1      MSRFIRASQRRTLLATLIAATLAQPLLAA--ESLDSKPASAITAAKNAEV
SdsA1      MSR--------LLALLALA----PLLAGAAETTAPKPPSAFTVEAQRRV
Consensus  MSR        LLA+L++A    PL A+   E+   KP SA+T+  +  V

          49                                             97
Pisa1      LKNLPFADREEFEAAKRGLIA-PFSGQIKNAEGQVVWDMGAYQFLNDKDA
SdsA1      EAELPFADRADFERADRGLIRRPERLLIRNPDGSVAWQLGGYDFLLDGKP
Consensus     LPFADR +FE A RGLI P I+N +G V W +G Y FL D

          98                                             147
Pisa1      ADTVNPSLWHQAQLNNIAGLFEVMPKLYQVRGLDPANMTIIEGDSGLVLI
SdsA1      RDSINPSLQRQALLNLKYGLFEVAEGIYQVRGFDLANITFIRGDSGWIVV
Consensus   D++NPSL QA LN  GLFEV  +YQVRG D AN+T I GDSG +++

          148                                            197
Pisa1      DTLTTAETARAALDLYFQHRPKKPIVAVVYSHSHIDHFGGARGIIDEADV
SdsA1      DTLTTPATARAAVELVSRELGERPIRTVIYSHAHADHFGGVRGLVEPQQV
Consensus   DTLTT  TARAA++L     ++PI  V+YSH+H DHFGG RG+++   V
```

```
            198                                                    247
Pisa1      KAGKVKVFAPSGFMEHAVSENILAGTAMARRGQYQSGVMVPRGAQAQVDS
SdsA1      ASGAVQIIAPAGFMEAAIKENVLAGNAMMRRATYQYGTQLPKGPQGQVDM
Consensus   +G V++ AP+GFME A+ EN+LAG AM RR   YQ G  +P+G Q QVD-

            248                                                    297
Pisa1      GLFKTTATNATNTLVAPNVLIEKPYERHTVDGVELEFQLTLGSEAPSDMN
SdsA1      AIGKGLARGPLS-LLAPTRLIEGEGEDLVLDGVPFTFQNTPGTESPAEMN
Consensus   + K  A       L+AP LIE   E  +DGV   FQ T G+E+P++MN

            297                                                    347
Pisa1      IYLPQFKVLNTADNAPPAMHNLLTPRGAEVRDAKAWAGYIDASLEKYGDR
SdsA1      IWLPRQKALLMAENVVGTLHNLYTLRGAEVRDALGWSKYINQALHRFGRQ
Consensus  I+LP+ K L  A+N   +HNL T RGAEVRDA  W+ YI+ +L ++G +

            348                                                    397
Pisa1      TDVLIQQHNWPVWGGDKVRTYLADQRDMYAFLNNRALNLMNKGLTLHEIA
SdsA1      AEVMFAVHNWPRWGNAEIVEVLEKQRDLYGYLHDQTLHLANQGVTIGQVH
Consensus   +V+  HNWP WG  ++    L  QRD+Y +L+++ L+L N+G+T+ ++

            398                                                    447
Pisa1      AEVSKLPGELDRKWYLRSYYGALSTNLRAVYQRYLGFYDGNPANLDPFPP
SdsA1      NRL-RLPPSLDQEWYDRGYHGSVSHNARAVLNRYLGYYDGNPATLDPLSP
Consensus    + +LP  LD++WY R Y+G++S N RAV  RYLG+YDGNPA LDP  P

            448                                                    497
Pisa1      VEAGKRYVEAMGGADAVLKQMRAAIDKGDYRWAVQLGNHLVFADPANKDA
SdsA1      EDSAGRYVEYMGGAERLLEQARASYARGEYRWVVEVVNRLVFAEPDNRAA
Consensus   ++  RYVE MGGA+ +L+Q RA+  +G+YRW V++ N LVFA+P N+ A

            498                                                    544
Pisa1      RALQADAMEQLGYQTENALWRNMYMTGAMELRHGVPTYD---SRGKSEMG
SdsA1      RELQADALEQLGYQAENAGWRNSYLSAAYELRHGVPRDQPTMKAGSADAL
Consensus  R LQADA+EQLGYQ ENA WRN Y++ A ELRHGVP           G ++

            545                                                    594
Pisa1      RALTPDMFFDLLAIRLDTDKAVGHDMTLNWVFEDLKQDIALTLRNGVLTQ
SdsA1      AAMDTGLLFDYLGVRLDAGAAEGKALSINLRLPDIGENYLLELKNSHLNN
Consensus   A+   + FD L +RLD  A G  +++N    D+ ++  L L+N  L

            595                                                    644
Pisa1      RVGSLNPKADVTVKLTKPTLDQIAARKLDLPTAIKQGTVKLDGDGKKLGE
SdsA1      LRGVQSEDAGQTVSIDRADLNRLLLKEVSAVRLVFEGKLKSSGNPLLLGQ
Consensus     G  +  A  TV + +  L+++  +++      + +G +K  G+   LG+

            645              661
Pisa1      FFGLLDSFSPKFNIVEP
SdsA1      LFGMLGDFDFWFDIVTP
Consensus  FG+L--F---F+IV-P
```

References

1. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool, *J. Mol. Biol.* **215**, 403-410.

Chapter 10

Flavins and Flavoproteins

10.1   Riboflavin and Flavocoenzymes FMN and FAD

Riboflavin, also named vitamin B_2, is built up by a tricyclic isoalloxazine moiety which is covalently linked at its N(10) position with a ribityl side chain as shown in (Figure 1).[1, 2] It is a precursor for the coenzymes FMN (flavin mononucleotide) and FAD (flavin adenine dinucleotide). FMN is additionally phosphorylated at the terminal hydroxyl group of the ribityl side chain and in FAD adenosine monophosphate (AMP) is linked to FMN by a phosphate group. Although riboflavin is produced by all plants and most microorganisms, animals and vertebrates lack the enzymes to synthesize this vitamin. However, they are capable to generate FMN and FAD from the precursor.
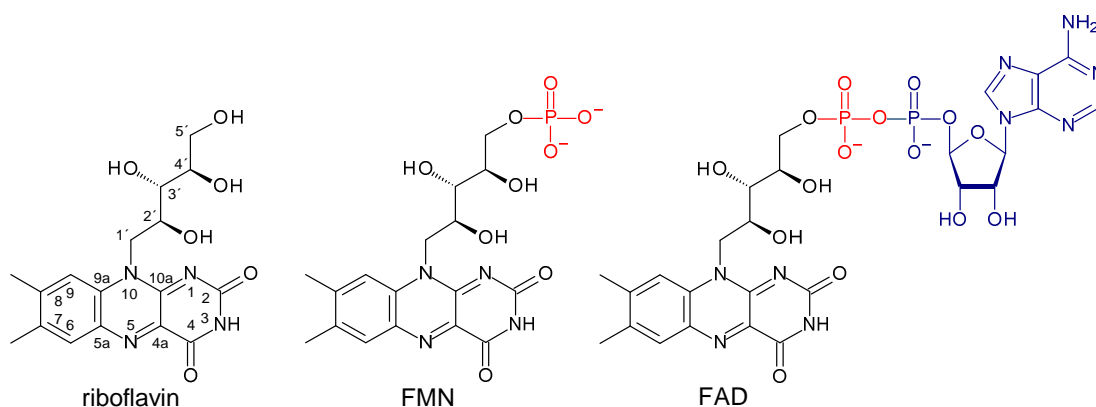
Figure 1. Chemical structures of riboflavin, FMN and FAD and numbering scheme for the isoalloxazine moiety.

10.1.1 Biosynthesis of Riboflavin, FMN and FAD

The biosynthesis of riboflavin and the flavocoenzymes FMN and FAD is briefly summarized in Figure 2.[3] The initial step is the conversion of GTP (1) to the first intermediate 2-5-diamino-6-ribosylamino-4(3H)-pyrimidinone-5´-phosphate (2) by hydrolysis of pyrophosphate and formal cleavage of C-8 of the imidazole ring. Both reactions are catalyzed by GTP cyclohydrolase II (reaction I). This compound is further transformed into 5-amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedione-5´-phosphate (5) via structurally different precursors in different organisms in two steps. In Archaea and in fungi[4, 5] step one is a NADPH dependent reduction process (reaction II), where the ribosyl side chain is transformed into the ribityl side chain (3). Step two performs a deamination

of the amino group at position 2 of the pyrimidine ring (reaction III) to yield compound 5. In plants and eubacteria,[6-9] these reaction steps occur in inverse order (reaction IV and V) via the ribosylaminopyridine intermediate 3. In a further step, the phosphate residue of intermediate 5 is released (reaction VI) to generate 5-amino-6-ribitylamino-2,4(1*H*,3*H*)-pyrimidinedione (6). However, the mechanism of dephosphorylation is still unknown.

The enzyme 3,4-dihydroxy-2-butanone-4-phosphate synthase catalyzes a complex rearrangement/elimination sequence (reaction VII) to convert ribulose-5-phosphate (7) to 3,4-dihydroxy-2-butanone-4-phosphate (8). This product is then condensed by lumazine synthase (reaction VIII) with the previously obtained dephosphorylation intermediate (6) to give 6,7-dimethyl-8-ribityllumazine (9).
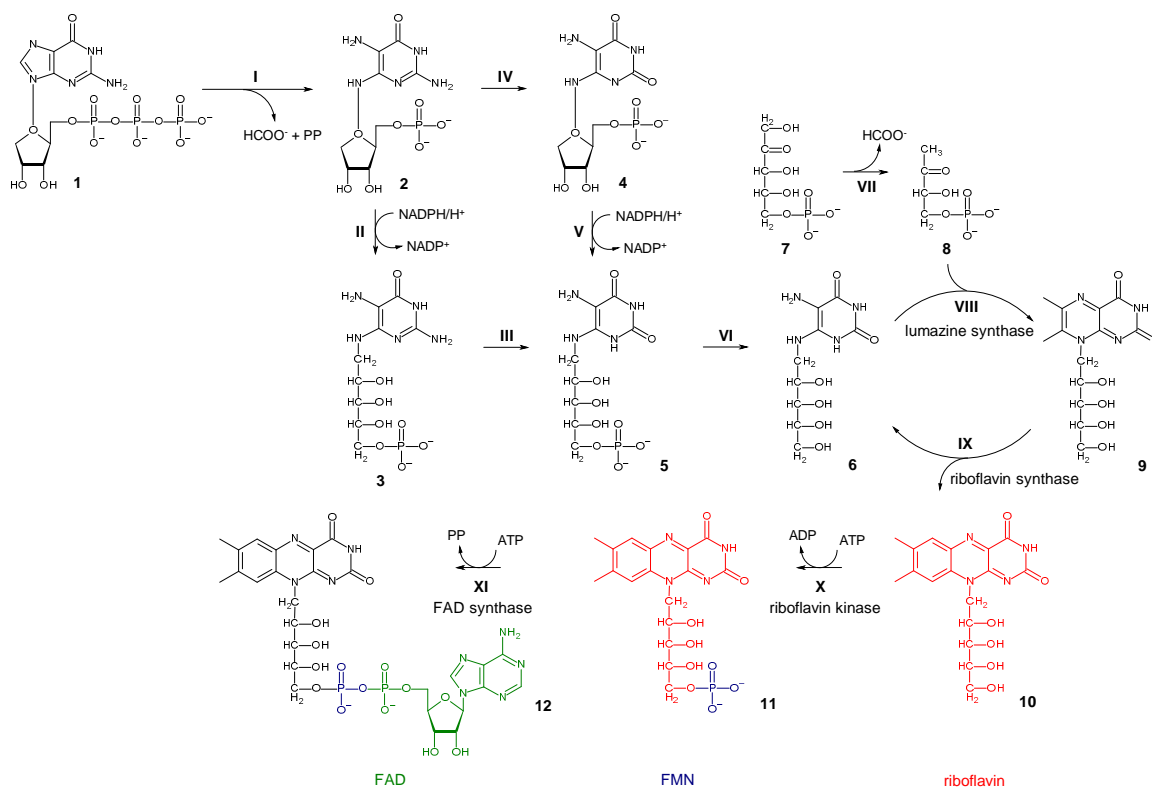
Figure 2. Biosynthesis of riboflavin, FMN and FAD (adapted from reference 3).

The final step in the biosynthesis of riboflavin (10) is catalyzed by riboflavin synthase (reaction IX) in an unusual dismutation of two molecules 6,7-dimethyl-8-ribityllumazine (9). Thereby an exchange of a four-carbon unit occurs, to obtain the tricyclic isoalloxazine moiety of riboflavin (10) and 5-amino-6-ribitylamino-2,4(1*H*,3*H*)-pyrimidinedione (6) which can again be used as a substrate for lumazine synthase (reaction VIII). Lumazine synthase and riboflavin synthase generate thereby 1 mol of riboflavin from 1 mol of GTP

and 2 mol of ribulose-5-phosphate.[3] It is worth to mention that the complex reaction catalyzed by riboflavin synthase (reaction IX) can also proceed nonenzymatically just in boiling aqueous solution at neutral or acidic pH.[10-12]

Finally, riboflavin is phosphorylated at the 5′-OH group of the ribityl side chain by the enzyme riboflavin kinase (reaction X) to yield FMN (11), which can further be converted by FAD synthase to FAD (12). This last step is obtained by a transfer of an adenylyl moiety from ATP (reaction XI). Reaction X and XI exist in all organisms, independently if they take up their required riboflavin content from nutritional sources, or they synthesize it autonomously.[3]

10.2   Other natural occurring Flavin derivatives

Besides the most common known flavins riboflavin, FMN and FAD, other flavin derivatives are found in nature, which will be briefly discussed in this chapter.

Lumichrome and Lumiflavin. Lumichrome and lumiflavin (and derivatives) have been identified to be products of the photolysis of riboflavin, FMN and FAD in the 1960s.[13]
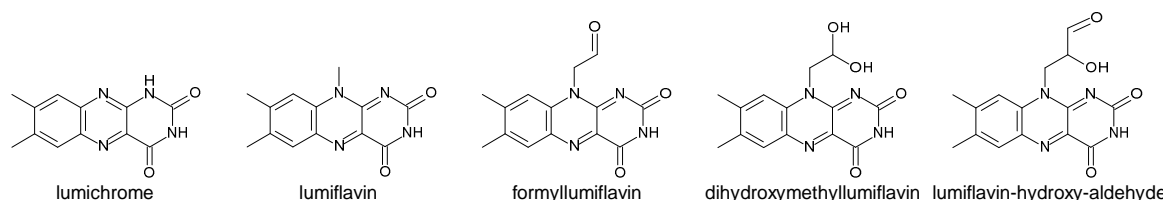
Figure 3. Chemical structures of lumichrome, lumiflavin and lumiflavine derivatives.

It has been shown that these degradation products (for their chemical structures see Figure 3) may be involved in various biological processes. For example lumichrome may inhibit flavin reductase in *E. coli* cells, thereby enhancing the antiproliferative effect of hydroxyurea, a cell specific ribonucleotide reductase inactivator.[14] Another example of their biological activity is that lumichrome and lumiflavin may inhibit the uptake process of riboflavin in liver cells by human-derived Hep G2 cells.[15]

5-Deazaflavins. 5-Deazaflavins are lacking of the nitrogen in the position 5 of the isoalloxazine moiety in flavin derivatives. The structures of coenzymes F_0 and F_{420} are shown in Figure 4. Although coenzyme F_{420} is structural related to riboflavin, its chemical

properties are more similar to that of nicotinamides than to that of FMN/FAD due to the exclusively simultaneous transfer of two electrons.[16] It is a signature molecule for methanogenic bacteria because it is involved in the hydride transfer during the reductive transformation of carbon dioxide and acetate into methane.[17, 18]
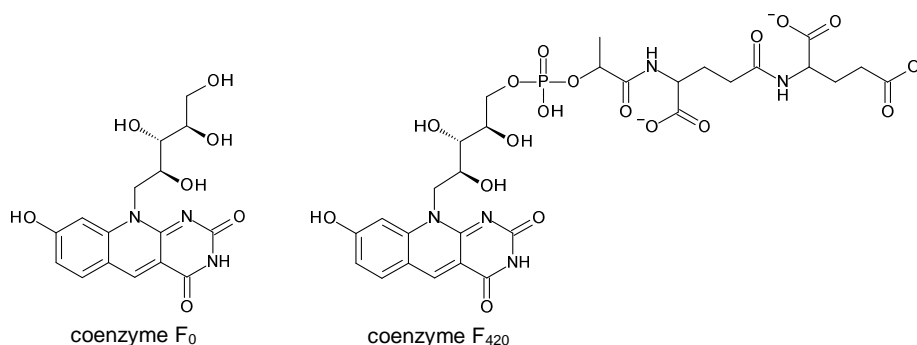
Figure 4. Chemical structures of coenzymes F_0 and F_{420}.

Moreover coenzyme F_{420} is found in *Streptomyces lincolnensis* where it is involved in the biosynthesis of the antibiotic lincomycin[19] and it has been shown to be required for the activation of drugs for the treatment of tuberculosis caused by *Mycobacterium tuberculosis*.[20]

Roseoflavin. The pink to red colored flavin derivative roseoflavin [7-methyl-8-dimethylamino-(1´-D-ribityl)isoalloxazine] is produced by the soil bacterium *Streptomyces davawensis* and has an antibiotic effect on gram-positive (e.g. *Bacillus subtilis* or *Staphylococcus aureus*) as well as gram-negative microorganisms. Up to date it is the only known naturally occurring flavin derivative with antibiotic activity.[21, 22]
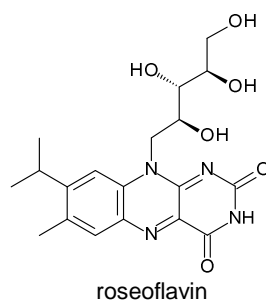
Figure 5. Chemical structure of the antibiotic roseoflavin.

10.3 Flavoproteins

Up to the present more than 370 flavin dependent proteins have been identified. Approximately three-fourth of these proteins utilize FAD as cofactor and only one-fourth are FMN dependent. It should be mentioned that proteins show the ability to bind either FMN or FAD in a noncovalently (about 90%) or covalently (about 10%) manner.[23,24] Although riboflavin is not known to be used as a cofactor in enzymes (except in one case[25]), it seems to be the preferred storage form for the biological active derivatives, for example in chicken eggs and archaeal dodecins.[26, 27]

Due to the chemically reactive heterocyclic isoalloxazine moiety of flavins it is obvious that flavin dependent enzymes are involved in a broad spectrum of biological processes. The heterocyclic three ring system can exist in three different redox states with significantly different chemical properties as shown in Figure 6: in (i) the fully oxidized (quinone) state, (ii) the one-electron reduced or semiquinone state and in (iii) its fully reduced (two-electron reduced) state, largely dependent on the protein environment and solvent used.[28]
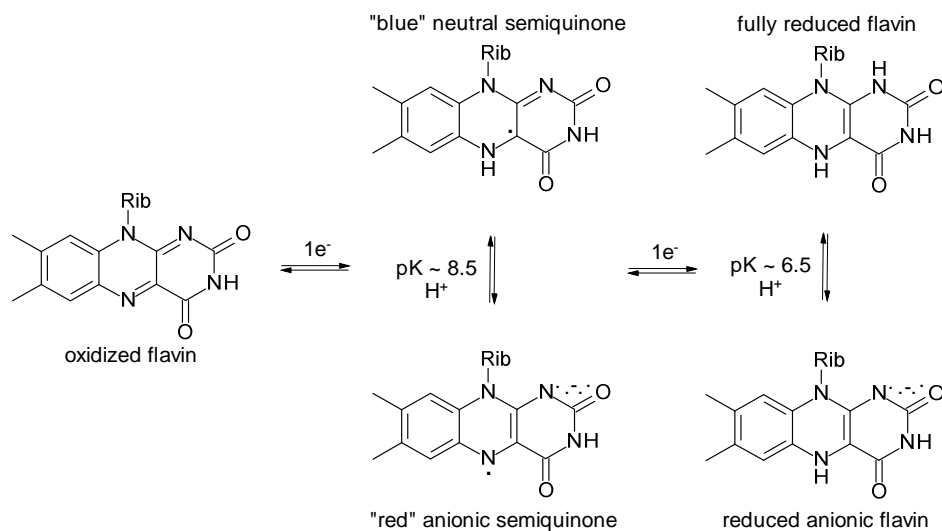
Figure 6. Redox and ionization states of flavin.

A recent study on flavin dependent enzymes has shown that 276 of fully classified flavoenzymes can be divided into oxidoreductases (91%), transferases (4.3%), lyases (2.9%), isomerases (1.4%) and ligases (0.4%).[24] Due to this versatile application of flavoproteins they are part of intensive studies by many enzymologists and are the subject of a various number of review articles.[29-33]

10.4   Storage of Flavins

Dodecin from *Halobacterium salinarum*. In 2003 the structure of a novel flavin binding protein from the halophilic marine gram-negative obligate aerobic archaeon *Halobacterium salinarum* could be solved (PDB code: 1MOG).[34] It was characterized as the smallest flavoprotein until now with only 68 amino acids. Crystals of this protein showed that monomers assembled to dodecamers, whereby four trimers coassembled with 6 riboflavin dimers resulting in a molecular weight of 89.3 kDa (Figure 7A). Due to its ability to form dodecamers with bound flavin cofactors, the protein was named dodecin (dodecamer and flavin). The isoalloxazine moieties of two riboflavins are sandwiched between the indole groups of two tryptophan residues and contact each other via their *re*-faces. Thereby the ribityl side chains of the riboflavin dimer are directed towards the outside of the protein (Figure 7B).
