On Catalytic Mechanisms for Rational Enzyme Design Strategies

Charlotte Kürten

Doctoral Thesis

KTH, Royal Institute of Technology
School of Engineering Sciences in Chemistry, Biotechnology and Health
Stockholm 2018
Für Conny
Abstract

Enzymes enable life by promoting chemical reactions that govern the metabolism of all living organisms. As green catalysts, they have been extensively used in industry. However, to reach their full potential, engineering is often required, which can benefit from a detailed understanding of the underlying reaction mechanism. In Paper I, we screened for an esterase with promiscuous amidase activity capitalizing on a key hydrogen bond acceptor that is able to stabilize the rate limiting nitrogen inversion. In silico analyses revealed the esterase patatin as promising target that indeed catalyzed amide hydrolysis when tested in vitro. While key transition state stabilizers for amide hydrolysis are known, we were interested in increasing our fundamental understanding of terpene cyclase catalysis (Paper II-V).

In Paper II, kinetic studies in D2O-enriched buffers using a soluble diterpene cyclase suggested that hydrogen tunneling is part of the rate-limiting protonation step. In Paper III, we performed intense computational analyses on a bacterial triterpene cyclase to show the influence of water flow on catalysis. Water movement in the active site and in specific water channels, influencing transition state formation, was detected using streamline analysis. In Paper IV and V, we focused on the human membrane-bound triterpene cyclase oxidosqualene cyclase. We first established a bacterial expression and purification protocol in Paper IV, before performing detailed in vitro and in silico analyses in Paper V. Our analyses showed an entropy-driven reaction mechanism and the existence of a tunnel network in the structure of the human enzyme. The influence of water network rearrangements on the thermodynamics of the transition state formation were confirmed. Introducing mutations in the tunnel lining residues severely affected the temperature dependence of the reaction by changing the water flow and network rearrangements in the tunnels and concomitant the active site.

Keywords: catalytic mechanisms, terpene cyclase, triterpene cyclase, solvent dynamics, protein hydration, thermodynamics, quantum tunneling, polycyclization, natural compounds, α/β-hydrolase, esterase, amidase, enzyme engineering, biocatalysis
Popular Scientific Summary

Enzymes are the cell’s catalysts. They accelerate chemical reactions, crucial for metabolic processes, and thereby enable life for all organisms. Most often they are proteins that perform chemistry at very mild conditions, such as a neutral pH, modest temperatures and atmospheric pressure. Additionally, in comparison to many traditionally used chemical catalysts, they are characterized by their selectivity. Their potential as green catalysts, performing incredible chemistry at environmentally-friendly conditions, has been explored in industry. In addition to using natural enzymes for industrial purposes, efforts to custom-tailor the biocatalysts for specific needs has fascinated biochemists for decades and has generated a number of industry-applied biocatalysts. A detailed understanding of the performed reaction can build a solid foundation for engineering processes.

Different enzymes catalyze different reactions but all of them accelerate the reaction by lowering the activation energy, which is the energy barrier between the reactants and the transition state on the way to product formation. A lowering of the activation energy can be achieved using different strategies. In the studies comprised within this thesis, we focused on two different enzyme classes: hydrolases and triterpene cyclases.

Hydrolases perform hydrolytic reactions, where a chemical bond is cleaved using water. Many hydrolases are essential for our digestion, like lipases that break down lipids (fats). They are also commonly used in industry, for example in laundry detergents to remove stains from clothing, as a greener alternative to conventionally used chemicals.

In Paper I, we applied the knowledge of a key interaction that lowers the activation energy in the hydrolytic cleavage of amides to identify an esterase (a hydrolase that uses esters as substrates and not amides) that can perform amide hydrolysis. Computational screening for a protein scaffold, which stabilizes this key interaction, revealed the potato enzyme patatin as a promising target. In vitro analysis of patatin indeed showed that this enzyme exhibits amidase activity. This study highlights, how a deeper understanding of the reaction mechanism can help in finding enzymes that can perform a certain reaction, even if only with low activity. Those hits can be a good starting point for further engineering strategies.

While the key activity shifter of hydrolases has been known, we were interested in increasing our fundamental understanding of terpene cyclase catalysis for future engineering (Paper II to V).
Terpene cyclases, perform elaborate reactions to produce complex scaffolds, which can have different beneficial functions. These multicyclic compounds can possess anti-microbial, anti-cancer, and anti-viral properties and are therefore of high industrial and pharmaceutical value. The human triterpene cyclase produces the precursor to all steroids, which forms the backbone of many hormones, vitamins, and signaling compounds. In Paper II to V, we focused on elucidating fundamentals of the terpene cyclase reaction mechanism.

In Paper II, analyses with a soluble terpene cyclase using heavy water (D₂O) showed that hydrogen tunneling, describing an atom moving through instead of over the above-mentioned energy barrier, is part of the reaction. Paper III to V evolve around a subclass of terpene cyclases – the triterpene cyclases. Computational analyses of a bacterial triterpene cyclase performed in Paper III, revealed that water molecules in the enzyme structure move in distinct streamlines. This water movement seems to be specific at different stages of the catalytic cycle, which could affect the lowering of the activation energy. The protein structures of triterpene cyclases are very conserved, that is why in Paper IV and Paper V, we put our attention towards the human oxidosqualene cyclase and analyzed if water movement also impacts the catalytic reaction performed by the human enzyme. First, we had to establish an easy-to-handle bacterial expression and purification protocol (Paper IV) to produce large quantities of the enzyme, before we performed the extensive in vitro and in silico analyses presented in Paper V. Our analyses showed that the enzyme’s structure is honeycombed with a tunnel network that allows water molecules to move within the structure. Furthermore, we were able to show that water network rearrangements immensely influence the activation barrier between substrates and products. We introduced tunnel obscuring mutations in the tunnel linings, which severely affected the reaction energetics by changing the water flow in the tunnels and concomitant active site.

In summary, the presented studies highlight the potential of understanding reaction mechanisms for enzyme engineering and the influence of solvent rearrangements on enzyme catalysis.
Populärvetenskaplig Sammanfattning


Terpencyklaser har förmågan att syntetisera komplexa molekylära strukturer genom katalys av mycket sofistikerade reaktioner. Dessa producerar multicykliska molekyler med en rad intressanta biologiska egenskaper såsom anti-mikrobiella, anti-cancer och anti-virala, och är därför av stort industriellt och farmaceutiskt värde. Det mänskliga triterpencyklaset producerar en molekyl som är föregångare till alla steroider, vilka utgör stommen i många hormoner, vitaminer och signalämnen.


Sammanfattningsvis betonar dessa studier potentialen av fundamental kunskap om enzymers reaktionsmekanismer för utvecklandet av nya tekniker samt effekten av omordning av lösningsmedel (vatten) inom enzymkatalys.
Thesis Defense

This thesis will be defended on the 26th of October 2018 at 1 PM in room K1, located at Teknikringen 56 at the KTH main campus for the degree of Doctor of Philosophy in Biotechnology (PhD in Biotechnology).

Respondent

Charlotte Kürten graduated as Master of Science from the Ruhr-University in Bochum, Germany in August 2014 and started her PhD at Kungliga Tekniska Högskolan (KTH) in September 2014.

Faculty Opponent

Rudolf Alleman, Professor, Pro Vice-Chancellor and Head of College of Physical Sciences and Engineering, School of Chemistry, Cardiff University.

Evaluation Committee

Mikael Widersten, Professor in Biochemistry, Department of Chemistry, Uppsala University.

Alessandra Villa, Associate Professor, Department of Biosciences and Nutrition, Karolinska Institute.

Katja Petzold, Associate Professor, Department of Medical Biochemistry and Biophysics, Karolinska Institute.

Chairman of the Thesis Defense

Peter Nilsson, Professor in Proteomics, Department of Protein Science, School of Engineering Sciences in Chemistry, Biotechnology and Health, Science for Life Laboratory, KTH.

Supervisors

Per-Olof Syrén, Associate Professor in Chemistry for Life Science, Department of Fibre and Polymer Technology, School of Engineering Sciences in Chemistry, Biotechnology and Health, Science for Life Laboratory, KTH.

Lucie Delemotte, Assistant Professor in Computational Biophysics, Department of Applied Physics, School of Engineering Sciences, Science for Life Laboratory, KTH.

Mathias Uhlén, Professor in Microbiology, Department of Protein Science, School of Engineering Sciences in Chemistry, Biotechnology and Health, Science for Life Laboratory, KTH.
List of Publications and Manuscripts

The presented thesis is based on the following articles (I-IV) and manuscript (V). Full versions of the papers are in the appendix of the thesis and are reprinted with permission from the copyright holders.

**Paper I**
Kürten, C.; Carlberg, B.; Syrén, P-O. Mechanism-Guided Discovery of an Esterase Scaffold with Promiscuous Amidase Activity.


*Respondent’s contribution: Performed part of the experiments and assisted in manuscript writing.*

**Paper II**


*Respondent’s contribution: Assisted the planning and supervision of the experiments and data analysis, prepared figures and assisted in manuscript writing.*

**Paper III**

ACS Omega. 2017, doi: 10.1021/acsomega.7b01084

*Respondent’s contribution: Contributed to experimental planning and data interpretation, assisted in manuscript writing.*

**Paper IV**
Kürten, C.; Uhlén, M.; Syrén, P-O. Overexpression of Functional Human Oxidosqualene Cyclase in *Escherichia coli*.


*Respondent’s contribution: Planned and conducted experiments, performed data analysis, major part in manuscript writing.*
Paper V

Kürten, C.; Eriksson, A.; Maddalo, G.; Edfors, F.; Uhlén, M.; Syrén, P-O. Engineering of Water Networks in Class II Terpene Cyclases Underscores the Importance of Amino Acid Hydration and Entropy in Biocatalysis and Enzyme Design.

Manuscript

Respondent’s contribution: Planned and conducted majority of experiments, performed data analysis, main part in manuscript writing.

Related Work not Included in the Thesis

Kürten, C.; Syrén, P-O. Unraveling Entropic Rate Acceleration Induced by Solvent Dynamics in Membrane Enzymes.

J. Vis. Exp. 2016, doi:10.3791/53168


J. Proteome Res. 2018, doi: 10.1021/acs.jproteome.7b00599
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-terminus, N-terminus</td>
<td>Carboxy-terminus and amino-terminus</td>
</tr>
<tr>
<td>CAI</td>
<td>Codon adaptation index</td>
</tr>
<tr>
<td>CAST</td>
<td>Combinatorial active site saturation test</td>
</tr>
<tr>
<td>CPS</td>
<td>ent-copalyl diphosphate synthase</td>
</tr>
<tr>
<td>Cryo-EM</td>
<td>Cryo-electron microscopy</td>
</tr>
<tr>
<td>DMAPP</td>
<td>Dimethylallyl diphosphate</td>
</tr>
<tr>
<td>E</td>
<td>Enzyme</td>
</tr>
<tr>
<td>$E_a$</td>
<td>Activation energy</td>
</tr>
<tr>
<td>ES</td>
<td>Enzyme-substrate complex</td>
</tr>
<tr>
<td>ff</td>
<td>Force field</td>
</tr>
<tr>
<td>FPP</td>
<td>Farnesyl diphosphate</td>
</tr>
<tr>
<td>$G, G^*$</td>
<td>Gibbs free energy, $G$ of activation</td>
</tr>
<tr>
<td>GC</td>
<td>Gas chromatography</td>
</tr>
<tr>
<td>GFPP</td>
<td>Geranyl farnesyl diphosphate</td>
</tr>
<tr>
<td>GPCR</td>
<td>G protein-coupled receptor</td>
</tr>
<tr>
<td>GPP, GGPP</td>
<td>Geranyl (geranyl) diphosphate</td>
</tr>
<tr>
<td>$H, H^*$</td>
<td>Enthalpy, $H$ of activation</td>
</tr>
<tr>
<td>hOSC</td>
<td>Human oxidosqualene cyclase</td>
</tr>
<tr>
<td>HPLC</td>
<td>High performance liquid chromatography</td>
</tr>
<tr>
<td>IPP</td>
<td>Isopentenyl diphosphate</td>
</tr>
<tr>
<td>KIE</td>
<td>Kinetic isotope effect</td>
</tr>
<tr>
<td>LC</td>
<td>Liquid chromatography</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular dynamics</td>
</tr>
<tr>
<td>MM</td>
<td>Molecular mechanics</td>
</tr>
<tr>
<td>MS</td>
<td>Mass spectrometry</td>
</tr>
<tr>
<td>NAC</td>
<td>Near attack conformer</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>PA clan</td>
<td>Proteases of mixed nucleophile, superfamily A</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein data bank</td>
</tr>
<tr>
<td>PET</td>
<td>Polyethylene terephthalate</td>
</tr>
<tr>
<td>PTM</td>
<td>Post-translational modification</td>
</tr>
<tr>
<td>QM</td>
<td>Quantum mechanics</td>
</tr>
<tr>
<td>RMSD</td>
<td>Root mean square deviation</td>
</tr>
<tr>
<td>RMSF</td>
<td>Root mean square fluctuation</td>
</tr>
<tr>
<td>$S, S^*$</td>
<td>Entropy, $S$ of activation</td>
</tr>
<tr>
<td>SHC</td>
<td>Squalene hopene cyclase</td>
</tr>
<tr>
<td>THz</td>
<td>Terahertz</td>
</tr>
<tr>
<td>TLC</td>
<td>Thin layer chromatography</td>
</tr>
<tr>
<td>TS</td>
<td>Transition state</td>
</tr>
<tr>
<td>WT</td>
<td>Wild type</td>
</tr>
</tbody>
</table>

XV
Table of Contents

ABSTRACT ......................................................................................................................... V
POPULAR SCIENTIFIC SUMMARY .................................................................................... VII
POPULÄRVTENSKAPLIG SAMMANFATTNING ................................................................ IX
THESIS DEFENSE ............................................................................................................. XI
LIST OF PUBLICATIONS AND MANUSCRIPTS .............................................................. XIII
ABBREVIATIONS ............................................................................................................... XV
THESIS OUTLINE ............................................................................................................. 1
I. ENZYMES ...................................................................................................................... 3
   Enzyme Structure ........................................................................................................... 4
   Biocatalysis and Enzyme Engineering ........................................................................... 6
II. Enzyme Catalysis ......................................................................................................... 15
   Catalytic Mechanisms ................................................................................................... 15
   Enzyme Kinetics ............................................................................................................ 21
   Transition State Theory and Thermodynamics of Enzymatic Reactions .................... 23
   Protein Structural Dynamics and Influence on Catalytic Function ............................. 28
   Protein Hydration and Solvent Influence on Protein Function .................................... 29
III. α/β-Hydrolases and Esterases .................................................................................. 39
IV. Terpenes and Terpene Synthases .............................................................................. 43
   Terpenes ....................................................................................................................... 43
   Terpene Cyclases ......................................................................................................... 46
V. Present Investigation ..................................................................................................... 55
   Paper I ......................................................................................................................... 55
   Paper II ....................................................................................................................... 58
   Paper III ..................................................................................................................... 60
   Paper IV ...................................................................................................................... 63
   Paper V ....................................................................................................................... 65
   Conclusions and Future Perspectives .......................................................................... 71
ACKNOWLEDGEMENTS ................................................................................................... 75
REFERENCES .................................................................................................................... 77
Thesis Outline

This thesis highlights the importance and potential of understanding catalytic mechanisms for the further development of comprehensive rational engineering strategies.

Chapter I gives an overall introduction to enzymes, their structure and use in biocatalysis and their engineering potential. It also includes a section about the basics and methods used for computer-aided protein design.

Chapter II deals with basic catalytic strategies, analytic tools to investigate enzyme catalysis and introduces thermodynamic principles. Protein dynamics and the influence of hydration and solvent movement on the thermodynamics of the system, with a focus on entropy of the reaction and catalysis, are elucidated in more detail.

Chapter III and IV introduce hydrolases and terpene cyclases, the enzyme classes and enzymes analyzed within the scope of this thesis.
I. Enzymes

Enzymes (from the Greek *enzymos*, leavened or in yeast [1]) are protein-based biological catalysts (biocatalysts). They can accelerate chemical reactions up to a factor of $10^{26}$ compared to the uncatalyzed reaction [2]. The fastest enzymes reach kinetic perfection and work at the diffusion limit, where the rate limiting step is the diffusion of the substrate or product in or out of the active site [3]. Almost all reactions occurring in cells and viruses need enzymatic catalysis for essential metabolic processes to occur at rates compatible with life. Essential reactions that would otherwise take up to hundreds or thousands of years or longer are accelerated up to sub-second timescales [4]. Humans have made use of enzymatic activity for centuries without knowing the molecular mechanisms governing the processes. The use of yeast in sugar fermentation to alcohol for the enzymatic production of wine predates 6000 BC and evidence for the production of cheese for dairy preservation dates back to 5500 BC [5]. Since then, enzymes in the form of yeast cultures, vegetable extracts or rennet have been used in many biotechnological applications, such as in the food production, in fermentation processes, dairy production, bread making, and meat tenderizing.

Jöns Jacob Berzelius was the first to acknowledge a special catalytic power of living tissue and sap coming from plants and animals that enhance chemical reactions in 1838 [6]. He was also the one to coin the word protein (from the Greek *proteios*, meaning primary) [7]. The connection that most enzymes are proteins came much later, with the first purification and crystallization of the enzyme urease from jack beans in 1926 by Sumner [8]. The new insight and understanding eventually led to the publication of the first 3D protein structure of myoglobin in 1958 [9] and the first enzyme structure of lysozyme in 1965 [10]. This propelled a structural revolution in protein science and enzymology. Today, 132,902 protein structures are reported in PDB (Protein Data Bank, as of August 2018).

Structure elucidation has had an incredible impact on the pharmaceutical industry as it allows a more targeted structural approach in drug design [11], but also on the biotechnology sector. Enzymes as biocatalysts are widely used in a variety of different industries [12]. However, there is often a need for enzyme engineering to generate the highly efficient and specialized enzymes used in industry today, especially for the production of novel compounds [13, 14]. The design process is simplified, when structural knowledge can be implemented.
Enzyme Structure

Most catalytic activities in the cell are performed by protein-based catalysts that are composed of varying sequences of smaller building blocks, the amino acids. There are 20 different proteinogenic amino acids with varying side chain properties that are connected via peptide bonds between their carboxy (C-) and amino (N-) -termini to form peptides and polypeptides (Figure 1 A and B). The polypeptide chain can consist of tens to tens of thousands of residues [15, 16]. The linear amino acid sequence is referred to as the primary structure of a protein (Figure 1 B). Interactions between backbone atoms of the biopolymer chain (mostly hydrogen bonds) dictate the formation of structural patterns called α-helices, β-sheets or loops (Figure 1 C). These secondary structural elements are arranged into a three-dimensional tertiary structure (Figure 1 D) that imparts protein function [17]. In addition to hydrogen bonds, the tertiary structure can be stabilized by other factors, such as hydrophobic effects, ionic interactions and salt bridges between the side chain residues, or by covalent interactions, such as disulfide bridges. Several tertiary structures of the same or different macromolecules can also cluster together and form a quaternary fold (Figure 1 E).

The amino acid sequence usually determines the structure of the protein, which in turn confers a protein’s function [17]. Despite the countless different primary structures, only a couple of hundreds of different tertiary protein folds have been observed in nature [18]. Enzymes performing similar reactions often share a common fold, such as the α/β-hydrolases (see α/β-Hydrolases and Esterases). The structure-function relationship puts larger evolutionary constraints on the structure compared to the sequence. If a mutation changes an amino acid, it doesn’t necessarily mean that the structure is changed. But if the structure is changed, the mutation is much more likely to impact the protein’s function. Catalytic residues, such as the catalytic triad in the α/β-hydrolases (see α/β-Hydrolases and Esterases), are an exception and are highly conserved; mutagenesis will directly affect activity. Under the selective pressure caused by evolution, the organism holding a functionally negative mutated protein will eventually be eradicated. This results in protein structures being more conserved than their amino acid sequences [19].

