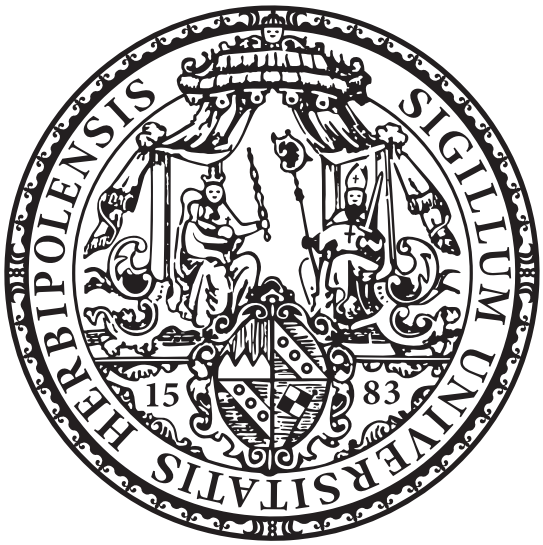
Theoretical investigations of proton transfer and interactions or reactions of covalent and non-covalent inhibitors in different proteins



Dissertation zur Erlangung

des naturwissenschaftlichen Doktorgrades

der Julius-Maximilians-Universität Würzburg

vorgelegt von

Thien Anh Le

aus Arolsen, jetzt Bad Arolsen

Würzburg 2018

Eingereicht bei der Fakultät für Chemie und Pharmazie am

_____

Gutachter der schriftlichen Arbeit

1. Gutachter: _____
2. Gutachter: _____

Prüfer des öffentlichen Promotionskolloquiums

1. Prüfer:_____
2. Prüfer:_____
3. Prüfer:_____

Datum des öffentlichen Promotionskolloquiums

_____

Doktorurkunde ausgehändigt am

_____

Contents

Chapter 1

Summary

Nowadays, computational-aided investigations become an essential part in the chemical, biochemical or pharmaceutical research. With increasing computing power, the calculation of larger biological systems becomess feasible. In this work molecular mechanical (MM) and quantum mechanical approaches (QM) and the combination of both (QM/MM) have been applied to study several questions which arose from different working groups. Thus, this work comprises eight different subjects which deals with chemical reactions or proton transfer in enzymes, conformational changes of ligands or proteins and verification of experimental data. In chapter 3 an introduction to enzymes and its classification is given to the reader. In this connection only enzymes are described which are part of this work. Additionally, the kinetic of enzymes as well as their inhibition are considered.

The theoretical background of used MM or QM approaches are described in chapter 4. Computational details about the simulations or calculations can be found in chapter 5. The following chapters refer to the different investigated subjects. If the reader wants to obtain a general overview to a specific topic, he (or she) may be advised that each chapter ends with a short summary which comprises the aim of the investigation and results of the respective study.

Chapter 6 deals with reaction mechanisms of aromatic inhibitors of cysteine proteases which can be found in many organisms. These enzymes are responsible for various cancer or diseases as for example Human African *Trypanosomiasis* (HAT) or the Chagas disease. Aromatic S_NAr-type electrophiles might offer a new possibility to covalently modify these proteases. Quantum mechanical calculations on Hartree Fock (HF), post-HF or DFT (density functional theory) level have been performed to gain insights into the energetics and possible mechanisms. Results show that different possible reaction products do not play an important role

and that DFT approaches are insufficient to describe the reaction energy of these inhibitors. The next chapter 7 also deals with *Trypanosomiasis* but the focus was set on a different enzyme. The particularity of *Trypanosomiasis* is the thiol metabolism which can also be modified by covalent inhibitors. In this context, the wild type and point mutations of the enzyme tryparedoxin have been investigated *via* molecular dynamic (MD) simulations to examine the influence of specific amino acids in regard to the inhibitor. The focus was also set to different protonation states of relevant cysteine residues in the active site. Experimental data showed that a dimerization of the enzyme occurs if the inhibitor is present. Simulations revealed that the stability of the dimer decreases in absence of the inhibitor and thus confirms these experiments.

Further investigations concerning cysteine proteases such as cruzain and rhodesain have been conducted in chapter 8 with respect to experimental kinetic data of covalent vinylsulfone inhibitors. Several approaches such as QM or QM/MM calculations and docking, MD or MMPBSA/MMGBSA simulations have been applied to reproduce these data. The utilization of force field approaches resulted in a qualitatively accurate prediction.

The kinase AKT (also known as protein kinase B) is involved in a range of intrinsic or extrinsic cell survival signals. A dysregulation of AKT leads to a series of diseases and plays an important role in the formation of cancer. Novel covalent-allosteric inhibitors have been developed and crystallized in complex with AKT. It was shown that depending on the inhibitor a different cysteine residue is modified. To investigate these differences in covalent modification computational simulations have been applied in chapter 9. As only negatively charged cysteine residues can react covalently, amino acids have been investigated which can stabilize the thiolate of cysteines or which might be responsible for a proton transfer.

Enoyl-(acyl carrier) (ENR) proteins are essential in the last step of the fatty acid biosynthesis II (FAS) and represent a good target for inhibition. The diphenylether inhibitor SKTS1 which was originally designed to target the ENR's of *Staphylococcus aureus* was also crystallized in InhA, the ENR of *Mycobacterium tuberculosis* (TB). Crystal structures indicate a change of the inhibitor's tautomeric form. This subject was investigated in chapter 10 *via* MD simulations. Results of these simulations confirmed the tautomerization of the inhibitor. Furthermore, environmental effects on the tautomeric form were investigated by conducting QM calculations in vacuum or solution. Also, another possible inhibitor was probed and its stability in complex with InhA was analyzed.

Chapter 11 deals with the development of a covalent inhibitor originating from a non-covalent ligand. The target FadA5 is an essential enzyme for the degradation of steroids in TB and is responsible for chronic tuberculosis. This enzyme was crystallized in complex with a

non-covalent ligand which served as starting point for this study. Computations on QM or QM/MM level and docking and MD simulations have been applied to evaluate potential candidates. Although no "hit" was obtained, insights into possible warheads, reactivity or docking procedures were gained.

Chapter 12 focuses on the modification of the product spectrum of *Bacillus megaterium levansucrase*, a polymerase which catalyzes the biosynthesis of fructans. The covalent modification of the wild type or mutants of the enzyme lead to an accumulation of oligosaccharides but also to polymers with higher polymerization degree. To understand these changes in product spectra MD simulations have been performed. They revealed on the one hand interactions between the modifier and important amino acids in the active site which are responsible for the enzyme's activity, on the other hand offered a possible explanation for the elongation process of the polymers.

Finally, the proton transfer in catalytic cysteine histidine dyads was investigated in chapter 13. The focus was set on the influence of the relaxation of the protein environment to the reaction. Calculations of the enzymes FadA5 and rhodesain revealed that the preferred protonation state of the dyade depends on the protein environment and has an impact on the reaction barrier. Furthermore, the adaptation of the environment to a fixed protonation state was analyzed *via* MD simulations. Dynamics reveal that at least one nanosecond is required until the relaxation of the surroundings to the respective protonation state occurs.

Chapter 2

Zusammenfassung

Heutzutage sind computergestützte Untersuchungen ein essentieller Teil in der chemischen, biochemischen oder pharmazeutischen Forschung. Durch die in den Jahren gestiegene Rechenleistung ist die Berechnung biologischer Systeme möglich. Im Rahmen dieser Arbeit wurden molekularmechanische (MM) und quantenmechanische (QM) Methoden sowie die Kombination beider (QM/MM) für verschiedene Studien eingesetzt, die teilweise aus Fragestellungen verschiedener Arbeitsgruppen hervorgegangen sind. Dadurch umfasst diese Arbeit acht verschiedene Themenkomplexe, bei denen chemische Reaktionen, aber auch der Protonentransfer in Enzymen, Konformationsänderungen von Liganden oder Proteinen und die Verifizierung experimenteller Daten im Fokus standen.
In Kapitel 3 soll dem Leser eine Einführung in Enzyme und deren Klassifizierung gegeben werden. Hierbei werden nur Enzyme genauer beleuchtet, welche Bestandteil dieser Arbeit sind. Ebenfalls wird auf die Kinetik von Enzymen und auch auf deren Inhibition eingegangen.
Die theoretischen Grundlagen der hier verwendeten MM oder QM Methoden werden in Kapitel 4 erläutert. Generelle Informationen zu den Simulationen oder Berechnungen können in Kapitel 5 gefunden werden. Die darauffolgenden Kapitel beziehen sich auf die verschiedenen Themenkomplexe, welche untersucht wurden. Sofern der Leser lediglich einen groben Überblick zu einer bestimmten Thematik erhalten will, wird darauf hingewiesen, dass das Ende jedes Kapitels eine kurze Zusammenfassung beinhaltet, welche Fragestellung und Ergebnisse umfasst.
Kapitel 6 befasst sich mit Reaktionsmechansimen aromatischer Inhibitoren für Cysteinproteasen, Enzyme, welche in vielen Organismen enthalten sind. Diese Enzyme sind für verschiedene Karzinome oder Krankheiten wie der Afrikanischen *Trypanosomiasis* oder der

Chagas-Krankheit verantwortlich. Aromatische S_NAr-Elektrophile bieten hierbei eine neue Möglichkeit der kovalenten Modifikation dieser Proteasen. Quantenmechanische Berechnungen auf Hartree Fock (HF), post-HF oder DFT (Dichtefunktionaltheorie) Niveau wurden durchgeführt, um Einblicke in die Energetik und mögliche Mechanismen zu erhalten. Die Ergebnisse zeigen, dass hierbei verschiedene mögliche Produkte ausgeschlossen werden können. Ebenfalls kristallisierte sich heraus, dass oft verwendete DFT Ansätze die Reaktionsenergie solcher Inhibitoren unzureichend beschreiben.
Das darauffolgende Kapitel 7 befasst sich ebenfalls mit *Trypanosomiasis*, setzt aber den Fokus auf ein anderes Enzym. Die Besonderheit von *Trypanosomiasis* ist der Thiol Metabolismus, welcher durch kovalente Inhibitoren modifiziert werden kann. In diesem Kontext wurden der Wildtyp und Punktmutationen des Enzyms Tryparedoxin mittels Molekulardynamik Simulationen untersucht, um Interaktionen einzelner Aminosäuren mit dem kovalenten Inhibitor C1 zu evaluieren. Auch die verschiedenen Protonierungszustände relevanter Cysteine im aktiven Zentrum standen im Fokus. Experimentelle Daten zeigten, dass eine Dimerisierung des Enzyms in Anwesenheit des Inhibitors stattfindet. Durch MD-Simulationen konnte gezeigt werden, dass die Stabilität des Dimers in Abwesenheit des Inhibitors sinkt, wodurch experimentellen Daten bestätigt wurden.
Weitere Untersuchungen zu Cysteinproteasen wie Cruzain und Rhodeasin wurden in Kapitel 8 durchgeführt, um experimentelle kinetische Daten von kovalenten Vinylsulfon Inhibitoren zu reproduzieren. Hierbei wurden Methoden wie QM oder QM/MM Rechnungen aber auch Docking, MD und MMPBSA/MMGBSA Simulationen angewandt, um diese Daten zu reproduzieren. In den Untersuchungen zeigte sich, dass die Verwendung der Kraftfeld-basierten Methoden zu qualitativ richtigen Vorhersagen führte.
Die Kinase AKT (auch bekannt als Protein Kinase B) ist in einer Reihe von Krankheiten involviert und spielt eine wichtige Rolle bei der Entstehung von Krebs. Neue kovalent-allosterische Inhibitoren wurden entwickelt und im kovalenten Komplex mit AKT kristallisiert. Die Kristallstrukturen zeigten, dass je nach Inhibitor ein anderes Cystein adressiert wurde. Um diese Unterschiede zu untersuchen, wurden in Kapitel 9 computergestützte Simulationen verwendet. Da erst ein negativ geladenes Cystein kovalent reagieren kann, wurden ebenfalls Aminosäuren untersucht, welche entweder stabilisierend auf die Thiolate der Cysteine wirken können oder für die Deprotonierung zuständig sein können.
Enoyl-(acyl carrier) (ENR) Proteine sind essentiell für den letzten Schritt in der Fettsäurebiosynthese II (FAS) und bilden ein gutes Target zur Inhibition. Der Diphenylether Inhibitor SKTS1, welchen man ursprünglich als Target für den ENR von *Staphylococcus aureus* entwarf, wurde auch in InhA, dem ENR von *Mycobacterium Tuberculosis* (TB), kristallisiert.

Die Kristallstrukturen weisen je nach Protein auf einen Wechsel der tautomeren Form des Inhibitors hin. In Kapitel 10 wurde dieser Sachverhalt mittels MD Simulationen untersucht. Hierbei zeigten die Ergebnisse eine Übereinstimmung mit den experimentellen Daten. Ebenfalls wurden Umgebungseffekte, welche Einfluss auf die tautomere Form haben können, mittels QM Rechnungen im Vakuum oder im Solvens untersucht. Zusätzlich wurde ein weiterer möglicher Inhibitor getestet und seine Stabilität im Komplex mit InhA analysiert.
Kapitel 11 befasst sich mit der Entwicklung eines kovalenten Inhibitors ausgehend von einem nicht-kovalenten Liganden. Das *Target* FadA5 ist ein integrales Enzym zur Degradation von Steroiden in TB und ist für die chronische Tuberkulose verantwortlich. Dieses Enzym wurde im Komplex mit einem nicht-kovalenten Liganden kristallisiert, welches als Startpunkt dieser Untersuchungen diente. QM, QM/MM, Docking und MD Simulationen wurden hierbei verwandt, um potentielle Kandidaten zu evaluieren. Obgleich kein „Hit“ erzielt wurde, konnten wichtige Erkenntnisse hinsichtlich möglicher Warheads, Reaktivität oder Docking Prozeduren gewonnen werden.
Das Kapitel 12 befasst sich mit der Modifikation des Produktspektrums von *Bacillus megaterium Levansucrase*, eine Polymerase, welche die Biosynthese von Fruktanen katalysiert. Durch kovalente Modifikatoren im Wildtyp oder bei Mutanten des Enzyms konnte sowohl eine Anreicherung von Oligosacchariden, aber auch von Polymeren mit höherem Polymerisationsgrad erzielt werden. Um diese Änderungen im Produktspektrum zu verstehen, wurden MD Simulationen durchgeführt. Diese zeigten einerseits Wechselwirkungen zwischen den Modifikatoren und wichtigen Aminosäuren in der aktiven Tasche, welche für die Enzymaktivität verantwortlich sind, andererseits bieten sie eine mögliche Erklärung des Elongationsprozesses der Polymere.
Schließlich wird in Kapitel 13 die Untersuchung des Protonentransfers in katalytischen Cystein Histidin Dyaden beschrieben. Hierbei stand der Einfluss der Relaxation der Proteinumgebung auf diese Reaktion im Fokus. Berechnungen in den Enzymen FadA5 und Rhodesain zeigten, dass der präferierte Protonierungszustand der Diade von der Proteinumgebung abhängt und einen großen Einfluss auf die Reaktionsbarriere hat. Um dynamische Effekte einzubeziehen, wurde die Adaption der Umgebung auf einen fixierten Protonierungszustand mittels MD Simulationen analysiert. Die Simulationen zeigten, dass mindestens eine Nanosekunde notwendig ist, bis die Relaxation der Umgebung hinsichtlich des Protonierungszustandes stattfindet.

Chapter 3

Introduction

3.1 The history of enzymes – a brief overview

At the end of the 17th century it was already known that the transformation of starch to sugar is due to plant extracts or that the digestion of meat is caused by intestinal secretions, still the mechanism responsible for these reactions was unidentified [1]. Louis Pasteur conducted investigations in the 19th century in regard to the fermentation of sugar to alcohol in starch. At that time these unknown substances which promote this reaction were called "ferments" and it was assumed that they are solely active in living organisms [1].
In the year 1877 the term "enzyme" was used by Wilhelm Kühne to describe the fermentation process for the first time. The word "enzyme" is derived from the Greek language (en = in and zyme = yeast) and was used to describe the reaction of ethyl alcohol to carbon dioxide, which is catalyzed by the enzyme zymase [1]. 20 years later Eduard Buchner was able to show, that this yeast extract can also ferment sugar outside of the cells and thus he disproved the assumption that ferments are only able to show activity in living organisms [2]. He named ferments or more precisely this enzyme zymase. 1907, 10 years after he conducted these experiments for the cell-free fermentation he was awarded with the Nobel Prize in chemistry. In these days, the enzymatic activity was related to proteins, however, most of the people presumed that proteins solely serve as carriers for enzymes and are not the catalysts itself which was also believed by the Nobel Prize laureate Richard Willstätter [1]. In the 1930s Northrop and Stanley were able to prove that this assumption was wrong. In their work, they revealed that pure proteins as pepsin, trypsin and chromotrypsin can indeed be enzymes [3]. Another milestone in the history of enzymes was set in 1965 by David Chilton Phillips and co-workers who performed the crystallization and structural elucidation of lysozyme *via*

X-ray [4]. This technique is still of relevance nowadays for the structural elucidation of novel proteins, the understanding of their catalytic reactions and in this respect the development of new drugs.
To understand the nature of proteins itself, it is of necessity to obtain an overview of the general structure and functions of proteins. Therefore, the next section is dedicated to this topic.

3.2 Structural features and functions of proteins

Proteins are one of the most important macromolecules which can be found in living organisms. They have a variety of different and decisive functions in nearly all biological processes. They play a role in the transport or the storage of other molecules, possess mechanical functions, are responsible for the immunity, are involved in the signal transduction by transferring nerve impulses or control the growth and cell differentiation [5].
Proteins consist of monomeric sub units, so called amino acids, which can be linked hence forming peptides (oligomeric/polymeric) and resulting in a chain. This is exemplary shown for the tripeptide consisting of serine, glycine and leucine in Figure 3.1-A. If one solely sets his focus on the linear chain one refers to the primary structure. Still, proteins are able to form a three-dimensional structure which can take, depending on the amino acid sequence, different shapes. The secondary structure can be obtained if amino acids interact with each other due to close interactions as hydrophilic ones (e.g. hydrogen bonds) or hydrophobic interactions (Coulomb interactions or π-stacking) (Figure 3.1-B). The existence of these alpha helices and beta sheets were already proposed in 1951 as primary motif of proteins by Pauling Corey and Branson [1]. If above mentioned interactions are formed at larger distances one obtains the tertiary structure (Figure 3.1-C). For some proteins a quaternary structure is also possible when a group of monomers (tertiary structures) interact with each other (Figure 3.1-D). The manifold possibilities to combine amino acids give the specific shape of proteins which is deterministic for the function of the protein. The resulting structures can differ in form and flexibility. Rigid structures serve as structural components in the cytoskeleton or in connective tissues while flexible proteins allow a merger between proteins or other molecules. These interactions result in an exchange of signals or information between cells by signal transduction and thus evoke different effects.

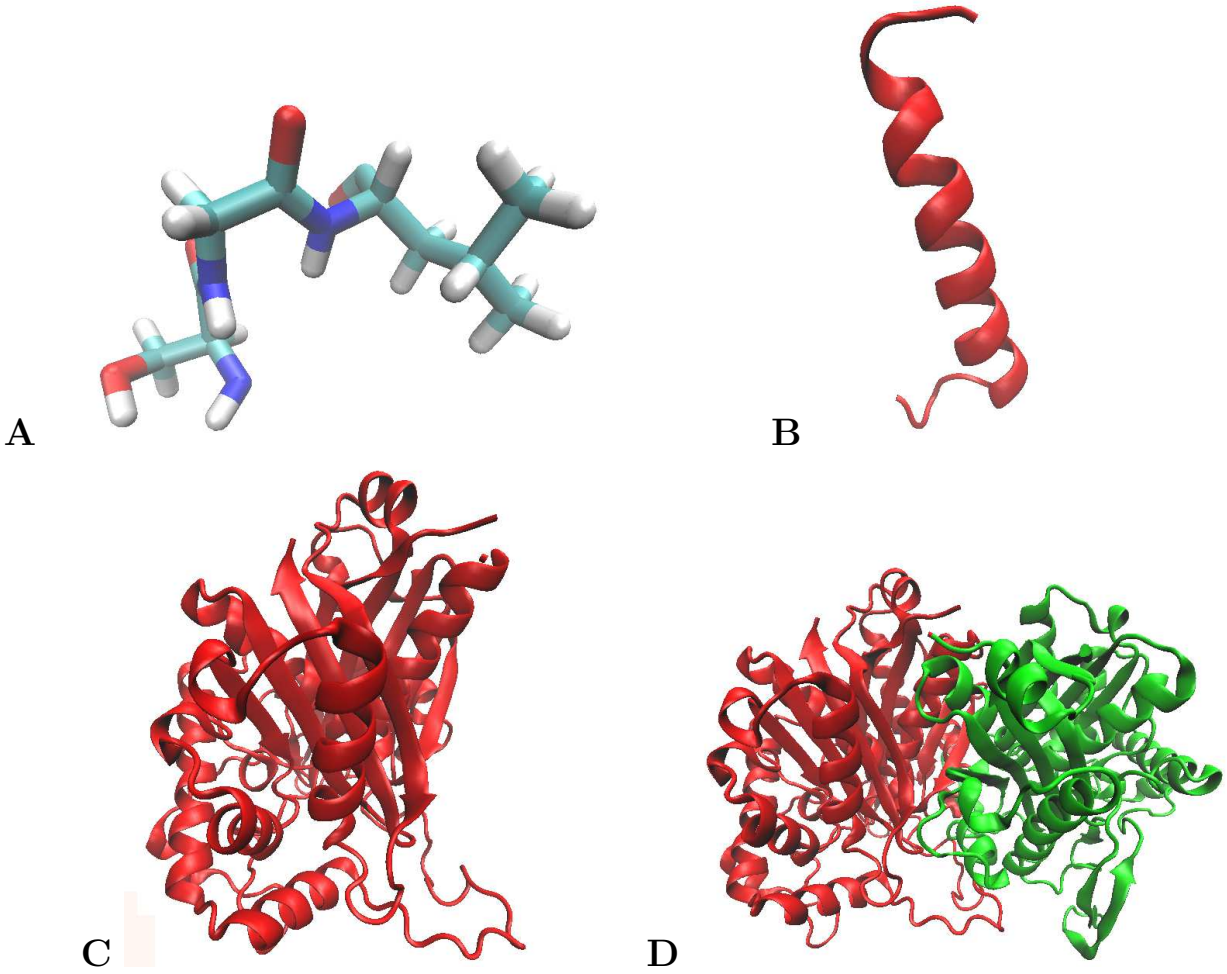


Figure 3.1: **A** – Primary structure of tripeptide Ser22, Gly23, Leu24 in FadA5 C93S variant (pdb code = 4UBT). **B** – Secondary structure of FadA5 C93S variant. **C** – Tertiary structure of FadA5 C93S variant **D** – quaternary structure of of FadA5 C93S variant (dimer).

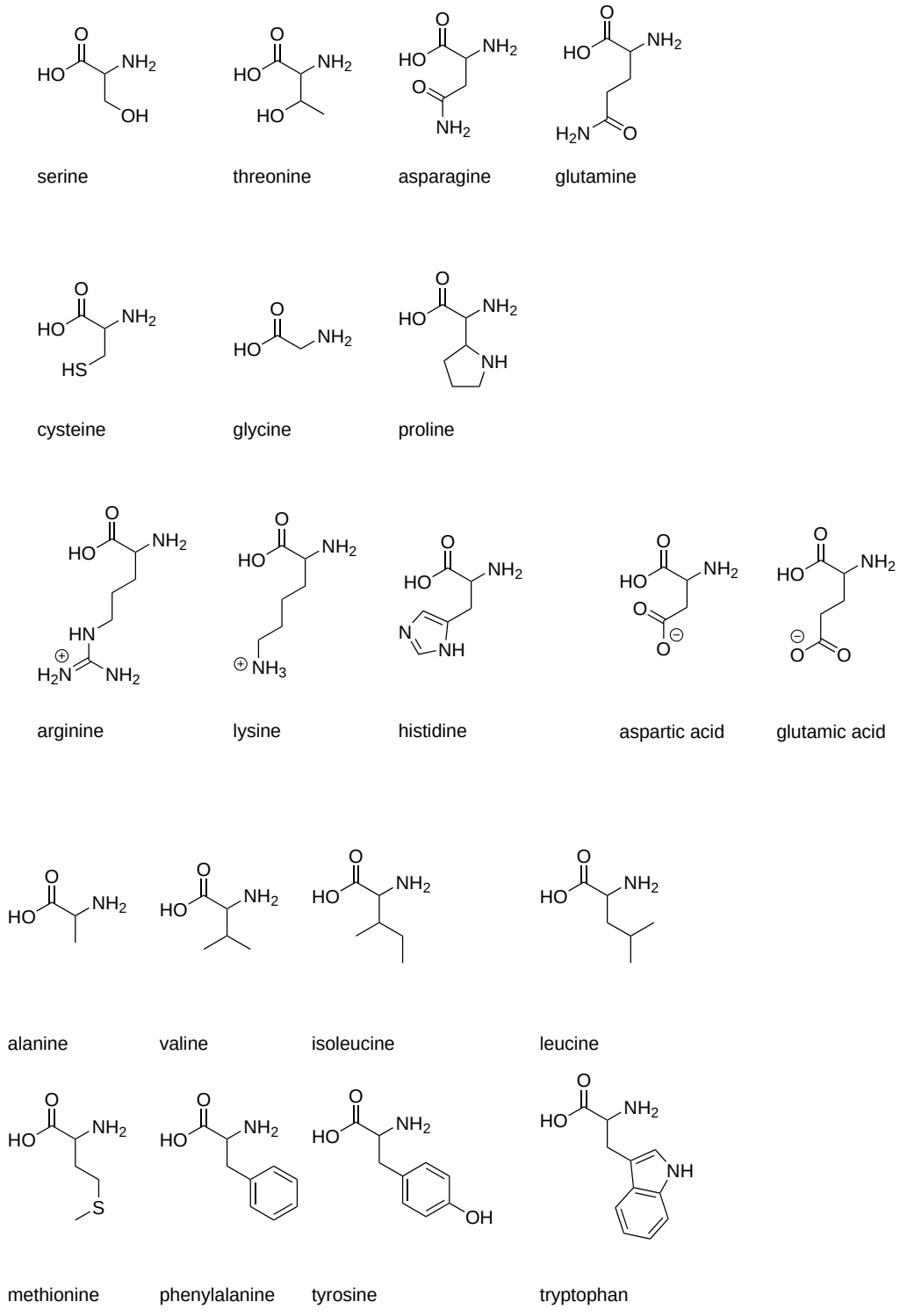


Figure 3.2: Chemical structures of proteinogenic amino acids at physiological pH 7.4.

The flexibility of proteins is determined by the functional groups of the amino acid side

chain (e.g. thiols, alocohols, carbon acids or amines). This set of amino acids differs in size, shape, hydrophilicity, hydrophobicity or charge (Figure 3.2). The spectra of interactions depend on the connection of these amino acids and thus dictate the structure of the protein. Apart from the structural role, side chains determine the catalyzed chemical reaction and its reactivity and are responsible for the function of enzymes.

3.3 The enzyme classes

Due to the variety of enzymes and their different catalytic characteristics the International Union of Biochemistry founded the Enzyme Commission in the year 1955 to classify these enzymes [1]. If one takes a look at the swissprot database one can find over 6000 different enzymes today which proves the relevance of the classification (https://enzyme.expasy.org/). Enzyme classes are organized by the enzyme commission number (EC) depending on the chemical reaction which they catalyze. The six main classes are:

- EC1: Oxidoreductases which catalyze the oxidation or reduction.
- EC2: Transferases which catalyze the transfer of groups.
- EC3: Hydrolases which promote the transfer of functional groups to water by a hydrolysis reaction.
- EC4: Lyases which add or remove groups to support the development of double bonds.
- EC5: Isomerases which lead to intramolecular transfer of groups (isomerization).
- EC6: Ligases which catalyze the ligation of two substrates due to ATP-hydrolysis.

In the following only enzymes are described which are part of this work. A more comprehensive description of the specific investigated enzyme can be found in each chapter.

3.3.1 Proteases

The metabolism of proteins in living organisms is an important process which leads to the "recycling" of proteins. Unnecessary proteins are degraded to amino acids which can be used for resynthesis of new proteins. This proteolytic process is catalyzed by proteases (EC 3.4). The general mechanism of proteases is subject of hydrolytic reactions, depicted in Figure 3.3, which can be described as the addition of a water molecule to a peptide.

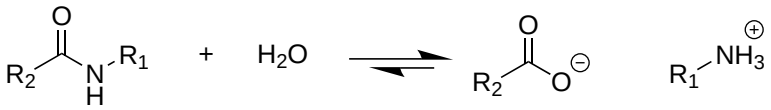


Figure 3.3: General reaction mechanism of proteases due to hydrolytic reactions.

The reaction is thermodynamically favored but the reaction rate is low. Due to the resonance structure of the peptide bond a resistance towards hydrolytic cleavages occurs [5]. The electrophilic character of the carbon is lowered by the partial double bond between the α-C atom and the nitrogen. In consequence, the reactivity of the nucleophilic reaction is decreased. Therefore, proteases are of necessity to activate these non-reactive carbonyl groups to favor the cleavage process of the peptide.
This large group of protease can be divided by two factors: the location where they act or the residue which they address. Endopeptidases (EC 3.4.21-99) split internal peptide bonds, while exopeptidases (EC 3.4.11-19) catalyze reactions at the N- or C-terminus. Furthermore, one can find serine (EC 3.4.21), cysteine (EC 3.4.22), aspartic (EC 3.4.23), metallo (3.4.24) and threonine proteases (EC 3.4.25).
In the course of this work rhodesain from *Trypanosoma brucei* (T.b.) *rhodesiense*, cruzain from *Trypanosoma cruzei* and cathepsin L have been investigated which belong to the class of cysteine proteases. Cysteine proteases can be found in many organisms such as viruses, bacteria or protozoa. Especially in the latter case most of the isolated proteases belong to the parasitic protozoa. A cysteine is located in the active center which is activated by a nearby histidine functioning as nucleophile. In contrast to serine proteases the histidine residue is sufficient and no catalytic triad is necessary (5).

3.3.2 Glycosyltransferases

Glycosltransferases (EC 2.4) are a subgroup of transferases which favor the production of complex carbohydrates. It becomes clear that a multitude of enzymes are needed since manifold glycosidic bonds can be formed. One can differentiate glycosyltransferases into three different groups: hexosyltransferases (EC 2.4.1), pentosyltransferases (EC 2.4.2) and transferases which transfer different glycosyl gropus (EC 2.4.99). In all examined organisms, glycosyltransferases solely account for 1-2% of the gene products (5).
The catalytic reaction of these enzymes can be schematically depicted as in Figure 3.4. Often, glucose is bound to nucleotides to achieve an increase of the molecules' energy conversation [5]. This monosaccharide nucleotide UDP-Glucose (uridine diphosphate) transfers a carbo-

hydrate to an acceptor substrate as amino acid residues of proteins (e.g. serine, threonine or asparagine), lipids or other carbohydrates (5). In this work the glycosyltransferase levansucrase (EC 2.4.1.10), a hexosyltransferase from *Bacillus megaterium* has been investigated.

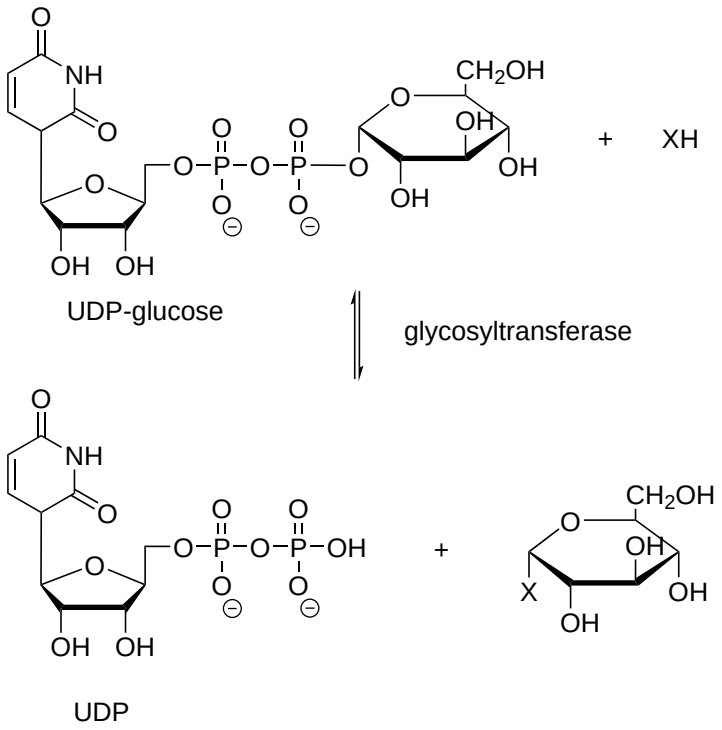


Figure 3.4: Transfer of glucose from UDP-glucose to acceptor substrate XH catalyzed by glycosyltransferase. This figure is adapted from [5].

3.3.3 Proteinkinases

Proteinkinases are another class of transferases and shall be considered in more detail. In eukaryotic cells the phosphorylation serves as regulatory mechanism for metabolic processes. It is known that 30% of the eukaryotic proteins are phosphorylated [5]. These phosphorylation reactions are catalyzed by protein kinases. These enzymes belong to one of the largest protein families with 100 homologous enzymes in yeast and over 500 enzymes in humans. Due to the variety of enzymes it is possible to ensure the precise regulation of place (specific tissue), time or substrate. Protein kinases are divided as tyrosine (EC 2.7.10), serine/threonine kinases (EC 2.7.11), dual-specifity kinases which act on serine/threonine and also tyrosine (EC 2.7.12), histidine kinases (EC 2.7.13), arginine kinases (EC 2.7.14) and other kinases (EC 2.7.15-99).

The reaction which is catalyzed by these enzymes is the transfer of a terminal phosphoryl group from adenosine triphosphate (ATP) to the respective amino acid of the acceptor protein. In eukaryotes a transfer to an alcohol group occurs as it is in serine, threonine or tyrosine (Figure 3.5). The serine/threonine kinase AKT1 (EC 2.7.11.1) was investigated in this work and will be further discussed in chapter 9.

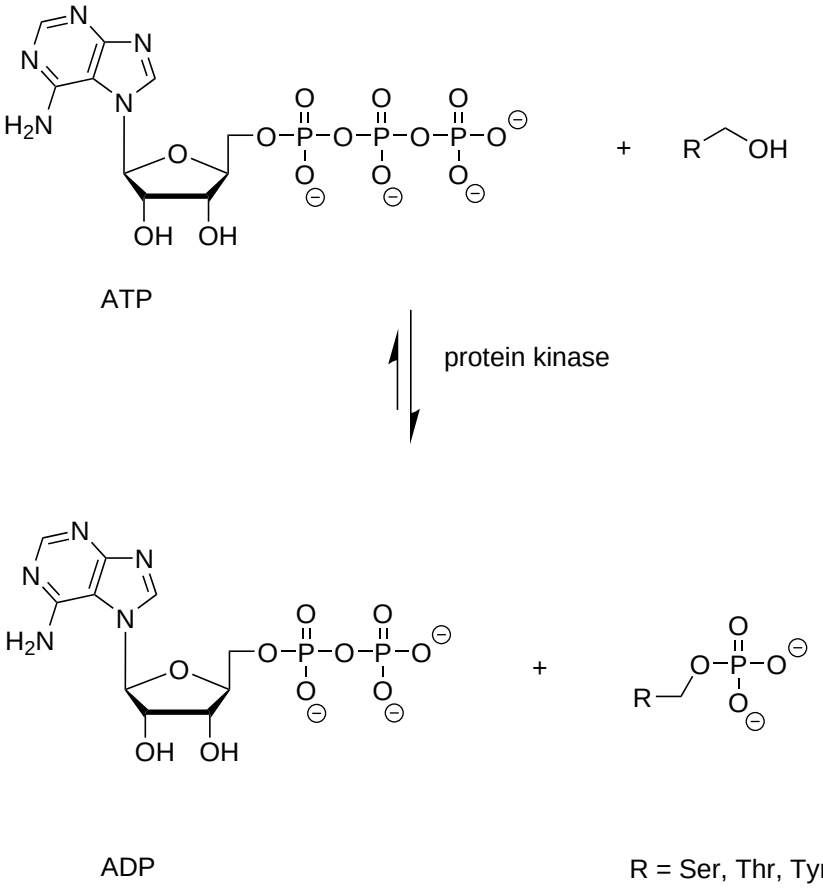


Figure 3.5: Transfer of phosphoryl group of ATP to residue (Ser, Thr, Tyr) catalyzed by protein kinase resulting in adenosine diphosphate (APD) and phosphorylated residue. This figure is adapted from [5].

3.3.4 Acetyltransferases

Acyltransferases are another representative of the group of transferases. One can split them in transferases which do not transfer amino-acyl groups (EC 2.3.1), aminoacyltransferases (EC 2.3.2) or such which convert acyl groups into alkyl groups on transfer (EC 2.3.3). In this work, FadA5 was investigated, an Acetyl-CoA C-acyltransferase (EC 2.3.1.9) or also known

as thiolase from *Mycobacterium tuberculosis*. Thiolases catalyze the reaction of 3-ketoacyl-CoA with CoA. On this occasion the thiol group of CoA reacts with 3-ketoacyl-CoA leading to a 2 C-atom shortened Acyl-CoA by cleavage (Figure 3.6).
Studies on the thiolase of *Zoogloea ramigera* showed that this reaction underlies a so-called ping-pong reaction [6]. In this case, the nucleophilic Cys89 reacts with the 3-ketoacyl-CoA and forms a covalent Acyl-CoA intermediate [7]. Due to another CoA the product is cleaved from the enzyme [8]. These reactions play an important role in the steroid synthesis or in the degradative pathways of the β-oxiation of fatty acids [9], [10].

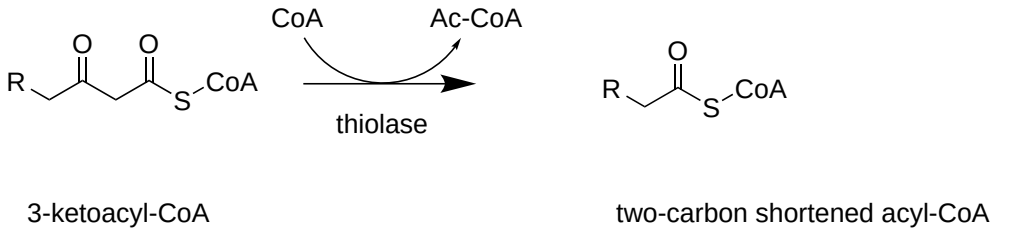


Figure 3.6: Degradation of 3-ketoacy-CoA to two-carbon shortened acyl-CoA catalyzed by thiolase.

3.3.5 Oxidoreductases

Oxidoreductases catalyze the electron transfer from one molecule (reductant) to the electron doner (oxidant) as depicted in Figure 3.7. Substrates can be various organic compounds as ketones, amines, or alcohols, anorganic molecules as sulphites or metals as mercury [11]. In most cases these enzymes utilize nicotinamide adinine dinucleotide cofactors as $NADP^+$ or NAD^+ to transfer a hydride (H^-) group to oxidize a substrate [11], [12], [13]. A prominent example is the mitochondrial electron transport chain which is essential for the respiratory chain in living organisms. In this case, electrons are transferred from NADH to oxygen by four different oxidoreductases: NADH-Q-oxidoreductase, Q-cytochrome-c-oxidoreductase, cytochrome-c-oxidase and the succinate-Q-reductase [14]. Tryparedoxin (TPX) of *Trypanosoma brucei* belongs to the class of reductases which catalyzes reactions where disulfide acts as an acceptor. It is part of a redox-cascade and will be further discussed in chapter 7.

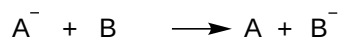


Figure 3.7: Catalyzed reaction of oxidoreductases by oxidation of one species and simultaneously reduce the other.

3.3.6 Oxidoreductases acting on the CH-CH group of donors

The enoly-(acyl-carrier-protein) reductase (EC 1.3.1.9) or abbreviated as ENR, belongs to the main enzyme class of the oxidoreductases and is an essential enzyme in the type II fatty acid synthesis (FAS) system [15]. It catalyzes the reduction of a trans-2-acyl-ACP (acyl carrier protein) to an acyl-ACP species by means of NADH or NADPH which act as reducing agent (Figure 3.8). These enzymes are a good target for the development of new antibiotics as they play an important role in the metabolism of many bacteria. Additionally, the bacterial ENR sequence is differently compared to the mammalian fatty acid biosynthesis, thus a selective targeting *via* drugs is possible. In chapter 10 the ENR of *staphyloccocus aureus* FabI and of *mycobacterium tuberculosis* InhA are further discussed.

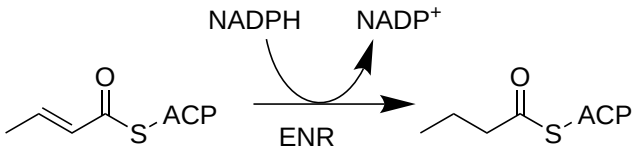


Figure 3.8: Catalyzed reduction of trans-2-acyl-ACP to a saturated acyl-ACP species.

3.4 Michaelis-Menten kinetic

Enzymes catalyze different reactions and lower reaction barriers to enable faster processes. To investigate the reaction rate of enzymes the best way is to observe the increase of reactands in the course of time. To achieve this experimentally one can determine the concentration of the substrate S and the over time increasing concentration of product P. At some point no change of S or P is observable. In this occasion the substrate is still converted into the product by the enzyme, though the backward reaction occurs contemporary so that one reaches a reaction equilibrium.

If one focuses on the forward reaction, one can define the reaction velocity V_0 which stands for the emerging mol of product per second at the beginning of reaction ($t \approx 0$) (Figure 3.9-A). In this case, we assume that the backward reaction is negligible thus the amount of available enzyme remains constant. When V_0 is applied to the substrate concentration one obtains the curve depicted in Figure 3.9-B. This behavior is observed in many enzymes [5].

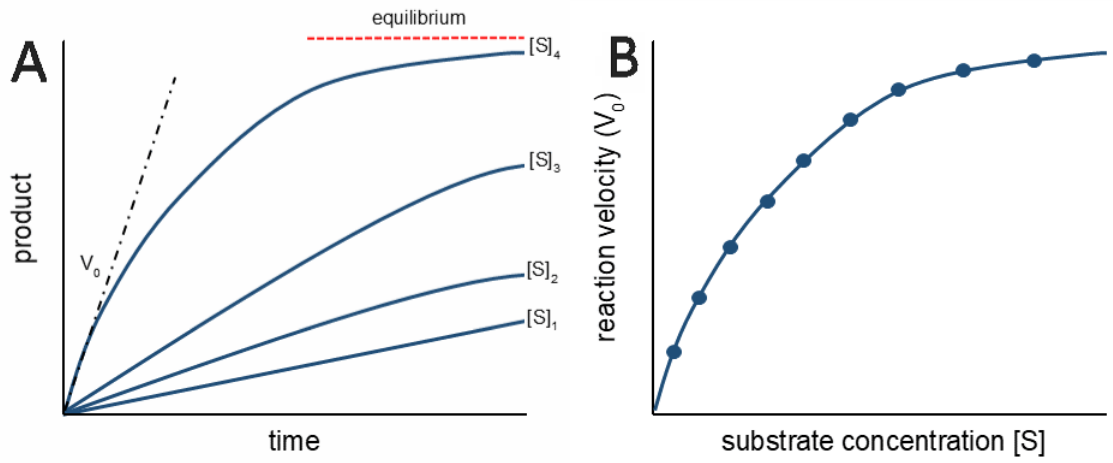


Figure 3.9: **A** - Amount of products obtained at different concentrations of substrate in the course of time. V_0 refers to the initial rate which is obtained by the slope of each curve at the beginning of the reaction. **B** – The reaction velocity V_0 obtained from the slopes at different concentrations plotted against the substrate concentration. This figure is adapted from [5].

An explanation for these kinetic properties of enzymes was proposed in 1913 by Leonor Michaelis and Maud Menten [16]. The crux of this matter is the consideration of an enzyme-substrate complex (ES) which occurs as an intermediate in the catalysis:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} E + P$$

As one can see in this chemical equation the ES complex is build with a rate constant k_1. This complex can be degraded back to $E + S$ again or be converted to E and P. Still in the latter case it can degrade back to the ES complex. Since we are solely considering the reaction rate at the beginning, we can set the rate constant k_{-2} to zero due to the fact that the concentration of the product has to be very low. This results in the following equation:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \underset{}{\overset{k_2}{\rightleftharpoons}} E + P$$

Consequently, one has to measure the reaction velocity V_0 (at different substrate concentrations) right at the beginning before the accumulation of P occurs. As a result, one obtains the kinetic of enzymes which obey the rules of the Michaelis-Menten kinetic (Figure 3.10).

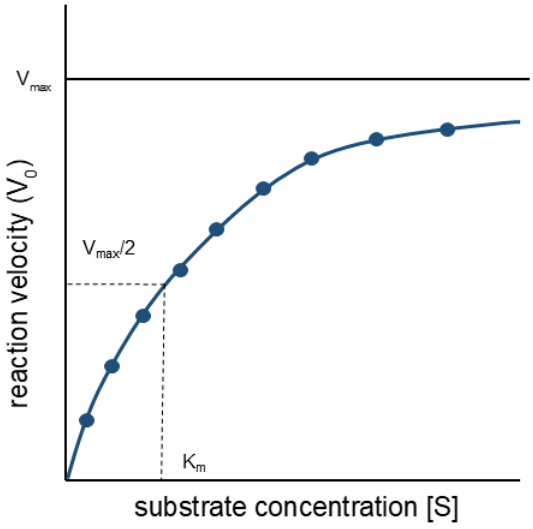


Figure 3.10: The reaction velocity V_0 plotted against the substrate concentration. The Michaelis constant K_M is the substrate concentration in which the reaction rate is half of V_{max}, whereas V_{max} is the maximum reaction rate. This maximum rate can be estimated by evaluating the value in the asymptotic limit. This figure is adapted from [5].

The graph depicted in Figure 3.10 can be described with the following equation, the Michaelis-Menten equation:

$$V_0 = V_{max} \frac{[S]}{[S] + K_M} \tag{3.1}$$

By means of equation 3.1 important properties can be derived. In this context V_0 stands for the reaction rate, V_{max} refers to the maximal possible reaction rate, K_M the Michaelis-Menten constant and [ ] represents the concentration, in this case of the substrate.
At low substrate concentrations ($[S] << K_M$) one acquires $V_0 = (V_{max}/K_m)/[S]$ and thus obtains a first-order reaction, since reaction rate and substrate concentration are proportional. If the substrate concentration is larger than K_M ($[S] >> K_M$) one obtains a reaction of zeroth order as $V_0 = V_{max}$ and in consequence the reaction is independent from the substrate concentration.
In case of $[S] = K_M$ we get $V_0 = \frac{V_{max}}{2}$ and on this occasion the reaction rate reaches half of its maximum value at the substrate concentration K_M. The Michaelis constant is an important characteristic in enzyme-catalyzed reactions and plays substantial role in the quantification of enzyme kinetic [5].

3.5 Inhibition of enzymes

Enzymes regulate various and important processes in organisms as in humans but as well as in bacteria, protozoa or in viruses. In case of humans, a defective function or a dysregulation of enzymes can result in diseases. If one disturbs enzymes of viruses or bacteria it is possible to contain these diseases (virostatic or bacteriostatic) or cure them (virucide or bactericide). Therefore, it is of importance to understand the function of each enzyme to modulate them respectively to e.g. counteract dysfunctions or specifically inhibit bacterial enzymes to gain a therapeutic benefit.

3.5.1 Different ways of inhibition

The inhibition of enzymatic catalysis can be divided into two different types. On the one hand, one refers to the reversible inhibition in which the inhibitor develops weak interactions with the enzyme to form an enzyme-inhibitor (EI) complex. This complex is not stable and dissociation can occur. On the other hand, the dissociation of the EI complex is weak or non-existent in case of an irreversible inhibition.
The reversible inhibition of enzymes can occur as competitive, uncompetitive or noncompetitive inhibition. If one determines the enzyme kinetics it is possible to draw conclusions in terms of the inhibition type. The irreversible inhibition of enzymes can be achieved by the formation of a covalent bond between inhibitor and enzyme. In this occasion, one distinguishes between different groups. Group-specific reagents modify the amino acide side-chain covalently which is located in the active site and also essential for the catalysis. Reactive substrate analogs also bind at the active site but additionally mimic the native enzyme substrate. Suicide inhibitors bind as substrate-analog at the active site and lead due to catalysis to an auto-inactivation of the enzyme [5].
Beside orthosteric inhibitors which interact with the active site of an enzyme, allosteric inhibitors have to be mentioned which can bind at a different location [17]. These inhibitors have two different mechanisms: In the first case one can consider the following scenario. If no allosteric inhibitor is bound to the active site, the substrate can bind and thus the enzyme can catalyze the formation of the product. If the allosteric inhibitor binds to the enzyme, the active site can be blocked so that the inhibitor hinders the entrance of the substrate. In the second case, the binding of the allosteric inhibitor modulates the active site and through this modulation the active site cannot be targeted by the substrate [18].

Reversible inhibition

Competitive inhibition – In case of the competitive inhibition a blockade of the enzyme's active site occurs due to the binding of the inhibitor. Consequently, no natural substrate can form a complex with the enzyme and the catalysis of the enzyme is reduced. As a concurrent binding of substrate and inhibitor is not possible, one can neutralize the inhibition by excess of substrate, since the developed $E - I$ complex solely forms weak interactions. The degree of dissociation of the $E - I$ complex can be measured by the dissociation constant K_i.

$$K_i = \frac{[E][I]}{[EI]} \tag{3.2}$$

If one looks at equation 3.3, it is evident that strong inhibitors possess a low dissociation constant. In this context K_M^{app} refers to the apparent K_M of the reaction. Furthermore, an increase of K_M occurs due to a competitive inhibitor although V_{max} remains the same in a saturated substrate-enzyme complex [5].

$$K_M^{app} = K_M + (1 + \frac{[I]}{K_i}) \tag{3.3}$$

Uncompetitive inhibition – One refers to uncompetitive inhibition if the inhibitor solely binds to an enzyme-substrate complex. Thus, a neutralization of inhibition cannot be achieved by an excess of substrate. In contrast to competitive inhibition K_M^{app} as well as V_{max} can be modified.

Noncompetitive inhibition – This form of inhibition occurs if a concurrent binding of substrate and enzyme occurs. Thus, the inhibitor can bind to the enzyme to form the EI complex or it binds to the ES complex. An excess of substrate cannot lower the inhibition. Although K_M remains the same after inhibition a reduction of V_{max} can occur and in consequence one obtains a new variable V_{max}^{app} which is defined as:

$$V_{max}^{app} = \frac{V_{max}}{1 + [I]/K_i} \tag{3.4}$$

Irreversible inhibition

The irreversible inhibition which is evoked by a covalent binding of an inhibitor and a aminoacid side chain of an active side residue, can be differentiated into two categories.
The first describes a covalent binding which results from a preceeding interaction of inhibitor and enzyme. Consequently, the inhibition is a two step mechanism which can be described

by the following chemical equation:

$$\mathrm{E} + \mathrm{I} \underset{}{\overset{K_I}{\rightleftharpoons}} \mathrm{E{\cdot}I} \xrightarrow{k_{inact}} \mathrm{E\text{-}I}$$

In a first step a reversible enzyme-inhibitor ($E \cdot I$) complex is formed which can be quantified by the binding constant K_I. K_I describes the necessary concentration of inhibitor to achieve half of the maximum potential rate of covalent bond formation [19]. This should not to be mixed up with the dissociation constant K_i. The second step is the formation of a covalent $E - I$ complex which depends on the concentration of the $E \cdot I$ complex. If solely the reversible $E \cdot I$ complex is available and no free enzyme remains, the observed rate of inactivation is k_{inact}. This rate describes the formation of a covalent $E - I$ complex and is defined as the first-order rate constant [19]. To describe the overall reaction from unbound enzyme to the covalent enzyme-inhibitor complex one uses the ratio of both, resulting in k_{inact}/K_I.
The second category of covalent-irreversible inhibition follows a different mechanism: by an immediate covalent formation without the presence of a non-covalent complex. In this occasion, the reactivity of the nucleophile of the enzyme and the electrophilic part of the inhibitor (also known as warhead) is so high that the efficiency of the reaction is equivalent to k_{inact}/K_I. This inhibition can be described with the following chemical equation:

$$\mathrm{E} + \mathrm{I} \xrightarrow{k_{inact}/K_I} \mathrm{E\text{-}I}$$

Covalent-reversible inhibition

Covalent-reversible inhibitors take a different approach. These inhibitors form a covalent bond with the enzyme but the activity of the enzyme can be regained if the bond is cleaved. Although the covalent binding is reversible, the binding affinity is about three magnitudes larger compared to non-covalent inhibitors [20]. One can divide covalent-reversible inhibitors in two types: The first one show a reversible inhibition because of their chemical mechanism which results in a weak bonding. Thus, the covalent complex can dissociate to the non-covalent complex or even the free enzyme. The second type are described as “slow substrates” which are rather recognized as substrates. The cleavage of the covalent bond results from a catalyzed reaction by the enzyme and leads to the release of a product which does not serve as an inhibitor anymore. This kind of reaction occurs at lower rates hence one refers to “slow

substrates" if such inhibitors are described [21].
Covalent-reversible inhibitors contain different groups as e.g. an aldehyde as electrophilic group which can form a hemiacetal with nucleophilic residues in the active site [22]. Beside aldehydes, there are other functional groups which can form a reversible covalent bond. Bortezomib, a proteasome inhibitor, undergoes a reversible boronic ester transesterification due to its boronic acid [22], [23]. Nitriles can also act as warheads to form a reversible covalent bond [24]. In case of Michael acceptors, α,β-unsaturated carbonyl systems, react with different nucleophiles which are carbon-, nitrogen-, sulfur- or oxygen-based [22]. To utilize them as covalent-reversible inhibitors one can change the electron density of the double bond to increase or decrease the reactivity to tune these kinds of warheads [25].

Chapter 4

Theoretical background

4.1 Molecular Mechanics

As biochemical systems are usually large (e.g macromolecules, organelles or cells), calculations applying quantum mechanical methodologies for the whole system are unfeasible. This is why one makes use of force field (FF) methods (also known as molecular mechanics). In contrast to quantum mechanics, the electronic energy is written as a parametric function of the nuclear coordinates which is parametrized by experimental data or high-level quantum mechanical calculations [26]. Additionally, force field methods use atoms and do not consider electrons as individual particles so that the binding information has to be explicitly specified. The dynamics of these atoms are described with classical mechanics, i.e. Newtons law of motion. If one connects these atoms to form a molecule one uses the "ball and spring" model. Atoms differ in this case by their size and the "softness" and the binding results from the bond's length and rigidity [27], [28].
Force field methods make use of the fact that molecules consist of subunits which show the same properties independent from the molecule which they are contained with. If one considers at the linear alkane chain $CH_3(CH_2)_nCH_3$ the amount of energy contribution for the heat of formation is practically the same for all CH_2 subunits [26]. The concept of molecules comprising structural similar subunits is realized in force field methods by the introduction of so-called atom types.

4.1.1 Molecular mechanical energy term

To describe the energy of molecules, the calculation of the force field energy E_{ff} is necessary. E_{ff} is the total energy of a system and can be described as a sum of different terms which

expresses the distortion of atoms and intermolecular interactions:

$$E_{ff} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{el} + E_{cross} \tag{4.1}$$

The individual terms are:

- the energy function for stretching the bond between two atoms (E_{str})
- the energy which needs to be provided to bend an angle (E_{bend})
- the torsional energy resulting from a rotation around a bond (E_{tors})
- non-bonded inter- and intramolecular interactions of the Van der Waals and Coulomb parts (E_{vdw} und E_{el})
- the coupling of the first three terms (E_{cross})

As the energy sub-terms depend on the atomic position and yield an energy with respect to the atomic coordinates R_i, it is also possible to determine the derivatives of the energy with respect to R. This is necessary in order to perform geometry optimization.

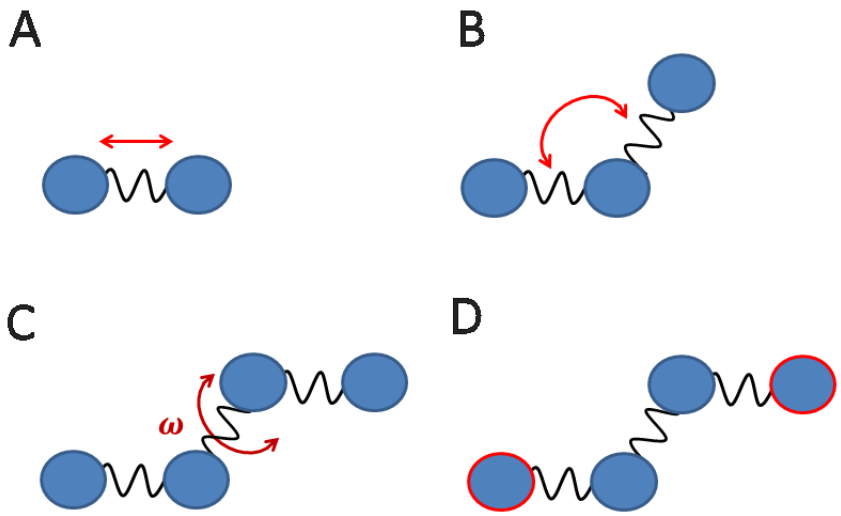


Figure 4.1: Illustration of force field energy terms of the **A** – stretch term, **B** – bending term, **C** – torsional term, **D** – non-bonded terms.