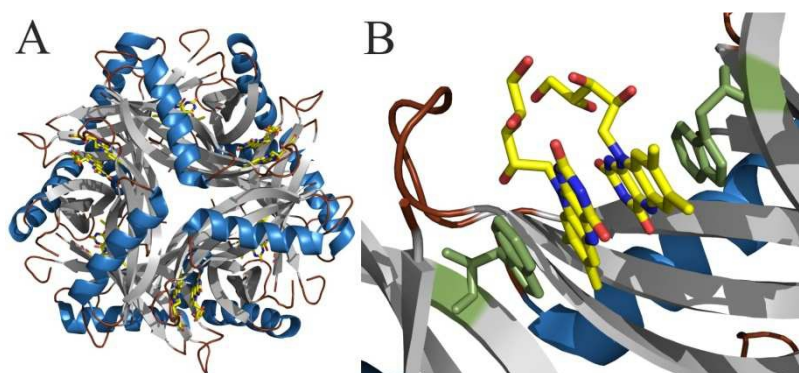
Figure 7. Structure representation of dodecin from *Halobacterium salinarum* (PDB code:1MOG). (A) Cartoon representation, showing the bound riboflavins in yellow sticks. (B) Close up of the flavin binding site with the tryptophan residues depicted in green sticks. Pictures were prepared with the program PyMol.[35]

Further studies on that protein showed that the dodameric assembly is not only able to bind riboflavin, but also lumichrome, lumiflavin and FMN in a 1:1 stoichiometry whereby the smallest K_D value could be obtained for lumichrome (9.88 nM) indicating that this flavin derivative is a ligand of dodecin *in vivo*. The dissociation constants for lumiflavin and riboflavin were calculated to be 17.57 nM and 35.76 nM, respectively. In the case of FAD a dodecin/ligand ratio of 2:1 was observed, indicating for a different binding mode. The authors suggested thereof, that dodecin from *H. salinarum* may prevent riboflavin from a possible degradation and moreover may act as a rescue mechanism to bind

lumichrome if a degradation of riboflavin occurs. Furthermore it may act as a "waste-collecting protein" to protect the cellular environment from possibly high amounts of the photo-toxic product lumichrome.[36] Recent studies on this protein showed that archaeal dodecin acts as a riboflavin binding protein that is of high functional versatility in neutralizing riboflavin to protect the cellular environment from uncontrolled flavin reactivities. Riboflavin is released into the biosynthesis of FMN and FAD when necessary for metabolic activities. Moreover it prevents light-induced riboflavin relativities and degradation of riboflavin by deactivation of the reactive excited state of riboflavin.[27, 37]

Dodecin from *Thermus thermophilus*. The structure of a second dodecin was solved from the extremely thermophilic gram-negative eubacterium *Thermus thermophilus* by Meissner *et al*. (PDB code: 2V18).[38] Although the overall structure (Figure 8A) was similar to the dodecin from *H. salinarum*, binding of FMN dimers (instead of riboflavin) was found in *si-si*-orientation rather than *re-re*-orientation as can be seen in Figure 8B.
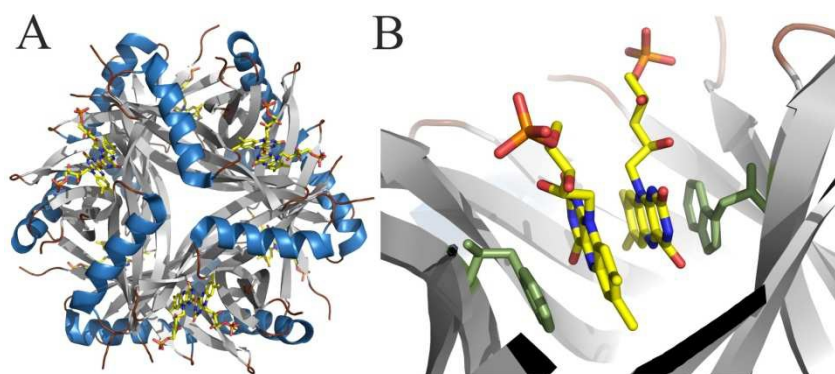
Figure 8. Structure representation of dodecin from *Thermus thermophilus* (PDB code: 2V18). (A) Cartoon representation, showing bound FMN in yellow sticks. (B) Close up of the flavin binding site with the tryptophan residues depicted in green sticks. Pictures were prepared with the program PyMol.[35]

Similar to the dodecin from *H. salinarum*, dodecin from *T. thermophilus* is able to bind riboflavin, FMN, FAD, lumichrome and lumiflavin with binding constants in the nmolar range. However, the affinity for bulkier flavins like FMN and FAD (K_D values were obtained to be 311 nM and 589 nM for *T. thermophilus* and 11300 nM and 440 nM for *H. salinarum*) was higher than for the small ligands lumichrome, lumiflavin and riboflavin (80 nM, 141 nM and 233 nM for *T. thermophilus* and 12, 18 and 36 nM for *H. salinarum*). Moreover the stoichiometry between dodecin and ligand was found to be 1:1 in all cases.

Additionally to the above mentioned flavin derivatives also coenzyme A serves as a ligand. These findings indicate that the functional role is different for bacterial and archaea dodecins. The authors suggest that dodecin from *T. thermophilus* may act as a bifunctional cofactor storage protein.[38]

10.5   A Flavoprotein in combination with Zinc

High throughput crystallization experiments by the Midwest Centre for Structural Genomics facility led to the structure of a protein from a gram-negative anaerobic microbe which resides in the human intestine, *Bacteroides thetaiotaomicron* (PDB code: 3CNE). Based on its overall topology, the protein was classified as a member of the DJ_1 superfamily which represents a heterogeneous family including chaperones and hydrolases.[39-42] Due to a possible protease activity, the protein was named ppBat (putative protease from *B. thetaiotaomicron*). ppBat occurs as a dimer with a flavin moiety bound between the dimeric interface as shown in Figure 9A.
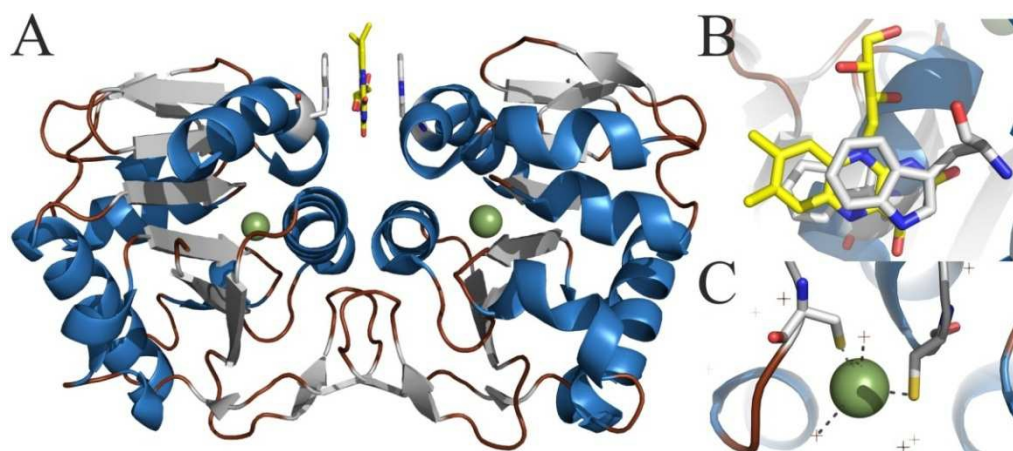
Figure 9. Structure representation of ppBat from *Bacteroides thetaiotaomicron* (PDB code: 3CNE). (A) Cartoon representation, showing the bound flavin in yellow sticks, the zinc ions in green and the tryptophan residues in grey sticks. (B) Close up of the flavin binding site with the tryptophan residues depicted in grey sticks. (C) Close up of the zinc binding site with the cysteine residues shown in grey sticks. Pictures were prepared with the program PyMol.[35]

Moreover each protomer features a metal-binding site which was identified to bind zinc in an approximately 16 Å distance away from the edge of the flavin. The flavin moiety is sandwiched between two tryptophan residues almost equidistant (Figure 9B). However, out of the structure, the flavin moiety cannot be identified to be riboflavin, FMN or FAD,

due to the missing 5th carbonyl group of the ribityl side chain. A closer look onto the zinc binding site in ppBat (Figure 9C) shows that the zinc ion is tetrahedral coordinated by two thiol groups of cysteine residues and two water molecules. Among the members of the DJ_1 superfamily, a hydrolase activity is associated with a cystidine, histidine or aspartic acid catalytic triad, and without any metal-ion.[40-42] Furthermore, such a complexation of zinc seems to be a novel as it has never been reported in other zinc dependent proteins so far. Moreover flavoproteins were not associated with hydrolase activity until to date.[24] Although nothing is currently known about the biological function of ppBat, the protein features an unusual combination of two cofactors, the flavin moiety and zinc ion.

10.6   References

1.   Kuhn, R., Reinemund, K., and Weygand, F. (1934) Synthese des Lumi-lactoflavins., *Ber. 67*, 1460-1462.

2.   Karrer, P., Schöpp, K., and Benz, F. (1935) Synthesen von Flavinen IV., *Helv. Chim. Acta. 18*, 426-429.

3.   Fischer, M., and Bacher, A. (2008) Biosynthesis of vitamin B2: Structure and mechanism of riboflavin synthase., *Arch Biochem Biophys. 474*, 252-265.

4.   Chatwell, L., Krojer, T., Fidler, A., Römisch, W., Eisenreich, W., Bacher, A., Huber, R., and Fischer, M. (2006) Biosynthesis of riboflavin: structure and properties of 2,5-diamino-6-ribosylamino-4(3H)-pyrimidinone 5'-phosphate reductase of *Methanocaldococcus jannaschii*., *J Mol Biol. 359*, 1334-1351.

5.   Graupner, M., Xu, H., and White, R. H. (2002) The pyrimidine nucleotide reductase step in riboflavin and F(420) biosynthesis in archaea proceeds by the eukaryotic route to riboflavin., *J Bacteriol. 184*, 1952-1957.

6.   Fischer, M., Römisch, W., Saller, S., Illarionov, B., Richter, G., Rohdich, F., Eisenreich, W., and Bacher, A. (2004) Evolution of vitamin B2 biosynthesis: structural and functional similarity between pyrimidine deaminases of eubacterial and plant origin., *J Biol Chem. 279*, 36299-36308.

7.   Richter, G., Fischer, M., Krieger, C., Eberhardt, S., Lüttgen, H., Gerstenschläger, I., and Bacher, A. (1997) Biosynthesis of riboflavin: characterization of the bifunctional deaminase-reductase of *Escherichia coli* and *Bacillus subtilis*., *J Bacteriol. 179*, 2022-2028.

8.   Chen, S. C., Chang, Y. C., Lin, C. H., Lin, C. H., and Liaw, S. H. (2006) Crystal structure of a bifunctional deaminase and reductase from *Bacillus subtilis* involved in riboflavin biosynthesis., *J Biol Chem. 281*, 7605-7613.

9.   Stenmark, P., Moche, M., Gurmu, D., and Nordlund, P. (2007) The crystal structure of the bifunctional deaminase/reductase RibD of the riboflavin biosynthetic pathway in *Escherichia coli*: implications for the reductive mechanism., *J Mol Biol. 373*, 48-64.

10.   Beach, R., and Plaut, G. W. (1969) The formation of riboflavin from 6,7-dimethyl-8-ribityllumazine in acid media., *Tetrahedron Lett. 40*, 3489-3492.

11.   Beach, R. L., and Plaut, G. W. (1970) Stereospecificity of the enzymatic synthesis of the o-xylene ring of riboflavin., *J Am Chem Soc. 92*, 2913-2916.

12.   Rowan, T., and Wood, H. C. (1968) The biosynthesis of pteridines. V. The synthesis of riboflavin from pteridine precursors., *J Chem Soc Perkin 1. 4*, 452-458.

13.   Smith, E. C., and Metzler, D. E. (1963) The photochemical degradation of riboflavin., *J Am Chem Soc. 85*, 3285-3788.

14.   Fieschi, F., Nivière, V., Frier, C., Décout, J. L., and Fontecave, M. (1995) The mechanism and substrate specificity of the NADPH:flavin oxidoreductase from *Escherichia coli*., *J Biol Chem. 270*, 30392-30400.

15.   Said, H. M., Ortiz, A., Ma, T. Y., and McCloud, E. (1998) Riboflavin uptake by the human-derived liver cells Hep G2: mechanism and regulation., *J Cell Physiol. 176*, 588-594.

16.   Jacobson, F., and Walsh, C. (1984) Properties of 7,8-didemethyl-8-hydroxy-5-deazaflavins relevant to redox coenzyme function in methanogen metabolism., *Biochemistry. 23*, 979–988.

17.   Graham, D. E., and White, R. H. (2002) Elucidation of methanogenic coenzyme biosyntheses: from spectroscopy to genomics., *Nat Prod Rep. 19*, 133–147.

18. Graham, D. E., Xu, H., and White, R. H. (2003) Identification of the 7,8-didemethyl-8-hydroxy-5-deazariboflavin synthase required for coenzyme F(420) biosynthesis., *Arch Microbiol.* **180**, 455-464.

19. Coats, J. H., Li, G. P., Kuo, M. S., and Yurek, D. A. (1989) Discovery, production, and biological assay of an unusual flavenoid cofactor involved in lincomycin biosynthesis., *J Antibiot (Tokyo).* **42**, 472-474.

20. Stover, C. K., Warrener, P., VanDevanter, D. R., Sherman, D. R., Arain, T. M., Langhorne, M. H., Anderson, S. W., Towell, J. A., Yuan, Y., McMurray, D. N., Kreiswirth, B. N., Barry, C. E., and Baker, W. R. (2000) A small-molecule nitroimidazopyran drug candidate for the treatment of tuberculosis., *Nature.* **405**, 962-966.

21. Kasai, S., Miura, R., and Matsui, K. (1975) Chemical Structure and Some Properties of Roseoflavin, *Bull. Chem. Soc. Jpn.* **48**, 2877–2880.

22. Petrolli, D. B., Nakanishi, S., Barile, M., Mansurova, M., Carmona, E. C., Lux, A., Gärtner, W., and Mack, M. (2011) The antibiotics roseoflavin and 8-demethyl-8-amino-riboflavin from *Streptomyces davawensis* are metabolized by human flavokinase and human FAD synthetase., *Biochem Pharmacol.* **82**, 1853–1859.

23. Massey, V. (2000) The chemical and biological versatility of riboflavin., *Biochem Soc Trans.* **28**, 283-296.

24. Macheroux, P., Kappes, B., and Ealick, S. E. (2011) Flavogenomics--a genomic and structural view of flavin-dependent proteins., *FEBS J.* **278**, 2625-2634.

25. Juárez, O., Nilges, M. J., Gillespie, P., Cotton, J., and Barquera, B. (2008) Riboflavin is an active redox cofactor in the Na+-pumping NADH: quinone oxidoreductase (Na+-NQR) from *Vibrio cholerae*., *J Biol Chem.* **283**, 33162-22167.

26. Monaco, H. L. (1997) Crystal structure of chicken riboflavin-binding protein., *EMBO J.* **16**, 1475-1807.

27. Grininger, M., Staudt, H., Johansson, P., Wachtveitl, J., and Oesterhelt, D. (2009) Dodecin Is the Key Player in Flavin Homeostasis of Archaea., *J Biol Chem.* **284**, 13068–13076.

28. Ghisla, S., and Massey, V. (1986) New flavins for old: artificial flavins as active site probes of flavoproteins., *Biochem J.* **239**, 1-12.

29. Ghisla, S., and Massey, V. (1989) Mechanisms of flavoprotein-catalyzed reactions., *Eur J Biochem.* **181**, 1-17.

30. Joosten, V., and van Berkel, W. J. (2007) Flavoenzymes., *Curr Opin Chem Biol.* **11**, 195-202.

31. Fraaije, M. W., and Mattevi, A. (2000) Flavoenzymes: diverse catalysts with recurrent features., *Trends Biochem Sci.* **25**, 126-132.

32. Rigby, S. E., Basran, J., Combe, J. P., Mohsen, A. W., Toogood, H., van Thiel, A., Sutcliffe, M. J., Leys, D., Munro, A. W., and Scrutton, N. S. (2005) Flavoenzyme catalysed oxidation of amines: roles for flavin and protein-based radicals., *Biochem Soc Trans.* **33**, 754-757.

33. Mansoorabadi, S. O., Thibodeaux, C. J., and Liu, H. W. (2007) The diverse roles of flavin coenzymes--nature's most versatile thespians., *J Org Chem.* **72**, 6329-6342.

34. Bieger, B., Essen, L. O., and Oesterhelt, D. (2003) Crystal structure of halophilic dodecin: a novel, dodecameric flavin binding protein from *Halobacterium salinarum*., *Structure.* **11**, 375-385.

35. DeLano, W. L. (2002) The PyMOL Molecular Graphics System   http://www.pymol.org., (DeLano Scientific, S. C., CA, USA., Ed.).

36. Grininger, M., Zeth, K., and Oesterhelt, D. (2006) Dodecins: a family of lumichrome binding proteins., *J Mol Biol.* **357**, 842-857.

37. Staudt, H., Oesterhelt, D., Grininger, M., and Wachtveitl, J. (2011) Ultrafast Excited-state Deactivation of Flavins Bound to Dodecin, *J Biol Chem. 287*, 17637-17644.

38. Meissner, B., Schleicher, E., Weber, S., and Essen, L. O. (2007) The dodecin from *Thermus thermophilus*, a bifunctional cofactor storage protein., *J Biol Chem. 282*, 33142-33154.

39. Honbou, K., Suzuki, N. N., Horiuchi, M., Niki, T., Taira, T., Ariga, H., and Inagaki, F. (2003) The crystal structure of DJ-1, a protein related to male fertility and Parkinson's disease., *J Biol Chem. 278*, 31380-31384.

40. Quigley, P. M., Korotkov, K., Baneyx, F., and Hol, W. G. (2003) The 1.6-A crystal structure of the class of chaperones represented by *Escherichia coli* Hsp31 reveals a putative catalytic triad., *Proc Natl Acad Sci U S A. 100*, 3137-3142.

41. Tao, X., and Tong, L. (2003) Crystal structure of human DJ-1, a protein associated with early onset Parkinson's disease., *J Biol Chem. 278*, 31372-31379.

42. Wei, Y., Ringe, D., Wilson, M. A., and Ondrechen, M. J. (2007) Identification of functional subclasses in the DJ-1 superfamily proteins., *PLoS Comput Biol. 3*, e10.

Chapter 11

Aims of the project ppBat from *Bacteroides thetaiotaomicron*

Aims of the project ppBat from *Bacteroides thetaiotaomicron*

The structure of a novel putative protease from *Bacteroides thetaiotaomicron*, named ppBat has been recently solved. The protein shows an intriguing combination of a flavin binding motif and a zinc binding site which has never been observed for any other protein until to date. Out of the characterized crystal structure, the flavin derivative bound between the dimeric interface of the protein could not be identified, because the flavin cofactor was truncated in proximity of the C5-carbon of the ribityl chain. Additionally one zinc atom is coordinated to the protein in a perfect 1:1 stoichiometry; the zinc ion binds two cysteine residues and two water molecules. This binding motif is completely unusual since no other protein possessing a similar coordination sphere for the zinc has been previously identified.

Initially, we were interested to shed light on the flavin species naturally bound to ppBat. Furthermore, we aimed to determine the affinity between the natural flavin and the protein and to understand whether the flavin is necessary for the dimerization of the protein. From the analysis of the crystal structure, it seemed that two tryptophan residues were fundamental for binding the flavin; therefore the function of these tryptophan residues was ascertained by site-directed mutagenesis studies. Additional points of our investigation concerning the chemical properties of the flavin moiety were: (i) investigation of the reduction process (ii) determination of the reoxidation rate (iii) quantification of the affinity between the reduced flavin species and ppBat. Finally, we were interested to identify the role of the zinc ion in ppBat regarding a possible catalytical or structural role.

It was quite unexpected to figure out that ppBat binds all naturally occurring flavin derivatives with high affinity, functioning therefore as a "flavin catcher". Hence this finding would pave the way to employ ppBat for the production of biosensors. In fact ppBat was applied for the determination of riboflavin in cow´s milk samples.

Chapter 12

Reverse Structural Genomics: An unusual flavin-binding site in a putative protease from *Bacteroides thetaiotaomicron*

AUTHOR CONTRIBUTIONS

The manuscript has been has been published on the Journal of Biological Chemistry (2012), Vol. 287, No. 33, pages 27490-27498. The research was carried out in cooperation with the Department of Chemistry and Chemical Biology (Cornell University, Ithaca, USA). My contribution consisted in the cloning, expression and purification of ppBat and its variants in presence and absence of flavins and flavin derivatives. Moreover, I performed fluorescence spectroscopy, oligomerization studies and determined the extinction coefficients, the dissociation constants as well as the reoxidation rate of ppBat. Furthermore, I identified the cofactors bound to the protein and performed a part of the difference titration experiments.

Reverse Structural Genomics:

An Unusual Flavin Binding Site in a Putative Protease from *Bacteroides thetaiotaomicron*.

Tanja Knaus[1], Elisabeth Eger[1], Julia Koop[1], Steve Stipsits[1], Cynthia L. Kinsland[2], Steven E. Ealick[2], and Peter Macheroux[1*]

[1]From the Institute of Biochemistry, Graz University of Technology, A-8010 Graz, Austria and the

[2]Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14850, U.S.A.