More recently it has been shown that, despite the fact that most catalytic function in the cell is executed by protein-based enzymes, RNA-based catalysts (Ribozymes) like the Ribosome exist [20]. Their discovery in the 1980s is seen as major evidence for the RNA-world hypothesis as the origin of life, in which ribozymes are proposed to be the most ancient biological catalysts [21].
Figure 1: Protein structure. A: Two amino acids with side chains $R_2$ and $R_3$ are linked via a peptide bond (blue) in a condensation reaction, resulting in a peptide and water (green). In the cell, this reaction is performed by the ribosome using aminoacyl-tRNAs that carry the amino acids. B: Primary protein structure in form of a linear, unfolded polypeptide chain. C: $\alpha$-helices (blue) and $\beta$-sheets (cyan) as secondary structural elements, represented as ribbons. D: Tertiary protein structure colored according to C with loops shown in grey. E: Quaternary protein structure with the different monomers colored in purple, green and cyan.
Biocatalysis and Enzyme Engineering

Biocatalysis is generally defined as the industrial use of enzymes and whole cells in green chemistry applications [13]. Enzymes are very specific and selective and can be employed under mild reaction conditions in terms of temperature, pressure and solvent and are therefore regarded as an attractive alternative to traditional chemical synthesis [13]. Enzymes can either be used as catalysts in the manufacturing of a commercial product, for example for the production of high fructose corn syrup using glucose isomerase [22], or the enzyme can be the commercial product itself, such as lipases that are a key component of detergents [23]. Drawbacks of using enzymes for industrial purposes compared to traditional chemical catalysis can be the high cost of enzyme production—which might include enzyme purification-, possible substrate, and product inhibition, low solubility of substrate or product [24] and limited operating conditions (pH, temperature, solvent tolerance) [25]. Therefore, cheap and easy expression systems are crucial to enable thorough enzymatic analysis (see Paper IV) and the large-scale production and industrial use of biocatalysts. Additionally, enzymes are often used for purposes or under conditions that they did not naturally evolve for, so that they require modifications to meet the desired needs [13, 26]. Molecular biology techniques, such as DNA mutagenesis but also the incorporation of unnatural amino acids [27], allow alteration of the protein sequence, leading to a modified amino acid chain with possibly new structure and/or function. Enzyme engineering studies can target the enzyme’s structural stability [28], activity in different environments [29, 30], substrate scope [29, 31] and product profile [32].

Generally, enzyme engineering strategies can be categorized into rational, random, semi-rational and de novo design approaches.

The random engineering approach (Figure 2, left), often in the form of directed evolution, does not require much knowledge about the protein structure and catalytic mechanism. Random mutagenesis, for example by error-prone Polymerase Chain Reaction (PCR), usually leads to large libraries (>1000) of mutated DNA sequences that can be transformed into the selected host organism. A screening process tailored to the enzymes’ catalytic activity can be used to select hit candidates that are expressed and tested in more detail. The best variant is then subjected to another round of random mutagenesis in an iterative process, until a protein variant with the desired features is generated. The screening process is pivotal for the selection of the catalyst, which is often summarized as “you get what you screen for” – the first law of directed evolution coined by Frances Arnold, a pioneer in this field [33]. One drawback of such a traditional directed evolutionary approach is the fact, that each generated variant is independent of the other variants in the library, meaning that beneficial substitutions (and also deleterious mutations) cannot be shared with the other library members and have to be “found” individually. To overcome this problem, a DNA shuffling step can be introduced, where the genes in the library are fragmented by digestion and randomly recombined by homologous recombination,
which enables lateral combination of beneficial mutations [34]. Random mutagenesis approaches can lead to mutations widely spread throughout the protein sequence and structure, and final enzyme variants can contain mutations far away from the catalytic active site. The consequence and influence of the introduced mutations on enzymatic activity is usually analyzed after obtaining the variant with the desired features and can lead to a rational understanding [35]. Such analysis can also reveal the influence of mutations in second or third shell residues on activity regulation mechanisms by for example protein [36, 37] or potentially solvent dynamics.

Random engineering approaches can generally be applied in cases, where no structural knowledge is prevailing or when the catalytic mechanism is unknown. In contrast, rational enzyme engineering strategies (Figure 2, right) require an in-depth understanding of the protein structure and the catalytic mechanism and often involve computational analyses (see Analysis of Computational Models and Applications for Enzyme Engineering) as well. Important residues that supposedly influence the targeted feature are changed purposely to generate a single or multiple number of variants that are expressed in the host of choice and analyzed with respect to the specific feature [38].

Figure 2: Random engineering and rational enzyme design. Left: The random engineering approach usually labelled as directed evolution does not require any prior knowledge of structure or catalytic mechanism. Random mutations are introduced to the DNA sequence, e.g. by error-prone PCR, the generated sequences (the library) are transformed into the host of choice, protein variants (yellow mutations) are expressed and eventually purified. The large number of generated sequences requires an additional screening and selection process and then further biochemical characterization of the generated variants. Usually, directed evolution is performed as an iterative process, where an overall very large number of variants is generated and tested, requiring a simple assay to test for the desired trait. Right: Rational enzyme design is a knowledge-based engineering strategy, where mutagenesis-sites are usually chosen in silico based on previously gathered information, such as structure elucidation. Specific mutations (blue) are introduced by site-directed mutagenesis; the altered sequence is transformed into the host of choice and the protein variant is expressed, purified and tested for the desired trait. Enzyme in grey, mutations in blue and yellow.
Semi-rational engineering strategies, a combination of both approaches, lead to a more targeted generation of a smaller library with mutations located in a certain selected area, which presumably impacts the trait to be optimized [39, 40]. One semi-rational enzyme engineering method based on structural knowledge widely applied in the field is the CAST (Combinatorial Active site Saturation Test) method [41, 42]. By targeting two to three spatially close amino acids simultaneously, this method allows cooperative, synergistic effects in a small mutant library. Followed by a traditional directed evolutionary approach, even more diverse but small libraries can be generated [41].

Engineering approaches do not only focus on improving already existing reactions, but also on generating catalysts that can promote reactions new to nature [14, 43]. For this purpose and in general semi-rational design strategies, choosing the right structural backbone as starting point for further engineering is crucial. A promising strategy to do so, is to exploit and evolve promiscuity [44-46], the ability of many enzymes to perform side activities in addition to their native catalyzed reaction [47]. It has been hypothesized that also in nature promiscuous activities serve as starting points for the natural evolution of enzymes with new catalytic functions [48]. Finding a backbone with potential side activities towards the desired reaction requires a deep understanding of the catalytic mechanism of both the natural and the new reaction (explored in Paper I).

However, not only natural enzyme scaffolds can be promising starting points for further engineering attempts. More recently such starting points have been designed de novo. In contrast to the above described methods, de novo protein design does not rely on already existing scaffolds but focuses on designing new structures from scratch. The design of the novel backbones is based on physical and not nature’s principles, which enables introducing enzymatic activity to sequences and structures new to nature [49]. Do novo protein design heavily relies on computational analysis (see Computational Methods for Enzyme Engineering) [50]. Besides designing whole novel protein scaffolds, de novo enzyme design, where a computed Transition State (TS) structure of a desired reaction is docked into an already existing backbone to find a promising scaffold, has been explored as well [51]. And although de novo enzymes have been designed [50, 51], their catalytic activities are usually low and they are often improved by including a subsequent directed evolution approach to optimize the designed backbone [52, 53]. These directed evolution experiments can lead to mutations far away from the active site, probably changing solvation and protein stability [54].

The fact that to date, we are not able to design high biocatalytic activity from scratch, resembling natural turnover numbers, highlights that our understanding of enzymatic activity and influences of the structure and solvation on catalysis is not complete. However, for semi-rational, rational engineering and de novo approaches to improve a catalyst or to introduce new reactivities, a profound knowledge of the catalytic mechanism as well as the structure-function relationship is crucial.
Computational methods can aid in gaining a deeper insight into the structure, dynamics and catalytic mechanisms of a desired reaction – both for improving an already existing reaction and for the de novo design of catalytic activities.

Computational Methods for Enzyme Engineering

Many modern enzyme engineering strategies include analysis of in silico models to understand, improve, and design enzymatic activity. A model, defined as larger (or smaller) representation of the object of interest, is easy to handle, simplified and helps to explain, control and predict events and allows the analysis of “what if” scenarios. The generation and analysis of in silico models, based on the protein structure, can thus help in investigating reaction mechanisms (Paper III, Paper V), in analyzing protein structure and dynamics and in redesigning catalytic activity. Furthermore, computational models can aid in creating more focused libraries in semi-rational approaches, or in finding promiscuous scaffolds (Paper I) and are crucial in de novo design. Results presented in this thesis contain conclusions drawn from analyses of static models and Molecular Dynamics (MD) trajectories, mostly focusing on properties of the protein structures and changes in solvation. Therefore, the following sections focus on the generation of MD trajectories, on analyses performed within this thesis, and on typical examples of in silico-based enzyme engineering approaches.

Model Generation and Dynamic Trajectories

Most commonly used approaches to generate enzyme models are either empirical, such as Molecular Mechanics (MM), MD and Docking methods or rely on ab initio calculations, based on quantum chemistry (Quantum Mechanics, QM). The modelling methods do not only differ in their fundamental principles but also in the time scales they allow to model (Figure 3). This is mostly limited by computer power. Static MM and progressive MD (evolution of the system over time) simulations use force fields (ff) to calculate the potential energies of all parts of the model. For simplification, atoms are treated as single particles with an assigned mass, radius, polarizability, and constant net charge. Bonds are treated as springs with an equilibrium distance equal to the experimental or calculated (by precise QM calculations) bond length. The same applies for bond angle bending and dihedral angles. Intramolecular and intermolecular interaction of the different atoms and molecules of the system are described by different equations comprised in the ff. They can be based both on empirical observations and quantum mechanics calculations and allow determination of the potential energies. The potential energies of the molecules equal the sum of energy terms for covalent and noncovalent interactions between atoms. Covalent interactions comprise energies for bond stretching, angle bending, rotation and the dihedral angle, whereas noncovalent interactions include electrostatic and van der Waals interactions. Ffs differ in functional forms and parameters and are therefore suitable for different structures.
Figure 3: Time scales of motions associated with enzyme structure and catalysis and in silico methods that can capture them. The length of ab initio MD and QM/MM MD mainly depends on the size of the regions treated using QM.

For the simulation of proteins most commonly used ffs are AMBER [55] (used in Paper III), CHARMM [56] and GROMOS (developed for GROMACS [57]). Some programs incorporate their own ffs, such as YAMBER [58], an AMBER based ff developed for the user-friendly simulation program YASARA [58] (used in Paper II, Paper V). MM energy minimizations are frequently used for the preparation of structures for following MD simulations. The experimentally determined structure, a common starting point for MD simulations of proteins, often does not represent the natural solvated form of the protein. There could be artefacts from crystal packing, for example. The natural structure usually corresponds to the most stable complex at the potential energy minimum [59]. Therefore, an energy minimization step that brings the structure to a low potential energy-geometry, can prepare the structure for further analysis. Different minimization techniques use different amounts of energetic information. An increasing amount of energetic information makes it easier to find the global energy minimum but comes at an increased computational cost. Energy minimizations in YASARA, using the YAMBER ff, start with a first-order steepest descent minimization, which gives a first approximation of the structure at a potential energy minimum. This is followed by a simulated annealing step until the potential energy of the structure converges and the structure is at its energy minimized fold. As mentioned above, ffs build not only the basis for static MM methods but also for the generation of MD trajectories, which simulate the dynamic evolution of the system over time (t). The trajectories of the atoms in the molecular system are calculated by numerically solving Newton’s equations of motion. The forces that the different molecules of the system exert on each other (calculated using the ff), determine their behavior in the system and with that the position of the particle at time t+Δt. Such ff-based in silico models can be analyzed in numerous different ways and
On Catalytic Mechanisms for Rational Enzyme Design Strategies

build the basis for many successful rational, semi-rational or *de novo* enzyme engineering studies [52, 60-62]. Because in such *ff*-based methods each atom is treated as single particle, energies are calculated as functions of the nuclear positions only and electronic motions are not considered. Therefore, electron distribution and with that bond breakage and formation (reactions) cannot be studied using MM and MD simulations. However, high-energy intermediates can be used to represent the TS during MD simulations.

In contrast, QM methods and *ab initio* MD, capitalize on the wave-particle duality of the atoms, so that electronic rearrangement processes -chemical reactions- can be followed. To simulate molecular systems and their potential energies and electronic structures on such detailed level, different quantum theories can be incorporated to solve the wavefunctions of the particles. A frequently used method is Density Functional Theory (DFT). The detail of these calculations allows their application mostly to smaller systems and is limited by computer power. Therefore, hybrid QM/MM [63] and QM/MM MD methods have been developed. For the simulation of enzymes and their catalyzed reactions, the active site region can be treated with QM, whereas the rest of the protein body is treated with lower cost MM methods. The QM/MM interaction region is very complex and the coupling of the two differently treated areas is an active field of development today.

Dynamic simulations (both MD, *ab initio* MD and hybrid QM/MM MD systems) are influenced by external parameters such as pH, temperature and solvation. These parameters need to be controlled. When modelling protein structures, usually a big number of water molecules has to be included in the model, which comes along with a computational cost. However, there are different solvent models that treat the solvent in the simulation box differently. If the movement of single water molecules and specific interactions with water molecules (hydrogen bonds for example) are not the center of the study, implicit solvent models can be applied. The solvent is thereby treated as polarizable continuum with a dielectric constant $\varepsilon$. In explicit solvent models, on the contrary, each single solvent molecule is simulated individually, which requires incorporation of empirical interaction potentials between solvent and solute and between the solvent molecules and makes the simulations longer and computationally more expensive. However, within the studies in this thesis, we were mostly interested in solvent dynamics and water network rearrangements, so that explicit simulations of water molecules were necessary (see Paper III and Paper V).

*Analysis of Computational Models and Applications for Enzyme Engineering*

Before an enzyme can be engineered rationally based on a computational model, the model or MD trajectory has to be analyzed accordingly. This is not limited to the protein structure and dynamics but also includes solvation and water network rearrangements.
Analyses based on the structure include a simple but effective visual inspection of the active regions (can be active site only or for example substrate access tunnel) of the generated model. Based on an understanding of the mechanism, such inspections allow rational placement of amino acid substitutions and analysis of introduced changes \textit{in silico} and subsequently \textit{in vitro} that can have major consequences on the catalytic activity or protein function [60]. An initial analysis of the protein structure often also includes determination of Root Mean Square Deviations (RMSD), the average distances between the atoms (usually atomic coordinates of the Ca, Figure 1 A) of superimposed proteins. This can give insight into the similarity of for example two different homologous proteins (used in Paper IV) or when analyzed with snapshots of an MD trajectory, the variation in the structure over time. Such changes in RMSD over time can also give valuable insight into the integrity and stability of the protein in the set-up system and is used to determine if the system reached equilibrium (applied in Paper III, Paper V). When analyzing a dynamic system where the atomic coordinates of the C\textalpha{} atoms of the amino acids fluctuate around a distinct average position over time, the RMSF (Root Mean Square Fluctuations) can additionally be calculated. RMSF analysis allows determination of large and small scale dynamics of single residues, structural elements and the protein backbone over time (used in Paper V) [64]. Such structural analyses can also reveal functional rearrangements of the structure over time, for example, large loop movements that can be crucial for catalysis and therefore depict a possible engineering target [65, 66].

However, one major goal of optimizing enzymes for their industrial use is to increase enzyme stability, the ability to remain active under challenging process conditions (e.g. pH, temperature, salt) [67]. It has been shown that the protein stability is related to its structural dynamics and that rigidification of flexible sites usually increases protein stability [62]. Such flexible regions and engineering targets can be detected by standard MD simulations (RMSF analysis), high-temperature MD that triggers the unfolding of the structure, or behavioral comparison of trajectories of the protein of interest to a homologous thermostable protein that can reveal different flexible regions [62]. Stability can also be increased by hydrophobic packing of the core, by introducing more charged amino acid side chains at the protein surface or by introducing disulfide bridges, with predictions usually based on the protein model [62]. Increasing protein stability, as an important concept in biocatalysis, has been widely studied and different algorithms have been designed that can predict beneficial mutations [68].

Structural rearrangements and dynamics can also drive activity regulation, as for example allostery (see \textbf{Protein Structural Dynamics and Influence on Catalytic Function}). Understanding dynamic regulatory mechanisms often involves the analysis of MD trajectories [69]. A deeper understanding of allosteric regulation can help in both redesigning or newly introducing allostery for regulated enzymatic activity using rational or semi-rational engineering strategies [69, 70].
Analyses within this thesis, include elucidating the influence of water movement on triterpene cyclase catalysis (see Paper III, Paper V). Based on the idea that specific water tunnels in the structure of triterpene cyclases exist [71], dynamic tunnel analyses using the CAVER algorithm [72, 73] were performed. CAVER identifies tunnels and channels in the structure that connect buried cavities, such as the active site, with the surroundings. Several programs developed for the detection of tunnels or channels in the structure (CAVER but also MOLE [74], or MolAxis [75], only CAVER is able to cluster tunnels and therefore allows detection of the tunnel dynamics [72]) generate Voronoi diagrams to analyze the geometry of the structure computationally. For the Voronoi diagram, the van der Waals radii of the atoms of the structure are approximated and adjusted to be equal to the size of the smallest atom in the structure. The atoms are treated as balls that then build the center for the Voronoi cells and the generation of the diagram. The identified tunnels represent the shortest pathway (with a certain radius) centered on the Voronoi cell walls connecting the seeding point with the bulk. The detected tunnels can then be analyzed according to the tunnel profile, the tunnel lining residues, the bottleneck radii and residues constructing the bottleneck [72, 73]. Bottleneck residues, comprising the narrowest point of a tunnel, can be hot spots for the modification of the tunnel and with that modification of the protein function or enzyme activity. Tunnels in protein structures enable transport of substrates and products to and from buried active sites, ions and water molecules and can therefore be modified to engineer an enzyme’s substrate specificity [76], enantioselectivity [77], and activity [78]. In Paper V, we show that a tunnel network in the human triterpene cyclase oxidosqualene cyclase exists. Redesign of the tunnel network by introducing single point mutations in the tunnel lining residues massively influences the thermodynamics (see Transition State Theory and Thermodynamics of Enzymatic Reactions) of the reaction, potentially due to altered water movement during catalysis. Such analyses can show how modifications of tunnel networks allows influencing enzymatic activity on an energetic level, which can be explored as potential engineering strategy.

However, CAVER analysis is purely based on the protein structure and does not detect water movement within the channels or localizes water molecules, and their energetic properties in the active site. As mentioned above, MD trajectories performed using an explicit solvent model do not only allow the analysis of the trajectory in terms of the structure but also in terms of the water movement over time. Multiple MD-based methods and algorithms to perform such analysis have been developed. One commonly used method, focusing on protein-ligand binding and therefore extensively used in drug design and development, is WaterMap [79]. WaterMap localizes hydration sites in the protein structure and computes their thermodynamic properties. Hydration sites are locations in the structure where water aggregates as detected from water densities in the MD trajectories. Excess entropy and enthalpy are calculated using ff parameters and are based on thermodynamic properties of the water molecules observed in the hydration site in comparison to water molecules.
in the bulk. WaterMap has been developed to improve ligand binding and is therefore focusing on cavities, where water shows different behavior compared to water in bulk.

Analysis of water network rearrangement and influence on triterpene cyclization, however, involved analysis of water movement and relocalization of water molecules in the previously detected water channels of a bacterial triterpene cyclase ([71], Paper III) and human triterpene cyclase (Paper V). We therefore employed two different strategies. For analysis of the water flow in the bacterial enzyme (Paper III), a biophysical analysis technique initially designed to track water movement in the Photosystem II has been adapted [80]. This algorithm originates from fluid dynamics and is designed for detecting areas with highly anisotropic movement of water (“streams”). Such directed movement of water molecules can be detected with diffusion tensor fields that describe the diffusion pattern of water molecules in a certain area (voxel) of the protein (1 Å³) over the whole trajectory. To detect the diffusion tensor, the direction of the fastest water movement in each voxel, the initial and future position of the water molecules in each voxel (at time t and t+Δt), are determined. Then, tensor elements are calculated, averaged over the time frame of the MD simulation and the diffusion tensor is computed. Voxels that show increased linear, directed water streams (movement in the detected direction needs to be at least 1.5 times faster than movement in any other direction) are identified, connected to neighboring voxels and water movement is visualized as streamlines [80, 81].