The stretch energy term

The energy function E_{str} describes the stretching of a bond between two atoms as depicted in Figure 4.1-A. E_{str} is expressed as a Taylor series expansion around the equilibrium bond length R_0 which is terminated after the second order resulting in the following equation:

$$E_{str}(R^{AB} - R_0^{AB}) = E(0) + \frac{dE}{dR}(R^{AB} - R_0^{AB}) + \frac{1d^2E}{2dR^2}(R^{AB} - R_0^{AB})^2 \tag{4.2}$$

The first term of equation 4.2 is usually set to zero, as this is the zero point of the energy scale [26]. In consequence E_{str} can be simplified to:

$$E_{str}(R^{AB} - R_0^{AB}) = k^{AB}(R^{AB} - R_0^{AB})^2 = k^{AB}(\Delta R^{AB})^2 \tag{4.3}$$

In this context k^{AB} represents the parametrized force constant between two atoms. The equation is written in the form of a harmonic oscillator, in which the potential for displacement around the equilibrium has a quadratically form. R_0 is parametrized within each molecular subunit and does not reflect the equilibrium pair distance for every molecule, as variations may occur during a minimization. It is possible to extend the Taylor expansion by including higher-order terms. However, this leads to a concurrent increase of parameters which need to be fitted. Furthermore, such an extension does not necessarily lead to an improvement especially in case of the description of distorting bond lengths [26]. Another problem in the harmonic approximation are the energies obtained for large bond displacements far away from the equilibrium arrangement. For such cases it is of importance to determine a function which describes the convergence of the dissociation energy at infinite distortion of the bond. With the help of a morse potential this criterion can be fulfilled.

$$E_{Morse}(\Delta R) = D(1 - e^{-\alpha \Delta R})^2 \tag{4.4}$$

$$\alpha = \sqrt{\frac{k}{2D}} \tag{4.5}$$

In this context D represents the dissociation energy and α is also related to the force constant k. Despite this, problems can occur by using a morse potential because of weak forces in long bonds [26]. Thus, a slow convergence result for obtaining the equilibrium bond length if poor (non-meaningful e.g. very long bond-lengths) starting structures are used.

The bending energy term

The bending energy can be described as the necessary energy to vary the angle between three atoms (ABC) (Figure 4.1-B). As it is the case for the stretch energy, a taylor expansion is applied up to second order:

$$E_{bend}(\theta^{ABC} - \theta_{\beta}^{ABC}) = k^{ABC}(\theta^{ABC} - \theta_0^{ABC})^2 \tag{4.6}$$

where k^{ABC} is the bending parameter, θ^{ABC} the bending angle and θ_0^{ABC} the ideal angle for the ABC triple of concerned atoms.

Out of plane bending energy term

A special case in terms of the bending energy is the out of plane bending energy. Looking at the angle of three atoms ABC of a molecule $ABCD$ and assuming that B is a sp^2-hybridized atom, it is understandable that a huge amount of energy is needed to obtain a pyramidalization [26]. To describe this distortion, an out-of-plane energy bending term (E_{oop}) is introduced as a quadratic function of the out of plane bending angle. This energy term depends on the angle χ or on the quadratic function of the distance d (Figure 4.2).

$$E_{oop}(\chi) = k^B\chi^2 \qquad E_{oop}(d) = k^B d^2 \tag{4.7}$$

The force constant k^B has to be individually determined for each sp^2-hybridized atom. The force constants can be very large to ensure the non-bending out of plane which is not often included in standard FFs. A balanced description with respect to the other bending/stretching parameters is needed.

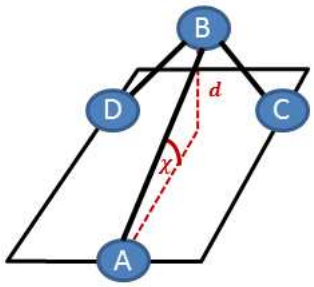


Figure 4.2: Illustration of out of plane variables (angle χ or distance d).

The torsional energy term

The rotation around dihedrals is best described by periodic functions i.e. the rotation within a full circle (360°) or half circle (± 180°) needs to be captured well. Thus, the rotation by 360° has to result in the same value. The torsional energy E_{tors} differ from $E_{stretch}$ and E_{bend} in further aspects. The rotation barrier does not only depend on E_{tors} but also on non-binding interactions, i.e. electrostatic or Van der Waals terms. Also, the variation of these torsions leads to low energetic differences [26]. Therefore, Taylor expansions are not suitable for the torsional energy as they lead to big changes of the minimum structure [26]. E_{tors} is usually incorporated as a Fourier series which accounts for the above mentioned periodicity:

$$\begin{aligned} E_{tors}(\omega^{ABCD}) = &\frac{1}{2}V_1^{ABCD}[1 + cos(\omega^{ABCD})] \\ &+\frac{1}{2}V_2^{ABCD}[1 - cos(2\omega^{ABCD})] \\ &+\frac{1}{2}V_3^{ABCD}[1 + cos(3\omega^{ABCD})] \end{aligned} \tag{4.8}$$

To describe the torsional energy correctly, the torsional parameters V_1^{ABCD}, V_2^{ABCD} and V_3^{ABCD} have to be determined for each combination of the four atom types A, B, C and D. These parameters reflect the height of barrier for a rotation around a dihedral [26].

Van der Waals energy term

The Van der Waals energy E_{vdw} describes non-bonding interactions between atoms, i.e. repulsive or attractive forces. On this occasion, electrostatics do not play a role and it can rather be described as non-polar interactions between atoms. At intermediate distances attractive dipole-dipole interactions can arise which are referred as dispersion or as "London" forces [29]. As a consequence, E_{vdw} approaches zero for large pair-distances and shows a minimum if atoms "touch" each other and increases at short distances. To describe these repulsive and attractive interactions the Lennard Jones potential is often used which possesses a R^{-6} and a R^{-12} dependence.

$$E_{LJ}(R^{AB}) = \frac{C_1}{(R^{AB})^{12}} - \frac{C_1}{(R^{AB})^6} \tag{4.9}$$

C_1 and C_2 are constants which are needed to define this energy function. The application of the exponent 12 and 6 is due to an enhanced computational performance, although the exponent of 9 or 10 can give better results [26].

Electrostatic energy

The other non-bonded interactions arise due to (partly) positively and negatively charged atoms. This can be accomplished by the assignment of partial charges to each atom. The interaction of each atom is described by the Coulomb potential:

$$E_{el}(R^{AB}) = \frac{Q^A Q^B}{\epsilon R^{AB}} \tag{4.10}$$

Q^A and Q^B refer to the partial charges of atom A and B, R^{AB} expresses the distance between both atoms and ϵ stands for the dielectric constant.

Cross terms

Cross terms describe the coupling of the previous mentioned terms. They are hence an interplay of these terms to provide a correct description when e.g. the bond depends on the angle (and vice versa). The E_{cross} terms are expressed as a product of a first order Taylor expansion. Exemplary the stretch/bend term is shown:

$$E_{str/bend} = k^{ABC}(\theta^{ABC} - R_0^{ABC})[(R^{AB} - R_0^{AB}) - (R^{BC} - R_0^{BC})] \tag{4.11}$$

These constants do not depend on all atoms but mostly from the central atom, in case of equation 4.11 of atom B ($k^{ABC} = k^B$).

4.1.2 Molecular Dynamic simulations

To describe mechanisms and interactions of biological systems, it is necessary to include dynamical effects. Proteins for example can show motions of several to hundreds of Ångström [26]. Some biological mechanisms can only be fully understood by inclusion of a dynamic description. The binding of oxygen to myoglobin or haemoglobin for example can only occur due to dynamical changes of these proteins [30]. Molecular Dynamic (MD) simulations are advantageous to perform investigations of such biological systems. The term Molecular Dynamics refers to the simulation of the motion of molecules but they are also used as as a tool to refine crystal structures or NMR data [30]. For these simulations it is of great importance to have a satisfactory description of the interactions between the particles, hence a valid description of the inter- and intramolecular forces is needed. Once calculated, these forces are applied in the equation of motion. In most cases the classical Newtonian equations of motion are sufficient to describe the movement of these particles. MD simulations start

with the acquisition of the system's energy as a function of the atom coordinates. The forces which act on the atoms are the first derivatives of the potential energy with respect to the atomic coordinates. These forces can be used to solve the Newtonian equation of motion and thus one obtains the motion of each particle as a function of time. For certain problems this technique can offer a deeper insight and delivers ready to use interpretation.
One can run simulations for the period of several picoseconds up to microsecond timescale to simulate the properties like densities but the cost ratio is comparably low and so it can be conducted within a time period of less than one day on server computer system. Simulations of a homogenous system such as a water cluster in a box with periodic boundaries can be calculated within seconds to determine structural or dynamic properties [31]. Taking a closer look at proteins, one has a heterogeneous system, hence an increased amount of computational effort is needed. Naturally the size of the examined system is also of importance in terms of computational costs. To perform these simulations one usually uses an X-ray structure as a starting point. The general procedure can be described as follows:
At the beginning preliminary calculations are carried out to obtain the relaxation of the system. Firstly, one starts by applying an optimization algorithm to minimize "local stress" which arise due to close atoms or bond length or angle distortions [30]. Afterwards a random velocity ϑ is assigned to the atoms which results from a maxwellian distribution at low temperature and the simulation is conducted for a few picoseconds [30]. In this context one uses the Newton equation

$$F_i = m_i a_i \tag{4.12}$$

where a_i represents the acceleration of an atom which is determined by the force F_i of the particular atom. F_i is calculated by the derivatives of equation 4.1 and depends on the position of the atom while m_i stands for the respective atomic mass. By obtaining the acceleration a_i it is possible to calculate the position r_i at a specific time t:

$$r_i(t + \Delta t) = r_i(t) + \vartheta_i \Delta t + \frac{1}{2} a_i (\Delta t)^2 \tag{4.13}$$

The time step Δt is usually set to one to two femtoseconds. By alterating the assigned velocity obtained from a maxwelian distribution at a chosen temperature the equilibration is continued. The temperature T can be measured by the mean of kinetic energy of the system:

$$\frac{1}{2} \sum_{i=1}^{N} m_i \left\langle \vartheta_i^2 \right\rangle = \frac{3}{2} N k_b T \tag{4.14}$$

In equation 4.14 N represents the number of atoms, $\langle \vartheta_i^2 \rangle$ refers to the square of mean velocity of the ith atom and k_b stands for the Boltzmann constant. If the temperature stays stable for a certain amount of time, one can assume that the atomic momenta are governed by a Maxwell distribution and each region of the protein has the same temperature [30].
To control the course of simulation one can calculate the root mean square deviation (RMSD) of the protein backbone which is mostly a "distance RMSD" [32]. The RMSD is a single-number measure which quantifies differences of structures A (e.g. starting structure) and B (structure after a certain amount of simulation time). It can be described as average distance of simulated structures in relation to a reference position r_i^{ref} (e.g backbone of crystal structure or starting structure of simulation) according to the following equation:

$$RMSD = \sqrt{\frac{1}{n}\sum_{i=1}^{N}(r_i - r_i^{ref})^2} \tag{4.15}$$

On the one hand the RMSD value can show the stability of the crystal structure and on the other hand it can evaluate the convergence of the MD simulation analyzing the fluctuations in the course of simulation time.

4.1.3 Docking

To model interactions between small molecules and a protein, docking simulations can be conducted. This model is used to investigate the binding of molecules (e.g. drugs) in the binding site and thus one can draw conclusions on biochemical processes [33]. To perform these docking simulations two steps are needed: Firstly, one predicts the ligands conformation, position and also orientation in the binding site of the protein (referred as pose). Secondly, one evaluates the stability of docked poses by applying scoring functions. Both steps are known as sampling and scoring in literature. If the binding site is well described the prediction is rather accurate. The first docking approaches were initially based on Fischer's lock-and-key theory [34], [35]. To obtain a better description, one can think of a rigid model which includes the flexibility of the protein. This can be related to the ligand-protein interaction as described in the "induced-fit" theory of Koshland [36]. Nowadays, the usage of a flexible ligand and a rigid protein is a popular method to perform docking simulations [37]. This procedure is used because it is not as time consuming as the simulation of a flexible ligand and a flexible protein. Still, a huge number of possible poses occurs because of the large number of internal as well as translational and rotational degrees of freedom.

Therefore, different algorithms have been developed to give the best prediction. This section summarizes the most common algorithms used in docking programs.

Matching algorithms (MA) [38], [39]
MA uses the ligands' molecular shape map to insert it into the active site of a protein, while both are represented as pharmacophores (the part which is responsible for the pharmacological effect). The match is hence calculated by the distances of each pharmacophore while chemical properties such as hydrogen-bond acceptors or donors are included in the mapping procedure.

Incremental construction (IC) [40]
This methodology splits the ligand into fragments while rotational bonds are broken. Usually the largest fragment (or the one with the most significant role) is initially docked to serve as an anchor. Afterwards the remaing fragments are attached incrementally. The flexibility of the ligand is thus guaranteed by the generation of different orientations which are supposed to fit into the active site.

Genetic Algorithms (GA) [41], [42], [43]
Genetic algorithms are well-known stochastic methods. Inspired by genetic information of living organisms', genes in GA refer to the ligand's degree of freedom which are described by binary strings. The pose of the ligand is therefore represented by the "chromosome" which are made up of these genes. Two different operators are of importance for GA. One is enabled to insert mutations randomly to evoke changes of the genes. The other performs crossover which leads to a change between two chromosomes and both leading to a new ligand. Afterwards the resulting ligands are scored and those who pass the scoring will be used for the next evaluation. This algorithm is also used in the Gold docking program package [42].
As mentioned before, it is of necessity to evaluate the resulting poses, hence scoring functions are used to differentiate between correct or incorrect poses. These function estimate the binding affinity between ligand and protein. They can be divided in force-field-based, empirical or knowledge based scoring functions. The empirical ChemPLP function which is the default scoring function in the Gold program package uses a piecewise linear potential (f_{PLP}) to evaluate steric complementarity of protein and ligand [44]. The implementation of

ChemPLP in the Gold program package is realized by the following equation:

$$\begin{aligned} fitness_{PLP} = -(&W_{PLP} \cdot f_{PLP} + W_{lig-clash} \cdot f_{lig-clash} \\ &+W_{lig-tors} \cdot f_{lig-tors} + f_{chem-cov} + W_{prot} \cdot f_{chem-prot} \\ &+W_{cons} \cdot f_{cons}) - (f_{chem-hb} + f_{chem-cho} + f_{chem-met}) \end{aligned} \tag{4.16}$$

The equation considers the heavy atom clash potential or the torsional potential of the ligand ($f_{lig-clash}$ and $f_{lig-tors}$). Additionally, the fitness function includes covalent docking ($f_{chem-cov}$), flexible side chains or explicit water molecules ($f_{chem-prot}$) and also constraints (f_{cons}). The second part of the equation refers to the distance- and angle dependent hydrogen or metal bonding terms ($f_{chem-hb}$, $f_{chem-cho}$ and $f_{chem-met}$). These functions are additionally weighted by W_i, where i stands for the different potentials.

4.1.4 MMPBSA / MMGBSA

The methodologies MM/PBSA and MM/GBSA are a combination of molecular mechanic, the Poisson-Boltzman equation or generalized Born model and the surface area continuum solvation. The main goal of both methods is estimating the binding affinities of ligands in biological macromolecules, i.e. proteins. MM/PBSA or MM/GBSA rely on sampling but the focus is set on the end states resulting in a decrease of computational effort. The calculation of the free binding energy G_{bind} can be described by the following equation:

$$\Delta G_{bind} = \langle G_{PL} \rangle - \langle G_P \rangle - \langle G_L \rangle \tag{4.17}$$

The ensemble averages are indicated by the brackets, PL refers to the protein-ligand complex, P for the protein and L for the ligand. The free energy of each part can be calculated as a sum of different terms [45], [46].

$$G = E_{bnd} + E_{el} + E_{vdW} + G_{pol} + G_{np} - TS \tag{4.18}$$

The first three terms of equation 4.18 refer to the MM energy terms for bond, angle, torsion, electrostatic and van der Waals interaction. The polar contributions which result from the generalized Born model (or by solvation of the Poisson Boltzman equation) are represented by G_{pol}. The non-polar term G_{np} is estimated from a linear relation to the solvent accessible surface area (SASA). Furthermore, T stands for the temperature and the entropy S can be obtained by the normal-mode analysis of vibrational frequencies [47].

In the Poisson Boltzman (PB) approach one tries to solve the Poisson Boltzman equation which consists of two parts. On the one hand, it describes the change of the electrostatic potential in a medium consisting of a constant dielectric in relation to the charge density. On the other hand it contains the Boltzman distribution which is responsible for the ion distribution in the respective system [48]. The basis of this approach is the solvation of the partial differential Poisson equation [49], [50], [51]:

$$\nabla\epsilon(r)\nabla\theta(r) + 4\pi\rho(r) = 0 \tag{4.19}$$

In equation 4.19 $\nabla\theta(r)$ represents the electrostatic potential, $\epsilon(r)$ a dielectric function which depends on the coordinates. The charge density caused by the solute is expressed as $\rho(r)$.
The generalized Born approach is an approximation of the Poisson Boltzman equation and is expressed as follows [52]:

$$\nabla G^{GB}_{Born} = (1 - \frac{1}{\epsilon}) \sum_{ij} \frac{q_i q_j}{r^2_{ij} + \alpha^2_{ij} exp(-D_{ij})^{0.5}} \tag{4.20}$$

where $\alpha_{ij} = (\alpha_i \alpha_j)^{0.5}$ and $D_{ij} = \frac{r^2_{ij}}{2\alpha^2_{ij}}$
The atomic partial charge is represented by q_i and q_j while the double sum runs over every atom pair, also in the case $i = j$. The effective Born radius of an atom i is expressed by α_i while the ∇G^{GB}_{Born} term approximates the dielectric part of Coulomb's law, when atoms move beyond the contact distance of their Born radii. Atoms are modeled as charged spheres with a lower internal dielectric compared to the environment [45]. The screening of each atom is therefore regulated by the local environment, i.e. the more atoms surround e.g. atom i the larger it will be descreened (less of its electrostatic will be screened). Due to the descreening the Born radius α is therefore obtained for each atom. A large α value refers to a strong electric field, thus it is described with a strong electric field and reflects the situation of an atom in vacuum. In contrast to that, a small α value represents the opposite situation as if the atom was surrounded by a bulk of water [45]. Depending on the applied GB model the atomic descreening is calculated differently.
The MM/PBSA or MM/GBSA can be classified between docking and scoring and alchemical perturbation (AP) methods. Docking and scoring have low computational costs but are in many cases not as accurate in their prediction of the binding affinity. AP methods calculate the free energy difference between two states but contrary to MM/PBSA or MM/GBSA they do not set the focus only on the end states but also on the reaction coordinate connecting

both states. Just like MM/PBSA or MM/GBSA, AP methods are very accurate but the bottleneck of these methods is that an extensive sampling is needed. This leads to higher computational costs [47], [53].

4.2 Quantum mechanics

All quantum mechanical methods try to find a solution of the Schrödinger equation. The basic expression in that respect is the time-independent Schrödinger equation:

$$\hat{H}\psi = E\psi \tag{4.21}$$

$\hat{H}$ is the Hamilton operator (also called Hamiltonian) which acts on a wave function ψ so that one obtains the eigenenergy E. All information about the system is comprised in the wave function ψ as the total energy. The non-relativistic Hamilton Operator for molecular systems can be described as a sum of potential and kinetic energy operators of nuclei and electrons [26]:

$$\hat{H} = T_n + T_e + V_{ne} + V_{ee} + V_{nn} \tag{4.22}$$

T_n and T_e stand for the kinetic operator of the nuclei or the electrons, V_{ne}, V_{ee} and V_{nn} for the potential operators of the nuclei-electron, electron-electron or nuclei-nuclei, so that one can write the Hamilton operator explicitly as:

$$\hat{H} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \frac{1}{2}\sum_{A=1}^{M}\frac{1}{M_A}\nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>1}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}} \tag{4.23}$$

in atomic units.
In equation 4.23 i and j refer to the N electrons, A and B for the M nuclei, R for the distance between the nuclei and r for the distance between electron-electron or electron-nuclei. The charge of the nuclei is expressed by Z. The two first terms contain the laplacian operator which is a differential operator and is connected to the kinetic energy of electrons respectively nuclei.

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{4.24}$$

If one applies the Hamiltonian on the exact wave function the total energy of the system is yielded. The expectation value of the Hamiltonian is:

$$\langle E \rangle = \langle \psi | \hat{H} | \psi \rangle \equiv \int \psi^* \hat{H} \psi dV \tag{4.25}$$

One central aspect in quantum chemistry is the Born-Oppenheimer approximation. The motion of nuclei is much slower compared to electrons as they are much heavier [54]. Thus, one can think of a molecule where electrons move in the field of nuclei which are fixed. Therefore, the kinetic energy operator for the nuclei is not included in the electronic Hamiltonian description. Hence, one can often neglect the second term of equation 4.23 (the kinetic energy of the nuclei), while the repulsion of the nuclei can be seen as a constant (last term of equation 4.23). Therefore, the electronic Hamiltonian can then be written as:

$$\hat{H} = -\frac{1}{2} \sum_{i=1}^{N} \nabla_i^2 - \sum_{i=1}^{N} \sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N} \sum_{j>1}^{N} \frac{1}{r_{ij}} \tag{4.26}$$

Thus, the solution to a Schrödinger equation is the electronic wave fuction ψ_{el} if an electronic Hamiltonian is applied:

$$\hat{H}_{el} \psi_{el} = E_{el} \psi_{el} \tag{4.27}$$

The electronic wave function ψ_{el} describes the electron's motion and depends parametrically on the coordinates of the nuclei but explicitly on the coordinates of the electrons. The electronic energy E_{el} contains the approximated total energy of the system. QM calculations aim to find the wave function, usually for the ground state wave function, which is an eigenfunction of the Hamiltonian (in this context, the electronic Hamiltonian) to obtain relevant information of the system. The exact solution is only possible for an one-electron system, thus further approximations have to be made. One starts with a trial wavefunction ψ_{trial} and applies the variational principle which states that the energy expectation value of this trial wave function is larger or the same as the actual energy of the ground state [26]:

$$E_0 = \langle \psi_0 | \hat{H} | \psi_0 \rangle \leq E_{trial} = \langle \psi_{trial} | \hat{H} | \psi_{trial} \rangle \tag{4.28}$$

The search of this (closest to the exact) wave function is conducted within a subsets of possible wave functions. In the Hartree Fock (HF) model these subsets are Slater determinants which are made up of one electron functions (orbitals). They consider one important factor: The total electronic wave function has to be antisymmetric due to the spin of electrons. The

reason why two electrons cannot have the same quantum number results because of the antisymmetry which is anchored by the Pauli principle. This antisymmetry is enabled by Slater determinants.

$$\psi_0 \approx \Phi_{SD} = \frac{1}{\sqrt{N!}} \begin{bmatrix} \chi_1(\overrightarrow{\chi}_1) & \chi_2(\overrightarrow{\chi}_1) & \cdots & \chi_N(\overrightarrow{\chi}_1) \\ \chi_1(\overrightarrow{\chi}_2) & \chi_2(\overrightarrow{\chi}_2) & \cdots & \chi_N(\overrightarrow{\chi}_2) \\ \vdots & \vdots & \vdots & \vdots \\ \chi_1(\overrightarrow{\chi}_N) & \chi_2(\overrightarrow{\chi}_N) & \cdots & \chi_N(\overrightarrow{\chi}_N) \end{bmatrix} \tag{4.29}$$

In equation 4.29 $\chi(\overrightarrow{\chi})$ represents a spin orbital and is expressed as:

$$\chi(\overrightarrow{\chi}) = \phi(\overrightarrow{r})\sigma(s) \qquad \sigma = \alpha, \beta \tag{4.30}$$

The spin orbital contains a spatial part $\phi(\overrightarrow{r})$, while the spin is represented by $\sigma(s)$ and can be α or β. Another approximation is the assumption that the trial wave function consists of a single Slater determinant. This has the consequence that the electron-electron repulsion is included as an average effect. If one uses this single-determinant trial wave function one can apply the variational principle to minimize the energy using the Fock equation. Therefore, one expresses the energy in the HF method (in consideration of the Born-Oppenheimer approximation) by the following equation:

$$E = \sum_i^N \langle \phi_i | \, h_i \, | \phi_i \rangle + \frac{1}{2} \sum_{ij}^N (\langle \phi_j | \, J_i \, | \phi_j \rangle - \langle \phi_j | \, K_i \, | \phi_j \rangle) + V_{nn} \tag{4.31}$$

In equation 4.31 h_i represents the one-electron operator, J is the Coulomb operator describing the electrostatic repulsion between two electrons and K refers to the exchange operator. Both operators J and K are written as:

$$J_i \, |\phi_j(2)\rangle = \langle \phi_i(1) | \, g_{12} \, | \phi_i(1) \rangle \, |\phi_j(2)\rangle \tag{4.32}$$

$$K_i \, |\phi_j(2)\rangle = \langle \phi_i(1) | \, g_{12} \, | \phi_j j(1) \rangle \, |\phi_i(2)\rangle \tag{4.33}$$

where g_{12} is a two-electron operator which accounts for the electron-electron repulsion.
The next step is the determination of a set of molecular orbitals (MO) which minimizes the energy or make it stationary in regard to a change in the orbitals. A requirement of these variations is that the MOs has to remain orthonormal after change. These constrained

optimizations can be realized by Lagrange multipliers (λ_{ij})

$$\delta\mathcal{L} = \delta E - \sum_{ij}^{N} \lambda_{ij}(\langle\delta\phi_i\,|\phi_j\rangle + \langle\phi_i\,|\delta\phi_j\rangle = 0 \tag{4.34}$$

where $\mathcal{L}$ is the Lagrange function [26]. If one inserts the Fock Operator F_i one obtains

$$\delta E = \sum_{i}^{N}(\langle\delta\phi_i|\,F_i\,|\phi_i\rangle + \langle\phi_i|\,F_i\,|\delta\phi_i\rangle) \tag{4.35}$$

with

$$F_i = h_i + \sum_{j}^{N}(J_j - K_j) \tag{4.36}$$

F_i is an effective one electron operator which describes the kinetic energy of an electron and the interaction with all nuclei (attraction) and all other electrons (repulsion) realized by the J and K operators. If one combines the equations 4.34 and 4.35 and also diagonalizes the Lagrange multipliers one obtains Fock equations which are a set of pseudo-eigenvalue equations:

$$F_i\phi_i' = \epsilon_i\phi_i' \tag{4.37}$$

In equation 4.37 ϕ_i' refers to canonical MOs, while ϵ_i stands for the for the corresponding MO energies. This set of functions can be employed to determine the solution of one orbital ϕ_i' under the premise that all other orbitals are known. Hence, one electron feels the mean field of the other electrons (also referred as mean-field approximation).
Usually for investigating chemical problems one invokes a further approximation, the linear combination of atomic orbitals (LCAO) approximation. Therein, the orbitals are expanded as a linear combination of atomic centered basis functions:

$$\phi_i = \sum_{\alpha}^{M} c_{\alpha i}\chi_\alpha \tag{4.38}$$

where M stands for the number of basis functions. By inserting these basis functions, multiplying them from the left side and integrating equation 4.37, one obtains the Roothaan-Hall equation:

$$FC = SC\epsilon \tag{4.39}$$

The Fock matrix elements are present in the F matrix, overlap integrals between basis functions are included in the S matrix. The matrix C contains the coefficients of the expansion of the molecular orbitals into the atomic basis. The solution of the Roothan-Hall equation is accomplished by an iterative way. One starts with an initial guess of matrix C. Out of it the Fock matrix is built, diagonalized whereby a new set of better coefficients results. These new coefficients are used to built up a new Fock matrix which is diagonalized and hence optimized, until the change of the coefficients is below a given convergence criterion.

4.2.1 Møller Plesset perturbation theory

To further improve the solution of the Schrödinger equation, i.e. to come closer to the exact solution, one can expand the HF equations with further determinants. One possibility is the utilization of the Møller Plesset pertubation theory of second order (MP2) [55]. Originating from Hartree Fock the Hamiltonian $\hat{H}_0$ is extended with the perturbation operator $\hat{H}'$.

$$\hat{H} = \hat{H}_0 + \lambda \hat{H}' \tag{4.40}$$

This operator refers to the correlation energy and scales with λ [26]. The reference Hamiltonian $\hat{H}_0$ can be obtained if $\lambda = 0$, i.e. if no perturbation is present. This perturbation can be extended from zeroth order (MP0), first order (MP1) up to n-th order (MPn).

$$E(MP0) + E(MP1) + E(MP2) = E(HF) + E(MP2) \tag{4.41}$$

$MP0$ is merely the summation of the orbital energies.

$$MP0 = E(MP0) = \sum_{i=1}^{N_{elec}} \epsilon_i \tag{4.42}$$

In the sum of the orbital energies the electron-electron repulsion is counted twice. The first-order energy correction fixes the problem of the electron-electron repulsion at zeroth order.

$$W_1 = \langle \Phi_0 | H' | \Phi_0 \rangle = \langle V_{ee} \rangle - 2 \langle V_{ee} \rangle = - \langle V_{ee} \rangle \tag{4.43}$$

In consequence the sum of the energies of $MP0$ and $MP1$ result in the Hartree Fock energy.

$$MP1 = E(MP0) + E(MP1) = E(HF) \tag{4.44}$$

Most of the correlation energy is already recovered in MP2 (80-90%) and it scales only to the fitfth power, thus it has an excellent cost-benefit ratio (in relation to the system's size). In case of MP3 which accounts 90-95 % of the correlation energy, the computational effort increases to sixth power and worsen the ratio [26].

4.2.2 Coupled Cluster approaches

In Coupled Cluster (CC) methods corrections are included to the reference wave function in an infinite order [26]. The excitation operator $\hat{T}$ generates all possible electronic excitations of the order i (number of electrons which are present in the system) by acting on the reference wave function. Besides the Hartree Fock determinant, the CC wave function also includes additionally excited determinants on i-th order.

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + ... + \hat{T}_i \tag{4.45}$$

$\hat{T}_1$, $\hat{T}_2$, etc. are excitation operators for single, double or higher excitations respectively. $\hat{T}_1$ for example can be written as:

$$\hat{T}_1 \Phi_{HF} = \sum_{j}^{occupied} \sum_{k}^{virtual} t_{jk} \Phi_{jk} \tag{4.46}$$

i.e. it describes all single excitations from the occupied orbitals j to all virtual orbitals k. Usually these excitation operators act on HF reference wave functions Φ_{HF}. Thereby a set of excitations is generated which contain the expansion coefficients t_{jk} (in this context also called amplitudes) are later determined by solving the coupled cluster equation. The coupled cluster wave function is written as follows:

$$\psi_{CC} = e^{\hat{T}} \Phi_0 \tag{4.47}$$

$$e^{\hat{T}} = 1 + \hat{T} + \frac{1}{2}\hat{T}^2 + \frac{1}{6}\hat{T}^3 + ... = \sum_{k=0}^{\infty} \frac{1}{k!}\hat{T}^k \tag{4.48}$$

The abbreviation CCS, CCSD, etc. give the form of the excitation operator $\hat{T}$ which contain only single ($\hat{T} = \hat{T}_1$) or single and doubly ($\hat{T} = \hat{T}_1 + \hat{T}_2$), etc. excitation. Inserting its form in the exponential ansatz we obtain for example for the CCS:

$$\psi_{CCS} = e^{\hat{T}} \Phi_0 = e^{\hat{T}_1} \Phi_0 = [1 + \hat{T}_1 + \frac{1}{2}\hat{T}_1^2 + \frac{1}{6}\hat{T}_1^3 + ...] \Phi_0 \tag{4.49}$$

From equation 4.49 it is clear that ψ_{CCS} does not only contain single excitation but because of the $\hat{T}_1^k$ expression also all higher excitations which are obtained if $\hat{T}_1$ is applied twice ($\hat{T}_1^2$), three times ($\hat{T}_1^3$) and so on. If one focuses on the term $\hat{T}_1^2$ it becomes clear that the CCS wavefunction does not contain all double excitation but only the so-called unlinked clusters, i.e. those which can be generated by a double application of $\hat{T}_1$. Due to the inclusion of these unlinked cluster excitations, coupled cluster has some very favorable properties, e.g. it is size-consistent.
The disadvantage of these methods is that even CCSD (coupled cluster single double excitations) calculations scale with a power of 6 (in relation to the number of basis functions); calculation applying CCSDT, i.e. including triplet excitations scale with a power of 8 [56], [57]. Thus, this methodology is unfeasible for larger systems in terms of computational costs. One possibility to reduce the costs is to treat triplet excitations by perturbation theory. This method is called CCSD(T) which only scales with a power between 6 and 7 but still delivers precise results provided that a large basis set is used [58], [59]. The brackets in CCSD(T) indicate that the triplet excitation is treated with perturbation theory. CCSD(T) is one of the most important methods in CC approaches, as errors in energy calculations are of 1 kcal/mol and it is thus often referred as gold standard of quantum chemistry [60]. In this context one can mention the CC2 method, which is an approximation of the CCSD model and is also based on the perturbation theory [61]. This method is comparable to the MP2 method in terms of the precision as well as in scaling [26], [62], [63].

4.2.3 Spin component scaled calculations

The concept of the spin component scaling (SCS) is based on the separate scaling of the correlation energy of anti-parallel spin components $E_{\uparrow\downarrow}$ and the parallel spin components $E_{\uparrow\uparrow+\downarrow\downarrow}$. In this occasion two scaling factors are used [64].

- p_S is associated with dynamical correlation effects (short-ranged, $p_S = \frac{6}{5}$)
- p_T for non-dynamical ones (long-ranged, $p_T = \frac{1}{3}$)

$$E_{SCS} = p_S E_{\uparrow\downarrow} + p_T E_{\uparrow\uparrow+\downarrow\downarrow} \tag{4.50}$$

As solely empirical numbers are used, the scaling procedure can easily be integrated in many quantum mechanical computations without an increase in computational costs. With the SCS procedure one can improve results of MP2 as well as of CC2 [64], [65].

4.2.4 Density functional theory

Density functional theory (DFT) is an alternative to wave function-based approaches. The basis of DFT is the description of the ground state energy by the electron density ρ [66]. DFT methods do not calculate the energy as the expectation value of the wave function but as a functional of the electron density. In contrast to functions, which generate a number from a set of variables, functionals produce a number from a function.
In wave function approaches, one has 4N variables for each wave function (in a N-electron system): three spatial ones and one for the spin for each electron. Thus, wave functions become more complicated as the number of electrons increases. In contrast the electron density it is solely dependent on three spatial coordinates i.e. its complexity does not increase with the size of the system. However, while the one-to-one correspondence of wave function and energy is given by the Schrödinger equation, the corresponding relationship between the density and the ground state energy of a given nuclei orientation (the so-called external field) were only proven by the Hohenberg-Kohn theorem.
The problem which nowadays still remains is the connection between the electron density and the energy by an unknown functional. This is still an important topic in research [26], [67], [68]. If one starts to formulate an energy functional $E[\rho]$ it can be divided in three parts, the kinetic energy $T[\rho]$, the nuclei-electron interaction $E_{ne}[\rho]$ and the electron-electron repulsion $E_{ee}[\rho]$ which can be split into the Coulomb and exchange part $J[\rho]$ and $K[\rho]$.

$$E[\rho] = T[\rho] + E_{ne}[\rho] + E_{ee}[\rho] \tag{4.51}$$

with

$$E_{ee}[\rho] = J[\rho] + K[\rho] \tag{4.52}$$

In an uniform electron gas $T[\rho]$ and $K_D[\rho]$ can be described as

$$T[\rho] = C_F \int \rho^{\frac{5}{3}}(r)dr \tag{4.53}$$

$$K_D[\rho] = -C_X \int \rho^{\frac{4}{3}}(r)dr \tag{4.54}$$

with

$$C_F = \frac{3}{10}(3\pi^2)^{\frac{2}{3}} \tag{4.55}$$

$$C_X = \frac{3}{4}(\frac{3}{\pi})^{\frac{2}{3}} \tag{4.56}$$

The nuclei-electron interaction $E_{ne}[\rho]$ as well as the Coloumb part $J[\rho]$ can be written as:

$$E_{ne}[\rho] = -\sum_{a}^{N_{nuclei}} \int \frac{Z_a(R_a)\rho(r)}{|R_a - r|} dr \tag{4.57}$$

$$J[\rho] = \frac{1}{2} \int \int \frac{\rho(r)\rho(r')}{|r - r'|} dr dr' \tag{4.58}$$

The resulting energy functional is known as Thomas-Fermi theory [26].

$$E[\rho] = T[\rho] + E_{ne}[\rho] + J[\rho] \tag{4.59}$$

If one includes the exchange part $K_D[\rho]$ of Dirac (equation 4.54) one obtains the Thomas-Fermi-Dirac model. Still both models fail to predict chemical bonding as the assumption of the uniform electron gas seems to be too crude. The applicability of DFT in computational chemistry is due to the introduction of orbitals stated by Kohn and Sham [69]. In orbital-free models the kinetic energy is not well described thus the idea of this formalism is to split the kinetic energy functional into two parts: an exact calculable one and a correction term. Similar to HF method the formulas for the kinetic, Coloumb electron-electron and nuclei-electron energies are identical and the following general DFT energy expression results:

$$E_{DFT}[\rho] = T_S[\rho] + E_{ne}[\rho] + J[\rho] + E_{xc}[\rho] \tag{4.60}$$

In a hypothetical system, where electrons do not interact, $T_S[\rho]$ is the exact kinetic energy functional which calculates the kinetic energy from a Slater determinant (indicated by the subscript S), i.e. as HF.

$$T_S = \sum_{i=1}^{N_{elec}} \langle \phi_i | -\frac{1}{2}\nabla^2 | \phi_i \rangle \tag{4.61}$$

In real systems electrons do interact, thus equation 4.61 does not result in the total kinetic energy, still it is an improvement with respect to the uniform electron gas expression. The remaining part of equation 4.60 lies in the exchange correlation term $E_{xc}[\rho]$ which is composed of the kinetic energy connection term, the electron correlation and exchange energy terms.

$$E_{xc}[\rho] = (T[\rho] - T_S[\rho])(E_{ee}[\rho] - J[\rho]) \tag{4.62}$$

Although the DFT formalism is exact in principle, the crux of this approach is that the form of the exchange part is unknown thus it needs to be approximated by empirical functionals. The quality of DFT methods is described by Jacob's ladder of density functional approximations for the exchange-correlation energy [70]. Local Density Approximations (LDA) is the simplest approach and provides only a moderate accurate description, in terms of molecules. Generalized gradient approximation (GGA) methods offer an improvement to LDA as they include not only the electron density but also the derivatives of the density.
Generally, GGA methods approximate E_{xc} by the following equation:

$$E_{xc} = -\frac{1}{2}\sum_{\sigma}\int \rho_{\sigma}^{\frac{4}{3}} K_{\sigma} d^3r \tag{4.63}$$

with the dimensionless coefficient K_σ. K_σ is described by another dimensionless coefficient x_σ with

$$x_\sigma = \frac{|\nabla\rho_\sigma|}{\rho_\sigma^{\frac{4}{3}}} \tag{4.64}$$

The main difference of GGA exchange functionals results from on the dependence of K_σ and on x_σ [71]. The first GGA exchange functionals was introduced by Becke in 1988 as a correction of the LDA exchange energy ϵ_X^{LDA} [72].

$$\epsilon_X^{B88} = \epsilon_X^{LDA} + \Delta\epsilon_X^{B88} \tag{4.65}$$

$$\epsilon_X^{B88} = -\beta\sum_{\sigma}\int \rho_\sigma^{\frac{4}{3}} \frac{x_\sigma^2}{1 + 6\beta x_\sigma sinh^{-1}x_\sigma} \tag{4.66}$$

The parameter β was specifically determined to fit best to the exchange energy of the six noble gas atoms [72].
Meta-GGA methods are the next step to improve the DFT methodology. These methods include the dependency of exchange and correlation functionals on the second derivative of the electron density ($\nabla^2\rho$). Hybrid-GGA tackle another important problem of DFT which is the self-interaction, i.e. electrons interacting with each other. For HF this problem does not occur, as classical and non-classical electron-electron interactions negate each other for the same electron, which is not the case in KS-DFT [73]. To minimize resulting errors, HF exchange (often named exact exchange in a DFT context), is mixed in [73]. In general, those

hybrid functionals can be described by the following equation:

$$E_{xc}^{GH} = c_x E_x^{HF} + (1 - c_x) E_x^{DFT} + E_c^{DFT} \tag{4.67}$$

with c_x as "tunable" parameter. One of the most popular hybrid-GGA methods is the functional B3LYP after Becke, Lee, Yang and Parr [72], [74], [75], [76].

$$E_{xc}^{B3LYP} = (1 - a) E_x^{LDA} + a E_x^{exact} + b \Delta E_x^{B88} + (1 - c) E_c^{LDA} + c E_c^{LYP} \tag{4.68}$$

B3LYP already belongs to the semi-empirical hybrid GGA's, as the parameters a, b and c are fitted to experimental data. Hybrid-GGA methods can additionally be extended to hybrid-meta GGA as in the Minnesota functionals, e.g. M05, M06, M08 or M011 [77], [78], [79], [80].
One of an important improvement to DFT methods is the dispersion correction introduced by Grimme in the year 2004, which has been further improved in the course of years [81], [82], [83]. As many gradient-corrected functionals describe dispersive interactions poorly he introduced a model-dependent empirical quantity which describes the missing dispersion [84], [85], [86]. This additional empirical dispersion energy is described as:

$$E_{disp} = -s_6 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{C_6^{ij}}{R_{ij}^6} f_{dmp}(R_{ij}) \tag{4.69}$$

and

$$f_{dmp}(R) = \frac{1}{1 + e^{-\alpha(R/R_0 - 1)}} \tag{4.70}$$

In the equations 71 and 72 N refers to the number of atoms, C_6^{ij} to the dispersion coefficient between the atoms ij, s_6 to scaling factors referring to the specific atom type and R_{ij} to the interatomic distance, while R_0 stands for the atomic van der Waals radii [81]. The improved performance of DFT due to this correction has been tested in various cases and is always recommended to include in calculations [87], [88], [89], [90].

4.2.5 COSMO

The inclusion of explicit solvent molecules can be computationally demanding as an increase of system size occurs. One solution is offered by continuum solvent models which describe the embedding of a solute molecule by a dielectric continuum with the permittivity of ϵ forming a cavity within the dielectric. They are also referred as implicit solvent models. This interface

between dielectric and cavity is also known as solvent accessible surface (SAS)and can be obtained by applying probe spheres which consider the size of the solvent molecules [91]. Due to the charge distribution of the solute a polarization of the dielectric medium occurs resulting in the generation of screening charges on the surface as response of the medium. In COSMO the polarizable continuum is defined as a perfect conductor, by setting $\epsilon = \infty$ [92]. Hence, surface charges are screened perfectly and additionally electrostatic problems at the boundary are simplified. The introduction of an empirical scaling function $f(\epsilon)$ enables a subsequent scaling of surface charges from an ideal screened density $\sigma(s)^{ideal}$ to consider the actual solvent permitivitty ϵ [91]. Thus one obtains the apparent surface charge $\sigma(s)^{app}$.

$$\sigma(s)^{app} = f(\epsilon)\sigma(s)^{ideal} = \frac{\epsilon - 1}{\epsilon + k}\sigma(s)^{ideal} \tag{4.71}$$

In equation 4.71 k refers to an empirical factor which is obtained by experimental data [92].

4.3 QM/MM Methodologies

In 1975 Warshell and Levitt introduced the QM/MM (quantum mechanical/molecular mechanics) concept, which is still essential for the calculation of enzymatic reactions [93]. In 1990 the coupling between semi-empircal QM methods and the CHARMM (Chemistry at Harvard Macromolecular Mechanics) force field has been conducted and compared with ab initio and experimental data with respect to the precision and effectiveness [94], [95]. Nowadays QM/MM methods are established, and one does not only find them applied in biomolecular systems but also in inorganic/organometallic or solid state systems [96].

4.3.1 QM-MM partitioning

In the QM/MM approach the system (e.g. protein-ligand complex) is generally divided into two parts. The "inner" part is described with quantum mechanical methods, the "outer" one with molecular mechanics (Figure 4.3). Within the inner part the atoms which are involved in reactions are included, so that bond breaking and forming can be treated by using QM methods in a sufficient accurate way. The outer part, treated e.g. by a force field, folds in the influence of the environment on the reaction. Due to the severe interaction between both parts, the energy cannot be calculated as a sum of both subsystems and therefore coupling terms are needed. Especially at the boundary one has to be cautious, as covalent bonds are

cut. The QM part can be extended with so called "link atoms" which are not part of the system and will be described later on.

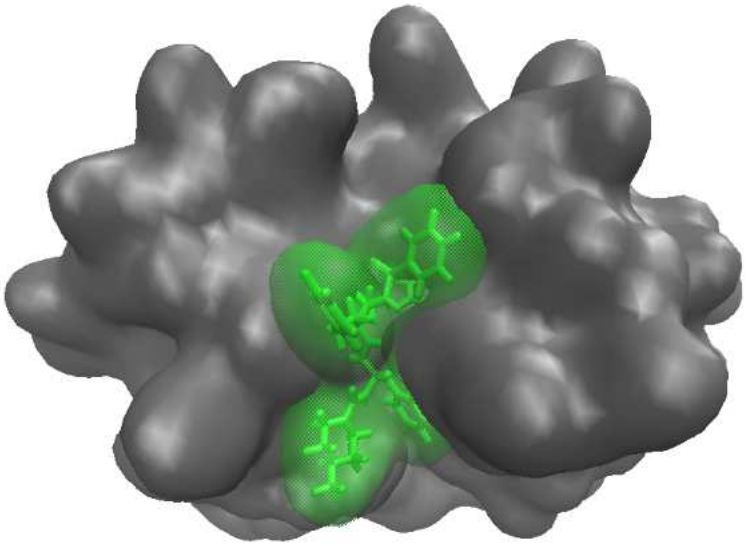


Figure 4.3: QM/MM partitioning of a protein ligand complex. The whole system (S) consists of two parts: The outer region (O) depicted in gray and the inner region (I) depicted in green.

4.3.2 QM/MM energy Expression

Subtractive and additive QM/MM schemes

To describe the QM/MM energy one has two different possibilities: the subtractive and additive QM/MM scheme. In case of the subtractive scheme one needs a MM calculation of the whole system as well as of the inner subsystem and a quantum mechanical one of the inner part. Hence, the energy is calculated by the following equation:

$$E^{sub}_{QM/MM}(S) = E_{MM}(S) + E_{QM}(I+L) - E_{MM}(I+L) \tag{4.72}$$

S refers to the whole system, I to the inner part and L for the link atoms. The advantage of the subtractive procedure is the simplicity as no QM/MM coupling terms are needed, so that standard QM and MM methods can be applied without any modifications [96]. As long as MM force terms and link atoms do provide acceptable QM potentials, artifacts which are caused by the link atoms are corrected by the subtraction. The disadvantage of this method is that a good parametrization of the inner MM part is needed. Additionally, the coupling of both parts is only performed on MM level. Especially in the latter case the problem of

electrostatic interactions arises, which are caused by the Coulomb interactions between fixed atomic charges in the QM and MM areas. Still, the application of electrostatic embedding enables the inclusion of MM charges into the QM Hamiltonian so that a better description is possible [96].
Another option to obtain the QM/MM energy is by using the additive QM/MM scheme.

$$E^{add}_{QM/MM}(S) = E_{MM}(O) + E_{QM}(I+L) - E_{QM-MM}I, O \tag{4.73}$$

In contrast to the subtractive procedure the MM calculation is only performed for the outer part and one uses explicitly a coupling term, in this case $E_{QM-MM}(I, O)$, which gathers the interaction of both terms [96]. The inner part including link atoms are described as in the subtractive scheme by quantum mechanical approaches. The coupling term can be described by the following equation:

$$E_{QM-MM}(I, O) = E^{b}_{QM-MM} + E^{vdW}_{QM-MM} + E^{el}_{QM-MM} \tag{4.74}$$

and contains besides bonded interactions also van der Waals and electrostatic interactions between the QM and MM region.

Electrostatic QM/MM interaction

The coupling between the QM charge density and the charge model of the MM region can be accomplished by mechanical embedding, electrostatic embedding or polarized embedding. The mechanical embedding treats the QM-MM electrostatic interactions at the same basis as MM-MM electrostatics. That means that the charge model of the MM part (in most cases atomic point charges) is transferred to the QM part [96]. The computational costs are low but disadvantages do occur. Charges of the outer part do not interact with the QM density, which are in consequence not directly polarized by the electrostatic environment [96]. If a change in the charge distribution of the QM region occurs (e.g due to a reaction) it is necessary to update the charges, though it would lead to large fluctuations in the potential energy surface. Also, MM point charges of the inner region have to be described well. These disadvantages can be avoided if a QM calculation is performed in the presence of the MM charge models.
In the electrostatic or electronic embedding scheme the electronic structure of the inner region is adapted to the changes of the charge distribution of the environment, so that they are automatically polarized by it. This can be calculated by inserting MM point charges as

one-electron terms into the QM-Hamiltonian.

$$\hat{H}^{el}_{QM-MM} = -\sum_{i}^{N}\sum_{J\epsilon\Phi}^{L}\frac{q_J}{|r_i - R_J|} + \sum_{a\epsilon I}^{M}\sum_{J\epsilon\Phi}^{L}\frac{q_J Q_a}{|R_a - R_J|} \tag{4.75}$$

The MM point charges q_J are located at R_J, while the nuclear charges Q_a of the QM atoms are placed at R_a. The positions of the elctrons are described with r_i. The double sum runs over i, J and a of N electrons, L point charges and M QM nuclei. As mentioned before, one has to be cautious if boundaries run through covalent bonds. In this instance, MM charges are set to the close proximity of the QM electron density and this leads to overpolarization. Electrostatic embedding schemes include the polarization of the QM part by the MM part but neglects variation arising from the polarization of the MM part by the QM part. This can be included by the usage of polarizable force fields.

Non-bonded and bonded QM/MM interactions

The non-bonding van der Waals interaction of the QM part are described by a Lennard Jones potential; hence suitable parameters are needed. Often one uses the atom types of the force field but problems might occur if QM atoms change their character during a reaction. In most cases this does not represent a problem as short-ranged interactions are involved. Still, this can be solved by moving the QM-MM boundary further away from the QM atoms.

Overview of Boundary schemes

As mentioned before within QM/MM the treatment of covalent bonds which crosses the QM/MM boundary is an critical point. In case of enzymatic reactions in biological system this is inevitable and can result into problems. Firstly, the affected QM atom has to be capped and this cannot be achieved by heterolytically or homolytically cleavage. The overpolarization between the QM density and the MM charges has to be prevented, especially if link atoms are used. Also, double-counting of the interactions of both parts has to be circumvented. It is of importance to cut a unipolar bond which is possible for an aliphatic C-C bond but should be avoided in case of an amide bond. In general, three approaches can solve these problems:
In the link atom scheme an additional artificial atomic center L is included which is usually a hydrogen atom. The calculation of the inner part takes the link atom into account ($I + L$), while the generated bond is described on MM level. The problem is that due to the

insertion of the link atom three artificial degrees of freedom are generated. Also, the link atom is close to the QM electron density so that overpolarization can occur. Additionally, in chemical and electronical terms the link atom differs from the replaced atom. The problem of overpolarization can be reduced by deleting one-elctron integrals which are associated with the link atoms [94]. Another possibility is to delete the M point charges in the link region of the Hamiltonian or to shift/distribute the point charges in the link region [97], [98]. By these techniques the handling of link atoms improves so that they are widely used [96].
The boundary-atom scheme replaces the MM atom of the bond by a special boundary atom which occurs in the QM and MM calculation. On the MM side it is viewed as a normal MM atom, while in the QM calculation it represents the cleaved bond and the electronic character of the MM part attached to the QM atom.
In the localized-orbital scheme hybrid orbitals are set to the boundary and frozen. Hence, they cap the QM region and represent the cleaved bond.

Chapter 5

Computational details

5.1 MD simulations

The MD simulations were conducted using the AMBER14 software package [99]. Protein monomers or dimers were described by the AMBER forcefield ff14SB [100]. Ligands, inhibitors and cofactors have been parameterized by utilizing the generalized AMBER forcefield (GAFF) with the program antechamber [101], [102]. Electrostatic potential (ESP) charges of non-protein solutes have been calculated by the Gaussian 03 pgroam package on Hartree Fock level, employing the 6-31G(d) basis set, except for one case, which is explicitly mentioned in Chapter 7 [103], [104]. The protein(-inhibitor) complex was solvated in an octahedral shell in a 10 Å radius around the system employing the TIP3P water model [105]. Charges have been neutralized by Na^+ or Cl^- ions. The solvent was minimized for 1000 cylces while restraining the complex. Subsequently, another 2500 cycles were run without restraints on the solute. After gradually heating to 300 K over 100 ps and fixating the solvent, the simulation was continued at constant temperature (300 K) and pressure (1 atm). The duration of the simulation were adopted to the given question and is explicitly mentioned in the respective chapters. Time steps of 1 fs were chosen. Bonds involving hydrogen were constrained with the SHAKE algorithm and simulations were performed under periodic boundary conditions.

5.2 QM calculations

QM calculations as single point calculations or optimzations of investigated structures were performed by employing the TURBOMOLE V6.6 or Gaussian09 software package [106], [107], [108], [109]. 3d model structures were prepared with the Avogadro program package

[110]. Nudged elastic band (NEB) calculations have been performed with the CAST program package which utilized the MOPAC 2016 energy interface [111], [112].

5.3 QM/MM calculations

QM/MM calculations have been conducted by utilizing the Chemshell [113], [114]. Geometries have been optimized by the DL-Find geometry optimizer [115]. The QM part was calculated with the TURBOMOLE V6.6 software package while the MM part was described by the AMBER14 force field. The shift embedding scheme was utilized to describe the coupling. In all calculations the optimization of the MM part solely included atoms in an inner layer of 10 Å around the reactive thiol group, whereas the rest of the system was not optimized. The QM part is described in respective sections.

5.4 Graphics and data evaluation

The presented graphics have been prepared with the Origin 2016 software package [116]. The visualization of presented structure as well as the calculation of RMSD values were conducted by the molecular visualizationion program VMD [117]. Structural formulas wre generated with the Chemdraw software. In presented figures the color code of 3d visualized atoms are as follows: light blue refers to carbons, yellow to sulfurs, red to oxygens, blue to nitrogens and white to hydrogens.
The mean absolute error (MAE) has been calculated as

$$MAE = \frac{\sum_{i=1}^{n} |C - x_i|}{n} \tag{5.1}$$

where n is the number of data points, C the reference value and x the calculated values.
The standard deviation has been determined by the following equation:

$$\sqrt{\frac{\sum_{i=1}^{n} (\bar{x} - x_i)^2}{n}} \tag{5.2}$$

The number of data points is given with n, $\bar{x}$ is the mean of all values and x_i refers to the calculated value.

Chapter 6

Theoretical investigations of aromatic warheads for cysteine cathepsin protease inhibitors employing quantum mechanical calculations

6.1 Introduction

Cysteine cathepsin proteases are involved in nearly all proceses of lysosymes as for example the degradation of proteins, metabolism of lipids and proteins, antigen presentation, cellular stress signalling and lysosome-mediated cell death [118]. Cathepsin belongs to the C1 family (papain family) of the clan CA as most of cysteine proteases [119], [120]. They influence additionally the activation, processing or degradation of chemokines, cytokines or different growth factors [120], [121].

One can find 11 different members of the cathepsin family among humans [118]. Except for two representatives, cysteine cathepsins are mostly endopeptidases located within endolysomal compartments [118]. There is on the one hand cathepsin C (CTSC) or also known as dipeptidyl peptidase I and on the other hand cathepsin Z, a carboxymonopeptidase which cleaves dipeptides of the C terminus of substrate proteins [122], [123]. Still some extracellular functions can be found for specific cysteine cathepsins as e.g. cathepsin K, which is responsible for the bone remodeling by osteoclasts [124].