Background: The structure of a putative protease from *Bacteroides thetaiotaomicron* has a unique binding site for a flavin.

Results: The protein tightly binds to lumichrome, riboflavin, FMN, FAD and flavin derivatives.

Conclusion: The putative protease is a scavenger of flavin and may function as a storage protein in gut bacteria.

Significance: The wealth of information available for cofactors is invaluable to assess protein structures.

SUMMARY

The structure of a putative protease from *Bacteroides thetaiotaomicron* features an unprecedented binding site for flavin mononucleotide. The flavin isoalloxazine ring is sandwiched between two tryptophan residues in the interface of the dimeric protein. We characterized the recombinant protein with regard to its affinity for naturally occurring flavin derivatives and several chemically modified flavin analogs. Dissociation constants were determined by isothermal titration calorimetry. The protein has high affinity to naturally occurring flavin derivatives, such as riboflavin, FMN and FAD as well as lumichrome, a photodegradation product of flavins. Similarly, chemically modified flavin analogs showed high affinity to the protein in the nanomolar range. Replacement of the tryptophan by phenylalanine gave rise to much weaker binding while in the tryptophan to alanine variant flavin binding was abolished. We propose that the protein is an unspecific scavenger of flavin compounds and may serve as a storage protein *in vivo*.

Structural genomics efforts have led to an increasing number of structures of proteins that are not characterized on a functional biochemical level. In the case of homologous structures, biochemical function may be inferred to some extent and thus provide an initial lead for further functional investigations. In a recent survey of flavin-dependent proteins, we identified a putative protease from *Bacteroides thetaiotaomicron*, a microbe inhabiting the human gut[1, 2], that presumably binds FMN[3]. The structure of the protein was solved to 2 Å resolution by the Midwest Centre for Structural Genomics (MCSG) facility (pdb code 3cne, Figure 1A) and features a mono nuclear metal binding site, presumably occupied by zinc although other metals such as cobalt may be the native metal in the protein. In addition, a flavin was found in the protein, which is sandwiched by two tryptophan residues provided by each of the two protomers in the dimeric protein (Figure 1B). Based on its overall topology, the protein was classified as a member of the DJ-1 superfamily (Pfam DJ-1_PfpI), a heterogenous family of hydrolases and chaperones with many uncharacterized and poorly understood members[4-7]. The hydrolase activity found in some members appears to be associated with a catalytic triad formed by a conserved cysteine, histidine and aspartic acid[5, 8]. Other members of the family, such as Hsp31 and YajL from *Escherichia coli* were reported to display activity as a holding and covalent chaperone,

respectively[9, 10]. The close structural relationship suggests that the zinc ion might be the catalytic centre of a hydrolytic activity although none of the other family members appear to require metal ions for catalytic activity. Furthermore, complexation of

the zinc is atypical for Zn-dependent hydrolases with two cysteine-derived thiol groups (Figure 1C, [11]). In addition, the zinc is liganded by water molecules, however, this interaction varies in the four chains of the asymmetric unit.
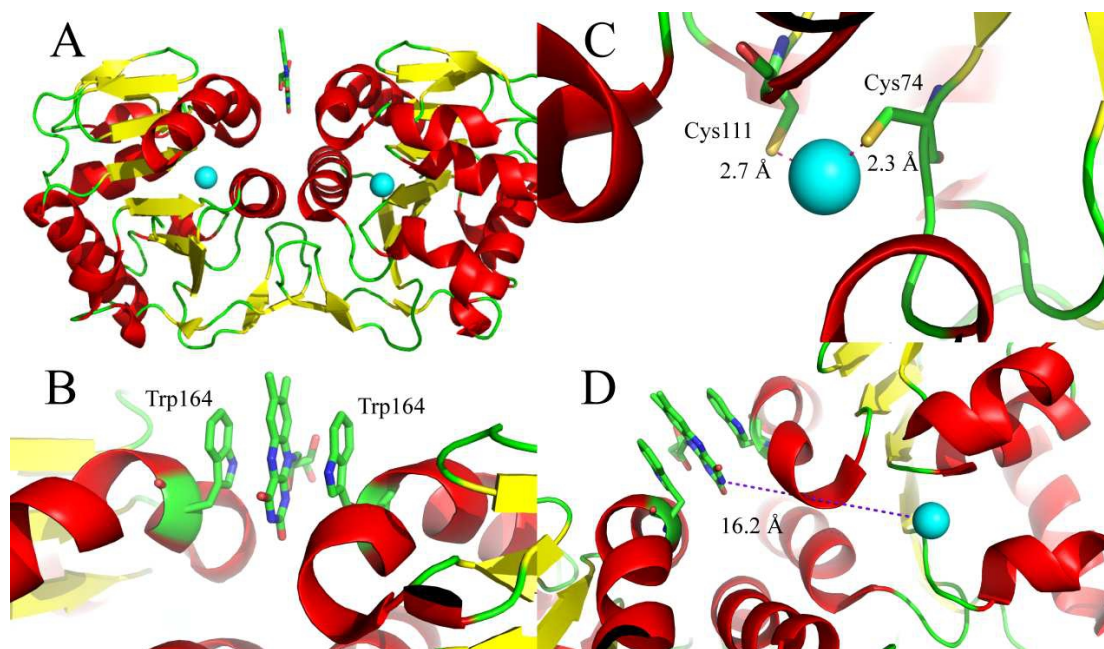
Figure 1. Structural representation of the putative protease from *Bacteroides thetaiotaomicron* (pdb code 3cne). Panel A: Overall topology of the dimeric protein in a cartoon representation with α-helices in red and β-strands in yellow. Zinc ions are shown as cyan spheres. The flavin is shown as a stick model. Panel B: Flavin binding site in the dimer interface of the putative protease with the two tryptophan residues (one from each protomer) sandwiching the isoalloxazine ring shown in a stick model. Panel C: Zn ion coordination sphere (one in each of the two protomers). Zn is shown as a sphere with the two coordinating cysteine residues (Cys74 and Cys111) and two coordinating water molecules in 3.6 and 3.1 Å distance from the zinc ion. Panel D: Spatial relationship between the zinc ion and the flavin. The closest atom to the zinc ion is N(3) of the isoalloxazine ring with a distance of 16.3 Å.

In contrast to zinc, the flavin cofactor occurs predominantly in oxidoreductases (ca. 90% of flavin-dependent enzymes) and not hydrolases. Although the flavin cofactor is also present in other enzyme classes, like transferases, lyases and isomerases, it has never been reported in a hydrolase[3]. The large distance of ≈ 16 Å between the edge of the flavin cofactor and the zinc argues against a direct cooperation between these two cofactors (Figure 1D). This intriguing combination of unusual cofactors sparked our interest in the protein. Here, we report the biochemical characterization of the flavin binding site of the protein and demonstrate that the recombinant protein is capable of binding not only naturally occurring flavin derivatives, such as lumichrome, riboflavin, FMN and FAD (supplementary figure S1 panel A) but also a variety

of chemically modified "artificial" flavin analogs, such as iso-riboflavin (6.7-dimethyl-8-nor-riboflavin) and iso-FMN as well as 8-amino-, 1- and 5-deaza-riboflavin[12-14] (supplementary figure S1 panel B). Thus it appears that this putative protease from *B. thetaiotaomicron* (ppBat) unspecifically binds tricyclic aromatic systems such as the alloxazine and isoalloxazine ring systems. Furthermore, we showed that replacement of Trp164 (Figure 1C) by alanine completely abolishes the ability to bind the flavin ring system, while the Trp164Phe variant retained some affinity. Sequence alignment of ppBat with other related putative proteases from the genus *Bacteroides* and related bacteria revealed that Trp164 is conserved in the majority of homologs. Therefore, we propose that ppBat may serve as a flavin storage protein in gut bacteria.

EXPERIMENTAL PROCEDURES

Reagents

All chemicals were of the highest grade commercially available from Sigma-Aldrich (St. Louis, MO, U.S.A.), Fluka (Buchs, Switzerland) or Merck (Darmstadt, Germany). Ni-NTA agarose was from GE Healthcare, Bio-Sciences AB (Uppsala, Sweden) and Sephadex resin was from Pharmacia Biotech AB (Uppsala, Sweden). Flavin analogs were a generous gift from Prof. Sandro Ghisla and the late Prof. Peter Hemmerich (University of Konstanz, Germany).

Expression of recombinant protein in E. coli host cells

The gene *gat1* of *Bacteroides thetaiotaomicron* was cloned into a derivative of pET-28 for IPTG-inducible expression in *E. coli* BL21 (DE3) host cells (kanamycin resistance). The N-terminus of the expressed protein possesses a hexa-histidine tag to facilitate purification. Precipitated cells were resuspended in lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0), using 2 ml buffer per gram of wet cells. At this stage lumichrome, riboflavin, FMN, FAD or modified flavin analogs were added in excess to saturate the protein with the respective ligand. In order to obtain the apo-protein (protein preparation "as isolated") no ligands were added. Resuspended cells were disrupted by 0.5 s sonication pulses for 10 min while cooling on ice. The cell debris was removed by centrifugation at 18,000 g for 30 min at 4 °C. The hexa-histidine tagged protein was purified by Ni-NTA affinity chromatography, loading the supernatant onto a Ni-NTA HisTrap^{TM}HP column (GE Healthcare), previously equilibrated with lysis buffer. After loading of the filtrated lysate, the column was washed with ten column volumes of wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH 8.0) and bound protein was recovered with elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 150 mM imidazole, pH 8.0). The purity of the eluted fractions was determined by SDS/PAGE. Fractions containing ppBat were pooled, dialyzed against 20 mM Tris/HCl buffer containing 100 mM NaCl (pH 8.0) over night and concentrated using Centripreps (Millipore). The final concentration was determined at 280 nm using an ε_{280} of 13,075 M^{-1} cm^{-1}. The protein was flash-frozen and stored at -20 °C.

Site-directed mutagenesis

The Trp164Phe and Trp164Ala variants were prepared using the QuikChange$^{(R)}$ XL-site directed mutagenesis kit (Stratagene) with a slight modification of the protocol. A two-step PCR reaction was carried out, where separate reactions for forward and reverse primer were performed for 4 cycles prior to another 18 cycles where an equal amount of the forward and reverse reaction were mixed. Following primers were used for the insertion of the mutations: 5´-CTGCACAAGATGAGAATAC GATCTTTACAATGTTGCCGAAAGTCATAG-3´ and 5´-CTATGACTTTCGGCAACATTGTAAAGA TCGTATTCTCATCTTGTGCAG-3´for Trp164Phe and 5´-CTGCACAAGATGAGAATACGATCGCG ACAATGTTGCCGAAAGTCATAG-3´and 5´-CTA TGACTTTCGGCAACATTGTCGCGATCGTATTC TCATCTTGTGCAG-3´for Trp164Ala, respectively. After verification of positive transformants by sequencing, expression and purification was achieved as described for the wild-type protein.

UV/vis absorbance and fluorescence spectroscopy

UV/visible absorbance spectra were recorded with a Specord 210 spectrophotometer (Analytik Jena, Jena, Germany). Fluorescence emission spectra were recorded with a Shimadzu RF301 PC spectrofluorophotometer. Difference titrations were carried out at 25 °C in tandem-cuvettes by the addition of ppBat to the flavin in the measurement cell and to buffer (20 mM Tris/HCl, 100 mM NaCl, pH 8.0) in the reference cell at 2 min intervals. Fluorescence titrations were performed at 25 °C in 20 mM Tris/HCl, 100 mM NaCl, pH 8.0 buffer. For quenching of the FMN fluorescence, 1.8 ml of 1 μM FMN solution were titrated with 2 μl aliquots of 100 μM protein solution every 2 min under stirring. Emission at 524 nm was measured at an excitation of 467 nm until an endpoint was reached. Spectra were corrected for dilution as well as for fluorescence of enzyme bound flavin. In case of tryptophan fluorescence, 2 ml of 15 μM protein solution were titrated with 2 μl aliquots of 1.5 mM FMN solution every 2 min (emission at 343 nm, excitation at 282 nm).

Isothermal titration calorimetry (ITC)

Dissociation constants for the binding of flavins to ppBat were determined using a VP-ITC system (MicroCal). All experiments were performed at 25 °C

in 20 mM Tris/HCl, 100 mM NaCl, pH 8.0 buffer and solutions were degassed immediately before measurements. Titration experiments for FMN, FAD, riboflavin and lumichrome consisted of 21 injections (7 µl, duration time 14 s, spacing time 300 s) of a flavin solution (530 µM) to 1.495 ml ppBat (40 µM). 40 µM enzyme were titrated with iso-FMN (370 µM, 21x10 µl, duration time 10 s, spacing time 300 s), 5-deaza-riboflavin (200 µM, 25x10 µl, duration time 20 s, spacing time 300 s) and 25 µM protein were titrated with 8-amino-riboflavin and iso-riboflavin (300 µM, 25x7 µl, duration time 14 s, spacing time 300 s), respectively. 40 µM of the Trp164Phe variant were titrated with 7 µl aliquots of 300 µM riboflavin for 25 times (duration time 14 s, spacing time 300 s). One set of sites fitting with Origin version 7.0 (MicroCal) for ITC data analysis was used to obtain dissociation constants.

Determination of oligomeric state by size exclusion chromatography

Determination of the oligomerization state for the wild-type protein and variants was carried out with a Superdex200 analytical 10/300 GL column (GE Healthcare) using a ÄKTA explorer system. The calibration curve was prepared according to the manufacturer´s protocol. Approximately 1 mg of protein samples as isolated and saturated with FMN, respectively, were dissolved in 20 mM Tris/HCl, 100 mM NaCl, pH 8.0 buffer prior to size exclusion chromatography (flow rate 0.5 ml/min).

Photoreduction

Photoreduction was carried out as outlined in a previous study [15]. The protein was diluted in 20 mM Tris/HCl, pH 8.0 containing 100 mM NaCl, 1 mM EDTA and 2 µM benzyl viologen. Quartz cuvettes were rendered anaerobic by repeated cycles of evacuation and flushing with nitrogen. Reduction of the flavin cofactor was achieved by irradiation of the sample with a 250-W halogen lamp which was cooled to 15 °C during the time of irradiation. Spectra from 300 to 700 nm were recorded at the same temperature until no further changes were observed. Then the cuvette was opened to air and the absorbance spectrum was recorded.

Rapid reaction studies

Stopped-flow measurements were carried out with a Hi-Tech (SF-61DX2) stopped-flow device (TgK

Scientific Limited, Bradford-on-Avon, UK) positioned in a glove box from Belle Technology (Weymouth, UK) at 25 ˚C. Reoxidation of FMN free in solution and bound to ppBat was measured by monitoring changes at A451 (ppBat) or A449 (FMN), respectively, with a KinetaScanT diode array detector (MG-6560, TgK Scientific Limited). Reduction of FMN was achieved by addition of solid sodium dithionite. Reoxidation was then monitored by rapidly mixing the reduced solution with buffer previously equilibrated with air.
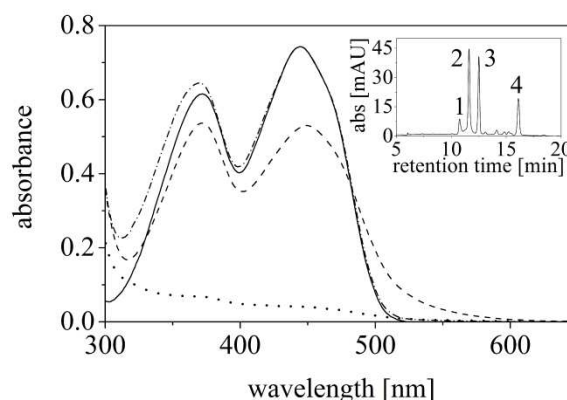
Figure 2. UV/vis-absorbance spectrum of ppBat as isolated (dotted line), ppBat saturated with FMN (dashed line) and after denaturation (dashed-dotted line). The UV/vis absorbance spectrum of free FMN is shown for comparison (solid line). The inset shows the HPLC analysis of bound ligands. The trace shown was recorded at 370 nm: peak 1 = FAD (retention time = 10.7 min), peak 2 = FMN (11.6 min), peak 3 = riboflavin (12.5 min) and peak 4 = lumichrome (16.1 min).

HPLC determination of flavins and lumichrome

Cofactors were released by thermal precipitation of ppBat at 95 °C for 10 min, cooling on ice for 5 min and centrifugation for 5 min. The flavin extracts were loaded onto an Atlantis® dC18 8 µm 4.6 x 250 mm column (Waters) and eluted with a water/acetonitril multi-step gradient (0-1.5 min 0-5% acetonitrile; 1.5-20 min 5-60% acetonitrile; 20-22 min 95% acetonitrile) at a flow rate of 1 ml/min. The elution was monitored by UV-absorption at 370 nm and cofactors were identified according to their elution time and UV/vis absorbance spectra.

Zn-determination

The amount of Zn in ppBat samples was determined by means of inductively coupled plasma optical emission spectrometry (ICPOES).

RESULTS

Expression and purification of ppBat

Cells transformed with the plasmid carrying the gene of ppBat were cultured and protein production was induced by IPTG. The majority of the target protein was soluble and purified by a single Ni-NTA affinity chromatography step (the progress of purification is shown in supplementary Figure S2). The purified protein was obtained with an average yield of 30 mg/g wet cells (ca. 140 mg/l bacterial culture).
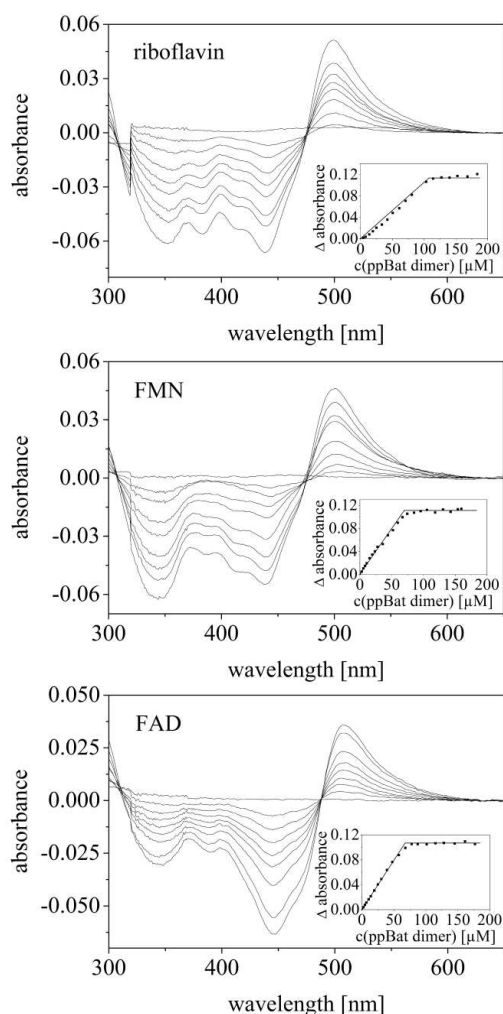
Figure 3. UV/vis absorbance difference titration of riboflavin (80 μM, top panel), FMN (61 μM, central panel) and FAD (60 μM, bottom panel) with ppBat. The data points shown in the insert were manually fitted by straight lines to indicate the titration end points.

Characterization of the isolated protein

The purified protein was pale yellow and had absorbance maxima at 372 and 449 nm indicative for bound flavin. The absorbance ratio A280/A449 was very high (≈ 45) for a small flavoprotein and suggested that the protein was predominantly present in a flavin-free apo-form (Figure 2, dotted line). When FMN was added to the extraction buffer, the protein was isolated with a reproducibly lower A280/A449 ratio of 5 indicating that the apo-protein present in the crude extract was reconstituted to the holo-protein (Figure 2). The extinction coefficient of bound flavin was determined by denaturation of the protein with subsequent determination of the UV/vis absorbance spectrum of the released FMN (ε_{free} = 11,500 M^{-1} cm^{-1}) to ε_{bound} = 8,960 M^{-1} cm^{-1}. Using this extinction coefficient, it was estimated that only 8-10% of the isolated protein is present in the holo-form when no flavin was added to the crude extract. The hypochromic effect at both absorbance maxima is accompanied by a long wavelength absorbance extending beyond the typical flavin absorbance (Figure 2, compare dashed and solid line). In contrast to the low flavin content of the purified protein, zinc was found in stoichiometric amounts, *i.e.* one Zn atom per protomer.

FMN bound to ppBat showed fluorescence excitation and emission maxima similar to that of free flavin, however, fluorescence intensity was approximately 47% lower as compared to free FMN. Similarly, FMN binding to ppBat quenched the tryptophan fluorescence emission at 343 nm to ca. 28%.

Since the deposited structure of the protein has a truncated N(10)-side chain (only four carbon atoms of the side chain were modelled, Figure 1B), it is not clear which flavin derivative is bound to the protein. Therefore, we also tested riboflavin and FAD as potential binding partners of the recombinant protein. We found that riboflavin and FAD also bind to the protein with spectral changes similar to those seen with FMN. Further analysis by difference UV/vis-absorbance spectroscopy revealed that all three flavin compounds bind tightly to the apo-protein with similar spectral perturbations (Figure 3) and molar extinction coefficients (Table 1). All three flavin derivatives produced a sharp endpoint at equimolar concentrations of flavin and dimeric protein in agreement with a 1:1 stoichiometry as suggested by the three-dimensional structure.