For the analysis of the human enzyme (see Paper V), we were interested in analyzing specific changes in water networks during catalysis and effects of introducing tunnel obscuring mutations. To be able to correlate specific water network rearrangements to the formation of the TS, we analyzed the MD trajectory for hydrogen-bonded water networks, using a previously designed algorithm [82]. It detects and clusters water networks, based on hydrogen bonding, starting from a seeding point. Water networks were compared for the empty enzyme without ligand bound in the active site and the structure with the product (TS-analogue) localized in the active site. We additionally analyzed and compared water networks in the CAVER detected tunnels of the Wild Type (WT) to the networks in tunnels of variants with obscured tunnels.
II. Enzyme Catalysis

Catalytic Mechanisms

Traditionally, the analysis of an enzyme’s operational mode centered around discerning the catalytic mechanism, which is mostly focused on the action in the active site [83, 84]. Enzyme-catalyzed reactions can generally be conceived as the succession of three reaction steps: enzyme-substrate binding, product formation and product release (Figure 4). Enzymes enhance the rate of the performed chemical reactions by reducing the activation energy ($E_a$) or TS free energy compared to the energy of the uncatalyzed reaction. A stabilization of TS, or decrease of $E_a$, is generally considered to increases the abundance of the TS structure, which will accelerate product formation. TS stabilization can be achieved via many different routes, both during the binding step, but also via different mechanisms during the product formation.

![Figure 4: Reaction scheme of enzyme catalyzed reactions. The enzyme (grey) binds to the Substrate (S) and forms the Michaelis complex (ES) followed by transition state formation (ES‡). The Product (P) is formed and released to yield the free, unchanged enzyme and product.](image)

Binding

The initial model for enzyme catalysis, the “lock and key” model, focusing on the binding step, was proposed by Emil Fischer in 1894 [85]. He suggested that the high specificity in both substrate recognition and product profile of enzyme catalyzed reactions arises from the complementarity of the geometric shape of the active site (the lock) and the substrate (the key). The lock and key model became the standing model of enzyme catalysis until Pauling proposed in 1946 [86, 87] that enzymes bind the activated substrate (the TS) and thereby promote the reaction. Pauling’s hypothesis was influenced by Eyring’s transition state theory, published in 1935 [88], concerning rates of chemical reactions. Based on the increasing recognition that protein structures are flexible and not static, Koshland proposed the “induced-fit model for binding” in 1958 [89]. Both, the induced-fit model and the model of conformational selection proposed around a decade later [90, 91] are the currently standing models.
for substrate binding. They both consider protein flexibility and the involved structural changes during the ligand binding step. The induced-fit model assumes that the enzyme binds the substrate and subsequently changes its conformation accordingly to fit the substrate or its TS perfectly. The conformational selection model states instead that the enzyme is periodically sampling different structural conformations and the substrate or its TS binds to the conformation that is suitable to undergo the enzymatic reaction [92, 93]. Different enzymes apply different binding strategies, mostly induced-fit or conformational selection, or a combination thereof [92].

Active site topology, stereochemistry, hydrophobicity and electrostatic profile define which molecules can bind, be brought into the TS and undergo catalysis. A reduction of $E_a$ upon binding can be achieved by a preferred binding of the substrate TS through better active site complementarity (preorganized active site) to the TS, by approximation and proper orientation of the reactants, or by bond distortion upon binding [94]. Enzyme-substrate binding usually results in a net-stabilization of the system, where hydrogen bonds, van der Waals interactions, and hydrophobic and electrostatic forces are energetically more favorable in the enzyme-substrate complex and compensate for the unfavorable energetic contributions of losing translational and rotational entropy [94].

Catalytic Principles

Enzymes can employ different chemical strategies to lower the TS free energy. Proximity effects influence most enzymatic reactions: Enzyme-substrate binding brings reactive groups in close proximity, which gives the reaction an intramolecular character. This increases the reaction rate similarly to the effect of an increased concentration of the reactants. Other principles include covalent catalysis, electrostatic effects (in the preorganized active site involving hydrogen bonding, stabilization by ion pairs or metal ions) and general acid/base catalysis [94]. Usually, a combination of different catalytic principles leads to the enzymatic rate acceleration. The enzymes analyzed in this thesis, $\alpha/\beta$-hydrolases and terpene cyclases, mostly apply covalent catalysis, acid/base catalysis, stabilization by electrostatic effects and hydrogen tunneling, which are explained in more detail below (Figure 5).
Figure 5: Catalytic mechanisms. A: Covalent catalysis implemented by a catalytic triad with six stages of catalysis. If the second nucleophile is an activated water molecule as depicted in the figure (4.), the reaction is of hydrolytic nature B: General acid catalysis as performed by class II terpene cyclases with an aspartate residue (R) as catalytic acid protonating a double bond. The transition state of the reaction is indicated in square brackets. Note that the transition states of terpene cyclizations are stabilized by cation-π interactions with side chains of aromatic amino acids in the active site (electrostatic effects, not shown). C: Reaction sequence of a proton (magenta) transfer mediated via proton tunneling through the transition state barrier. Probabilities of the localization of the proton are depicted in cyan. The probability of finding a particle/atom at a certain position can be calculated using the squared wave function that relates the probability of finding the atom in a certain position (on the substrate or product side or even in the barrier) to the mass of the particle, the particle energy and the potential energy. Instead of forming the high energy transition state (indicated in square brackets) a tunneling proton is transferring through the barrier.
Covalent Catalysis

Covalent catalysis is characterized by the substrate, or part of the substrate, transiently binding to a residue (or cofactor) of the enzyme in a covalent manner. By employing this strategy, the enzyme is providing an alternative route with a TS lower in energy instead of decreasing the TS energy of the reaction in solution [95]. Covalent catalysis using a catalytic triad (or more seldomly catalytic dyad) is mostly performed by transferases and hydrolases. This well-studied mechanism is very versatile and is involved in the degradation of different biomacromolecules, but also man-made materials, such as polyethylene terephthalate (PET) [96, 97].

A catalytic triad usually consists of a nucleophile (serine, cysteine, threonine), which initiates the reaction, a base that polarizes and eventually deprotonates the nucleophile, increasing the reactivity (histidine, lysine, serine), and an acid that positions and polarizes the base. The acid is not always necessary and is not present in catalytic dyads. Covalent catalysis can be separated into two main phases with six sub-stages (Figure 5 A). In phase one (Figure 5 A 1-3), the nucleophile attacks the carbonyl carbon of the substrate, which forces the electrons towards the carbonyl oxygen, creating a negative charge that is usually stabilized by an “oxyanion hole”. This leads to the formation of the first tetrahedral intermediate, where the substrate is covalently bound to the enzyme. The intermediate disintegrates back to reform the C=O double bond, which leads to the release of part of the original substrate, usually assisted by donation of a hydrogen from the catalytic base. The release of the second half of the substrate in phase two of the reaction (Figure 5 A 4-6) requires the formation of a new tetrahedral intermediate assisted by the attack of a second substrate (water in hydrolases). The second tetrahedral intermediate collapses in a similar manner as the first tetrahedral intermediate, leading to the release of the second substrate which completes the catalytic cycle [98].

Electrostatic Effects

The lowering of the TS energy by an electrostatic stabilization of the TS structure has been described as major component of the rate-enhancement achieved by enzyme catalysis [95]. Such electrostatic stabilization of a charged TS can be achieved, for example, by formation of ionic bonds between the substrate and amino acid side chains with acidic or basic properties, between the substrate and metal cofactors or by the formation of hydrogen bonds for the orientation of the substrate (as exemplified in Figure 4). A preorganized active site with positioned polar groups enables the formation of strong electrostatic fields (dielectric constants average around 4 F/m in enzyme active sites [99]), which has been shown to correlate with the catalytic rate-enhancement achieved during enzyme catalysis [100, 101]. The electrostatic stabilization of carbocationic TS in terpene cyclase catalysis by aromatic side chains that possess partial negative charges in the ring systems (see Terpene Cyclases) drives the reaction forward and can be decisive for the product outcome [102]. Additionally, in class I terpene cyclases (see Terpene Cyclases) that act on
phosphorylated substrates, the free pyrophosphate group, emerging after reaction initiation, can act as a counter-ion and thereby influence the reaction outcome (see *Terpene Cyclases*) [103]. Electrostatic effects are also a major contributor to catalysis performed by $\alpha/\beta$-hydrolases where, for example, the developing negative charge at the substrate oxygen is stabilized in an “oxyanion hole”. Additionally, a water molecule is activated by electrostatic effects to act as nucleophile to finalize the hydrolytic reaction (see *Covalent Catalysis*).

**General Acid-Base Catalysis**

For acid-base catalysis, side chains with acidic or basic properties participate in proton transfer reactions to stabilize the TS and thereby enhance the speed of the reaction ([Figure 5 B](#)) [104]. This can occur via the counter-balancing of developing charges, the activation of nucleophiles or electrophiles needed for catalysis, or an increase of electric interactions. Nearly every enzymatic reaction requires some type of proton transfer, involving acidic or basic groups (aspartate, glutamate, histidine, cysteine, tyrosine, lysine), where the $pK_a$ that determines the reactivity of the side chain can be perturbed by the local protein environment to enhance the acidic or basic strength [105].

**Hydrogen Tunneling**

Some enzymatic reactions involving hydrogen transfers have been described, where the reaction proceeds under seemingly impossible conditions (e.g. not sufficient energy to be able to surpass the TS). Such phenomena cannot be explained by the classical mechanics “over the barrier” model of chemical reactions that is based on the particle nature of atoms. To explain those discrepancies, the wave properties (quantum mechanics) of atoms have to be considered. In quantum mechanics “through the barrier” movements are allowed ([Figure 5 C](#)). Atoms obey probability laws, where a quantum particle can be found in different areas with some regions having higher probabilities of harboring the particle than others. This probability that can be calculated using wave functions, can penetrate the activation barrier, especially for lighter atoms, such as hydrogen. This means that there is a probability for the hydrogen to cut through and be located at the other side of the barrier, a property known as tunneling. Lighter particles, such as hydrogen/protons possess more quantum mechanical characteristics and thus can tunnel easier, which is why tunneling contributes mostly to hydrogen/proton transfer reactions. The tunneling probability is not only dependent on the height and width of the TS barrier, but also on the donor-acceptor distance. This allows enzymes to partially influence tunneling by altering the distance with structural flexibility and dynamics [106-108].

The influence of tunneling on enzymatic protonation reactions is mostly measured indirectly, for example, by the determination of kinetic isotope effects (KIE = $k_{\text{Hydrogen}}/k_{\text{Deuterium}}$). Therefore, the transferred proton (H) is replaced by deuterium (D, or tritium) and the effect on the rate of the catalyzed reaction is measured (see *Enzyme*
Kinetics). In classical mechanics, based on the particle nature of atoms, the rate for the deuterium transfer is expected to be lower compared to the rate for the hydrogen transfer because of reduced bond vibration frequencies due to the heavier mass of deuterium. These reduced vibrations for a C-D bond lead to a reduced ground state energy (of around 1.15 kcal/mol) compared to a C-H bond. Thus, a maximum KIE of 7 (based on Equation 2, with 1.15 kcal/mol lower ground state energy for the C-D bond and the assumption that \( A_H \) and \( A_D \) are similar) can be explained by classical mechanics. If tunneling is involved, much larger KIEs of up to 300 can be observed. As mentioned above, the lighter the particle, the higher the probability for tunneling, which is especially notable for the isotopes H and D, where D has twice the mass of H. In transfer reactions, where tunneling is involved, the rate for the H-transfer (\( k_{\text{Hydrogen}} \)) will thus be much faster (up to 300 times faster compared to 7 times faster for reactions without tunneling) compared to the rate of the D-transfer (\( k_{\text{Deuterium}} \)). However, also lower KIEs do not exclude a potential contribution of hydrogen tunneling to the reaction, if the KIE is temperature-independent [109]. The temperature-dependence of the KIEs for classical reactions and reactions where tunneling is involved can be explained using the Arrhenius equation (Equation 1), proposed by Svante Arrhenius 1889 [110, 111], which relates reaction rates to temperature.

\[
k = A \times e^{-\frac{E_a}{RT}} \quad \text{(Equation 1)}
\]

where \( k \) is the rate constant, \( A \) the pre-exponential Arrhenius factor, \( E_a \) the activation energy, \( T \) the absolute temperature in K, and \( R \) the universal gas constant. For analysis of the temperature-dependence of the KIE, Equation 1 can be reformulated to Equation 2.

\[
\ln \text{KIE} = \ln \frac{k_H}{k_D} = -\frac{E_{aH} - E_{aD}}{R} \times \frac{1}{T} + \ln \frac{A_H}{A_D} \quad \text{(Equation 2)}
\]

According to Equation 1, a transfer reaction following the classical over the barrier scheme, will give a straight line with a negative slope when plotting \( \ln k \) against \( 1/T \) (Arrhenius plot). The slope of a plot for a classic over the barrier mechanism based on Equation 2 depends on the difference in the activation energies (\( E_{aH} - E_{aD} \)), which is, in some cases, the difference of the different ground state energies between the two reactants (C-D bond and the C-H bond). As previously mentioned, this difference corresponds to maximum 1.15 kcal/mol leading to a maximum KIE of 7 that can be explained by classical mechanics (see above). The pre-exponential factor \( A \) describes the frequency of collisions in the correct orientation and is not very sensitive to a change from C-H to C-D, so that the \( A_H/A_D \) ratio is expected to be close to 1 for a reaction following a classical scheme.

In contrast, for a transfer reaction where tunneling is involved and the particles cross through the barrier, \( E_a \) reduces and effectively at very low temperatures or extensive tunneling, becomes 0. This results in a curved or temperature-independent slopes (for both plots based on Equation 1 and Equation 2) and with that temperature-
independent KIEs (when $E_{a,H} = E_{a,D}$). Classical over the barrier moves become less likely at very low temperatures, so that under those conditions, tunneling predominates the reaction, which gives slopes that eventually becomes 0 at low temperatures, giving $A_H/A_D$ values above 1 and close to the KIE [112]. Analysis of the KIE, the temperature-dependence of the KIE together with analysis of the Arrhenius factor is used in Paper II to analyze the contribution of tunneling to the hydrogen transfer reaction performed by a class II diterpene cyclase (see also Terpene Cyclases).

Usually, enzymes use a combination of different catalytic principles. The non-canonical $\alpha/\beta$-hydrolase patatin (Paper I) relies on covalent catalysis performed by a catalytic dyad, which also involves acid/base catalysis and electrostatic effects. Paper II to V, revolve around class II terpene cyclases, which require a specific pre-folding of the substrate upon binding and capitalize on general acid catalysis executed by a conserved aspartic acid residue in combination with a preorganized electrostatic environment in the active site. In Paper II, we additionally demonstrate the importance of tunneling in a class II diterpene cyclase based on temperature-independent KIEs.

**Enzyme Kinetics**

Generally, the study of enzyme catalysis and the influence of different catalytic mechanisms involves the determination of the reaction rate - an enzyme's kinetics. Varying conditions and analysis of their influence on the rate does not only allow determination of the employed catalytic mechanism, but can also give insight into activity regulation, metabolic role and is crucial in drug design and development [97, 113, 114].

The kinetic parameters $K_M$, the Michaelis constant that describes how much substrate is needed to reach half maximum velocity, and $k_{cat}$, the rate constant of the rate determining step, can both be determined from measuring the initial conversion rate ($v_0$, at $t=0$) of product formation at different substrate concentrations [$S$]. The hyperbolic correlation of increasing initial rate with increasing [$S$] was first described by Victor Henri in 1902, who also paved the way for the later mathematical description of enzymatic reactions by Leonor Michaelis and Maud Menten in 1913 [115]. The description is based on separation of the reaction in two different steps - formation of the ES complex (described by the constants $k_1$ and $k_{-1}$ for association and dissociation, respectively) and product formation (Equation 3, $k_{-2}$ (backwards reaction) is negligible at $t=0$).

\[
E + S \rightleftharpoons_{k_1}^{k_{-1}} ES \rightarrow E + P \quad (\text{Equation 3})
\]
The initial conversion rate of this reaction is dependent on the concentration of the ES complex and the rate of product formation \( k_2 \) and can be described by Equation 4.

\[
\nu_0 = k_2 [\text{ES}]
\]

\textbf{(Equation 4)}

The concentration of [ES] is unknown. But under the assumption that [ES] is stable (steady-state assumption), the association of ES from free enzyme E and free substrate S (described by the rate constant \( k_1 \)) and the dissociation of the ES complex (described by the constants \( k_1 \) and \( k_2 \)) are equal (Equation 5).

\[
k_1 [\text{E}][\text{S}] = k_1 [\text{ES}] + k_2 [\text{ES}]
\]

\textbf{(Equation 5)}

However, to solve this equation for [ES], the concentration of the free enzyme [E] needs to be derived using the total enzyme concentration \([E_{\text{tot}}]\) according to Equation 6.

\[
[E_{\text{tot}}] = [\text{E}] + [\text{ES}]
\]

\textbf{(Equation 6)}

Solving Equation 6 for [E], substitution into Equation 5 and rearrangement to solve for [ES] gives Equation 7 as description for the [ES] concentration.

\[
[\text{ES}] = \frac{[E_{\text{tot}}][S]}{k_1^{-1} + k_2^{-1} + [S]}
\]

\textbf{(Equation 7)}

The relation of the rate constants \( \frac{k_1^{-1} + k_2^{-1}}{k_1} \) is summarized in the Michaelis constant \( K_M \), and the so defined concentration of the [ES] complex can be substituted into Equation 4 to give Equation 8, a description of the enzymatic velocity.

\[
\nu_0 = \frac{k_2 [E_{\text{tot}}][S]}{K_M + [S]}
\]

\textbf{(Equation 8)}

\( k_2 \) corresponds to the catalytic rate \( k_{\text{cat}} \) and multiplied by \([E_{\text{tot}}]\) gives the maximum velocity \( V_{\text{max}} (k_{\text{cat}} \times [E_{\text{tot}}] = V_{\text{max}}) \). Substitution of this information into Equation 8 gives the final form of the Henri-Michaelis-Menten equation (Equation 9) that describes the velocity of enzymatic reactions. It is valid under the assumption that E + S and ES are under quasi-equilibrium, that the total enzyme concentration \([E_{\text{tot}}]\) is constant and much lower than [S], and that the reaction follows saturation kinetics.

\[
\nu_0 = \frac{k_{\text{cat}} [E_{\text{tot}}][S]}{K_M + [S]} = \frac{V_{\text{max}} [S]}{[S] + K_M}
\]

\textbf{(Equation 9)}

This mathematical model of enzymatic reactions allows for the determination of the catalytic efficiency of unimolecular reactions measuring the product formation over time using different starting substrate concentrations. If the reaction can be saturated and the rate is approaching the maximum rate \( V_{\text{max}} \) asymptotically, \( k_{\text{cat}} \) and \( K_M \) can both be extracted so that effects of different conditions on the efficiency can be determined to be on either \( k_{\text{cat}} \) or \( K_M \) (Paper II). Generally, \( k_{\text{cat}}/K_M \) represents a common measure for an enzyme’s efficiency as it describes both substrate binding as well as TS stabilization. Therefore, under unsaturated conditions, \( k_{\text{cat}}/K_M \) can be
used instead of the first-order rate constants. Under our experimental conditions, triterpene cyclases (studied in Paper III, Paper IV and Paper V) as membrane enzymes with hydrophobic substrates were unsaturatable, probably because both enzyme and substrate have to be solubilized in micelle-forming detergent. Hence, in Paper V, $k_{\text{cat}}/K_M$ was used to compare designed variants to WT.

For more complex reactions, such as kinetic models involving more than one substrate Equation 9 has to be modified. Kinetics of bimolecular reactions that follow a ping-pong bi-bi reaction scheme (which would describe covalent catalysis involving catalytic triads employed by hydrolases), are mathematically described by Equation 10, where $S_1$ and $S_2$ represent the two different substrates.

$$V = \frac{k_{\text{cat}}[E][S_1][S_2]}{K_M^S [S_2] + K_M^S [S_1] + [S_1][S_2]}$$  \hspace{1cm} (Equation 10)

Often however, determination of kinetic parameters of bimolecular reactions can be simplified by measuring the rate under conditions where one substrate is in excess ($[S_1] >> K_M$, naturally the case for hydrolases where water as substrate usually is in excess) and the concentration is constant. The reaction rate is then showing pseudo-first order behavior and is only dependent on varied concentration of the other substrate ($[S_2]$), which enables estimation of kinetic constants using the traditional Michaelis-Menten equation (Equation 10).