It is known for a long time that cathepsin expression is increased in tumor tissues compared to normal ones [125]. The prognosis for survival of cancers e.g. colorectal, head, neck breast

and lung cancer is related to the cathepsin expression [126], [127]. Cathepsins have a big contribution in respect to the progression of tumors in various cancer by different mechanisms [118]. Secreted cathepsins are potent effectors which can modify the environment of the tumor by turnover or degradation of the extracelluar matrix (ECM). Those secreted cathepsins bind on the cell membrane aft specific regions and are localized at endolysomal vesicles [128]. Depending on the region one obtains different substrates or functions [128].
Beside the progression another important effect of these cysteine proteases has to be mentioned. It was shown that they are involved in the response of anticancer therapy in the tumor microenvironment and thus, are of relevance in the development of resistance [129], [130].
Still, an increased expression is not only observable in tumor cells but also in tumor-associated cells e.g. endothelial cells or macrophages [124]. Additionally, dysfunctional cathepsin activity is not restricted to induce cancer but also various diseases as parasitic and viral pathogensis, cystic fibrosis, cardiovascular or neurodegenerative diseases, osteoporosis or arthritis [131], [132], [133], [134], [135], [136]. Therefore, it is understandable why cysteine cathepsins proteases represent good targets not only for the therapy of cancer but also for a lot of other diseases.
One strategy to modify these cathepsins is addressing them by means of covalent inhibitors. These inhibitors consist of a recognition unit which is responsible for the non-covalent interactions and a warhead which form a covalent bond with the respective protein. Residues as cysteine, serine or threonine side chains are often covalently modified by Michael acceptor systems as vinyl sulfones, unsaturated esters, nitrioles or ketones [22], [137], [138], [139], [140], [141], [142]. In contrast to that, one can rarely find warheads which react by electrophilic aromatic substitution (S_NAr) [143], [144].
The group of Opatz and Schirmeister from the University Mainz conducted investigations in terms of these warheads for cysteine proteases of the papain family, as cathepsin. They used a (S)-Phe-(S)-Leu-OBn dipeptide sequence as a recognition unit and added various aromatic S_NAr-type electrophiles at the N-terminus to obtain the covalent inhibitor (Figure 6.1). It was shown in the past, that this dipeptide sequence is suitable for cathepsin-like cysteine proteases [141], [142], [145]. Additionally, to various aromatic S_NAr-type electrophiles chloro- or nitrile-subsituted 1,4-naphtoquinones (compound 4 and compound 8) have been used as warheads by Schirmeister and Opatz group (Figure 6.2). As known in literature quinone, which are maleimide derivatives, can react with nucleophiles as thioles in proteins. Thus, a nucleophilic addition reaction at the Cβ-atom of the α,β-unsaturated carbonyl function with the sulfur of an active site cysteine residue is likely [146]. N-ethylmaleimide for example is

useful for the titration of thiole groups as the reaction occurs rapidly and is specific to these groups [147].
In this study experimental data indicate that both compounds inhibit cathepsin L in a competitive way and furthermore, dilutaion assays proven that the inhibition of cathepsin is reversible. As one can see in Figure 6.3 different reaction products are conceivable and therefore QM calculations have been performed to identify the product in lowest energy. As no crystal structure of the protein-inhibitor complex was available environmental effects were considered by employing the COSMO approach.

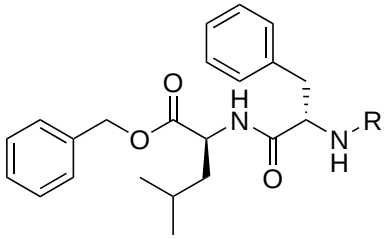


Figure 6.1: Dipeptide sequence (S)-Phe-(S)-Leu-Obn recognition unit. Aromatic warheads are attachaed at R.

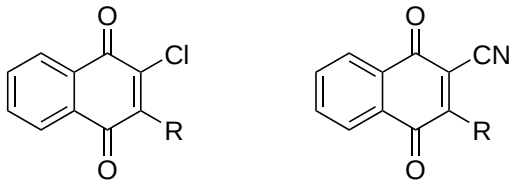


Figure 6.2: Chloro-subsituted 1,4-naphtoquinones (compound 4) and nitrile substituted (compound 8)

6.2 Benchmark of ab initio and DFT methods for geometry optimization

The initial step of the investigations started with the geometry optimization of reactants and possible reaction products of compound 4 and 8 with methanethiol (Figure 6.3). Therefore, different approaches like ab initio methods as HF and MP2 or the DFT approach B3LYP with Grimme's dispersion correction D3 were applied. These calculations employed the def-TZVP basis set and considered implicit solvation by COSMO (ϵ=78) [148]. To reduce computational

costs solely the 1,4-napthoquionone moiety (warhead) in connection to the methylamine of the recognition unit was considered.

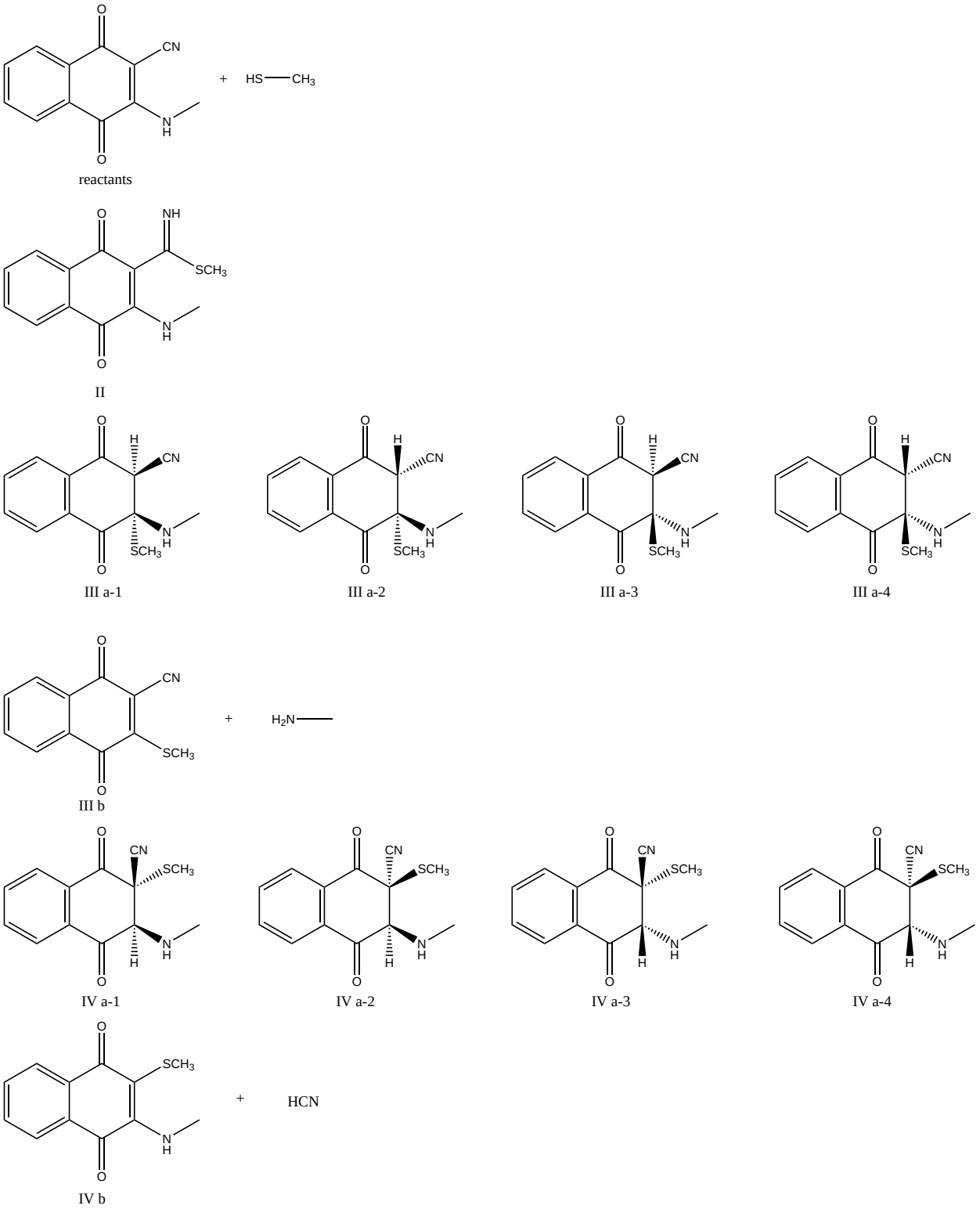


Figure 6.3: Reactans and products of the reaction of methanethiol with the inhibitor 8. The corresponding products for inhibitor 4 are obtained by replacing CN with Cl.

As one can see in Table 6.1 the calculated RMSD values of reactants and possible products of both investigated inhibitors are shown in relation to MP2 optimized structures. In case of HF the largest MAE of 0.242 Å is observed, mainly because of III b or IV b for inhibitor 4 and IV b for inhibitor 8. In contrast to these findings the calculated MAE of RMSD values for B3LYP optimized structures is lower by a value of 0.129 Å. The largest deviation was observed for III a-4 (inhibitor 4) with a value of 0.341 Å or IV b (inhibitor 8) with a value of 0.334 Å. In terms of structure optimization B3LYP appears to be an adequate methodology.

Table 6.1: RMSD values in relation to MP2 optimized structures employing the def-TVZP basis set. B3LYP-D3/def-TZVP shows a MAE of 0.129 Å, HF/def-TZVP a MAE of 0.242 Å. Structure optimizations employed the COSMO approach (ϵ=78). All values are given in Å.

compound 4	**B3LYP-D3**	**HF**	**compound 8**	**B3LYP-D3**	**HF**
educt	0.116	0.114	**educt**	0.153	0.172
III a-1	0.096	0.169	**II**	0.161	0.207
III a-2	0.109	0.196	**III a-1**	0.162	0.143
III a-3	0.086	0.139	**III a-2**	0.122	0.12
III a-4	0.341	0.463	**III a-3**	0.112	0.138
III b	0.127	0.967	**III a-4**	0.127	0.326
IV a-1	0.115	0.166	**III b**	0.224	0.151
IV a-2	0.089	0.146	**IV a-1**	0.160	0.173
IV a-3	0.172	0.203	**IV a-2**	0.088	0.390
IV a-4	0.097	0.127	**IV a-3**	0.175	0.146
IV b	0.162	0.854	**IV a-4**	0.120	0.148
			IV b	0.334	0.834
CH_3SH	0.013	0.011			
H_2NCH_3	0.009	0.009			
HCl	0.007	0.002			
HCN	0.008	0.020			
MAE[1]	0.129	0.242			

[1] Mean average error of RMSD with respect to MP2 optimization averaged of all compounds.

Although the optimized geometries are comparable by their RMSD values for B3LYP and MP2, a significant difference can be seen in terms of the calculated reaction energy ΔE_{reac}. As shown in Table 6.2 product IV b for compound 4 was predicted as most stable product applying HF or B3LYP showing up with a reaction energy of -5.2 or -5.8 kcal/mol. Additionally, the reaction energies ΔE_{reac} of the remaining products resulting in positive values. This contrasts with the results from MP2 calculations. In this case III a-1 was the most energetically stable product with a value of -15.7 kcal/mol. Furthermore, only III b was

positive in reaction energy. These differences are also observable for compound 8. HF shows only positive reaction energies, whereas B3LYP reveals for product II a difference in energy by -0.7 kcal/mol between product and reactants which is in the error of the method. In all other cases positive values are obtained. Again, MP2 calculated energies predict product III a-1 as lowest in energy by a value of -10.6 kcal/mol. Only products III a-4, III b and IV a-1 are described higher in energy compared to the reactants. Thus, the question arises if those discrepancies in the reaction energy occur due to the difference in optimized structure or being caused by the methodology. Therefore, single point energy calculation employing B3LYP and BHLYP in combination with Grimme's dispersion correction or ab initio methods as MP2, CC2 and SCS-CC2 have been conducted. Additionally, the basis set has been increased to Truhlar's basis set aug-cc-pVDZ. Environmental effects have been described by COSMO (ϵ=78) again.

If one takes a closer look at Table 6.3 and Table 6.4 a similar trend to previous calculations can be seen. Single point energy calculations which employ DFT methodologies predict product IV b for compound 4 as lowest in reaction energy. In this case an abstraction of the chlorine by a S_N2 reaction is described as favorable. In contrast to that ab initio methods show that product III a-1 is lower by $\approx$ 2-3 kcal/mol in energy compared to IV b. Interestingly all these results are independent from the employed methodology of structure optimization. Differences in single point energy which occur due to the usage of a different optimization method are below 2.0 kcal/mol (Table 6.7). Still, the trend in prediction is in line with the robusteness for the used methods.

This behavior can also be observed for single point energies of compound 8. In terms of B3LYP or BHLYP only the nucleophilic attack on the nitrile group on compound 8 leads to negative values of ΔE_{reac}. Once more, calculations on ab initio level predict different results. Reaction energies of product III a-1 are about 2-3 kcal/mol lower independent of the applied methodology for geometry optimization. And again, if one compares single point energies of the same method by varying B3LYP or MP2 for geometry optimization differences are at most of 2.0 kcal/mol (Table 6.7).

Table 6.2: Summary of the reaction energies (ΔE_{reac}) of possible reaction products of compound 4 or 8 with methanethiol (Figure 6.3) obtained from geometry optimization. All calculations were performed with a def-TZVP basis sets and employing the COSMO approach (ϵ=78) to mimic environmental effects. All energy values are depicted in kcal/mol.

compound 4	**HF**	**B3LYP-D3**	**MP2**
III a-1	5.7	1.8	-15.7[1]
III a-2	7.8	4.9	-12.0
III a-3	6.9	4.2	-13.2
III a-4	9.0	6.6	-10.0
III b	9.1	11.9	5.0
IV a-1	6.9	3.6	-14.1
IV a-2	7.3	4.2	-12.9
IV a-3	8.7	6.2	-10.2
IV a-4	11.2	8.9	-7.8
IV b	-5.2[1]	-5.8[1]	-10.9

compound 8	**HF**	**B3LYP-D3**	**MP2**
II	13.4	-0.7[1]	-2.3
III a-1	13.1[1]	7.2	-10.6[1]
III a-2	14.1	8.0	-8.1
III a-3	18.4	14.1	-4.9
III a-4	22.0	17.1	0.2
III b	18.0	16.5	9.9
IV a-1	23.0	18.4	0.9
IV a-2	13.6	11.8	-5.7
IV a-3	17.8	15.1	-5.0
IV a-4	15.1	13.0	-6.8
IV b	17.2	14.9	8.3

[1] Lowest reaction energy ΔE_{reac} of calculated reaction products of compound 4 or 8 with methanethiol.

Table 6.3: Reaction energies (ΔE_{reac}) of possible reaction products for the reaction of compound 4 with methanethiol (Figure 6.3) obtained from singlepoint energies on ab initio or DFT level employing the COSMO approach (ϵ=78) with an aug-cc-pVDZ basis set. Energies calculated from B3LYP-D3/def-TZVP + COSMO (ϵ=78) optimized structures. All energy values are presented in kcal/mol.

compound 4	**B3LYP-D3**	**BHLYP-D3**	**MP2**	**CC2**	**SCS-CC2**
III a-1	0.7	-3.5	-15.5[1]	-15.1[1]	-12.1[1]
III a-2	3.6	-1.0	-12.6	-12.3	-9.2
III a-3	3.1	-1.1	-12.9	-12.7	-9.9
III a-4	5.6	1.3	-10.0	-9.7	-7.0
III b	11.9	12.3	6.3	7.1	7.1
IV a-1	2.3	-1.3	-13.5	-13.6	-10.5
IV a-2	3.1	-0.6	-12.5	-12.6	-9.8
IV a-3	5.0	0.7	-10.3	-10.0	-7.3
IV a-4	7.6	3.3	-8.2	-7.8	-5.3
IV b	-5.9[1]	-6.2[1]	-11.4	-11.7	-9.8

[1] Lowest reaction energy ΔE_{reac} of possible reaction products of compound 4 with methanethiol.

Table 6.4: Reaction energies (ΔE_{reac}) of possible reaction products for the reaction of compound 4 with methanethiol (Figure 6.3) obtained from singlepoint energies on ab initio or DFT level employing the COSMO approach (ϵ=78) with an aug-cc-pVDZ basis set. Energies calculated from MP2/def-TZVP + COSMO (ϵ=78) optimized structures. All energy values are given in kcal/mol.

compound 4	**B3LYP-D3**	**BHLYP-D3**	**MP2**	**CC2**	**SCS-CC2**
III a-1	0.6	-4.1	-16.8[1]	-16.5[1]	-12.9[1]
III a-2	3.4	-1.6	-14.0	-13.8	-10.1
III a-3	2.8	-2.1	-14.4	-14.2	-11.0
III a-4	5.1	0.3	-11.8	-11.7	-8.5
III b	11.4	11.9	5.8	6.5	6.6
IV a-1	2.2	-2.0	-15.1	-15.2	-11.6
IV a-2	2.8	-1.6	-14.0	-14.1	-10.8
IV a-3	5.1	0.4	-11.5	-11.3	-8.0
IV a-4	7.3	2.5	-9.4	-9.1	-6.1
IV b	-6.2[1]	-6.2[1]	-11.8	-12.3	-10.0

[1] Lowest reaction energy ΔE_{reac} of possible reaction products of compound 4 with methanethiol.

Table 6.5: Reaction energies (ΔE_{reac}) of possible reaction products for the reaction of compound 8 with methanethiol (Figure 6.3) obtained from singlepoint energies on ab initio or DFT level employing the COSMO approach (ϵ=78) with an aug-cc-pVDZ basis set. Energies calculated from B3LYP-D3/def-TZVP + COSMO (ϵ=78) optimized structures. All energy values are given in kcal/mol.

compound 8	B3LYP-D3	BHLYP-D3	MP2	CC2	SCS-CC2
II	-4.4[1]	-4.2[1]	-7.0	-8.6	-3.7
III a-1	5.9	3.1	-9.9[1]	-10.1[1]	-6.6[1]
III a-2	6.5	3.7	-8.7	-9.1	-5.4
III a-3	12.5	9.1	-4.4	-4.4	-1.5
III a-4	15.4	12.5	0.2	0.0	2.8
III b	16.3	17.6	11.0	11.2	11.5
IV a-1	17.1	13.5	1.1	1.2	4.0
IV a-2	10.6	6.7	-5.5	-5.3	-2.7
IV a-3	13.7	10.1	-3.9	-3.8	-1.1
IV a-4	11.6	7.5	-6.0	-5.7	-3.3
IV b	15.5	15.9	9.8	8.9	10.0

[1] Lowest reaction energy ΔE_{reac} of possible reaction products of compound 8 with methanethiol.

Table 6.6: Reaction energies (ΔE_{reac}) of possible reaction products for the reaction of compound 8 with methanethiol (Figure 6.3) obtained from singlepoint energies on ab initio or DFT level employing the COSMO approach (ϵ=78) with an aug-cc-pVDZ basis set. Energies calculated from MP2/def-TZVP + COSMO (ϵ=78) optimized structures. All energy values are shown in kcal/mol.

compound 8	B3LYP-D3	BHLYP-D3	MP2	CC2	SCS-CC2
II	-4.1[1]	-4.5[1]	-6.4	-7.8	-3.0
III a-1	6.7	3.3	-11.4[1]	-11.8[1]	-7.7[1]
III a-2	7.0	3.5	-10.1	-10.5	-6.4
III a-3	12.8	9.0	-6.1	-6.2	-2.8
III a-4	15.4	11.9	-1.4	-1.6	1.6
III b	-16.2	17.7	-6.4	-7.8	-3.0
IV a-1	17.2	13.4	-0.7	-0.7	2.6
IV a-2	10.6	6.4	-7.0	-6.9	-3.8
IV a-3	14.0	10.0	-5.5	-5.4	-2.2
IV a-4	11.7	7.4	-7.5	-7.3	-4.3
IV b	15.2	15.9	8.8	7.8	9.4

[1] Lowest reaction energy ΔE_{reac} of possible reaction products of compound 8 with methanethiol.

Table 6.7: Differences (ΔE_{diff}) in DFT or ab initio calculated reaction energies (ΔE_{reac}) between B3LYP-D3 and MP2 optimized structures of possible reaction products for compound 4 or 8 with methanethiol. Singlepoint calculations employed the COSMO approach (ϵ=78) and an aug-cc-pVDZ basis set. All energy values are given in kcal/mol.

compound 4	$\Delta E_{diff}^{B3LYP-D3}$	$\Delta E_{diff}^{BHLYP-D3}$	ΔE_{diff}^{MP2}	ΔE_{diff}^{CC2}	$\Delta E_{diff}^{SCS-CC2}$
III a-1	0.1	0.6	1.3	1.5	0.9
III a-2	0.2	0.6	1.4	1.6	1.0
III a-3	0.3	1.0	1.5	1.6	1.1
III a-4	0.5	1.0	1.9	2.0	1.5
III b	0.5	0.4	0.5	0.6	0.5
IV a-1	0.1	0.7	1.6	1.6	1.1
IV a-2	0.3	1.0	1.4	1.4	1.0
IV a-3	0.1	0.4	1.2	1.3	0.7
IV a-4	0.3	0.8	1.2	1.3	0.8
IV b	0.3	0.1	0.5	0.6	0.3

compound 8	$\Delta E_{diff}^{B3LYP-D3}$	$\Delta E_{diff}^{BHLYP-D3}$	ΔE_{diff}^{MP2}	ΔE_{diff}^{CC2}	$\Delta E_{diff}^{SCS-CC2}$
II	0.3	0.3	0.6	0.8	0.8
III a-1	0.8	0.2	1.6	1.6	1.1
III a-2	0.5	0.2	1.4	1.4	1.0
III a-3	0.3	0.2	1.7	1.8	1.2
III a-4	0.0	0.6	1.5	1.6	1.2
III b	0.1	0.1	0.5	0.5	0.4
IV a-1	0.1	0.1	1.8	1.9	1.4
IV a-2	0.1	0.3	1.5	1.6	1.1
IV a-3	0.3	0.0	1.5	1.7	1.1
IV a-4	0.1	0.2	1.4	1.6	1.0
IV b	0.3	0.0	1.0	1.0	0.5
MAE[1]	0.3	0.4	1.3	1.4	1.0

[1]Mean average error with respect to MP2/def-TZVP optimized structures averaged of both compounds.

Moreover Table 6.7 shows, energies which are calculated with B3LYP or BHLYP have a lower MAE of 0.3 or 0.4 kcal/mol. In terms of ab initio methods SCS-CC2 shows the lowest MAE with a value of 1.0 kcal/mol, followed by MP2 and CC2 with 1.3 and 1.4 kcal/mol. In this case the difference in prediction of the most stable product does not result from the methodology for optimization but rather from the methodology employed for reaction energy evaluation. Thus, the next step was a benchmark of ab initio methods in comparison to CCSD(T) the gold standard for reaction energies.

6.3 Comparison of reaction energies for ab initio methods compared to CCSD(T) in *vacuo*

In the previous section differences in the prediction of possible reaction products were revealed which depend on the usage of whether ab initio, post-HF or DFT methods were used. Usually one can expect that MP2 or CC2 results are more reliable to B3LYP or BHLYP calculations. Still, it is of necessity to prove this statement for this specific case. Therefore, a benchmark was conducted using CCSD(T) as reference. As no COSMO implementation was available for the Turbomole program package, single point calculations were performed in *vacuo*. Additionally, the def-TZVP basis set was employed to decrease computational costs with respect to CCSD(T) calculations. Only for product III a-1 calculations on CCSD(T) level with the augmented basis set aug-cc-pVDZ (double-zeta, polarized and augmented) were conducted [149]. For the remaining ab initio methods solely the aug-cc-pVDZ basis set was used as well as MP2 optimized geometries.

Firstly, one can see that the results of the CCSD(T) calculations lead to the same conclusions as before: product III a-1 shows the lowest reaction energy. For compound 4 this product is 8 kcal/mol more stable compared to IV b which was described by DFT methods as energetically lowest product, for compound 8 it is $\approx$ 6 kcal/mol lower in energy compared to product II (Table 6.8). This confirms the assumption of the reliability of MP2 and CC2 calculated energies as the results reflect the same trend as determined by former calculations.

However, how do the remaining ab initio methods perform? In terms of MP2 and CC2 one obtains a MAE of 2.6 and 2.8 kcal/mol compared to CCSD(T)/def-TZVP calculations. Both methodologies describe the reaction energy quite similar, as they differ by at most 0.6 kcal/mol. Only once a difference of 1.3 kcal/mol is obtained for product II of compound 8. By employing the spin-component scaling approaches the MAE decreases. SCS-MP2 shows a MAE of 0.3 kcal/mol, SCS-CC2 a slightly higher one by 0.4 kcal/mol. If one compares

both methods, one can see that the calculated energies differ by a largest value of about 0.8 kcal/mol. Although the differences in MAE are below the accuracy of the calculations (methods/basis set), SCS-MP2 was chosen as reference for further calculations due to its lower MAE. Still it is to be expected that SCS-CC2 could also be utilized.

Table 6.8: Reaction energies (ΔE_{reac}) of possible reaction products of compound 4 or 8 with methanethiol from different ab initio calculations compared to CCSD(T)/def-TZVP calculations (for reaction product III a-1 the aug-cc-pVDZ basis set was additionally considered, marked as [3]). All reaction energies are calculated in vacuum employing the aug-cc-pVDZ basis set. All energy values are shown in kcal/mol.

method	CCSD(T)	MP2	CC2	SCS-MP2	SCS-CC2
compound 4					
III a-1	-16.2[2] / -17.1[2,3]	-19.4[2]	-19.3[2]	-16.1[2]	-15.7[2]
III a-2	-10.1	-13.9	-13.9	-10.6	-10.2
III a-3	-12.9	-15.6	-15.5	-12.7	-12.4
III a-4	-10.3	-13.3	-13.4	-10.5	-10.1
III b	6.7	6.5	7.0	6.7	7.3
IV a-1	-12.9	-16.0	-16.3	-12.9	-12.7
IV a-2	-12.6	-15.2	-15.4	-12.3	-12.2
IV a-3	-11.0	-14.1	-14.1	-11.0	-10.7
IV a-4	-7.9	-10.8	-10.7	-8.0	-7.6
IV b	-8.2	-10.8	-11.3	-8.8	-9.0
compound 8					
II	-6.8	-10.9	-12.2	-7.0	-7.4
III a-1	-12.7[2] / -13.5[2,3]	-15.8[2]	-16.3[2]	-12.4[2]	-12.3[2]
III a-2	-10.0	-13.7	-14.3	-10.3	-10.2
III a-3	-6.8	-9.5	-9.8	-6.6	-6.4
III a-4	-2.2	-4.8	-5.1	-2.2	-2.0
III b	9.4	8.9	9.0	9.3	9.6
IV a-1	-2.5	-5.5	-5.6	-2.7	-2.3
IV a-2	-8.2	-11.0	-11.0	-8.3	-7.9
IV a-3	-7.6	-10.0	-10.0	-7.3	-6.9
IV a-4	-9.6	-12.1	-12.0	-9.6	-9.1
IV b	7.5	8.0	7.0	9.1	8.6
MAE[1]		2.6	2.8	0.3	0.4

[1]Mean average error with respect to CCSD(T)/def-TZVP averaged of all compounds. [2] Lowest calculated reaction energy ΔE_{reac} of possible reaction products of compound 4 or 8 with methanethiol. [3] CCSD(T) calculation employing the aug-cc-pVDZ basis set.

To evaluate the errors caused by the usage of different basis sets, additional calculations of the reaction energies have been conducted by employing the def-TZVP basis set. As one can see in Table 6.9 a decrease of 1-2 kcal/mol occurs for most of the products calculated with SCS-MP2 if aug-cc-pVDZ is used. A similar trend can be seen for CCSD(T). Products III a-1 for compound 4 and 8 are lower in energy if a larger basis set is used.

Table 6.9: Calculated reaction energies (ΔE_{reac}) of possible reaction products of compound 4 or 8 with methanethiol in vacuum employing the def-TZVP or aug-cc-pVDZ basis set. For CCSD(T) solely product III a-1 has been calculated with an aug-cc-pVDZ basis set. All energy values are given in kcal/mol.

method **basis set**	CCSD(T) **aug-cc-pVDZ**	CCSD(T) **def-TZVP**	SCS-MP2 **aug-cc-pVDZ**	SCS-MP2 **def-TZVP**	$\Delta_{basisset}^{SCS-MP2}$
compound 4					
III a-1	-17.1[2]	-16.2[2]	-16.1[2]	-14.6[2]	1.4
III a-2	NA	-10.1	-10.6	-8.4	2.2
III a-3	NA	-12.9	-12.7	-11.3	1.4
III a-4	NA	-10.3	-10.5	-8.6	1.9
III b	NA	6.7	6.7	6.4	0.3
IV a-1	NA	-12.9	-12.9	-11.6	1.3
IV a-2	NA	-12.6	-12.3	-11.0	1.3
IV a-3	NA	-11.0	-11.0	-9.5	1.5
IV a-4	NA	-7.9	-8.0	-6.2	1.8
IV b	NA	-8.2	-8.8	**-7.6**	1.2
compound 8					
II	NA	-6.8	-7.0	-2.7	4.3
III a-1	-13.5[2]	-12.7[2]	-12.4[2]	-11.22	1.3
III a-2	NA	-10.0	-10.3	-8.1	2.2
III a-3	NA	-6.8	-6.6	-5.2	1.4
III a-4	NA	-2.2	-2.2	-0.3	1.8
III b	NA	9.4	9.3	9.1	0.1
IV a-1	NA	-2.5	-2.7	-1.0	1.8
IV a-2	NA	-8.2	-8.3	-6.7	1.6
IV a-3	NA	-7.6	-7.3	-6.5	0.9
IV a-4	NA	-9.6	-9.6	-8.5	1.1
IV b	NA	7.5	9.1	8.6	0.6
					MAE[1] = 1.5

[1]Mean average error with respect to SCS-MP2/aug-cc-pVDZ averaged of all compounds in relation to SCS-MP2/def-TZVP. [2] Lowest calculated reaction energy ΔE_{reac} of possible reaction products of compound 4 or 8 with methanethiol. $\Delta_{basisset}^{SCSMP2}$ represents E(SCS-MP2/def-TZVP)− E(SCS-MP2/aug-cc-pVDZ).

Additionally, there is a need to prove the reliability of the results of CCSD(T) which represents the Gold standard for ground state closed shell systems. In case of open shell or biradical systems, Multi-Reference approaches are needed to describe the reaction correctly. In these systems one can exclude Multi-Reference character in the CCSD(T) computations.

The D_1 diagnostics show values below 0.08 and thus reaction energies obtained *via* the couple cluster approach can be considered as valid (Table 6.10).

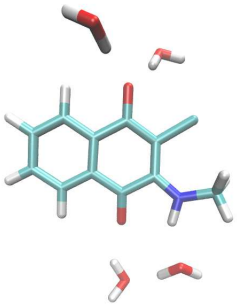
Table 6.10: Summary of CCSD D_1 diagnostic of reactants and possible reaction products for the reaction of compound 4 or 8 with methanethiol. All calculations were performed with the CCSD(T)/def-TVP or CCSD(T)/aug-cc-pDVZ approach.

basis set	**compound**	**4**	**8**	**compound**	
def-TZVP	**educt**	0.0767	0.0779	**CH_3SH**	0.0190
	II		0.0779	**HCl**	0.0101
	III a-1	0.0588	0.0612	**HCN**	0.0306
	III a-2	0.0577	0.0625	**H_3CNH_2**	0.0156
	III a-3	0.0597	0.0582		
	III a-4	0.0571	0.0611		
	III b	0.0675	0.0706		
	IV a-1	0.0640	0.0587		
	IV a-2	0.0603	0.0601		
	IV a-3	0.0577	0.0581		
	IV a-4	0.0578	0.0591		
	IV b	0.0783	0.0779		
basis set	**compound**	**4**	**8**	**compound**	
aug-cc-pVDZ	**educt**	0.0754	0.0770	C**H_3SH**	0.0208
	III a-1	0.0578	0.0603		

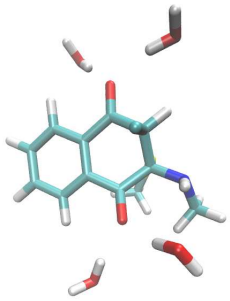
6.4 Influence of solvation effects on reactants, products and reaction energy

As no implementation of COSMO was available for CCSD(T) calculations in Turbomole, previous single point energy calculations were conducted in vacuum. To investigate the effects of implicit solvation, COSMO has been applied on MP2, CC2 or their spin-component scaled variants with an ϵ-value of 78.

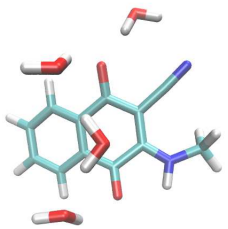
If one compares Table 6.8 and Table 6.11 one can observe that the use of COSMO leads to a decrease in reaction energy for nearly all possible products. The effect of the solvation is larger for compound 8 as differences are in the range of $\approx$ 3-5 kcal/mol for products II, III a-1 to III a-4 and IV a-1 to IV a4. For products of the S_N2 reaction III b and IV b the influence is smaller. The values are of about 1.7 and 0.8 kcal/mol. This trend can also be seen for



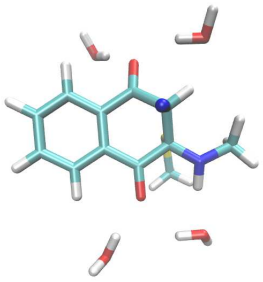
compound 4 educt



compound 4 product III a-1



compound 8 educt



compound 8 product III a-1

Figure 6.4: Orientation of the water molecules around the carbonyl moieties of educt and product III a-1 of compound 4 and 8 after geometry optimization with B3LYP-D3/def-TZVP employing COSMO (ϵ=78).

compound 4. In case of III b and IV b implicit solvation increases the reaction energy by 0.7 or 1.0 kcal/mol. For most of the remaining cases reaction energies are slightly higher compared to vacuum calculated values by 1.5 to 2.8 kcal/mol. Interestingly, one can observe, that the difference due to implicit solvation varies only by 0.1 kcal/mol for the above applied ab initio methods. Hence the decrease or increase in reaction energy by solvent effects is independent of the employed ab initio methodology.

To gain further insight on the influence of the solvent, additional calculations employing an ϵ-value of 4 have been conducted to simulate a non-polar environment (e.g protein). Additionally, four explicit water molecules have been considered for the reaction of compound 4 and 8 to product III a-1. SCS-MP2 was used for calculation as it leads to the lowest MAE value in comparison to the CCSD(T)/def-TZVP reaction energies.

As one can see in Table 6.12 the reaction energy decreases by about 1-3 kcal/mol with increasing permittivity. The only exceptions are the products III b and IV b of compound 4. The S_N2-reaction products reveal the opposite trend as their reaction energy increases by 0.9 to 1.6 kcal/mol as the stabilization *via* COSMO is larger on the product site. Additionally,

Table 6.11: Calculated reaction energies (ΔE_{reac}) of possible reaction products of compound 4 or 8 with methanethiol with COSMO (ϵ=78) employing the def-TZVP or aug-cc-pVDZ basis set. CCSD(T) has been calculated in vacuum employing the def-TZVP basis set (except for product III a-1 which has been calculated with an aug-cc-pVDZ basis set , marked as [3]). All energy values in kcal/mol.

method	**CCSD(T)**	**MP2**	**CC2**	**SCS-MP2**	**SCS-CC2**
compound 4					
III a-1	-16.2[2] / -17.1[2,3]	-16.8[2]	-16.5[2]	-13.5[2]	-12.9[2]
III a-2	-10.1	-14.0	-13.8	-10.7	-10.1
III a-3	-12.9	-14.4	-14.2	-11.5	-11.0
III a-4	-10.3	-11.8	-11.7	-8.9	-8.4
III b	6.7	5.8	6.5	5.9	6.6
IV a-1	-12.9	-15.0	-15.2	-11.9	-11.6
IV a-2	-12.6	-14.0	-14.0	-11.1	-10.8
IV a-3	-11.0	-11.5	-11.3	-8.5	-8.0
IV a-4	-7.9	-9.3	-9.1	-6.6	-6.1
IV b	-8.2	-11.8	-12.3	-9.8	-10.0
compound 8					
II	-6.8	-6.4	-7.8	-2.5	-2.9
III a-1	-12.7[2] / -13.5[2,3]	-11.4[2]	-11.8[2]	-7.9[2]	-7.7[2]
III a-2	-10.0	-10.1	-10.5	-6.6	-6.4
III a-3	-6.8	-6.1	-6.2	-3.1	-2.7
III a-4	-2.2	-1.4	-1.6	1.3	1.6
III b	9.4	10.6	10.7	10.9	11.2
IV a-1	-2.5	-0.7	-0.7	2.1	2.6
IV a-2	-8.2	-6.9	-6.9	-4.3	-3.8
IV a-3	-7.6	-5.5	-5.4	-2.7	-2.2
IV a-4	-9.6	-7.5	-7.3	-4.8	-4.3
IV b	7.5	8.8	7.8	9.9	9.4
MAE[1]		1.5	1.4	2.7	2.9

[1]Mean average error with respect to CCSD(T)/def-TZVP averaged of all compounds. [2] Lowest calculated reaction energy ΔE_{reac} of possible reaction products of compound 4 or 8 with methanethiol. [3] CCSD(T) calculation employing the aug-cc-pVDZ basis set.

a decrease of reaction energy is observable if four explicit water molecules are considered which are in the vicinity of each carbonyl moiety (Figure 6.4). To conclude this section, it is observable, that an increased solvent polarity leads to a decrease of the reaction energy. Differences of up to 6 kcal/mol were observable depending on environmental effects. This shows that the enzyme environment will influence the results. But as stated before experimental

information regarding orientation of the compound in enzyme is missing thus no calculations in that direction have been performed.

Table 6.12: Summary of the reaction energies (ΔE_{reac}) of possible reaction products for the reaction of compound 4 and 8 with methanethiol (see Figure 6.3) as a function of the environment. All calculations were performed with the SCS-MP2/aug-cc-pVDZ approach. All energy values in kcal/mol. The orientation of the water molecules is given in Figure 6.4.

compound 4	**ϵ=1 (vacuum)**	**ϵ=4**	**ϵ=78**	**e=78 & 4 H_2O**
III a-1	-16.1	-14.5	-13.5	-10.6
III a-2	-10.6	-10.7	-10.7	
III a-3	-12.7	-11.9	-11.5	
III a-4	-10.5	-9.6	-8.9	
III b	6.7	6.1	5.9	
IV a-1	-12.9	-12.4	-11.9	
IV a-2	-12.3	-11.5	-11.1	
IV a-3	-11.0	-9.5	-8.5	
IV a-4	-8.0	-7.1	-6.6	
IV b	-8.8	-9.5	-9.8	

compound 8	**ϵ=1 (vacuum)**	**ϵ=4**	**ϵ=78**	**e=78 & 4 H_2O**
II	-7.0	-4.3	-2.5	
III a-1	-12.4	-9.8	-7.9	-4.2
III a-2	-10.3	-8.2	-6.6	
III a-3	-6.6	-4.6	-3.1	
III a-4	-2.2	-0.2	1.3	
III b	9.3	10.1	10.9	
IV a-1	-2.7	0.0	2.1	
IV a-2	-8.3	-5.9	-4.3	
IV a-3	-7.3	-4.6	-2.7	
IV a-4	-9.6	-6.8	-4.8	
IV b	9.1	9.6	9.9	

6.5 Benchmarking – comparison of DFT/COSMO with SCS-MP2/COSMO

The applied DFT methods described product IV b for compound 4 or product II of compound 8 as lowest in terms of the reaction energy instead of product III a-1. The question arises if this discrepancy usually occurs for DFT methodologies or just in case of B3LYP and

BHLYP. Therefore, calculations have been extended by using further hybrid functionals as M06 and M06-2X or long-range-corrected functionals as CAM-B3LYP or ωB97-xD [150], [151]. As those methods were not available in the Turbomole 6.6 program package, the following calculations have been conducted by using the Gaussian09 program package. As mentioned before implicit solvation was included using COSMO with an epsilon value of 4 or 78. Furthermore SCS-MP2/aug-cc-pVDZ//COSMO results were used as reference as they showed the lowest MAE compared to the CCSD(T)/def-TZVP calculation.

As indicated in the previous calculations the applied DFT methods are not sufficient to describe the reaction energy. For compound 4 and a dielectric constant value of 78 one can see a positive thus endothermic description of the reaction in case of B3LYP and CAM-B3LYP except for product IV b with values of -6.2 or -4.1 kacl/mol (Table 6.13). Only the Minnesota functional MO6-2X predicts a reaction energy of -7.5 kcal/mol for product III a-1 and describes it as most stable as the SCS-MP2 calculations. Still, it has to be mentioned that IV b is only lower by 0.5 kcal/mol thus the separation between these two values is very low. For the other DFT methods IV b is still described falsely as the most stable reaction product is of about 1-2 kcal/mol lower in energy than III a-1 (except for B3LYP which is lower by $\approx$ 6 kcal/mol). If one applies a less hydrophilic environment a slight decrease in terms of energetics is observable for nearly all products except for most of the results of IIIb or IV b (Table 6.14). In contrast to this, a slight increase in reaction energy is observable leading to a stabilization of III a-1 for BHLYP which becomes less stable by 0.1 kcal/mol compared to IV b. Additionally, the separation in energy between these products is larger for M06-2X by applying a smaller value for the permittivity.

Contrary to these results, product II of compound 8 is always described as lowest in terms of the reaction energy. For a hydrophilic implicit solvation only M06-2X describes III a-1 and III a-2 exothermic but still not favorable as products. The remaining calculated reaction energies are positive. In contrast to compound 4 the difference in reaction energy compared to II is larger. The deviation ranges from 6 to 10 kcal/mol if one compares II and III a-1. As mentioned before the largest energetic difference between IV b and III a-1 of compound 4 was 6 kcal/mol in case of B3LYP. This leads to the conclusion that applied DFT methods are not sufficient and can not properly describe the reaction of compound 8. By including a nonpolar environment, the reaction energies decrease in a range of 0.5 to 2 kcal/mol mirroring the same trend as obtained from the post HF calculations. Furthermore, no destabilization of the SN2-reaction products III b or IV b occurred as for compound 4 and the predicted trend remains the same regardless of the environment.

Table 6.13: Calculated reaction energies (ΔE_{reac}) of possible reaction products of compound 4 or 8 with methanethiol in COSMO (ϵ=78) employing the aug-cc-pVDZ basis set. All energy values in kcal/mol.

method	**SCS-MP2**	**B3LYP-D3**	**BHLYP-D3**	**CAM-B3LYP**	**M06**	**M06-2X**	**ωB97-XD**
compound 4							
III a-1	-13.5[2]	0.6	-4.1	2.9	-3.9	-7.5[2]	-4.7
III a-2	-10.7	3.4	-1.6	4.7	-2.0	-5.7	-2.8
III a-3	-11.5	2.8	-2.1	4.2	-2.0	-5.6	-3.5
III a-4	-8.9	5.1	0.3	5.6	-0.2	-4.2	-1.5
III b	5.9	11.4	11.9	12.6	11.0	10.4	11.9
IV a-1	-11.9	2.2	-2.0	3.7	-2.3	-5.5	-3.4
IV a-2	-11.1	2.8	-1.6	3.7	-2.5	-5.3	-3.2
IV a-3	-8.5	5.1	0.4	5.3	-0.7	-3.8	-1.3
IV a-4	-6.6	7.3	2.5	7.1	1.6	-1.9	1.1
IV b	-9.8	-6.2[2]	-6.2[2]	-4.1[2]	-5.3[2]	-7.0	-6.3[2]
compound 8							
II	-2.5	-4.0[2]	-4.5[2]	-3.1[2]	-7.8[2]	-7.6[2]	-7.7[2]
III a-1	-7.9[2]	6.7	3.3	7.8	0.8	-1.8	0.7
III a-2	-6.6	7.0	3.5	8.0	0.7	-1.9	1.1
III a-3	-3.1	12.8	9.0	13.5	6.7	3.4	5.8
III a-4	1.3	15.4	11.9	16.2	9.3	6.6	9.1
III b	10.9	16.2	17.7	17.8	15.5	15.8	17.2
IV a-1	2.1	17.2	13.4	17.6	11.5	7.2	9.9
IV a-2	-4.3	10.6	6.4	10.2	4.3	1.6	4.0
IV a-3	-2.7	14.0	10.0	13.5	7.9	4.5	6.7
IV a-4	-4.8	11.7	7.3	11.9	5.8	2.5	5.1
IV b	9.9	15.2	15.9	17.0	15.1	12.3	15.0
MAE[1]		12.1	9.0	13.0	8.0	5.3	7.5

[1]Mean average error with respect to SCS-MP2/aug-cc-pVDZ averaged of all compounds in relation. [2]Lowest calculated reaction energy ΔE_{reac} of possible reaction products of compound 4 or 8 with methanethiol.

Table 6.14: Calculated reaction energies (ΔE_{reac}) of possible reaction products of compound 4 or 8 with methanethiol in COSMO (ϵ=4) employing the aug-cc-pVDZ basis set. All energy values in kcal/mol.

method	**SCS-MP2**	**B3LYP-D3**	**BHLYP-D3**	**CAM-B3LYP**	**M06**	**M06-2X**	**ωB97-XD**
compound 4							
III a-1	-14.5[2]	-0.6	-5.7	2.2	-4.6	-8.2[2]	-5.4
III a-2	-10.7	3.3	-1.7	4.8	-1.9	-5.6	-2.8
III a-3	-11.9	2.2	-2.7	3.9	-2.2	-5.8	-3.7
III a-4	-9.6	4.3	-0.5	5.2	-0.5	-4.6	-1.9
III b	6.1	11.5	12.1	12.6	11.0	10.4	11.8
IV a-1	-12.4	1.7	-2.6	3.4	-2.5	-5.8	-3.7
IV a-2	-11.5	2.2	-2.1	3.5	-2.7	-5.5	-3.4
IV a-3	-9.5	4.0	-0.7	4.8	-1.1	-4.3	-1.8
IV a-4	-7.1	6.7	1.9	6.9	1.4	-2.1	0.8
IV b	-9.5	-5.8[2]	-5.8[2]	-4.0[2]	-5.2[2]	-6.9	-6.2[2]
compound 8							
II	-4.3	6.2[2]	-6.6[2]	-4.4[2]	-9.0[2]	-8.7[2]	-9.0[2]
III a-1	-9.8[2]	4.5	0.9	6.5	-0.5	-3.1	-0.7
III a-2	-8.2	5.2	1.6	6.9	-0.3	-2.8	0.1
III a-3	-4.6	10.9	7.1	12.4	5.7	2.4	4.7
III a-4	-0.2	13.8	10.2	15.1	8.3	5.6	8.1
III b	10.1	15.3	16.8	17.2	15.0	15.2	16.5
IV a-1	0.0	14.7	10.8	16.0	10.0	5.7	8.4
IV a-2	-5.9	8.6	4.3	9.1	3.2	0.5	2.9
IV a-3	-4.6	11.8	7.8	12.3	6.7	3.3	5.4
IV a-4	-6.8	9.4	4.9	10.5	4.5	1.1	3.7
IV b	9.6	14.7	15.4	16.6	14.7	12.0	14.7
MAE[1]		12.0	8.8	13.2	8.3	5.6	7.8

[1]Mean average error with respect to SCS-MP2/aug-cc-pVDZ averaged of all compounds in relation. [2]Lowest calculated reaction energy ΔE_{reac} of possible reaction products of compound 4 or 8 with methanethiol.

6.6 Conclusion and discussion

This chapter comprises the results of the theoretical study of the chloro- or nitrile-substituted 1,4-naphtoquinones 4 and 8 (see Figure 6.2) and their reaction products with methanethiol. The group of Schirmeister and Opatz from the University Mainz used these structures as warheads to target cysteine proteases of the papain family. In this study the performance of the methods in terms of structure optimization, energy computations and the influence of the environment has been evaluated. Firstly, possible structures have been modeled and optimized on different QM levels (products and their naming shown in Figure 6.3). In terms of geometry optimization, the performance of B3LYP is comparable to MP2 and shows a MAE of 0.129 Å, while geometries obtained by applying HF show a MAE of 0.242 Å (Table 6.1). It was observable that the energy evaluation is not very sensitive with respect to the used geometry optimization methods. The energy difference was at most 2 kcal/mol whether B3LYP or MP2 optimized structures were used while the resulting trends were the same. In these initial calculations it was observable that applied DFT methods revealed a different prediction in the stability of possible reaction products compared with wave function-based approaches. Therefore, it was necessary to conduct further investigations in this matter.
In benchmark calculations in *vacuo* employing CCSD(T)/def-TZVP as reference the results of SCS-MP2/aug-cc-pVDZ showed the lowest MAE with 0.3 kcal/mol followed by SCS-CC2 with 0.4 kcal/mol (Table 6.8). As no Multi-Reference character was observed for CCSD(T) calculations, obtained results can be accepted as valid. The question arises if DFT approaches in general lead to a different result compared to ab initio methods. Therefore, caculations have been extended to Minnesota and long-range corrected functionals. In these cases, SCS-MP2 was used as reference due to its low MAE. The difference in prediction was seen for nearly all functionals. In case of compound 4 the S_N2 product IV b was overestimated in stability while for compound 8 product II instead of III a-1 was described as lowest in reaction energy. The only exception is MO6-2X which described product III a-1 of compound 4 like SCS-MP2 as lowest in energy (Table 6.13).
Additionally, environmental effects have been investigated by employing the COSMO approach at different epsilon values and by inclusion of explicit water molecules (depicted in Figure 6.4). Ab initio calculations revealed a decrease in reaction energy by 1-3 kcal/mol with increasing dielectric constant values. Only reaction energies of the S_N2 products showed an opposite behavior as an increase of 1-2 kcal/mol. Additionally, explicit water molecules were considered and a further decrease of about 3 kcal/mol was observable. These findings indicate environmental effects will likely influence the results. But as no experimental

information regarding the orientation of the inhibitor within the active site was available corresponding computations couldn't be conducted.
But one question still remains. How do these results fit with the experimental findings? As one can see in Figure 6.3 except for the S_N2 products III b and IV b it is to be expected that the remaining products occur due to a reversible reaction as long as the exothermicity of the addition reaction remains small. The reaction to the products III b and IV b is followed by the abstraction of a leaving group, in these cases methylamine or hydrogen cyanide. As a dissociation of these groups is to be expected and a reattack appears unlikely one can consider that these products result from an irreversible reaction. In terms of DFT calculations they are only excluded in case of inhibitor 8 as one obtains positive reaction energies. For compound 4 the S_N2 product IV b was described as the most stable product. These findings indicate that DFT based approaches must be handled with care as they often overestimate stabilities due to the aromaticity.
In terms of the prediction of the formation it is not possible to distinguish between product III a or IV as the deviation of the reaction energy is relatively low. Still, in comparison to the experimental findings computations could exclude the formation of products which would lead to an irreversible inhibition. These findings are in line with experimental data.

Chapter 7

Analysis of protonation state and protein-inhibitor interactions of C1 in tryparedoxin monomer, dimer and mutants

7.1 Introduction

The Human African Trypanosomiasis (HAT, or known as African sleeping sickness) is mostly located in sub-saharan Africa and belongs to the so called neglected tropical diseases (NTD) [152]. The occurrence of HAT in this region is due to tsetse fly, which is a suitable habitant for its vector. Within the past years a change in the prevalence occurred caused by increased controls. Although the disease was nearly eradicated in the middle of the 1960s, cases of HAT increased in the 1990s. The number of actual cases nowadays (50.000 – 70.000) appears in global context in a negligible magnitude. Still, untreated HAT ends fatal for the infected patients and the negatively socialeconomic effects on affected areas are not to be underestimated [152].

The source of HAT is reasoned by kinetoplastids, a group of flagellated protozoans, single-celled eukaryotes, which are hidden in the kinetoplast, a DNA-rich region [153]. These protozon parasites are transmitted *via* different insect vectors and change within their life cycle between the insect vector and the mammalian host, while contemporaneous huge changes appear in their mitochondrial morphology or metabolism [154]. Besides of HAT, which is caused by the infection of the kinetoplastides *Trypanosoma brucei* (T.b.) *gambiense* or T.b.

rhodensie, also the Chagas disease (known as south-american trypanosomiasis) should be mentioned, induced by *Trypanosoma cruzei* [152].
The particularity of trypanosomias is the thiol redox metabolism which is different from the human body and thus represents a promising target. In T.b. the peroxide damage of the parasite is prevented due to a cascade of enzyme reactions, with NADPH as main electron source [155]. The delivery of reducing equivalents is carried out by the flavoenzyme trypanothione reductase (TR), trypanothione ($T(SH)_2$), tryparedoxin (Tpx) and finally by a peroxidase (Px) which lead to the detoxification of hydroperoxide substrates (Figure 7.1) [156], [157]. Additionally, this cascade plays an important role in the synthesis of DNA precursors and consequently in the replication of the parasite [158].
The medication of these diseases is limitied to a few number of drugs which go along with high toxicity and additionally increased drug resistances are observable [159]. This shows the relevance of the development of new drugs for the medication of HAT. In the course of a high-throughput screening approach with nearly 80.000 chemicals by Fueller et. al the molecule 2-(chloromethyl)-5-(4-fluorophenyl)thieno[2,3-d]pyrimidin-4(3H)-one (compound 1, C1) was identified as a Tpx targeting drug, which interacts covalently with Tpx active site residue Cys40 of the active site WCPPC motif (Figure 7.2) [159].

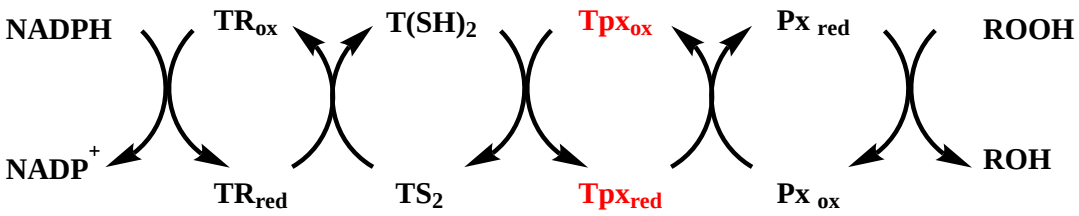


Figure 7.1: Detoxifiaction of hydroperoxides (ROOH) in African trypanosomias by a cascade of a peroxidase (Px), trypareodixin (Tpx), trypanothione ($T(SH)_2$) and the trypanothione reductase (TR).

The group of Hellmich from the University in Mainz was able to crystallize the first structure of Tpx in complex with C1 in covalent complex with the protein. The X-ray structure did not contain the leaving group hydrochloride which indicates that it diffuses after the nucleophilic reaction. After covalent modification of Tpx with C1, a dimerization of the protein was observable which was experimentally proven while Tpx wild type (oxidized or reduced form) stays in absence of the inhibitor in a monomer form. Chemical induced dimerization (CID) which is caused by so called dimerizers, is able to affect e.g. signal transmission and therefore stimulate target gene activation [160]. On this occasion dimerizers as rapamycin,

which is used as immunosuppressant, coumermycin or gibberlin are to mention [161], [162], [163]. Generally dimerizers are large molecules but in case of C1 it is so far one of the smallest known.
Thus, crystallization of Tpx with C1 offers the opportunity to conduct further investigations on computational side. Questions like inhibitor-protein interactions in the monomer or inhibitor-inhibitor or protein-protein interactions in Tpx dimer can be tackled *via* MD simulations and will be discussed in detail in the following sections.

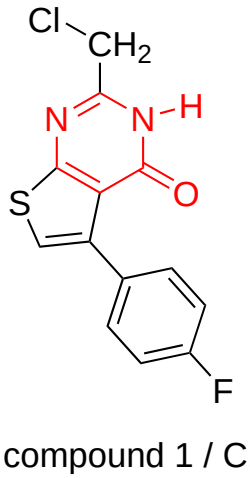


Figure 7.2: Chemical structure of compound 1 / C1 identified by Fueller et al. [159] . Pyrimidone unit of C1 is colored in red.

7.2 Motivation

To investigate C1 in complex with Tpx, MD simulations have been conducted to analyze the dynamics and the interactions with the surrounding protein environment. At the same time interactions between the cysteine residues Cys40 and Cys43 have been considered. Furthermore, nearby amino acids which might establish non-covalent interactions between inhibitor and protein have been included in the analysis as well.
Firstly, two different binding modes of C1 (pose I and pose II) were observed in case of the covalent protein-inhibitor crystal structure. The crystal structure consists of three different chains (A - C) where chain A and B form a dimer and C is present as a monomer. Pose I of the inhibitor C1 was obtained by considering solely chain A of the dimer, while pose II results from chain C. The initial question was to figure out which pose is more rigid and what kind of interactions are established in the course of simulation. Therefore, one can

conclude which pose is more likely in enzyme and additionally, these computations served as reference for the following investigations. The group of Hellmich mutated specific residues in the active site of Tpx and these variants have also been computed to analyze possible effects on the inhibitor-protein interaction and thus, orientation of the inhibitor in Tpx. Another point of interest was the computation of the non-covalent complex of C1 in wild type Tpx. The group of Opatz from the University of Mainz modified C1 by exchanging the chlorine with a hydrogen, thus no covalent reaction of the C1 derivative and Tpx was to be expected. Experimental data showed, that no decrease of the activity of Tpx was observable and in consequence, leading to the conclusion that non-covalent protein-inhibitor interactions should be weak. Therefore, the non-covalent complex of C1 with TPX was computed to gain further insight on this matter. Finally, dynamics of the dimer complex in presence and absence of C1 were studied to analyze drug-drug, protein-drug and protein-protein interaction which are of importance for the dimerization.
The protein environment was described *via* the AMBER14 force field whereas C1 was described by the GAFF force field. In terms of the simulations of the non-covalent complex, the description of the ESP charges was performed differently which will be discussed later on. The protocol of minimization and equilibration was conducted in the same manner as decribed in section 5.1. The duration of each production run consists of 30 ns, except for the MD simulations of the non-covalent complex which were performed for 10 ns.

7.3 C1 in covalent complex with Tpx in pose I or pose II

First of all, chain A of the X-ray structure obtained by Hellmichs group has been used as pose I, chain C of the same pdb structure for pose II. The difference between both binding modes is that C1 in pose II is exposed to the solvent whereas in pose I is parallel to the second inhibitor in chain B (Figure 7.3).
The MD simulation of both poses revealed several findings. First of all, C1 in pose I was more stable during the simulation. If one looks at the RMSD values one can see that in case of that pose that the RMSD value increases up to $\approx$ 2-3 Å. At certain times (for example during 6 and 7 ns) it decreases to 1 Å (Figure 7.5-A).