In order to identify the ligand(s) bound to ppBat as isolated from *E. coli* cells, the protein was denatured and the released compounds subjected to HPLC analysis. As shown in the insert of Figure 2, four peaks were observed. Based on the retention times

and UV/vis absorbance properties these compounds were identified as FAD (peak 1), FMN (peak 2), riboflavin (peak 3) and lumichrome (peak 4).

Determination of dissociation constants

Due to the apparently tight binding of riboflavin, FMN and FAD to apo-ppBat (Figure 3), we employed isothermal titration calorimetry (ITC) to determine their dissociation constants. As shown in Figure 4, binding of these compounds to apo-protein is exothermic at 25 °C. The experimental data was fitted with a single site binding model and yielded dissociation constants in the range of 74.3 nM and 800.6 nM for riboflavin and FAD, respectively (Table 2). Obviously, flavin binding in the dimer interface of ppBat is adversely influenced by the size of the side chain attached to N(10).

Table 1. Molar extinction coefficients (ε) of flavins and flavin analogs free and bound to ppBat.

Flavin	ε (free/bound) (M^{-1} cm^{-1})	λ_{max} (free/bound) (nm)	reference
Lumichrome	8,200/6,400	347/348	16
Riboflavin	12,500/9,800	444/449	16
iso-riboflavin	7,300/5,650	448/454	17
1-deaza-riboflavin	6,800/6110	536/558	18
5-deaza-riboflavin	11,500/9,620	399/398	14
8-amino-riboflavin	42,000/24,570	479/487	19
2'-deoxy-riboflavin	12,500/9950	443/452	ε of riboflavin was used
FMN	12,500/8,960	444/449	16, 20
FAD	11,300/8,500	448/449	16, 17, 20
iso-FMN	7,200/5,600	449/454	21

Binding of chemically modified flavins

In addition to the naturally occurring flavin compounds, ppBat also hosted a variety of derivatized riboflavin and FMN compounds, such as iso-riboflavin and iso-FMN (6.7-dimethylisoalloxazine), 2-thio-, 8-amino-, 1- and 5-deaza-riboflavin as well as 5-deaza-lumiflavin and lumichrome. Lumazine did not bind to ppBat. UV/vis absorbance spectra of four selected examples are shown in Figure 5 (see also Figures S3 and S4). In order to determine the stoichiometry and affinity of these flavin analogs, difference titrations were performed with an UV/vis-absorbance spectrophotometer. As shown in Figure 6, we found tight binding with all tested flavin analogs and a stoichiometry of 1:1 (flavin:dimer) for complex formation with ppBat. The binding of five flavin analogs was further examined by ITC. The three riboflavin analogs, 5-deaza-riboflavin, iso-riboflavin and 8-amino-riboflavin showed dissociation constants similar to riboflavin and iso-FMN bound even tighter than FMN (Table 2). Interestingly, 2'-deoxy-riboflavin lacking the hydroxyl group at the 2'-position of the ribityl side chain had only a marginally higher dissociation constant (117.4 vs. 74.3 nM, see also Figure S5) suggesting that the carbonyl oxygen of Asn161 forms only a weak hydrogen bond to the C2'-OH group (Table 2).

Properties of variants Trp164Phe and Trp164Ala

To investigate the importance of the indole side chain of Trp164 for dimerization of the protein and for flavin binding, two single replacement variants were generated by site-directed mutagenesis: Trp164Phe and Trp164Ala. Similar to wild-type ppBat, both variants were found to form dimers in solution as judged by size exclusion chromatography (see supplementary Figure S6). In the presence of FMN in the crude extract, the Trp164Phe variant retained the ability to bind FMN whereas the Trp164Ala variant was essentially devoid of FMN (Figure 7). In the absence of FMN in the crude extract, only trace amounts of flavin could be detected in both variants. Detailed HPLC-analysis of ligands released upon denaturation of the protein variants revealed that small amounts of FAD, FMN, riboflavin and lumichrome were bound (inserts in Figure 7). For the Trp164Phe variant the dissociation constant of riboflavin was determined by ITC to 1.19 ± 0.15 mM, i.e. 16-times higher than for wild-type ppBat.

Recombinant ppBat loaded with FMN was found to be present as a dimer in solution using analytical size exclusion chromatography. Both tryptophan variants yielded a virtually identical peak indicating that neither of the two amino acid substitutions affected dimerization of ppBat (supplementary Figure S6).
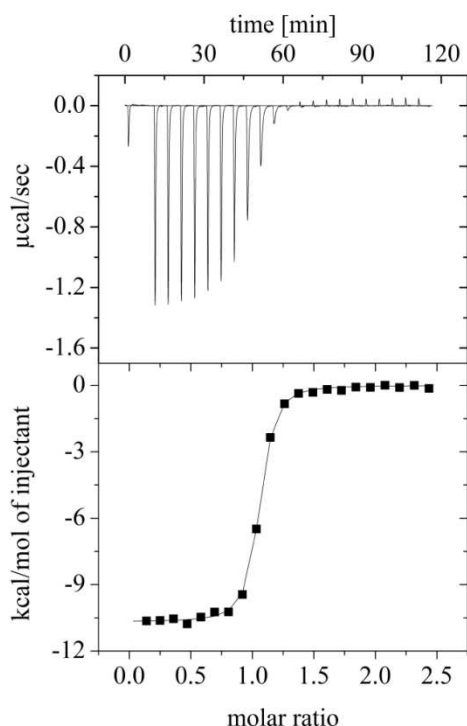
Figure 4. Representative example for the determination of a dissociation constant by isothermal titration microcalorimetry. The measurement shows the titration of 40 μM ppBat with riboflavin in 20 mM Tris/HCl, 100 mM NaCl, pH 8.0 at 25 °C.

Table 2. Dissociation constants for binding of lumichrome, riboflavin, FMN and FAD as well as modified flavin analogs to ppBat as determined by ITC. The values given are the average of three measurements carried out in 20 mM Tris/HCl, 100 mM NaCl, pH 8.0 buffer at 25 °C.

Flavin ligand	K_d (nM)
lumichrome	272 ± 61
riboflavin	74 ± 5
5-deaza-riboflavin	84 ± 5
iso-riboflavin	96 ± 2
8-amino-riboflavin	227 ± 26
2'-deoxy-riboflavin	117 ± 3
FMN	487 ± 19
iso-FMN	295 ± 21
FAD	801 ± 82

Photoreduction and reoxidation

Wild-type ppBat reconstituted with FMN was subjected to photoreduction under anoxic conditions. The course of the reduction is shown in Figure 8. The spectral changes upon photoreduction featured a loss of absorption at 449 nm and an isosbestic point at 345 nm. The final spectrum is characteristic for the fully reduced dihydro quinone form of the flavin. During reduction, spectral hallmarks of a flavin radical species, like an increased absorbance at 370 nm or

absorbance above 550 nm were absent indicating that no radical species is thermodynamically stabilized during reduction. The fully reduced dihydroflavin could be readily reoxidized by allowing access of oxygen to the cuvette (Figure 8). The rate of reoxidation was determined in a stopped-flow apparatus by rapid mixing of sodium dithionite reduced ppBat with air-equilibrated buffer (inset Figure 8). This yielded a $k_{obs} = 1.1 \pm 0.02$ s^{-1} (average of five independent measurements). Using the same experimental set-up, the reoxidation of free FMN yielded a $k_{obs} = 6.7 \pm 0.09$ s^{-1}, i.e. ca. 6-times faster than for FMN bound to ppBat (all measurements were performed at 25 °C).

DISCUSSION

Flavins are widely used coenzymes for reduction-oxidation processes with well over 300 different enzymatic reactions described in the literature [3, 22-24]. In addition to their role as redox catalysts, flavins are also found in some none redox enzymes, such as isomerases, lyases and transferases[3, 25]. However, the occurrence of a flavin (purportedly FMN) in a putative protease from *B. thetaiotaomicron* (ppBat) was a novelty since hydrolases have never been associated with this cofactor. The discovery that ppBat features a flavin prompted us to investigate the properties of the flavin binding site in this unusual family of proteases/chaperones (Figure 1). At first, the low flavin content of isolated ppBat was surprising in view of the fact that the deposited crystal structure clearly showed the presence of flavin bound between two tryptophans in the dimer interface of the protein (Figure 1A and B). Addition of a flavin, for example FMN, increased the flavin content close to one flavin per protein dimer, as could be demonstrated by titration experiments (see Figures 3 and 6). According to the information given by the authors of the crystal structure, no flavin (neither riboflavin, FMN nor FAD) was added in the crystallization trials and hence it follows that the population of ppBat with flavin bound (ca. 10%) preferentially crystallized, *i. e.* was selected by the applied conditions. Since our analysis of ppBat isolated without flavin addition has revealed the presence of several chromophores (lumichrome, riboflavin, FMN and FAD) the nature of the bound species obviously does not matter with regard to the propensity to generate crystals. The deposited crystal structure is also lacking information on the exact nature of the side chain as only the four carbon atoms

proximal to N(10) were modelled. Therefore the structure cannot distinguish between riboflavin, FMN and FAD. In the present study, we have shown that lumichrome as well as riboflavin, FMN and FAD all bind to ppBat with dissociation constants in the nanomolar range (see Table 2).

The UV/vis absorbance spectra of all flavins bound to ppBat showed pronounced hypochromicity and long wavelength absorbance indicative of a

charge-transfer complex (Table 1). This is in contrast to the rather modest quenching of both the flavin and the tryptophan fluorescence emission. Interestingly, virtually complete fluorescence quenching was observed for both fluorophores in model compounds, where stacking interactions between the isoalloxazine and indole ring were assumed[26]. Apparently, aromatic stacking interactions do not necessarily lead to complete fluorescence quenching.
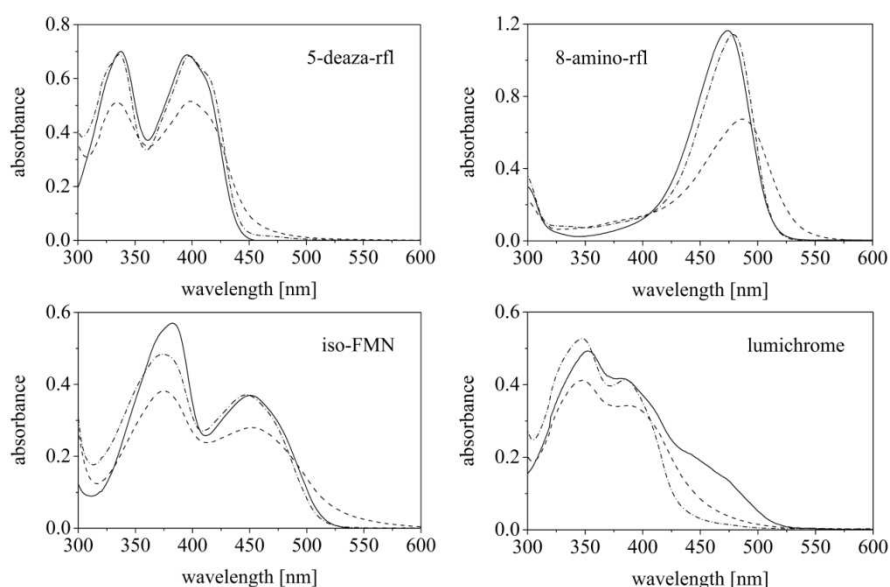
Figure 5. UV/Vis-absorbance spectra of ppBat with three different flavin analogs and lumichrome. Dashed lines represent spectra of ppBat saturated with the ligand, dashed-dotted lines were recorded after denaturation of the protein and solid lines are the spectra of the free chromophore.
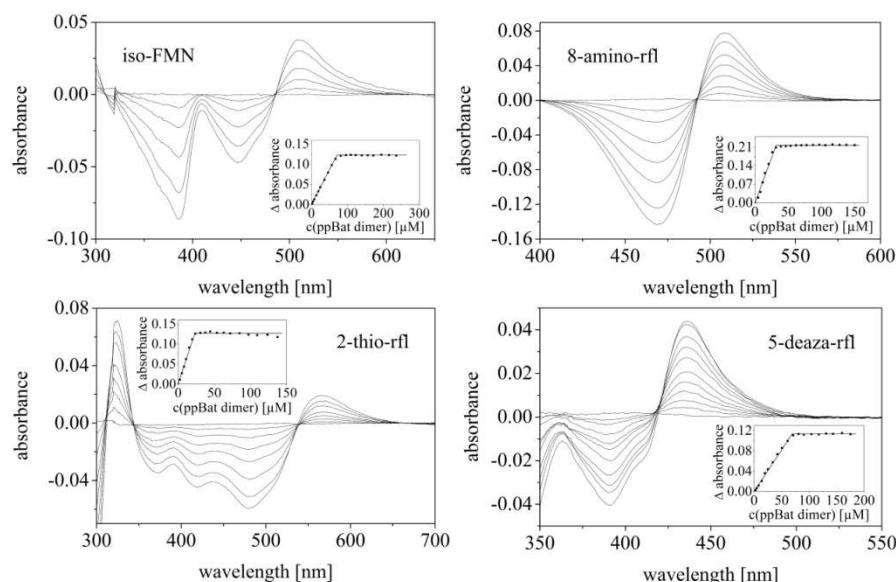
Figure 6. UV/Vis absorbance difference titration of iso-FMN (76 µM), 8-amino-riboflavin (rfl) (17 µM), 2-thio-riboflavin (32 µM) and 5-deaza-riboflavin (48 µM).

A salient feature of flavoproteins is the stabilization of either the red (anionic) or blue (neutral) flavin radical (semiquinone)[12]. In the case of ppBat,

reduction of bound flavin to the hydroquinone proceeded without the observable occurrence of a radical species (Figure 7). Similarly, reoxidation of

the reduced flavin yielded the oxidized form without any observable radical intermediate. Interestingly, the observed rate of reoxidation was approximately six times slower than for free flavin indicating that the mode of flavin binding in ppBat protects the flavin from reoxidation. This is in keeping with the current understanding of the control of oxygen reactivity by flavoenzymes. In recent years, it has emerged that enhanced oxygen reactivity is effected by an oxygen binding site near the C(4a)-position of the isoalloxazine ring where a superoxide anion is stabilized by a near-by positive charge, *e.g.* a lysine side-chain[27, 28]. The flavin binding site in ppBat neither provides interactions necessary to stabilize a flavin radical nor a superoxide anion near the C(4a) position. Thus the observed properties are in full agreement with the flavin binding mode.
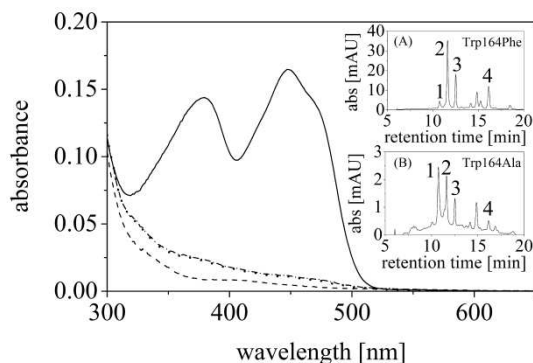
Figure 7. UV/Vis-absorbance spectra of the Trp164Phe and Trp164Ala variants.The solid and dashed lines represent the absorbance spectra of the Trp164Phe and Trp164Ala variant, respectively, purified with FMN added to the crude extract. The dotted and dash-dotted lines were obtained in the absence of additional FMN ("as isolated"). The insets show the HPLC analysis of bound ligands for the two variants as isolated. The traces show the absorbance at 370 nm: peak 1 = FAD, peak 2 = FMN, peak 3 = riboflavin, peak 4 = lumichrome.

Our studies have clearly shown that Trp164 is crucial for binding of flavins to the dimer interface of ppBat. In contrast to the aromatic stacking interaction, the assumed contact between the C2'-hydroxyl group of the ribityl side chain and the main chain carbonyl of Asn161 contributes very little (1.14 kJ/mol) to the binding energy (Figure 9). On the other hand, lumichrome which features an alloxazine rather than an isoalloxazine π-electron system and has no side chain at N(10), binds weaker than riboflavin (ca. 6 kJ/mol) and 2'-deoxy-riboflavin (ca. 4.86 kJ/mol) suggesting that the side chain favourably affects binding to ppBat. Alignment of 29 sequences with at

least 50% overall sequence identity showed that Trp164 and Asn161 are conserved in 18 and 22 proteins, respectively (Figure 10). On the other hand, the two cysteines that complex the zinc ion are present in all of the 29 proteins. This suggests that the (catalytic?) function of the protein is related to the presence of the zinc centre and flavin binding may have evolved as an additional asset in a subset of homologous proteins.
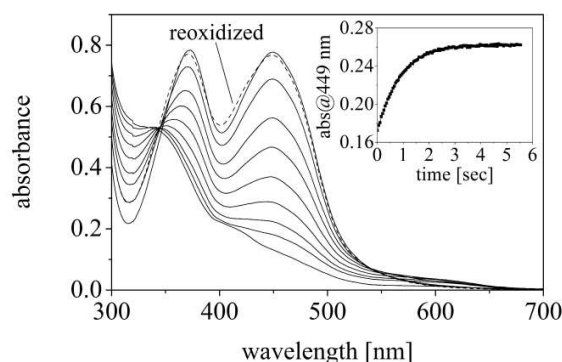
Figure 8. Anerobic photoreduction of ppBat in the presence of EDTA and benzylviologen shows that the FMN moiety can be directly reduced to its fully reduced form in a two-electron reduction process (isosbestic point at 343 nm) without the formation of either the red or the blue semiquinone radical. Absorbance spectra were recorded prior to illumination (top, solid line) and after 5, 7, 9, 11, 14, 21, 33 and 103 min, respectively (from top to bottom). The spectrum after reoxidation with dioxygen is shown as a dashed line. Inset shows reoxidation in the stopped-flow apparatus by molecular oxygen.

The binding mode in ppBat also rationalizes our findings that the protein is rather insensitive to the N(10) side chain carried by the isoalloxazine ring. Most flavoenzymes exhibit pronounced preference for either FMN or FAD and riboflavin is not used at all except in one case[3]. In addition, flavins with modifications in the isoalloxazine ring have lower affinity due to adverse effects on specific interactions with amino acids in the flavin binding pocket. As demonstrated by our affinity measurements, binding of modified flavins is in the same range as for the naturally occurring flavins and mainly governed by the side chain (*i.e.* iso-riboflavin and 5-deaza-riboflavin have similar affinities than riboflavin and iso-FMN is similar to FMN). Therefore it appears that ppBat does not utilize the bound flavin for catalysis. Likewise, since flavin binding is also not required to maintain the dimeric structure of ppBat, we conclude that flavin binding to ppBat does not concern the potential catalytic function of the

protein. On the other hand the strong binding of flavins, in particular of riboflavin, suggests that ppBat may play a role as a storage protein for gut bacteria.
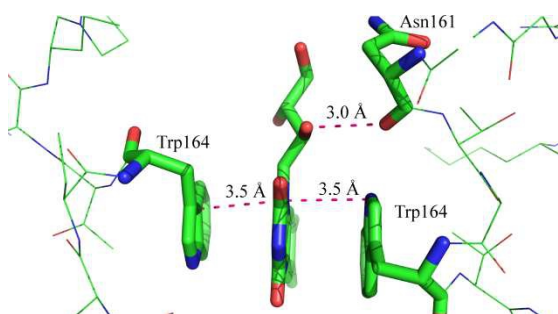
Figure 9. Close-up of the flavin binding site of ppBat. The two tryptophans are almost equidistant to the plane of the isoalloxazine ring system (orange lines). The main chain carbonyl group of Asn161 engages in a hydrogen bond (magenta) to C'2-OH of the ribityl side chain. Distances are given in Ångstrom. The graphic was generated using PyMol.

A related flavin binding mode was recently discovered in a family of archaeal and bacterial

proteins. These so-called dodecins consist of twelve protomers of 68-71 amino acids with six flavin binding sites[29]. As in ppBat, two tryptophans sandwich the isoalloxazine or alloxazine (lumichrome) ring system, albeit, each binding site accommodates two ligands instead of one. In addition, other amino acids engage in hydrogen-bond interactions to the flavin ligand, for example an arginine side chain forms hydrogen bonds to $C(4)=O$ of the isoalloxazine ring[29-31]. In the case of ppBat the only contacts are between the tryptophans of the two protomers and the isoalloxazine ring and a (weak) hydrogen bond between the carbonyl group of Asn161 of one protomer and the C'2-hydroxyl group of the ribityl side chain. Contacts between the pyrimidine ring and ppBat are not observed in the structure and hence the affinity of flavins to ppBat is mainly governed by the interaction with the indole side chains of the two tryptophans. Therefore it is not surprising that replacement of the tryptophans by alanine and phenylalanine has marked effects on the binding affinity.