### Transition State Theory and Thermodynamics of Enzymatic Reactions

While kinetic studies investigate how fast a catalytic process occurs, thermodynamic studies generally explore if the reaction does occur ($\Delta G < 0$). They can also be used to determine different energetic contributions to the Gibbs free energy ($G$) (Figure 6), a thermodynamic state function that is composed of an enthalpic ($H$) and entropic ($S$) part (Equation 11).

$$\Delta G = \Delta H - T\Delta S$$  \hspace{1cm} (Equation 11)

This concept of both an enthalpic and an entropic contribution to the free energy has been incorporated into the description of the temperature-dependence of chemical reactions in Eyring’s transition state theory (Equations 12 and 13, with $k$ as the rate constant, $\kappa$ as transmission coefficient that describes the fraction of the transition state complexes that proceeds to become product and is here assumed to be 1, $T$ as the absolute temperature in K, $k_B$ as the Boltzmann constant, $h$ as the Planck constant and $R$ as the universal gas constant [88, 116]). Under the assumption that there is a quasi-equilibrium between the substrate and the TS, it is possible to decompose the enthalpic ($\Delta H^t$) and entropic ($\Delta S^t$) contributions to TS formation (Figure 6).

$$k = \kappa \frac{k_B T}{h} e^{-\frac{-\Delta G^t}{RT}}$$  \hspace{1cm} (Equation 12)
\[
\ln \left( \frac{k}{k_B T} \right) = \frac{-\Delta H^\ddagger}{RT} + \frac{\Delta S^\ddagger}{R}
\]
(Equation 13)

Equation 12 is very similar to the Arrhenius equation (Equation 1), that also describes the temperature-dependence of chemical reactions, but it is based on the Gibbs free energy of activation (\(G^\ddagger\)) instead of \(E_a\), which enables derivation of \(\Delta H^\ddagger\) and \(\Delta S^\ddagger\). Based on Equation 13, measuring the temperature-dependence of the enzymatic reaction and performing linear regression of \(\ln \left( \frac{k}{k_B T} \right)\) against \(1/T\) reveals the activation enthalpy \(\Delta H^\ddagger\) (-slope \(\times R\)) and the activation entropy \(\Delta S^\ddagger\) (intercept \(\times R\)) \cite{116} (Paper V). It should be mentioned that also hydrogen tunneling contributions can be incorporated into TS theory via the pre-exponential factor \(\kappa\) \cite{117}. Depending on which kinetic constants \((k)\) are measured and inserted in Equation 8, energetic contributions can technically be determined for ES-complex formation \((K_M)\), the catalytic steps (combined in \(k_{cat}\)) or the overall catalytic efficiency \((k_{cat}/K_M,\ Figure 6)\).

Figure 6: Energy profile of an enzymatic reaction (black) in comparison to the uncatalyzed reaction (grey) and associated changes in the respective Gibbs free energies (colored grey for uncatalyzed reaction, green, cyan and blue for catalyzed reaction). For the enzymatic reaction, substrate (S) binds to the enzyme (E) to form the enzyme-substrate-complex (ES), which is accompanied by a usually favorable change in free energy \((\Delta G_{KM})\) associated with the kinetic parameter \(K_M\) (cyan). Upon binding follows the transition state formation \((ES^\ddagger)\), which is the high free energy-intermediate associated with the free energy change \(\Delta G_{cat}^\ddagger\) and the catalytic rate constant \(k_{cat}\) (blue). Alternatively, \(k_{cat}/K_M\) can be used to determine \(\Delta G_{cat}^{\ddagger,K_M}\) (green), which then includes both the ES-complex formation described by \(K_M\) and the catalytic rate constant \(k_{cat}\). The uncatalyzed reaction passes through a high free energy transition state \((S^\ddagger)\) that is associated with the rate \(k_{non}\). Enzymes lower the transition state energy (by stabilizing the TS structure by \(\Delta \Delta G_{cat}^{\ddagger}\)) to speed up the equilibrium between substrates and products, but they do not change the equilibrium of the reaction determined by \(\Delta G\). \(\Delta G\) as a state function is pathway-independent and does not change in the presence of a catalyst. If determined under standard conditions, the energy change is referred to as \(\Delta G^0\).
However, most often the focus of scientific studies lies either on enzyme-ligand binding, essential in drug design and development [118], or on the performed chemistry in biocatalysis [71, 119]. Comparing the thermodynamic parameters of the uncatalyzed reaction to those of the catalyzed reaction can give valuable insight into enzymatic strategies for lowering the TS energy [120, 121]. For instance, a lowering in enthalpy can usually be attributed to electrostatic effects and hydrogen bond formation. A (rarer) increase in entropy is often ascribed to proximity, dynamic effects or to solvent contribution (see below). However, it often is a challenge to decipher the contributions and effects of the different parts of the system (enzyme, substrate, solvent). Analysis of entropic and enthalpic contribution to the energy barrier of an enzyme catalyzed reaction can not only be used to compare uncatalyzed and catalyzed reactions but can also be informative when comparing enzyme variants with mutations that supposedly alter catalysis (Paper V).

Enthalpic Catalysis

Enthalpy is a thermodynamic potential describing the systems internal energy at a certain pressure and volume. It is often referred to as the heat of the reaction, as it can be increased or decreased by absorbing or releasing energy in form of heat. Enthalpy in enzymatic reactions is mainly connected to bond formation and breaking and electrostatic effects, such as hydrogen bonding or ionic interactions [95]. Such enthalpic, electrostatic effects in a preorganized active site have in fact been discussed as major contributors to the reduced $E_a$ of enzyme-catalyzed reactions as described by Warshel based on computational studies [95, 122]. A series of comparisons of activation parameters of uncatalyzed and enzyme catalyzed reactions revealed that indeed many enzymes increase the reaction rate mainly by lowering the activation enthalpy ([123-128], see Table 4 in [129]). Differential activation enthalpies when comparing uncatalyzed and corresponding catalyzed reaction, reactions of an enzyme with different substrates, or enzyme variants to wild type is thus an indication of a difference in steric and electrostatic interactions in the system.

Because of the prevalence of described enthalpically regulated catalysis and the advantage of temperature-dependence, enthalpic rate-enhancement has also been described as the evolutionary origin of the catalytic power associated with enzyme catalysis [4, 129]. Wolfenden et al. showed in several publications that most enzymatic rate-enhancement over the corresponding uncatalyzed reaction comes from a decrease of the enthalpy of activation. The decrease in enthalpy of activation to an average of 12 kcal/mol (catalyzed) from 28 kcal/mol (uncatalyzed) allows a massive increase of the reaction rate ($\sim 10^{12}$) and entails an automatic increase of rate-enhancement with the decrease of the temperature on a cooling earth [4, 129, 130], according to **Equation 12 and 13.**
Entropic Catalysis

Entropy is a more complex thermodynamic concept that describes the disorder of a system, which translates to the number of accessible states of the system. The higher the number of accessible microstates that can be populated, the higher and more favorable the entropy of the system. The different configurations arise from translational, rotational and vibrational degrees of freedom. According to Equation 11, a lowering of the activation energy can not only be achieved by a decrease in enthalpy but also by a favorable entropic contribution to catalysis, which corresponds to an increase in entropy (Figure 7).

In fact, before Warshel’s description of electrostatic effects and preorganized active sites as major contributors to catalysis [95, 122] and Wolfenden’s theory that enzymes evolved to lower the enthalpy of activation (see above), entropy had been anticipated as dominating factor to lower the activation barrier, both by proximity effects and the so-called Circe-effect. Proximity effects during the enthalpically favorable enzyme-substrate binding, allow the enzymatic reaction to proceed as a unimolecular reaction and the disadvantage in terms of losing translational and rotational degrees of freedom is already paid for upon binding [131]. The favorable entropic effect on catalysis of the substrate ground state destabilization - the Circe-effect- has been described by Jencks in 1975 [132]. The reactants are tightly bound and their shape changes less than in solution – corresponding to a reduced entropy.

However, many typical examples for entropic catalysis, where unfavorable entropic effects in binding have been believed to cancel out with the favorable enthalpic effects in binding and thereby decrease the entropic disadvantage in the activation barrier, have been shown to operate by enthalpic, electrostatic catalysis instead [133, 134]. Moreover, a comparison of the uncatalyzed to the corresponding enzyme-catalyzed reaction (that can proceed through different transition states) often shows a negative, unfavorable activation entropy, corresponding to a loss of translational and rotational motions when going through the TS [134]. However, such experiments can also reveal rare, less unfavorable, or even favorable entropic contributions to TS formation in the catalyzed reaction, as for example described for peptide bond formation at the ribosome, terpene cyclase squalene-hopene cyclase or cytidine deaminase [71, 120, 135]. It can be difficult to unravel the origin of the entropic contributions that can arise from enzyme, substrate and/or solvent (Figure 7). For enzymatic systems, most changes in degrees of freedom moving along the reaction coordinate are supposedly associated with solvent molecules (see Protein Hydration and Solvent Influence on Protein Function). Nevertheless, also the protein and substrate contribute to the entropy of the system, which can be associated with for example the conformational degrees of freedom of the protein [136, 137] (see Protein Structural Dynamics) or changes in bond vibrations of the reactants or products (Figure 7) [131].
Figure 7: Possible entropic effects in enzymatic reactions. Entropy can affect enzyme-substrate binding (left, going from free enzyme and substrate to the ES complex) and catalysis (right, going from the ES complex to the TS) in a favorable (top) and unfavorable (bottom) fashion. Entropic contributions can thereby arise from enzyme, substrate or solvent. Consequences of the entropic effects on the binding free energy ($\Delta G_{\text{ES}}$) or the free energy of activation ($\Delta G^\ddagger$) and the reaction rate ($k$) are labelled with arrows. For example, a favorable entropic effect in binding will lower the energy of the ES complex, which corresponds to a decrease in $\Delta G_{\text{ES}}$ (↓) and an increase in $E_\text{r}$, which will be concomitant with a decrease in overall rate, labelled as $k_m$ (see Figure 6). The influence of changes in the structural flexibility [136, 137] and water release upon enzyme-substrate association [138] have been described to influence binding energetics in a positive way. On the other hand, a decrease in dynamics, a destabilization of the ground state, a loss in translational and rotational entropy by forming one complex from several molecules [131] and trapping water molecules during the binding process [139] can influence binding unfavorably. However, it is important to keep in mind that usually overall binding is energetically favorable so that these unfavorable entropic effects are counterbalanced by favorable enthalpic contributions. For example, a trapped water molecule is entropically unfavorable but enables enthalpically favorable hydrogen bonding between binding partners [120]. Favorable entropic effects in the catalytic steps (usually referring to $\Delta G^\ddagger$ based on $k_{\text{cat}}$ but can also refer to $\Delta G^\ddagger$ based on $k_{\text{cat}}/K_M$) can arise from increased configurational entropy, large conformational changes or dynamic flexibility [140, 141]. Increased bond vibrations in the TS or product and the fact that a bimolecular reaction can proceed as unimolecular reactions [131] can have favorable effects on catalysis. Solvent release upon TS formation [71] and water network reorganization (Paper V) have additionally been discussed as favorable entropic contributors to catalysis. On the other hand, a decreased dynamic flexibility, trapped water molecules and unfavorable water molecule reorganization upon TS formation can have unfavorable effects on enzyme catalysis.

Most examples where favorable entropic effects in the association step have been observed are linked to an allosteric configurational change, structural changes and desolvation of the protein (see Table 6 in Supporting Information of Paper V).

A series of similar reactions, such as a comparison of the association of different substrates with the same enzyme or comparison of WT to variants, often shows
compensatory effects in entropic and enthalpic contributions to catalysis. This is called enthalpy-entropy compensation, a phenomenon not only observed for biochemical reactions [71, 142]. A more favorable enthalpy of activation will usually come hand in hand with a less favorable entropy of activation and vice versa, to an extent that the effects compensate each other. For example, an enthalpically more favorable TS binding leads to a more rigid, entropically less favorable TS [142]. The existence and source of the widely observed compensatory effects have been discussed vividly [143, 144] and the importance of solvation has been mentioned [145, 146]. Despite the fact that the physical background for the compensatory effects are not completely understood, the concept has been used in drug discovery and enzymology to investigate the relatedness of different binding events or catalytic reactions [71, 142, 147].

A thermodynamic analysis of enzymatic reactions can thus give insight into enthalpic and entropic contributions to catalysis and reaction mechanisms and can further reveal influence of protein dynamics and solvent movement on enzymatic activity.

The entropically driven triterpene cyclization performed by human oxidosqualene cyclase has been analyzed in Paper V: a combination of in vitro and in silico analyses of WT and designed mutants led to the conclusion that water network reorganization allows an entropically-driven catalysis (see also Protein Hydration and Solvent Influence on Protein Function).

**Protein Structural Dynamics and Influence on Catalytic Function**

Many studies elucidating enzyme mechanisms often focus on the active site composition (both structure and nature of amino acids forming the active site) and on understanding the underlying chemical mechanisms that lower the activation barrier. This has been and still is analyzed with kinetic studies, comparing WT and active-site-mutants (see Enzyme Kinetics) [148, 149]. The active site, however, constitutes often just a small portion of the protein (around 10-20% of the volume) [150] and the purpose of the large size of enzymes has been fascinating protein-biochemists for decades. The influence of the protein structure outside the active site on catalysis, besides putting the catalytic residues in place, is often connected to protein structural dynamics and associated solvent motions. Such contributions of protein or solvent movement on catalysis can be detected by a favorable entropic effect on catalysis (see Entropic Catalysis).

Generally, catalytic function requires balancing between structural flexibility and stiffness. Flexibility is necessary to allow substrate binding, TS formation, correct positioning of reacting groups, and product release. Stability keeps the reacting groups in proximity and allows correct interaction between enzyme and substrate. A direct link between structural motions (fast and slow (pico- to millisecond)) that can induce rearrangements (local and global) and the catalytic activity has been observed
Structural movement can affect catalytic activity in a direct way by inducing motion of participating groups (such as reacting groups in hydrogen tunneling, see *Hydrogen Tunneling*) or by rearrangements that enable the substrate to enter the active site (for example side chain or loop movements that allow unhindered substrate entrance to the active site) [102]. Movements can also influence protein function in a more subtle way by, for example, enhanced configurational entropy [140] or fluctuations in protein dynamics that have been shown to be intrinsic properties of the structure [152]. Protein dynamics and altered structural entropy have additionally demonstrated to be crucial in protein activity regulation via allostery [153-155].

Despite the strong coupling of protein and solvent movement [156, 157], most studies focus on protein dynamics to explain regulatory mechanisms such as allostery regulation. Advances in experimental techniques, such as Nuclear Magnetic Resonance (NMR) [158], Cryo-Electron Microscopy (EM) [159], Liquid Chromatography coupled with Mass Spectrometry (LC-MS/MS) [160] (*Paper V*), and TeraHertz (THz) spectroscopy [161] allow a deeper understanding. Experimental analysis in combination with computational tools [82, 162] (*Paper V*), can help to untangle protein and solvent contribution to enzyme activity and regulation.

**Protein Hydration and Solvent Influence on Protein Function**

The main project of this thesis involved elucidating the influence of water network reorganization on terpene cyclase catalysis (see *Terpenes and Terpene Synthases*). The effect of protein hydration and solvent dynamics on enzyme catalysis is rarely studied. Therefore, the following chapter is giving an overview of solvent influence on proteins and protein function in general, information that can give insight into its potential influence on enzyme catalysis as well.

Despite water being the “matrix of life” [163] that contributes to all reactions in the cell, the thermodynamic consequences and influence of water and water dynamics on protein function are not completely understood and often neglected. The unique properties of water, such as the high boiling and freezing points and the high surface tension, consequences of the formation of hydrogen bonds, enable the prevailing water-solvent-based biochemistry on earth. In the earlier days of enzymology, water was seen as a passive solvent. Nowadays, water is thought of as an active contributor that governs the structure, stability, dynamics, and activity of biomolecules and thereby dictates life in the cell [157]. The concentration of macromolecules in the cell is around 400 g/L [164], making the cell a very crowded place. As a consequence, two molecules in the cell will be only 1-2 nm apart, separated by only a few layers of water [165]. A point of discussion is the behavior of these water molecules in the cell compared to bulk water behavior. It is often assumed and stated that water in the cell is so confined that it displays slowed down reorientational dynamics and a lower average coordination number, giving the cytoplasm a gel-like consistency [165, 166].
However, in vivo NMR measurements and neutron scattering experiments show that around 85% of the water molecules in the cell behave like bulk water molecules and only the water surrounding solutes shows differing behavior [167, 168]. Even in perturbed water networks, the hydrogen bonds are constantly rearranging, influencing the structure, function, and dynamics of biomolecules and other solutes in the cell. Thus, water is not a passive diffusion medium, but it interacts with and determines the capability of the solvated molecules (Figure 8).

Figure 8: Water-influence on protein structure and function. A: Water exclusion can drive protein folding via the hydrophobic effect. A hydration layer surrounding the protein is formed that stabilizes the structure and is important for interaction with other molecules by either enabling interaction by water expulsion or by providing hydrogen bonding interactions. B: Water is filling clefts and pockets on the in- and outside of the protein. Internal water molecules mediate interactions and stabilize the structure. C: Water is part of enzyme catalysis on several levels as exemplified here by driving protein-substrate-binding through, for example, entropic gain upon water release (for more details see Water in Enzymatic Reactions). Protein in grey, water in elemental colors, substrate in blue.
Water Drives Protein Folding

One of the most profound influences of water on many protein structures and functions is the water-driven folding of the amino acid chains to their 3D tertiary structures (see Enzyme Structure). The strong polarity (dielectric constant = 80 F/m) and consequent hydrogen bonding properties of water molecules have a severe impact on the arrangement of water networks around apolar solutes, such as hydrophobic amino acid side chains, upon solvation. The interaction between water molecules is more favorable than interacting with hydrophobic side chain residues. This results in the water molecules arranging in a cage-like structure around the particle, which constitutes to a hypothetical entropic cost. If then two hydrophobic molecules are solvated in close proximity, “entropically unhappy” water molecules can be liberated if the two molecules are surrounded by a single cage - the so-called hydrophobic effect (as depicted in Figure 8 A). The association of the apolar side chain residues to form the hydrophobic inner core of the protein, i.e. the folding of the amino acid chain, would thus be driven by a favorable entropic contribution. The general assumption of the hydrophobic effect as a pure entropic effect by water expulsion, as initially stated by Frank and Evans [169], is more and more questioned [170]. Large changes in both enthalpy and entropy seem to drive the hydrophobic association of particles [170]. Protein folding by the hydrophobic effect seems to be more complex than initially postulated but connected to solvent rearrangement and movement [171].

The Hydration Layer – Water Surrounding the Protein

The folded solvated protein is surrounded by a hydration layer that also fills gaps, clefts, and potentially exposed active sites of enzymes. It forms a protective layer by maintaining the local conformational dynamics of the protein [172] and it determines the function of biomolecules [173]. Hydration water - also referred to as biological water [174] - is often analyzed in comparison to water behavior in bulk solvent. Changes in water behavior induced by the protein surface are still discussed vividly. Overall water density, relaxation times, and dynamics of the hydration layer are believed to be perturbed in comparison to bulk water. It has been shown that water in the hydration layer (around 5 Å, 2 layers of water [173]) shows a 10% increase in density due to shortening of water-water distance and an increased coordination number - the global-surface density effect, induced by the generated electrostatic field on the protein surface [175]. The dynamics of the hydration layer are mainly dependent on the amino acid composition, the electrostatic field, the side chain movement [176], and also the topology [175, 177] of the surface. Water relaxation times can be faster [178] or slower than in bulk (3–5 times [157]) and range from a few hundred femto- to a few hundred picoseconds [179]. Counter-intuitively, water in proximity to hydrophobic residues can show slower relaxation times than water close to hydrophilic residues [177] or vice versa [176, 180], leading to the conclusion that mainly geometrical surface topology governs relaxation times and not the type
of residue [181]. This also demonstrates the heterogeneity of biological water dynamics that are unique to a protein as they are dependent of the surface composition and topology. Protein hydration of peripheral membrane proteins, such as triterpene cyclases (Paper III, Paper IV, Paper V), is different compared to that of soluble proteins. The phospholipid head groups of the membrane strongly interact with the water molecules, inducing water retardation (3-8 times slower relaxation than in bulk) up to 10 Å above the membrane, including the hydration layer of the solvated protein. The observed entropy-gradient of perturbation perpendicular to the membrane is stronger than the observed retardation in the hydration shell of soluble proteins (2.5 times) [181].

The importance of the hydration layer of proteins for their function is gaining increasing attention [182] and it has been implied that instead of studying the protein structure as it is, the "active volume" of the substance (van der Waals surface) and the consequences on the water networks and dynamics surrounding them should be considered. This is because hydration water mediates interactions between molecules and influences protein activity on several different levels [165]: it has, for example, been shown that water dynamics on the surface of the molecular chaperone GroES are on average faster compared to other protein surfaces, which enables unhindered interactions with proteins that need folding assistance by the chaperone [178]. More generally, all enzymatic activity seems to be dependent on the presence of the hydration layer that ensures that the enzyme keeps its structural integrity and flexibility. Although there are many benefits of using enzymes in organic solvent [183], it has been shown that enzymes usually benefit from an increased concentration of water in the reaction [184], demonstrating the general importance of water for catalysis.