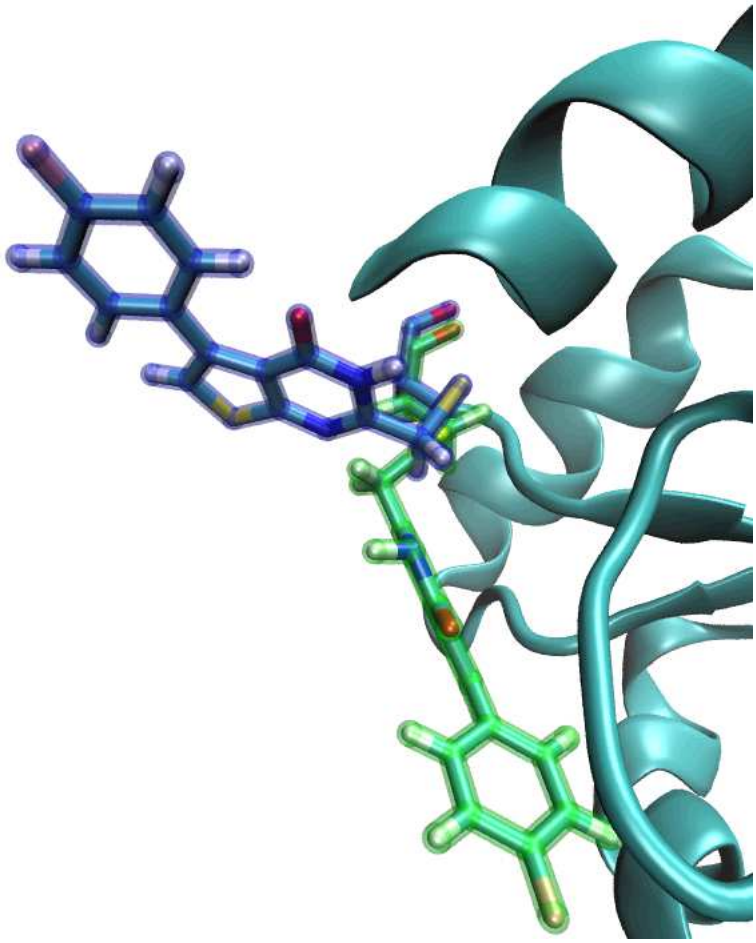


Figure 7.3: Covalent protein-inhibitor complex of inhibitor C1 and Tpx. Initial structure of pose I (green) and pose II (blue) of inhibitor C1 depicted in licorice representation.

This can be explained by the rotation of the pyrimidone ring of C1 (depicted in red in Figure 7.2) which will later one be discussed. In pose II the values increased up to 6 Å and show larger fluctuations. This is understandable as a movement of the inhibitor in direction of the protein is observable (Figure 7.4-C). Still C1 occasionally returns to a pose comparable the one at the beginning of the MD simulation accompanied by a decrease of RMSD values to 1.5 Å (Figure 7.5-B).

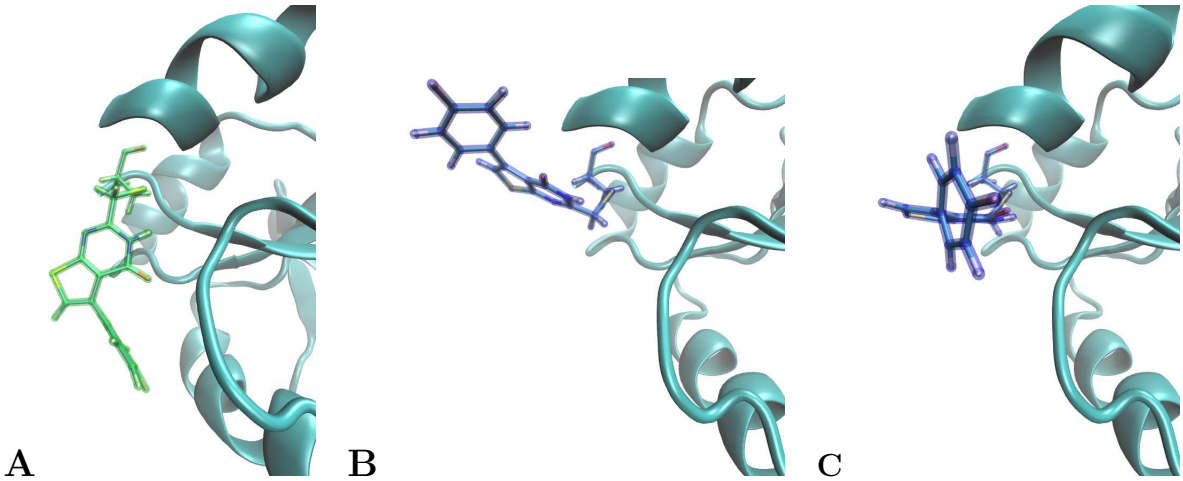


Figure 7.4: **A** – Snapshot from MD simulation of C1 after 30 ns of Tpx in pose I (green). **B** – Initial location of C1 in pose II (blue). **C** – Snapshot from MD simulation of C1 after 30 ns of Tpx in pose II (blue).

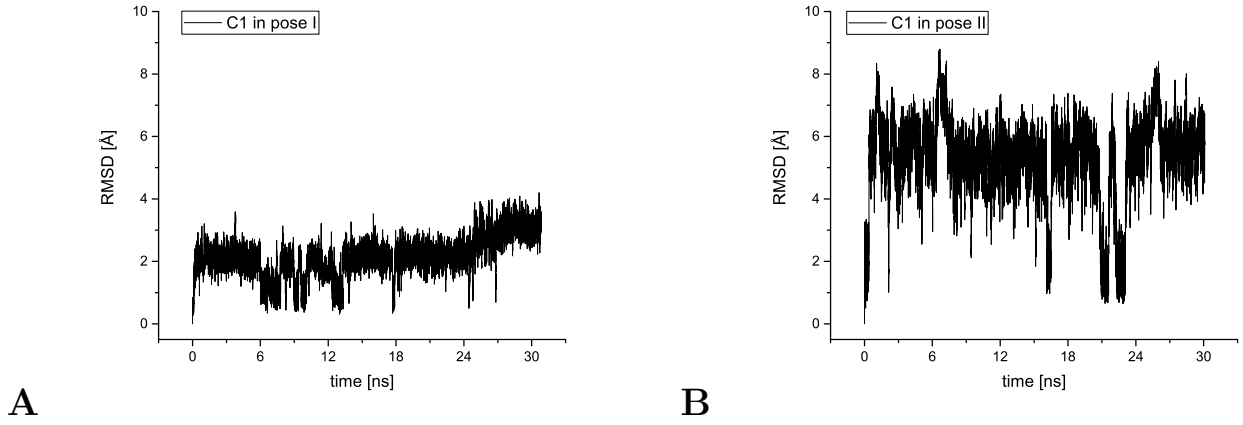


Figure 7.5: Course of RMSD values versus time plots are given for inhibitor C1 in covalent complex with Tpx in **A** – C1 in pose I or **B** – C1 in pose II during the 30 ns production run.

To understand the fluctuations of C1 in pose II one has to analyze interactions with the protein surrounding (Figure 7.6). In case of the first pose one can see that backbone nitrogen and oxygen of Ile109 is of importance. Both can establish hydrogen bonds with the nitrogen or oxygen of the pyrimidone ring of the inhibitor.

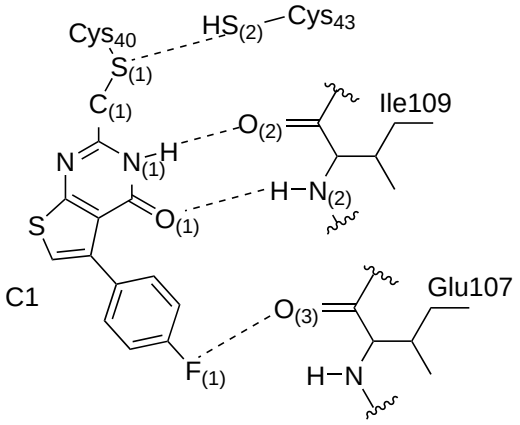


Figure 7.6: Interaction of covalent protein-inhibitor complex of C1 and Tpx depicted with dashed lines. Location of amino acids and C1 serves for clarification and does not reflect actual position in MD simulations.

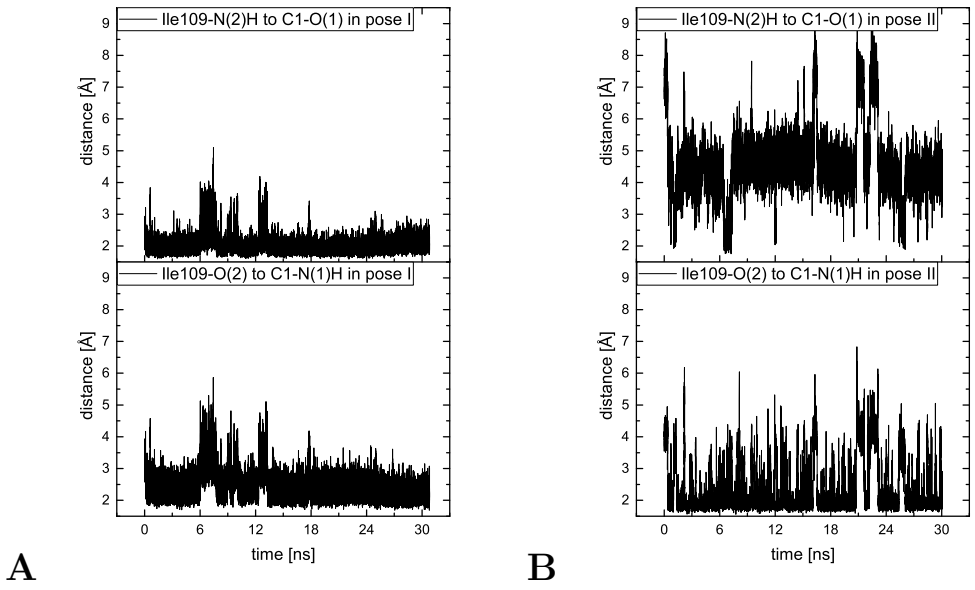


Figure 7.7: The variation in distance as a function of simulation time for inhibitor C1 in Tpx corresponding to Figure 7.6 during a 30 ns production run. **A** – Distance between Ile109-N(2)H to C1-O(1) or Ile109-O(2) to C1-N(1)H in pose I. **B** –Distance between Ile109-N(2)H to C1-O(1) or Ile109-O(2) to C1-N(1)H in pose II.

In terms of pose I the distance between backbone-N(2)H and C1 oxygen O(1) has a mean of 2.1 Å and shows a good interaction (Figure 7.7-A). Fluctuations which occur at ≈ 6 and 12 ns are due the change of orientation of C1 which will be discussed under the section 7.4.

For pose II this is not the case. As one can see in Figure 7.7-B the mean distance between Ile109-N(2)H and C1-O(1) is 4.6 Å and no water-mediated interaction can be observed. Still in pose II one can see a large decrease right at the beginning of the simulation from 7.4 Å to a mean of 4.6 Å. This is explained by the fact of the interaction of Ile109's oxygen and C1's pyrimidon nitrogen. For the interaction between Ile109-O(2) and C1-N(1) both poses show similar mean values, while the standard deviation of pose II is slightly larger. Another observation can be made in terms of the RMSD value of the inhibitor in pose II and its interaction with Ile109. The decrease of distance between C1 and Ile109 and therefore increase of the RMSD value shows a relation between these two quantities (Figure 7.8).

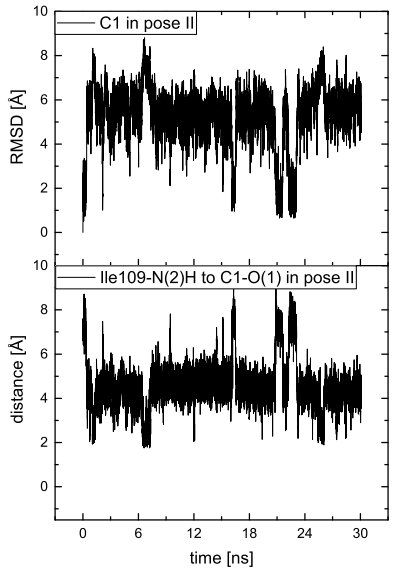


Figure 7.8: Course of RMSD values versus time plots are given for inhibitor C1 in covalent complex with Tpx during the 30 ns production run (top plot). The variation in distance as a function of simulation time for inhibitor C1 in Tpx between Ile109-N(2)H and C1-O(1) corresponding to Figure 7.6 during a 30 ns production run (bottom plot). Both plots are shown for clarification in terms of their relation.

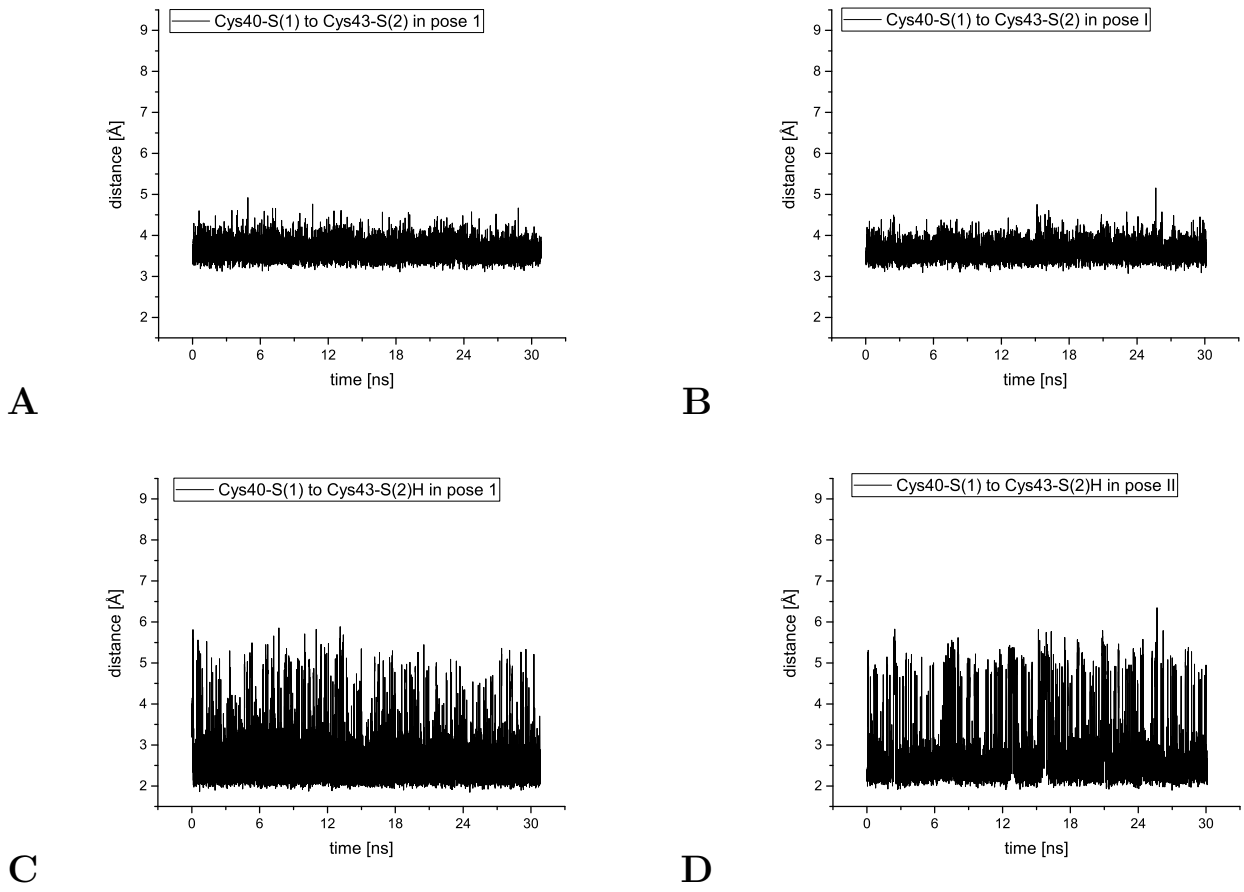
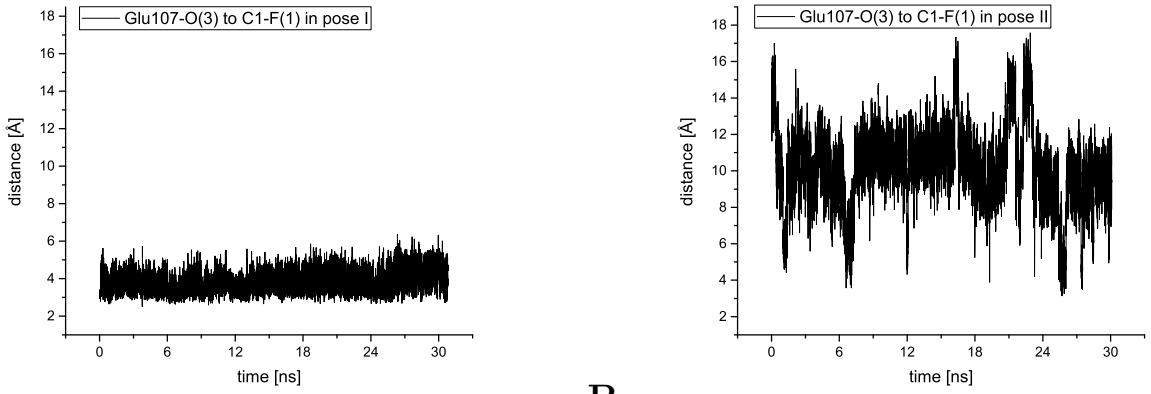


Figure 7.9: The variation in distance as a function of simulation time for inhibitor C1 in Tpx between Cys40 and Cys43 corresponding to Figure 7.6 during a 30 ns production run. **A** –Distance between Cys40-S(1) and Cys43-S(2) in pose I. **B** –Distance between Cys40-S(1) and Cys43-S(2) in pose II. **C** – Distance between Cys40-S(1) and Cys43-S(2)H in pose I. **D** – Distance between Cys40-S(1) and Cys43-S(2)H in pose II.

In terms of the interactions between the cysteine residues 40 and 43 nearly no differences are observable. The mean of distance between the two sulfurs of Cys40 and Cys43 as well as the standard deviation are the same. Only in case of the distance between Cys40 sulfur and Cys43 SH-group the mean and the standard deviation are slightly higher (see Table 7.1).
In regard of further possible interactions the backbone oxygen of Glu107 can be relevant. The distance between the fluorine of C1 and Glu107-O(3) has a mean of 3.8 Å and by variation of the substituents of the phenyl ring further interactions might be established (Figure 7.10-A). Especially the standard deviation of this distance is only 0.6 Å, indicating a possible stable interaction. Still, for pose II those interactions can't be seen as it shows a mean of distance of 10.2 Å although rarely a decrease to $\approx$ 4 Å does occur.



A **B**

Figure 7.10: The variation in distance as a function of simulation time for inhibitor C1 in Tpx between Glu107-O(3) and C1-F(1) corresponding to Figure 7.6 during a 30 ns production run. **A** – Distance between Glu107-O(3) to C1-F(1) in pose I. **B** – Distance between Glu107-O(3) to C1-F(1) in pose II.

Table 7.1: Mean and standard deviation of RMSD values or distances during a 30 ns production run of C1 in covalent complex with wild type Tpx. Mean and standard devations are calculated from 10,000 data points. All values are given in Å.

wild type		**C1 covalent pose I**	**C1 covalent pose II**
		mean / SD	mean / SD
RMSD	C1	2.1 / 0.6	5.2 / 1.4
distance	Ile109-N(2)H to C1-O(1)	2.1 / 0.4	4.6 / 1.1
distance	Ile109-O(2) to C1-N(1)H	2.6 / 0.6	2.3 / 0.8
distance	Glu107-O(3) to C1-F(1)	3.8 / 0.6	10.2 / 2.2
distance	Cys40-S(1) to Cys43-S(2)	3.6 / 0.2	3.6 / 0.2
distance	Cys40-S(1) to Cys43-S(2)H	2.6 / 0.5	2.7 / 0.8

7.4 Influence of point mutations of Tpx on the binding mode of C1

The focus of the following MD simulations lies on the mutation of specific amino acids which might have an effect on the binding mode of C1 in Tpx. Crystal structure revealed that C1 is embedded in the hydrophobic side chains of Trp39, Trp70 and Ile109. Furthermore, Cys43 is of importance in terms of the redox activity of TPX. Therefore, the residues have been modified at each time, i.e. Trp39, Cys43, Trp70 or Ile109 have been mutated to alanine. In

the previous MD simulation of the monomer it was shown that C1 in pose I has lower RMSD values and thus a higher rigidity in Tpx. Hence, chain A has been used for the simulations with respect to the mutants. The duration of the MD simulations consists of 30 ns and have been performed for each point mutation.

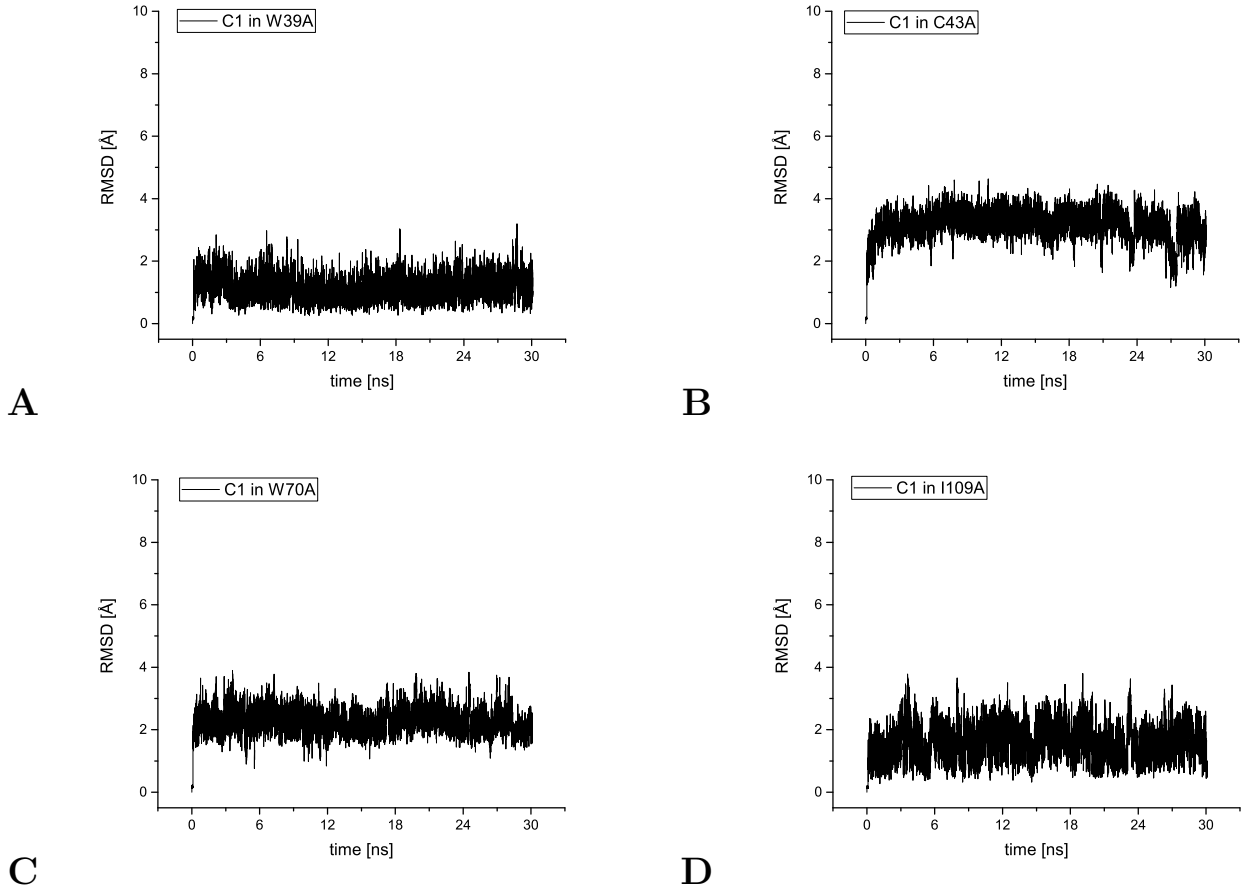


Figure 7.11: Course of RMSD values versus time plots are given for inhibitor C1 in covalent complex with Tpx mutants during the 30 ns production run. **A** – RMSD values of C1 in W39A. **B** – RMSD values of C1 in C43A. **C** – RMSD values of C1 in W70A. **D** – RMSD values of C1 in I109A.

The RMSD values of C1 in Figure 7.11 show that the mutation of these amino acids leads to a decrease in the standard deviation and lessened fluctuations compared to the simulation of C1 in wild type Tpx. In terms of the mean of the RMSD values the modification of Trp39 and also Ile109 to alanine are lower by 0.9 and 0.5 Å. The change of Cys43 or Trp70 to alanine leads to an increase of the mean RMSD value of C1. This shows that in case of Trp39 and Ile109 less fluctuations in the orientation of C1 occurs with respect to the crystallized

inhibitor.

To understand this situation, one has to take a look at Figure 7.12. In terms of Figure 7.12-A one can see that Trp39 in wild type Tpx is able to interact with C1 in a T-shaped stacking which leads to a change of the inhibitor's orientation. In this case the distances between C1 and Ile109 also decrease. Furthermore, it occasionally can have another orientation which is shown in Figure 7.12-B. In this case no interaction between Trp39 and C1 occur, thus allowing the inhibitor to rotate. At the same time the distance to Ile109 increases. This is the explanation for the increase of RMSD in the previous simulation of C1 in pose I in wild type Tpx (compare Figure 7.7-A). While one has a T-shaped stacking between Trp39 and C1 RMSD values of C1 are about 3 Å in the other case they are about 1.5 Å. These findings correlate with the RMSD plots in Figure 7.11.

The mutants C43A and W70A are comparable to the simulation of wild type Tpx. Both simulations show solely the T-shaped stacked orientation (Figure 7.12-C and Figure 7.12-D). For these cases the distance to Ile109 are of the same amount as in the previous simulation of the wild type. Of course, there is also the possibility that C1 might rotate in both mutants, but this event was not observable in performed simulations.

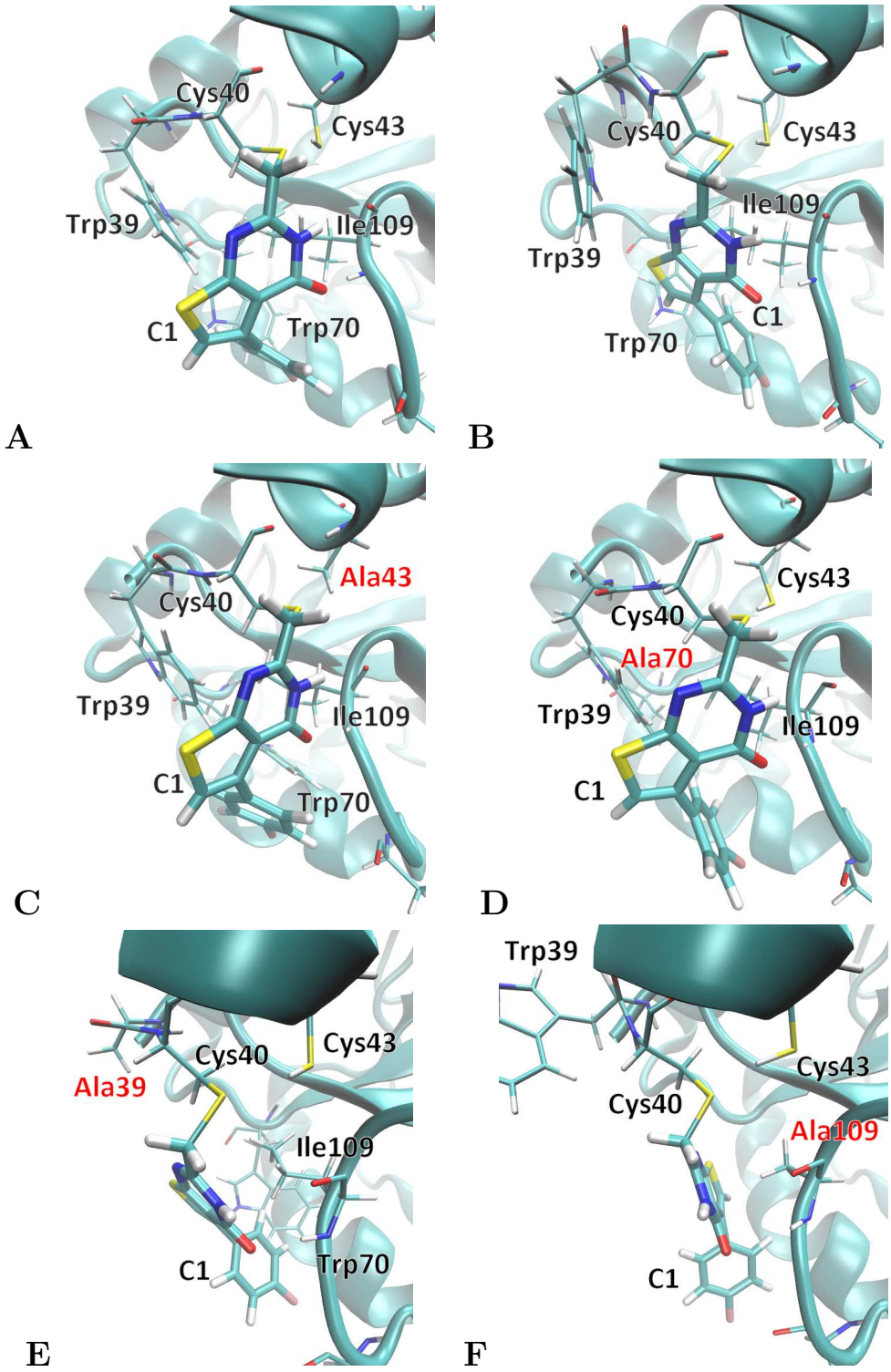


Figure 7.12: **A** – Snapshot of C1 (pose I) in wild type Tpx at 6 ns. **B** – Snapshot of C1 (pose I) in wild type Tpx at 12.6 ns **C** – Snapshot of C1 (pose I) in Tpx C43A mutant at 30 ns. **D** – Snapshot of C1 (pose I) in Tpx W70A mutant at 30 ns. **E** – Snapshot of C1 (pose I) in Tpx W39A mutant at 30 ns. **F** – Snapshot of C1 (pose I) in Tpx I109A mutant at 30 ns.

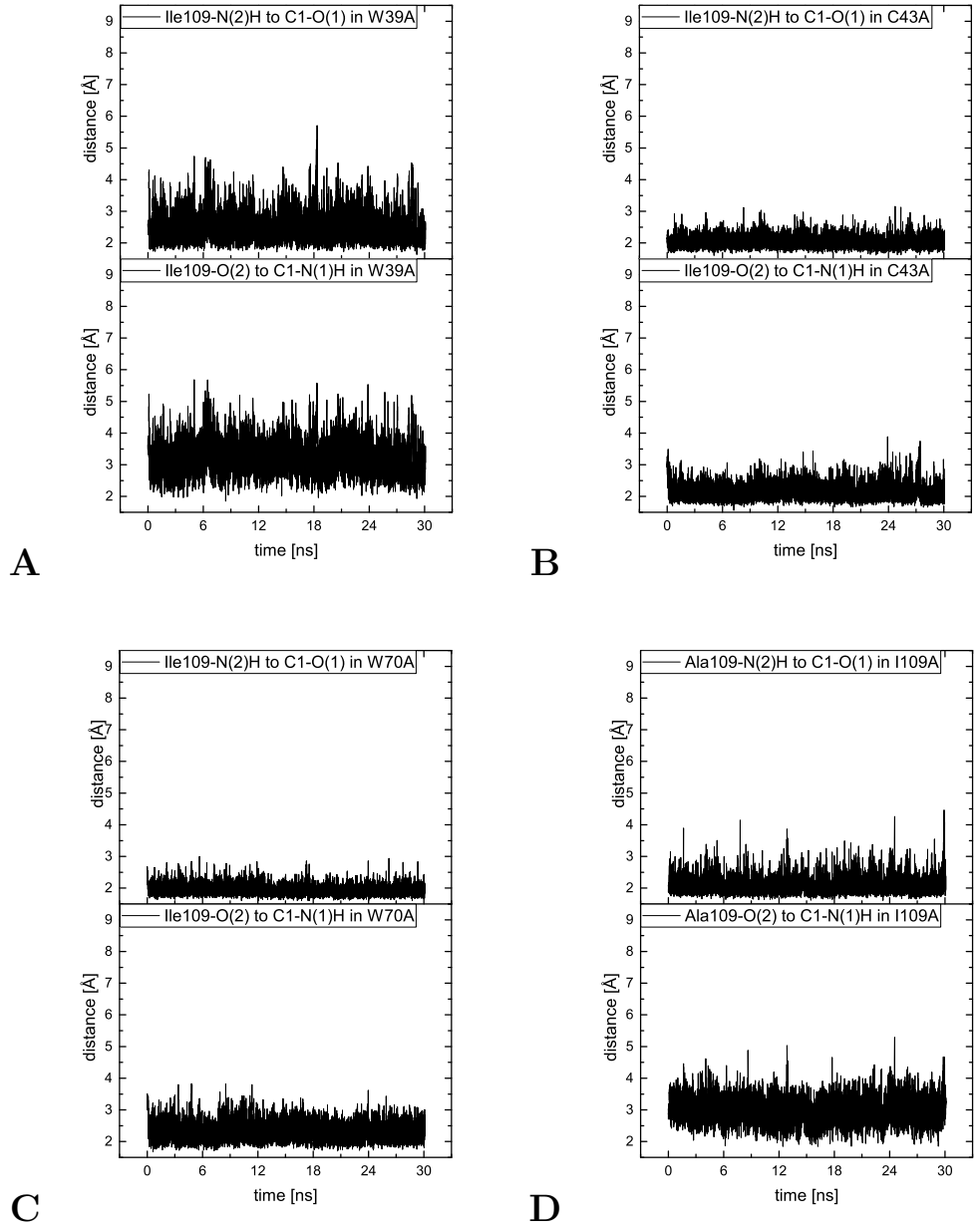


Figure 7.13: The variation in distance as a function of simulation time for inhibitor C1 in Tpx mutants corresponding to Figure 7.6 (in case of I109A it is Ala109) during a 30 ns production run. **A** –Distance between Ile109-N(2)H to C1-O(1) or Ile109-O(2) to C1-N(1)H inW39A. **B** – Distance between Ile109-N(2)H to C1-O(1) or Ile109-O(2) to C1-N(1)H in C43A. **C** –Distance between Ile109-N(2)H to C1-O(1) or Ile109-O(2) to C1-N(1)H in W70A. **D** –Distance between Ala109-N(2)H to C1-O(1) or Ala109-O(2) to C1-N(1)H in I109A.

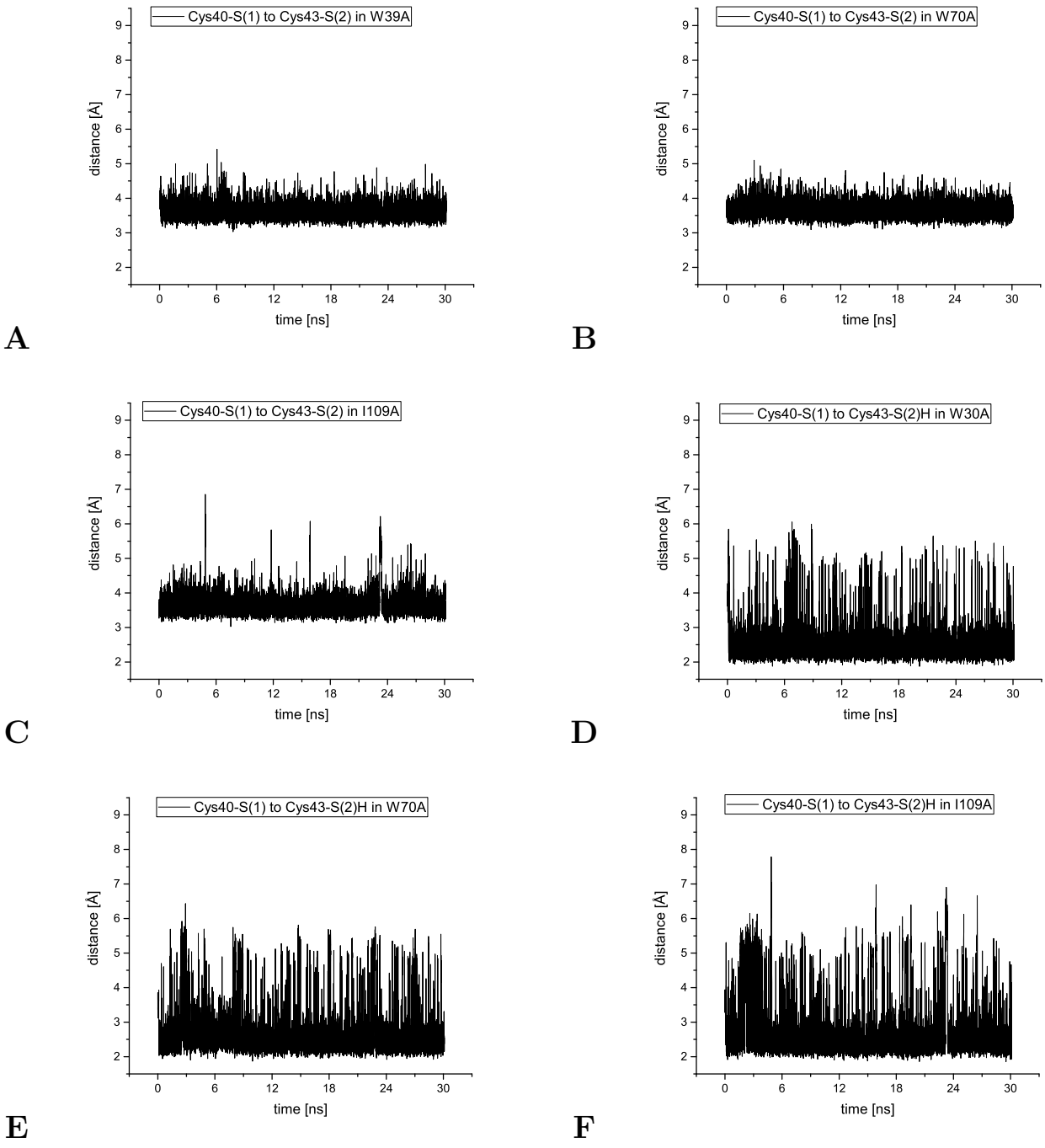


Figure 7.14: The variation in distance as a function of simulation time for inhibitor C1 in Tpx between Cys40 and Cys43 corresponding to Figure 7.6 during a 30 ns production run. **A** –Distance between Cys40-S(1) and Cys43-S(2) in W39A. **B** –Distance between Cys40-S(1) and Cys43-S(2) in W70A. **C** – Distance between Cys40-S(1) and Cys43-S(2) in I109A. **D** –Distance between Cys40-S(1) and Cys43-S(2)in W39A. **E** –Distance between Cys40-S(1) and Cys43-S(2)H in W70A. **F** - Distance between Cys40-S(1) and Cys43-S(2)H in I109A.

The simulation of the variant W39A in complex with C1 shows an orientation in which the RMSD values are lower and the distance to Ile109 is increased in contrast to the wild type simulation (Figure 7.13-A). This is understandable as no interaction was possible now since the bulky side chain of Trp39 was missing.
Also, in case of I109A the RMSD value is slightly lower because of a similar reason. As one can see in Figure 7.12-F the mutation of isoleucine to alanine lead to increased space which allowed the inhibitor to move further into the pocket. At the same time one can observe a movement of Trp39 towards the solvent.
In terms of the distance between Cys40 and Cys43 (except for the C43A mutant) the mean values of the sulfur-sulfur distance are comparable (Figure 7.14). The hydrogen-sulfur interaction is comparable in all three mutants as they show differences below 0.2 Å for the mean and below 0.3 Å for the standard deviation (Table 7.2).
Just as in the simulations before the distance between the fluorine and Glu107 backbone oxygen has been measured. For the variants W39A, W70A and I109A the mean distances are slightly lower compared to the wild type simulation whereas C43A shows a slightly larger mean distance of 4.2 Å (see Figure 7.15 and Table 7.2).

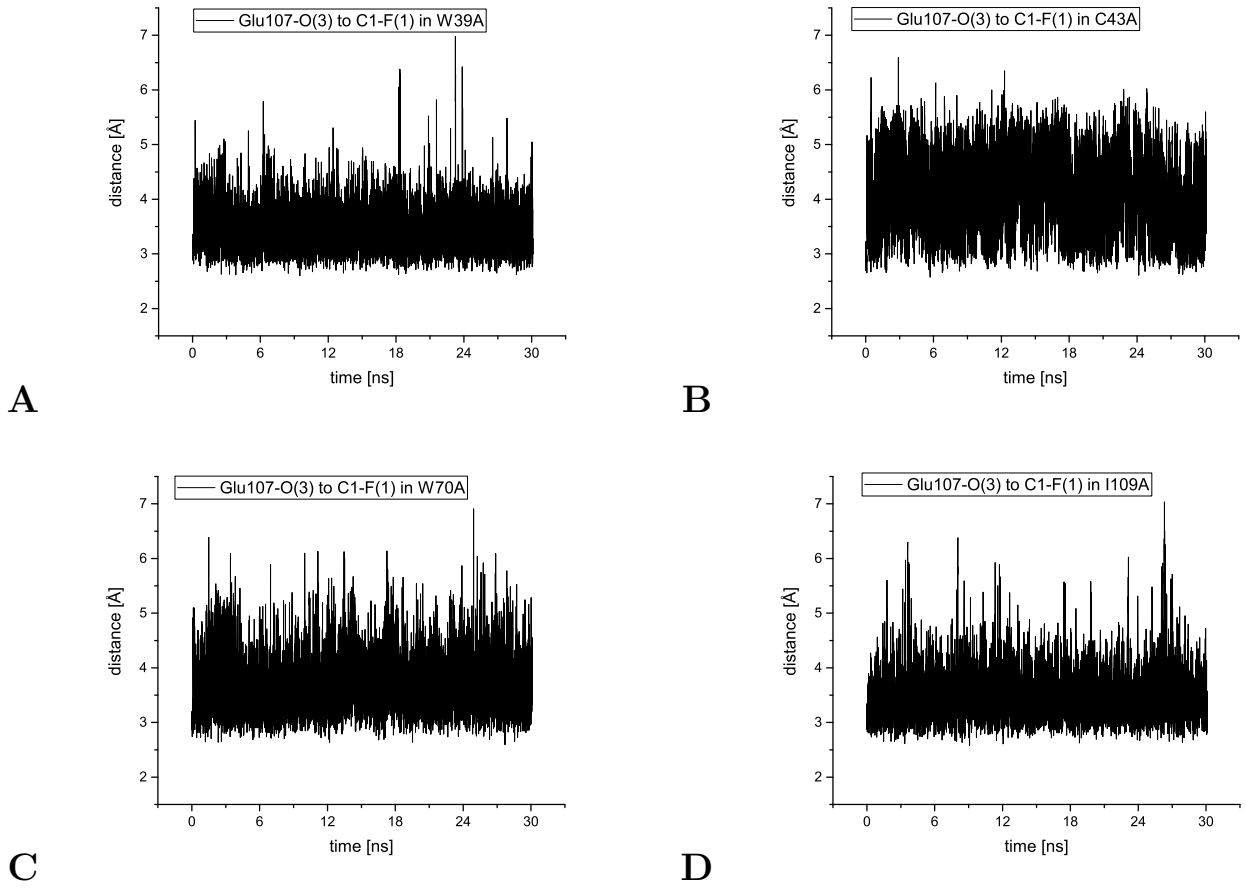


Figure 7.15: The variation in distance as a function of simulation time for inhibitor C1 in Tpx mutants between Ile109-N(2)H and C1-O(1) (in case of I109A it is Ala109) corresponding to Figure 7.6 during a 30 ns production run. **A** –Distance between Glu107-O(3) to C1-F(1) in W39A. **B** – Distance between Glu107-O(3) to C1-F(1) in C43A. **C** – Distance between Glu107-O(3) to C1-F(1) in W70A. **D** – Distance between Glu107-O(3) to C1-F(1) in I109A.

Table 7.2: Mean and standard deviation of RMSD values or distances during a 30 ns production run of C1 in covalent complex with Tpx mutants. Mean and standard devations are calculated from 10,000 data points. All values are presented in Å.

mutants		**Trp39_Ala** mean / SD	**Cys43_Ala** mean / SD	**Trp70_Ala** mean / SD	**Ile109_Ala** mean / SD
RMSD	C1	1.2 / 0.4	3.1 / 0.5	2.7 / 0.4	1.6 / 0.6
distance	Ile109-N(2)H to C1-O(1)	2.5 / 0.5	2.0 / 0.2	1.9 / 0.2	2.1 / 0.3
distance	Ile109-O(2) to C1-N(1)H	3.2 / 0.5	2.1 / 0.3	2.3 / 0.3	3.0 / 0.4
distance	Glu107-O(3) to C1-F(1)	3.4 / 0.4	4.2 / 0.7	3.7 / 0.5	3.5 / 0.5
distance	Cys40-S(1) to Cys43-S(2)H	3.6 / 0.2	- / -	3.6 / 0.2	3.6 / 0.3
distance	Cys40-S(1) to Cys43-S(2)H	2.5 / 0.5	- / -	2.6 / 0.7	2.7 / 0.8

7.5 Stability of the non-covalent complex of C1 in wild type Tpx

Another aspect, which has been examined, was the stability of the non-covalent protein-inhibitor complex. These investigations are conducted to gain further insight on the course of inhibition, as the crystal structure only consists of the final state of reaction. Chain B of the pdb structure Tpx_wt_inh_cc463a_pos1_final.pdb has been used as starting point as C1 in absence of chlorine was present. It is suspected that due to the radiation of the X-ray a cleavage of the covalent bond occurred and this lead to the crystallized non-covalent form. In this case, a methyl group instead of a chloromethyl group was at position 2 of the pyrimidin-4(3H)-one ring. Additionally, no chlorine was located in the vicinity of the binding site. It is to be expected that after formation of the covalent bond chlorine diffuses out of the binding pocket.

In order to analyze the non-covalent complex of C1 in Tpx it was of necessity to replace the hydrogen of the crystallized structure by chlorine. The parametrization has been changed in this case as the HF/6-31G(d) obtained charges showed one possibly incorrect value: the electrophilic carbon of C1 had a charge of -0.359 (see Figure 7.16). Because of that the method was changed and the basis set was expanded to M06-2X/aug-cc-pVDZ. This led to more meaningful values as the charge of electrophilic carbon changed to 0.382. The non-

covalent complex has been investigated for three different protonation states of Cys40 and Cys43 depicted in Figure 7.17. Additionally, the distance between the nucleophilic sulfur of Cys40 and the electrophilic carbon of C1 has been measured to estimate the possibility of covalent modification. Another difference regarding the other simulations is the production run time. This was reduced to 10 ns for each state. In these simulations solely the wild type was considered.

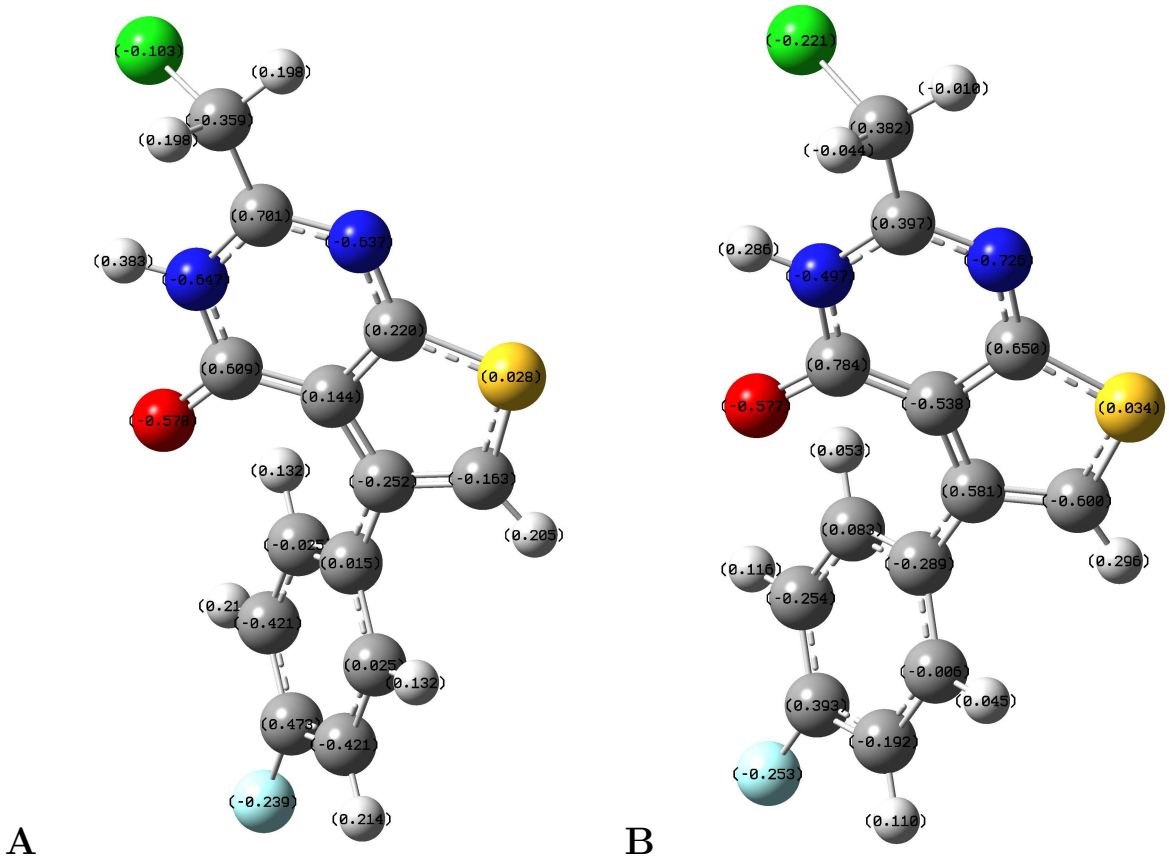


Figure 7.16: **A** – Calculated charges of inhibitor C1 using HF and the 6-31G(d) basis set. **B** – Calculated charges of inhibitor C1 using DFT functional M06-2X and the aug-cc-pVDZ basis set.

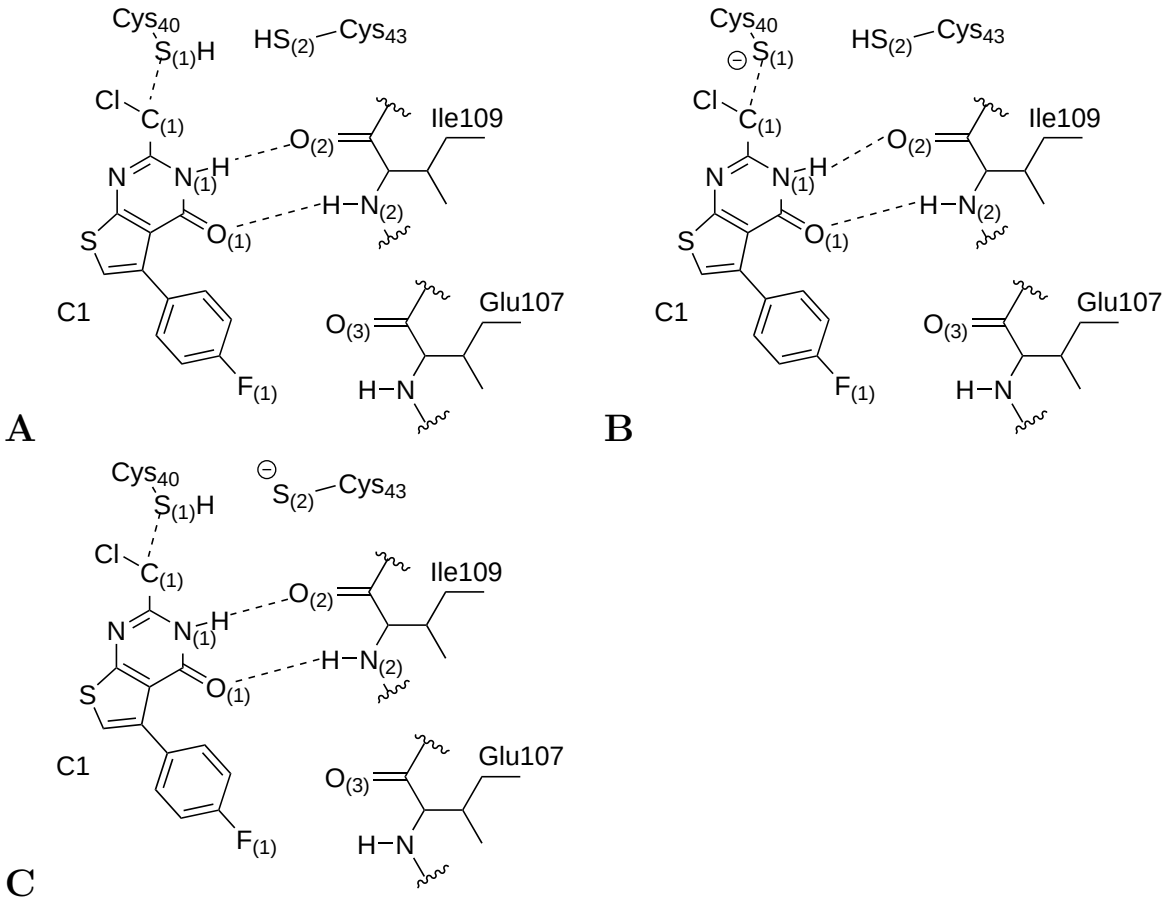


Figure 7.17: Interaction of non-covalent protein-inhibitor complex of C1 and Tpx depicted with dashed lines. Location of amino acids and C1 serves for clarification and does not reflect actual position in MD simulations. Variation of protonation states of Cys40 and Cys43 as followed: **A** – protonated Cys40 and protonated Cys43, **B** – deprotonated Cys40 and protonated Cys43, **C** – protonated Cys40 and deprotonated Cys43.

If one takes a look at the RMSD plots (Figure 7.18), all three protonation states show higher RMSD values compared to the simulation of the covalent complex. This is to be expected, as the non-covalent complex should show more fluctuations, it was observable that in case of a deprotonated Cys43 and especially for the neutral state (protonated Cys40 and Cys43) the RMSD values have a mean of 5.9 Å or 7.5 Å and are much larger in relation to the covalent complex. For the simulation of a deprotonated Cys40 the mean is of 4.2 Å and revealing a lower standard deviation of 0.9 Å.
In terms of the neutral system (Figure 7.17-A) the RMSD value increases strongly after 2 ns and at the same time the distance between Cys40-S(1) and C1-C(1) increases up to 16 Å due to a movement into a flipped orientation. Also, in case of the deprotonated Cys43 system

(Figure 7.17-C) an increase to 6 Å in RMSD is observable after 1 ns which lowers to 3 Å at 9 ns.

Although the simulation with the deprotonated Cys40 (Figure 7.17-B) showed the lowest mean in respect to the RMSD values, the distance between Cys40-S(1) and the electrophilic carbon C1-C(1) increases up to 12 Å (mean of simulation = 7.4 Å) and therefore leading to the conclusion that the covalent modification only occurs if the deprotonated sulfur is close enough to the electrophilic carbon of C1 to bind irreversible.

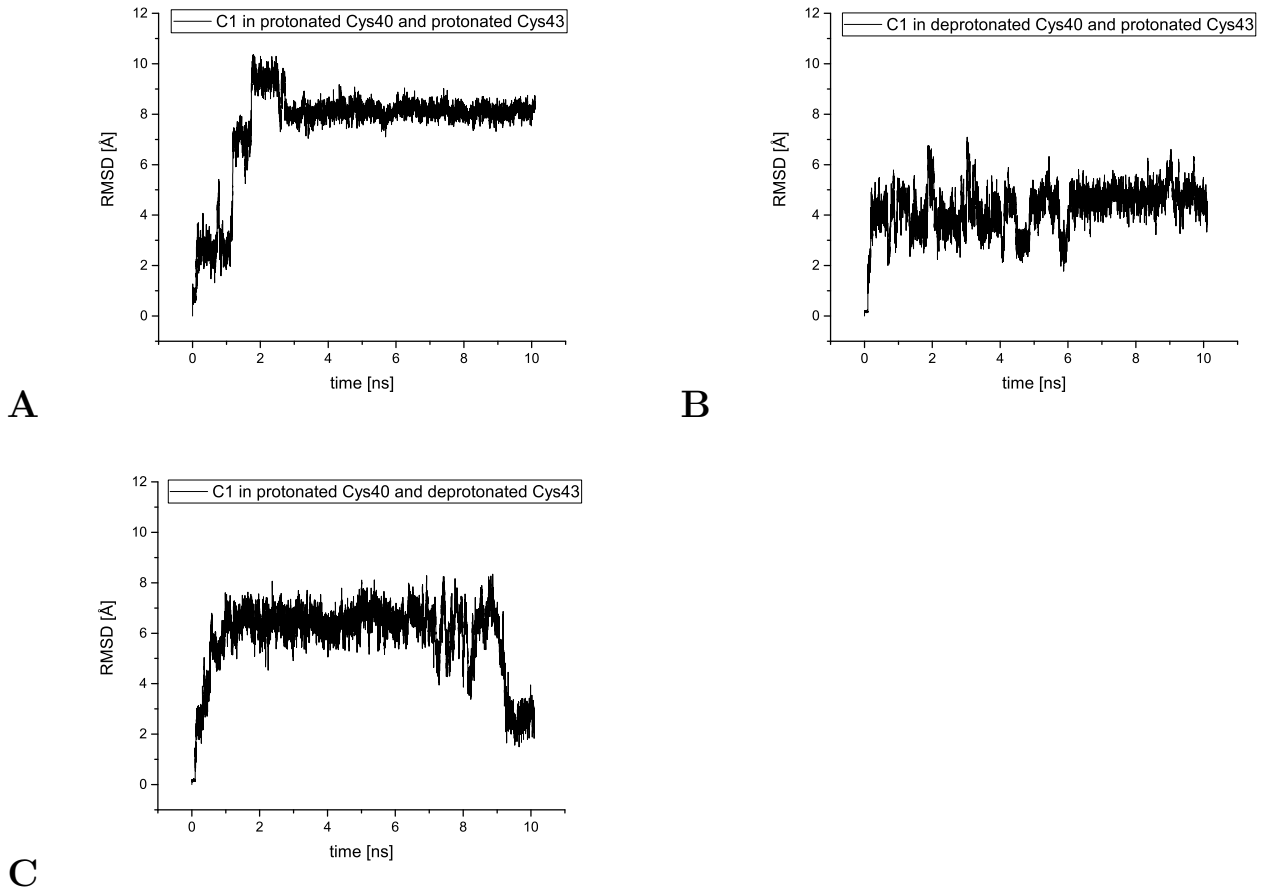


Figure 7.18: Course of RMSD values versus time plots are given for inhibitor C1 in non-covalent complex with wild type Tpx during the 10 ns production run. **A** – RMSD values of C1 in protonation state Cys40-SH and Cys43-SH. **B** – RMSD values of C1 in protonation state Cys40-S^- and Cys43-SH. **C** – RMSD values of C1 in protonation state Cys40-SH and Cys43-S^-.