```
                      C-terminus (156-175)  identity  similarity
                                               %          %
B.thetaiotaomicron    TAQDENTIWTMLPKVIEALK    100         -
B.caccae              TAQDENTIWTMLPQVIEALK     93         97
B.xylanisolvens XB1A  TAQDENTIWTMLPHIIEALK     93         97
B.sp. 3_1_23          TAQDENTIWTMMPQVIEALK     92         96
B.ovatus              TAQDENTIWTMMPQVIEALK     92         96
B.sp. D2              TAQDENTIWTMLPQVIEALK     92         97
B.sp. D22             TAQDENTIWTMLPHIIEALK     92         96
B.finegoldii          TAQDENTIWTMMPQVIKALK     91         97
B.xylanisolvens SD    TAQNENTIWTMMPQVTEALK     91         96
B.sp. D1              TAQNENTIWTMMPQVTEALK     91         96
B.eggerthii           TAQDENNIQTMIEKIVAALK     84         93
B.eggerthii DSM20697  TAQDENNIRTMIEKIVTALK     84         93
B.fragilis            TAQDENTLWTMMPQLIEALK     82         91
B.clarus              TAQDENNIRSMIEKVVAALK     81         90
B.sp. 20_3            TAKEEHYICTMMPQVLEALK     78         92
B.sp. 3_1_19          TAKEEHYICTMMPQVLEALK     77         91
Pb.johnsonnii         TAQDENSIWTMMSQVLDALK     77         88
B.vulgatus DSM1447    TAQSENFVWTMMPQVIEALK     75         89
B.vulgatus PC510      TAQSENFVWTMMPQVIEALK     75         89
B.dorei               TAQSENFVWTMMPQVIEALK     75         89
B.dorei DSM17855      TAQSENFVWTMMPQVIEALK     75         89
B.sp. 3_1_40A         TAQSENFVWTMMPQVIEALK     74         89
B.sp. 4_3_47FAA       TAQSENFVWTMMPQVIEALK     74         89
B.splanchnicus        TAQCEHTLPVLMPRLVEAL      68         83
A.shahii              TAQEEHAIPAMIGRVVEALK     67         81
A.putredinis          TAQNENTISCLLPELLKVLK     61         81
B.coprocola           TAQCEHTLPSMMPQLIEVLK     62         82
D.gadei               TAQTEDAIPSLIPGLLKAL      58         77
S.termiditis          TAKTEESAMEMIPYFIWAL      54         67
```

Figure 10. Alignment of the last 20 amino acids of the C-terminus of ppBat (amino acids 156 to 175) and related proteins from *Bacteroides* (B.), *Parabacteroides* (Pb.), *Alistipes* (A.), *Dysgomonas* (D.) and *Sebaldella* (S.). The tryptophan at position 164 is highlighted in green and the asparagine 161 in magenta. The percentage identity and similarity are given for the full-length proteins (175 amino acids). Two cysteines (Cys 74 and Cys111, see Figure 1C) are conserved in all sequences shown in this figure. Additionally, Asp76, His131 and Glu160, which are between 5 and 6 Å from the metal binding site, are conserved in all sequences but *B. vulgatus*.

In the case of alanine, binding was virtually abolished whereas the Trp164Phe exchange has retained some affinity (see Figure 7). In contrast, replacement of Trp38 in dodecin from *Thermophilus* has apparently no consequence on the propensity to bind flavins as several other interactions still ensure stable complex formation (see pdb entry 2vyx[29]).

This protein also binds coenzyme A and it appears that the dodacamer binds a mixture of cofactors *in viv* [29, 31]. Therefore dodecins appear to serve as cofactor storage proteins and may also be important as scavengers of potentially harmful cofactor degradation products, such as lumichrome, which is produced in the course of photodegradation of riboflavin *in vivo*[29, 31, 32]. The dissociation constants determined for dodecin from *Thermophilus* are similar to those found for ppBat and support the notion that ppBat may as well serve as a storage protein in gut bacteria *in vivo*. A BLASTp search using the dodecin sequence of *T. thermophilus* against the *B. thetaiotaomicron* genome failed to retrieve a dodecin homolog.

The distance between the plane of the indole ring of the two tryptophans and the isoalloxazine in ppBat is ca. 3.5 to 3.7 Å (Figure 9). A similar distance for π-stacking interaction of a tryptophan and an isoalloxazine ring was seen in the dodecins (3.4 to 3.8 Å) and was also found in the case of FMN adenylyltransferase[33]. Apparently, this distance is optimal for π-stacking interactions although the relative orientation of the interacting aromatic ring systems differs in the above mentioned proteins. Because the main contribution for flavin binding stems from the π-stacking interaction with the two tryptophan residues we assume that the asymmetric isoalloxazine binds in two possible orientations in the symmetric binding site (two-fold axis) provided by the two protomers of the dimeric protein. The electron density of pdb entry 3cne is consistent with 50% occupancy of the two possible orientations of the tricyclic ring system.

ACKNOWLEDGEMENTS

REFERENCES

1. Comstock, L. E., and Coyne, M. J. (2003) *Bacteroides thetaiotaomicron*: a dynamic, niche-adapted human symbiont, *Bioessays* **25**, 926-929.

2. Mahowald, M. A., Rey, F. E., Seedorf, H., Turnbaugh, P. J., Fulton, R. S., Wollam, A., Shah, N., Wang, C., Magrini, V., Wilson, R. K., Cantarel, B. L., Coutinho, P. M., Henrissat, B., Crock, L. W., Russell, A., Verberkmoes, N. C., Hettich, R. L., and Gordon, J. I. (2009) Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla, *Proceedings of the National Academy of Sciences of the United States of America* **106**, 5859-5864.

3. Macheroux, P., Kappes, B., and Ealick, S. E. (2011) Flavogenomics - a genomic and structural view of flavin-dependent proteins, *The FEBS journal* **278**, 2625-2634.

4. Honbou, K., Suzuki, N. N., Horiuchi, M., Niki, T., Taira, T., Ariga, H., and Inagaki, F. (2003) The crystal structure of DJ-1, a protein related to male fertility and Parkinson's disease, *J. Biol. Chem.* **278**, 31380-31384.

5. Quigley, P. M., Korotkov, K., Baneyx, F., and Hol, W. G. J. (2003) The 1.6-A crystal structure of the class of chaperones represented by *Escherichia coli* Hsp31 reveals a putative catalyitc cleft, *Proceedings of the National Academy of Sciences of the United States of America* **100**, 3137-3142.

6. Tao, X., and Tong, L. (2003) Crystal structure of human DJ-1, a protein associated with early onset Parkinson's disease, *J. Biol. Chem.* **278**, 31372-31379.

7. Wei, Y., Ringe, D., Wilson, M. A., and Ondrechen, M. J. (2007) Identification of functional subclasses in the DJ-1 superfamily proteins, *PLoS computational biology* **3**, e10.

8. Malki, A., Caldas, T., Abdallah, J., Kern, R., Eckey, V., Kim, S. J., Cha, S. S., Mori, H., and Richarme, G. (2005) Peptidase activity of the *Escherichia coli* Hsp31 chaperone, *The Journal of biological chemistry* **280**, 14420-14426.

9. Le, H. T., Gautier, V., Kthiri, F., Malki, A., Messaoudi, N., Mihoub, M., Landoulsi, A., An, Y. J., Cha, S. S., and Richarme, G. YajL, prokaryotic homolog of parkinsonism-associated protein DJ-1, functions as a covalent chaperone for thiol proteome, *The Journal of biological chemistry* **287**, 5861-5870.

10. Mujacic, M., Bader, M. W., and Baneyx, F. (2004) *Escherichia coli* Hsp31 functions as a holding chaperone that cooperates with the DnaK-DnaJ-GrpE system in the management of protein misfolding under severe stress conditions, *Molecular microbiology* **51**, 849-859.

11. Parkin, G. (2004) Synthetic analogues relevant to the structure and function of zinc enzymes, *Chem. Rev.* **104**, 699-767.

12. Ghisla, S., and Massey, V. (1986) New flavins for old: artificial flavins as active site probes of flavoproteins, *Biochem. J.* **239**, 1-12.

13. Matsui, K., and Kasai, S. (1974) Effect of 8-amino-8-dimethyl-D-araboflavin, 8-amino-8-dimethyl-D-riboflavin and 8-hydroxy-8-demethyl-D-riboflavin on mice, *Journal of nutritional science and vitaminology* **20**, 411-420.

14. Spencer, R., Fisher, J., and Walsh, C. (1976) Preparation, Characterization, and Chemical Properties of the Flavin Coenzyme Analogues 5-Deazariboflavin, 5-Deazariboflavin 5´-Phosphate, and 5-Deazariboflavin 5´-Diphosphate, 5´-5´-Adenosine Ester, *Biochemistry* **15**, 1043-1053.

15. Massey, V., and Hemmerich, P. (1978) Photoreduction of flavoproteins and other biological compounds catalyzed by deazaflavins, *Biochemistry* **17**, 9-17.

16. Müller, F., Ghisla, S., and Bacher, A. (1988) Vitamin B2 und natürliche Flavine, in *Vitamine II* (Isler, O., Brubacher, G., Ghisla, S., Kräutler, B., Ed.), pp 50-159, Georg Thieme Verlag, Stuttgart, New York.

17. Whitby, L. G. (1954) Transglucosidation reactions with flavins, *Biochem. J.* **57**, 390-396.

18. Spencer, R., Fisher, J., and Walsh, C. (1977) One- and Two-Electron Redox Chemistry of 1-Carba-1-deazariboflavin, *Biochemistry* **16**, 3586-3594.

19. Fitzpatrick, P. F., Ghisla, S., and Massey, V. (1985) 8-Azidoflavins as photoaffinity labels for flavoproteins, *The Journal of biological chemistry* **260**, 8483-8491.

20. Macheroux, P. (1999) UV-visible spectroscopy as a tool to study flavoproteins, *Methods in molecular biology (Clifton, N.J* **131**, 1-7.

21. Choong, Y. S., and Massey, V. (1981) Studies on lactate oxidase substituted with synthetic flavins, iso-FMN lactate oxidase, *J. Biol. Chem.* **256**, 8671-8678.

22. Fagan, R. L., and Palfey, B. A. (2010) Flavin-dependent enzymes, in *Comprehensive Natural Products II* (Begley, T. P., Ed.), pp 37-114, Elsevier.

23. Joosten, V., and van Berkel, W. J. (2007) Flavoenzymes, *Current opinion in chemical biology* **11**, 195-202.

24. Mansoorabadi, S. O., Thibodeaux, C. J., and Liu, H. W. (2007) The diverse roles of flavin coenzymes--nature's most versatile thespians, *The Journal of organic chemistry* **72**, 6329-6342.

25. Bornemann, S. (2002) Flavoenzymes that catalyse reactions with no net redox change, *Natural product reports* **19**, 761-772.

26. McCormick, D. (1977) Interactions of flavins with amino acid residues: assessments from spectral and photochemical studies, *Photochemistry and photobiology* **26**, 169-182.

27. Kommoju, P.-R., Chen, Z.-w., Bruckner, R. C., Mathews, F. S., and Schuman Jorns, M. (2011) Probing oxygen activation sites in two flavoprotein oxidases using chloride as an oxygen surrogate, *Biochemistry* **50**, 5521-5534.

28. McDonald, C. A., Fagan, R. L., Collard, F., Monnier, V. M., and Palfey, B. A. (2011) Oxygen reactivity in flavoenzymes: context matters, *J. Am. Chem. Soc.* **133**, 16809-16811.

29. Meissner, B., Schleicher, E., Weber, S., and Essen, L. O. (2007) The dodecin from *Thermus thermophilus*, a bifunctional cofactor storage protein, *J. Biol. Chem.* **282**, 33142-33154.

30. Grininger, M., Seiler, F., Zeth, K., and Oesterhelt, D. (2006) Dodecin sequesters FAD in closed conformation from the aqueous solution, *J. Mol. Biol.* **364**, 561-566.

31. Grininger, M., Staudt, H., Johansson, P., Wachtveitl, J., and Oesterhelt, D. (2009) Dodecin is the key player in flavin homeostasis of archaea, *J. Biol. Chem.* **284**, 13068-13076.

32. Cairns, W., and Metzler, D. (1971) Photochemical degradation of flavins. VI. A new photoproduct and its use in studying the photolytic mechanism, *J. Am. Chem. Soc.* **93**, 2772-2777.

33. Huerta, C., Borek, D., Machius, M., Ghrishin, N. V., and Zhang, H. (2009) Structure and mechanism of a eukaryotic FMN adenylyltransferase, *J. Mol. Biol.* **389**, 388-400.

Chapter 13

Supplementary Information

Reverse Structural Genomics: An unusual flavin-binding site in a putative protease from *Bacteroides thetaiotaomicron*

Supplemental Data

Reverse Structural Genomics: An Unusual Flavin Binding Site in a Putative Protease from *Bacteroides thetaiotaomicron*.

Tanja Knaus, Elisabeth Eger, Julia Koop, Steve Stipsits, Cynthia L. Kinsland, Steven E. Ealick, and Peter Macheroux

Table of Contents

Supplementary Figure 1

Chemical structures of naturally occurring flavins (panel A) and chemically modified flavin analogs (panel B).

A

riboflavin

lumazine

lumichrome

FMN

FAD

B

1-deaza-riboflavin

5-deaza-riboflavin

iso-riboflavin

8-amino-riboflavin

2´-deoxy-riboflavin

iso-FMN
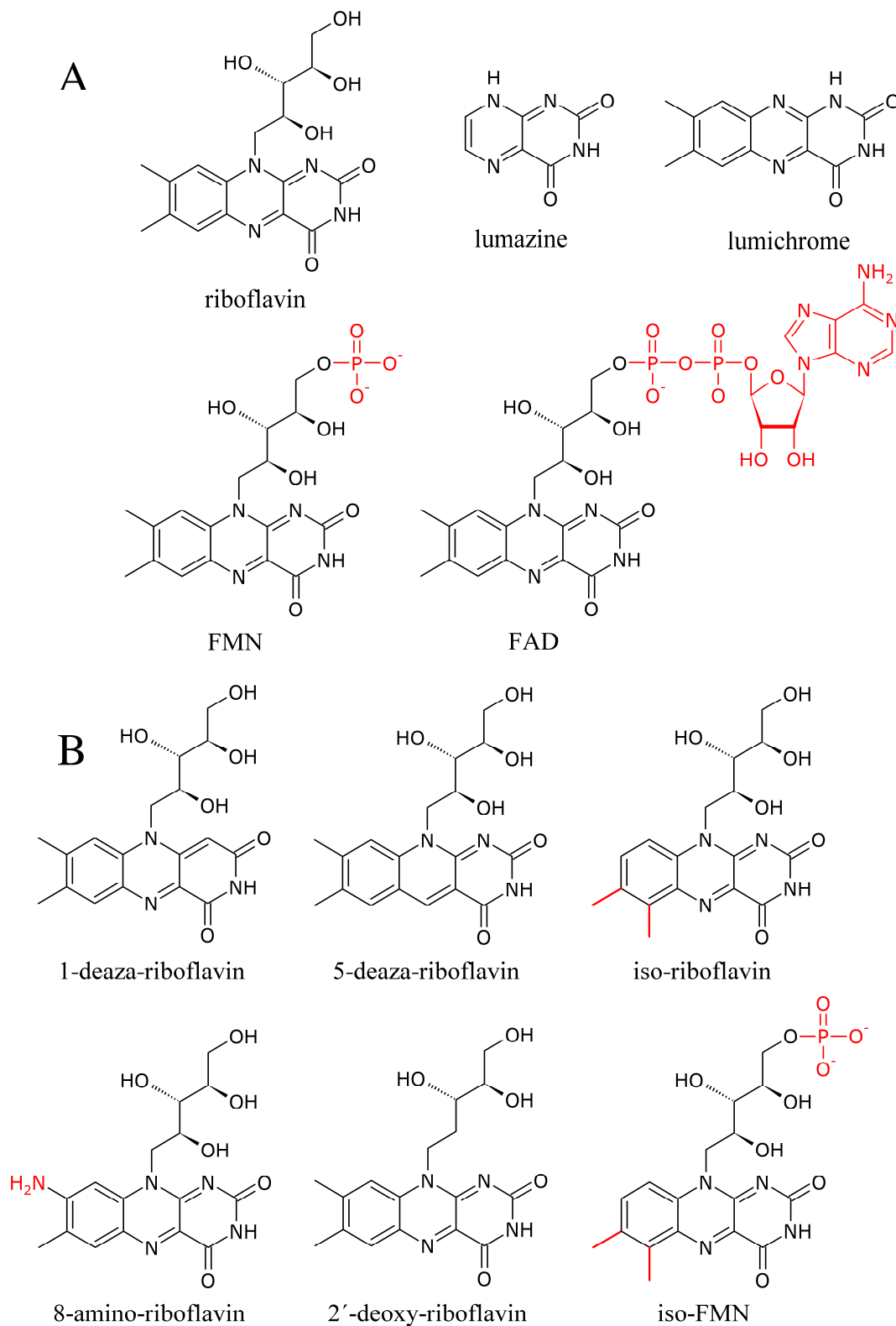
Supplementary Figure 2

SDS-PAGE of ppBat purification using Ni-NTA affinity chromatography.

Lane 1: molecular mass standard, lane 2: crude extract, lane 3: cell pellet after sonification and centrifugation, lane 4: supernatant, lane 5: flow through, lane 6: washing fraction, lane 7: pooled ppBat.
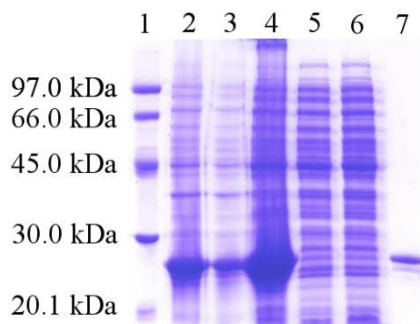
Supplementary Figure 3

UV/vis absorbance difference titration of 1-deaza-riboflavin (41 µM) with ppBat
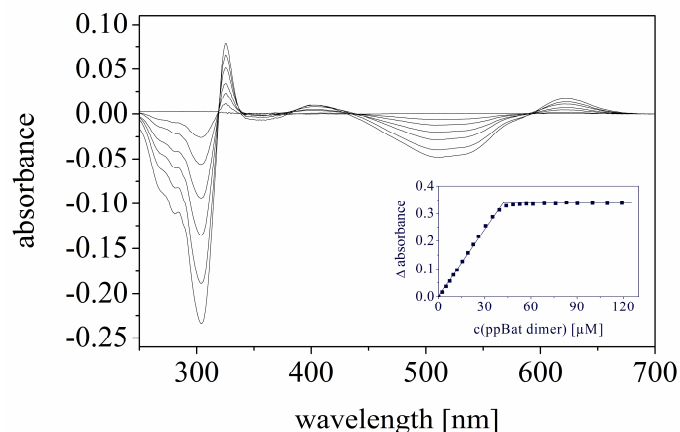
Supplementary Figure 4

UV/vis absorbance difference titration of 2'-deoxy-riboflavin (55 µM) with ppBat
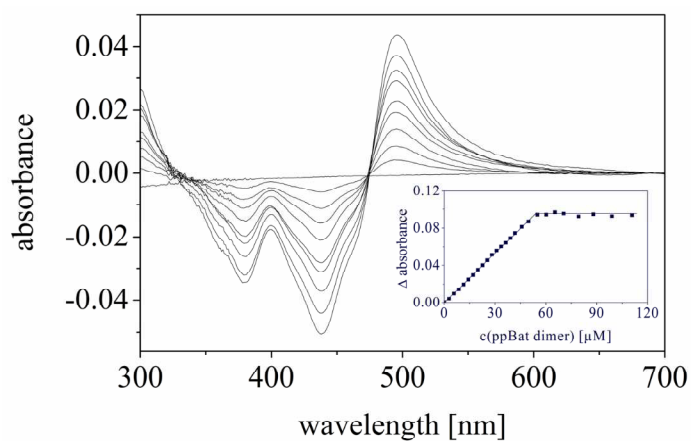
Supplementary Figure 5

ITC-determination of the dissociation constant for 2'-deoxy-riboflavin at 25 °C
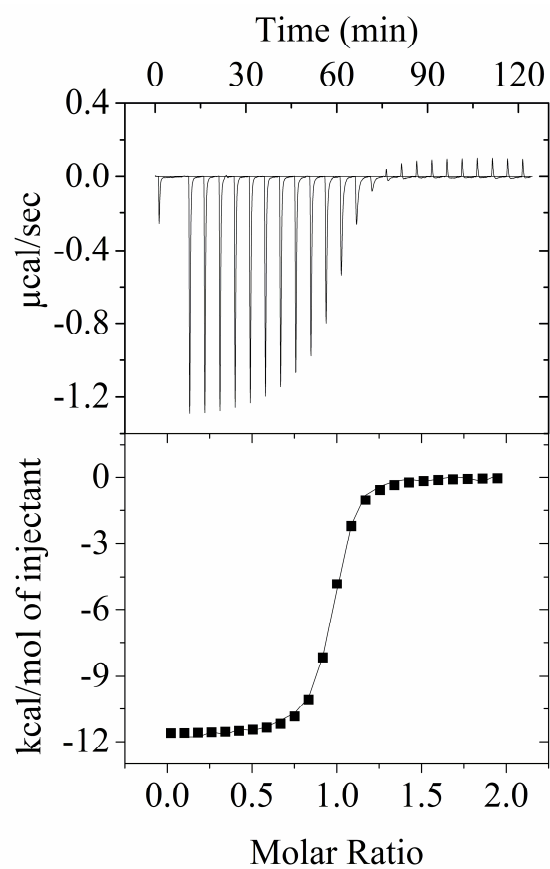
Supplementary Figure 6

Size exclusion chromatography of ppBat (solid line), ppBat with bound FMN (dashed line), Trp164 Phe (dotted line) and the Trp164Ala (dashed-dotted line) variant.