Overall, water in proximity to the protein surface seems to be altered on many different levels, which can affect thermodynamic properties. However, this applies not only to water on the outside of the protein, but also to internally captured water molecules.

Internal Water Molecules – Solvent Dynamics

Despite the expected entropic gain upon water exclusion, proteins are not completely dry and desolvated [185]. The analysis of several crystal structures revealed that 30-60 vol% of the protein crystal is water [186], including water on the out- (The Hydration Layer – Water Surrounding the Protein) and inside of the protein structure. The water captured inside the protein structure can be divided into structurally fixed and loose, flexible water molecules. The former rigid water molecules are well hydrogen bonded, mediate structural interactions between amino acids [187] and thus have been described as the “twenty-first” proteinogenic amino acid [180]. They are active components of the structure, stability, and dynamics and it has been demonstrated that inclusion of water molecules into the modelling of protein structures based on amino acid sequences is crucial for correct structure
prediction [188]. Such localized, structured water molecules can be observed by high resolution (around 2 Å) static X-ray structures and by NMR dispersion. They show residence times, a measure of how much time a water molecule spends in the same space, on the millisecond timescale [157] and can exchange with the bulk solvent within microseconds [185].

However, not all water localized inside the protein body is that structured. The looser, flexibly bound water molecules can only be detected by density or relaxation properties using neutron scattering [189], dielectric spectroscopy [190], NMR [191] or via computational methods such as MD simulations [192, 193] (see Computational Methods for Enzyme Engineering). They show exchange rates on a nanosecond timescale [157] (Paper V). As opposed to the structural water, internal flexible water molecules can have higher entropy than in the bulk environment [157]. The localization and environment of the water determine the thermodynamic properties of such internal water molecules [194].

The direct influence of perturbed water dynamics on protein function has, for example, been shown for water molecules in the central cavity of the SecY translocon. Water dynamics showed strongly retarded behavior in diffusive motions and rotational dynamics compared to bulk water, leading to a dipole alignment of the water molecules that is crucial for the SecY function of folding and partitioning of membrane proteins to the membrane [195]. Other studies showed that water movement in G Protein-Coupled Receptors (GPCRs) determines receptor activity [196, 197]. A more complex, indirect influence of water movement on protein function comes from the connection of solvent and protein dynamics. It has been observed that the water molecule exchange rate with the bulk solvent correlates with slow rearrangements of the protein matrix, giving the first indication that protein and solvent movement could be intertwined. Additionally, correlation of the thermal dependence of protein and water motions at low temperatures by a few Å and in the pico- to nanosecond timescale [198], together with the discovery that the entropy-driven protein dynamics underlie internal water rotational dynamics, allowed the conclusion that some protein motions are slaved by solvent dynamics. Movement of surface hydration water induces the motion of surface residue side chains, which then transmits to the protein core and evokes slower movements of the protein backbone - dynamic motions necessary for the proper function of many proteins (Protein Structural Dynamics and Influence on Catalytic Function). However, the general applicability of solvent-slaved protein dynamics has also been questioned, leading to a proposed model where motions in proteins can either be slaved or be independent of the solvent movement [157, 199].

Water in Enzymatic Reactions

Water is essential for protein folding, structure, stability, dynamics, and solubility, but as discussed below can also be a key player in all steps of enzyme catalysis - binding, the catalytic step, product release - and activity regulation (Figure 9).
Figure 9: Water contribution to enzyme catalysis. A: Water influences enzyme-substrate binding on several levels. Partial or complete desolvation of the binding pocket and the substrate upon association can give favorable energetic effects. Captured water molecules can support molecular recognition through hydrogen bonding or can aid in shaping the active site for correct substrate arrangement. B: Water can be chemically involved in catalysis as, for example, substrate itself (hydrolases). It can also mediate protonation, stabilize transition states or provide an entropic driving force by being expelled or relocated (Paper V). C: Water movement has been shown to be involved in activity regulation in, for example, allosteric networks or by inducing configurational changes that enable catalysis. Enzyme in grey, water in elemental colors, substrate in blue, hydrogen bonds in magenta, product in purple, allosteric binder in green.
Protein-Ligand Binding

Water is acknowledged as an integral part of most protein-protein, protein-DNA and protein-ligand interactions. The interaction of an enzyme with its ligand or substrate entails a partial or complete desolvation of the ligand and the binding site and a reorganization of the water-networks surrounding both binding partners (Figure 9A). The free energies of the water molecules localized in the active sites differ markedly from the free energies of water molecules in bulk, depending on the pocket environment (see Internal Water Molecules – Solvent Dynamics) [194]. Side chains in apolar pockets cannot provide hydrogen bonding interactions, leaving the water molecules in a potentially enthalpically and/or entropically unfavorable state. Upon ligand binding, displacement or rearrangement of these water molecules that mostly form hydrogen bonds amongst themselves, can be energetically favorable with several kcal/mol and can thus increase protein-ligand affinity. An example for an entropy-driven protein-ligand binding event that follows the classical view of an entropic hydrophobic effect is the binding of ligands to thrombin. Increasing size and hydrophobicity of the ligand has been shown to correlate with the number of water molecules expelled upon binding, which coincides with an increasing entropy of binding (up to around 5 kcal/mol) [138]. However, as described above (see Water Drives Protein Folding), the energetic origin of the hydrophobic effect is not necessarily entropic. An example for an enthalpy-driven association event is the binding of ligands to human carbonic anhydrase [200]. Reorganization of well-ordered water networks in the hydrophobic binding pocket leads to an enthalpy-dominated process.

In contrast, polar pockets do provide hydrogen bonding interactions between amino acid side chains and water molecules, leading to enthalpically favored water localizations that are kept in place by three or four hydrogen bonds. Instead of being expelled upon ligand-binding, these water molecules can mediate an enthalpy-driven enzyme-substrate recognition and binding, such as observed for nevirapine binding to HIV1 reverse transcriptase (one bridging water molecule accounts for around 7 kcal/mol in binding) [201]. It has been shown that up to 21 water molecules can be ligand-bound in the active site [202], either bridging interactions of substrate with the active site lining side chains or they function as extensions of the active site lining to enable correct folding of hydrophobic substrates [203]. However, these structured active site water molecules usually possess an unfavorable entropy. The theoretical entropy increase upon release of such a highly ordered water molecule from the active site to the bulk has been determined to be 7 cal mol⁻¹ K⁻¹, around 2.2 kcal mol⁻¹ at a physiological temperature [204].

A delicate balance of enthalpic and entropic contributions, majorly influenced by the localization and movement of the involved water molecules, determines the energetics of enzyme-substrate binding. Energetics of the binding event are thus dependent on the environment and very specific for each interacting pair. It has, for example, been shown that water network-perturbating mutations lead to altered
thermodynamics of ligand binding in human carbonic anhydrase [118]. Both entropy and enthalpy of binding were affected by the introduced mutations, although most analyses (different enzyme variants and ligand combinations) showed the same trend with an unfavorable change in binding enthalpy and a favorable change in binding entropy compared to WT. This is probably attributable to a strengthening effect on water-networks by the introduced mutations. This study furthermore highlights the importance of water molecules and network rearrangements in the energetics of the binding process.

The understanding of the immense impact of water displacement or relocation on the free energy change upon ligand binding is not only of fundamental biochemical interest but also became a standing pillar in modern, computer-aided drug design [205] (see also Computational Methods for Enzyme Engineering). The binding of a drug, often enzyme inhibitors, that replaces water molecules and additionally mimics the water-active site interactions, will have a more favorable binding free energy. This leads to a tightly bound ligand, which is generally a more effective drug. Small changes in the water-hydrogen-bonding network can have large impact on the interaction energy, such that the binding free energy is dominated by the contributions of the water instead of direct interactions between the ligand and the binding pocket [206]. Programs like WaterMap, which can determine the thermodynamic properties of water molecules in the binding pocket, can give reliable information about the increase in affinity upon displacement of active site water by the ligand [207] (see Computational Methods for Enzyme Engineering). Not only the displacement of single water molecules affects the binding affinity, but also rearrangement or stabilization of water-networks can direct protein-ligand binding [208]. Water dynamics in enzyme-ligand binding, coupled water-protein motion that assist enzyme-substrate interaction [209] and also hydration funnels mediating ligand binding have been described [210].

Water molecules do not only determine the energetics of the enzyme-substrate interaction. Different water networks can also mediate different active site-ligand interactions and accordingly allow binding of different substrates, increasing the catalytic scope of a biocatalyst [211, 212]. Water involved in protein-ligand binding thus can affect catalysis on several levels: it drives ligand binding and unbinding kinetics [213], binding thermodynamics and active site water often determines substrate recognition [214], specificity [215], enantioselectivity [216] and product profile [203] of an enzyme.

**The Catalytic Step**

As discussed above, water molecules are usually expelled when they do not mediate binding or have catalytic function. This is often due to energetic or steric reasons. Additionally, a dry protein environment has many advantages for enzyme catalysis, such as enabling stronger electrostatic fields (dielectric constant in enzyme active sites average around 4 F/m [99]) or shielding of unstable intermediates from being
quenched by active site water molecules. However, analysis of nearly 400 protein-ligand structures has shown that there are on average 4.6 water molecules per protein-ligand interface with a wide diversity of 0 to 21 water molecules per structure [202], usually depending on the composition of the catalytic pocket. These remaining internal water molecules can either be involved in catalysis, they can contribute to the stabilization of the ligand or TS, they can be of importance for the energetics of the reaction (Figure 9 B) or they can be part of activity regulation, as described below.

Water molecules can participate in chemical reactions as substrates themselves, for example in hydrolytic reactions [217], or they can be metal-ionized nucleophiles, as in the hydration of carbon dioxide performed by carbonic anhydrase [218]. Furthermore, the amphoteric nature of water molecules allows them to be involved in proton transfer reactions [102].

Trapped water molecules in the active site have been shown to serve as extension of the active site residues to help in stabilizing the TS structure [219, 220] or smoothen out the active site wall to ensure a proper fit of the substrate, as demonstrated for a terpene cyclase [203]. Different solvation of the TS and water networks in the active site have been retrospectively described as a source of increased enzymatic activity in different engineering attempts [221, 222]. An intended TS stabilization by internal water molecules has been recently explored as enzyme engineering strategy [82]. The significance of active site solvation and rearrangement of hydration pattern in TS formation has also been acknowledged in a de novo designed kemp eliminase [222]. However, theses active site water molecules need to be controlled precisely to avoid unwanted side reactions. Water displacement to and from the active site at the right moment during a catalytic cycle can be promoted via the formation of water channels like in all trypsin-like proteases [223] or FabI Staphylococcus aureus enoyl-acyl carrier protein reductase [224]. The formation of such water-wire harboring water channels that connect the bulk with the active site [225] has also been observed to allow re- or deprotonation of amino acids, cofactors, substrate or product [102].

A subtler solvent contribution to catalysis has been observed for peptide hydrolysis by a metalloprotease. By following catalysis time-dependently, it has been shown that solvent dynamics are substrate-specific and generate slow protein-water-coupled motions that persisted longer than one catalytic cycle, which influences enzyme reactivity [161]. Such solvent movement in form of relocation of water molecules and water network reorganization, which can occur in tandem with dynamic protein movements, can improve the energetics of the reaction [226]. Water relocation in the active site but also from the active site to the bulk solvent during the transition state can influence the Gibbs free energy of activation by several kcal/mol in both enthalpy and entropy of activation [226]. A lowering of the free energy by water network rearrangements and water movement from energetically unfavored to energetically favored locations in the active site has been studied for protein-ligand binding
(see **Protein-Ligand Binding**) but rarely for the catalytic step (as explored in **Paper V**).

Thermodynamically, all parts of the system, protein, substrate and solvent, need to reach the TS. However, solvent contribution is usually neglected and the influence of solvent dynamics on enzyme activity remains poorly understood. A deeper understanding of solvent distribution and water network rearrangements upon transition state formation (**Paper V**) could help in extending rational enzyme engineering strategies in the future.

**Activity Regulation**

Enzyme activity can be regulated by different strategies, such as by allostery, the long-distance communication between distinct binding sites that can be promoted by water molecules (**Figure 9 C**) [227, 228]. The interconnection of protein and solvent dynamics makes it difficult to untangle both effects experimentally. However, there are examples where allostery has been shown to be transmitted via structural water molecules [227, 228], the rearrangement of water networks or solvent dynamics, including examples where the allosteric regulation is not accompanied by large conformational changes. Examples for allosteric regulation of enzymatic activity based on hydration water, structural water, or water-networks include aurora kinase [229] and GTPase allostery mediated via structural water networks [230, 231]. Allosteric regulation by intertwined solvent and protein dynamics has, for example, been reported for fluoroacetate dehalogenase (FACD) [155].
III. $\alpha/\beta$-Hydrolases and Esterases

Paper I presented in this thesis investigates the promiscuous activity of a non-canonical $\alpha/\beta$-hydrolase from potatoes. The following chapter is therefore introducing $\alpha/\beta$-hydrolases and more specifically the esterase patatin.

Hydrolases are essential enzymes that break down biomacromolecules, using water to cleave chemical bonds. They can be further classified based on the type of bond they can hydrolyze into, for example, amidases or esterases, cleaving amide or ester bonds, respectively. Many hydrolases have been analyzed and engineered for industrial purposes, as for example detergent ingredients or for applications in the pharmaceutical or food industry [23, 232]. Promising engineering examples include the substantial increase of thermostability of Bacillus subtilis lipase LipA [233], the complete inversion of the enantioselectivity of a Bacillus subtilis esterase [234] but also complete transformation of catalytic activity, for example, from esterase to epoxidase hydrolase activity [235]. Different hydrolases cover a large array of bond-cleavage reactions, which are performed by different structures and mechanisms. Based on their catalytic mechanism and fold, they can be clustered into different groups. Large subgroups are the PA (proteases of mixed nucleophile, superfamily A) clan that have a chymotrypsin fold or the $\alpha/\beta$-hydrolase-fold superfamily.

The $\alpha/\beta$-hydrolase superfamily is one of the largest structurally-related group of enzymes [236]. They share a common fold with a central mostly parallel $\beta$-sheet core, interconnected and surrounded by $\alpha$-helices [236, 237] (Figure 10). They have very low sequence similarities but high conservation of catalytic residues. Catalysis by $\alpha/\beta$-hydrolases is most commonly performed by a catalytic triad (see Covalent Catalysis) that is able to degrade a variety of different substrate bonds. In brief, a nucleophilic residue, polarized by the base of the triad, is attacking the electrophilic carbon (carbonyl carbon in amides or esters) of the substrate. This yields the first tetrahedral intermediate with a charge build-up at the carbonyl oxygen that is usually stabilized in an oxyanion hole. The tetrahedral intermediate collapses back to the carbonyl carbon, which yields in a covalently-bound acyl-enzyme intermediate and the first product that leaves the active site. Attack of a water molecule triggers the formation of a second tetrahedral intermediate, which collapses, yielding the free enzyme and second product (Figure 5). Most important conserved residues are the members of the catalytic triad and the residues comprising the oxyanion hole. Recent studied examples belonging to the $\alpha/\beta$-hydrolases and performing such covalent catalysis include the PET degrading hydrolase PETase [97] (Figure 10).
The esterase patatin analyzed in **Paper I** is a non-canonical $\alpha/\beta$-hydrolase. It is a non-specific lipid acyl hydrolase that degrades phospholipids, glycolipids, sulfolipids and mono- and diacylglycerols. Similar to canonical $\alpha/\beta$-hydrolases, it possesses a central $\beta$-sheet core kept in place between a front and back $\alpha$-helical domain [238, 239] (**Figure 11**). However, patatin possesses only six central $\beta$-sheets ($\beta_{1,2,5,6,9,11}$, **Figure 11**) in its core domain, which is topologically different from canonical $\alpha/\beta$-hydrolases.

Additionally, the catalytic mechanism of patatin is distinct from the typical catalytic triad-based mechanism, using a catalytic Ser-Asp dyad instead. Similar to other hydrolytic enzymes performing catalysis with a catalytic dyad [239], the Oγ atom of the catalytic serine is very close to Oδ2 of the aspartate, allowing direct activation by abstraction of the serine proton prior to nucleophilic attack of the substrate carbonyl carbon. The following catalytic cascade is similar to the reaction performed by a catalytic triad (**Figure 5**). An oxyanion hole is formed to stabilize the developing negative charge at the substrate carbonyl oxygen during formation of the tetrahedral...
intermediate. The first product leaves supported by a proton transfer from the aspartate. Hydrolysis and regeneration of the free enzyme occurs through a similar mechanism [239].

The differences in structure and mechanism of patatin compared to other hydrolases and especially esterases enables the formation of a hydrogen bond between the protein backbone and the substrates’ amide group. This hydrogen bond is key to why it shows high promiscuous activity towards amides, a fact described and explored in Paper I.
IV. Terpenes and Terpene Synthases

The majority of the work performed within this thesis focused on analyzing the catalytic mechanism of different terpene cyclases. These enzymes are key to terpene synthesis and are introduced in the following chapter.

Terpenes

Terpenes and terpenoids comprise the largest, most ancient and structurally diverse class of natural products with more than 80,000 registered compounds in the natural product dictionary [240, 241]. They are produced by organisms of all kingdoms of life and new compounds are still discovered regularly [242]. The word terpene originates in the word “turpentine” that comes from the tree species Pistacia terebinthus, the first source for harvesting the terpene-rich turpentine resin. Today however, according to the terpene rule initially formulated by Ruzicka in 1953 [243], all chemical compounds that descend from an isoprene precursor and possess the basic molecular formula (C₅H₈)ₙ belong to the chemical class of terpenes (also called isoprenes). Terpenes, hydrocarbon-based compounds, are often modified or further functionalized, giving rise to a large diversity of natural products with different properties. If they contain additional functional groups, they are classified as terpenoids (for simplification, terpenes and terpenoids will be referred to as terpenes). Their natural role is often unknown, but many terpenes are plant secondary metabolites and can act, for example, as attractants, expellents, anti-feedants, signaling compounds and growth regulators and can have anti-microbial, anti-viral, anti-malaria and anti-cancer properties [244].

All these various natural attributes give rise to the usage of many terpenes in a variety of different industries as drugs, solvents, flavors, fragrances, synthetic intermediates or as polymers in rubber [245, 246]. Traditionally, commercially useful terpenes are harvested from their plant source, associated with high costs and a negative environmental outcome. An increasing number of terpene synthesis routes are recreated chemically but the chemical synthesis is very challenging and often involves many steps and harsh chemicals [247, 248], so that the biotechnological terpene production is gaining more and more attention [24, 46, 249, 250]. Although the basic molecular structure of terpenes is a multiple of isoprene units, it should be noted that not all terpenes are carbon multiples of five: This is because subsequent modifications can involve cleavage or further addition of carbons and can give structures with various carbon numbers. They are, however, divided according to the number of fused five-carbon subunits into hemi (C₅)-, mono (C₁₀)-, sesqui (C₁₅)-, di (C₂₀)-, sester (C₂₅)-, tri (C₃₀)-, sesquar (C₃₅)- and poly (C₅₀)-terpenes (Figure 12).
Their biosynthetic production can be separated into four phases (Figure 12). In the first phase, the five-carbon precursor molecules are generated. In the second phase, the isoprene subunits are fused to produce the linear hydrocarbon isoprene backbone that is then cyclized in the third phase and further functionalized in the last phase by downstream tailoring enzymes such as cytochrome P450 or methyltransferases.

The precursor molecules DMAPP (dimethylallyl-diphosphate) and IPP (isopentenyl-diphosphate, Figure 12, Phase I) are produced in the mevalonate (eukaryotes and archaea) or in the non-mevalonate (MEP pathway in prokaryotes and cyanobacteria) pathway (both pathways exist in plants). Formation of the linear isoprene precursor for subsequent cyclization, the second phase of terpene production, is performed by prenyltransferases (Figure 12, Phase II). They are terpene synthases that couple isoprenyl subunits in a most commonly head-to-tail (C₃-C₄) or rarer tail-to-tail (C₄-C₄, squalene synthase) manner. Irregular isoprenyl coupling in a non-head-to-tail or tail-to-tail manner, especially for the generation of longer isoprenes (C₃₀), leads to an even larger variety of precursor molecules. The generation of longer isoprenes has been reported to be sometimes split into two or even three steps, such as for bacterial squalene synthesis [251]. Otherwise, prenyltransferases are processive enzymes that elongate the chain by subsequently adding IPP units. They possess an α-helical fold with a central active site pocket that determines the length of the product (“molecular ruler”) [240].