The findings in terms of the course of RMSD values are in relation to the interactions

with Ile109. If one compares the distances depicted in Figure 7.20 with the RMSD plots in Figure 7.18 one can see that the variations in value behave similarly. While the RMSD value increases one can observe an increase in distance thus weakening of the hydrogen bonds. The lowest mean in distance can be seen for the simulation with the deprotonated Cys40. Still, these interactions might not be sufficient for a stable non-covalent protein-inhibitor complex which would already lead to a deactivation of TPX.

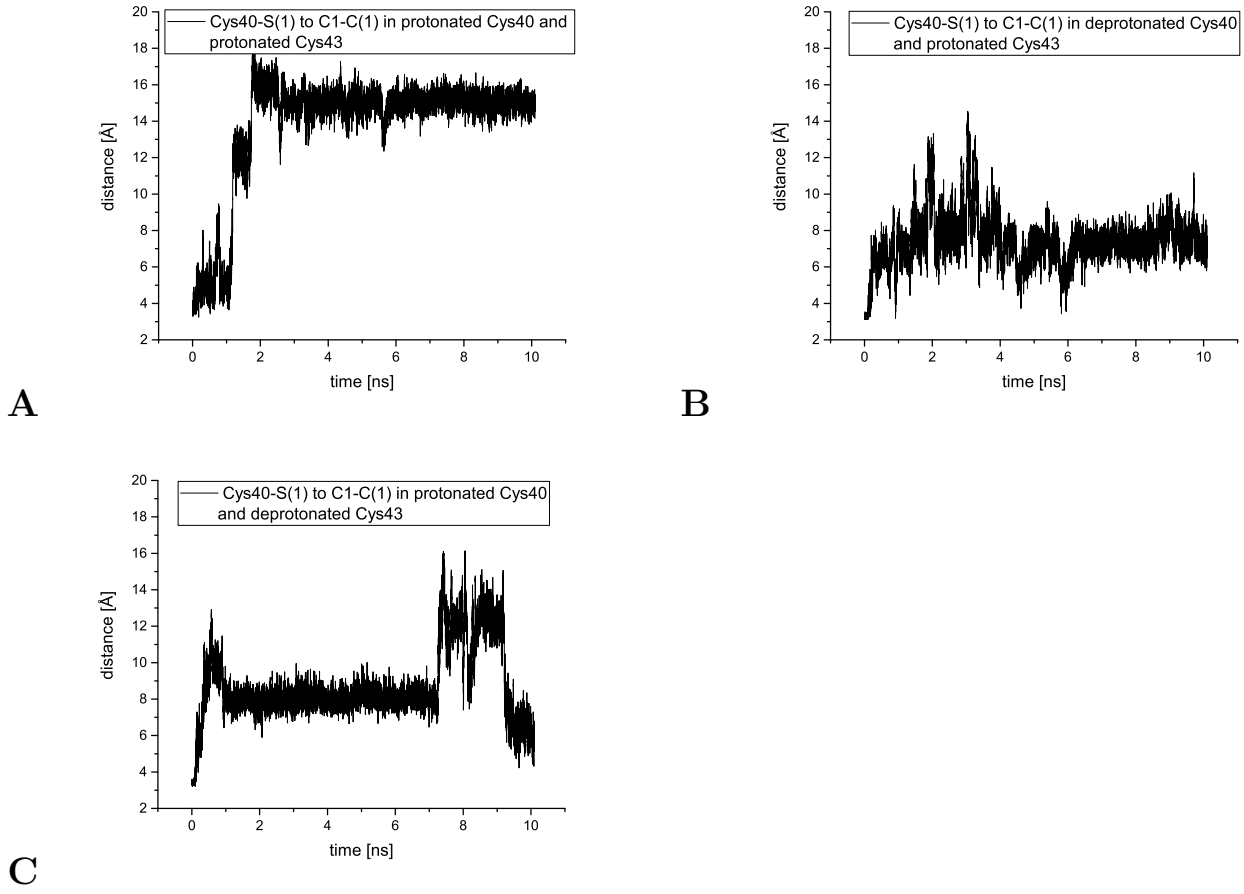


Figure 7.19: The variation in distance as a function of simulation time for inhibitor C1 in Tpx between Cys40-S(1) and C1-C(1) corresponding to Figure 7.17 during a 10 ns production run. **A** – Distance between Cys40-S(1) and C1-C(1) n the protonation state Cys40-SH and Cys43-SH. **B** – Distance between Cys40-S(1) and C1-C(1) in the protonation state Cys40-S^- and Cys43-SH. **C** – Distance between Cys40-S(1) and C1-C(1) in the protonation state Cys40-SH and Cys43-S^-.

Therefore, the simulations are in line with the experimental findings of C1's derivative (chlorine was exchanged with hydrogen). The simulations showed solely in one case hydrogen

interactions between C1 and the active site of Tpx. Still, the course of RMSD reveals larger fluctuations compared to the covalent complex of wild type Tpx or its mutants, leading to the conclusion that the non-covalent complex is low in stability. In consequence, no inhibition is to be expected, as mirrored in the experimental data.

Table 7.3: Mean and standard deviation of RMSD values or distances during a 10 ns production run of C1 in non-covalent complex with Tpx (in different protonation states for Cys40 and Cys43) corresponding to section 3. Mean and standard devations are calculated from 10,000 data points. All values are in Å.

non covalent		**C1 in Cys40-SH / Cys43-SH** mean / SD	**C1 in Cys40-S$^-$ / Cys43-SH** mean / SD	**C1 in Cys40-SH / Cys43-S$^-$** mean / SD
RMSD	C1	7.5 / 1.9	4.2 / 0.9	5.9 / 1.5
distance	C1-C(1) to Cys40-S(1)	13.8 / 3.3	7.4 / 1.4	8.6 / 2.0 l
distance	Ile109-N(2)H to C1-O(1)	6.7 / 1.8	3.0 / 1.5	7.3 / 2.0
distance	Ile109-O(2) to C1-N(1)H	10.7 / 2.6	5.4 / 1.8	8.6 / 2.1

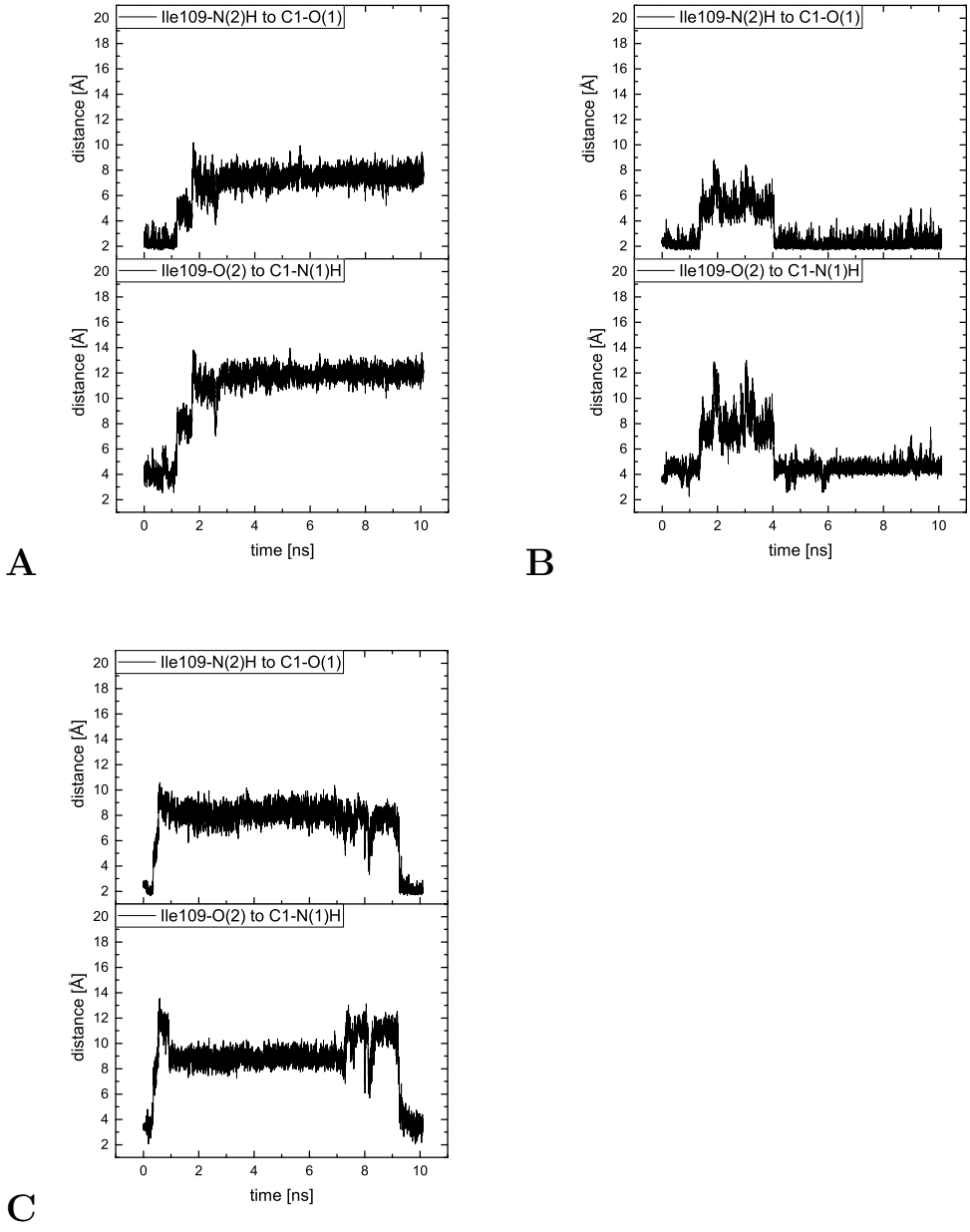


Figure 7.20: The variation in distance as a function of simulation time for inhibitor C1 in Tpx corresponding to Figure 7.17 during a 10 ns production run. **A** – Distance between Ile109-N(2)H to C1-O(1) or Ile109-O(2) to C1-N(1)H In the protonation state Cys40-SH and Cys43-SH. **B** – Distance between Ile109-N(2)H to C1-O(1) or Ile109-O(2) to C1-N(1)H in the protonation state Cys40-S^- and Cys43-SH. **C** – Distance between Ile109-N(2)H to C1-O(1) or Ile109-O(2) to C1-N(1)H in the protonation state Cys40-SH and Cys43-S^-.

7.6 Wild type Tpx in absence of inhibitor in different protonation states

In this section interactions of both active site cysteine residues have been analyzed in absence of the inhibitor C1. Crystal structures do not contain hydrogens and thus the protonation state is not known. Still, this information is needed to get further insights as only a deprotonated cysteine can be modified. Therefore, MD simulations were conducted to determine the most likely protonation state of Cys40 and Cys43. The system was prepared by removal of the inhibitor and preparation of three different protonation states (see Figure 7.21). Afterwards, hydrogen bonds and sulfur-sulfur distances have been calculated based on the 30 ns production runs.

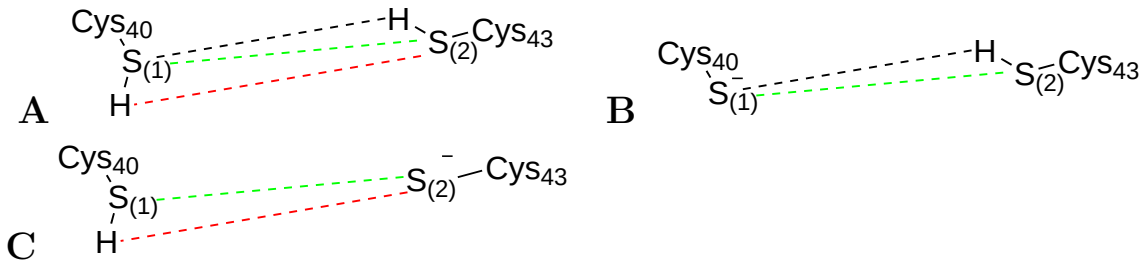


Figure 7.21: Interaction of Cys40 and Cys43 in wild type Tpx in absence of inhibitor depicted with dashed lines. Location of amino acids serves for clarification and does not reflect actual position in MD simulations. Variation of protonation states of Cys40 and Cys43 have been conducted as followed: **A** – protonated Cys40 and protonated Cys43. **B** – deprotonated Cys40 and protonated Cys43. **C** – protonated Cys40 and deprotonated Cys43.

The distance plots show that the strongest interactions between both Cys residues are obtained for the neutral state as the standard deviations are below 0.5 Å (Figure 7.22). Also, the sulfur-sulfur distance (depicted in Figure 7.21) with a mean of 3.6 Å is comparable to the MD simulations of the covalent complex (wild type and mutants). For a deprotonated Cys40 the mean values are slightly higher compared to the neutral state and they also show fluctuations (Table 7.4). The increase in distance which can be seen in Figure 7.22-B occurs due to the movement of Trp39 towards the water shell. It is not clear if the weaker interaction leads to that motion or the movement itself weakens the interaction of both cysteine residues. For the last protonation state, the behaviour is completely different. Firstly, the deprotonated Cys43 orientates towards Tyr80's hydroxyl group at $\approx$ 5 ns. This leads to an increase in distance to Cys40 and thus weaker interactions. Subsequently, Cys40 starts to orientate

towards the solvent at the same time as Trp39 and a return to its initial position was not observable. In consequence, no hydrogen bond to Cys43 was established again.

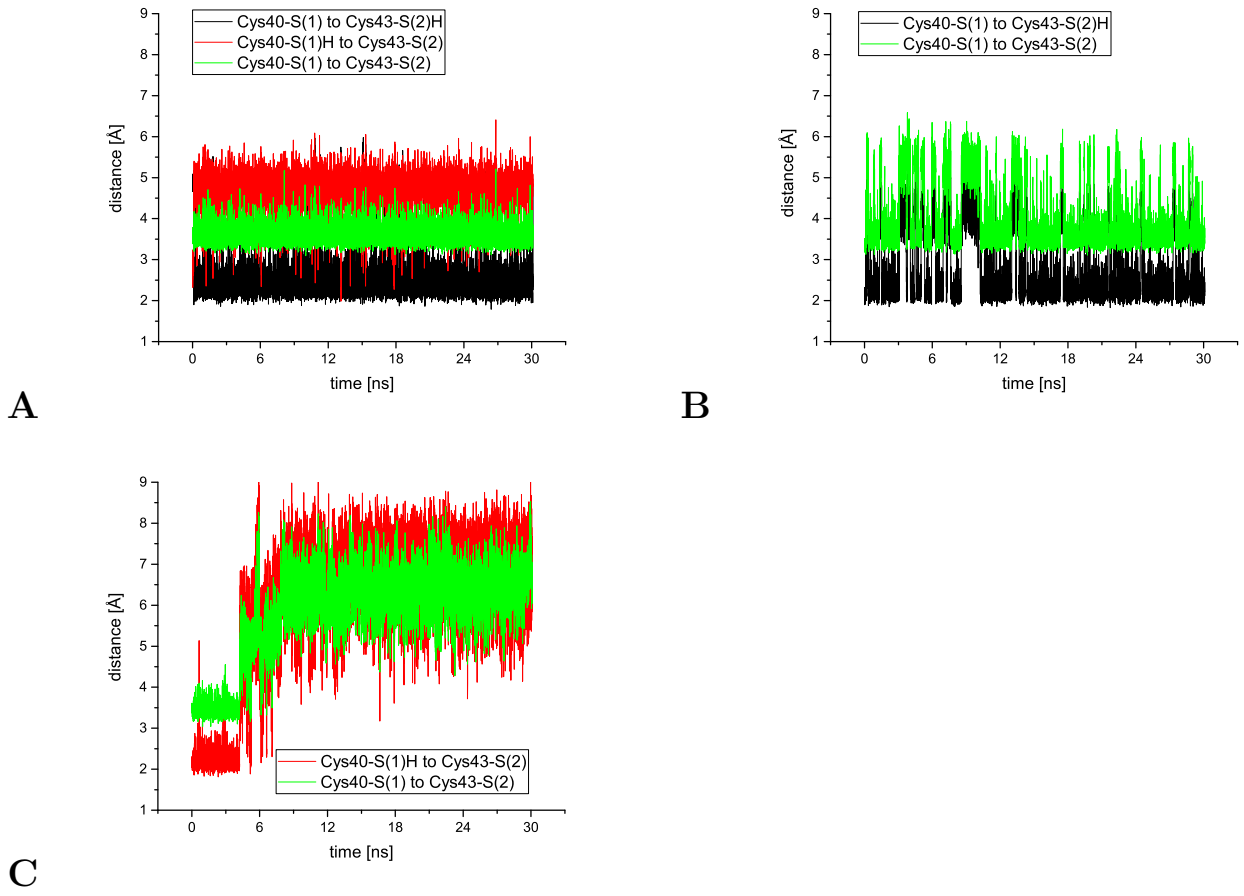


Figure 7.22: The variation in distance as a function of simulation time for wild type C1 in absence of Tpx between Cys40 and Cys43 corresponding to Figure 7.21 during a 30 ns production run. **A** –Distance between Cys40-S(1) and Cys43-S(2)H (black curve), Cys40-S(1)H to Cys43-S(2) (red curve) or Cys40-S(1) to Cys43-S(2) (green curve) for protonated Cys40 and Cys40. **B** – Distance between Cys40-S(1) and Cys43-S(2)H (black curve) or Cys40-S(1) to Cys43-S(2) (green curve) for deprotonated Cys40. **C** –Distance between Cys40-S(1)H to Cys43-S(2) (red curve) or Cys40-S(1) to Cys43-S(2) (green curve) for deprotonated Cys43.

Table 7.4: Mean and standard deviation of RMSD values or distances during a 30 ns production run of C1 in non-covalent complex with Tpx (in different protonation states for Cys40 and Cys43). Mean and standard devations are calculated from 10,000 data points. All values are given in Å.

Tpx in absence of inhibitor C1		**Cys40-SH** / Cys43-SH mean / SD	**Cys40-S$^-$** / Cys43-SH mean / SD	**Cys40-SH** / Cys43-S$^-$ mean / SD
distance	Cys40-S(1) to Cys43-S(2)H	2.5 / 0.5	2.7 / 0.9	-
distance	Cys40-S(1)H to Cys43-S(2)	4.7 / 0.4	-	6.0 / 1.8
distance	Cys40-S(1) to Cys43-S(2)	3.6 / 0.2	3.9 / 0.8	5.8 / 1.2

7.7 Analysis of the dimerization of Tpx in covalent complex with C1 and in absence of C1

Experimental data reveal that the dimer formation of Tpx solely occurs in presence of a covalently bound C1 with a high affinity as no monomer was detected after the drug was added. Furthermore, the dimer appears to be stable as dilution experiments did not reverse the dimerization. To probe the stability of the Tpx dimer, two different simulations have been conducted. On the one hand in presence of two C1 molecules which are in covalent complex with Tpx and on the other in absence of both inhibitors. In the latter case both cysteine residues were protonated.

At first, investigations focused on interactions of C1 with Tpx dimer as well as on drug-drug interactions in the dimerized Tpx. The dynamics show that both C1 inhibitors have a mean RMSD value of 1.1 Å along with a standard deviation of 0.4 Å which is lower compared to C1 in pose I monomer. These findings are understandable as both molecules interact with Tpx as well as with each other due to π-stacking, thus leading to increased stability. In terms of distance of C1 to backbone-NH group of Ile109 in chain A and chain B, a slight increase in mean by 0.3 Å or 0.6 Å is observable if one compares the values with C1 (pose I) in covalent complex with Tpx monomer.

Interactions between both C1 molecules to the carbonyl group of Ile109 are reduced as mean distances of 3.4 Å or 3.6 Å are obtained in the course of the simulation. This is understandable as due to the intermolecular π-π interactions of both C1 molecules (which apparently appear to be stronger than the interactions to Ile109's backbone) their positions are more

restricted compared to C1 bound to the Tpx monomer.

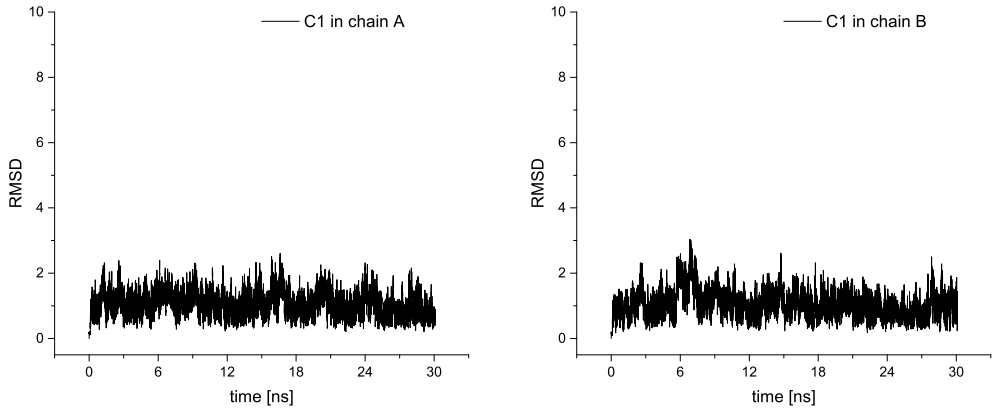


Figure 7.23: Course of RMSD values versus time plots are given for inhibitor C1 in covalent complex with Tpx dimer. (**Left**) C1 covalently bound to Cys40 of chain A. (**Right**) C1 covalently bound to Cys40 of chain B during the 30 ns production run.

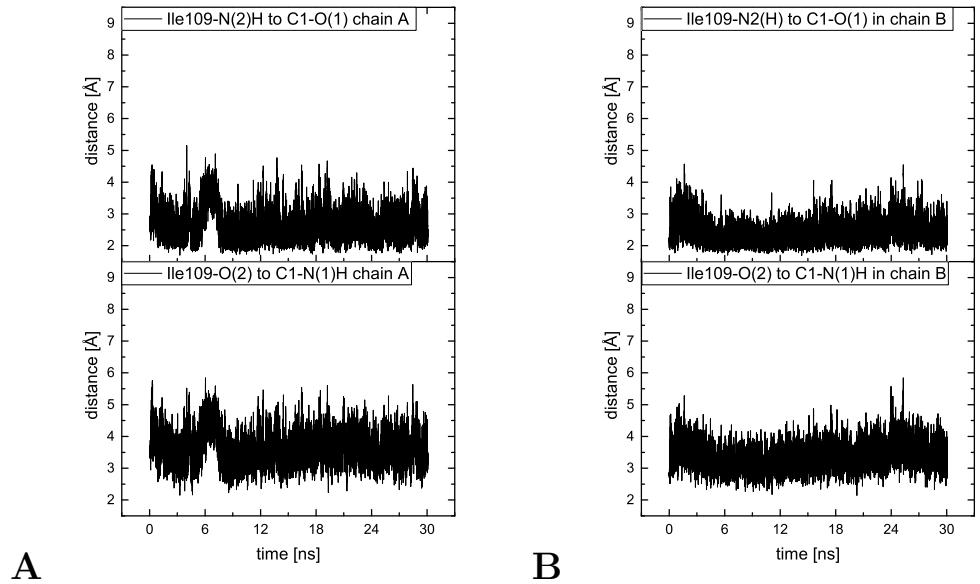


Figure 7.24: The variation in distance as a function of simulation time for inhibitor C1 in Tpx dimer corresponding to Figure 7.6 during a 30 ns production run. **A** – Distance between Ile109-N(2)H to C1-O(1) or Ile109-O(2) to C1-N(1)H in chain A. **B** –Distance between Ile109-N(2)H to C1-O(1) or Ile109-O(2) to C1-N(1)H in chain B.

These interactions are observable if one calculates the variation in distance of perpendicular atoms of C1 in chain A to atoms of chain B (Figure 7.25 and Table 7.6). Interactions between both thiophene rings of C1 in chain A and chain B show a mean in distance of 3.7 Å or 3.8 Å with a low standard deviation of 0.2 Å or 0.3 Å (Table 7.6). This indicates that a stable π-π interaction between these two heterocycles is present. In case of the flourophenyl-rings of both C1 the flexibility is slightly higher as the standard deviation is larger. These differences might occur due to the sandwich conformation of both thiophene rings while the flourophenyl-rings are stacked in a parallel displaced manner. These findings reveal a stable intermolecular interaction of both C1 molecules as well as an appropriate interaction to Ile109's backbone which results into lower mean RMSD values of both inhibitors compared to C1 in Tpx monomer as mentioned before. Additionally, the distance of the fluorine of both molecules to Glu107's backbone oxygen show slightly lower distances and fluctuations compared to the results of Tpx monomer in complex with C1. This can also be explained by the increased immobility of both inhibitors.

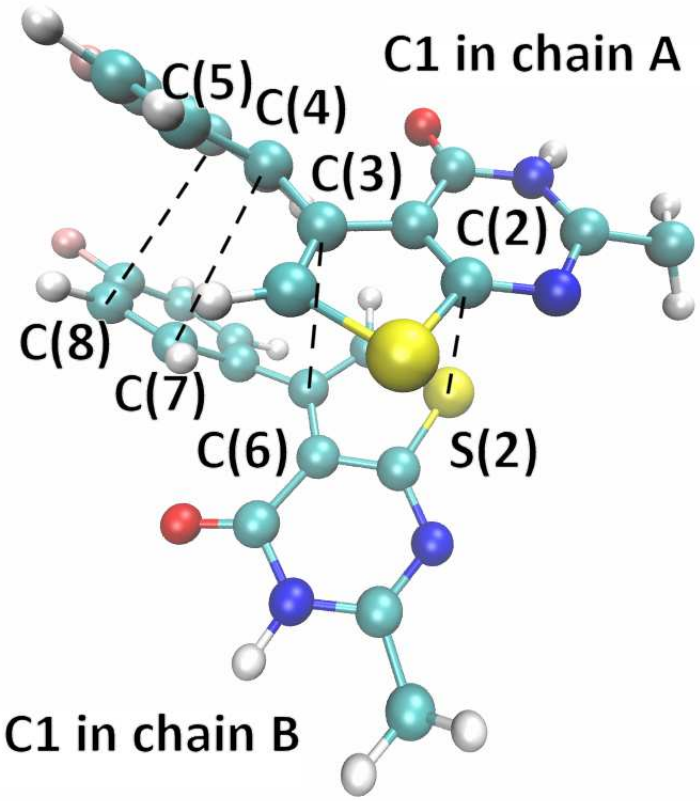


Figure 7.25: Selected perpendicular atoms of two C1 molecules in covalent complex with chain A or chain B of Tpx depicted with dashed lines.

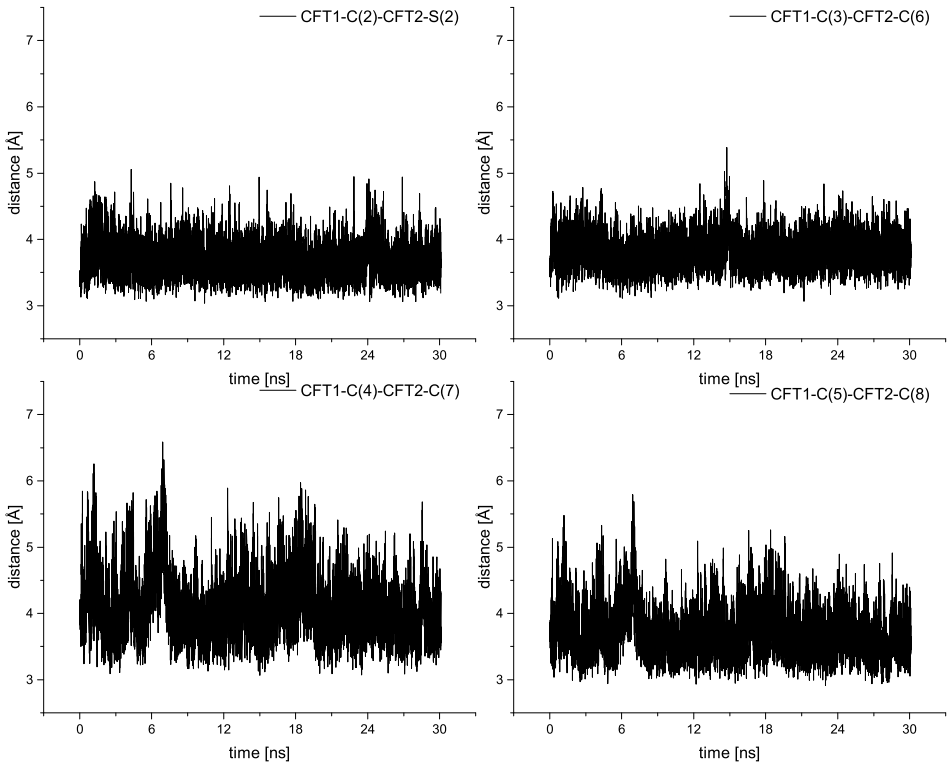


Figure 7.26: The variation in distance as a function of simulation time for both C1 molecules in chain A or chain B of Tpx corresponding to Figure 7.25 during a 30 ns production run. **A** – Distance between C1-C(2) in chain A to C1-S(2) in chain B. **B** – Distance between C1-C(3) in chain A to C1-C(6) in chain B. **C** – Distance between C1-C(4) in chain A to C1-C(7) in chain B. **D** – Distance between C1-C(5) in chain A to C1-C(8) in chain B.

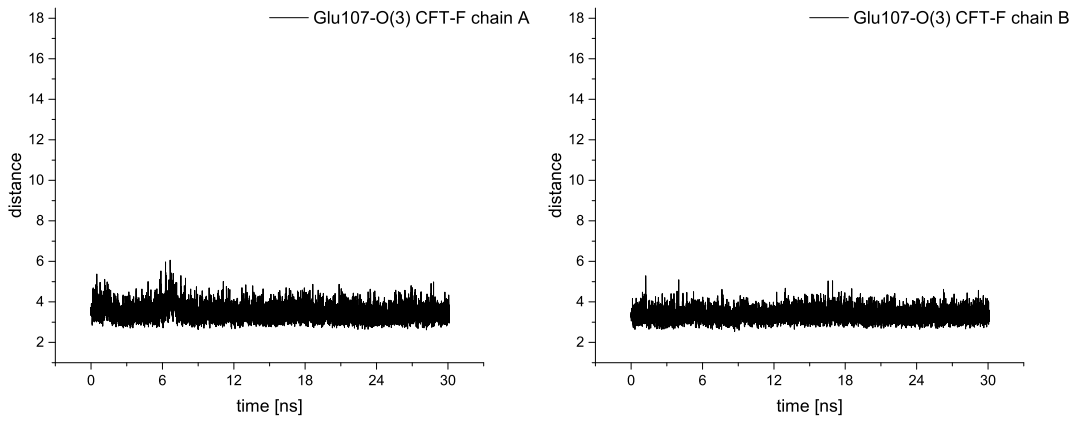


Figure 7.27: The variation in distance as a function of simulation time for inhibitor C1 in Tpx between Glu107-O(3) and C1-F(1) corresponding to Figure 7.6 during a 30 ns production run. Distance between Glu107-O(3) to C1-F(1) in chain A (left). Distance between Glu107-O(3) to C1-F(1) in chain B (right).

Another point of interest was the stability of the Tpx dimer. To probe the necessity of C1 in terms of dimer stabilization the simulation was relaunched in absence of the inhibitor. If one compares the mean and especially the standard deviation of the RMSD of both chains one can see larger fluctuations if C1 is not covalently bound to Tpx (Figure 7.28). If one looks at Table 7.7, chain A and B show low fluctuations with a standard deviation of 0.2 Å with a mean of 1.8 Å or 1.9 Å. Contrary to these results larger means of 2.8 Å (for chain A) or 3.1 Å (for chain B) are obtained in absence of the inhibitors and an increase of about 0.8 Å in standard deviation is observable. Moreover, after 24 ns an increase of $\sim$ 2 Å occurs if no inhibitor molecules are present.

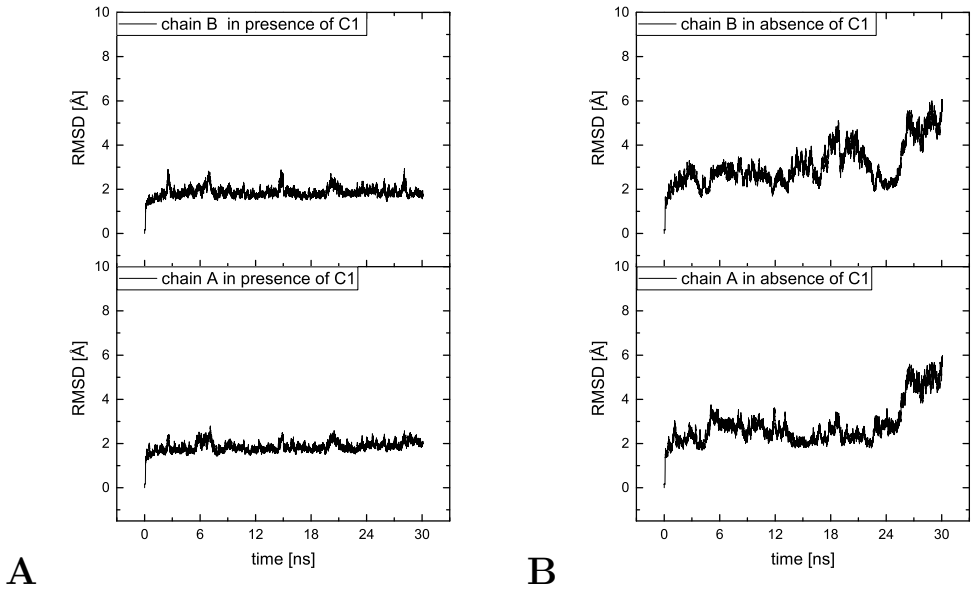


Figure 7.28: Course of RMSD values versus time plots are given for chain A and chain B of Tpx dimer in **A** – presence of two covalently bound C1 molecules or in **B** – absence of both C1 molecules during the 30 ns production run.

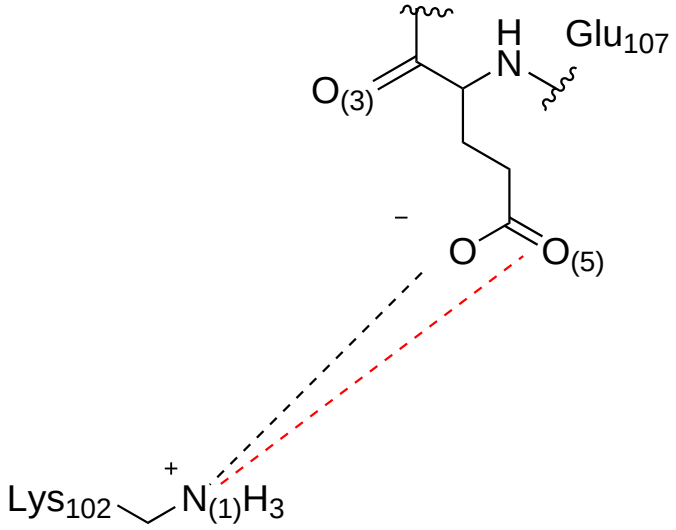


Figure 7.29: Interaction of Lys102 (chain A) and Glu107 (chain B) or Lys102 (chain B) and Glu107 (chain A) of Tpx dimer depicted with dashed lines. Location of amino acids and C1 serves for clarification and does not reflect actual position in MD simulations.

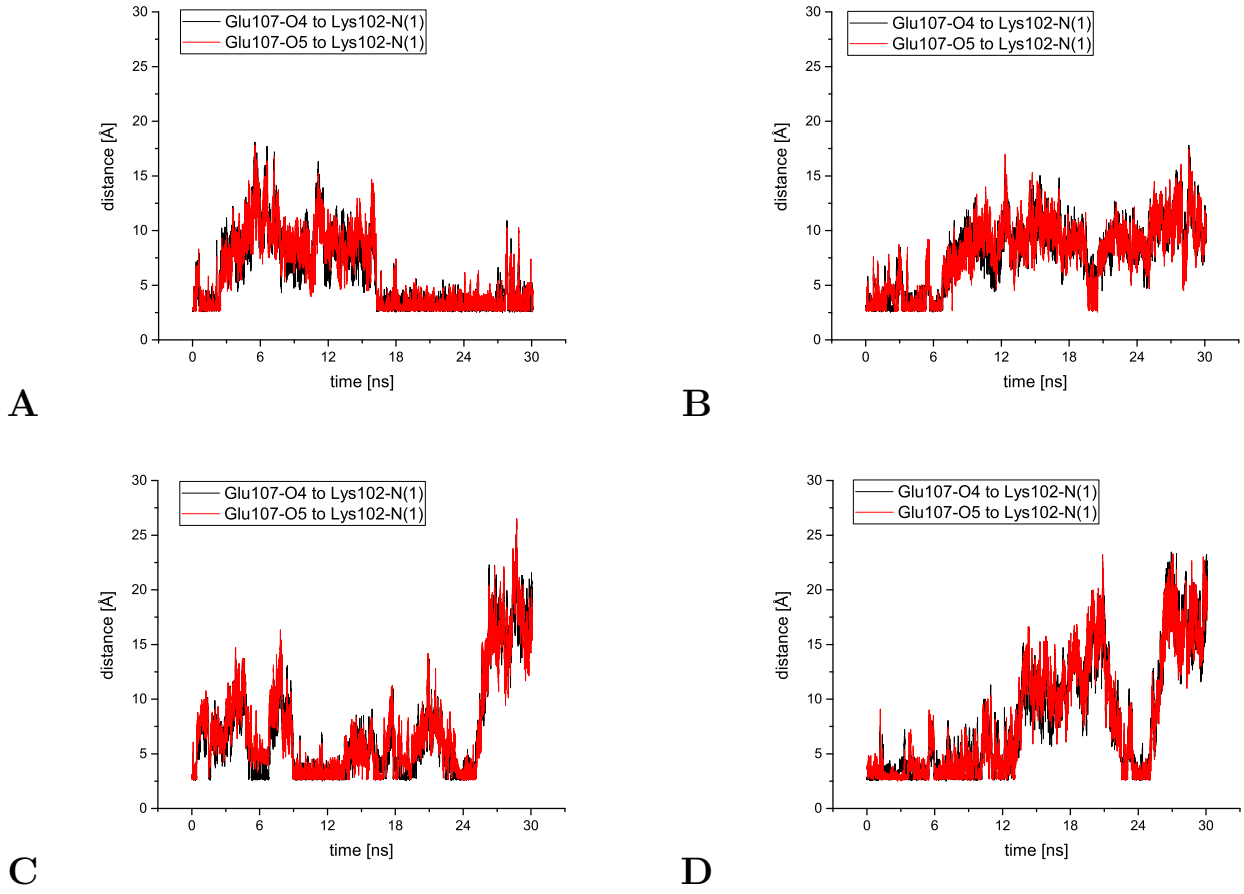


Figure 7.30: The variation in distance as a function of simulation time for **A** - Glu107 (chain A) to Lys102 (chain B) and **B** – Glu107 (chain B) to Lys102 (chain A) in presence of two C1 molecules or **C** – Glu107 (chain A) to Lys102 (chain B) and **D** – Glu107 (chain B) to Lys102 (chain A) in absence of both C1 inhibitors in Tpx corresponding to Figure 7.29 during a 30 ns production run.

Table 7.5: Mean and standard deviation of RMSD values or distances during a 30 ns production run of two C1 in covalent complex with chain A or chain B of wild type Tpx dimer. Mean and standard devations are calculated from 10,000 data points. All values are in Å.

wild type dimer		**chain A** mean / SD	**chain B** mean / SD
RMSD	C1	1.1 / 0.4	1.1 / 0.4
distance	Ile109-N(2)H to C1-O(1)	2.7 / 0.5	2.4 / 0.4
distance	Ile109-O(2) to C1-N(1)H	3.6 / 0.5	3.4 / 0.4
distance	Glu107-O(3) to C1-F(1)	3.5 / 0.4	3.3 / 0.3

To gain further insight into this increase in RMSD intermolecular protein-protein interactions have been analyzed. It was observable that Lys102 of chain A and Glu107 of chain B (and vice versa) can establish hydrogen bonds in the course of simulation (Figure 7.30). In presence of C1 hydrogen bonds are formed between Glu107 (chain A) and Lys102 (chain B) for the first 5 ns and after 16 ns. In the other case, a stable interaction is only maintained for the first 6 ns. In comparison to these observations one obtains interaction for Glu107 (chain A) and Lys102 (chain B) from 9 to 24 ns or for the first 12 ns in case of the interaction of Glu107 (chain B) and Lys102 (chain A). After 24 ns an increase to 18 Å occurs as both chains are moving apart. This motion is in correlation with the course of RMSD values and indicates the dissociation of the dimer. Hence, one can conclude that the stability of Tpx dimer is maintained due to the presence of covalently bound C1 and the resulting drug-drug and drug-protein interactions.

Table 7.6: Mean and standard deviation of intermolecular distances of both C1 molecules during a 30 ns production run in covalent complex with wild type Tpx dimer as depicted in Figure 7.25. Mean and standard devations are calculated from 10,000 data points. All values are in Å.

wild type dimer		**mean / SD**
distance	C1-C(2) to C1-S(2)	3.7 / 0.3
distance	C1-C(3) to C1-C(6)	3.8 / 0.2
distance	C1-C(4) to C1-C(7)	4.1 / 0.5
distance	C1-C(5) to C1-C(8)	3.7 / 0.4

Table 7.7: Mean and standard deviation of RMSD values of chain A or B in presence or absence of C1 during a 30 ns production run. Mean and standard devations are calculated from 10,000 data points. All values are in Å.

wild type dimer	**RMSD**	**chain A**	**chain B**
presence of inhibitor	**mean / SD**	1.8 / 0.2	1.9 / 0.2
absence of inhibitor	**mean / SD**	2.8 / 0.9	3.1 / 1.0

Table 7.8: Mean and standard deviation of intermolecular distances of both Glu107 in chain A and Lys102 in chain B (and vice versa) in presence (*) and absence (**) of two C1 molecules during a 30 ns production run in covalent complex with wild type Tpx dimer as depicted in Figure 7.29. Mean and standard devations are calculated from 10,000 data points. All values are in Å.

wild type dimer		**Glu107 of chain A to Lys102 of chain B*** mean / SD	**Glu107 of chain B to Lys102 of chain A*** mean / SD	**Glu107 of chain A to Lys102 of chain B**** **mean / SD**	**Glu107 of chain B to Lys102 of chain A**** **mean / SD**
distance	Glu107-O4 to Lys102-N(1)	5.9 / 3.3	8.0 / 3.0	7.0 / 4.7	8.2 / 5.3
distance	Glu107-O5 to Lys102-N(1)	6.0 / 3.2	7.9 / 3.0	7.3 / 4.6	8.3 / 5.4

7.8 Conclusion

The focus of this chapter lies on the investigation of tryparedoxin (TPX) which is part of the thiol redox metabolism in *Trypanosoma brucei* and its modification by the covalent inhibitor C1 (Figure 7.2). MD simulations have been conducted to gain insight on the drug-protein interaction pattern, the influence of point mutations on C1, the protonation state of TPX and the dimerization of the enzyme.

First of all, the dynamics of the covalent protein-inhibitor complex of the wild type for two different poses have been investigated which were present in the crystal structure. The mean RMSD value of the inhibitor in pose I is 3.1 Å lower than in pose II (see Table 7.1). C1 in pose II shows a movement towards the active site and interactions with Ile109 backbone oxygen increased. For pose I, two different orientations are observed which are caused by the movement of Trp39. In terms of the interaction between Cys40 and Cys43 both poses show similar mean values in distance. Also a possible interaction with backbone oxygen of Glu107 might be established if different polar substituents are introduced to the phenyl ring of C1.

The MD simulations of the Tpx variants (based on the covalent protein-inhibitor complex of C1 in pose I) revealed that the orientation of the inhibitor can change in two directions depending on the mutant. As mentioned before two different orientations were observed for pose I in wild type Tpx. The C43A and W70A variant revealed a T-shaped stacked conformation of C1 to Trp39 which was observable for most parts of the wild type simulation. In case of W39A that interaction is missing and does not cause a rearrangement of the inhibitor. The Tpx variant I109A offers more space in the active site. Due to this fact it enabled a

motion of C1 into the pocket and in consequence lessened the interaction with Trp39 which orientates towards the water shell. The mean RMSD values of W39A and I109A are lower than in case of the wild type calculations as they do not change their orientation because of above mentioned findings (Table 7.2). For C43A and W70A these values are higher as they solely show the second orientation which occurs due to the interaction of C1 and Trp39. Contrary to that, C1 in pose I wild type shows both conformations leading to a mean RMSD value which is in between. Still, it is to be expected that C1 in C43A and W70A might also change its conformation though it was not observed in these simulations.

The investigations of the non-covalent complex (wild type) revealed that the MD simulation containing a deprotonated Cys40 and a protonated Cys43 show the lowest mean RMSD values. Still, the mean distance between the Cys40-S(1) and C1-C(1) was 7.4 Å in that case. Therefore, one would not conclude a covalent modification as the distance for a nucleophilic attack was too far away. In consequence, one can suggest that the formation of the covalent bond is not determined by the non-covalent complex but rather occurs randomly. In terms of interactions with Ile109 backbone NH, an interaction can be seen for most times. For the other two protonation states C1 shows larger RMSD values and the distances were not reasonable for covalent modification or interaction with Ile109. This is in line with observations of Hellmichs group as no inhibition occurred in absence of chlorine in C1 thus, a weak non-covalent protein-inhibitor complex was expected.

Also, the wild type was investigated in absence of the inhibitor by varying the protonation states of Cys40 and Cys43. These simulations showed that the lowest fluctuations are obtained if both cysteine residues are protonated (see Table 7.4). The distances measured for the deprotonated Cys40-S system are similar to the neutral state but the mean values and standard deviations are slightly higher. The third investigated protonation state Cys40-SH and Cys43-S- shows the largest mean in distance or standard deviation with respect to the Cys40-Cys43 interaction. The movement of deprotonated sulfur of Cys43 towards Tyr80 leads to a destabilization of the interaction and a movement of Trp39 and Cys40 in direction of the water shell. Therefore, it is to be expected, that this protonation state is unlikely. This is in line with the crystal structure of Tpx in covalent complex with C1. Only a modification of Cys40 was observed and simulations showed that a deprotonated Cys43 is unlikely and in consequence, no nucleophilic attack is to be expected at this residue.

Lastly, the stability of the Tpx dimer has been evaluated in presence and in absence of C1. The obtained RMSD values show that the stability of both Tpx chains is decreased if the inhibitor is missing. This goes along with a rise in distance of both chains after 24 ns which indicates the dissociation of the protein dimer. This is in agreement with experimental find-

ings as no dimerization process occurred if inhibitor molecules were absent. Furthermore, simulations show that drug-drug interactions as π-π interactions are maintained in the course of simulation. Additionally, drug-protein interactions with Ile109 appear weaker compared to the monomer but due to the π-stacked conformation of both inhibitors a decrease in RMSD values for both C1 molecules is observable in the course of simulation time. In terms of protein-protein interaction, both chains show occasionally intermolecular hydrogen bonds due to Lys102 and Glu107 side chains which can be preserved for a certain period of time.

Chapter 8

Comparison of Docking and combined classical and quantum mechanical approaches for evaluation of kinetic data

8.1 Introduction: Rhodesain and cruzain– family and pathogenetic

The death of thousands of people in Sub-Saharan Africa is attributable to the infectious disease HAT, as mentioned in the previous chapter. Here, one differentiates between different subspecies of the protozoan parasite. Its transmission to humans occurs due to the tsetse fly, a specific species of the genus Glossina.

The temporal course of symptoms and the life expectancy is strongly related to the subspecies by which the patient is infected [164]. T.b. rhodesiense, which leads to the acute form, can progress to a lethal end during months or even within weeks, due to the infestation of the parasites in the central nervous system (CNS) [165]. The subspecies T.b. gambiense is the cause for the chronical form which can be asymptomatic for months or years [164]. The perfidious thing in this respect is the fact that symptoms occur after infection of the CNS.

In the first stage, infected humans show symptoms such as headache, fever or joint pains. The second stage is more severe as a cerebral infection is developed after invasion of the blood-brain-barrier [164]. The symptoms in the second stage are worse because the CNS is

affected. Tremors of the tounge, hands or feet, severe headaches, mental disorientation and neurological psychiatric manifestations are the consequences. Lastly, infected patients fall into a coma and die because of the infection [166].

Another famous representative of the Trypanosomias family is T. cruzi the cause of Chagas disease. Even after 100 years of its discovery, Chagas disease is still leading to an increased morbidity in Central and South America [167]. Although the distribution of this disease is sterned due to increased measures, still millions of humans remain, who are chronically infected. Especially for patients with human immunodeficiency virus (HIV) Chagas disease is an opportunistic infection [168]. Acute infections usually are asymptomatic or can lead to mild symptoms which still can be accompanied with anemia or thrombocytopenia. After an incubation time of 1-2 weeks patients show non-specific febrile illness [167]. Additionally, symptoms as chills, vomiting, diarrhea or conjunctivitis can occur [169].The insidious circumstance about the infection is its detection: T. *cruzi*-specific antibodies can be negative during acute infection [170]. In the indeterminate state of the Chagas disease chronic complications can developed due to the infection with a probability of 15-30 %. The entry in the chronic stage of this disease can occur over a period of decades which is presented by cardiac or gastrointestinal complications. Heart diseases caused by the chronic Chagas stadium lead to arrhythmias, thromboembolic events or even heart failure [171].

Treatment of Chagas disease is carried out by chemotherapy with benznidazole or nifurtimox, which is under medical supervision due to the increased toxicity and the occurrence of malignant lymphomas in experimental animals [172], [173]. The chance of success in terms of both drugs is 60% in case of acute infection but they are not used for the medication of the chronic form of Chagas [172]. This shows the importance of the development of novel drugs for the medication of Chagas disease.

By dint of the development of the cysteine protease inhibitor (CPI) K777 (also known as K11777) a possible foundation stone was laid for the treatment of parasitic diseases [172]. Engel et. al were able to treat mice in the acute or chronic phase of infection in absence of toxicity by the mammalian host. This CPI was efficacious in preclinical trials in which even immunocompetent and immunodeficient mice and dogs have been treated [174], [175]. Classic CIPs compromise di- or tripeptides in combination with a warhead, which binds covalently (reversible or irreversible) at the active-site cysteine side chain. Bryant et al. developed further possible CIPs (originating from the K11777 scaffold) and evaluated the IC_{50} values of these compounds in cruzain, rhodesain and TbCatB (cathepsin B-like protease in T.brucei) [176]. Additionally, experimental Ki_{app} and k_{inact} values were available in case of rhodesain (Table 8.2).

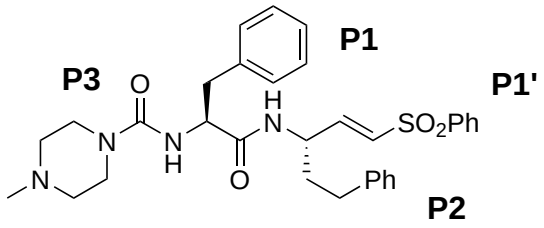
The main focus of the investigation described in this chapter lies in the reproduction and evaluation of kinetic data for K11777 and three selected derivatives (3a, 7a and 7b) *via* computational methodologies (Figure 8.1). The question arises whether it is possible to qualitatively reproduce the k_{inact} values computationally for the above mentioned inhibitors and also to transfer these findings to predict k_{inact} values for cruzain.
Why is this of interest? By establishing a protocol for the prediction of kinetics or generally speaking the suitability of novel CPIs one can beforehand filter out candidates and therefore pick the best compounds.
Firstly, the idea was to calculate the reaction energy and also the height of barrier for the reactions by using various QM/MM methodologies. It is understandable that the reaction rate of the reaction is determined by the energy of the transition state and therefore conclusions can be drawn in respect to k_{inact}. Computational details will be discussed in the following sections.

Table 8.1: Experimental IC_{50} values of the protease inhibition of cruzain and rhodesain by K11777 and different vinylsulfone derivatives measured by [176] are listed.

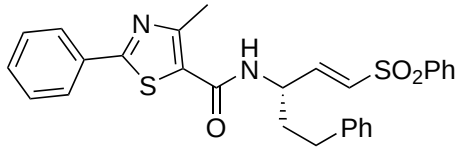
compound	**protease Inhibition IC_{50} [μM]**	**protease Inhibition IC_{50} [μM]**
	Cruzain	**Rhodesain**
K11777	0.004	0.007
3a	10	< 0.006
7a	0.05	0.1
7b	> 100	56

Table 8.2: Experimental inhibition kinetics obtained for different vinylsulfone inhibitors for rhodesain measured by [176].

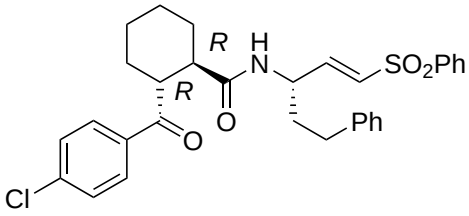
compound	$k_{inact}/K_i(\mathbf{M^1\ s^1})$	$K_{iapp}(\mu\mathbf{M})$	$k_{inact}(s^1)$
K11777	555	0.78	0.029
3a	84	0.32	0.0018
7a	9750	0.80	0.00051
7b	2160	7.9	0.0007



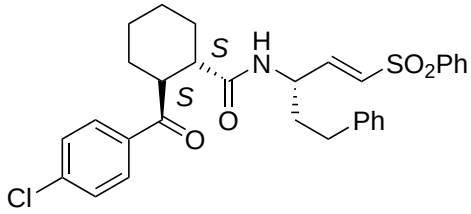
K11777



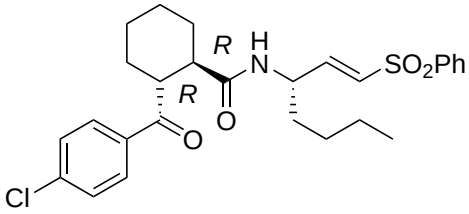
3a



7a



7b



8a

Figure 8.1: Inhibitors of cruzain and rhodesain of the vinylsulfone class. Solely the structures of K11777, 3a, 7a and 7b were investigated. The positions of P1'-P3' are depicted for K11777 for clarification.

8.2 System preparation and MD simulation

The first step of the investigation consists in the conduction of MD simulations to obtain relaxed structures which serve as starting points for QM/MM calculations. For K11777 in complex with cruzain the pdb structure 2OZ2 was used. As no crystal structure was available for 3a, 7a or 7b in complex with the proteins a manual modification was needed. Therefore, another crystal structure containing inhibitor 8a has been manually modified to respective ligands (pdb code 3HD3). Especially in case of 7a and 7b it was advantageous to use this crystal structure due to the similarity of 8a's structure to both (Figure 8.1). The protein-inhibitor complex of K11777 and rhodesain was prepared by employing the existing

crystal structure (pdb code 2PZU). The remaining inhibitors have been manually created by modification of K11777 as no crystal structure was available.
The proteins cruzain and rhodesain were parameterized using the AMBER14 force field ff14SB while the inhibitor K11777, 3a, 7a and 7b by the generalized AMBER force field (GAFF). For all cases a covalent protein-inhibitor complex has been prepared to calculate the backwards reaction to the non-covalent complex. The duration of each production run contained 10 ns resulting in a total simulation time of 80 ns. As the interest lies on the creation of relaxed structures as starting points for QM/MM calculations no detailed investigations have been conducted with respect to the MD simulations.