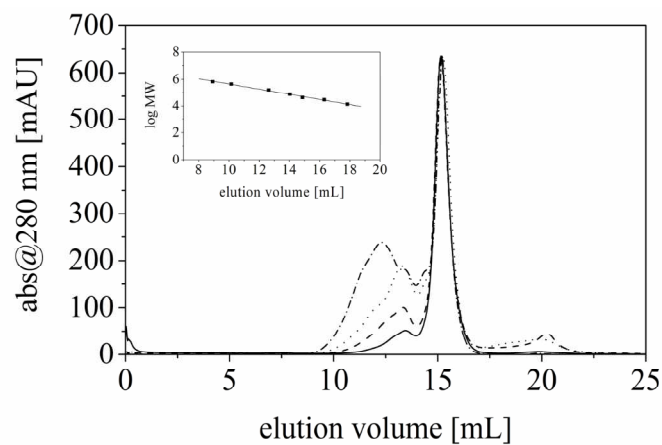
Chapter 14

Determination of free and bound riboflavin in Cow´s Milk using a novel flavin-binding protein

AUTHOR CONTRIBUTIONS

The manuscript has been submitted to Food Chemistry in September 2012 for reviewing. Based on the idea of PETER MACHEROUX, I designed the method to determine the flavin content in cow´s milk and incorporated a master student (JULIA KOOP) in this study. She performed most of the experiments under my direct supervision.

Determination of free and bound riboflavin in cow's milk using a novel flavin-binding protein

Julia Koop, Stefanie Monschein, E. Pauline Macheroux, Tanja Knaus* and Peter Macheroux*

Institute of Biochemistry, Graz University of Technology, A-8010 Graz, Austria

Running title: *Riboflavin in cow's milk*

To whom correspondence should be addressed: Peter Macheroux or Tanja Knaus, Graz University of Technology, Institute of Biochemistry, Petersgasse 12/II, A-8010 Graz, Austria; Tel.: +43-316-873 6450; Fax: +43-316-873 6952; E-mail: peter.macheroux@tugraz.at or tanja.knaus@tugraz.at

ABSTRACT

A recently described putative protease from the gut bacterium *Bacteroides thetaiotaomicron* (termed ppBat) exhibits two tryptophan residues in the interface which enable specific binding of the isoalloxazine heterocycle of riboflavin and its two cofactor forms, FMN and FAD. Recombinant ppBat was used to capture riboflavin from bovine milk directly without any prior preparation steps. The flavin-loaded protein was then re-isolated by means of affinity chromatography to identify and quantify the captured flavins. Free riboflavin concentrations were determined to 197 and 151 µg/L for milk with 3.5 and 0.5% fat content, respectively. Total riboflavin concentrations were also determined after acid-treatment of milk and were 4-5 times higher than for free riboflavin. Free FMN and FAD were not detectable and only trace amounts of FMN were found in milk following acid treatment. The method appears to be amenable to develop a direct assay for free riboflavin in milk and other foods.

INTRODUCTION

Milk is a very complex biological fluid generated by the mammary glands that serves as the sole nutrient for the newborn. It contains carbohydrates, lipids, proteins and micronutrients such as the vitamins. Among these, vitamin B_2 or riboflavin, is present in approximately up to 1 mg/L amounting to 70% of the daily recommended intake.[1] In fact, the vitamin was first isolated by Richard Kuhn in the early 1930s from 50.000 L of whey.[2] In recent years several methods based on HPLC analysis[3] or fluorometry[4] were established to analyse riboflavin and its two active coenzymes, FMN and FAD. The developed procedures suffer from several drawbacks such as the requirement for time consuming

precipitation and extraction steps and as a consequence the inability to distinguish between free and bound riboflavin and its derivatives. In addition, fluorometric quantification of riboflavin, FMN and FAD is difficult due to the vastly different fluorescence emission efficiency of these compounds [5]. In the study presented here we have chosen a different approach to determine the flavin content in cow's milk that is based on a recently discovered flavin-binding protein.[6] In contrast to riboflavin binding protein (RBP) from chicken egg white the recombinant protein from *Bacteroides thetaiotaomicron* (ppBat) binds riboflavin (K_D = 70 nM), FMN (K_D = 400 nM) and FAD (K_D = 800 nM) and hence can be used to scavenge all three naturally occurring forms of the vitamin. In addition, the recombinant protein is equipped with an N-terminal hexa-histidine tag enabling rapid separation, spectral characterisation and quantification of protein-bound flavins. Using this protein, we could quantify and distinguish free and bound flavins in cow's milk based on absorbance properties of the isolated flavin-protein complex.

EXPERIMENTAL PROCEDURES

Reagents & proteins

All chemicals were of the highest grade commercially available from Sigma-Aldrich (St. Louis, MO, U.S.A.), Fluka (Buchs, Switzerland) or Merck (Darmstadt, Germany). Ni-NTA agarose was from GE Healthcare, Bio-Sciences AB (Uppsala, Sweden) and Sephadex resin was from Pharmacia Biotech AB (Uppsala, Sweden). Cow's milk (3.5% and 0.5% fat content) was from *Stainzer Milch* cooperation, Stainz, Styria. Raw milk was obtained from a local farm.

The protein used to scavenge flavins was prepared by recombinant expression in *E. coli* host cells as described previously.[6] Purification of the protein was performed as follows: Resuspended cells were disrupted by 0.5 s sonication pulses for 10 min while cooling on ice. The cell debris was removed by centrifugation at 18,000 g for 30 min at 4 ºC.

The hexa-histidine tagged protein was purified by Ni-NTA affinity chromatography, loading the supernatant onto a Ni-NTA HisTrap[TM]HP column (GE Healthcare), previously equilibrated with lysis buffer. After loading of the filtrated lysate, the column was washed with ten column volumes of wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH 8.0) and bound protein was recovered with elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 150 mM imidazole, pH 8.0). The purity of the eluted fractions was determined by SDS/PAGE. Fractions containing ppBat were pooled, dialyzed against 20 mM Tris/HCl buffer containing 100 mM NaCl (pH 8.0) over night and concentrated using Centripreps (Millipore). The final concentration was determined at 280 nm using an ε_{280} of 13,075 M^{-1} cm^{-1}. The protein was flash-frozen and stored at -20 °C.

Extraction of flavins

15 ml of a buffer containing 50 nmol riboflavin, FMN and FAD was incubated with 560 nmol ppBat for 30 min. In a second experiment 30 nmol riboflavin and 60 nmol of each FMN and FAD were used. After incubation, ppBat was isolated by affinity chromatography and the pooled protein fractions were concentrated to 1 mL microfiltration (Centripreps). An UV/vis-absorbance spectrum was then recorded to determine the concentration of bound flavin. Heat denaturation at 95 ℃ for 10 min was employed to release bound flavins from ppBat followed by HPLC analysis.

Extraction of flavins from cow's milk was performed as follows: To 100 mL milk 360 nmol of ppBat were added and incubated for 30 min, then loaded onto a Ni-NTA resin (bed volume ca. 1 mL). The column was washed with ten column volumes of washing buffer and then bound ppBat was removed from the column by elution buffer (see above). The protein containing fractions were pooled and concentrated to 1 mL by microfiltration (Centripreps). Protein-bound flavins were analyzed by UV/vis-absorbance spectroscopy. In addition, the flow through was collected and precipitated by adding 5 mL 100% trifluoroacetic acid at 4 °C. The precipitate was removed by centrifugation (30 min at 18,000 rpm) and the supernatant was brought to pH 8 by slowly adding 1 M NaOH. 25 mL of the supernatant were diluted with an equal volume of lysis buffer and 460 nMol ppBat was added. Following 30 min incubation ppBat was isolated as described before. This two-step isolation of flavins was complemented by a single extraction experiment. In this case, 100 mL of milk was precipitated with 5 mL trifluoroacetic acid and the total flavin was determined using the isolation procedure described above.

UV/vis absorbance spectroscopy

UV/visible absorbance spectra were recorded with a Specord 210 spectrophotometer (Analytik Jena, Jena, Germany).

HPLC determination of flavins

Bound flavins were released by thermal precipitation of ppBat at 95 °C for 10 min, cooling on ice for 5 min and centrifugation for 5 min. The flavin extracts were loaded onto an Atlantis® dC18 8 µm 4.6 x 250 mm column (Waters) and eluted with a water/acetonitril multi-step gradient (0-1.5 min 0-5% acetonitrile; 1.5-20 min 5-60% acetonitrile; 20-22 min 95% acetonitrile) at a flow rate of 1 mL/min. The elution was monitored by UV-absorption at 370 nm and cofactors were identified according to their elution time and UV/vis absorbance spectrum.

RESULTS & DISCUSSION

The flavin-binding site of ppBat is unusual because it does not strictly distinguish flavins according to the N(10) side chain attached to the isoalloxazine ring system as is typical for flavoproteins. This is due to the formation of a sandwich complex where the isoalloxazine ring system is stacked by two indole side chains of tryptophan residues in the interface of the protein dimer. Figure 1 shows a cartoon representation of the re-determined crystal structure of ppBat in the presence of riboflavin (compare with pdb code 3cne). Details of the X-ray structure analysis will be reported elsewhere (manuscript in preparation).
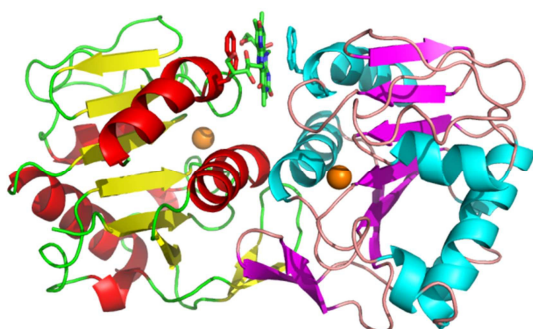
Figure 1. Structural representation of the flavin-binding protein from *Bacteroides thetaiotaomicron* (ppBat). The dimeric protein is shown in a cartoon representation with α-helices shown in red (left protomer) and blue (right protomer) and β-sheets in yellow (left protomer) and magenta (right protomer). The putative zinc atoms in each of the protomers are shown as orange spheres. Riboflavin (top center) is sandwiched between two tryptophans, one from each protomer, and is shown as a stick model. Riboflavin is bound in two orientations with the N(10)-side chain pointing either pointing into the plane or out of plane corresponding to the C2 axis of the dimer. The figure was prepared using the program PyMol.[7]

Except for the π-stacking interaction the crystal structure of the protein does not show any further binding contacts. This rationalizes the finding that all three naturally occurring

flavins bind to the protein. To proof that ppBat scavenges riboflavin, FMN and FAD, we incubated a cocktail containing these compounds with excess protein and then isolated the His-tagged protein by affinity chromatography. Bound flavins were released by heat denaturation and then analyzed by HPLC. As shown in Figure 2, all three natural flavins were recovered by ppBat, however the recovery of FAD was less than expected based on the ratio of flavins in the incubated samples (Table 1). This finding is likely to reflect the weaker binding of FAD (K_D = 800 nM) relative to FMN (K_D = 400 nM) and riboflavin (K_D = 74 nM).[6]
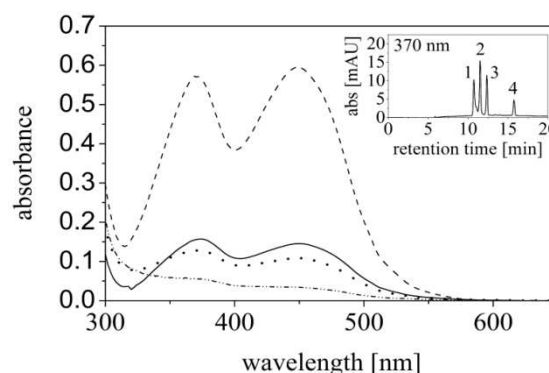
Figure 2. Extraction of riboflavin, FMN and FAD using ppBat. Shown are the UV/vis absorbance spectra of ppBat as isolated (dashed dotted dotted line), ppBat saturated with riboflavin (dashed line), ppBat incubated with 50 nmol riboflavin, FAD and FMN (solid line), and ppBat incubated with 30 nmol riboflavin, 60 nmol FAD and FMN. The inset shows the HPLC analysis of bound ligands from the first incubation mixture. The trace shown was recorded at 370 nm: peak 1 = FAD (retention time 10.7 min), peak 2 = FMN (retention time 11.6 min), peak 3 = riboflavin (retention time 12.4 min) and peak 4 = lumichrome (retention time 15.7 min).

Following this initial experiment, we employed ppBat to scavenge free flavins present in milk with different fat content (raw, 3.5% and 0.5%). In all cases, ppBat efficiently

trapped free flavins although the procedure was less amenable to raw milk since the high fat content apparently reduced the flow rate of the affinity column resulting in low protein recovery. Isolation of the flavin-loaded ppBat yielded a UV/vis absorbance spectrum with peaks at 370 and 445 nm previously reported for riboflavin bound to ppBat[6] (Figure 3).

Analysis of the bound flavins by HPLC showed that riboflavin is the dominant flavin species extracted with this method (Figure 3, inset). Using the molar absorption coefficient of bound riboflavin (10,900 M^{-1} cm^{-1}) the amount of free riboflavin was calculated to 197 and 151 µg/L for 3.5% and 0.5% milk, respectively.

Table 1. Binding of riboflavin, FMN and FAD by ppBat. The protein (560 nmol) was incubated with the indicated amount/concentration of Riboflavin (Rfl), FMN and FAD in 15 mL buffer, pH 8.0. Each experiment was performed in triplicate (A-C) and the relative peak areas were determined at 370 and 450 nm, the maxima of the UV/vis-absorbance spectrum.

Peak area (mAU * min)	Rfl	FMN (50 nmol)	FAD	Rfl (30 nmol)	FMN (60 nmol)	FAD
A						
370 nm	3.78	3.68	2.47	1.86	2.57	2.70
450 nm	4.65	4.52	3.01	2.37	3.75	3.84
B						
370 nm	3.95	3.69	2.40	1.71	2.36	2.06
450 nm	4.86	4.52	2.94	2.11	2.90	2.51
C						
370 nm	2.02	2.70	2.30	3.61	3.32	2.20
450 nm	2.48	3.31	2.81	4.43	4.08	2.69

Since this is considerably lower than previously reported values[3], we reasoned that most of the flavin is not available for binding to ppBat as it is present in a bound form. In order to evaluate this assumption, we collected the flow-through from the affinity column and denatured the milk proteins by acid precipitation. After adjustment of the pH to 8.0, the supernatant was again incubated with ppBat to bind any released flavins. The scavenger protein was then isolated and the flavin content determined as described before. In a parallel experiment, milk was first treated with acid and then incubated with ppBat after pH adjustment to determine the total amount of flavin (in 100 mL). The UV/Vis absorbance spectra of the flavins scavenged by ppBat are shown in Figure 3. The result clearly confirms our assumption that the fraction of free flavin amounts to 20-25% of the total flavin and that the remainder is present in a bound form and therefore not scavenged by ppBat. HPLC analysis of the total flavin demonstrates that most of the total flavin is found as riboflavin with only minor traces present as FMN (see inset of Figure 3).
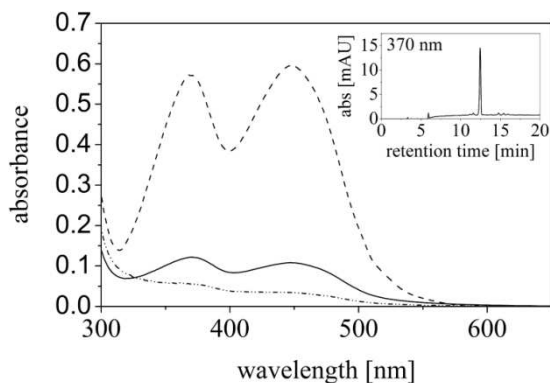
Figure 3. Extraction and isolation of flavins from cow's milk. Panel A shows a typical spectrum of ppBat after incubation with milk and subsequent purification by affinity chromatography (free flavins isolated from 25 mL 3.5 % milk. Shown are the UV/vis absorbance spectra of ppBat as isolated (dashed-dotted-dotted line), ppBat saturated with riboflavin (dashed line) and ppBat incubated with 25 mL of 3.5% fat bovine milk (solid line). After incubation ppBat was isolated by affinity purification as detailed in materials and methods. The inset shows the HPLC analysis of bound ligands. The trace was recorded at 370 nm and shows one main peak at the expected retention time for riboflavin (12.5 min).

Figure 4. Analysis of total flavin content of cow's milk. Shown are the UV/vis absorbance spectra of ppBat as isolated (dashed-dotted-dotted line), ppBat saturated with riboflavin (dashed line) and ppBat incubated with 25 mL of 0.5% fat bovine milk: free riboflavin in milk (solid line), protein-bound riboflavin in milk (dashed-dotted line) and total amount of riboflavin in milk (dotted line). The inset shows the HPLC analysis of flavins released from ppBat as described in materials and methods. The trace was recorded at 370 nm: peak 1 = FMN (retention time 11.6 min), peak 2 = riboflavin (12.4 min) and peak 3 = unidentified compound (14.6 min).

As summarized in Table 2, the amount of flavin released by acid treatment exceeds the amount of free flavin by a factor of 4-5 resulting in an overall flavin content of ca. 800 µg/L which is close to the reported flavin concentration of cow's milk [3]. Similar flavin concentrations were obtained when milk was denatured first and the total flavin content was determined by means of binding to ppBat (last line in Table 2). Interestingly, the flavin composition of the free and acid-released fraction slightly varies in that the latter also contains FMN, which is basically absent in the free flavin fraction (Figure 3 and 4, compare insets). It appears that charged flavin species such as FMN and FAD are readily bound by milk components, such as the calcium-rich micelles, and therefore not available for binding to ppBat.

To further evaluate this assumption, we added 300 µg/L FMN and FAD to a milk sample prior to analysis with our method. Surprisingly, none of the extra flavins were captured by ppBat indicating that the flavins were sequestered by milk components preventing binding to ppBat.

Previous studies have shown that the distribution of flavins strongly varies with the milk fraction. Kanno *et al.* found the largest percentage of flavins in whey (67%) followed by the casein fraction (18%), skim milk membrane (9%) and cream (8%).[8] In addition, the ratio of free to bound flavin depends on the fraction with whey containing 92.4% free flavin while 72.4% of flavins found in the cream are in a bound form.[8] The association of flavins is also complex as some flavins in their coenzyme forms FMN and FAD are present in milk redox enzymes such as the FAD-

dependent enzymes xanthine oxidase and sulfhydryl oxidase.[9-11] However, the majority of bound flavins appear to be unspecifically associated with proteins, in particular riboflavin as it is not present in enzymes, which have either FMN or FAD specifically bound to their active sites.[12] Taking the complexity of flavin distribution and association into account our results can be readily rationalized.

Table 2. Determination of free and bound flavin in cow's milk

Sample	3.5% milk	0.5% milk
	μg/L of total riboflavin	
free	197 ± 74	151 ± 22
bound	625 ± 72	619 ± 132
Σ free + bound	822 ± 14	770 ± 116
total	623 ± 60	813 ± 61

Since we aimed to demonstrate that ppBat scavenges flavins directly from milk without any preparatory steps such as centrifugation, precipitation or extraction we could clearly identify and quantify free riboflavin in bovine milk. The lower amount of riboflavin found is due to the fact that most (ca. 75%) of flavins (mostly riboflavin) is associated with protein components of the milk fat globule membranes.[8] Release of flavins from these proteins by acid treatment enabled the determination of the ratio of free to bound riboflavin.

Although ppBat has a lower affinity to FMN and FAD our pilot experiments have clearly shown that these two forms are also scavenged from a flavin-containing cocktail. However, we could only detect very small traces of FMN and FAD in acid-treated milk (see Figure 4). Interestingly, even the addition of FMN and FAD to milk samples did not lead to recovery of the added extra flavins by ppBat suggesting that they were bound by protein components of the milk, *e.g.* the milk fat globule membrane fraction and hence are not available for the ppBat flavin binding site.

In this study we have shown that ppBat is capable of extracting free flavins, most specifically riboflavin, from milk without any prior treatment of the sample. In our view, ppBat could be used to rapidly detect riboflavin in milk and other food samples by fluorometry, for example by immobilization of ppBat (e.g. by its hexa-histidine tag) and measurement of the flavin fluorescence emission as a function of sample concentration.

FOOTNOTES

To whom correspondence may be addressed: Peter Macheroux or Tanja Knaus, Graz University of Technology, Institute of Biochemistry, Petersgasse 12, A-8010 Graz, Austria; Tel.: +43-316-873 6450; Fax: +43-316-873 6952; E-mail: peter.macheroux@tugraz.at or tanja.knaus@tugraz.at

The abbreviations used are: FMN, flavin mononucleotide; FAD, flavin adenine dinucleotide; HPLC, high performance liquid chromatography; ppBat, putative protease from *Bacteroides thetaiotaomicron*;

REFERENCES

1. Board, F. N. (1989) Recommended Dietary Allowances., Vol. **10th edition**, National Academy Press, Washington DC.

2. Kuhn, R., György, P., and Wagner-Jauregg, T. (1933) Über Lactoflavin, den Farbstoff der Molke, *Ber. Dtsch. Chem. Ges.* **66**, 1034-1038.

3. Roughead, Z. K., and McCormick, D. (1990) Qualitative and quantitative assessment of flavins in cow's milk, *J. Nutr.* **120**, 382-388.

4. Zandomeneghi, M., Carbonaro, L., and Zandomeneghi, G. (2007) Biochemical fluorometric method for the determination of riboflavin in milk, *J. Agric. Food Chem.* **55**, 5990-5994.

5. Harvey, R. A. (1980) Flavin 1, N6-ethaneoadenine dinucleotide, *Methods in enzymology* **66**, 290-294.

6. Knaus, T., Eger, E., Koop, J., Stipsits, S., Kinsland, C. L., Ealick, S. E., and Macheroux, P. (2012) Reverse structural genomics: An unusual flavin binding site in a putative protease from Bacteroides thetaiotaomicron, *J. Biol. Chem.* **287**, 27490-27498.