Regular prenyltransferases are metal-dependent enzymes (Mg²⁺), where three metal ions are required to provide efficient ionization of the substrate and for carbocation formation that allows product generation. The metal ions are coordinated by two aspartate-rich metal ion binding motifs (DDXXD; number of Xs can vary) and by the remaining diphosphate group of the substrate. A structural rearrangement caps the active site and leads to water exclusion to avoid premature quenching of the generated carbocation. The enormous structural diversity of the terpene class, however, comes from the cyclization (Figure 12, Phase III) of the generated linear isoprene molecules. The cyclization reactions performed by terpene cyclases are outstandingly complex and have been repeatedly described as the most complicated existing enzymatic reactions [240] with more than half of the substrate carbons undergoing bond changes and hybridization during catalysis.
**Phase I** - DMAPP and IPP production

![DMAPP and IPP reaction](image)

**Phase II** - Generation of linear isoprenoid precursors

<table>
<thead>
<tr>
<th>Carbon atoms</th>
<th>Terpene prefix</th>
<th>Linear isoprene precursor</th>
<th>Class</th>
<th>Fold</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>hemi-</td>
<td>OPP OPP OPP OPP</td>
<td>I</td>
<td>αβ</td>
</tr>
<tr>
<td>10</td>
<td>mono-</td>
<td>OPP OPP OPP OPP OPP OPP</td>
<td>I</td>
<td>αβ</td>
</tr>
<tr>
<td>15</td>
<td>sesqui-</td>
<td>OPP OPP OPP OPP OPP OPP</td>
<td>I</td>
<td>α, αα, αβ, αβγ</td>
</tr>
<tr>
<td>20</td>
<td>di-</td>
<td>OPP OPP OPP OPP OPP OPP</td>
<td>I, II</td>
<td>α, αα, αβγ</td>
</tr>
<tr>
<td>25</td>
<td>sester-</td>
<td>OPP OPP OPP OPP OPP OPP</td>
<td>I, II</td>
<td>α, αα</td>
</tr>
<tr>
<td>30</td>
<td>tri-</td>
<td>OPP OPP OPP OPP OPP OPP</td>
<td>II</td>
<td>βγ</td>
</tr>
</tbody>
</table>

**Phase III** - Cyclization

![GGPP cyclization](image)

**Phase IV** - Functionalization

![Dammarenediol-II and Protopanaxadiol](image)

**Figure 12: Overview of terpene biosynthesis.** Phase I: DMAPP and IPP as final products of the mevalonate or MEP pathway. Phase II: Linear isoprene precursor generation by fusion of IPP to DMAPP to generate the C_{10} molecule GPP (geranyl diphosphate), or IPP fusion to GPP to generate the C_{15} molecule FPP (farnesyl diphosphate), or IPP fusion to FPP to generate the C_{20} molecule GGPP (geranyl-geranyl diphosphate), or IPP fusion to GGPP to generate the C_{25} molecule GFPP (geranyl-farnesyl diphosphate). The C_{30} precursor molecule squalene is produced by fusion of two FPP molecules in a tail-to-tail manner. Table layout according to Christiansen, 2017 [240]. Phase III: Linear precursor molecules are cyclized by class I and class II terpene cyclases, here exemplified by GPP cyclization to ent-copalyl diphosphate by PtmT2. Phase IV: Cyclic terpenes can be further modified (magenta) by tailoring enzymes, such as for the production of the anti-cancer compound protopanaxadiol by protopanaxadiol synthase, a cytochrome P450.
**Terpene Cyclases**

Traditionally, all terpene cyclases are categorized into two different classes according to the catalytic strategy they employ. Class I enzymes use activated substrates and initiate the reaction by metal-assisted ionization, whereas class II cyclization is protonation-initiated (Figure 13).

Class I mechanism

![Class I mechanism diagram]

Class II mechanism

![Class II mechanism diagram]

*Figure 13: Initiation of terpene cyclization by class I and class II terpene cyclases [252]. Top: Class I cyclases require a trinuclear Mg\(^{2+}\) cluster to perform ionization-initiated dephosphorylation of an activated substrate for the generation of the initial carbocation that triggers ring closure reactions. Bottom: Class II cyclases protonate the terminal alkene (or epoxide group) of the substrate by a conserved aspartate to generate the first carbocation that triggers ring closure reactions.*

Despite the differences in substrate composition, initiation of the reaction, protein fold and amino acid sequence, there are many noticeable common features across the classes. Both classes follow the same reaction scheme, where the substrate is acquired and prefolded in the active site, followed by initiation of the reaction, carbocation stabilization and final deprotonation.

In both class I and class II enzymes, substrate binding and arrangement into the active prefold with correct stereochemical configuration is crucial for the product outcome. The active site serves as a template that guides the flexible isoprene substrate into a reactive conformation that resembles the product structure. Initiation of the reaction is class-specific but in both classes leads to a carbocation that propagates through the substrate and induces all following ring-closure reactions. The active site has to stabilize the generated high-energy carbocationic intermediates and prevents premature quenching by providing cation-\(\pi\)-interactions with at least two to three aromatic side chains in the hydrophobic active site lining. In fact, it is the preciseness of the prefold and the stabilization of the carbocationic intermediates that determine the fidelity of the cyclase. High-fidelity cyclases, such as human oxidosqualene cyclase, provide a snug fit and allow only a single substrate prefold [84,
On Catalytic Mechanisms for Rational Enzyme Design Strategies

Other cyclases, such as γ-humulene synthase from Abies grandis, produce over 52 different products from the same substrate by being able to accommodate different prefolds [254]. This has been explored as a rational enzyme engineering strategy, where single amino acid substitutions in the active site that remold the active site contour have led to different product outcome [255, 256] without changing the catalytic strategy itself. Terpene cyclization is finalized by deprotonation by a suitable base or water addition using a bulk water molecule [257]. For final deprotonation by water addition, a highly coordinated water molecule has to be present in the active site that needs to be controlled so that it cannot quench previous intermediates. Alternatively, it has been proposed that water accesses the active site by loop movements that allow water to enter the active site only for deprotonation and formation of the alcohol product [257, 258]. Overall, terpene cyclases chaperone the cyclization but interfere little chemically.

The increasing number of available crystal structures in the last 20 years has revealed that terpene cyclases do not only apply similar catalytic strategies but are also structurally related. Terpene synthases (including cyclases of both classes and prenyltransferases) generally are composed of so-called α-, β- and γ-folds that are combined to a number of different tertiary structures [240]. Triterpene cyclases with an α-helical βγ-fold are believed to be the most ancient terpene cyclases. The γ- and the β-fold of the triterpene cyclase squalene hopene cyclase from Alicyclobacillus acidocaldarius share 23% sequence identity and structural homologies. This led to the conclusion that the γ-domain arose from a primordial gene duplication and fusion to the β-unit followed by the development of the catalytic activity at the domain interface. The class I α-fold is evolutionary distinct and arose from a gene duplication event of a four helical bundle metal-binding protein that initially performed solely prenyltransferase activity. Fusion of the βγ- to the α-fold gave αβγ-structures, the typical fold of plant diterpene cyclases with class II catalytic activity and possibly bifunctional activity [259, 260]. Cyclase activity in α developed later, giving the class I cyclase mechanism. A subsequent loss of the γ-fold led to the αβ-structures of today’s class I terpene cyclases. This was probably followed by a loss of the β-unit, giving α-fold cyclases. This evolutionary relationship is apparent in the same folds and chemistries, originating from the common primordial ancestor [240, 261, 262]. It has been shown that domain architecture is crucial for the catalytic outcome of terpene cyclization reactions. For example, swapping of domains can yield novel products and deletion of catalytically inactive domains (such as deletion of the βγ-domain in an αβγ-class I cyclase) has been shown to compromise cyclization fidelity [263].

It should be mentioned that examples of non-canonical terpene cyclization reactions by structurally non-related enzymes with different catalytic mechanisms have been reported recently [264, 265]. However, the focus of this thesis lies on traditional class II terpene cyclases and their employed catalytic strategies.
Class II Cyclases

Class II cyclases protonate the terminal carbon-carbon $\pi$-bond (alkene group) or epoxide group of the isoprene substrates GGPP (diterpene cyclases) or (oxido)squalene (triterpene cyclases) (Figure 13). They employ general acid catalysis using a Bronsted acid, a strictly conserved aspartate located in the sequence motif DXDD (prokaryotes) or XXDCX (eukaryotes). For class II terpene cyclases that protonate a terminal double bond, the proton has been reported to be oriented in an anti-position, making the acid $10^4$-fold more acidic than with a syn-oriented proton [266, 267], which enables the protonation of the alkene bond. This anti-orientation can be stabilized by hydrogen bonds, including incorporation of a water molecule as shown in Arabidopsis thaliana CotB2 [268]. The reprotonation of the catalytic acid has been reported to occur via a water wire connecting the catalytic acid with the bulk that can allow proton transport via a Grotthuss mechanism [269].

Protonation of the alkene or epoxide group yields a tertiary carbocation that then reacts with the closest remaining $\pi$-bond of the substrate to form a 6-membered ring (A-ring, see Figure 14 and Figure 15). The reaction proceeds by iterative carbocation formations and ring closure reactions to form the multicyclic product. This emphasizes the importance of the correct prefold of the flexible linear substrate and the shielding of the generated carbocations from premature quenching. The product diversity comes from subsequent reactions such as 1,2-hydride or 1,2-alkyl shifts and rearrangements. Final deprotonation occurs by a base or the capture of a water molecule to form the alcohol product.

Class II terpene cyclases perform activity in a $\beta\gamma$-fold scaffold with the large hydrophobic active site cavity lying at the interface of the two domains and the catalytic aspartate located in the $\beta$-fold. Many class II cyclases are monotopic membrane proteins with the membrane helix located in the $\gamma$-fold. They can have $\alpha\beta\gamma$-folds with either the $\alpha$-fold inactive or, for bifunctional cyclases such as Abietadiene synthase that performs tandem class I and class II diterpene cyclization, with activity in both active sites [270].

Class II Diterpene Cyclases

Class II diterpene cyclases promote cyclization of the 20-carbon isoprene GGPP to a variety of different products with promising properties [271]. The pyrophosphate moiety of the substrate that is not part of the initiation of the reaction serves as a handle to fix the substrate in the large hydrophobic active site. The active site cavity lies at the interface of the $\beta\gamma$-fold. Substrate stabilization can occur via a Mg$^{2+}$-cluster bound to an EDXXD-motif in the $\gamma$-fold [241]. Protonation of the terminal double bond by an anti-oriented proton of the catalytic aspartate at the opposite end to the diphosphate group yields a carbocation. This is accompanied by a concerted C$_{10}$-C$_{13}$ and C$_6$-C$_{11}$ bond formation (Figure 14 B) to form the bicyclic carbocationic intermediate. Aromatic residues in the active site provide stabilizing $\pi$-electron...
interactions. As previously described, the contour of the active site enforces prefolding of the substrate and thus determines the product outcome. Subsequent modifications such as hydride transfers and methyl shifts occur before termination by proton elimination or solvent addition. The remaining diphosphate stays intact and can be the substrate for further cyclization by a class I cyclase. This “double cyclization” occurs naturally in plant diterpene biosynthesis and together with the modular nature of terpene cyclase folds has been explored for the construction of different synthetic routes to a repertoire of functional diterpenes [272]. Different class II and class I terpene cyclase folds have been combined to enable first a class II cyclization reaction, initiated at the GGPP terminal double bond, followed by a class I cyclization with the remaining diphosphate, which led to the generation of 41 novel diterpene compounds [272].

Figure 14: PtmT2 structure (A, PDB: 5BP8 [273]) and enzymatic reaction mechanism (B). Enzyme colored according to secondary structural elements (helices in blue, sheets in cyan) with the catalytic aspartic acid (Asp313) in magenta in the active site cavity in the βγ-interface.

Aside from the structure of the bifunctional abietadiene synthase that exerts class II diterpene cyclization in the βγ-interface [274], crystal structures of two pure class II terpene cyclases have been solved [266]. Both are ent-copalyl diphosphate synthases (CPS) using the same catalytic strategy. Plant CPS from Arabidopsis thaliana reveals the typical αβγ-fold of plant class II diterpene cyclases with an inactive α-unit [269]. Structural elucidation of PtmT2 [273], a CPS from the soil bacterium Streptomyces
platensis (analyzed in Paper II) shows $\beta\gamma$-architecture (Figure 14). The $\beta\gamma$-fold of PtmT2 has been shown to be more closely related to the A. thaliana CPS $\beta\gamma$-fold than to the $\beta\gamma$-fold of the triterpene cyclase squalene hopene cyclase (SHC) although the sequence is more similar to the SHC sequence (24% sequence identity and 39% sequence similarities to SHC vs 16% sequence identity and 26% sequence similarities to AtCPS). The active site in the interface of the PtmT2 $\beta\gamma$-fold harbors the catalytic DXDD motif. It contains seven aromatic and four aliphatic residues that form the hydrophobic active site walls. Nevertheless, it is filled with water molecules. The negatively charged diphosphate group of the substrate is stabilized by two lysine residues and a $\text{Mg}^{2+}$ ion, coordinated by three aspartate residues [273].

**Class II Triterpene Cyclases**

Triterpene cyclases occur in all kingdoms of life - prokaryotes, eukaryotes and archaea [275]. Ancient linear and cyclic triterpenes, used as biomarkers to analyze sediments and fossils, have been proposed to be part of the evolution of biomembranes as primitive membrane components that clustered and essentially formed the first cell-like structures [276]. Modern triterpene cyclases generate a large diversity of different compounds with diverse function in many organisms, such as providing membrane fluidity and integrity [277]. Additionally, they are precursor molecules for the generation of steroids, other hormones and secondary metabolites [102]. Triterpene cyclases use the class II mechanism (Figure 13), with a reaction sequence very similar to the diterpene cyclase mechanism described above. Instead of the $\text{C}_{20}$ GGPP, they use a $\text{C}_{30}$ linear isoprene substrate that can be a pure hydrocarbon such as squalene or the epoxidized form – 2,3-oxidosqualene. Generally, squalene cyclases can be found in bacteria and oxidosqualene cyclases (OSC) are present in higher organisms. However, OSC with eukaryotic origin can occur in bacteria as well, which supposedly arose from lateral gene transfer [275].

Differences in triterpene cyclase active site geometries give rise to different prefolds and with that to the catalysis of more than 80 different carbon skeletons, divided into several groups such as sterols, hopanes and lupanes [278]. Triterpene cyclases are inherently promiscuous and therefore promising engineering targets for the generation of novel compounds with potent properties [252, 279-281]. Compounds based on natural polycyclic triterpenes are already widely used, or in the drug discovery pipeline. For example, as antibiotics (zeylasterone [282]), as anti-cancer compounds (ursolic acid [283], cucurbitacin [284], botulin [285]), or as insecticides (azadiron [286]). Human triterpene cyclase OSC is also a drug target for hypercholesterolemia, atherosclerosis [287] and tumor growth [288].

Despite differences in molecular structure and prefold of the substrate, the catalytic active motif, and the rearrangements occurring after initiation, there are many similarities in the reaction sequence of squalene and oxidosqualene cyclases that generally follow the above presented terpene cyclization scheme (see **Terpene Cyclases**). A deeper understanding of the structure-function relationship and
details of the catalytic mechanism arose from the two resolved triterpene cyclase crystal structures of SHC from *A. acidocaldarius* (best resolved structure: 1UMP [104]) and human OSC (hOSC, 1W6K [102]) ([**Figure 15**](#)), the latter being the main focus of this thesis.

**Figure 15**: Structure of SHC ([A, PDB 1UMP [289]]) and structure and reaction mechanism of human OSC ([B and C, PDB 1W6K [102]]). Enzymes colored according to secondary structural elements (helices in blue, sheets in cyan) with the catalytic aspartic acid in magenta (Asp376 in SHC and Asp455 in OSC) in the active site cavity lying in the $\beta\gamma$-interface.

Substrate acquisition of the monotopic membrane enzymes usually occurs from the membrane, as both substrates and products are highly hydrophobic. The substrate, squalene for SHC and oxidosqualene for OSC, enters the active site and is forced into its prefold by the shape of the active site. The large central active site cavity of SHC,
with aromatic amino acids in the walls, binds squalene in the required chair-chair-chair conformation [290]. For OSC it has been shown that the substrate acquisition channel is separated from the active site via a constriction site, such that side chains have to move to allow the passage of the substrate. The energetically unfavorable chair-boat-chair conformation of 2,3-oxidosqualene is enforced by Tyr98, located in the upper half of the active site that pushes the methyl group at C10 below the molecular plane (one residue insertion above and one residue deletion below the molecular plane in sequence comparison to SHC) to create the B-ring (Figure 15) [102].

This is followed by the initiation of the reaction by the catalytic active aspartate at the top of the active site, which generates the initial carbocation [291]. The catalytic active DXDD motif of SHC with Asp376 is backed up by His451 as strong electrophile. The anti-position of the transferred proton is ensured by hydrogen bonding to Asp374 and Asp377 that make it a very strong Brønsted acid, able to protonate the terminal carbon-π bond. SHC also accepts the more readily protonated epoxide group of oxidosqualene [292]. The human OSC reaction is initiated by Asp455 that is hydrogen bonded to Cys456 and Cys533, which increases its acidity. However, its syn-oriented proton is less acidic compared to the SHC counterpart (Asp376) and is not able to activate the C=C π-bond of squalene [102].

The initiation leads to carbocation formation, which results in the first ring formation and delocalization of the carbocation that then migrates through the substrate skeleton and initiates all following ring closure reactions (Figure 15). For SHC, the polycyclization occurs in eight substeps: A- and B-ring formation, C-ring closure as a 5-membered ring, ring expansion, formation of a 5-membered D-ring, ring expansion and E-ring closure as 5-membered ring, finished by final deprotonation to either hopene or hopanol, which are functional analogues of steroids in bacterial membranes [293]. A water network polarized by residues G-R-N-N, at the base of the active site, functions as catalytic base for the deprotonation [104].

In human OSC, carbocation formation triggers A- and B-ring closure. This is followed by an anti-Markovnikov 5-membered C-ring formation that is rearranged to a 6-membered ring, followed by ring formation of the 5-membered D-ring to generate the protosterol cation with the carbocation at C20 (Figure 15). Other than in SHC, this carbocation cannot be stabilized as there are no corresponding aromatic amino acids providing negative point charges, so that there is no fifth ring formed in human OSC [102]. Computational QM/MM MD (see Computational Methods for Enzyme Engineering) studies showed that A- to C-ring formation occurs in a concerted asynchronous manner and gives a 6-6-5 intermediate which is followed by ring expansion of the C-ring, concomitant with the 5-membered D-ring formation [294, 295]. Following skeletal rearrangements, such as 1,2-methyl and 1,2-hydride shifts, lead to the carbocation being shifted to C8/C9, where it is eliminated by His232 as catalytic base that is hydrogen bonded to Tyr302 [102].
The heat released during the highly exergonic reaction has been discussed to widen up the substrate channel so that the bulky pentacyclic hopene or hopanol (SHC) or the tetracyclic lanosterol (OSC) can be released back to the membrane [290]. As described for all terpene cyclizations, aromatic amino acids, lining the active site walls, provide stabilizing cation-π interactions and prevent early elimination of the generated carbocation by nucleophilic attacks through water in the active site or amino acid side chains and ensure the full cyclization to the respective product [102, 296]. Reprotonation of the catalytic acid of SHC, Asp376, occurs via Tyr495 and a positioned water molecule [290]. The catalytic acid of hOSC, Asp455, is reprotonated via Glu459 that is connected to the bulk solvent by a chain of water molecules or by the proton from the final deprotonation step that can be transferred back to Asp455 [102].

The great acceleration of the reaction compared to the reaction in solvent supposedly arises from the stabilization of the carbocations in an electron rich environment [292]. This purely enthalpic origin of the reaction has been questioned and the involvement of solvent dynamics has been discussed [71] (additionally explored in Paper III and Paper V).