8.3 QM/MM calculations

In the first place, snapshots have been extracted of the MD simulations preparation runs mentioned in section 8.2 which serve as starting point for QM/MM calculations. The preparation of these computations will be explained in the following. As shown in Figure 8.2 the QM area consists of the ligand as well as cysteine and histidine. The importance of both residues is given by the fact that the nucleophilic sulfur of cysteine form a covalent bond to the ligands and histidine acts as proton acceptor (in case of the backwards reaction). Ensuring that the computations are still feasible within a reasonable time only the side chain has been included into the QM part. In regard to the inhibitors only QM atoms were selected which are available in all inhibitors to ensure comparability of the results.
Firstly, a local optimization has been conducted by the usage of the described QM/MM scheme. The software package ChemShell 3.6.0 was utilized applying electrostatic embedding and the charge shift method. The MM part was treated by the AMBER14 force field while the QM part was optimized employing B3LYP in combination with the def-TZVP triple- basis set. Due to computational costs, only molecules in a radius of 10 Å around the active site cysteine residue were optimized, while the rest of the system was kept fix.

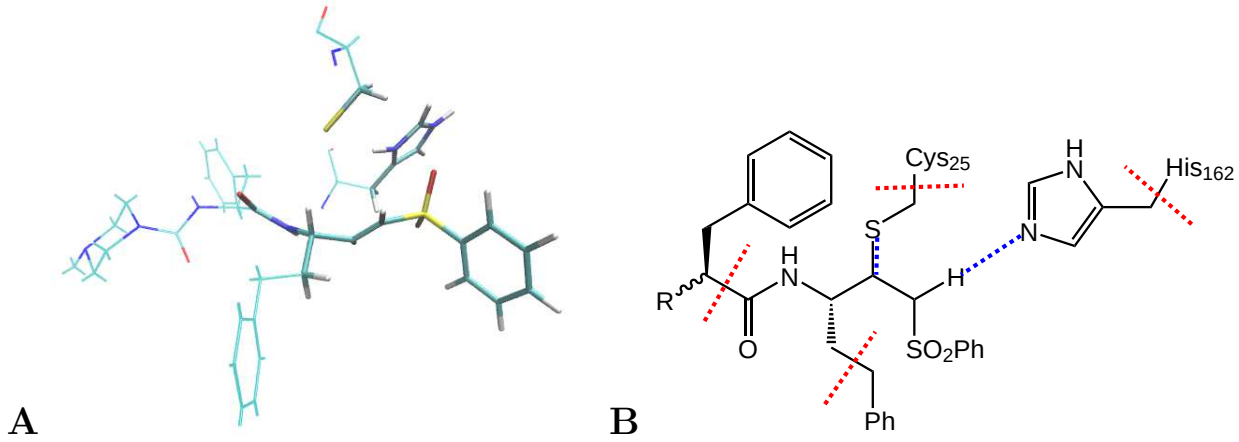


Figure 8.2: Selected QM area for QM/MM optimization. **A** – QM part of His162 and inhibitor K11777 covalently bound to Cys25 in licorice representation. **B** – Selected QM area, red dashed lines indicate C-C bond cuts. Blue dashed lines indicate variation of distance for 2-dimensional relaxed scan by increasing the distance of Cys25's sulfur and K11777's electrophilic carbon and decreasing the distance between His162 nitrogen and K11777's hydrogen.

To obtain the reaction energy and the height of barrier (of the backward reaction) the distance between Cys25's sulfur and the ligands electrophilic carbon has been widened while the distance between His162's nitrogen and the ligands hydrogen was decreased which is depicted in Figure 8.2-B by highlighting the related distances with blue dashed lines. As all resulting potential energy surfaces show a similar profile. Therefore, the pathway only for the K11777-cruzain complex is exemplary shown (Figure 8.3-A). The formation of the covalent bond is computed by decreasing the distance of the nucleophilic sulfur related to the cysteine and the electrophilic carbon of K11777. Afterwards a proton transfer occurs from the doubly protonated histidine to K11777 (Figure 8.3-B).

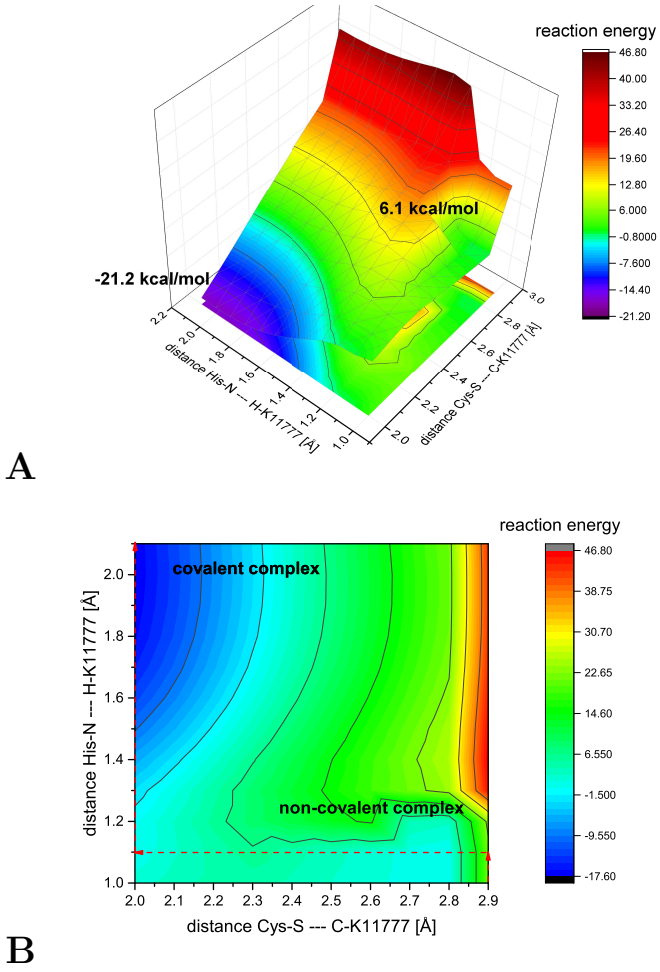


Figure 8.3: **A** - Potential energy surface obtained from 2-dimensional relaxed scan of K11777 in cruzain from covalent to non-covalent protein inhibitor complex. Variation of distance of selected atoms is depicted in Figure 8.2-B. **B** – Reaction coordinate of nucleophilic attack of K11777 indicated by red dotted arrows. Energies are obtained by QM/MM optimization employing B3LYP-D3/def-TZVP with Grimme's dispersion correction and the AMBER14 force field. Energies are in relation to the non-covalent protein-inhibitor complex of the relaxed scan. All energy values are in kcal/mol.

The summary of obtained reaction energies and height of barriers is given in Table 8.3. The energy of the transition state in relation to the non-covalent protein-inhibitor complex of 7a is lowest by a value of 1.6 kcal/mol, followed by 3a (3.3 kcal/mol) and K11777 (5.8 kcal/mol). For inhibitor 7b the largest barrier is predicted which is of 6.6 kcal/mol. As one can see the results are contrary to experimental findings with respect to rhodesain. As shown in Table 8.2, K11777 has the largest k_{inact} value which consequently should lead to the smallest height of the barrier which is not the case. Additionally, 7a and 7b differ by 5

Table 8.3: Reaction energies and transition state energies of K11777 and its derivatives 3a, 7a and 7b obtained from QM/MM optimization employing B3LYP-D3/def-TZVP and the AMBER14 force field are summarized. Energies are in relation to the non-covalent protein-inhibitor complex corresponding to Figure 8.3. All energy values are given in kcal/mol.

protein	**compound**	**transition state**	**reaction energy**
cruzain	K11777	6.1	-21.2
	3a	0.6	-37.8
	7a	1.7	-34.4
	7b	7.5	-27.6
rhodesain	K11777	5.8	-31.6
	3a	3.3	-33.1
	7a	1.6	-38.0
	7b	6.6	-30.0

kcal/mol and 7a's barrier is predicted as to be the lowest in energy. These findings conflict with the experimental data. The trend of the kinetic data could not be reproduced by these computations. Therefore, singlepoint calculations applying different DFT or ab initio methods have been conducted to shed some further light on these findings.

8.4 QM benchmark calculations

In this section, benchmark calculations applying different wavefunction and density functional based methods were conducted and referenced to CCSD(T) energies. The calculations initially started on QM level in absence of the protein environment. Due to the size of the QM region previously defined within the QM/MM calculations, computational costs using CCSD(T) would be too high. Therefore, solely the reaction energy of the QM region of the vinylsulfone-inhibitor scaffold and methanethiol (representative for cysteine) has been computed to evaluate the best methodology (Figure 8.4). It must be mentioned that the energetically description of the transition state is not considered in these computations.

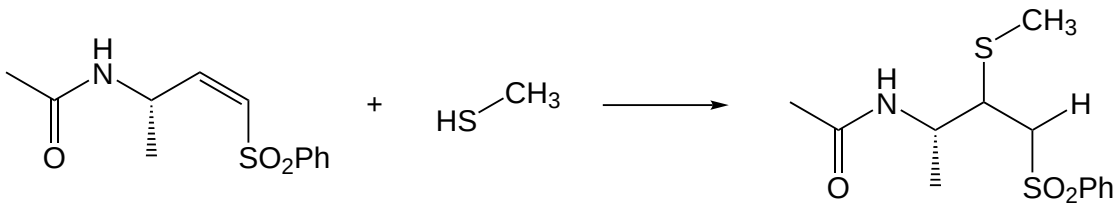


Figure 8.4: Reactans and product of the reaction of methanethiol with the vinylsulfone scaffold.

Firstly, the vinylsulfone scaffold, methanethiol and the resulting product has been optimized with B3LYP-D/def-TZVP. Two cases have been considered: the optimization in vacuo as well as an implicit solvent by using the COSMO approach, with water parameters (ϵ=78) simulating a hydrophilic environment. Afterwards singlepoint energy calculations were computed for both cases. As no COSMO implementation is available for the couple cluster approach in the Turbomole V6.6 program package solely singlepoint energy computations of CCSD(T) in vacuo were used for comparison.
As one can see in Table 8.4 best results are achieved by SCS-CC2 and are underpinned by the smallest MAE values of 0.1 and 0.2 kcal/mol (Figure 8.5). For ab initio methods the application of spin-component scaling leads to an improvement of results. In case of the DFT approaches BP86 performs best and lead to a comparable result as SCS-CC2 (and a better one than SCS-MP2) which might appear due to error compensation. Especially the tested hybrid GGA methods show larger MAE values compared to PBE or BP86 while they only outperform LDA. Figure 8.5 shows that MAE values compared to CCSD(T)/aug-cc-pVDZ single point energies, do not differ largely despite of the presence or absence of implicit solvation within the structure optimization procedure. In contrast to that it is observable that a decrease in reaction energy of 5-6 kcal/mol occurs by applying COSMO (if the same settings for structure optimization are utilized). Lastly, it is of importance to have a closer look on the D1 diagnostic of all computations on CCSD(T)/aug-cc-pVDZ level to ensure that no multi-reference character within the wavefunctions description plays a role. As shown in Table 8.5 all values of the D1 diagnostic are below 0.065 and are in an acceptable range.

Table 8.4: QM reaction energies of vinylsulfone scaffold with methanethiol. For all methods the aug-cc-pVDZ basis set was applied. For * marked reaction eneriges ΔE_{reac} singlepoint energy calculation were performed on vacuo optimized structures or for ** marked on COSMO (ϵ=78) optimized structures. Grimme's dispersion correction was applied for B3LYP and BHLYP. All energy values in kcal/mol and with to the non-covalent complex.

method	* ΔE_{reac} (**vacuo**)	* ΔE_{reac} (**COSMO**)	** ΔE_{reac} (**vacuo**)	** ΔE_{reac} (**COSMO**)
CCSD(T)	-32.5	NA	-32.8	NA
MP2	-36.0	-30.8	-36.3	-30.6
CC2	-37.7	-32.7	-38.1	-32.5
SCS-MP2	-31.4	-26.3	-31.7	-26.1
SCS-CC2	-32.3	-27.3	-32.7	-27.1
LDA	-39.3	-34.0	-39.4	-33.5
PBE	-30.0	-25.0	-30.5	-24.8
BP86	-31.8	-26.7	-32.4	-26.7
B3LYP	-27.0	-21.8	-27.4	-27.4
BHLYP	-28,1	-22.6	-28.4	-28.4

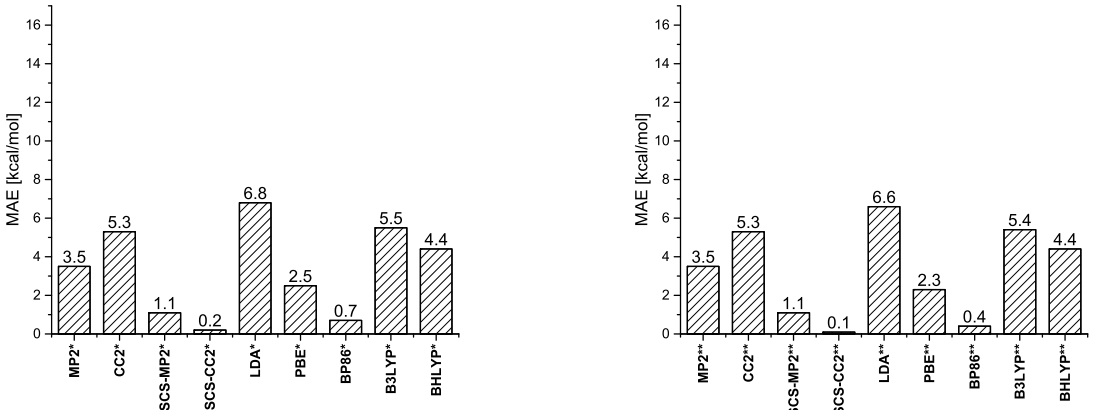


Figure 8.5: Mean average error with respect to CCSD(T)/aug-cc-pVDZ reaction energy. Singlepoint energy calculation were performed on vacuo optimized structures for * marked reaction ΔE_{reac} or on COSMO (ϵ = 78) optimized structures for ** marked. Grimme's dispersion correction was applied for B3LYP, B3LYP_G and BHLYP. All MAE values in kcal/mol.

Table 8.5: Summary of CCSD D_1 diagnostics of educt, methanethiol and product of nucleophilic attack of methanethiol and vinylsulfone scaffold. All calculations were performed applying the CCSD(T)/aug-cc-pDVZ approach.

Optimization in vacuum	**CCSD D_1 diagnostics**	**Optimization with COSMO (ϵ=78)**	**CCSD D_1 diagnostic**
educt	0.0585	educt	0.0647
methanethiol	0.0211	methanethiol	0.0212
product	0.0608	product	0.0612

8.5 QM/MM benchmark calculations

QM benchmark calculations showed that spin-component scaled CC2 is the most suitable methodology in terms of the description of the reaction energy in comparison with CCSD(T). As mentioned before QM/MM singlepoint computations applying the 2nd order coupled cluster approximated triples approach is too costly and therefore SCS-CC2 has been used as reference for the following QM/MM benchmark. In the previous section it was not possible to reproduce the experimental k_{inact} values qualitatively by the QM/MM computations on B3LYP-D3 level. Therefore, different DFT and ab initio approaches have been applied to the QM/MM optimized structures of the transition state as well as on the non-covalent protein-inhibitor complex. Additionally, the augmented triple-ζ basis aug-cc-pVDZ was utilized instead of the def-TZVP basis set, while the QM region and the force field description remained the same as it was for geometry optimization.

The results of the benchmark given in Table 8.6 show a similar trend in prediction of the kinetics for the tested inhibitors in rhodesain as obtained with B3LYP-D3/def-TZVP. All ab initio methods describe the reaction barrier for K11777 as highest in energy followed by inhibitor 7b while the transition states of 3a and 7a are about 3-4 kcal/mol lower in energy. In case of the DFT approaches barriers were also predicted in the wrong order. Especially computations on LDA, PBE and BP86 level resulted in transition state energies which are lower compared to the non-covalent complex. Therefore, it is not possible to make a statement concerning the order of the kinetics in terms of the tested inhibitors and transfer these findings to cruzain.

In terms of the MAE values, DFT approaches show an error which is of about 2-3 kcal/mol larger compared to the wave function-based methods. Interestingly, it is observable that the MAE values of B3LYP are lower if the def-TZVP basis set is used instead of the aug-cc-pVDZ. Although the order of kinetics could not be reproduced by the application of higher

ranked methods and basis sets one can see that ab initio methods are with respect to MAE values superior compared to the tested DFT approaches.

Table 8.6: QM/MM reaction energies and transition state energies of K11777 and its derivatives 3a, 7a and 7b obtained from single point calculations employing wave function-based and DFT methodologies in rhodesain. For all tested methods the aug-cc-pVDZ basis set was applied. In case of the * marked method energies were obtained from QM/MM geometry optimization employing B3LYP-D3/def-TZVP and the AMBER14 force field. All energy values are in kcal/mol and in relation to the non-covalent complex.

compound	**K11777**		**3a**		**7a**		**7b**	
method	**TS**	ΔE_{reac}	**TS**	ΔE_{reac}	**TS**	ΔE_{reac}	**TS**	ΔE_{reac}
MP2	6.2	-38.1	2.6	-40.2	1.8	-44.3	5.1	-36.5
CC2	3.2	-37.6	0.7	-39.0	-0.6	-43.6	2.7	-35.8
SCS-MP2	8.1	-37.7	3.6	-40.6	3.7	-43.4	7.4	-36.1
SCS-CC2	5.9	-37.1	2.2	-39.5	1.9	-42.6	5.8	-35.3
LDA	1.5	-31.2	0.4	-31.7	-3.1	-39.1	-2.3	-31.6
PBE	2.6	-31.3	1.1	-32.2	-1.5	-37.6	3.8	-29.0
BP86	1.6	-31.3	0.6	-31.9	-2.1	-37.4	3.1	-28.6
B3LYP	5.2	-31.6	2.5	-33.4	1.3	-37.0	6.3	-29.4
BHLYP	11.2	-32.6	6.0	-36.0	7.0	-38.2	7.9	-33.7
B3LYP*	5.8	-31.6	3.3	-33.1	1.6	-38.0	6.6	-30.0

Table 8.7: QM/MM reaction energies and transition state energies of K11777 and its derivatives 3a, 7a and 7b obtained from single point calculations employing wave function-based and DFT methodologies in cruzain. For all tested methods the aug-cc-pVDZ basis set was applied. In case of the * marked method energies were obtained from QM/MM geometry optimization employing B3LYP-D3/def-TZVP and the AMBER14 force field. All energy values are in kcal/mol and in relation to the non-covalent complex.

compound	**K11777**		**3a**		**7a**		**7b**	
method	**TS**	ΔE_{reac}	**TS**	ΔE_{reac}	**TS**	ΔE_{reac}	**TS**	ΔE_{reac}
MP2	3.8	-27.9	0.9	-43.5	5.5	-40.9	5.7	-34.2
CC2	1.2	-27.7	0.4	-42.3	-0.9	-40.8	3.1	-33.6
SCS-MP2	6.6	-26.9	-0.8	-43.6	-4.0	-44.6	8.0	-33.8
SCS-CC2	4.7	-26.6	-0.6	-42.2	-1.8	-41.6	6.1	-33.1
LDA	-1.5	-31.2	1.5	-36.4	-1.5	-34.2	2.2	-27.0
PBE	1.7	-22.2	1.3	-34.6	-0.2	-32.9	4.1	-27.1
BP86	1.0	-22.2	1.0	-34.8	-1.1	-32.9	3.4	-26.7
B3LYP	5.5	-21.0	0.4	-31.8	1.6	-33.8	6.7	-27.4
BHLYP	12.0	-20.4	0.6	-38.7	7.2	-35.2	12.3	-29.6
B3LYP*	6.1	-21.2	0.6	-37.8	1.7	-34.4	7.5	-27.6

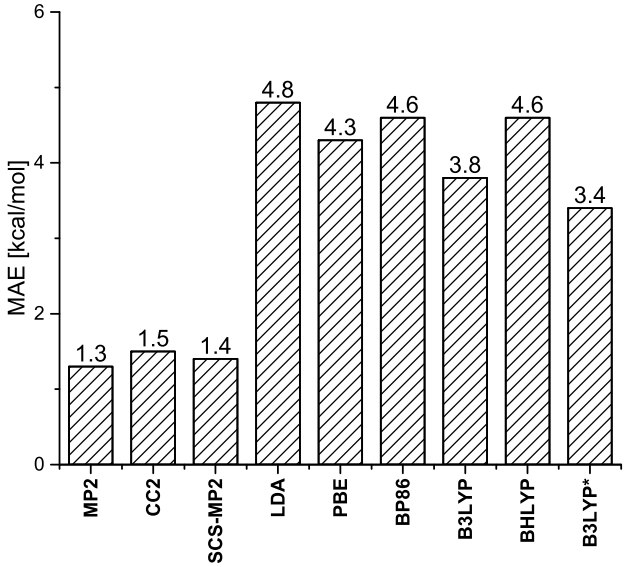


Figure 8.6: Mean average error with respect to SCS-CC2/aug-cc-pVDZ averaged of all reaction energies and reaction barrier of all investigated inhibitors in rhodesain and cruzain. For all tested methods the aug-cc-pVDZ basis set was applied. In case of the * marked method, energies were obtained from QM/MM geometry optimization employing B3LYP-D3/def-TZVP and the AMBER14 force field. All MAE values in kcal/mol.

8.6 MD Simulation of non-covalent protein-inhibitor complex after relaxed scan

As the computations of the backward reaction were not able to reproduce the right order of the experimental kinetic data a different approach was tested in this part. The non-covalent complex obtained from the relaxed scan has been used as starting point for a classical MD simulation. Aim of these simulations is to measure the distance of the nucleophilic sulfur of the cysteine residue and the electrophilic carbon of the vinylsulfone inhibitors in the course of simulation time. It is understandable that the height of barrier for a nucleophilic attack should increase at the same time as the distance although it has to be mentioned that observations are of course qualitatively. To obtain meaningful results each inhibitor-protein complex has been simulated thrice and for 10 ns.

Table 8.8: Calculated mean and standard deviations of the distance between electrophilic carbon of respective inhibitor and nucleophilic sulfur of Cys25 in cruzain and rhodesain during a 10 ns MD simulation are given. Mean and standard deviations are calculated from 10,000 data points. All values are presented in Å.

cruzain								
compound	**K11777**		**3a**		**7a**		**7b**	
	mean	SD	mean	SD	mean	SD	mean	SD
MD1	3.5	0.2	5.6	1.9	3.5	0.2	3.6	0.3
MD2	3.5	0.2	3.8	0.6	3.5	0.2	3.5	0.3
MD3	3.5	0.2	4.8	0.7	4.2	0.7	3.5	0.2
rhodesain								
compound	K11777		3a		7a		7b	
	mean	SD	mean	SD	mean	SD	mean	SD
MD1	3.6	0.3	6.1	2.9	3.7	0.4	3.5	0.3
MD2	3.4	0.2	3.6	0.2	3.8	0.4	3.4	0.2
MD3	3.4	0.2	10.1	2.2	4.1	0.5	3.5	0.3

In contrast to the prediction in the QM/MM calculations MD simulations indicate that K11777 and 7b would result in the lowest barrier as the mean distance of both inhibitors (in both proteins) are about 3.5 Å with a standard deviation of 0.2-0.3 Å. Slightly larger mean values are obtained for 7a while 3a appears to be more unstable in terms of the distance. Therefore, mean distances of 6.1 Å or 10.1 Å occur with larger standard deviations of about 2-3 Å. One might conclude that K11777 and 7b would react at the same speed due to the

same mean distance and thus the same height of barrier. Also, the nucleophilic reaction of 7a would be slightly slower, while 3a kinetics would be worse. These findings still do not fit to the experimental data and a different solution is needed.

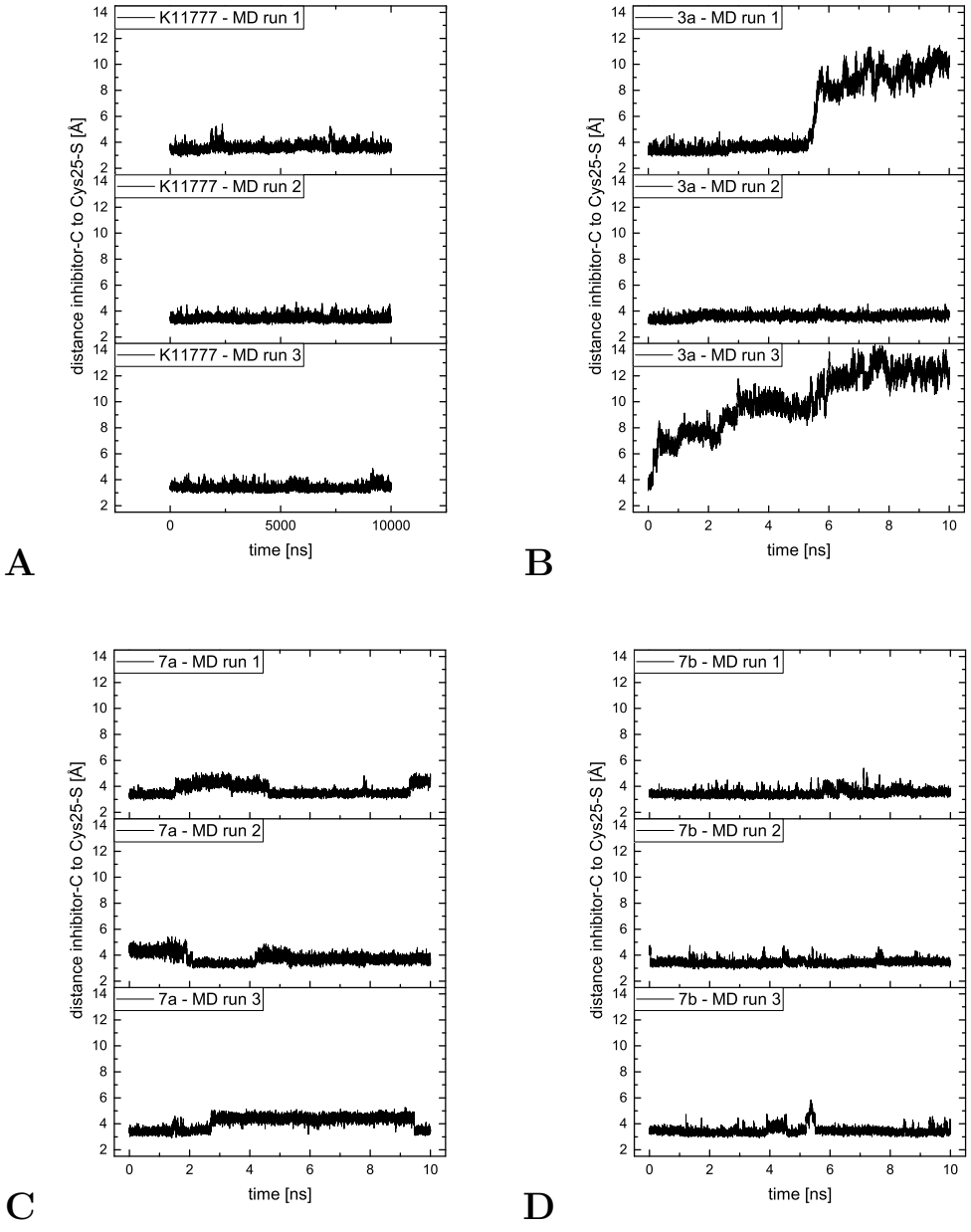


Figure 8.7: The variation in distance as a function of simulation time is shown for **A** – K11777, **B** – 3a, **C** – 7a and **D** – 7b to Cys25's sulfur in non-covalent complex with rhodesain during 10 ns production run. Calculated mean and standard deviation are summarized in Table 8.8.

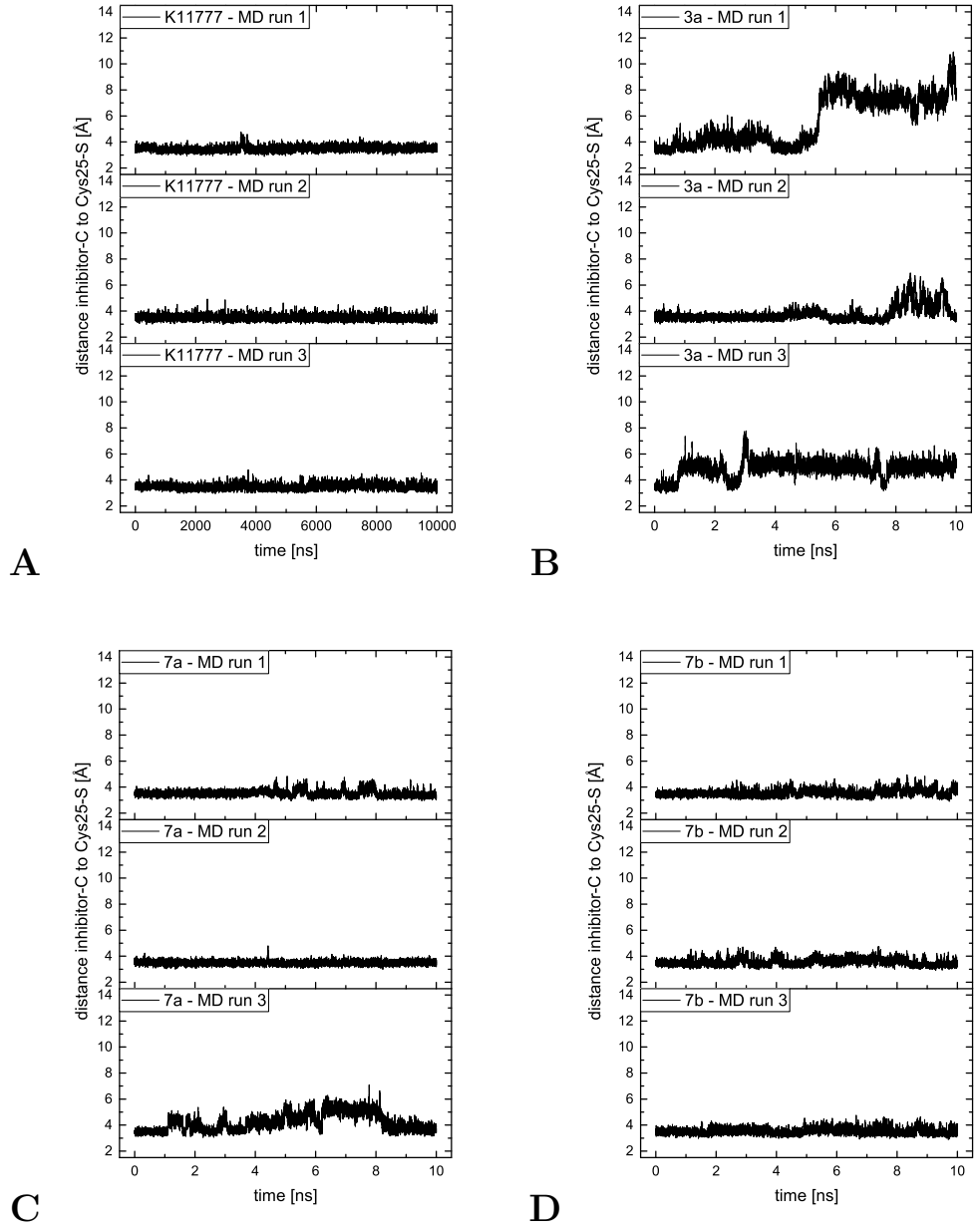


Figure 8.8: The variation in distance as a function of simulation time is shown for **A** – K11777, **B** – 3a, **C** – 7a and **D** – 7b to Cys25's sulfur in non-covalent complex with cruzain during 10 ns production run. Calculated mean and standard deviation are summarized in Table 8.8.

8.7 Docking of K11777 and 7a in rhodesain

Up to this point all computations were not able to predict the right order of experimental rates. As a result, a different strategy is needed. Firstly, QM/MM computations starting from the covalent protein-inhibitor complex were conducted. Therefore, it is conceivable that the protein environment is relaxed to the covalent state and overestimates its stability. Secondly, MD simulations of the non-covalent complex utilized the end point of the relaxed scan as initial structure. It is possible that the original non-covalent complex is differently in terms of conformation and position. Therefore, the idea is to start from the non-covalent complex by using docking approaches. Additionally, the focus of investigated inhibitors was laid on K11777 and 7a for a specific reason. If one looks at the experimental kinetic data in Table 8.2 not only differences in k_{inact} values can be seen but also different K_{iapp} values. Solely K11777 and 7a have nearly the same Ki_{app} value and thus computations should be more comparable. To obtain a proof of concept the focus was brought on rhodesain.
The evaluation of the docking process was realized by redocking of K11777 in rhodesain in cooperation with the Sotriffer group, University of Würzburg, utilizing the GOLDv5.2.2 software package. The typical preparation starts with the removal of all crystal water and the assignment of hydrogen atoms at pH=6.0, which was reported for the crystal structure (pdb code 2OZ2) employing the MOE software package [177]. In case of Cys25 a deprotonated form was considered while His162 was doubly protonated. Before the docking procedure was performed both inhibitors were protonated by the MOE software package and afterwards minimized with a convergence criterion of 0.05 kcal/mol. In order to receive a sufficient amount of poses the number of genetic algorithms runs was increased to the number of 50. GA settings have additionally been set to be very flexible and no constraints were applied. The variable was the specification of the approximate radius r of the binding site. Cys25's sulfur has been selected as center of the active site and four different values for r were assigned (10, 15, 20 and 30 Å). The definition of r is crucial as it should be big enough to enable complementary fitting points between protein and ligand. But by selecting too large radii unusable information are included in the docking process due to the fact that more atoms are defined as active site. This can lead to worse docked poses. The scoring of obtained poses was conducted with the CHEMPLP fitness function and the best 10 poses were rescored by DSX with CSD pair-potential utilizing the python script of Dr. Michael Hein and Maximilian Kuhn [178]. Afterwards RMSD values of the non-covalent inhibitor were calculated in comparison to the covalent one from the crystal structure.
The results of the redocking in Table 8.9 show that the choice of a r-value of 20 Å leads

to seven poses with RMSD values below 3 Å while three of them are ranked in the top 3 poses. Slightly worse results are obtained if a r-value of 10 Å is used although it should be mentioned that one pose with this setting lead to a pose with a RMSD value below 2 Å. In case of r=15 Å or 30 Å only poses were obtained in which K11777 was flipped thus leading to RMSD values larger than 5 Å. Therefore, one can conclude that best results are obtained by the usage of r=20 Å.
If one takes a closer look at the top ten poses one can see that there are six different ones (Figure 8.9). Flipped structures occur in conformation II, IV and V while I, III and VI are in the right position for a nucleophilic attack. For these flipped structures the distance between the sulfur of Cys25 and the electrophilic carbon is larger than 5 Å so that a reaction is not possible.

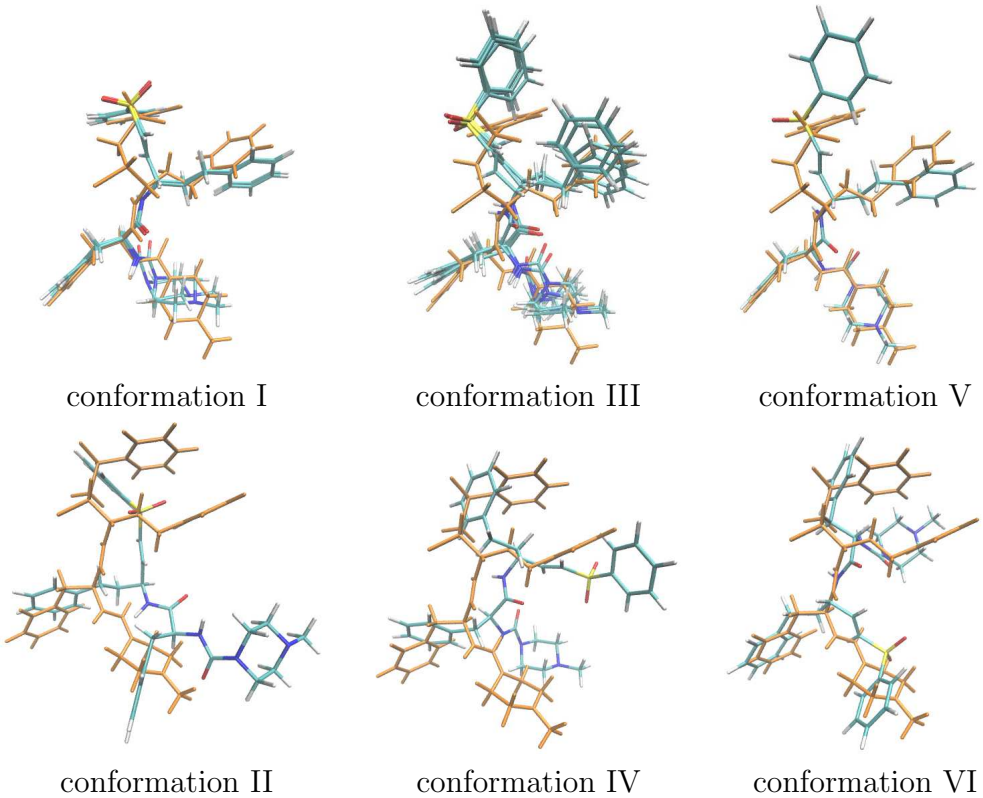


Figure 8.9: Comparison of different conformations of K11777 obtained from redocking compared to X-ray structure of K11777 in covalent complex with rhodesain (orange licorice representation). Poses refer to Table 8.10.

The same setting has been utilized for the docking of inhibitor 7a as the application of a r-value of 20 Å leads to the best results. As shown in Figure 8.10 three different poses were

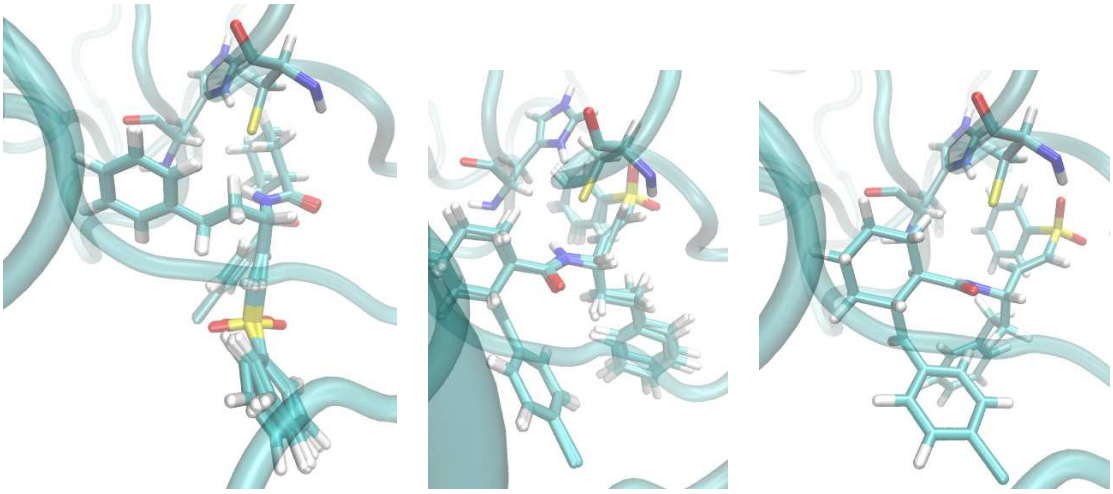
Table 8.9: Docking results of K11777 in rhodesain by varying r to 10 Å, 15 Å, 20 Å and 30 Å. The poses were scored by CHEMPLP and rescored by DSX. Order of conformations on the basis of DSX ranking. Negative DSX values are shown for clarity. Distance is measured between electrophilic carbon center of warhead and Cys25 sulfur. RMSD is calculated in relation to crystal structure of K11777 in covalent complex with rhodesain. Distance is given in Å.

r		**10 Å**				**15 Å**		
Ranking	**distance**	**RMSD**	**-DSX**	**PLP**	**distance**	**RMSD**	**-DSX**	**PLP**
1	6.4	10.9	140.893	78.12	5.7	10.8	130.455	77.9
2	3.4	1.9	140.194	78.02	6.5	5.1	129.129	73.75
3	3.4	2.9	137.149	75.89	7.1	10.8	128.749	72.01
4	4.9	10.0	132.86	77.06	5.9	10.1	127.432	72.67
5	5.6	7.1	129.754	77.8	7.2	10.5	125.867	73.29
6	3.2	2.6	129.376	76.69	6.1	9.6	123.267	72.69
7	4.7	6.7	125.634	78.25	6.0	9.6	122.309	74.57
8	3.7	6.3	123.95	78.11	10.2	11.5	121.116	71.9
9	4.1	2.3	120.297	78.15	8.0	5.9	119.41	72.13
10	4.1	3.2	119.64	75.72	7.7	4.9	118.231	73.71

r		**20 Å**				30 Å		
Ranking	**distance**	**RMSD**	**-DSX**	**PLP**	**distance**	**RMSD**	**-DSX**	**PLP**
1	3.2	2.7	139.898	77.3	6.9	11.0	140.149	75.28
2	3.4	2.0	138.298	84.64	5.5	7.0	135.426	81.61
3	3.4	2.0	138.178	82.79	5.5	6.7	131.762	73.58
4	6.2	7.0	134.109	79	5.7	7.0	130.681	74.78
5	3.8	2.5	132.747	79.53	6.2	10.1	130.028	74.05
6	3.1	2.7	129.126	78.13	5.7	6.8	126.79	75.98
7	5.1	10.1	128.359	77.31	7.1	10.8	124.754	73.89
8	5.3	6.6	128.083	79.64	5.0	6.8	124.68	74.05
9	4.0	2.4	125.3	77.57	11.5	10.1	113.024	75.53
10	4.1	2.5	125.285	77.58	6.8	4.5	112.766	74.56

obtained by the docking procedure. As no crystal structure of 7a in complex with rhodesain was available solely the distance between Cys25 nucleophilic sulfur and 7a's electrophilic carbon was calculated. The top four poses (conformation I) showed a distance of about 5 Å while the remaining were closer for a nucleophilic attack (conformation II and III, see Table 8.10). This indicates that 7a appears to be farther away for a nucleophilic attack and in consequence should react slower compared to K11777. To evaluate the findings of the

docking procedure, MM-PB(GB)SA calculations have been performed and will be discussed in the next section.



conformation I conformation II conformation III

Figure 8.10: Different conformations obtained from docking of inhibitor 7a in rhodesain. Poses refer to Table 8.10.

Table 8.10: Docking results of K11777 and 7a in rhodesain by using r value of 20 Å. The poses were scored by CHEMPLP and rescored by DSX. Order of poses are on the basis of DSX ranking. Negative DSX values are shown for clarity. Distance is measured between electrophilic carbon of inhibitor and Cys25's sulfur. Distance is given in Å.

inhibitor		**K11777**				**7a**		
Ranking	**distance**	**-DSX**	**PLP**	**confor-mation**	**distance**	**-DSX**	**PLP**	**confor-mation**
1	3.2	139.898	77.3	VI	5.1	130.666	74.05	I
2	3.4	138.298	84.64	I	5.1	128.787	73.82	I
3	3.4	138.178	82.79	I	5.1	128.363	73.77	I
4	6.2	134.109	79.00	IV	5.1	128.171	73.35	I
5	3.8	132.747	79.53	III	3.1	128.142	74.91	II
6	3.1	129.126	78.13	III	5.1	128.141	73.11	I
7	5.1	128.359	77.31	V	3.1	127.238	75.87	II
8	5.3	128.083	79.64	II	3.5	126.643	74.89	III
9	4.0	125.3	77.57	III	5.1	126.633	73.13	I
10	4.1	125.285	77.58	III	3.3	126.189	75.19	II

8.8 MM-PB(GB)SA calculations

In order to verify the differences between the obtained conformations of K11777 and 7a, investigations in terms of the free binding energy (ΔG_{bind}) have been performed by utilizing the MM-PB(GB)SA approach which is implemented in the AMBER14 program package. Solely the highest ranked pose of each conformation was considered. Firstly, the gas phase topology and coordinates were created for the inhibitors, the protein and the protein-inhibitor complex for the MM-PBSA calculation as well as the solvated protein-inhibitor complex for the MD simulations. The protein was described with the AMBER14 force field while the inhibitor was parametrized *via* GAFF. Afterwards the solvated complex was equilibrated by a short minimization containing 1000 steps, 50 ps of heating and 50 ps of density equilibration while restraining the protein-inhibitor complex by a value of 2 kcal·mol^{-1}·Å^{-2}. An equilibration at constant pressure was conducted at 300K for 2 ns. The production run was performed at the same conditions as the equilibration for 5 ns while each 10 ps the coordinates were recorded to obtain uncorrelated structures. The estimation of the binding free enthalpy was conducted by the calculation of the interaction and solvation energy for the inhibitors, the protein and the protein-inhibitor complex. It has to be mentioned that the results do not reflect the true free energy rather the free binding enthalpy, as the entropy contribution was not considered. Still, these values should be comparable as long as similar systems are compared (in this case different docking poses of the same system).
Table 8.11 shows the calculated free binding enthalpies for all six poses of K11777 obtained from the previous docking simulation. The flipped conformations II, IV and V which resulted in larger RMSD values than 6 Å in the docking also showed more unfavorable ΔH_{bind} values as these conformations range from -11.7 to -23.0 kcal/mol for the generalized Born model or from 1.0 to 10.5 kcal/mol if the Poisson-Boltzman equation was employed. In contrast to that, conformation I, III and VI were described as more stable with free binding enthalpies of -36.7 to -42.3 kcal/mol while conformation III was lowest in energy for MM-PBSA or MM-GBSA. These results strengthen the validity of applied calculations as docked structures with low RMSD values also show more exothermic free binding enthalpies. If the prediction would be the other way around flipped structures would be preferred which cannot react covalently and this would be in contradiction to experimental findings.

The corresponding calculations have also been conducted for the three conformations of inhibitor 7a which were predicted by the docking procedure. Contrary to the results of the rescoring, conformation II was described lowest in free binding enthalpy despite of the used methodology. Conformation I which was scored in the top four poses is highest in free binding

Table 8.11: Calculated free binding enthalpies, standard deviations and standard error of mean of K11777 in rhodesain employing the generalized Born model or Poisson-Boltzmann equation. Conformations refer to docking procedure. All values in kcal/mol.

	conformation I			**conformation II**		
	average ΔH_{bind}	std. dev	std. err. of mean	average ΔH_{bind}	std. dev	std. err. of mean
GB	-36.7	3.4	0.5	-11.7	3.1	0.4
PB	-2.4	4.2	0.6	10.5	3.5	0.5

	conformation III			**conformation IV**		
	average ΔH_{bind}	std. dev	std. err. of mean	average ΔH_{bind}	std. dev	std. err. of mean
GB	-42.3	2.8	0.4	-23.0	2.3	0.3
PB	-8.2	4.8	0.7	1.0	2.5	0.3

	conformation V			**conformation VI**		
	average ΔH_{bind}	std. dev	std. err. of mean	average ΔH_{bind}	std. dev	std. err. of mean
GB	-15.9	2.7	0.4	-40.0	3.2	0.4
PB	7.5	3.7	0.5	-6.7	4.1	0.6

enthaply values regarding the applied generalized Born model as pose III and also about 10 kcal/mol higher in relation to conformation II concerning the ΔH_{bind} value if the Poisson Boltzmann equation is utilized.

The MM-PBSA and MM-GBSA calculations of K11777 and 7a showed that the ΔH_{bind} value of the flipped conformations are always less exothermic. In case of K11777 these results are in line with the docking simulations as VI and I are best ranked. For 7a the flipped conformation I was scored as best pose. However, MM-PBSA and MM-GBSA calculations showed that this conformation is less exothermic as the conformation I which is lower in distance for a nucleophilic attack. Up to this point the question about rates remained unanswered as 7a's conformation II shows a distance between nucleophilic sulfur and electrophilic carbon of 3.1 Å which is 0.7 Å closer than the pose used for conformation III of K11777.

Table 8.12: Calculated free binding enthalpies of 7a in rhodesain employing the generalized Born model or Poisson-Boltzmann equation. Conformations refer to docking procedure. All values in kcal/mol.

	conformation I			**conformation II**		
	average ΔH_{bind}	std. dev	std. err. of mean	average ΔH_{bind}	std. dev	std. err. of mean
GB	-17.5	2.0	0.3	-33.0	2.5	0.4
PB	12.8	4.9	0.7	2.3	3.5	0.5

	conformation III		
	average ΔH_{bind}	std. dev	std. err. of mean
GB	-24.4	2.2	0.3
PB	11.9	4.5	0.6

8.9 MD Simulation of docked structures

In consequence of the MM-PB(GB)SA calculations, MD simulations of the conformation with the lowest ΔH_{bind} value have been conducted for both inhibitors to evaluate fluctuations in distance between Cys25's sulfur and the electrophilic carbon of the vinylsulfone inhibitor. The same poses employed in the MM-PB(GB)SA calculations have been utilized as starting point for the simulations. The sytem preparation as well as minimization, heating and production run have been conducted in the same manner as in the section 5.1. Each MD simulations were performed thrice with a duration of 10 ns.

As mentioned before, the distance between the electrophilic carbon of the vinylsulfone inhibitor and the nucleophilic sulfur of cysteine strongly influences the kinetics of the reaction because either the inhibitor or cysteine moiety or both have to move to reach distances at which the interaction between both reacting centers is strong enough to initiate the chemical reaction leading to the bond formation. This is obvious as the larger the distance is, the slower the reaction takes place. The resulting mean distances of K11777's MD simulations (Figure 8.11-A) show that they are between 3.5 to 3.7 Å which is slightly smaller than the distance of 3.8 Å of the docked conformation (Table 8.13). Still fluctuations in distance occur in the course of simulation leading to standard deviations of about 0.3 Å (Figure 8.11-A). In contrast to K11777, inhibitor 7a reveals two different scenarios (Figure 8.11-B). In the first MD simulation the distance between both reacting centers fluctuates either around 3.5 or around 4.5 Å, thus leading to a slightly larger mean in distance compared to K11777 (Figure

8.11 and Table 8.13). Still, due to the flucutations the standard deviation is of 0.5 Å. In the remaining simulations the mean in distance is 4.5 Å with low fluctuations (standard deviation of 0.2 Å). These findings explain the experimental kinetics, as inhibitor 7a shows mostly larger distances to Cys25 compared to K11777. Thus, the nucleophilic attack is favored for K11777 due to the dynamic of the non-covalent enzyme-inhibitor complex.

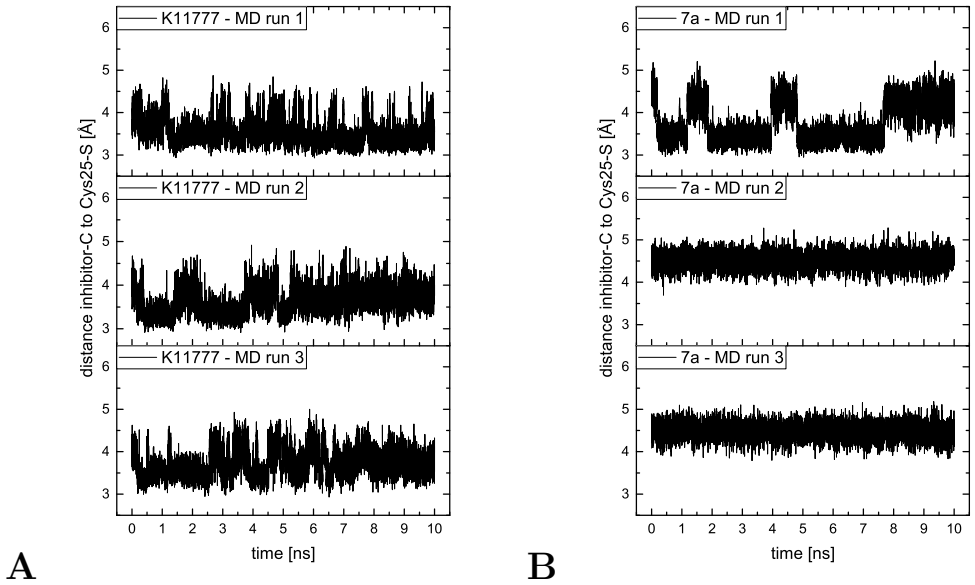


Figure 8.11: The variation in distance is plotted as a function of simulation time for **A** – K11777's electrophilic carbon to Cys25's sulfur starting from pose III and **B** – 7a's electrophilic carbon to Cys25's sulfur starting from pose II in non-covalent complex with rhodesain during 10 ns production run. The related calculated mean and standard deviation are summarized in Table 8.13.

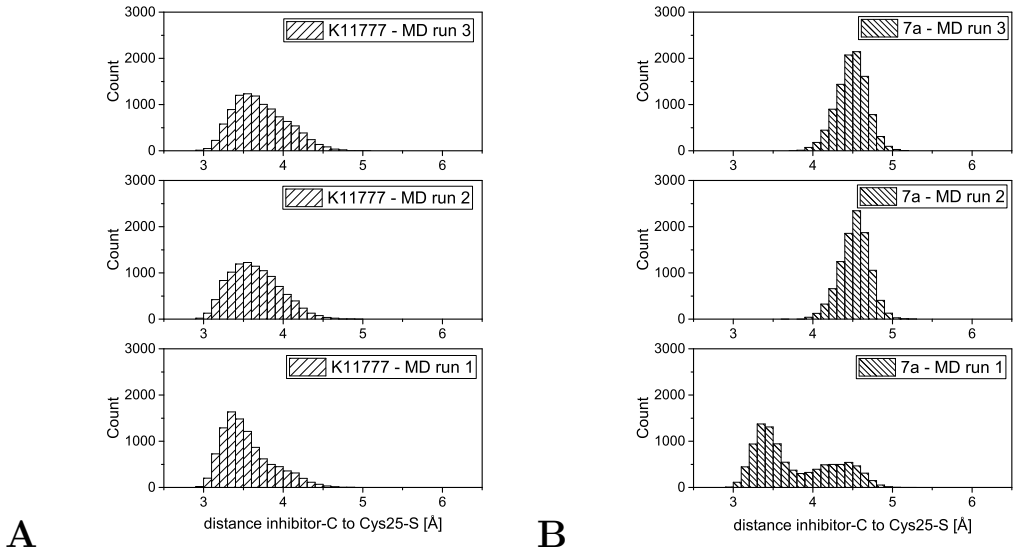


Figure 8.12: Frequency of the measured distance between **A** – K11777's or **B** – 7a's electrophilic carbon and Cys25's sulfur. The related calculated mean and standard deviation are summarized in Table 8.13.

Table 8.13: Calculated mean and standard deviation of the distance between the electrophilic carbon of K11777 or 7a to Cys25's sulfur in rhodesain. Mean and standard devations are calculated from 10,000 data points. All values are given in Å.

K11777	**MD run 1**	**MD run 2**	**MD run 3**
mean	3.54	3.64	3.71
standard deviation	0.31	0.31	0.33
7a	**MD run 1**	**MD run 2**	**MD run 3**
mean	3.77	4.53	4.49
standard deviation	0.48	0.18	0.19

8.10 Conclusions

Aim of this study was to qualitatively reproduce experimental kinetic data (k_{inact} values) obtained by Bryant et. al [176]. The investigation focused on the vinylsulfone inhibitor K11777 and three selected derivatives (7a, 7b and 8a, see Figure 8.1) which target cysteine proteases such as rhodesain of *Trypanosoma brucei rhodensie* and cruzain of *Trypanosoma*

cruzi. Therefore, different computational approaches on MM, QM or QM/MM level have been conducted. Why are k_{inact} values of interest? If these values are computationally accessible (or at least differences between inhibitors can be estimated) one can predict the suitability of novel covalent vinylsulfone inhibitors or would be able to differentiate between acceptable or unsuitable ones.

Investigations started with MD simulations which have been conducted to obtain relaxed structures for each covalent protein-inhibitor complex. For each simulation a certain snapshot has been extracted and the backward reaction from the covalent to non-covalent complex has been calculated *via* a QM/MM relaxed scan on B3LYP-D3/def-TZVP level and applying the ff14SB force field. Results indicate that 7a and 3a are superior in respect to the kinetics compared to K11777 or 7b in rhodesain which is contrary to experimental values. Thus, it was not possible to transfer these results to cruzain. The question arised if the applied methodology and basis set was sufficient to describe the reaction. Therefore, a suitable QM methodology was evaluated by a benchmark, in which CCSD(T) was used as reference and additionally the basis set size was increased to aug-cc-pVDZ. The spin-component scaled CC2 approach showed the lowest MAE values in relation to CCSD(T). The benchmark calculations also revealed that the application of implicit solvation does not affect the resulting optimized geometries but decreases reaction energies by a value of 5-6 kcal/mol.

In a next step the energetics of the transition state as well as the non-covalent state have been computed on QM/MM level by different ab initio or DFT approaches and have been compared to SCS-CC2, as computational costs of CCSD(T) were too high. The order of the transition state energies remained the same for all tested inhibitors in rhodesain like in the case of the B3LYP-D3/def-TZVP computations. Consequently, it was not possible to transfer these findings to cruzain. The next attempt was to perform MD simulations of the non-covalent protein-inhibitor complex obtained from the relaxed scan. Aim of these simulations was to probe the variation in distance between Cys25's nucleophilic sulfur and the inhibitors electrophilic carbon to gain insight into kinetic data. The lowest mean in distance was obtained for K11777 which is in line with experimental findings, but for 7b, too which does not reflect the experimental data. Inhibitor 7a shows higher mean distances, while 3a appear to be the most unstable in terms of the distance.

The focus of investigation was shifted on the one hand to the analysis of K11777 and 7a in rhodesain as both showed similar Ki_{app} values but different k_{inact} values. On the other hand, the simulations started from the non-covalent complex. The first step was the redocking of K11777 and the docking of 7a to obtain the best poses. To validate these findings MM-PB(GB)SA calculations have been conducted for each conformation to filter out the best

pose. Lastly MD simulations of the best pose have been carried out to assess the variation of distance between Cys25's sulfur and K11777's or 7a's electrophilic carbon. The results indicate that the distance of K11777 is in most cases shorter while 7a is for the longest peroid about 1 Å farther apart. In terms of the kinetics results of the MD simulations do reflect qualitatively experimental data with respect to the k_{inact} value.