7. Delano, W. L. (2002) The PyMOL molecular graphics system, DeLano Scientific, San Carlos, CA.

8. Kanno, C., Kanehara, N., Shirafuji, K., Tanji, R., and Imai, T. (1991) Binding form of vitamin B2 in bovine milk: its concentration, distribution and binding linkage, *J. Nutr. Sci. Vitaminol.* **37**, 15-27.

9. Briley, M. S., and Eisenthal, R. (1975) Association of xanthine oxidase with the bovine milk-fat-globule membrane, *Biochem. J.* **147**, 417-423.

10. Jaje, J., Wolcott, H. N., Fadugba, O., Cripps, D., Yang, A. J., Mather, I. H., and Thorpe, C. (2007) A flavin-dependent sulfhydryl oxidase in bovine milk, *Biochemistry* **46**, 13031-13040.

11. Jarasch, E.-D., Bruder, G., Keenan, T. W., and Franke, W. W. (1977) Redox constituents in milk fat globule membranes and rough endoplasmic reticulum from lactating mammary gland, *J. Cell Biol.* **73**, 223-241.

12. Macheroux, P., Kappes, B., and Ealick, S. E. (2011) Flavogenomics - a genomic and structural view of flavin-dependent proteins, *The FEBS journal* **278**, 2625-2634.

Chapter 15

Reverse Structural Genomics II: Stability of a putative protease from *Bacteroides thetaiotaomicron*

15.1   Introduction

Additionally to previous studies on the flavin binding site in ppBat from *Bacteroides thetaiotaomicron*[1], we were further interested in the complete biochemical characterization of the flavin binding site in this novel putative protease. In ppBat, the flavin moiety is sandwiched between the dimeric interface of the protein by the indole rings of two tryptophan residues (Trp164) almost equidistant. Previous studies showed that the flavin moiety is not responsible for a dimerization of the protein since the Trp164Ala variant, which was not able to bind a flavin cofactor anymore, still occurred as a dimer in solution. In the Trp164Phe variant the dissociation constant between riboflavin and the enzyme was 16 fold increased in comparison to the wild type protein, indicating, that the tryptophan residues are crucial for a tight binding of the isoalloxazine-moiety of naturally occurring flavins as well as several chemically modified flavin analogues in ppBat.[1]

Moreover the function of the zinc ion regarding a possible catalytic or structural role in the protein has not been investigated. The metal binding site in the protomer is assembled by two cysteine residues (Cys74 and Cys111, respectively) and two water molecules, coordinating the zinc ion in a way which does not occur in any known protein. Single and double mutants were prepared by exchanging the hydrophilic cysteine residues to hydrophobic alanine residues, which might not be able to coordinate the zinc ion anymore.

During this study of the flavin binding site in ppBat the redox potentials of oxidized/reduced FMN (enzyme bound and free in solution) have been determined. From these data the binding affinities between ppBat wild type and its variants and the reduced cofactors have been calculated, starting from the dissociation constants between the enzyme and the oxidized flavin. For the identification of a potential structure stabilizing role of the zinc ion in the enzyme, temperature stability studies of ppBat wild type and variants were investigated using circular dichrism (CD) measurements and thermofluor experiments. Measurements were additionally performed in the presence of ethylene diamine tetra acetic acid (EDTA) which acts as a chelating agent and forms complexes with metal ions.

15.2   Experimental Procedures

Site-Directed Mutagenesis: Modification of the zinc binding site was achieved by substitution of the cysteine residues to alanine residues. Therefore the single variants Cys74Ala and Cys111Ala as well as the double mutant Cys74Ala Cys111Ala were prepared by STEFANIE MONSCHEIN during her master thesis.

Expression and Purification: The expression and purification of ppBat and its variants was performed as described by Knaus *et al.*[1]

Zinc Determination: The amount of zinc in ppBat variants was determined by inductively coupled plasma optical emission spectrometry (ICP-OES) at the Institute of Analytical Chemistry and Food Chemistry, Graz University of Technology by Dipl.-Ing. Dr. techn. HELMAR WILTSCHE.

Determination of the Dissociation Constants between Oxidized Flavin Species and ppBat Variants: Dissociation constants for the binding of oxidized flavins to ppBat variants were determined using a VP-ITC system (MicroCal) as described previously.[1] All experiments were performed at 25 °C in 20 mM Tris/HCl, 100 mM NaCl, pH 8.0 buffer and solutions were degassed immediately before measurements. Titration experiments for riboflavin and FMN consisted of 24 injections (7 µL, duration time 14 sec, spacing time 300 sec) of a flavin solution (480 µM riboflavin and 530 µM FMN, respectively) to 1.495 mL of the Cys74Ala, Cys111Ala and Cys74Ala Cys111Ala variants (40 µM). 40 µM of ppBat Trp164Phe were titrated with 7 µL aliquots of 530 µM FMN for 40 times (duration time 7 sec, spacing time 300 sec). One set of sites fitting with Origin version 7.0 (MicroCal) for ITC data analysis was used to obtain dissociation constants.

Redox Potential: The redox potential (E_0) of oxidized/reduced FMN free in solution and protein bound was determined by the dye-equilibration method using the xanthine/xanthine oxidase electron delivering system as described in a previous study.[2] Concentrations of redox dye and protein were chosen in a way that the absorbance maxima were roughly in the same range. Reactions were carried out anoxic in 20 mM Tris/HCl, 100 mM NaCl buffer at pH 8.0 with a Hi-Tech (SF-61DX2) stopped-flow device (TgK Scientific Limited, Bradford-on-Avon, UK) positioned in a glove box from Belle Technology (Weymouth, UK) at 25 °C. The simultaneously reduction process of

FMN (free in solution and bound to ppBat and its variants, respectively) and the redox dye, was measured by monitoring changes in the absorbance with a KinetaScanT diode array detector (MG-6560, TgK Scientific Limited). The reactions were started by rapidly mixing of a solution containing xanthine (300 µM), benzyl viologen (5 µM, mediator) and the appropriate amount of enzyme and a mixture of xanthine oxidase (from bovine milk, Grade III, SIGMA Aldrich, approximately 2 nM) and redox dye (phenosafranin, E_0 = -252 mV). Redox potentials were calculated from plots of log([ox]/[red]) of the enzyme versus log([ox]/[red]) of the redox dye according to Minnaert.[3]

Determination of the Dissociation Constants between Reduced Flavin and Protein: From the determined redox potentials between $FMN/FMNH_2$ and the dissociation constants between the oxidized flavin and ppBat, the dissociation constants between the reduced form of the cofactor and the enzymes could be calculated according to the thermodynamic relationship which is described in Scheme 1.[4]

$$\Delta G_4 = R * T * \ln K_B$$
$$K_{D,\ oxidized}$$

FMN$_{free,\ oxidized}$ $\longrightarrow$ FMN$_{protein\ bound,\ oxidized}$

$$\Delta G_1 = z * F * E_{0,\ FMN/FMNH2,\ free}$$
$$\Delta G_3 = z * F * E_{0,\ FMN/FMNH2,\ protein\ bound}$$

FMN$_{free,\ reduced}$ $\longrightarrow$ FMN$_{protein\ bound,\ reduced}$

$$K_{D,\ reduced}$$
$$\Delta G_2 = R * T * \ln K_B$$
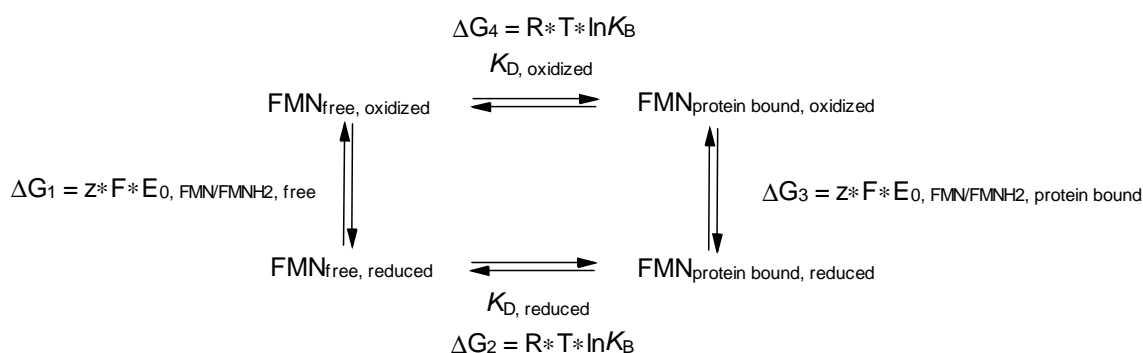
Scheme 1. Thermodynamic relationship between the redox potentials and the dissociation constants (adapted from reference 4).

The abbreviations used in Scheme 1 are explained in the table below:

Table 1. Explanation of abbreviations used in Scheme 1.

abbreviation	name	unit
K_D	dissociation constant	[M]
ΔG	Gibbs energy	[J·mol^{-1}]
R	gas constant	8.314472 [J·mol^{-1}·K^{-1}]
T	temperature	[K]
K_B	binding constant	[M^{-1}][*]
z	number of electron transferred simultaneously	-
F	Faraday constant	96,485.34 [J·mol^{-1}·V^{-1}]

[*] $K_B = K_D^{-1}$

The Gibbs energy for the binding affinity between the reduced cofactor and the enzyme can be calculated according to the correlation as described in Equation 1:

Equation 1 $\quad\quad\quad\quad\quad\quad \Delta G_1 + \Delta G_2 = \Delta G_3 + \Delta G_4$

Temperature Stability of ppBat and Variants: The temperature stability of the enzymes has been studied using two different kinds of methods: (i) *CD-Measurements* were carried out in quartz cells of 0.02 cm thickness on a JASCO J-715 spectropolarimeter (Essex, UK). Approximately 1 mg/mL of protein (dissolved in 20 mM Tris/HOAc buffer, pH 8.0) was heated up from 25 °C to 95 °C (heat rate 1 °C·min^{-1}, regulated with a Julabo F25 water bath – Seelbach, Germany) and changes in the secondary structures of the proteins according to thermal denaturation were monitored at a wavelength of 210 nm. Additionally CD spectra were recorded between 190 and 260 nm every 5 °C of heating. Spectra were corrected for the buffer background and converted to molar ellipticity. (ii) *Thermofluor*[®] *and ThermoFAD Studies:* The unfolding of ppBat and variants due to increasing temperatures was further determined by a monitoring of the fluorescence using a solvatochromic dye (*Thermofluor*[®]) and the intrinsic fluorescence of the protein bound flavin cofactor (*ThermoFAD*) as described previously.[5,6] In the case of *Thermofluor*[®] measurements, SYPRO Orange was used as fluorescence indicator. An increasing fluorescence signal can be measured, due to the binding of the dye on hydrophobic parts of the protein which are released to the surface when the protein gets unfolded. *ThermoFAD* measures the increase of the intrinsic flavin fluorescence when the cofactor is released from the protein and its fluorescence is therefore not quenched by the protein environment anymore. Measurements were carried out in 20 mM Tris/HCl, 100 mM NaCl buffer, pH 8.0 with a protein concentration of 1 mg/mL using a reaction volume of 25 µL. In the case of *Thermofluor*[®] experiments additionally 5 µL SYPRO orange dye (Sigma Aldrich, diluted 1:100) were added. Additional experiments were performed by the addition of EDTA (5-fold molar excess) into the protein solutions. Samples were pre-heated to 25 °C for 60 sec, before the temperature was raised to 95 °C in 0.5 °C every 30 seconds. The fluorescence emission was determined using a SYBR Green fluorescence emission filter for intrinsic flavin fluorescence and a HEX fluorescence emission filter for the SYPRO orange dye.

15.3   Results

Zinc Content: In contrast to ppBat wild type where stoichiometric amounts of zinc have been determined by ICP-OES measurements[1], only traces of zinc have been found in the double mutant Cys74Ala Cys111Ala as well as in single mutants Cys74Ala and Cys111Ala, respectively.

Redox Potential: The redox potentials of ppBat and variants have been determined in order to further calculate the dissociation constants between the protein and reduced FMN. The simultaneous reduction process for ppBat wild type saturated with FMN using phenosafranin as redox dye is shown in Figure 1. In the inset the Nernst-plot used for the calculation of the redox potential is displayed. The slope of approximately one shows that a simultaneous two-electron-transfer occurs in this system.
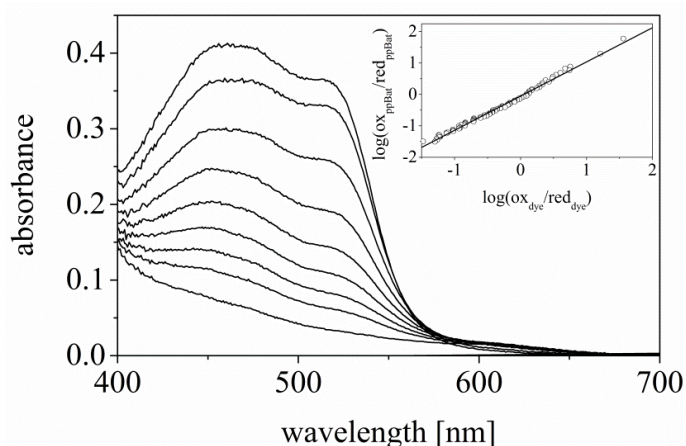
Figure 1. Representative example for the determination of a redoxpotential. The measurement shows the simultaneous reduction of ppBat wild type with bound FMN and the phenosafranin dye using the xanthine/xanthine oxidase system. Selected spectra of the course of reduction are displayed. In the inset the Nernst plot, which was used for calculating the redox potential of the enzyme-bound FMN cofactor is shown. Data points for evaluation were measured at 414 and 520 nm for flavin and dye, respectively, where no significant influence of the other chromophore could be observed.

The redox potentials calculated for ppBat and variants are summarized in Table 2. Hereby it should be pointed out that the redox potentials for the wild type enzyme and the cysteine variants are almost in the same range and approximately 30 mV different from the redox potential of FMN free in solution. However, FMN bound to the Trp164Phe variant shows a similar redox potential to free in solution.

Table 2. Calculated redox potentials

enzyme	redox potential E_0 [mV]
FMN_{free}	-222
ppBat wild type (FMN)	-252
ppBat Trp164Phe (FMN)	-229
ppBat Cys74Ala (FMN)	-256
ppBat Cys74Ala Cys111Ala (FMN)	-247

Dissociation Constants for ppBat and Variants: The dissociation constants for the wild type enzyme and the variants were determined by isothermal titration calorimetry. A typical result from a titration experiment is depicted in Figure 2 which shows the titration of the Cys74Ala variant with FMN.
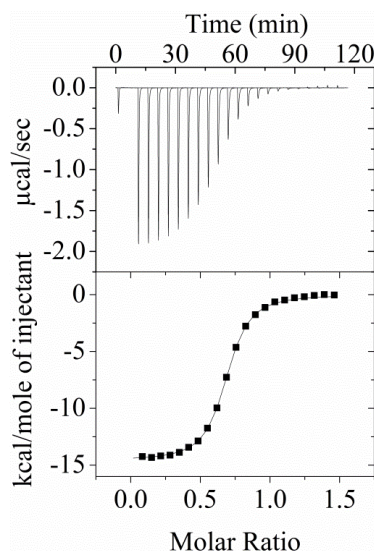
Figure 2. Representative example for the determination of a dissociation constant by isothermal titration calorimetry. The measurement shows the titration of the ppBat Cys74Ala variant (40 µM) with FMN in 20 mM Tris/HCl, 100 mM NaCl buffer, pH 8.0 at 25 °C.

From the analysis of the dissociation constants reported in Table 3, one can deduce that the zinc ion does not affect the affinity of the enzyme with the flavin cofactors as expected. In contrast the Trp164 is fundamental for the binding of the flavin moiety since dissociation constants are approximately 15 fold increased. Taking into account the redox potential reported in Table 2 and the dissociation constants for the oxidized FMN reported in Table 3, we could determine the dissociation constant for the reduced forms of FMN. Interestingly, the values of the $K_{D, reduced}$ increased roughly 10 fold for the wild type and cysteine variants, whereas only 1.7 fold for the Trp164Phe variant.

Table 3. Dissociation constants between oxidized/reduced flavin and ppBat and variants.

	K_D oxidized [nM]		K_D reduced [nM]
enzyme	riboflavin	**FMN**	**FMN**
ppBat wild type	$74 \pm 5^*$	$487 \pm 19^*$	5030
ppBat Trp164Phe	$1190 \pm 150^*$	8460 ± 400	14600
ppBat Cys74Ala	87 ± 5	435 ± 17	6140
ppBat Cys111Ala	91 ± 10	445 ± 15	n.d.□
ppBat Cys74Ala Cys111Ala	121 ± 11	488 ± 40	3414

* data are already published[1] and shown for comparison

□ not determined

Temperature Stability Studies: For the determination of the melting point of ppBat, initially CD-measurements were performed for the wild type enzyme and the Trp164Phe variant as isolated and saturated with FMN, respectively, and for the Trp164Ala variant. Figure 3A shows the changes of the molar ellipticity at 210 nm at an increasing temperature for all performed experiments. Interestingly, none of the proteins investigated shows a complete unfolding of their secondary structures. In Figure 3B selected CD-spectra recorded from 190 to 260 nm during the up-scan (25 to 95 °C) and down-scan (95 to 25 °C) for the wild type enzyme are shown. At approximately 75 °C (green line) the conformation of ppBat undergoes structural changes in the secondary structure, however, a complete disorganization due to melting of the enzyme cannot be observed. These results were obtained for all proteins measured.
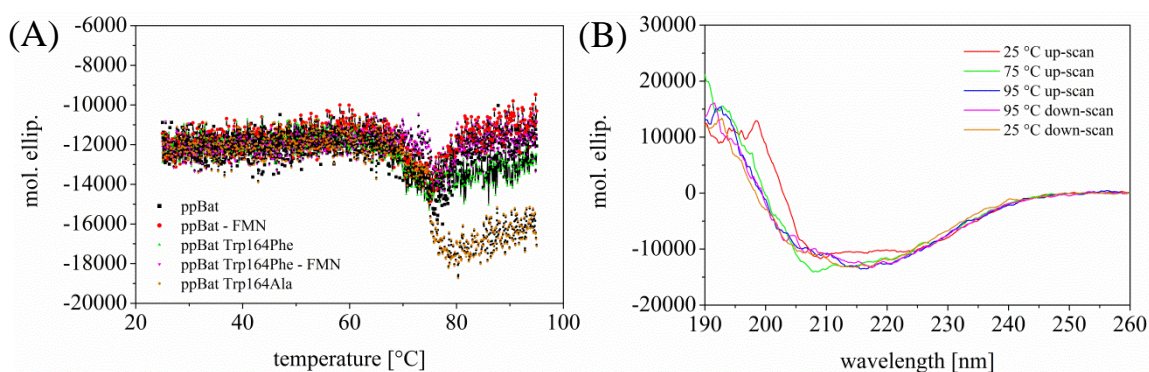
Figure 3. CD-measurements. (A) Changes in the molar ellipticity recorded at 210 nm at increasing temperatures for all enzymes tested. (B) Selected CD-spectra for ppBat wild type recorded during up-scan (25 to 95 °C) and down-scan (95 to 25 °C).

As the CD-measurements did not show clear results regarding the behavior of ppBat at increasing temperatures, *Thermofluor*® *and ThermoFAD Studies* were carried out. In these experiments the variants which showed modifications in the zinc binding have been investigated as well. The "unfolding" curves for ppBat and its variants using *Thermofluor*® measurements without EDTA and in the presence of the chelating agent are shown in Figure 3A and Figure 3B, respectively. It is clearly visible that the cysteine to alanine variants, which are not able to bind a zinc ion anymore, show significantly decreases in their rearrangement temperatures in comparison to the wild type enzyme and the tryptophan variants (Figure 3A). Moreover, in the presence of EDTA, only the rearrangement temperatures of ppBat wild type and Trp164 variants were decreased for almost 12 °C to the same temperature range as was observed for the cysteine to alanine variants. The determined rearrangement temperatures for all measurements prepared are listed in Table 4.
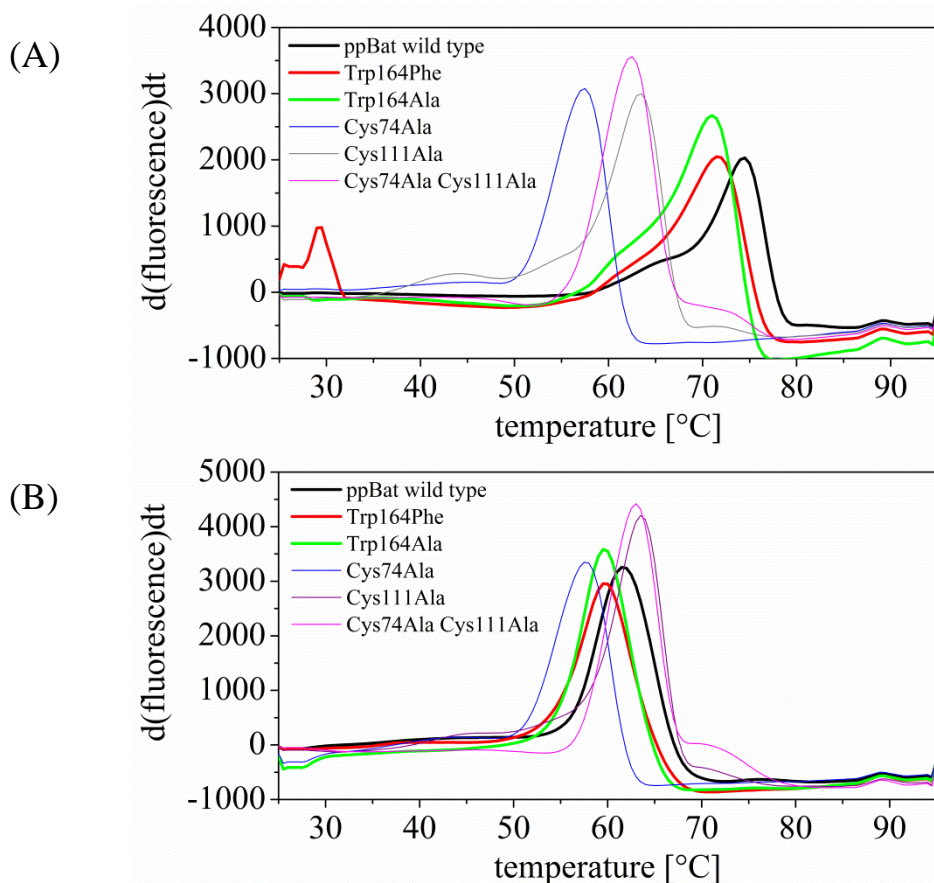
Figure 4. Rearrangement curves for ppBat and its variants using *Thermofluor*® studies. (A) Temperature studies on ppBat and variants using the SYPRO orange fluorescence dye. (B) Temperature studies on ppBat and variants using the SYPRO orange fluorescence dye in the presence of the chelating agent EDTA.