The described similarities of the catalyzed reactions emerge from a parallel evolution and high structural similarities (Figure 15). Triterpene cyclases are composed of two domains (βγ), giving them a dumbbell-shaped structure, where each domain is comprised of an alpha/alpha barrel fold with a large hydrophobic active site cavity at the interface [102, 290]. Interestingly, this fold is shared by the catalytically unrelated farnesyl transferase and nisin cyclase [296], highlighting the evolvability and promiscuity of the structural backbone. Despite the high structural similarities, sequences of triterpene cyclases are quite diverse (only around 40% sequence identities) and they only share a few common sequence motifs. One sequence motif that only occurs in triterpene cyclases with differing quantity (5-8) is the QW motif that is supposedly stabilizing the structure during the highly exothermic reaction (48 kcal/mol [290]). It probably prevents thermal denaturation of the structure by electrostatic forces [297]. Although it could be assumed that OSC evolved from SHC after the great oxidation event, it becomes more and more evident that they share the same common ancestry and evolved in parallel. According to an evolutionary model first expressed by Ourrisson, extant cyclases evolved from a primitive enzyme through gene duplication and divergence, while maintaining crucial elements in the active site (acid, base). Product specificity arose through mutations that changed the shape of the active site [298]. A classification of modern triterpene cyclases, based on phylogenetic analysis, reveals 20 homologous families arranged into 2 superfamilies (OSC and SHC) with 10 subfamilies in each clade [299].

Within the scope of this thesis, the triterpene cyclases SHC, hOSC and the diterpene cyclase PtmT2 were analyzed in more detail.
V. Present investigation

The work presented in this thesis aims to emphasize the importance of understanding catalytic mechanisms of enzymes for further development of improved, more comprehensive rational engineering approaches. In **Paper I**, we used the detailed mechanistic understanding of a reaction to find a promiscuous scaffold that is able to perform the desired reaction, as a promising starting point for further enzyme engineering. **Paper II to Paper V** evolve around class II terpene cyclases. Details of the employed catalytic mechanism and the previously unacknowledged importance of water movement in the active site for efficient catalysis are discussed. A combination of *in silico* and *in vitro* experiments build the basis for the presented studies.

**Paper I**

**Mechanism-Guided Discovery of an Esterase Scaffold with Promiscuous Amidase Activity**

Catalytic promiscuity is an enzyme’s capability to catalyze not only its main reaction but also various side reactions, usually with lower specificity or turnover. For natural enzyme evolution, promiscuity allows the development of novel catalytic activities that can be beneficial under certain evolutionary pressures. Prominent examples of promiscuous enzyme structures include the terpene cyclase γ-humulene synthase that produces 52 different terpene products from the same substrate [254] and the α/β-hydrolase fold superfamily (see **α/β-Hydrolases and Esterases**), including enzymes with the same fold that catalyze at least 17 different reactions [300]. Promiscuous activities are also promising starting points for both directed evolution and rational engineering strategies [12, 46] that further increase the specificity or efficiency of the catalyst towards the desired reaction. However, a key step in the beginning is the identification of such a promiscuous starting point. Within this study, we used amide bond hydrolysis as a model reaction to present a mechanism-guided identification of an esterase with potentially promiscuous amidase activity. We further expressed and analyzed the discovered enzyme scaffold that indeed displayed high amidase over esterase activity.

Both amidase and esterase activity can be performed by the α/β-hydrolase fold. However, proteases and amidases can usually cleave both amide and ester bonds, whereas esterases, if at all, perform poor amidase activity. This is due to the rate-limiting TS nitrogen inversion required for amidase activity that is stabilized by a hydrogen bond between the NH group of the substrate and either the amidase...
(enzyme-assisted) or the substrate itself (substrate-assisted) [301] (Figure 16 1.). This stabilizing interaction is usually missing in esterases, so that they can only perform poor amidase activity. Previous screening of esterases possessing a canonical α/β-hydrolase fold (see α/β-Hydrolases and Esterases) for a scaffold with a hydrogen bond acceptor for TS stabilization, hinting at a potential promiscuous amidase activity, did not give any results [301]. Thus, within this work, we focused on non-canonical α/β-hydrolase esterases (Figure 16 2.) to find an esterase backbone that allows such a hydrogen bond formation between the substrate and the enzyme.

Figure 16: Mechanism-guided identification of an esterase performing amidase activity. 1. Mechanistic key players of amidase activity were identified (hydrogen bond assisted nitrogen inversion in orange, hydrogen bonds in magenta). 2. In silico screening identified a non-canonical α/β-hydrolase as potential promiscuous scaffold. 3. 90 ns MD simulations showed that patatin provides the necessary backbone hydrogen bond donor and forms a productive TS with an amide substrate (A188, catalytic residues and residues of the oxyanion hole are shown in stick and elemental colors). 4. Expression and analysis of patatin revealed high promiscuous amidase over esterase activity.
To identify non-canonical $\alpha/\beta$-hydrolase esterases with a hydrogen bond acceptor able to stabilize the required nitrogen inversion for amidase activity, a previously calculated TS structure (calculated with quantum mechanics using DFT at the B3LYP/6-31G(d,p) level [82]) was used. The TS structure was superposed (using the catalytic serine and residues forming the oxyanion hole as guiding points) to non-canonical $\alpha/\beta$-hydrolase esterases to screen for enzymes providing potential hydrogen bond acceptors. This docking of the TS led to the identification of human phospholipase A2 and the structurally related acyl hydrolase patatin from potatoes (see $\alpha/\beta$-Hydrolases and Esterases). This analysis revealed the backbones of the amino acids G394 and A188, for human phospholipase A2 and patatin respectively, that could act as hydrogen bond acceptors for stabilization of the TS. To enable more detailed analysis of the potential promiscuity of the soluble esterase patatin, 90 ns MD simulations were performed. They were based on the 2nd tetrahedral intermediate of the reaction (see Covalent Catalysis and $\alpha/\beta$-Hydrolases and Esterases for details of the reaction mechanism) to focus on the rate-limiting nitrogen inversion stabilized by the discussed hydrogen bond. For the generation of this 2nd tetrahedral intermediate, the TS structure was covalently attached to the catalytic nucleophile (S77 in patatin). Additionally, to keep the catalytic acid D215 in a productive conformation, the reacting NH group was protonated. Analysis of the MD trajectory showed that a productive transition state was formed in 2% of all snapshots (Figure 16 3.).

Expression and analysis of the identified patatin scaffold for esterase and promiscuous amidase activity (Figure 16 4.), using p-nitrophenyl butyrate as ester and p-nitrobutyranilide as amide, confirmed high amidase over esterase activity (ratio of $2*10^{-5}$ with $k_{cat}/K_M$ (amide) = 0.45 M$^{-1}$ s$^{-1}$ and (ester) = 20900 M$^{-1}$ s$^{-1}$). In contrast to previously determined high promiscuous amidase activities performed by esterases, for example of Bacillus subtilis esterase (ratio of 1.1 $* 10^{-4}$ (with $k_{cat}/K_M$ (amide)= 0.46 M$^{-1}$ s$^{-1}$ and (ester) = 4200 M$^{-1}$ s$^{-1}$) [302] that was attributed to a water network mediating the interaction between the scissile nitrogen of the substrate and the amino acid E188, patatin enables a direct hydrogen bond between the substrate and the enzyme.

Previous engineering efforts to perform amidase activity in an esterase scaffold, both by directed evolution [303, 304] and rational design [82, 305], gave esterases with improved amidase activity. These studies either focused on introducing an amino acid residue to enable direct hydrogen bonding between the substrate and the protein [82, 305] or focused on engineering a water mediated hydrogen bond bridge between the nitrogen and the protein backbone [82, 305]. Directed evolution approaches led to amino acid substitutions that probably enabled promiscuous activity using other mechanisms: The promiscuous activity of a generated B. subtilis double mutant, for example, affected substrate binding [304]. In contrast, we herein applied the detailed mechanistic understanding of the amidase reaction to find a promiscuous scaffold that already possesses the internal hydrogen bond acceptor which could
be a promising starting point for further enzyme engineering efforts. Other strategies to find promiscuous scaffolds, such as de novo enzyme design or ancestral sequence reconstruction, can also give starting points for further engineering efforts [52, 306]. However, if the key mechanistic activity shifter is known, finding a promiscuous scaffold in already existing enzymes allows bypassing the disadvantages of such bioinformatic methods such as poor sequence quality or quantity or uncertainties in the constructed evolutionary tree, the base of ancestral sequence reconstruction.

**Paper II**

**Protonation-Initiated Cyclization by a Class II Terpene Cyclase Assisted by Tunneling**

A major goal of this thesis was to provide more insight into the terpene cyclase catalytic mechanism employed by class II terpene cyclases (see also Paper III to Paper V). Terpene cyclase catalysis has been described as the most complicated biochemical reaction [240]. A simple linear isoprene substrate is subjected to a number of stereospecific ring closure reactions followed by potential 1,2-methyl and 1,2-hydride shifts, giving rise to a countless number of polycyclic compounds with often potent bioactivities. Therefore, terpene cyclases are promising engineering targets, making a deeper understanding of the underlying mechanism desirable (see Terpenes and Terpene Synthases).

It has been previously suggested that the rate-limiting [71] proton transfer that initiates class II terpene cyclization is supported by proton tunneling [307]. To further study if proton transfer is affecting class II terpene cyclization, we turned our attention to the soluble diterpene cyclase ent-copalyl diphosphate synthase PtmT2 from Streptomyces platensis (Figure 14) [273]. Instead of deionizing the diphosphate moiety of the substrate (which would correspond to a class I mechanism), PtmT2 performs a class II bicyclization, initiated at the terminal double bond of the GGPP substrate.

We performed kinetic studies of GGPP cyclization to the bicyclic ent-copalyl diphosphate by PtmT2 to determine the KIE and to analyze the influence of tunneling on the enzymatic reaction (see Hydrogen Tunneling). To reduce experimental errors, in relation to determination of the actual deuterium concentration in the medium, KIEs were determined in varying H₂O/D₂O buffer concentrations. The ent-copalyl diphosphate product was analyzed using high-performance liquid chromatography (HPLC) coupled with UV detection. The liquid sample is thereby mixed with a carrier phase and then separated chromatographically into its components (mainly the substrate GGPP and the product ent-CPP) depending on their interactions with the stationary phase (the column) followed by UV-detection. Incorporation of deuterium was determined by electrospray ionization of the sample, followed by mass spectrometry, which gave m/z ratios for the different components in the sample. HPLC-MS measurements revealed a peak with an m/z of 273 or 274,
respectively, corresponding to the dephosphorylated product with either hydrogen or deuterium incorporated. Initial measurements gave a high KIE of 7 in 94% D₂O which leveled off to around 4 when using lower H₂O/D₂O ratios. KIEs <7 can be explained by classical mechanics (see Hydrogen Tunneling) and are not necessarily demonstrating the importance of tunneling in a reaction. Thus, we furthermore determined the temperature-dependence of the KIE. This analysis revealed a temperature-independent KIE (Figure 17 A) and Arrhenius factors A_H/A_D of 4 that are characteristic for hydrogen transfer reactions, where tunneling is involved (see Hydrogen Tunneling). A twofold reduction of k_cat in D₂O together with previously determined deuterium isotope effects for the triterpene cyclase SHC from Alicyclobacillus acidocaldarius [71] suggest that the hydrogen transfer is part of the rate-limiting step. Reduced values for k_cat/K_M imply that binding could also be affected by performing the reaction in heavy water. Simple in silico analysis of hydrogen transfer distances support the proposed hypothesis of hydrogen tunneling involvement in PtmT2 catalysis (Figure 17 A).

In summary, we demonstrated the importance of hydrogen tunneling in the GGPP cyclase PtmT2 (Figure 17 C).
Paper III
MD Simulations Reveal Complex Water Paths in Squalene Hopene Cyclase: Tunnel-Obstructing Mutations Increase the Flow of Water in the Active Site

As previously mentioned, one major goal of this thesis was to give more insight into class II terpene cyclization with a special focus on enzyme hydration and the influence of water movement on catalysis. The importance of water movement on triterpene cyclization achieved by Alicyclobacillus acidocaldarius SHC has been previously studied by combined in silico and in vitro analyses [71]. It has been shown that water tunnels in the SHC structure, connecting the large central active site and the surroundings, allow water passage from the active site to the bulk solvent [71]. Furthermore, it has been hypothesized that water expulsion during the substrate prefold and catalytic cycle allows the entropically driven polycyclization reaction of the WT enzyme [71]. To study this hypothesis, enzyme variants with a hindered water flow, achieved by insertion of bulky amino acids into the wall-lining of the water tunnels, were analyzed [71]. Obscured thermodynamic profiles with entropies of activations fully unfavorable for the reaction, as analyzed using Eyring transition state theory, were detected [71] (see Transition State Theory and Thermodynamics of Enzymatic Reactions). All variants with introduced tunnel obscuring mutations showed a decrease in activation entropy concomitant with a lowering in activation enthalpy [71].

The present in silico study was a follow-up study to further analyze these previously designed and thermodynamically analyzed SHC variants [71] concerning the water flow within the active site and between active site and bulk solvent. Therefore, a computational method [80] was redesigned to visualize the most likely pathways of water in SHC and designed variants as streamlines, based on detected water movement in the MD trajectory (see Analysis of Computational Models and Applications for Enzyme Engineering).

For this purpose, 150 ns to 300 ns MD trajectories were computed with AMBER, using the SHC crystal structures 1UMP [289] and 2SQC [104]. Water flow in the empty enzyme, enzyme with substrate located in the entrance channel, substrate squalene in the active site, protonated squalene in the active site, product in the active site, and product in the substrate entrance channel (which supposedly is also the gateway of the product back to the membrane) of the WT structures were compared. Additionally, to analyze the effect of the mutations lying in the previously identified tunnels, mutations (S38W, S168F, S168W, F605W, V440F (control)) were introduced in silico, MD simulations were performed similarly, and water flows with substrate in the active site were identified and compared to WT analyses.

Streamline analysis of all WT trajectories (empty, substrate, protonated substrate, product) showed a continuous flow through certain regions of the structure with most streamlines detected in and around the active site. Water molecules were
moving in and out during all stages of catalysis as exemplified by analyzing the empty, substrate-bound, intermediate-bound and product trajectories. However, streamlines corresponding to water movement were distinct in the different stages. This could be due to steric reasons as the least movement is detected in the trajectory of the structure with the bulky product. However, differences in water movement can also largely affect thermodynamic properties of the system and with that severely affect catalysis (see Water in Enzymatic Reactions and Paper V). Remarkably, differences in water flow of the trajectory with the substrate in the active site and the protonated substrate in the active site (corresponding to the structures before and after the rate limiting proton transfer) were detected, which are most likely not evoked by steric hindrance.

Figure 18: Streamlines in the active site of WT (A and B) and variants and concomitant changes in the streamlines in the active site evoked by introducing bulkier residues in the tunnel linings (C-H). Protein in grey, streamlines in red, substrate in green, catalytic aspartate in cyan and mutated residues in yellow (S38W), orange (S168F), magenta (S168W) or cyan (F605W).

Comparison of the water flow in the WT structure to the previously constructed variants with obscured thermodynamic parameters showed that introducing bulkier side chains forces water to reroute, which concomitantly changes the water movement in the active site (Figure 18). Initially surprising was the detection of an increased water flow and movement in the active site of the designed variants (except for the control) and also in the region between the catalytic active aspartate (Asp367) and the substrate (Figure 18). Additionally, the number of water molecules in a radius of 5 Å around the catalytic aspartate and the protonation distance between catalytic aspartate and substrate seems to be increased by introducing mutations in the wall-linings of the detected tunnels (Table 1). However, this is not valid for all variants and also does not show the same pattern as observed for the in vitro [71] determined entropies of activation (Table 1).
Table 1: Previously determined entropies of activation [71], average number of water molecules in a 5 Å radius (with and without cut-off of the first 50 ns of the trajectory, justified by the observed diverting behavior for some of the trajectories, where equilibrium was reached after 50 ns) and protonation distance between acid and C=C bond in the substrate (with and without cut-off).

<table>
<thead>
<tr>
<th></th>
<th>Entropy of activation based on $k_{cat}/K_M$ (at 328 K) [71] [kcal/mol]</th>
<th>Average number of water molecules in 5 Å radius of Asp367</th>
<th>Average number of water molecules in 5 Å radius of Asp367 (cut-off)</th>
<th>Protonation distance [Å]</th>
<th>Protonation distance (cut-off) [Å]</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>16</td>
<td>2.5</td>
<td>2.9</td>
<td>4.14</td>
<td>4.29</td>
</tr>
<tr>
<td>S38W</td>
<td>-41</td>
<td>3.5</td>
<td>4.0</td>
<td>4.77</td>
<td>4.88</td>
</tr>
<tr>
<td>S168F</td>
<td>-16</td>
<td>3.2</td>
<td>3.2</td>
<td>5.33</td>
<td>5.45</td>
</tr>
<tr>
<td>S168W</td>
<td>-13</td>
<td>2.6</td>
<td>3.1</td>
<td>4.93</td>
<td>5.29</td>
</tr>
<tr>
<td>F605W</td>
<td>-7</td>
<td>3.6</td>
<td>4.0</td>
<td>5.37</td>
<td>5.37</td>
</tr>
<tr>
<td>V440F</td>
<td>n. d.</td>
<td>2.8</td>
<td>2.6</td>
<td>3.88</td>
<td>3.67</td>
</tr>
</tbody>
</table>

Based on these results, it was hypothesized that the increased protonation distance could potentially result from water molecules being fixed in between the aspartate and the substrate which could at least partially (2.3 kcal/mol in entropy for fixing one water molecule to a protein at 328 K, see Protein-Ligand Binding) explain the unfavorable entropy in the variants and the concomitant lowering in activation enthalpy. For further clarification, near attack conformer (NAC) analysis [308] has been performed for WT and variants. A NAC is a reactive substrate conformation [309], energetically between the Michaelis complex and the TS. The NAC concept has been used to explain enzyme catalysis: in an enzyme catalyzed reaction NACs will form more easily compared to the reaction in solution, which will enhance the rate of product formation for the enzyme catalyzed reaction. Within this study, a NAC was defined as a structure, where the protonation distance (catalytic hydrogen of Asp367 and reacting carbon in the substrate) is within the sum of the van der Waals radii of the interacting atoms. The protonation ready conformations (NAC) for the C-H bond in SHC should be close to 3 Å. The fraction of structures reaching NAC and the corresponding activation energies to reach the NAC ($\Delta G_{NAC}$) have been determined. A higher NAC-frequency for WT and control concomitant with a lower $\Delta G_{NAC}$ have been observed. However, the variants barely reach NAC status, according to this analysis. In contrast, in vitro analysis of the designed variants showed that some of the variants have a lower $\Delta G^*$ and should thus have higher frequencies of reaching NAC-status and a lower $\Delta G_{NAC}$.

Overall, it can be concluded that in the triterpene cyclase SHC, water tunnels exist that allow water movement from the active site to the bulk. Obscuring those tunnels by introducing bulkier side chains changes water movement around the mutation but also has severe effects on the water flow in the active site. In the present study, where all variants were designed to close the tunnels (by the bulkier side chains), an
increased number of water molecules and an increased flow of water has been detected in the active site.

Based on the above described similarities in structure and reaction sequence (see **Class II Cyclases**) it was of interest to analyze if other (tri)terpene cyclases also perform an entropy driven catalysis, where water movement is decisive. In **Paper IV** and **Paper V**, we therefore established the expression and investigated the influence of water movement on the catalysis performed by human triterpene cyclase.

**Paper IV**

**Overexpression of Functional Human Oxidosqualene Cyclase in *Escherichia coli***

The importance of water movement for triterpene cyclization in bacteria has been analyzed in **Paper III** on the basis of previously performed thermodynamic studies with *Alicyclobacillus acidocaldarius* SHC [71]. A key objective of this thesis was to additionally analyze and compare water influence on human triterpene cyclization by oxidosqualene cyclase. hOSC produces tetracyclic lanosterol from 2,3-oxidosqualene, which constitutes a key step in sterol biosynthesis (**Figure 15**). This makes it not only an interesting biotechnological engineering target for the generation of valuable polycyclic compounds but additional a promising pharmaceutical drug target for hypercholesterolemia and other cardiovascular diseases [102, 287]. A central pillar for drug design in general is a deep understanding of the reaction mechanism (**Paper II, Paper V**) and an overall well-studied and well-analyzed target. This usually also requires an easy-to-handle, stable and cheap expression and purification system. However, hOSC, as other triterpene cyclases as well, is a membrane-bound, insoluble protein which makes its expression in bacteria more challenging. Therefore, triterpene cyclases from origins other than bacteria are usually expressed in eukaryotic systems such as yeast or insect cells, often in time-consuming procedures leading to low yields of active enzyme [287, 310-313].