Chapter 9

Investigations on the proton transfer and covalent modification of kinase AKT1

9.1 Introduction kinase AKT1

More than 30 years ago the v-AKT oncogene of AKT8 transforming retrovirus was identified by Stephen Staal [179]. Four years later, three different groups have characterized independently the cellular homolog of v-AKT as a Ser/Thr protein kinase [180], [181], [182]. Woodgett and Coffer used library screening and named the identified protein kinase because of the analogy to the protein kinases A and C (PKA and PKC) protein kinase B (PKB) [182]. Nowadays it is known that AKT/PKB is conserved in three isoforms in mammalian genomes: AKT1 (PKBα), AKT2 (PKBβ) und AKT3 (PKBγ) [183]. Cantley et al. showed that the activity of AKT is induced by the phosphoinositide 3-kinase (PI3K), a lipid kinase which is activated by an extracellular stimulus [184]. A range of intrinsic or extrinsic cell survival signals are induced by the PI3K/AKT pathway which leads to an increased proliferation, the loss of the apoptosiss signal or cell growth and additionally is in relation with the insulin response [185], [186], [187].

The signal transduction starts with the stimulation of receptor tyrosine kinases (RTK) or G protein-coupled receptors (GPCRs), whereby PI3K isoforms are recruited and activated at the plasma membrane (Figure 9.1). By the class I PI3K a phosphorylation of the 3'hydroxyl of the inositol head group of phosphoinositides occurs. Thus, the production of lipid second messenger PtdIns-3,4,P_2 (PI3,4P_2) and PtdIns-3,4,5-P_3 (PIP_3) which results from the phos-

phorylation of PI4,5P_2 by PI3K, is increased [183], [188]. The binding of PIP_3 to AKT is enabled by the kinase's pleckstrin homology (PH) domain. This PH domain is present in many proteins and plays an important role for intracellular signalling or is also an essential component of the cyto skeleton [189], [190], [191], [192], [193], [194], [195]. At endomembranes PI3,4P_2 is produced after the reaction of PIP_3 with 5'phosphatase SH2-domain containing inosotol 5'phosphatase (SHIP), which can interact with the PH domain of AKT [196], [197]. This shows that PI3K and its lipid products are necessary for interaction with AKT and in consequence are rate limiting for proper AKT activation. Regardless of the membrane, the relocalization of inactive AKT in the cytosol after synthesis of PIP_3 or PI3,4P_2 and in consequence recruitment to the membrane is needed.

After activation of AKT *via* PI3K, phosphorylation of two "key residues" are necessary to enable the maximal activity. On the one hand, Thr308 (T-loop) of the catalytic protein kinase core has to be phosphorylated and on the other Ser473 of the C-terminal hydrophobic motif [198]. Thr308 is phosphorylated by the phosphoinositide-dependent protein kinase 1 (PDK1). This modification occurs by relocalization of AKT and PDK1 to the membrane site of PIP3 or by PI3,4P_2 synthesis induced change of conformation which enables the phosphorylation of AKT's Thr308 residue by PDK1 [199], [200]. The activity of AKT is strongly reduced in absence of a Ser473 phosphorylation because this modification stabilizes Thr308's phosphorylation [198], [201]. The phosphorylation of Ser473 occurs by mTORC2, although the role of PI3K activation and hence PIP_3 is not clarified yet.

The termination of AKT's signal can happen in several ways. On the one hand a reduced AKT activity is indirectly evoked by the tumor suppressor PTEN, a PIP_3 phosphatase which leads to a dephosphorylation of PIP_3 to PI4,5P_2. On the other hand, AKT can be regulated directly at the "key residues". Protein phosphatase 2A (PP2A) dephosphorylates AKT at Thr308 and in consequence leads to inactivation of AKT [202]. Furthermore, PH domain leucine-rich repeat protein phosphatases (PHLPP1 and PHLPP2) have been identified as physiological AKT Ser473 phosphatases [203], [204].

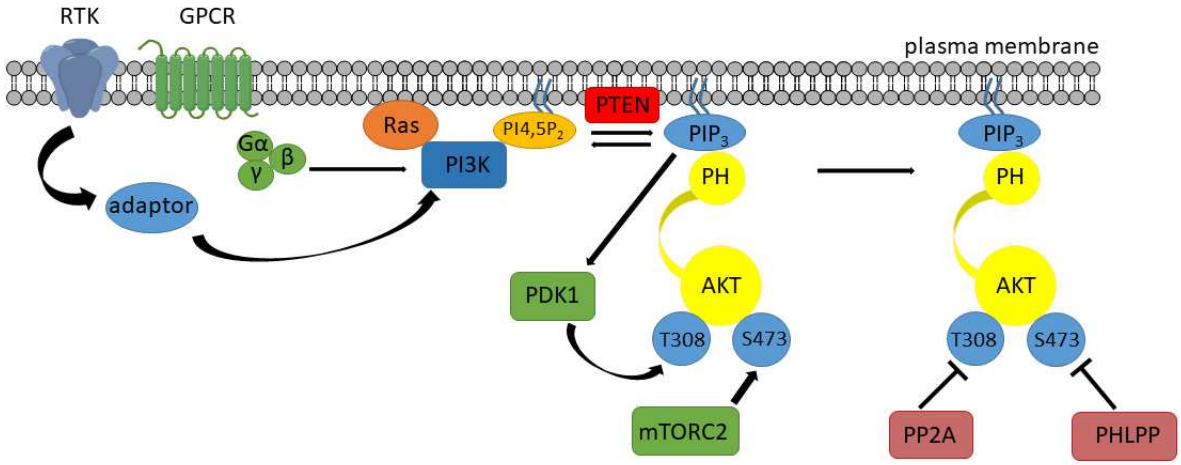


Figure 9.1: Signal transduction of AKT regulation at the plasma membrane adapted from [183].

The phosphorylation of further enzymes by AKT shows that a vast range of cellular functionalities are modulated. The main aspects are depicted in Figure 9.2. The glycogen synthase kinase 3 (GSK3) was the first substrate of AKT which was identified as multifunctional serine and threonine protein kinase. The phosphorylation by AKT leads to a hampered binding for substrates of GSK3. GSK3 has a vital role in cellular metabolism by e.g. the inhibition of metabolic enzymes as its substrate glycogen synthase [205], [206].
The phosphorylation by AKT can also supress the expression of Forkhead Box O Family Transcription Factors (FoxO) targets, which usually induce apoptosis, growth inhibition or tissue-specific metabolic changes [207], [208]. Furthermore, AKT can also modulate the tuberous sclerosis complex 2 (TSC2/tuberin), which is inactived by phosphorylation and thus enables an activation of mTORC1 [209], [210], [211]. Consequently, mTORC1's activity leads to an increased range of anabolic processes as the synthesis of protein, lipid and nucleotides and also an inhibition of autophagy [212].
Due to the diverse modulation induced by AKT it becomes evident that changes or dysregulation of the PI3K/AKT signalling pathway leads to manifold diseases. In this occasion, miscellaneous carcinomes as lung, prostate, endometrial or breast cancer arise from a disturbed signalling pathway [185]. In particular, some breast carcinomes show an overexpression or mutation of AKT which is associated with the progress in disease and additionally increased resistances to chemo- or radiotherapy [213], [214]. Also, a disturbance of AKT termination is related to cancer, as for a reduced or non-existent PHLPP1/PHLPP2 expression [215]. Furthermore, diseases as diabetes mellitus 2, autoimmune diseases, lipodystrophy, chronic

inflammatory disease are caused by disturbances of GSK3, FoxO or TSC-mTORC1 [210].

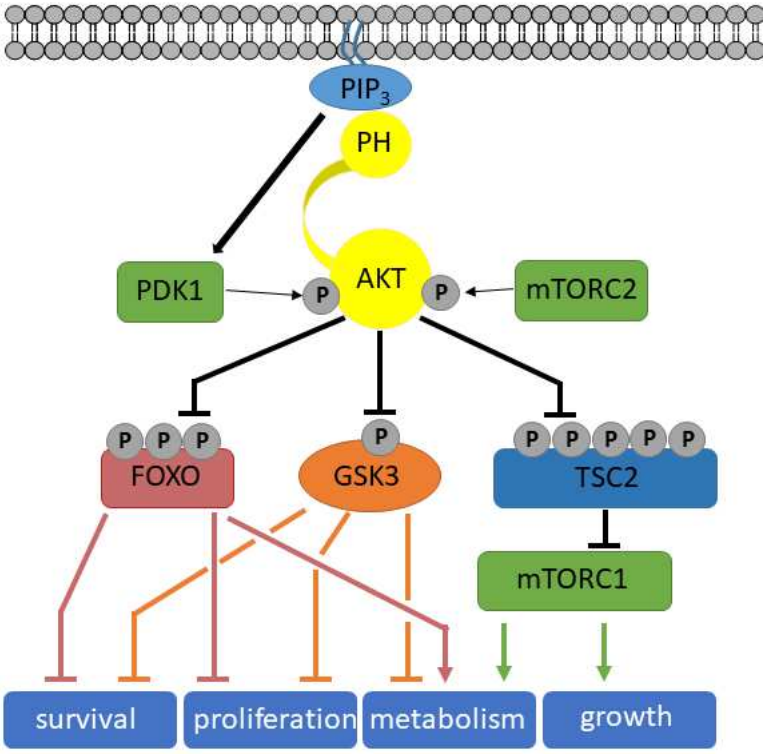


Figure 9.2: Regulation of cellular functions via AKT adapted from [183].

Based on the numerous diseases which result from the dysregulation of AKT, it becomes clear that AKT is an interesting target for the research of possible inhibitors. In case of protein kinases, one can find various orthosteric inhibitors, which occupy the active site [216]. In case of AKT one should mention ipatasertib or afursertib, a thiophenecarboxamide-derivative, which is already in phase I/II studies [217], [218], [219], [220], [221]. Both mentioned examples are ATP-competitive small molecules and hence selectivity is a severe problem, as this binding pocket is ubiquitous in the AGC kinase family. In contrast, allosteric inhibitors which bind on the protein surface and block the enzyme through an induced conformational change in the protein binding pocket, are able to enhance the selectivity because their binding sites are not conserved within the enzyme family. [222], [223].
In the inactive conformation of AKT the PH domain can be inhibited ("PH-in" conformation), while PI3K products lead to a "PH-out" conformation, enabling the phosphorylation of PDK1 [224], [225]. By targeting AKT *via* allosteric inhibitors one can achieve a stabi-

lization of the "PH-in" conformation, in which AKT stays inactive [226]. A novel method was introduced by the development of so called covalent-allosteric inhibitors (CAI's), which combine properties of covalent inhibitors and allosteric modulaters [227].
Rauh's group from the Technical University Dortmund developed different covalent-allosteric inhibitors and additionally crystallized them in covalent complex with AKT (Figure 9.3). These complexes show a binding in the interdomain region between the PH-domain or the kinase-domain due to covalent modification of Cys296 or Cys310. This is quite interesting as the ligands do not differ strongly. While RL1784 (also known as borussertib) was covalently bound to Cys296, the other two inhibitors form a covalent bond to Cys310.
While X-ray crystallographic data indicate a bonding to Cys296 or Cys310, mass spectrometry indicated a modification of both cysteine residues for inhibitors RL1784 or RL1780. For RL2132 no mass spectrometry data are available. The data of mass spectrometry can only give qualitative results (binding or non-binding) but quantitative data are missing.
Thus, computations were conducted to gain further insight. Each inhibitor was prepared in a non-covalent complex with AKT and dynamics have been obtained by MD simulations. Within the simulations distances between the nucleophilic cysteine and the electrophilic warhead are determined so that one can conclude qualitatively which residue is more likely modified. To further extend these simulations investigations in regard to the deprotonation of both cysteine residues were carried out. On this occasion nearby basic amino acids and their distances to both cysteine residues have also been analyzed.

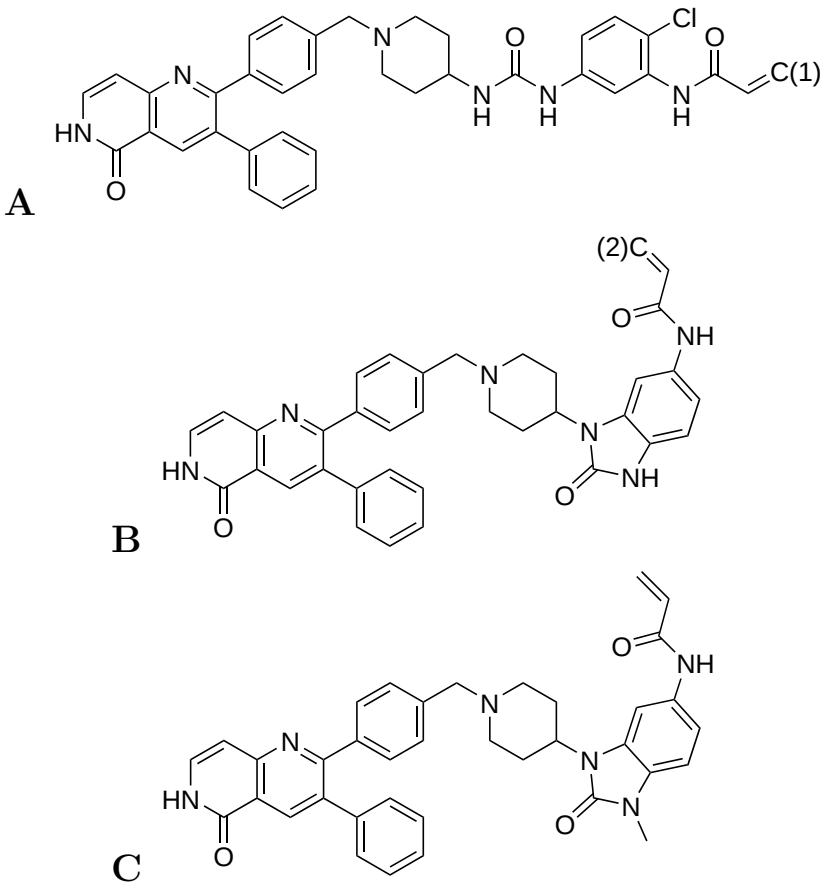


Figure 9.3: **A** - Structure of inhibitors N-(2-oxo-3-(1-(4-(5-oxo-3-phenyl-5,6-dihydro-1,6-naphthyridin-2-yl)benzyl)piperidin-4-yl)-2,3-dihydro-1H-benzo[d]imidazol-5-yl)acrylamide (RL1780), **B** - N-(2-chloro-5-(3-(1-(4-(5-oxo-3-phenyl-5,6-dihydro-1,6-naphthyridin-2-yl)benzyl)piperidin-4-yl)ureido)phenyl)acrylamide (RL1784, or known as borussertib) and **C** – N-(1-methyl-2-oxo-3-(1-(4-(5-oxo-3-phenyl-5,6-dihydro-1,6-naphthyridin-2-yl)benzyl)piperidin-4-yl)-2,3-dihydro-1H-benzo[d]imidazol-5-yl)acrylamide (RL2132).

9.2 System preparation

The main difference within the system preparation steps compared to the complexes in other chapters was the need to generate complete pdb structures as residues were missing. For the crystal structure of RL1784 in covalent complex with AKT1 only the residues 5-109, 145-188, 200-297 and 308-439, for RL1780 the residues 3-44, 49-112, 144-186 and 205-444 and for RL2132 the residues 5-45, 50-109, 145-188, 200-297 and 308-439 were available. The first consideration was to restrain the backbone of existing amino acids and simulate the dynamics. This method failed as the protein still moved, although, a huge force with respect to the

harmonic restraint was applied. In consequence it was not possible to draw any conclusions out of these simulations. It was of necessity to add the missing residues to obtain trustworthy results from MD simulations. Thus, the Modeller 9.18 program package was employed for the generation of the complete pdb structure [228], [229], [230].
First of all, the sequence information was extracted of the existing crystal structure to obtain the FASTA code. Afterwards an alignment between the original structure, with gap characters corresponding to the missing residues, and the known full sequence was performed. Afterwards a model containing all residues was generated while the former incomplete pdb file was used as a template (Figure 9.4). This resulted in nine different pdb structures. The structures Ia, Ib and Ic represent the X-ray of AKT1 with covalently bound RL1784, while structures IIa, IIb and IIc reflect the crystal structure of AKT1 in complex with RL1780 (Figure 9.5 and Figure 9.6). Finally, IIIa, IIIb and IIIc give the structure of RL2132 in covalent complex with the protein (Figure 9.7).
This analysis of the RMSD values show large differences in generated structures. This indicates the need of a longer simulation time compared to previous simulations to obtain meaningful results. Therefore, each production run was conducted with duration of 50 ns. The protein was described with the AMBER14 force field ff14SB as before, while the GAFF force field was applied to inhibitors RL1784, RL1780 or RL2132. The procedure of minimization and gradual heating was carried out as previously explained in section 5.1.

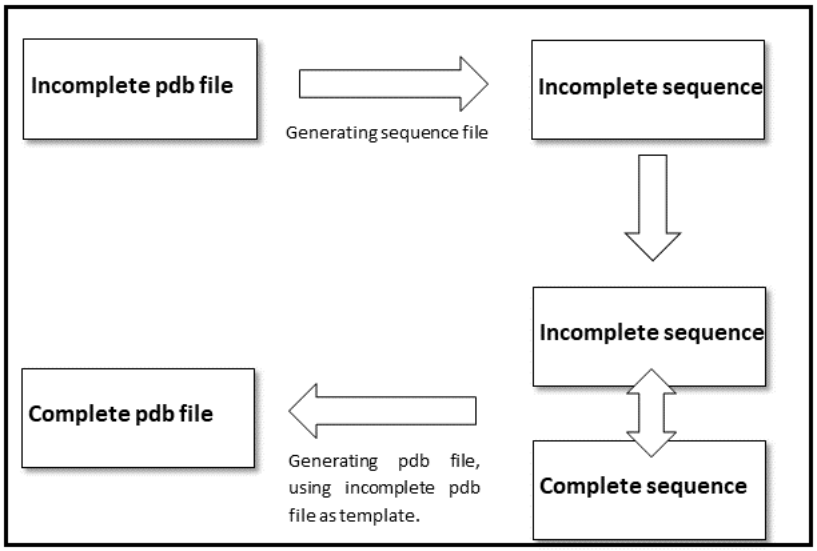


Figure 9.4: Procedure of generating complete pdb file by using incomplete X-ray as template.

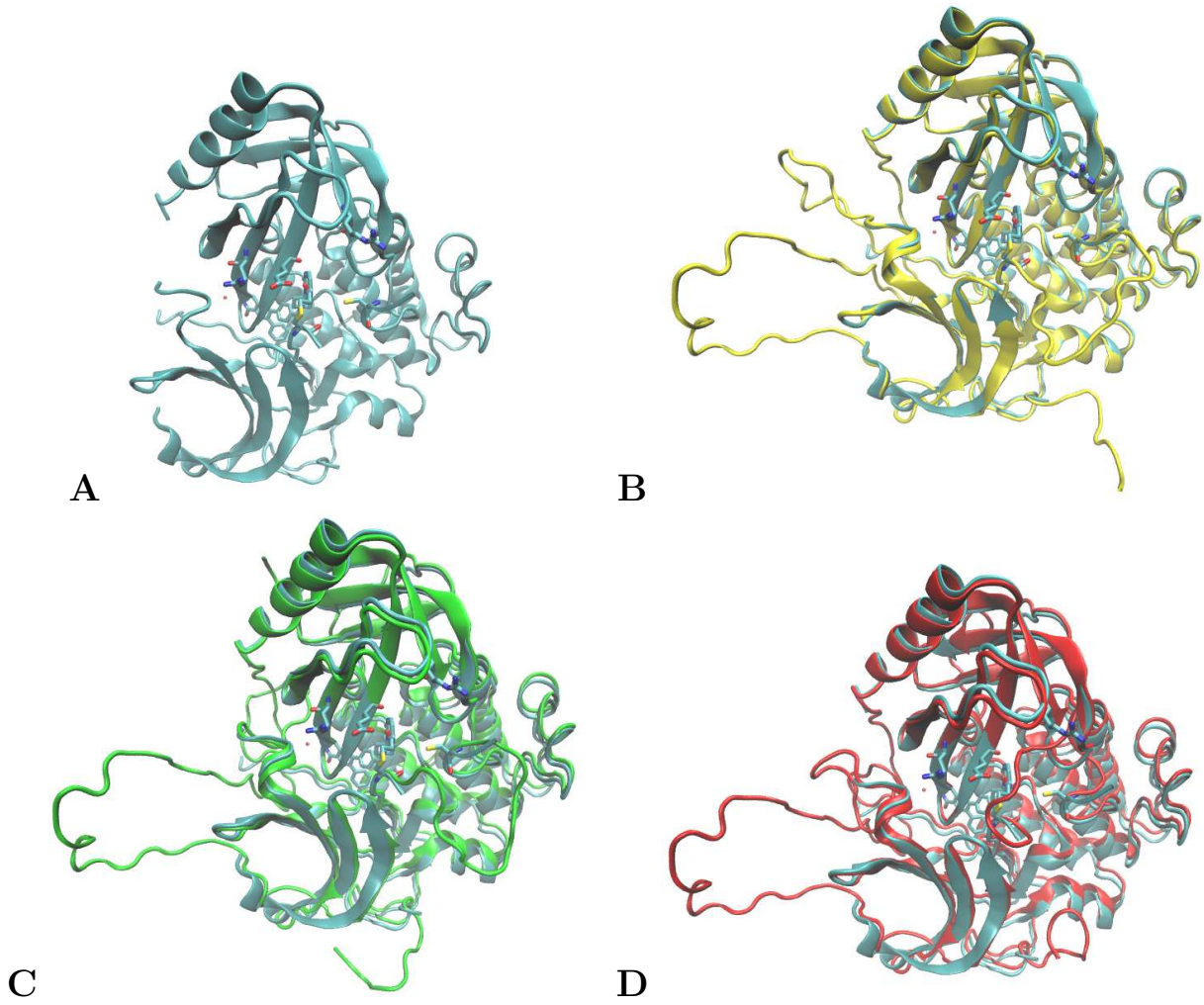


Figure 9.5: **A** – Incomplete pdb structure of AKT1 in covalent complex with RL1784. **B** – complete pdb structure of AKT1 structure Ia (yellow) aligned with incomplete structure. **C** – complete pdb structure of AKT1 structure Ib (green) aligned with incomplete structure. **D** – complete pdb structure of AKT1 structure Ic (red) aligned with incomplete structure.

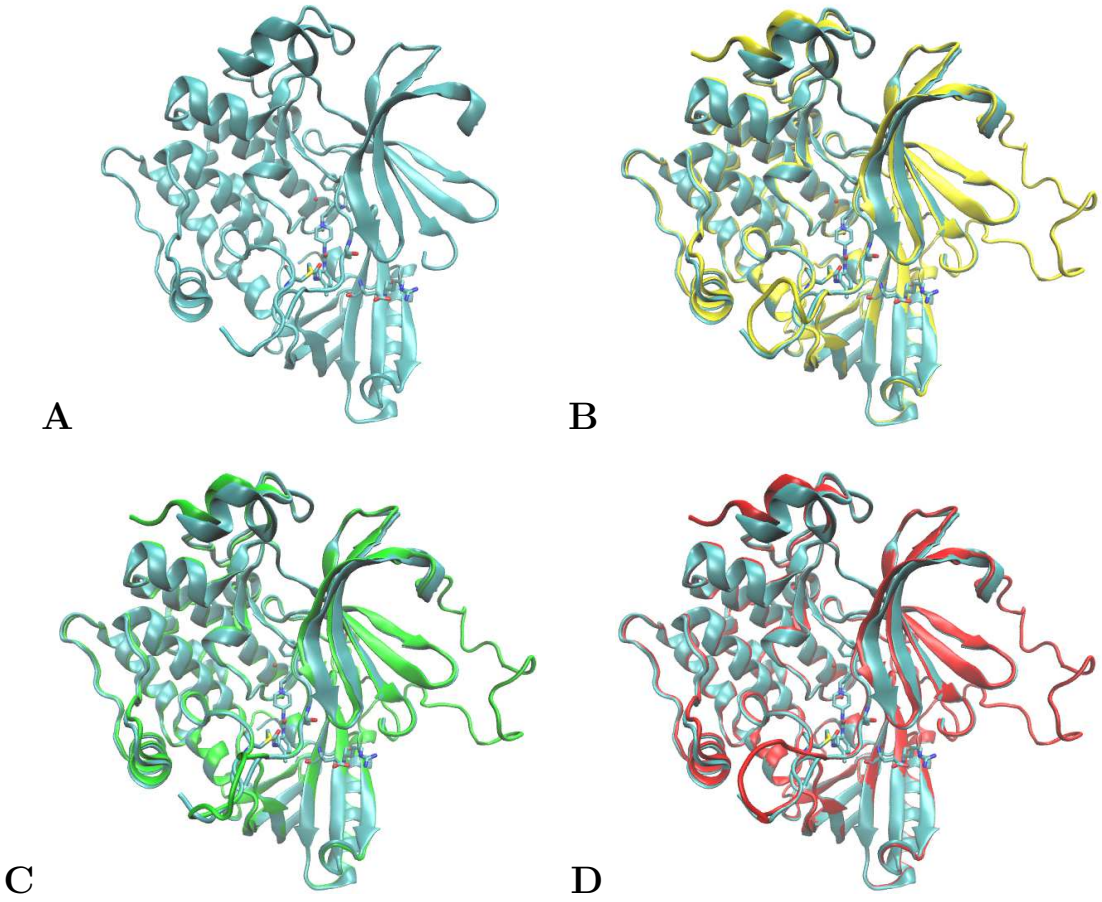


Figure 9.6: **A** – Incomplete pdb structure of AKT1 in covalent complex with RL1780. **B** – complete pdb structure of AKT1 structure IIa (yellow) aligned with incomplete structure. **C** – complete pdb structure of AKT1 structure IIb (green) aligned with incomplete structure. **D** – complete pdb structure of AKT1 structure IIc (red) aligned with incomplete structure.

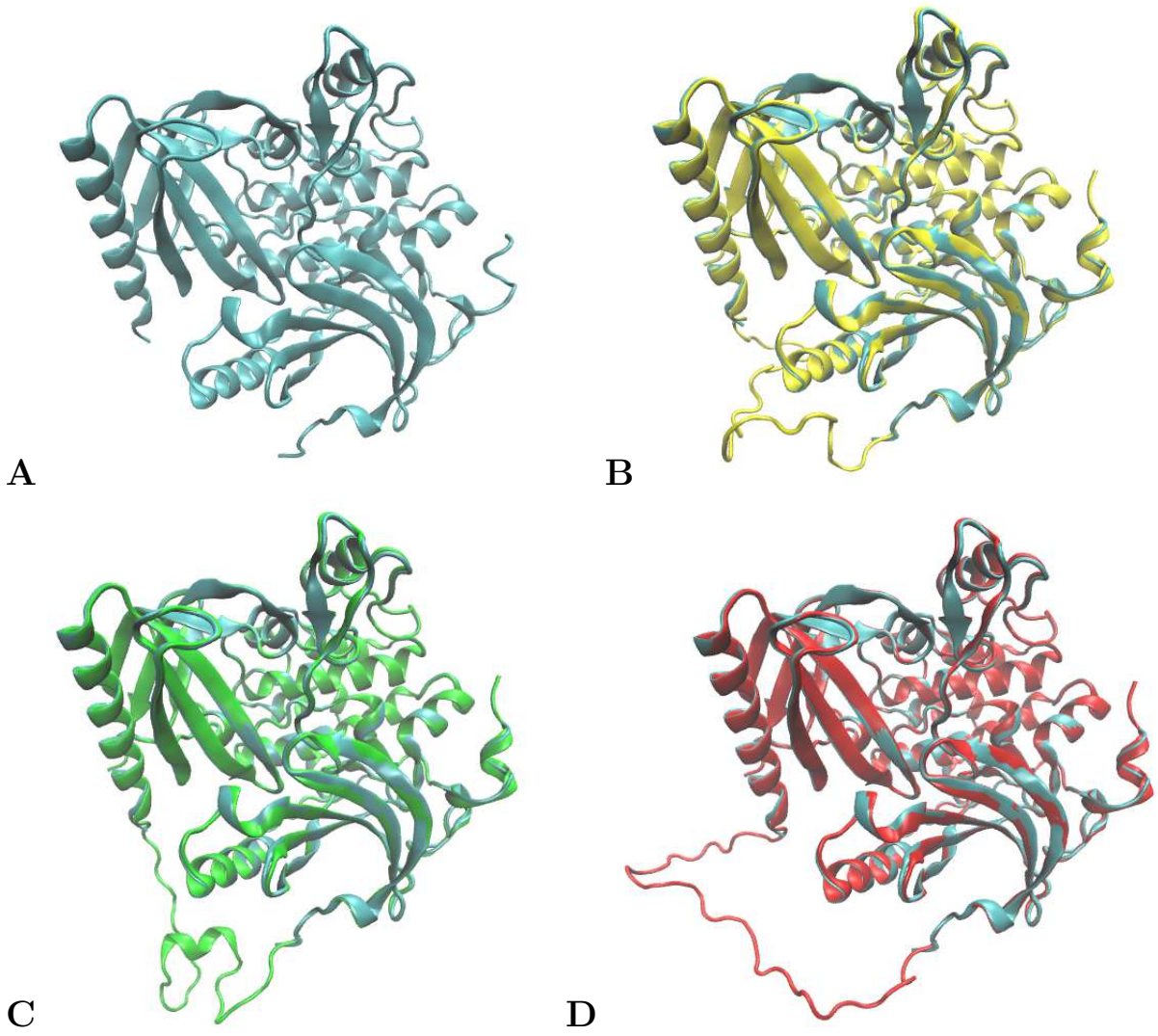


Figure 9.7: **A** – Incomplete pdb structure of AKT1 in covalent complex with RL2132. **B** – complete pdb structure of AKT1 structure IIIa (yellow) aligned with incomplete structure. **C** – complete pdb structure of AKT1 structure IIIb (green) aligned with incomplete structure. **D** – complete pdb structure of AKT1 structure IIIc (red) aligned with incomplete structure.

To investigate whether Cys296 and Cys310 can interact with the inhibitors, the AKT1-inhibitor complexes were parameterized as a non-covalent complex as mentioned before. Insights if a reaction can take place is given by the distance between the electrophilic center of the inhibitor and the nucleophilic sulfur of cysteine residue (Figure 9.8). As the distance is related to the probability of the reaction itself, it is possible to draw qualitative conclusions concerning the modified cysteine residue. Clearly, this variation in distance cannot be simulated in a covalent protein-inhibitor complex.

Table 9.1: Calculated RMSD values of protein between generated structures Ia, Ib, Ic, IIa, IIb and IIc. Mean and standard devations are calculated from 10,000 data points. All values in Å.

protein [Å]	**Ia**	**Ib**	**Ic**	**IIa**	**IIb**	**IIc**	**IIIa**	**IIIb**	**IIIc**
Ia	0.00	3.66	3.54	5.19	5.22	5.15	4.50	6.06	7.04
Ib	3.66	0.00	4.13	5.15	5.19	5.12	5.18	6.60	7.52
Ic	3.54	4.13	0.00	5.57	5.62	5.58	4.49	6.02	6.99
IIa	5.19	5.15	5.57	0.00	0.59	0.73	4.05	5.74	5.88
IIb	5.22	5.19	5.62	0.59	0.00	0.68	4.05	5.74	5.89
IIc	5.15	5.12	5.58	0.73	0.68	0.00	4.00	5.70	5.84
IIIa	4.50	5.18	4.49	4.05	4.05	4.00	0.00	3.91	3.96
IIIb	6.06	6.60	6.02	5.74	5.74	5.70	3.91	0.00	4.66
IIIc	7.04	7.52	6.99	5.88	5.89	5.84	3.96	4.66	0.00

Additional simulations have been performed by introducing RL1780 into the environments Ia, Ib or Ic, RL1784 into IIa, IIb or IIc and also RL2132 into Ia, Ib and Ic to observe the effects on the complex by inserting the inhibitor into the "wrong" environment. By aligning the proteins, RL1784 has been placed in the protein environment of IIa, IIb and IIc while RL1780 or RL2132 were inserted into Ia, Ib, Ic. All simulations of the protein-inhibitor complex only considered protonated cysteine residues. It has to be mentioned that a deprotonated Cys296 or Cys310 might result in different findings but these protonation states were not included due to computational efforts.

Protonated cysteine residues are not sufficiently reactive for a bond formation, hence the second question was to figure out responsible amino acids for the deprotonation of Cys296 or Cys310. In this context, MD simulations of structures Ia, Ib, Ic, IIa, IIb and IIc in absence of an inhibitor were conducted to investigate possible modulation of Cys296 or Cys310 by basic aminoacids. If one takes a look at the crystal structure of RL1784 in complex with AKT1 one can observe a distance of 4.4 Å between Cys296's sulfur and Glu85's oxygen. Therefore, the analysis of this distance during the course of the production run was conducted. In case of Cys310 no clear interactions to basic amino acids were observable in respect to the X-ray structure. Consequently, it was necessary to investigate the production runs in terms of developing interactions to basic amino acids during the course of simulation.

9.3 Protein-inhibitor interactions

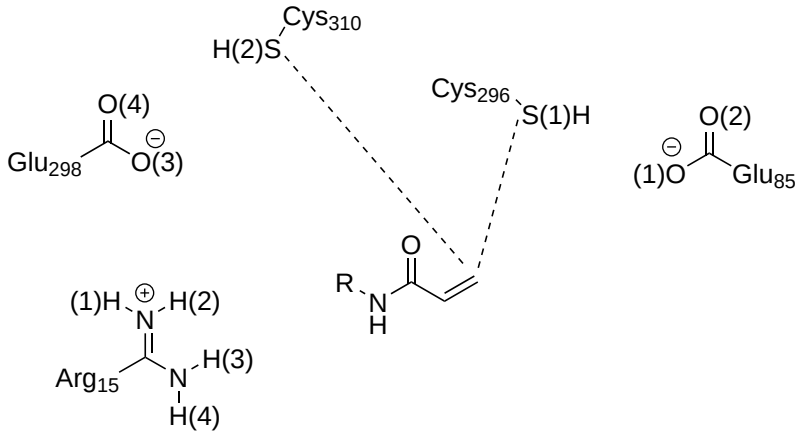


Figure 9.8: Interactions of active site Cys296 or Cys310 with RL1784 or RL1780 depicted with dashed lines, corresponding to the following distance plots. Location of amino acids and warhead serves for clarification and does not reflect actual position in MD simulations.

The simulations revealed that RL1784 in Ia and Ib show mean distances of 4.3 or 4.0 Å to Cys296 accompanied with low standard devations (0.9 and 0.6 Å) whereas larger values are obtained for the structure Ic (Table 9.2). Modification of Cys310 might occur in case of Ia and during 17 – 33 ns in Ib as the distance decreases below 4 Å, an appropriate value for a covalent attack (Figure 9.9-A and B). Still, the mean in distance is lower (especially in case of Ic) for Cys296 to the inhibitor compared to Cys310 which is observable in all three MD simulations. This is also the case for the complex of RL1784 in the generated structures IIa-IIc. The important difference one can observe is that the mean distance between the electrophilic carbon of RL1784 and both cysteine residue is increased compared to simulations in Ia to Ic. Due to the fact that the inhibitor was placed into the "wrong" protein environment, i.e. into the structures IIa - IIc which originates from the X-ray of RL1780 in complex with AKT1, this result is to be anticipated. Interestingly, the same trend is observable: In terms of the distance to Cys296 one obtains lower values independent of the utilized generated pdb structure.

For RL1780 in IIa – IIc a clear preference for Cys310 is observable. Mean distances between the electrophilic carbon of RL1780 and the sulfur of Cys310 are 3.5 or 3.6 Å for the structures IIa or IIb, clearly lower than RL1784 to Cys296 (Figure 9.10). Also, standard deviations of 0.3 Å are observable for both structures. Only for the protein-inhibitor complex IIc fluctuations up to 6 Å occur leading to a mean distance of 5.3 and a standard deviation of 1 Å.

The mean distance to Cys296 is about twofold of value compared to Cys310. A complete different picture can be seen if RL1780 is inserted into the "wrong" environment (Ia - Ic). In the cases of Ia to Ic the mean distances to Cys296 are lower as well as the values of the standard deviation. As a conclusion, one can state that it is significant to have a closer look at the course of distance.

In case of RL1780 in Ia one can observe during 0 to 12 ns and 48 to 50 ns a distance close enough for covalent modification of Cys310. At other times the distance increases up to 8 to 10 Å and exceeds the distance towards Cys296 which is about 7 Å (Figure 9.10-A). Although the mean distance to Cys296 is lower, the duration of an appropriate distance for covalent modification is longer for the distance between the inhibitor and Cys310. In terms of RL1780 in the generated pdb structure Ic (Figure 9.10-B), solely for the first 5 ns a modulation of Cys296 is expected while for Cys310 only few occasions for a covalent modification can be seen. A whole different picture can be seen for the structure Ib. At first glance it is obvious that Cys296 is favored in terms of distance and also showing appropriate distances for covalent modification during the course of simulation. Still a decrease in distance is observable after 10 ns for RL1780 to Cys310 and additionally lower fluctuations after 40 ns.

The crystal structure of RL2132 in covalent complex with AKT1 showed a covalent modification on Cys310. As one can see in Figure 9.11-D to Figure 9.11-F the inhibitor adopts a lower distance to Cys310 than to Cys296 if one considers the pdb structures IIIa, IIIb and IIIc. Fluctuations are lowest in case of IIIa, while in IIIb or IIIc the distance between RL2132 and Cys310 can increase up to 8 Å. If one compares the mean of the distance, it is observable that the distance to Cys296 is larger compared to Cys310 (Table 9.3). This is in contrast to the results obtained in structure Ia, Ib or Ic (Figure 9.11-A to Figure 9.11-C). The mean values of the distance between Cys310 to RL2132 are at the one hand larger than in IIIa, IIIb or IIIc. On the other hand Cys296 show mean distances which are lower than the one obtained from Cys310 to RL2132. In case of Ia one needs to take a closer look as the distance for a nucleophilic is given for Cys310 during $\sim$ 28 ns. Afterwards an increase to 12 Å occurs leading to a larger mean value. For Cys296 solely a decrease in distance to $\sim$ 4 Å is observable during 25-30 ns. For Ib the distance for nucleophilic attack increases for both cysteine residues. Interestingly a complete different picture can be seen in case of Ic. Cys296 shows an appropriate distance for covalent modification while the distance of Cys310's sulfur to the electrophilic carbon increases to 12-16 Å.

To shortly conclude the investigations in this section, MD simulations gave a qualitative answer whether Cys296 or Cys310 was preferably modified by the inhibitors RL1784, RL1780 or RL2132. In section 9.5 results are discussed in detail.

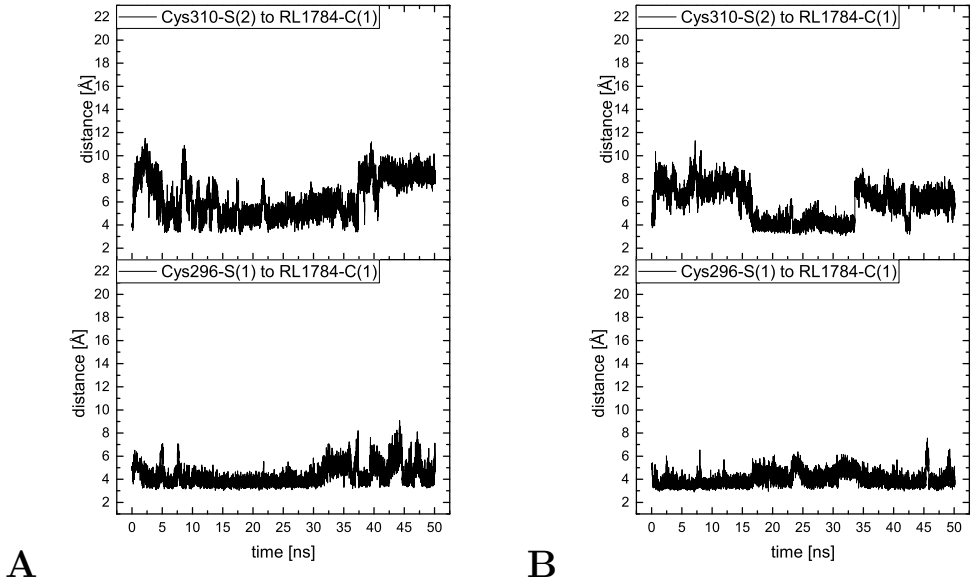


Figure 9.9: The variation in distance as a function of simulation time for RL1784 with Cys296 or Cys310 in structure Ia (**A**) and in structure Ib (**B**) corresponding to Figure 9.8.

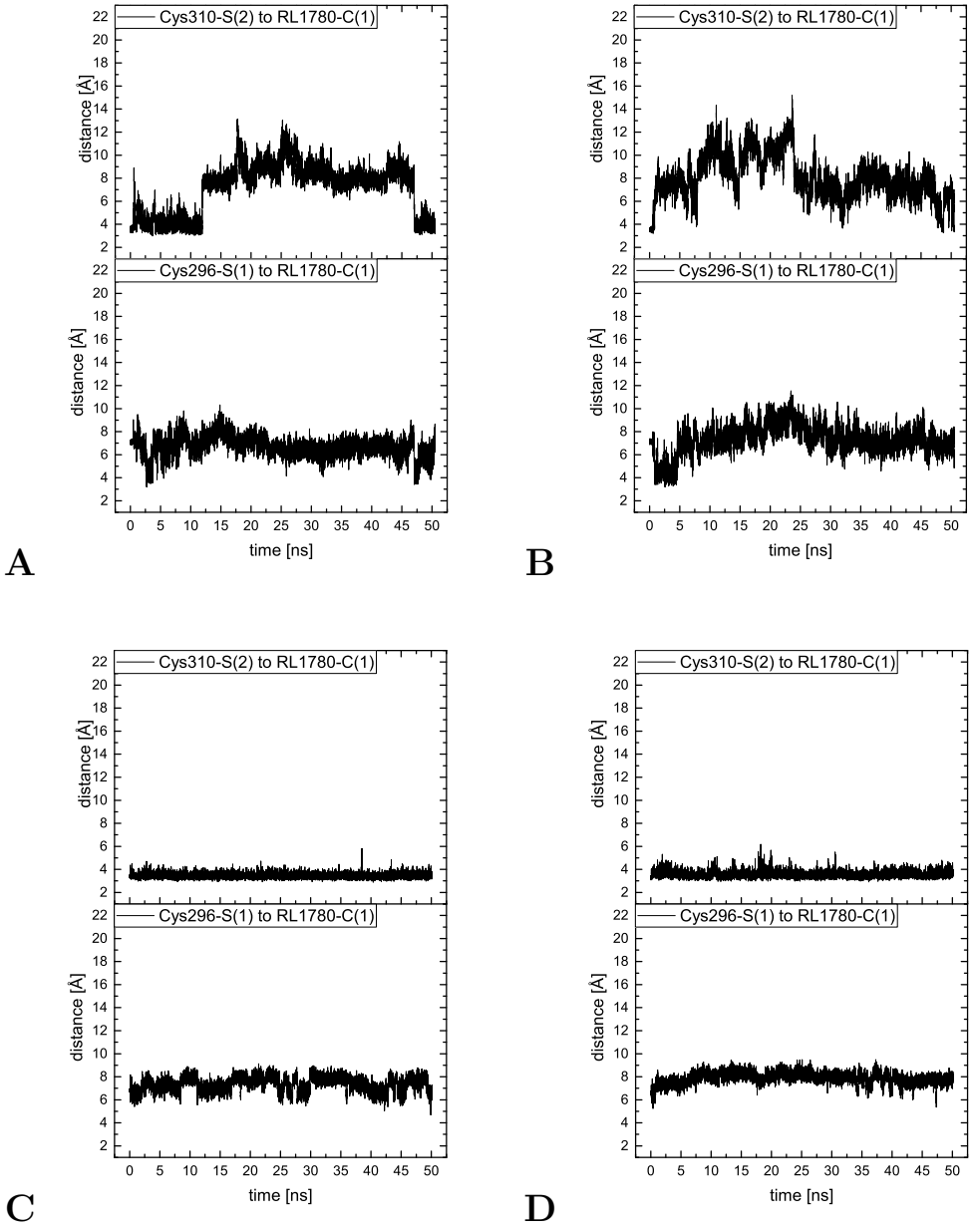


Figure 9.10: The variation in distance as a function of simulation time for RL1780 with Cys296 or Cys310 in structure Ia (**A**), in structure Ic (**B**), in structure IIa (**C**) and in structure IIb (**D**) corresponding to Figure 9.8.

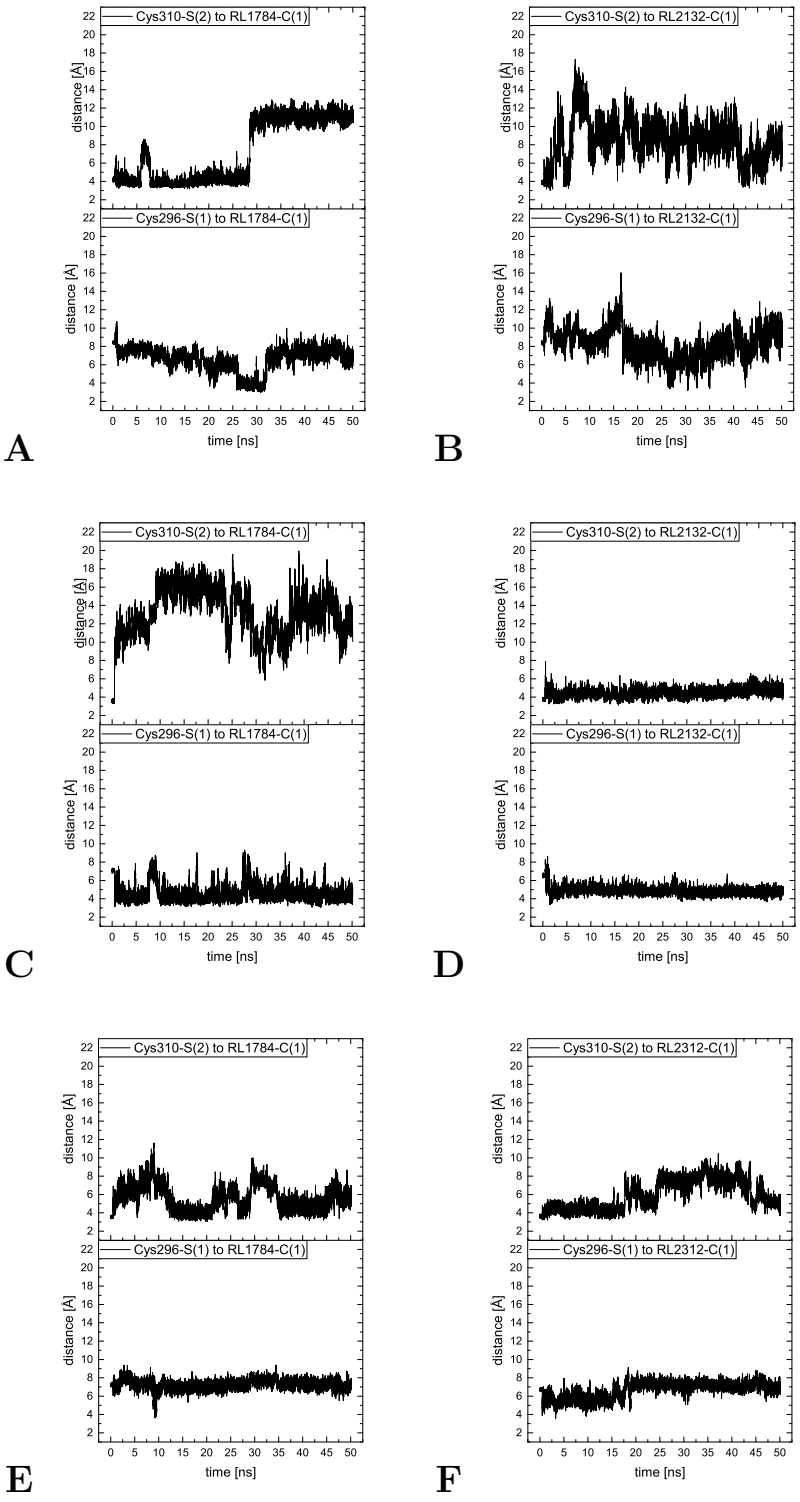


Figure 9.11: The variation in distance as a function of simulation time for RL2132 with Cys296 or Cys310 in structure Ia (**A**), in structure Ib (**B**), in structure Ic (**C**), in structure IIIa (**D**), in structure IIIb (**E**) and in structure IIIc (**F**), corresponding to Figure 9.8.

Table 9.2: Calculated mean and standard deviation of the distance between Cys296(S1) to RL1784-C(1) or RL1780-C(1) in protein kinase AKT-1. Mean and standard devations are calculated from 10,000 data points. All values in Å.

pdb	**distances**	**Cys296-S(1) to RL1784-C(1)**	**Cys310-S(2) to RL1784-C(1)**
Ia	**mean / SD**	4.3 / 0.9	6.3 / 1.7
Ib	**mean / SD**	4.0 / 0.6	5.8 / 1.5
Ic	**mean / SD**	5.2 / 2.0	13.0 / 2.4
IIa	**mean / SD**	5.5 / 0.6	6.6 / 0.6
IIb	**mean / SD**	5.0 / 0.7	7.0 / 1.3
IIc	**mean / SD**	5.6 / 1.2	10.1 / 1.7

pdb	**distances**	**Cys296-S(1) to RL1780-C(1)**	**Cys310-S(2) to RL1780-C(1)**
Ia	**mean / SD**	6.7 / 0.9	7.1 / 2.2
Ib	**mean / SD**	4.3 / 1.0	5.4 / 1.5
Ic	**mean / SD**	7.3 / 1.3	8.0 / 1.9
IIa	**mean / SD**	7.3 / 0.7	3.5 / 0.3
IIb	**mean / SD**	7.9 / 0.6	3.6 / 0.3
IIc	**mean / SD**	10.6 / 1.1	5.3 / 1.0

Table 9.3: Calculated mean and standard deviation of the distance between Cys296-S(1) or Cys310-S(2) to RL2132-C(1) in protein kinase AKT-1. Mean and standard devations are calculated from 10,000 data points. All values in Å.

pdb	**distances**	**Cys296-S(1) to RL2132-C(1)**	**Cys310-S(2) to RL2132-C(1)**
Ia	**mean / SD**	6.6 / 1.3	7.2 / 3.4
Ib	**mean / SD**	8.3 / 1.7	8.3 / 2.4
Ic	**mean / SD**	4.6 / 1.0	13.1 / 2.7
IIIa	**mean / SD**	4.9 / 0.5	4.4 / 0.5
IIIb	**mean / SD**	7.3 / 0.5	5.6 / 1.5
IIIc	**mean / SD**	6.7 / 0.9	5.9 / 1.6

9.4 Interactions of basic amino acids to Cys296 and Cys310 in non-covalent protein-inhibitor and apoprotein complex

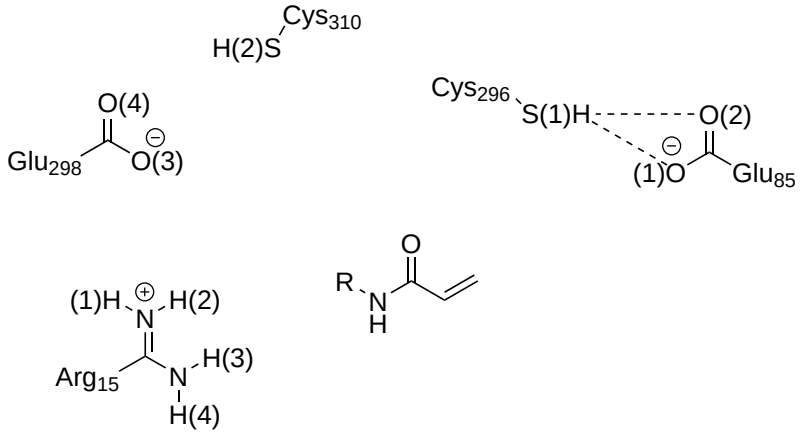


Figure 9.12: Interactions of active site Cys296 with Glu85 depicted with dashed lines, corresponding to the following distance plots. Location of amino acids and warhead serves for clarification and does not reflect actual position in MD simulations.

The next scope of investigation was to determine amino acids which might be able to deprotonate the cysteine residues 296 or 310. Why is this of importance? Only a deprotonated cysteine residue is able to form a covalent bond with an inhibitor. If the deprotonation process is understood, one can predict which cysteine residue is more likely modified.
As mentioned in the beginning of this section, a distance of 4.4 Å between Cys296 and Glu85 was measured in the crystal structure of RL1784 in AKT1. However, how does this distance evolve during the simulation time? To consider a possible rotation of the carboxylate group the distance between both oxygen and the thiol group is measured (Figure 9.12). In this respect the mentioned mean values are related to the lowest calculated one.
In most cases one can observe a movement of Glu85 opposite directed to Cys296 and thus to an increased distance towards Cys296. In case of RL1784 in Ib (Figure 9.13) one can observe a water-mediated interaction with Glu85 leading to distances of $\approx$ 5 Å. Similar but weaker interactions can be seen in RL1780 in IIb which occur infrequently (Figure 9.14). In case of RL2132 in Ib distances below 3 Å are occasionally observable indicating a possible direct proton transfer. Otherwise distances around 5 Å are found, however, because water molecules are inserted between both centers in the course of the simulation water-mediated

proton transfers between both amino acids might take place (Figure 9.15-A). This is also observable for RL2132 in IIIa leading to a similar variation in distance but slightly larger mean in distance compared to Ib (Figure 9.15-B and Table 9.5). Thus, a proton transfer might occur but distances for this occurrence are not given in every structure.
In absence of the inhibitors RL1784 or RL1780 mean distances between the investigated amino acids are partly lower. As in the presence of RL1784 one can see distances in Ib (Figure 9.16) which are close enough for a water-mediated proton transfer. In contrast to previous simulations this occurrence can also be seen for the structures IIa. After 5 ns of fluctuations one can observe a decrease of distance between Cys296 and Glu85 leading to an appropriate distance for a water-based proton transfer. These occurrences are rarer for IIc yet it shows possibilities of proton transfer compared to the simulations in presence of an inhibitor. Therefore, an occurrence of a proton transfer is possible.

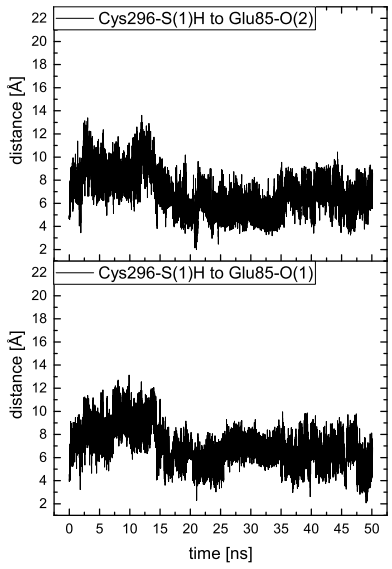


Figure 9.13: The variation in distance as a function of simulation time for Glu85 with Cys296 in presence of RL1784 in structure Ib corresponding to Figure 9.12.

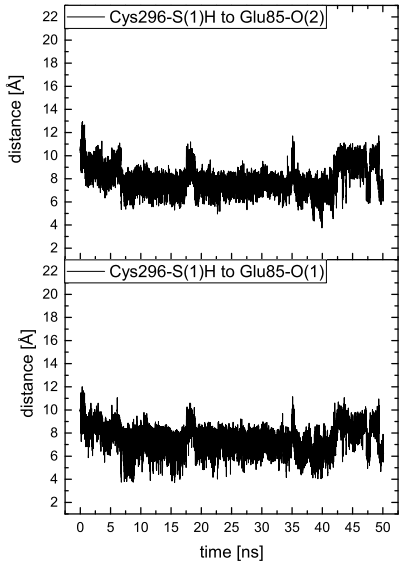


Figure 9.14: The variation in distance as a function of simulation time Glu85 with Cys296 in presence of RL1780 in structure IIb corresponding to Figure 9.12.

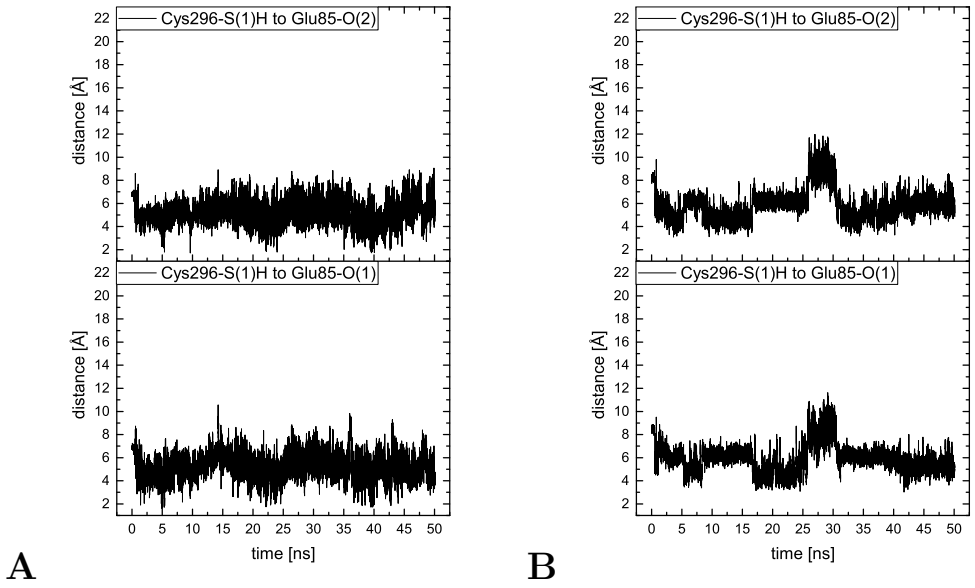


Figure 9.15: The variation in distance as a function of simulation time for Glu85 with Cys296 in presence of RL2132 in structure Ib (**A**) and in structure IIIa (**B**) corresponding to Figure 9.12.