Measurements were carried out for every enzyme (i) as isolated, (ii) saturated with riboflavin and (iii) saturated with FMN. Interestingly, the proteins saturated with riboflavin showed a slightly increased rearrangement temperature in comparison to the isolated form or the protein saturated with FMN. Moreover, the Trp164Phe variant, which shows a significantly higher dissociation constant for the binding of flavins in comparison to the wild type enzyme, shows a decreased rearrangement temperature (approximately 2.5 to 3 °C). These results indicate, that the binding of a flavin moiety between the tryptophan residues of the dimeric protein stabilizes the structure in a slightly way.

Table 4. Summary of *Thermofluor*® and *ThermoFAD* studies.

	$T_{rearrangement}$ [°C] flavin fluorescence	$T_{rearrangement}$ [°C] SYPRO orange	$T_{rearrangement}$ [°C] flavin fluorescence EDTA	$T_{rearrangement}$ [°C] SYPRO orange EDTA
ppBat	**73.5**	**72.0**	**61.0**	**60.0**
ppBat-RF	**75.5**	**74.5**	**62.0**	**61.5**
ppBat-FMN	**74.5**	**73.5**	**60.5**	**61.0**
Trp164Phe	-	**70.5**	**-**	**59.0**
Trp164Phe-RF	-	**71.5**	**-**	**59.5**
Trp164Phe-FMN	-	**71.0**	**-**	**59.5**
Trp164Ala		**71.0**	**-**	**59.5**
Cys74Ala	56.0	55.5	56.5	55.5
Cys74Ala-RF	58.0	57.5	56.5	57.5
Cys74Ala-FMN	56.5	56.5	55.0	57.0
Cys111Ala	63.0	61.0	63.5	61.0
Cys111Ala-RF	64.5	63.5	64.0	63.5
Cys111Ala-FMN	63.5	62.5	63.5	63.0
Cys74Ala Cys111Ala	62.5	60.0	63.0	60.5
Cys74Ala Cys111Ala-RF	64.0	62.5	64.0	63.0
Cys74Ala Cys111Ala-FMN	63.5	62.0	63.0	62.0

15.4   Discussion

ICP-OES studies showed that neither the cysteine to alanine single variants nor the double variant were able to bind a zinc ion anymore. That means that both cysteine residues are crucial for the binding of the metal ion in ppBat.

The redox potentials of FMN bound to ppBat wild type or the Cys74Ala variant or the Cys74Ala Cys111Ala variant are nearly the same. In contrast both free FMN and FMN bound to Trp164Phe variant possess a 30 mV higher redox potential. Additionally, from the analysis of the dissociation constants reported in Table 3, one can deduce that the zinc

ion does not affect the affinity of the enzyme with the flavin cofactors. In fact, the Cys74Ala and Cys111Ala variants are not able to bind zinc ion anymore. Nevertheless, the dissociation constants for riboflavin and FMN are essentially the same for both variants in comparison to the wild type. The situation is completely different in the case of the Trp164Phe variant, whereby a 15 fold increased binding constant was measured. The exchange of the Trp164 residue to phenylalanine also caused the remarkable increase of the redox potential (ca. 30 mV) compared to the wild type and the other variants, demonstrating that this residue is highly important for the binding of the cofactor and contributes to increase the electron density on the isoalloxazine moiety due to π-stacking.

From CD-measurements, we were unable to infer the melting point of ppBat and variants since the typical sigmoidal profile for the molar ellipticity was not observed. Although the CD-spectra recorded in the near UV range did not show a significant difference while raising the temperature. These preliminary observations already pointed out the extreme thermal stability of the wild type enzyme and the variants. Nevertheless, *Thermofluor®* *and ThermoFAD* studies were undertaken in order to evaluate whether the mutations could influence the stability of the enzyme.

In Figure 4A and B, the maximum of fluorescence for the SYPRO orange dye represents the temperature where the protein undergoes unfolding or significant rearrangement of the secondary structure. In this respect, the replacements of the Trp164 residue to phenylalanine or alanine provoked a slightly decrease of the rearrangement temperature in comparison to the wild type enzyme. In fact, the Trp164 is involved in the binding of the flavin cofactor to the enzyme; thus replacing this residue reduces the ability of the protein to bind the cofactor with high affinity leading to a moderate destabilization of the dimeric structure. The stabilizing effect of the flavin cofactor bound to the protein is clearly shown in Table 4 whereby, for all the variants tested, the rearrangement temperature was always higher when the protein was saturated with riboflavin in comparison to protein as isolated or saturated with FMN. This interpretation is also in agreement with the dissociation constants determined, as riboflavin shows the highest affinity to ppBat (Table 3).

The mutation of either Cys74 or Cys111 to alanine residues leads to zinc-free proteins, causing a striking decrease of the rearrangement temperature from 74 °C to around 60 °C and therefore indicating that zinc has a structural role. In another experiment, ppBat wild type and variants were incubated in presence of EDTA to remove the zinc ion from the

protein structure. In this case all the proteins showed a very similar rearrangement temperature supporting again the structural function of the zinc (Figure 4B). It is worth noting that in Figure 4A the Cys74Ala variant has the lowest rearrangement temperature in comparison to the other single and double variants. This finding can be interpreted on the basis of the crystal structure of ppBat. The Cys74 is exposed to the protein surface and located in proximity to a hydrophobic pocket consisting of Val78, Val10 and Phe72. One might expect that the alanine residue of the Cys74Ala variant flips positioning the hydrophobic group towards the inner hydrophobic pocket, hence provoking a significant rearrangement of the secondary structure and leading to lower thermostability. Conversely, Cys111 is positioned far away from this hydrophobic pocket thus the alanine residue of the Cys111Ala variant does not flip, which reflects the higher thermostability of this variant compared to the Cys74Ala variant.

15.5   References

1. Knaus, T., Eger, E., Koop, J., Stipsits, S., Kinsland, C. L., Ealick, S. E., and Macheroux, P. (2012) Reverse Structural Genomics: An Unusual Flavin-Binding Site in a Putative Protease from *Bacteroides thetaiotaomicron.*, *J Biol Chem.* **287**, 27490-27498.

2. Massey, V. (1991) in *Flavins and Flavoproteins* (Curti, B., Zanetti, G., and Ronchi, S., Eds.), pp 59-66, Walter de Gruyter, Como, Italy.

3. Minnaert, K. (1965) Measurement of the equilibrium constant of the reaction between cytochrome c and cytochrome a., *Biochim Biophys Acta.* **110**, 42-56.

4. Macheroux, P., Petersen, J., Bornemann, S., Lowe, D. J., and Thorneley, R. N. (1996) Binding of the oxidized, reduced, and radical flavin species to chorismate synthase. An investigation by spectrophotometry, fluorimetry, and electron paramagnetic resonance and electron nuclear double resonance spectroscopy., *Biochemistry.* **35**, 1643-1652.

5. Pantoliano, M. W., Petrella, E. C., Kwasnoski, J. D., Lobanov, V. S., Myslik, J., Graf, E., Carver, T., Asel, E., Springer, B. A., Lane, P., and Salemme, F. R. (2001) High-density miniaturized thermal shift assays as a general strategy for drug discovery., *J Biomol Screen.* **6**, 429-440.

6. Forneris, F., Orru, R., Bonivento, D., Chiarelli, L. R., and Mattevi, A. (2009) ThermoFAD, a Thermofluor-adapted flavin ad hoc detection system for protein folding and ligand binding., *FEBS J.* **276**, 2833-2840.

Chapter 16

Appendix

PERSONAL INFORMATION

Name	**TANJA KNAUS**
Address	**MONDSCHEINGASSE 3/9, 8010 GRAZ - AUSTRIA**
Telephone	**0043/6507813227**
E-mail	**tanja.knaus@tugraz.at**
Nationality	**AUSTRIAN**
Date and place of birth	**04 MAY 1978, Mürzzuschlag (Austria)**

WORK EXPERIENCE

JANUARY 2009 – PRESENT

(4) Teaching Assistant Professor: Department of Biochemistry – *GRAZ UNIVERSITY OF TECHNOLOGY, Graz (AUSTRIA)* – Group of research: Prof. Dr. Peter Macheroux

AUGUST 2001 – SEPTEMBER 2003

(3) Strategic Purchasing Department at a medical-technical company: Purchase of tool-related plastic parts – *ROCHE DIAGNOSTICS, Graz (Austria)*

JULY 2000 – JULY 2001

(2) Purchasing Department at an electrotechnical company: Technical Assistant for the Implementation of new Products – *ARTESYN AUSTRIA GmbH, Kindberg (Austria)*

JULY 1997 – JUNE 2000

(1) Purchasing Department at an electrotechnical company: Technical Assistant for the purchase of active and passive electronic parts – *ARTESYN AUSTRIA GmbH, Kindberg (Austria)*

EDUCATION

JANUARY 2009 – PRESENT

(5) PhD DISSERTATION at the Department of Biochemistry: *GRAZ UNIVERSITY OF TECHNOLOGY, Graz (AUSTRIA)* – Title: Biochemical Characterization and Mechanistic Studies on Two Zinc Dependent Proteins – Group of research: Prof. Dr. Peter Macheroux

DECEMBER 2008

(4) MASTER OF SCIENCE IN CHEMISTRY (spec. in Biotechnology, Biochemistry and Food Chemistry: *GRAZ UNIVERSITY OF TECHNOLOGY, Graz (AUSTRIA)*; (with honour)

AUGUST 2008 – DECEMBER 2008

(3) DIPLOMA THESIS at the Department of Biochemistry: *GRAZ UNIVERSITY OF TECHNOLOGY, Graz (AUSTRIA)* – Title: Characterization of the NADP(H):FMN Oxidoreductase YcnD from *Bacillus subtilis* – Group of research: Prof. Dr. Peter Macheroux

OCTOBER 2003 – DECEMBER 2008

(2) ACADEMIC STUDIES OF TECHNICAL CHEMISTRY (Diploma Programme): *GRAZ UNIVERSITY OF TECHNOLOGY, Graz (AUSTRIA)*

JUNE 1997

(1) INDUSTRIAL ENGINEER IN PLASTICS ENGINEERING (high school level): *SCHOOL OF TECHNOLOGY FOR PLASTICS, Kapfenberg (AUSTRIA)*

Teaching activities

2012   (9) Assistant Professor for the course "Methods in Immunology Laboratory Course", Graz University of Technology

(8) Assistant Professor for the course "Molecular Biology Laboratory Course", Graz University of Technology

2011   (7) Assistant Professor for the course "Biochemistry Laboratory Course I", Graz University of Technology

(6) Assistant Professor for the course "Methods in Immunology Laboratory Course", Graz University of Technology

(5) Assistant Professor for the course "Molecular Biology Laboratory Course", Graz University of Technology

2010   (4) Assistant Professor for the course "Biochemistry Laboratory Course I", Graz University of Technology

(3) Assistant Professor for the course "Methods in Immunology Laboratory Course", Graz University of Technology

(2) Assistant Professor for the course "Molecular Biology Laboratory Course", Graz University of Technology

2009   (1) Assistant Professor for the course "Biochemistry Laboratory Course I", Graz University of Technology

Graduate research students supervised

February 2012 – Present	(5) Stefanie Monschein, Master Diploma Thesis: Title: Functional Characterisation of the putative protease from *Bacteroides thetaiotaomicron*
September 2011 – June 2012	(4) Julia Koop, Master Diploma Thesis: Title: Characterization of flavin binding properties of the putative protease I from *Bacteroides thetaiotaomicron* and isolation of flavins from bovine milk
June 2011 – September 2011	(3) Elisabeth Eger, Bachelor Thesis: Title: Characterization of 3CNE from *Bacteroides thetaiotaomicron*
February 2010 – May 2010	(2) Julia Pitzer, Bachelor Thesis: Title: Pisa1 from *Pseudomonas* sp., the first inverting *sec*-alkylsulfatase
July 2009 – October 2009	(1) Martin Mario Faccinelli, Bachelor Thesis: Title: Pisa1, Expression, Purification and Analysis of a Secondary Alkylsulfatase

Others   Languages

Mother tongue	German
Other languages	English Level: Fluent

PUBLICATION LIST

PUBLICATIONS ON PEER-REVIEWED JOURNALS

6. Koop J., Monschein S., Macheroux E.P., **Knaus T.*** and Macheroux P.*: "Determination of free and bound riboflavin in cow's milk using a novel flavin-binding protein", paper submitted to *Food Chemistry.*

* corresponding author

5. **Knaus T.**, Schober M., Kepplinger B., Faccinelli M., Pitzer J, Faber K., Macheroux P.* and Wagner U. *: "Structure and mechanism of an inverting alkylsulfatase from *Pseudomonas* sp. DSM6611 specific for secondary alkylsulfates", *FEBS Journal,* in press (DOI: 10.1111/febs.12027).

4. **Knaus T.**, Eger E., Koop J., Stipsits S., Kinsland C.L., Ealick S.E. and Macheroux P.*: "Reverse Structural Genomics: "An Unusual Flavin Binding Site in a Putative Protease from *Bacteroides thetaiotaomicron*", *The Journal of Biological Chemistry* (2012), Vol. 287, No. 33, pages 27490-27498.

3. Schober M., **Knaus T.**, Toesch M., Macheroux P., Wagner U. and Faber K.*: "The Substrate Spectrum of the Inverting *sec*-Alkylsulfatase Pisa1", *Advanced Synthesis and Catalysis* (2012) Vol. 354, No. 9, pages 1737–1742.

2. Schober M., Gadler P., **Knaus T.**, Kayer H., Birner-Grünberger R., Gülly C., Macheroux P., Wagner U. and Faber K.*: "A Stereoselective Inverting *sec*-Alkylsulfatase for the Deracemisation of *sec*-Alcohols", *Organic Letters* (2011), Vol. 13, No. 16, pages 4296-4299.

1. Voss C.V., Gruber C.C., Faber K., **Knaus T.**, Macheroux P. and Kroutil W.*: "Orchestration of concurrent oxidation and reduction cycles for stereoinversion and deracemisation of *sec*-alcohols", *Journal of the American Chemical Society* (2008) Vol. 130, No. 42, pages 13969–13972.

ADDITIONAL MANUSCRIPTS IN PREPARATION

2. Langer S., Nakanishi S., Mathes T., Schleicher E., **Knaus T.**, Macheroux P., Mase T., Miyakawa T., Tanokura M. and Mack M.*: "The flavoenzyme AzoR from *Escherichia coli* is a target for the antibiotic roseoflavin produced by *Streptomyces davawensis*"

1. Mutti F.G., Sattler J.H., Pressnitz D., Fuchs C., **Knaus T.**, Macheroux P. and Kroutil W.*: "Asymmetric synthesis of pharmacologically relevant aromatic bicyclic amines employing ω-transaminases"

ORAL PRESENTATIONS TO INTERNATIONAL CONFERENCES AS CO-AUTHOR

3. Macheroux P., **Knaus T.**, Daniel B., Binter A., Pavkov-Keller T. and Gruber, K.: "Reverse structural flavogenomics: Discovery of function for uncharacterized flavoproteins." Trends in Enzymology - Göttingen (Germany), 2012

2. **Knaus T.**, Schober M., Faber K. and Macheroux P.: "Active site characterisation of the first *sec*-alkyl sulfatase: towards the deracemisation of *sec*-alcohols." The 5th central European conference - Chemistry toward Biology - Primosten (Croatia), 2010

1. Voss C.V., Gruber C.C., **Knaus T.**, Faber K., Macheroux P. and Kroutil W.: "Concurrent tandem redox reactions for deracemisation of *sec*-alcohols." International congress on biocatalysis (Biocat 2008) - Hamburg (Germany), 2008

POSTER PRESENTATIONS TO INTERNATIONAL CONFERENCES

Communications (2), (3), (5), (6) and (8) were presented by Tanja Knaus

9. Toesch M., Schober M., **Knaus T.**, Macheroux P., Wagner U. and Faber K.: "A two-step enantioconvergent process for the deracemisation of *sec*-alcohols via inverting and retaining alkylsulfatases." Multistep enzyme-catalyzed processes (MECP2012) - Graz (Austria), 2012

8. **Knaus T.**, Eger E., Koop J., Stipsits S., Kinsland C.L., Ealick S.E. and Macheroux P.: "Reverse structural genomics: An unusual flavin binding site in a putative protease from *Bacteroides thetaiotaomicron*." Trends in Enzymology – Göttingen (Germany), 2012

7. Schober M., **Knaus T.**, Macheroux P., Wagner U. and Faber K.: "A stereoselective inverting *sec*-alkylsulfatase for the deracemisation of sec-alcohols." Biotrans 2011 - Giardini Naxos, Sicily (Italy), 2011

6. **Knaus T.**, Schober M., Wagner U., Faber K. and Macheroux P.: "Structural and molecular characterisation of the first inverting *sec*-alkylsulfatase (Pisa1) from *Pseudomonas* sp." Biotrans 2011 - Giardini Naxos, Sicily (Italy), 2011

5. **Knaus T.**, Gadler P., Schober M., Faber K., Wagner U. and Macheroux P.: "Pisa1 from *Pseudomonas* sp., the first inverting *sec*-alkyl sulfatase." Conference on biocatalysis for chemical synthesis – Graz (Austria), 2010

4. Schober M., Gadler P., **Knaus T.**, Macheroux P. and Faber K.: "Molecular characterisation of the first inverting *sec*-alkyl sulfatase (Pisa1) from *Pseudomonas* sp." Novel Enzymes 2010 – Exeter, (U. K.), 2010

3. **Knaus T.**, Gadler P., Schober M., Faccinelli M., Wagner U., Macheroux P. and Faber K.: "Site-directed mutagenesis of Pisa1 reveals active site residues involved in enantio-preference." Novel Enzymes 2010 – Exeter, (U. K.), 2010

2. **Knaus T.**, Ehammer H., Sollner S. and Macheroux P.: "YcnD from *Bacillus subtilis* unleashed: a Flavin-Containing NAD(P)H:FMN Oxidoreductase, which also acts as a Quinone Reductase." 16th International Symposium on Flavins and Flavoproteins – Jaca (Spain), 2008

1. Steiner A.J., Kartusch C., **Knaus T.**, Stütz A.E. and Wrodnigg T.M.: "Glycoconjugation employing Cabohydrate Bearing N-Hydroxysuccinimide Esters." Benzon Symposium 54 – Glycosylation: Oppurtunities in Drug Development – Copenhagen (Denmark), 2007.

OTHER POSTER PRESENTATIONS

Communications (4) and (5) were presented by Tanja Knaus

5. **Knaus T.**, Eger E. and Macheroux P.: "An unusual flavin binding site in a putative protease from *Bacteroides thetaiotaomicron*." 7th DocDay NAWI Graz Doctoral School of Molecular Biosciences and Biotechnology - Graz (Austria), 2012

4. **Knaus T.**, Eger E., Stipsits S. and Macheroux P.: "A putative protease from *Bacteroides thetaiotaomicron* - the first zinc-dependent flavoprotein." 6th DocDay NAWI Graz Doctoral School of Molecular Biosciences and Biotechnology - Graz (Austria), 2011

3. Schober M., **Knaus T.**, Toesch M., Macheroux P., Wagner U. and Faber K.: "A stereoselective inverting *sec*-alkylsulfatase for the deracemisation of sec-alcohols." 6th DocDay NAWI Graz Doctoral School of Molecular Biosciences and Biotechnology - Graz (Austria), 2011

2. Schober M., **Knaus T.**, Toesch M., Gadler P., Wagner U., Macheroux P. and Faber K.: "Development of an enantioconvergent process for the deracemisation of secondary alcohols with the first inverting *sec*-alkylsulfatase (Pisa1) from *Pseudomonas* sp." 5th DocDay NAWI Graz Doctoral School of Molecular Biosciences and Biotechnology - Graz (Austria), 2011

1. Schober M., Gadler P., **Knaus T.**, Macheroux P. and Faber K.: "Molecular characterisation of the first inverting *sec*-alkyl sulfatase (Pisa1) from *Pseudomonas* sp." 3rd DocDay, NAWI Graz Doctoral School of Molecular Biosciences and Biotechnolgy - Graz (Austria), 2010