Thus, for further analysis, the aim of this study was to set up an easy expression, extraction and purification protocol for hOSC that does not require time-consuming experiments or expensive equipment (**Figure 19**).

We first assessed if post-translational modifications (PTMs), a high GC content or a low codon adaptation index (CAI) of hOSC could be a potential reason for the previously published unsuccessful expression of a homologous eukaryotic lanosterol synthase in *E. coli* (**Figure 19 1.**). Indeed, a high GC content in the N-terminal region and a low CAI (0.65) were found. Both issues could be prevented by using a codon optimized gene for *E. coli*. We additionally compared hOSC to SHC, the latter easily expressible in *E. coli*. Overall, the highly conserved structures, sequence similarities, and predicted common ancestry let us hypothesize that expression of hOSC in *E. coli* should be feasible, using a codon optimized gene (CAI=0.83, GC content adjusted). We chose *E. coli* BL21 (DE3) and BL21 (DE3) Star cells in combination with His-
tagged protein (both C- and N-terminal hexa-histidine tag) as cheap and easy to handle expression systems (Figure 19 2. and 3.) embedded in the in-house (former DNA2.0, now ATUM) plasmid pD861. The expression systems were established by test expressions in both strains using both C- and N-terminally tagged protein to determine the optimal growth conditions, correct inducer concentration (rhamnose) and to optimize the extraction and purification protocol for high hOSC yields (Figure 19 4.).

An optimized extraction procedure followed by metal affinity purification with Ni-NTA agarose beads (Figure 19 5.) led to fairly pure hOSC protein (Figure 20 A) and a yield of 2.9 mg/g bacterial cell pellet after 4 h of expression compared to the previously reported 2.7 mg/g yeast cells after 48 h of expression. Subsequent activity essays and product determination by thin layer chromatography (TLC) and gas
chromatography mass spectrometry (GC-MS, **Figure 19.6.**, **Figure 20 B and C**) confirmed the successful purification of active hOSC from *E. coli*, using this simple expression, extraction and purification protocol.

![Figure 20: Expression, extraction and purification of active hOSC from *E. coli*. A: SDS-Page of different fractions of hOSC purification by Ni-NTA affinity chromatography. L: Ladder with marked signals for 25, 50 and 75 kDa, MF: Membrane fraction, F: Flow through, W1 and W2: Wash fractions and E1 – E4: Elution fractions. B: Thin layer chromatography showing a lanosterol product reference (L), an oxidosqualene substrate reference (O) and the extracted reaction (R). C: GC-MS chromatogram of the derivatized extracted reaction (green) giving a lanosterol product peak (derivatized lanosterol reference in blue, marked by an arrow).](image)

**Paper V**

**Engineering of Water Networks in Class II Terpene Cyclases Underscores the Importance of Amino Acid Hydration and Entropy in Biocatalysis and Enzyme Design**

Using the bacterial expression, extraction and purification protocol established in **Paper IV**, we could further investigate the importance of water placement and repositioning during catalysis of class II terpene cyclases. The previously analyzed SHC [71] (**Paper III**), a thermophilic enzyme, has a high benefit of an entropy driven reaction ($\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$), so that we questioned if the highly favorable entropy is enabling a unique catalytic mechanism or if it is rather a general trait of class II terpene cyclases. We focused our attention on hOSC (see **Class II Triterpene Cyclases**), a crucial enzyme in the steroid biosynthesis and a homologous enzyme to SHC. We reasoned that hOSC could serve as an ideal model system to investigate the importance of entropy and water on biocatalysis. It requires a specific substrate prefold of the linear isoprene substrate, concomitant with a change in entropy. Similar to SHC, it is a membrane protein acting on a hydrophobic membrane-solubilized substrate that is directly acquired from the membrane. This means that no large desolvation effects that usually contribute to the entropy of the system (see **Protein-Ligand Binding**) are expected.

The free energies of water molecules within the enzyme structure depend on the positioning and the local environment of the molecules. Thus, a repositioning of water molecules or water-network-reconfiguration, where water molecules are relocated, are expected to contribute to the thermodynamic state of the system. The
repositioning of water molecules during transition state formation can therefore affect enthalpy and entropy of activation and can contribute favorably to biocatalysis (Figure 9). An enhanced understanding of the contribution of solvent dynamics to the Gibbs free energy of the transition state, which in turn determines the rate of the reaction, would enable the generation of a more fine-tuned rational enzyme engineering approach of terpene cyclases. The aim of this study was to further analyze how water movement on the ps- to ns-timescale contributes to biocatalysis that typically occurs on a μs- to ms-timescale with combined in vitro and in silico studies (Figure 21).

Kinetic and thermodynamic analysis of hOSC WT revealed strong temperature-dependence of $k_{cat}/K_M$ associated with a large favorable entropy of activation ($\Delta S^*$) of 7.6 kcal/mol (at 37 °C) and an unfavorable enthalpy of activation ($\Delta H^*$ = 22.7 kcal/mol). Probably because of the membrane-associated nature of triterpene cyclases, hOSC was not saturable under in vitro conditions. Thus, we additionally determined kinetic and thermodynamic parameters of PtmT2, a soluble, saturable diterpene cyclase, following a class II terpene cyclase mechanism (see also Paper II). PtmT2 analysis could confirm that both rates, $k_{cat}$ (1.4 kcal/mol) and $k_{cat}/K_M$
On Catalytic Mechanisms for Rational Enzyme Design Strategies

(5.5 kcal/mol), are accelerated by a favorable entropic contribution. Interestingly, enthalpies of activation of the thermodynamically analyzed hOSC, PtmT2 (16.5 kcal/mol) and SHC (31 kcal/mol) [71] are much higher than the enthalpies of activation of similar chemical reactions, showing the generally entropy-driven nature of class II terpene cyclases.

To investigate if the favorable entropic contribution is enabled by water movement in specific water tunnels that have been previously identified in the homologous SHC structure, we performed in silico CAVER [72] analysis (see Computational Methods for Enzyme Engineering). A tunnel network consisting of 15 tunnels connecting the active site with the bulk solvent was detected. The ten highest ranked tunnels (according to tunnel parameters such as the bottleneck radii of the tunnels) were selected for further inspection, whereof six tunnels were targeted for obstruction (Figure 22 A, showing tunnels T1 to T3). Tunnel obstructing mutations were introduced both in vitro by site-directed mutagenesis and in silico (Figure 22 B) by simple amino acid swaps. In vitro activity screening of all six variants showed dramatically perturbed thermodynamic activation parameters with some variants showing completely inversed temperature profiles (variants S580W and S443T, respectively, blocked tunnels referred to as T2 and T3). More detailed kinetic and thermodynamic analyses of those variants gave completely obscured activation parameters with entropies of activation of -13.1 kcal/mol (S580W) and -11.1 kcal/mol (S443T) and compensatory changes in enthalpies of activation (2.7 and -11.1 kcal/mol, respectively, Figure 22 C). Such dramatic changes in the entropy of a system (changing from very favorable to unfavorable) could theoretically be associated with all parts of the system (enzyme, substrate, solvent). The same substrate was used in all measurements, so that we concluded that either large structural rearrangements, a change in protein dynamics (configurational dynamics) or a change in solvent rearrangement caused the observed differences in entropy of activation. To exclude differences in entropy due to contributions from altered protein dynamics in the variants compared to the WT, we performed a limited proteolysis – mass spectrometry (LiP-MS) experiment [160, 314]. LiP-MS is based on the assumption that protease accessibility (under a limited amount of time) is linked to protein dynamics. With this relatively simple experimental approach, we could determine overall similar structural and dynamical landscapes.
Figure 22: Overview of the results gained from the combined in vitro and in silico approach presented in Figure 21. A: The three highest ranked tunnels (blue) detected by CAVER analysis based on hOSC structure (grey). Introduced mutations are marked in red. B: Illustration of a tunnel (T2, outer lines in black) that has been blocked by S580W (protein backbone in grey, serine in elemental colors and enlarged sticks, tryptophan in yellow and balls, water molecules in stick, substrate in purple). C: Gibbs free energies of activation for WT, S580W, and S443T show the drastically altered thermodynamic parameters. D: Reconfigured water networks in the S580W variant. The WT structure, residues and water molecules (1,2) are illustrated in purple. Introducing the S580W mutant (backbone in grey, side chains in cyan) reconfigured the water network to the water networks marked in cyan labelled A-E.

In silico comparison (RMSD) of substrate-analogue (PDB: 1W6J [102]) and product-bound hOSC structures (PDB: 1W6K [102]) allowed the conclusion that no large structural rearrangements occur during catalysis that could explain the large favorable entropic contributions to WT catalysis. Additionally, the product-bound structure and the corresponding structure with an empty active site of WT and in silico generated variants built the basis of 200 ns MD trajectories generated in YASARA (see Computational Methods for Enzyme Engineering). Average RMSF values determined over the whole 200 ns trajectories for WT and variants were very similar. Thereby, a contribution of protein dynamics via changes in configurational entropy to the dramatic in vitro observed changes in activation entropy was concluded to be minimal.

We therefore focused on the in silico analysis of differences in water tunnels and water networks between WT and variants and additionally compared the empty and product-bound (considered TS-like) trajectories to determine water network
rearrangements in reaching the TS. Time-dependent CAVER analysis of WT and variants showed that the tunnel network is very dynamic with tunnels opening and closing during the simulation. Introducing mutations in the tunnel-lining residues led to changed tunnel parameters for T2 (S580W, such as the radius) and to a roughly ten-times reduced speed of water movement within the tunnel and furthermore a completely blocked tunnel T3 (S443T). We also compared specific water network rearrangements that occurred in the detected tunnels by changing tunnel wall lining residues using a previously written algorithm [82]. Tunnel hydration was completely altered by introducing single amino acid changes. Based on the conclusions from Paper III, we hypothesized that changes in the water movement and networks in the tunnels should be accompanied by changes in solvation and water networks in the active site. Adjusting the algorithm for water network analysis of hOSC [82], we were able to detect changes in the water network configuration between the empty and the product-bound structures and between WT and variants.

In summary, we found substantial differences in the water-network-configurations in the ground and transition states in hOSC WT and variants. Water relocation in reaching the TS in variants with obscured water tunnels that allow water movement from the bulk to the active site and vice versa is very different to that of WT and could explain the previously determined in vitro differences in activation entropy. This let us conclude that the favorable entropic contribution to hOSC and potentially in general in class II terpene cyclase catalysis could be driven by water network rearrangements.
Conclusions and Future Perspectives

Enzymes, as biocatalysts, promote chemical reactions that govern the metabolism of all living organisms. Their ability to perform incredible chemistry at modest conditions has been exploited in industry and has contributed to the generation of a more sustainable society. However, not all natural enzymes can stand industrial conditions or perform the reactions needed, so that many different enzyme engineering techniques have been developed to custom-tailor the biocatalysts to our needs. A detailed understanding of the reaction mechanism can build a solid foundation for rational, semi-rational and de novo engineering strategies. Also directed evolutionary approaches rely on the selection of an appropriate starting point, which is facilitated by a certain understanding as well.

The overall aim of this thesis was to highlight the potential of a deep mechanistic understanding of enzyme catalysis for enzyme engineering, of both the chemistry performed in the catalytic center (Paper I and II) but also synergistic solvent dynamic effects that occur in the active site but also in the protein scaffold (Paper III to V).

The atomistic understanding of key differences of the performed hydrolytic chemistry of amidases and esterases, the rate-limiting nitrogen inversion that needs stabilization by a hydrogen bond acceptor, paved the way for finding an esterase scaffold able to perform promiscuous amide hydrolysis. Such promiscuous activities are prominent starting points for random directed evolutionary engineering approaches [12, 46]. The combination of using a deep understanding, translating it into the essential structural requirements with in silico screening for a scaffold and simple in vitro analysis has therefore proven as a successful strategy to predict promising targets for further engineering. More recently, instead of finding natural promiscuous scaffolds, de novo design strategies enabled the constructions of scaffolds new to nature that have been explored as starting points for directed evolutionary approaches as well [52]. Until now, de novo designed enzymes do not reach turnover numbers of nature’s own catalysts. Random engineering approaches or such combinatory design strategies (de novo + directed evolution or semi-rational) can lead to mutations scattered over the whole protein structure, not only directly in the active site. This highlights the importance of the protein scaffold. Outer regions can influence the activity on several levels including structural and solvent dynamics, which has been partly explored in Paper II to V.

Paper II to V evolve around terpene cyclases. They catalyze remarkable reactions that produce a countless number of bioactive polycyclic compounds, terpenes, whose biocatalytic production is of major interest for different industry sectors. We were interested in gaining fundamental insights into terpene cyclase catalysis that could build the basis for future enzyme engineering approaches.
To increase our understanding of terpene cyclase catalysis, we performed kinetic analysis of a soluble diterpene cyclase using D$_2$O-enriched buffers. This resulted in temperature-independent kinetic isotope effects, leading to the conclusion that quantum tunneling is important for the rate limiting initial protonation step. It has been argued that enzymes that promote quantum tunneling can influence the process by enabling a short donor-acceptor distance via dynamic movements of the participating atoms [106]. This in turn also suggests that engineering attempts affecting the protein dynamics can influence the tunneling process. In case of the analyzed terpene cyclase, such improved dynamics would influence the rate limiting protonation, which could increase product formation and thus represents an engineering target.

Not only protein dynamics can have large impacts on catalysis [152] but also solvent movement can play an important role [71]. It has been previously shown that the unusually large favorable entropy of activation of the bacterial triterpene cyclase squalene hopene cyclase is probably enabled by water movement during catalysis. Based on this hypothesis, in Paper III, we performed extensive in silico analysis to show that water in the squalene hopene cyclase structure flows in distinct streamlines. Such water movement occurs both in the active site and also in specific water channels in the structure that allow movement between the active site and the bulk. The water flow, which can have large influence on the energetics and thermodynamics of the reaction, was different for the substrate-bound and transition state structure.

To elucidate if the influence of water movement on the thermodynamics of catalysis is limited to the bacterial squalene hopene cyclase or rather a general concept, we included the structural homologue, human triterpene cyclase oxidosqualene cyclase into our analysis (Paper IV-V). With the in Paper IV established expression and purification protocol, we were able to produce large amounts of pure and active enzyme. This allowed a thorough in vitro thermodynamic analysis of human oxidosqualene cyclase presented in Paper V. This analysis showed a strong temperature-dependence and entropy driven reaction. In silico evaluation of MD simulations revealed a tunnel network in the enzyme structure that allows for water movement. By introducing single point mutations in the tunnel lining residues to disturb the water movement in the tunnels, we could additionally show that these changes in the tunnels lead to water network rearrangements in the active site, concomitantly changing the thermodynamics of the reaction.

As enthalpy and entropy of activation comprise the energy required to reach the transition state, a lowering of the TS barrier and with that increase of activity and product formation, could potentially be achieved by adjusting solvent dynamics in the tunnels of the structure. Modulating enzyme activity by altering solvent dynamics is not straight forward and has therefore not been extensively explored yet. However, redesigning water networks in the enzyme structure for improved biocatalysis has been highlighted recently [226]. Many bioactive polycyclic triterpene natural products are
generated by enzymes of the plant kingdom. Difficulties in the biotechnological production and analysis, however, lead to less identified and analyzed plant triterpene cyclases and no solved structures that would enable computational analysis to the same extent. A future aim would be to identify similarities in structure, catalysis and solvent dynamics to cover cyclases from the different kingdoms of life and further, extend the analysis to other enzyme classes as well. The analyses performed in Paper III and V highlight the often neglected importance of water networks and water rearrangements for catalysis, which can be explored as future enzyme engineering strategy.

Additionally, the human enzyme is a promising drug target for cholesterol lowering pharmaceuticals that can also have anti-cancer and anti-microbial properties [288, 315]. Obstruction of some water channels, by introducing mutations in the tunnel lining residues, led to loss of activity potentially by blocking water movement between bulk solvent and active site. This could be exploited further for drug design strategies that could focus on blocking the channels instead of purely focusing on the active site region.

In conclusion, this thesis emphasizes the importance of understanding reaction mechanisms on an atomistic level for all engineering strategies. We furthermore provide evidence that highlight the influence of water network rearrangements and dynamics on catalysis, which can build the basis for new enzyme design strategies.
Acknowledgements

I would like to thank all co-authors and colleagues who have helped me along the road to accomplish the work presented in this thesis.

Per-Olof, thank you for giving me the chance to work on such an interesting but also challenging project and for the opportunity to move to Stockholm, a city that I love and that will always be my second home.

Lucie, my PhD and life wouldn´t have been the same without you. I am so grateful that you became my second supervisor and gave me constant support during all PhD-life situations. I learnt so much from you, both scientifically but also some meaningful life lessons. Thank you for your open door, constant encouragement and guidance. You are a role-model to me and I think everyone should have a Lucie in their PhD-Life.

Peter, I am beyond grateful that you’ve always had an open door for me. Without your positivity, calmness and support this thesis probably wouldn’t have been written after all. Even though you’re so busy you always took the time to listen to me and guide me through some tough times. In my opinion not only a Lucie but also a Peter should be available for every PhD student!

Natalie and Björn, I am the luckiest PhD student to have had you as my group mates. To me it feels like we became family! I learnt so much from both of you and I am truly grateful to have you in my life. Natalie, from the moment we met, you were a great friend and later colleague. You are an inspiration to me and I think, no matter where we end up, we will be close. Björn, you were the best student (together with Adam ☺) and later colleague someone could wish for. I admire your strength, reliability and kindness. Thank you (also for helping me with the Sammanfattning)!

Arne, even though you weren’t in Scilife all the time, you integrated within seconds. I wish I’d be as relaxed as you are!

Camilla, we had a lot of fun working together, even though it wasn’t always easy. Thanks for trying to teach me how to be a better modeler – even though my membrane kept crashing in ns, I learned a lot.

Paul and the Cyano group, thank you for adopting me, especially in the beginning of my PhD. It was so much fun sharing lab and office space with all of you. Ivana, Josefine and Danuta, thank you for freeing up some precious fridge and freezer space for me, showing me around everywhere in Scilife and giving me the best start to my PhD. I am grateful for all your support throughout my PhD, not only scientifically! We should all reunite and have some serious discussions in the GC
room! **Kiyano**, we had the most fun teaching the gene technology course together, what a great way to become friends. I love hanging out with you in what I would consider to be MY hood.

**To the Caulobacter group**, thank you all for mixing up the lab a bit and bringing so much new knowledge and energy to Gamma 5. **Kristina H.**, I am constantly amazed by your wisdom. I was so happy that we were able to write our theses together, at least partially, which kept me sane and made me realize that probably everyone goes through the same stages during the writing of a PhD thesis. On a side note: I am still happy to have found someone that is on the same page when it comes to cleanliness and organization ☺️. **Deike**, thanks for always having an open ear and for taking 3000 pictures of me to get one good shot for my applications!

**Anders, Yasmin, Leif** and the rest of the drug discovery and development platform on Gamma 5, thank you for all your help and support for the protein expression and purification part of my thesis. I know that I was pretty lucky to be on the same floor with you, your knowledge and equipment ☺️.

**Everyone on Gamma 5**, thank you for a great time and so many funny lunch breaks!

**Laura** and **Christian, Lucia, Maria M.** and **Maria J.** on Alpha 2, thanks for being so supportive. **Sofia**, thanks for introducing me to the world of Kickboxing – sorry, for all the uncontrolled punches and kicks that you had to take. I am still learning!

**Lisa M.**, it was so much fun organizing the GRS with you. The whole planning and organization brought so much positivity into my life!

To all the students, PhD students, post docs and researchers I had the privilege to meet during my time as PhD student, I had so much fun getting to know you and your science. That was the best part of my PhD.

To all my friends and family: thank you for your support! I especially want to thank my sister **Phi**, my cousin **Lisa** and **Anna**: everything that I achieve in life is connected to your support and positive energy. Thank you for always being there and for giving me love, laughter and joy (and Helene). **Hannah**, thank you for sharing all the hipster interests with me ☺️.

To my parents, **Mama, Papa** and **Claudi**, thank you for your support. I wouldn´t be where I am without you! **Susanne and Conny**, thank you for being my additional parents!

**David**, I don´t know what to say – thank you for everything! You are the greatest and you make me who I am.
References


On Catalytic Mechanisms for Rational Enzyme Design Strategies


On Catalytic Mechanisms for Rational Enzyme Design Strategies


Publications and Manuscripts