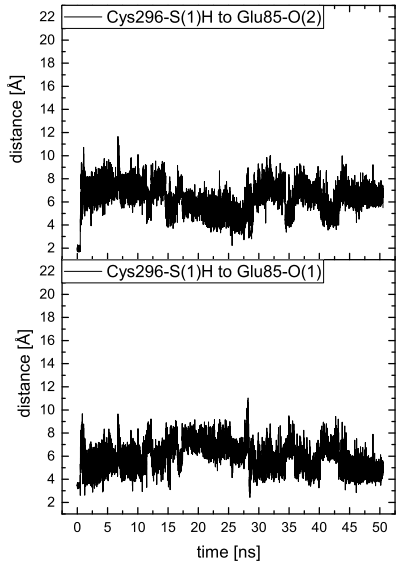


Figure 9.16: The variation in distance as a function of simulation time for Glu85 with Cys296 in absence of RL1784, RL1780 or RL2132 in structure Ib corresponding to Figure 9.12.

Table 9.4: Calculated mean and standard deviation of the distance between Cys296(S1) to Glu85(O1) or Glu85(O2) in protein kinase AKT-1. Mean and standard devations are calculated from 10,000 data points. All values in Å.

pdb	**distances**	**Cys296-S(1) to Glu85-(O1)***	**Cys296-S(1) to Glu85-(O2)***
Ia	**mean / SD**	9.8 / 1.8	9.7 / 1.8
Ib	**mean / SD**	7.0 / 1.7	6.9 / 1.8
Ic	**mean / SD**	10.4 / 2.1	10.6 / 2.3
IIa	**mean / SD**	10.2 / 1.4	10.4 / 1.4
IIb	**mean / SD**	9.7 / 1.5	9.8 / 1.5
IIc	**mean / SD**	10.5 / 2.1	10.8 / 2.1

pdb	**distances**	**Cys296-S(1) to Glu85-(O1)****	**Cys296-S(1) to Glu85-(O2)****
Ia	**mean / SD**	10.6 / 1.6	10.3 / 1.6
Ib	**mean / SD**	10.2 / 2.1	9.9 / 1.8
Ic	**mean / SD**	9.6 / 2.1	9.5 / 2.1
IIa	**mean / SD**	7.9 / 1.3	8.0 / 1.4
IIb	**mean / SD**	7.5 / 1.2	8.0 / 1.2
IIc	**mean / SD**	8.9 / 1.1	8.8 / 1.1

pdb	**distances**	**Cys296-S(1) to Glu85-(O1)*****	**Cys296-S(1) to Glu85-(O2)*****
Ia	**mean / SD**	11.7 / 2.2	11.8 / 2.1
Ib	**mean / SD**	5.9 / 1.2	6.3 / 1.3
Ic	**mean / SD**	8.9 / 1.7	9.1 / 1.7
IIa	**mean / SD**	6.7 / 1.5	7.1 / 1.6
IIb	**mean / SD**	9.2 / 1.7	9.4 / 1.4
IIc	**mean / SD**	6.7 / 1.3	6.9 / 1.4

*MD simulation in presence of RL1784. ** MD simulation in presence of RL1780. *** MD simulation in absence of inhibitor.

Table 9.5: Calculated mean and standard deviation of the distance between Cys296(S1) to Glu85(O1) or Glu85(O2) in presence of inhibitor RL2132 in protein kinase AKT-1. Mean and standard devations are calculated from 10,000 data points. All values in Å.

pdb	**distances**	**Cys296-S(1) to Glu85-(O1)**	**Cys296-S(1) to Glu85-(O2)**
Ia	**mean / SD**	7.5 / 1.7	7.5 / 1.8
Ib	**mean / SD**	5.2 / 1.2	5.2 / 1.1
Ic	**mean / SD**	9.1 / 1.8	9.2 / 1.6
IIIa	**mean / SD**	5.9 / 1.2	5.9 / 1.3
IIIb	**mean / SD**	7.5 / 1.8	8.1 / 1.4
IIIc	**mean / SD**	8.6 / 1.2	8.7 / 1.2

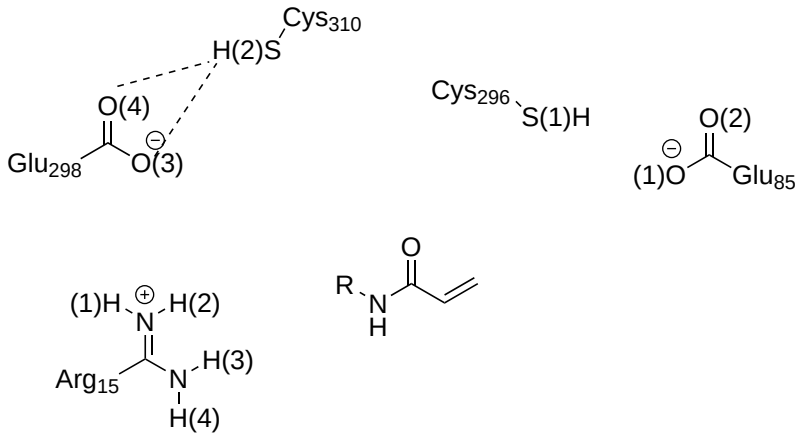


Figure 9.17: Interactions of active site Cys310 with Glu298 depicted with dashed lines, corresponding to the following distance plots. Location of amino acids and warhead serves for clarification and does not reflect actual position in MD simulations.

In terms of the deprotonation of Cys310, interactions with two amino acids are established in the course of the simulation time. A decrease of distance between Cys310 and Glu298 is observable in the non-covalent complex of AKT with RL1784 as well as in the absence of any inhibitor. Just as before the distance from Cys310 to both oxygens is measured and the given mean value is referred to the lower value. The distance between Glu298's oxygen and Cys310 thiol group in structure Ib shows a steady decrease to values between 2 to 4 Å observable after 35 ns (Figure 9.18-A) leading to a mean of 5.3 Å. In this case even a direct proton transfer might occur in contrast to interactions of Cys296 to Glu85. A quite similar behavior can be seen for structure IIa with the difference that the decrease in distance occurs after 25 ns but still leading to the same mean distance (Figure 9.18-B). This occurrence appears between 10-30 ns for IIb thus showing a slightly higher mean of 5.8 Å and only twice with a duration of 2 ns in structure IIc explaining the mean distance of 8.4 Å (Figure 9.18-C and D). In contrast to these findings distances which allow a direct or a water-mediated proton transfer are not found if RL1780's is present. In case of RL2132 a water-mediated proton transfer appears possible during 25-40 ns in structure Ib as well as occasionally in IIIc (Figure 9.19-A and Figure 9.19-B).

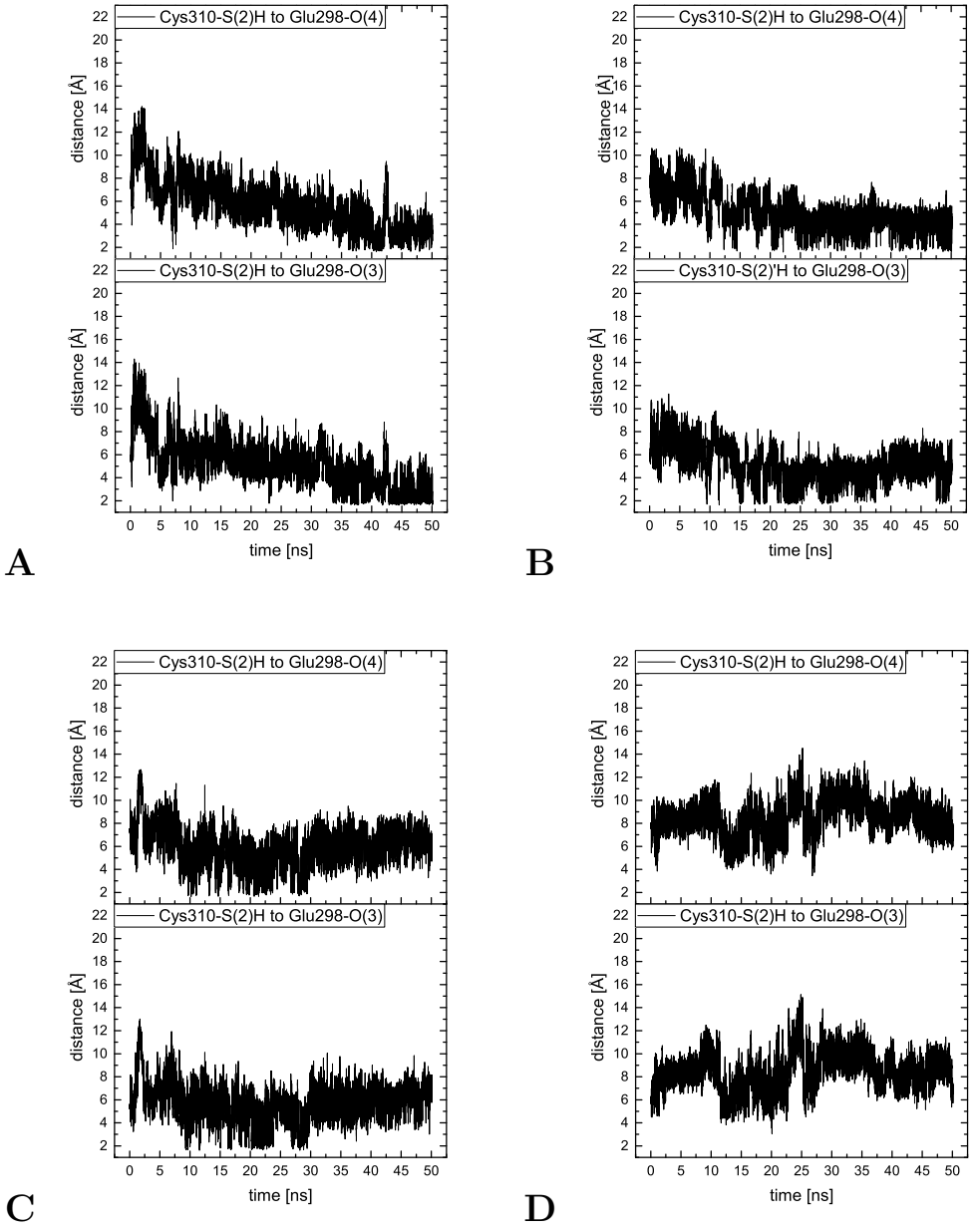


Figure 9.18: The variation in distance as a function of simulation time for Glu298 with Cys310 in presence of RL1784 in structure Ib (**A**), in structure IIa (**B**), in structure IIb (**C**) and in structure IIc (**D**), corresponding to Figure 9.17.

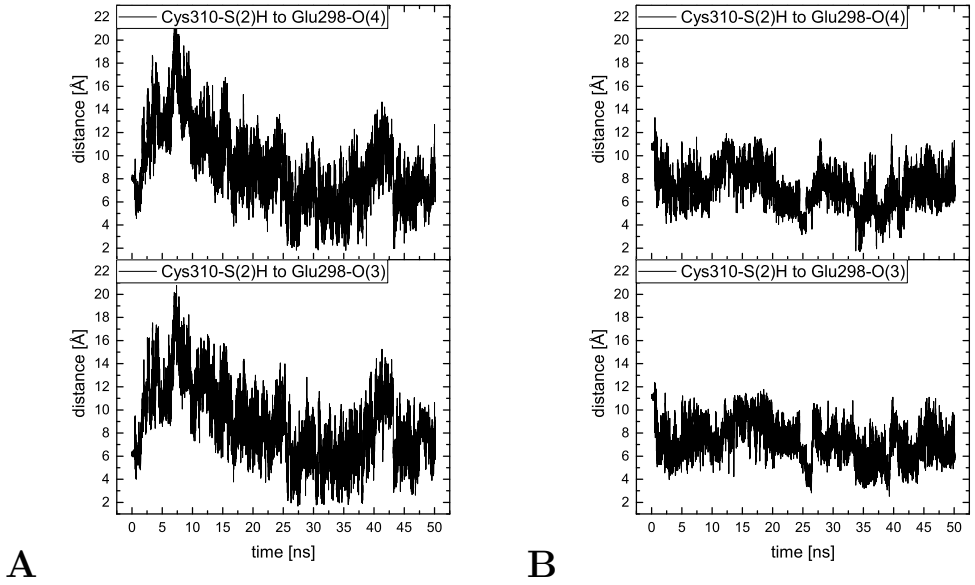


Figure 9.19: The variation in distance as a function of simulation time for Glu298 with Cys310 in presence of RL2132 in structure Ib (**A**) and in structure IIc (**B**) corresponding to Figure 9.17.

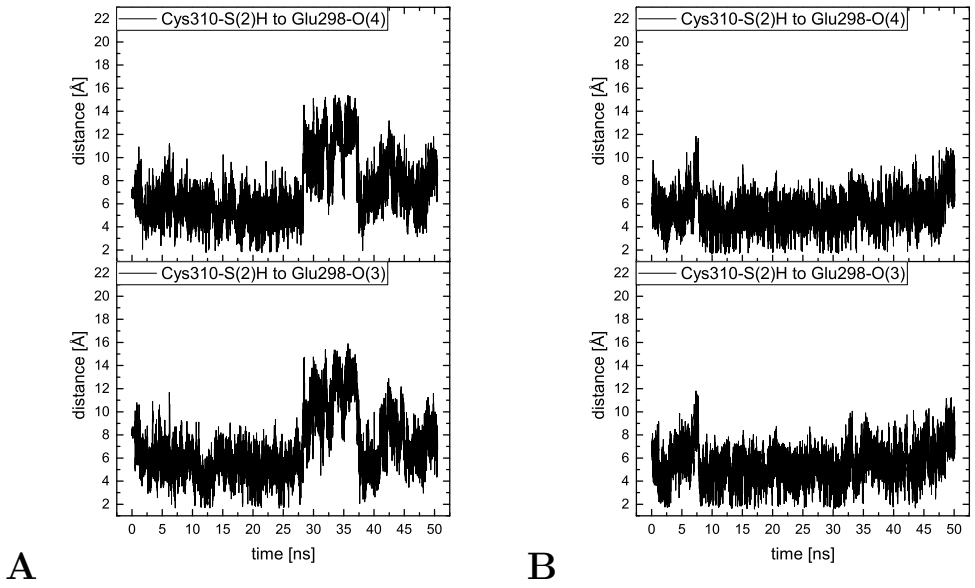


Figure 9.20: The variation in distance as a function of simulation time for Glu298 with Cys310 in absence of RL1784, RL1780 or RL2132 in structure IIa (**A**) and in structure IIb (**B**) corresponding to Figure 9.17.

In absence of the inhibitor, only in the cases IIa, IIb and occasionally in IIc a proton transfer is to be expected. IIa shows the largest mean for these structures due to the increase from ≈ 5 to 12 Å between 28 – 38 ns (Figure 9.20-A). For IIb the interaction between Cys310 and Glu298 is stable during the course of simulation thus resulting in a small deviation of 1.5 Å (in comparison to IIa or IIc) and a mean of 5.5 Å (Figure 9.20-B).

Table 9.6: Calculated mean and standard deviation of the distance between Cys310(S2) to Glu298(O3) or Glu298(O4) in protein kinase AKT-1. Mean and standard devations are calculated from 10,000 data points. All values in Å.

pdb	**distances**	**Cys310-S(2) to Glu298-(O3)***	**Cys310-S(2) to Glu298-(O4)***
Ia	**mean / SD**	11.0 / 1.7	11.0 / 1.7
Ib	**mean / SD**	5.3 / 2.2	5.7 / 2.2
Ic	**mean / SD**	11.7 / 2.9	11.8 / 2.9
IIa	**mean / SD**	5.3 / 1.5	5.1 / 1.5
IIb	**mean / SD**	5.8 / 1.7	6.0 / 1.8
IIc	**mean / SD**	8.4 / 1.6	8.4 / 1.6
pdb	**distances**	**Cys310-S(2) to Glu298-(O3)****	**Cys310-S(2) to Glu298-(O4)****
Ia	**mean / SD**	10.6 / 1.6	10.3 / 1.6
Ib	**mean / SD**	10.2 / 2.1	9.9 / 1.8
Ic	**mean / SD**	9.6 / 2.1	9.5 / 2.1
IIa	**mean / SD**	6.4 / 1.0	6.7 / 1.0
IIb	**mean / SD**	6.7 / 1.2	6.8 / 1.2
IIc	**mean / SD**	13.5 / 1.4	13.6 / 1.4
pdb	**distances**	**Cys310-S(2) to Glu298-(O3)*****	**Cys310-S(2) to Glu298-(O4)*****
Ia	**mean / SD**	8.7 / 1.7	8.5 / 1.8
Ib	**mean / SD**	9.7 / 1.4	9.9 / 1.4
Ic	**mean / SD**	11.4 / 2.3	11.4 / 2.3
IIa	**mean / SD**	6.9 / 2.7	7.0 / 2.6
IIb	**mean / SD**	5.5 / 1.5	5.5 / 1.5
IIc	**mean / SD**	6.7 / 1.7	6.7 / 1.6

*MD simulation in presence of RL1784. ** MD simulation in presence of RL1780. *** MD simulation in absence of inhibitor.

Table 9.7: Calculated mean and standard deviation of the distance between Cys310(S2) to Glu298(O3) or Glu298(O4) in presence of inhibitor RL2132 in protein kinase AKT-1. Mean and standard devations are calculated from 10,000 data points. All values in Å

pdb	distances	Cys310-S(2) to Glu298-(O3)	Cys310-S(2) to Glu298-(O4)
Ia	**mean / SD**	7.7 / 1.0	7.6 / 1.1
Ib	**mean / SD**	8.9 / 3.2	9.0 / 3.3
Ic	**mean / SD**	13.1 / 3.7	13.1 / 3.7
IIIa	**mean / SD**	7.7 / 1.2	7.8 / 1.2
IIIb	**mean / SD**	8.7 / 1.7	8.7 / 1.7
IIIc	**mean / SD**	7.3 / 1.6	7.3 / 1.7

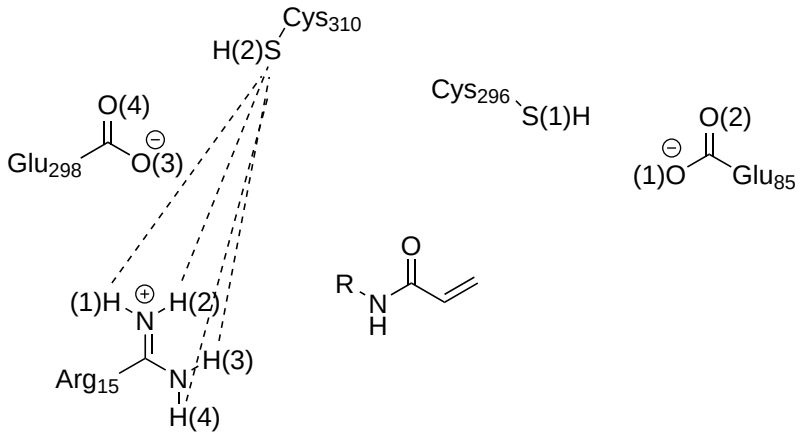


Figure 9.21: Interactions of active site Cys310 with Arg15 depicted with dashed lines, corresponding to the following distance plots. Location of amino acids and warhead serves for clarification and does not reflect actual position in MD simulations.

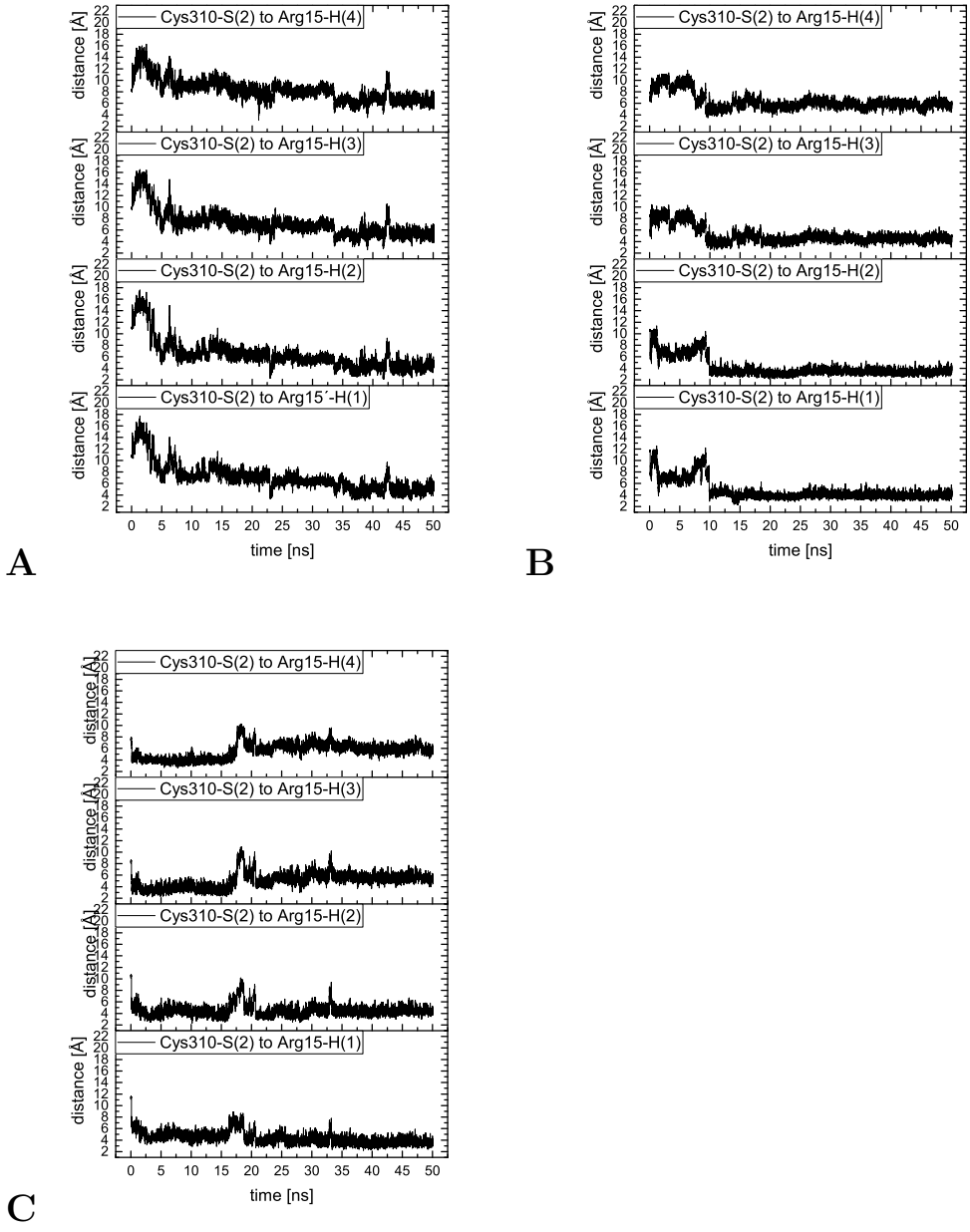


Figure 9.22: The variation in distance as a function of simulation time for Arg15 with Cys310 in presence of RL1784 in structure Ib (**A**), in structure IIa (**B**) and in structure IIb (**C**) corresponding to Figure 9.21.

Additional to Glu298 interactions further ones are formed between Arg15 to Cys310 during the simulations. Again, all hydrogens of Arg15's guanidine group have been considered to account possible rotations of the side chain. Interestingly a relationship between Arg15 and Glu298 are sometimes found. For example, one can observe a similar behavior as in struc-

ture Ib in complex with Rl1784, as a steady decrease is visible as before (Figure 9.22-A). For IIa the distance also lowers to values between 3.5 and 4.5 Å after 10 ns (Figure 9.22-B). For IIb the same values are observable. Only during 15-20 ns an increase to 8 Å occurs (Figure 9.22-C). In presence of inhibitor RL1780 only in structure IIc an interaction between Arg15 and Cys310 is formed revealing the lowest calculated mean in distance of 3.0 Å (Figure 9.23) while no interactions can be observed for the simulations of the protein-inhibitor complex of RL2132. In absence of an inhibitor distances between 3.5 and 4.5 Å are shown for the structure Ib and twice for Ic during 5 to 6 ns and 20 to 25 ns (Figure 9.24-B and C). In case of the structures IIa – IIc distances between 8 to 16 Å are observed (Figure 9.24-D to F).

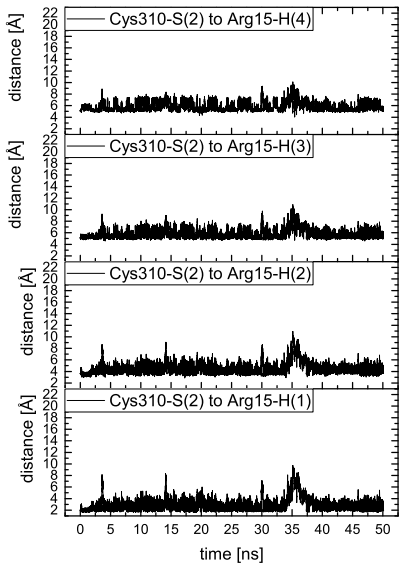


Figure 9.23: The variation in distance as a function of simulation time for Arg15 with Cys310 in presence of RL1780 in structure IIc corresponding to Figure 9.21.

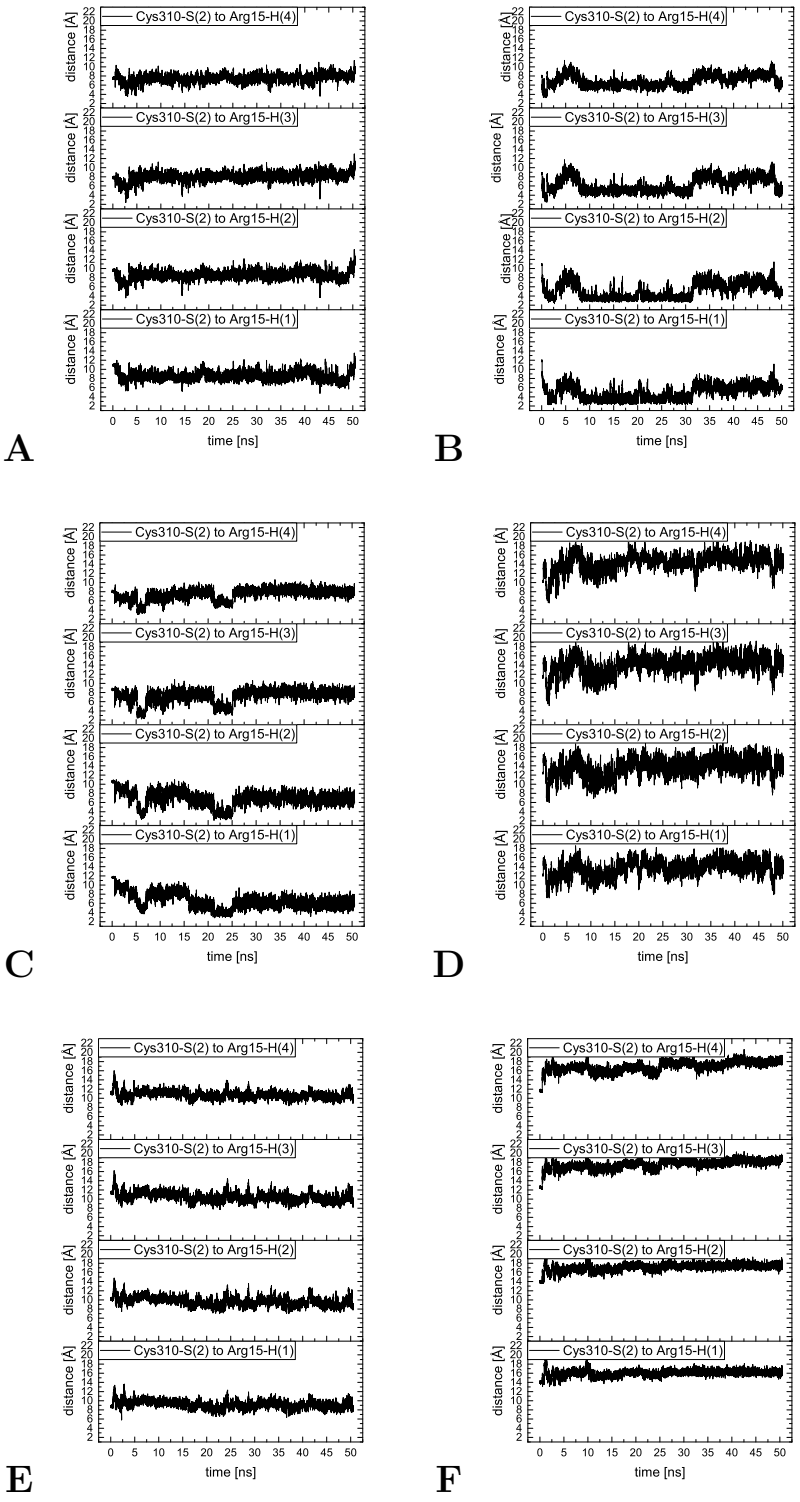


Figure 9.24: The variation in distance as a function of simulation time for Arg15 with Cys310 in absence of RL1784, RL1780 or RL2132 in structure Ia (**A**), in structure Ib (**B**), in structure Ic (**C**), in structure IIa (**D**), in structure IIb (**E**) and in structure IIc (**F**), corresponding to Figure 9.21.

Table 9.8: Calculated mean and standard deviation of the distance between Cys310(S2) to Arg15-H(1), Arg15-H(2), Arg15-H(3) or Arg15-H(4) in protein kinase AKT-1. Mean and standard devations are calculated from 10,000 data points. All values in Å.

pdb	**distances**	**Cys310-S(2) to Arg15-H(1)***	**Cys310-S(2) to Arg15-H(2)***	**Cys310-S(2) to Arg15-H(3)***	**Cys310-S(2) to Arg15-H(4)***
Ia	**mean / SD**	11.8 / 1.8	12.0 / 2.0	9.7 / 1.7	10.9 / 1.9
Ib	**mean / SD**	7.0 / 2.4	6.3 / 2.6	7.1 / 2.2	8.3 / 1.9
Ic	**mean / SD**	16.5 / 2.3	16.4 / 2.1	16.3 / 2.0	16.4 / 1.9
IIa	**mean / SD**	4.8 / 1.8	4.1 / 1.7	5.1 / 1.5	6.3 / 1.4
IIb	**mean / SD**	4.4 / 1.1	4.5 / 1.1	5.1 / 1.4	5.6 / 1.4
IIc	**mean / SD**	7.5 / 1.4	7.1 / 1.4	6.9 / 1.2	7.3 / 1.3

pdb	**distances**	**Cys310-S(2) to Arg15-H(1)****	**Cys310-S(2) to Arg15-H(2)****	**Cys310-S(2) to Arg15-H(3)****	**Cys310-S(2) to Arg15-H(4)****
Ia	**mean / SD**	11.5 / 1.8	12.2 / 2.1	14.0 / 2.2	14.7 / 2.0
Ib	**mean / SD**	12.4 / 2.5	12.5 / 2.4	12.6 / 2.1	12.5 / 1.7
Ic	**mean / SD**	12.3 / 2.7	12.3 / 2.7	12.9 / 3.0	13.4 / 3.0
IIa	**mean / SD**	6.7 / 1.1	6.3 / 1.0	7.6 / 1.0	8.9 / 1.0
IIb	**mean / SD**	8.4 / 1.1	8.0 / 1.3	8.9 / 1.4	9.8 / 1.2
IIc	**mean / SD**	3.0 / 1.4	4.5 / 1.3	5.9 / 1.1	5.9 / 1.0

pdb	**distances**	**Cys310-S(2) to Arg15-H(1)*****	**Cys310-S(2) to Arg15-H(2)*****	**Cys310-S(2) to Arg15-H(3)*****	**Cys310-S(2) to Arg15-H(4)*****
Ia	**mean / SD**	8.5 / 1.0	8.5 / 1.0	8.0 / 1.0	7.4 / 0.9
Ib	**mean / SD**	4.9 / 1.6	5.2 / 1.8	6.3 / 1.6	7.0 / 1.2
Ic	**mean / SD**	6.5 / 1.7	6.8 / 1.6	7.3 / 1.4	7.4 / 1.2
IIa	**mean / SD**	13.6 / 1.8	13.6 / 2.1	13.9 / 2.1	14.2 / 2.0
IIb	**mean / SD**	9.2 / 0.9	9.6 / 1.0	10.5 / 1.0	10.7 / 0.8
IIc	**mean / SD**	16.0 / 0.7	17.1 / 0.8	17.5 / 1.0	16.9 / 1.2

*MD simulation in presence of RL1784. ** MD simulation in presence of RL1780. *** MD simulation in absence of inhibitor.

Table 9.9: Calculated mean and standard deviation of the distance between Cys310(S2) to Arg15-H(1), Arg15-H(2), Arg15-H(3) or Arg15-H(4) in presence of inhibitor RL2132 in protein kinase AKT-1. Mean and standard devations are calculated from 10,000 data points. All values in Å.

pdb	distances	Cys310-S(2) to Arg15-H(1)	Cys310-S(2) to Arg15-H(2)
Ia	**mean / SD**	13.3 / 1.3	13.4 / 1.3
Ib	**mean / SD**	13.2 / 2.1	13.3 / 2.0
Ic	**mean / SD**	14.3 / 1.6	15.0 / 1.8
IIIa	**mean / SD**	8.9 / 0.8	8.2 / 0.9
IIIb	**mean / SD**	9.3 / 1.4	9.0 / 1.4
IIIc	**mean / SD**	8.9 / 1.0	8.5 / 1.2

pdb	distances	Cys310-S(2) to Arg15-H(3)	Cys310-S(2) to Arg15-H(4)
Ia	**mean / SD**	13.1 / 1.5	12.8 / 1.8
Ib	**mean / SD**	14.3 / 1.9	15.1 / 1.9
Ic	**mean / SD**	15.6 / 1.9	15.3 / 1.7
IIIa	**mean / SD**	7.7 / 0.8	7.9 / 0.7
IIIb	**mean / SD**	9.1 / 1.5	9.4 / 1.3
IIIc	**mean / SD**	8.8 / 1.3	9.3 / 1.1

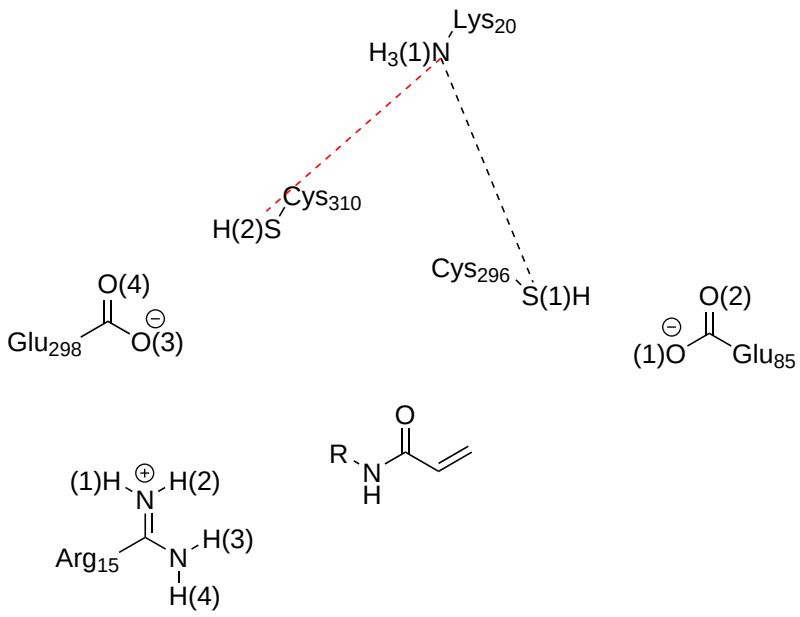


Figure 9.25: Interactions of active site Cys296 and Cys310 with Lys20 depicted with dashed lines, corresponding to the following distance plots. Location of amino acids and warhead serves for clarification and does not reflect actual position in MD simulations.

The analysis of the MD simulation of RL2132 in IIIa, IIIb and IIIc revealed another interaction between Cys296 or Cys310 and the basic amino acid Lys20. As one can see in Figure 9.26-D to Figure 9.26-F distances of around 5 Å are observable leading to the conclusion of an interaction between these residues. In all cases the mean distances between Cys310 to Lys20 are lower than between Cys296 and Lys20 (Table 9.10). This kind of interaction does not appear for the structures Ia, Ib or Ic (in Figure 9.26-A to Figure 9.26-C). It has to be mentioned, that all simulations considered a protonated Lys20 which cannot be involved in the deprotonation of cysteine. Still, interactions from lysine to cysteine can be relevant as as a stabilization of the thiolate after proton transfer can result.

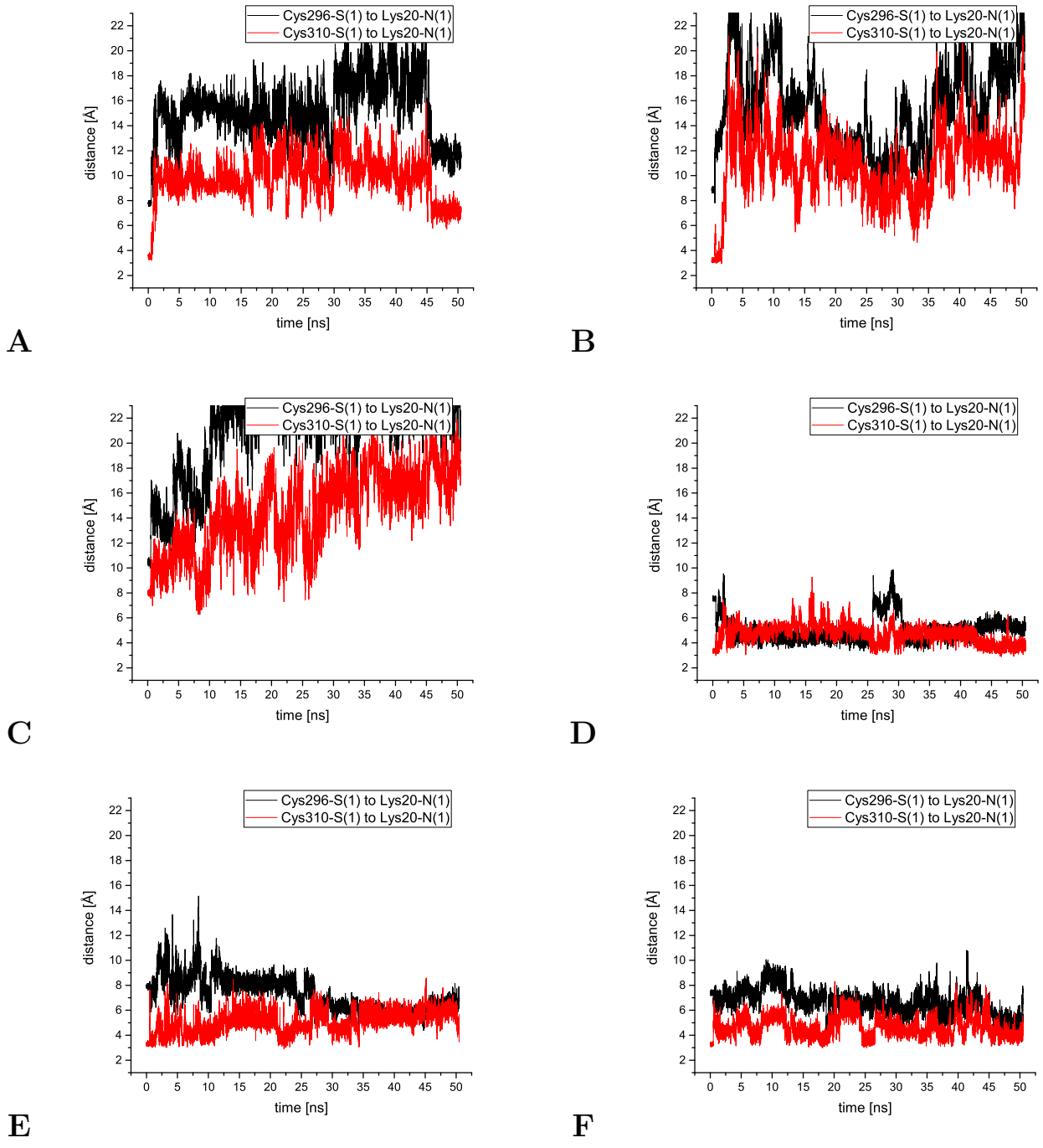


Figure 9.26: The variation in distance as a function of simulation time for Lys20 with Cys296 (black curve) or Cys310 (red curve) in presence of RL2132 in structure Ia (**A**), in structure Ib (**B**), in structure Ic (**C**), in structure IIIa (**D**), in structure IIIb (**E**) and in structure IIIc (**F**), corresponding to Figure 9.25.

Table 9.10: Calculated mean and standard deviation of the distance between Cys296-S(1) or Cys310-S(2) to Lys20-N(3) in presence of RL2132 in protein kinase. Mean and standard devations are calculated from 10,000 data points. All values in Å.

pdb	distances	Cys296-S(1) to Lys20-N(3)	Cys310-S(2) to Lys20-N(3)
Ia	**mean / SD**	15.3 / 2.5	9.9 / 1.7
Ib	**mean / SD**	15.5 / 3.6	10.9 / 2.7
Ic	**mean / SD**	21.7 / 3.9	14.5 / 3.2
IIIa	**mean / SD**	5.1 / 1.0	4.7 / 0.7
IIIb	**mean / SD**	7.3 / 1.4	5.1 / 0.9
IIIc	**mean / SD**	6.7 / 1.0	4.7 / 0.8

9.5 Conclusions

The group of Rauh from the Technical University Dortmund crystallized three different covalent-allosteric inhibitors (RL1784, RL1780 or RL2132) in covalent complex with AKT, a protein which is involved in the regulation cellular functions (Figure 9.3). This chapter comprises investigations these inhibitors in AKT1 and focused on two different questions.
Crystal structures revealed a covalent modification of Cys296 in case of RL1784 while the other two were covalently bound to Cys310. In case of RL1784 and RL1780 mass spectrometry indicated that both residues are modified by these inhibitors. However, these measurements cannot differentiate which residue is more active. Thus, the question arose which nearby cysteine residue is more likely modified. Therefore, MD simulations of the non-covalent complex have been conducted to analyze the variation of distance between both cysteine residues and the warhead. Experimentally these investigations cannot be performed because only information about the covalent enzyme-inhibitor are available. If a small distance towards the cysteine residue can be maintained for a longer period of time one can conclude that the respective residue is more likely modified.
Another point of interest is the deprotonation process of these cysteine residues which is needed to enable the covalent reaction. In case of Cys296 the crystal structure indicate that Glu85 is relevant for the abstraction of the proton. Hence, MD simulations in complex with the inhibitors as well as in absence of the inhibitors have been conducted and residues have been identified which might be responsible for the deprotonation.
Initially, it was necessary to build up the whole protein using theoretical approaches as many residues were missing in the crystal structure. The Modeller 9.18 program package which added the missing residues lead to different protein structures (Figure 9.5to Figure 9.7).

These structures were used for Molecular Dynamics simulation.
The results of these simulation offer a qualitative answer in respect to the modified cysteine residue. In the course of simulation one can observe lower values for the distance from RL1784 to Cys296 for the generated structures Ia-Ic compared to IIa-IIc. The mean distance to Cys296 is always lower as to Cys310. Still one can occasionally see appropriate distances for covalent modification of Cys310 although this does not occur as often as for Cys296. This explains the possible covalent bond to Cys310. It also explains why in the crystal structure a covalent bond is only found between RL1784 and Cys296.
For RL1780 the mean and standard deviation of the distance between inhibitor and Cys310 is lower for the structures IIa to IIc. This indicates that the formation of a covalent complex is more likely than for RL1784 with Cys296 in their initial crystal structures. Additionally, no close distance to Cys296 was found. This is not the case for the simulations in the structures Ia to Ic. In this connection RL1780 is farther away to Cys310 whereas also the frequency of a proper distance is lower than before. It appears that a certain amount of time is of necessity until the warhead is directed towards Cys310. Especially in one case Cys296 forms a permanent interaction with the inhibitor. This also is in line with the mass spectrometry revealing that a modification of Cys296 is also possible. Still, for the other structures the distance to Cys296 is not sufficiently close or appears infrequently compared to Cys310. These findings indicate that for RL1780 a modification of Cys310 is more likely.
In case of RL2132 distances for covalent modification of Cys310 are obtained in the structures IIIa, IIIb and IIIc. For IIIa fluctuations are low (standard deviation of 0.5 Å) in terms of the distance between Cys310 and the inhibitor. Still interactions between Cys296 and RL2132 are also observable with a slightly higher value in distance which indicates possible modification of both residues. This cannot be observed for IIIb and IIIc as the nucleophilic sulfur of Cys296 is too far away. Additionally, the distance between Cys310 and the inhibitor show larger fluctuations in contrast to IIIa. The insertion of RL2132 into the environment of Ia, Ib and Ic leads to different results. For Ia an appropriate distance for modification of Cys310 can be maintained for 25 ns and afterwards an increase to 12 Å occurs. Dynamics of the inhibitor in Ib show infrequent appropriate distances for a covalent modification of Cys310 as well as for Cys296. A complete different picture is obtained for Ic as Cys296 is closer to RL2132 than Cys310. These results indicate that the modification of Cys310 is slightly preferred although results are not as clear as for RL1780 or RL1784.
The evaluation of the deprotonation of Cys296 or Cys310 also revealed new findings. In terms of the interaction of Glu85 with Cys296 it is not as strong as expected from the interaction profile of the crystal structure. Still, it has to be mentioned that distances were

observed where a water-based abstraction of Cys296's hydrogen might appear although the likelihood is not very high. For the cysteine residue 310 two interactions are observed in the course of simulation. Firstly, Glu298 oriented in several simulations towards Cys310 leading to distances where an abstraction appears possible. Secondly, Arg15 establishes interactions to Cys310 and occasionally to Glu298. A deprotonation of the cysteine residue might be possible, but it is also conceivable that the protonated arginine functions as a stabilizer for the proton transfer. Another interesting finding is the role of Lys20. Dynamics of AKT1 in non-covalent complex with RL2132 in the structures IIIa, IIIb and IIIc reveal interactions of Cys296 and Cys310 to Lys20. This was not observed in case of the other inhibitors. It becomes clear that this observation was only made by the previous mentioned structures as they were generated by the crystal structure of RL2132. Thus, one can conclude that Arg15 and especially Glu298 are of importance for the deprotonation process of Cys310 while Lys20 plays an important role for the protein-inhibitor complex of RL2132.

Chapter 10

The binding mode of non-covalent diphenylether inhibitors in FabI and InhA

10.1 Introduction

With the discovery of penicillin in the year 1929 by Fleming the history of humanity entered a turning point [231]. It became clear that diseases caused by bacterial infections were manageable. But already in the year 1940, before penicillin was available on the market, bacterial penicillinases have been identified which proved Flemings observation that several bacteria have not been inhibited by penicillin [232]. With the rational design of methicillin 1959, the problem of penicillinases seemed to be solved. Still, three years afterwards methicillin-resistant *Staphylococcus aureus* (MRSA) appeared and in consequence further multiantibiotic-resitant variants occurred so that the abbreviation MRSA is also referred to multidrug-resistant S. *aureus* [233].

This leads to the necessity of the development of new antibiotics. In the period of the 1950s to the 1970s, many classes of antibiotics as tetracyclines, macrolides, aminoglycosides or quinolone have been discovered, just to mention a few of them [234]. The mechanism of antibiotics is manifold, may it be by the disturbance of the protein, RNA, DNA synthesis or of the bacterial cell wall [235]. The increased development of resistance towards these drugs was problematic and still is today. MRSA was in the late 1960 endemic in hospitals but spread out rapidly in the population in the 1990s so that it is prevelant in the world. Today, we are confronted with community acquired (CA)-MRSA which shows increased virulence

and distinctive characteristics concerning the transmission [236]. This can be traced back by the large consumption of antibiotics in medicine but also in agricultural economics and animal husbandry. Accordingly, resistances rose in alarming heights [237]. Especially the appearance of resistances of last-resort antibiotic vancomycin by so called vancomycin resistant S. *aureus* (VRSA) underlines the urgency of the development of new antibiotics [238].
A lot of bacterial pathogenes which are associated with epidemics, developed a multidrug-resistant (MDR) form due to the overused application of antibiotics. MDR *Mycobacterium tuberculosis* is not only present in the Third World countries but also in industrial countries. Already in 1944 when streptomycin was introduced as drug targeting tuberculosis (TB; also known as "the great white plague"), mutated strains have been found which showed resistances against this antibiotic [233].
The treatment of TB often turns out as complicated and includes an intensive initial treatment of two months with a fourfold combination of isoniazid, rifampicin, pyrazinamide and ethambutol. Subsequently a 4-month period treatment with isoniazid and rifampicin has to be continued. In case of MDR infections different drugs (depending on the results of resistance at least four effective drugs have to be chosen) have to be applied in combination for a time period of 18 to 24 months [239], [240]. To win the fight against MDR strains new targets or concepts have to be developed.
Copeland et al. showed 2006 that an increase in the drug residence time (tr) enhance the pharmocodynamic activity and alow lower drug concentrations due to elimination processes [241], [242]. This was also confirmed by studies which investigated the inhibition of ENR reductases (E.C. 1.3.1.9) *via* triclosan and different diphenyl ether (DPE) compounds [243]. It was shown that an increase of t_r results in a contemporaneous improvement in inhibition. ENRs catalyze the last step in the elongation in the bacterial fatty-acid biosynthesis (FAS II pathway) and offer a good drug-target which is addressed by triclosan and also by isoniazid. The reduction of trans-2-acyl-ACP to a saturated acyl-ACP species is catalyzed by this enzyme, while in most cases the cofactor NADH or NADPH serves as reducing agent (Figure 10.1). While the mammalian FAS I synthase executes many functions, the FAS II system of bacteria, protozoa or plants consists of many individual enzymes which are equipped with a certain functionality and thus a selective inhibition is possible as it is not found in humans [244].

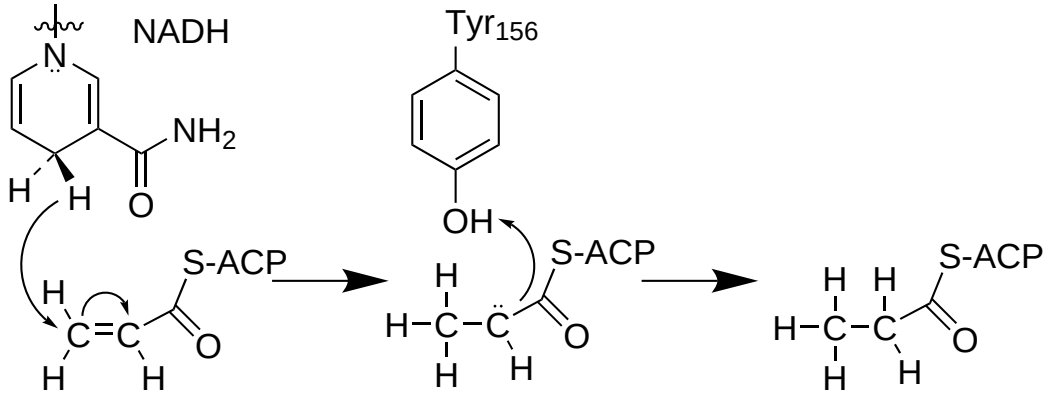


Figure 10.1: Reduction of double bond by addition of hydride ion of NADH (NADPH) of trans-2-butenoyl-ACP and abstraction of proton of Tyr156's phenol group in E. coli FabI. Proton of Tyr156's phenol group is restored through Lys163, hydroxyl groups of the ribose moiety of the coenzyme and the surrounding water molecules. Graphic is adapted from [245].

SKTS1 (Figure 10.2), a diphenyl ether (DPE) inhibitor which was originally designed for targeting the ENR of S. *aureus* (saFabI), was crystallized by the group of Kisker from the University of Würzburg in complex with saFabI and additionally with InhA, the ENR of M. *tuberculosis*. In terms of the binding position, SKTS1 occupies the same binding region in InhA as in FabI. Just like in case of FabI, Kiskers group showed that the measured kinetics indicate a favorable binding towards InhA. Although the binding mode of SKTS1 in InhA is similar to FabI, interactions resulting from the pyridone moiety are different. If one takes a look at the interactions of the pyridone ring of SKTS1 with Ala97 in FabI, distances between the inhibitors nitrogen or oxygen (of the pyridone ring) in to Ala97's backbone oxygen are different compared to SKTS1's interaction with Met98 in InhA (Figure 10.3). It appears that a tautomerization of the pyridone ring to a hydroxypyridine is of necessity to establish more favorable interactions with the protein environment in InhA in contrast to FabI. The interaction pattern obtained by the crystal structure between SKTS1 and the protein also indicates an interaction between Glu100 (for FabI) or Gln100 (for InhA) and the diphenylether inhibitor.

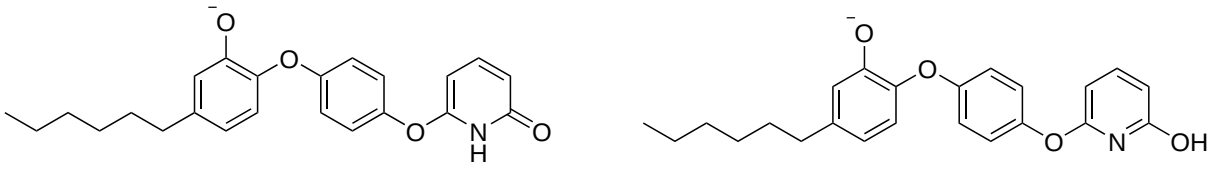


Figure 10.2: Pydridone form of SKTS1 (**left**) and hydroxypyridine form of SKTS1 (**right**).

As crystal structures do not contain hydrogen centers computational investigations are advantageous to shed light into the question of SKTS1's tautomeric form in the respective enzyme. QM calculations in vacuum and solution for the pyridone and hydroxypyridine ring were performed as well as for SKTS1 in both tautomeric forms to analyze the preference state of the inhibitor itself. Afterwards dynamics of SKTS1 were investigated via MD simulations with duration of 10 ns have been performed to investigate the different interactions to the protein environment depending on the tautomeric form. Additionally, a derivative of SKTS1 has been investigated by MD simulations to gain insight in regard of possible novel inhibitors. As this derivative consists of a 2-aminoimidazole instead of a pyridone ring it might show a better binding behavior as it is not affected by tautomerization.

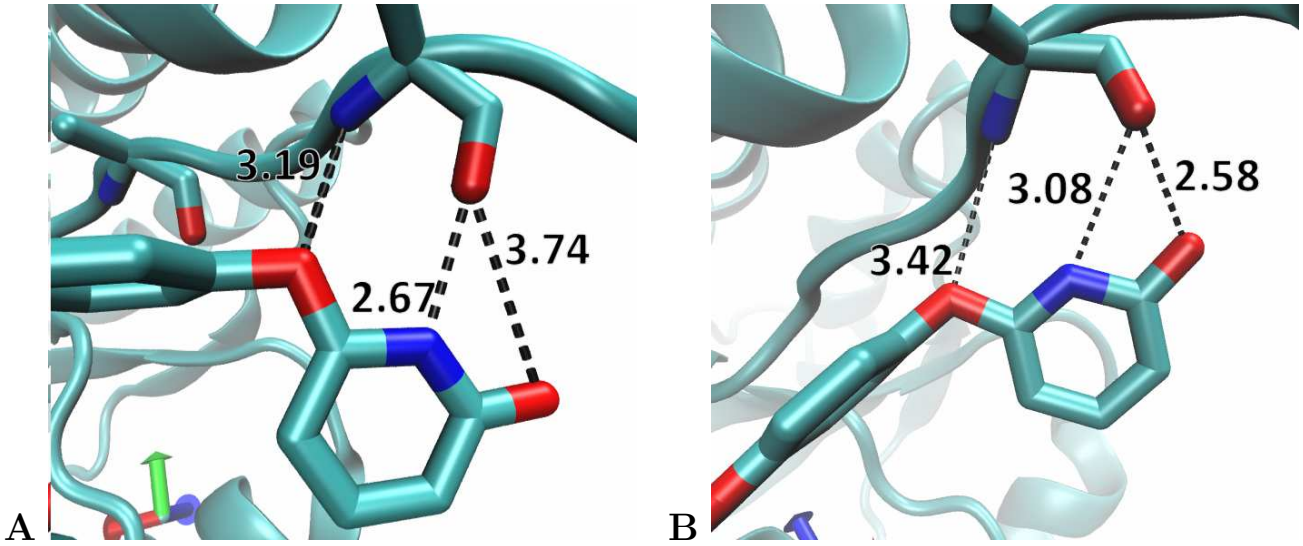


Figure 10.3: Interaction of the SKTS1 with Ala97's backbone in saFabI (**A**) or of SKTS1 with Met98's backbone (**B**).

10.2 QM calculations of pyridone and hydroxypyridine in vacuo and solution

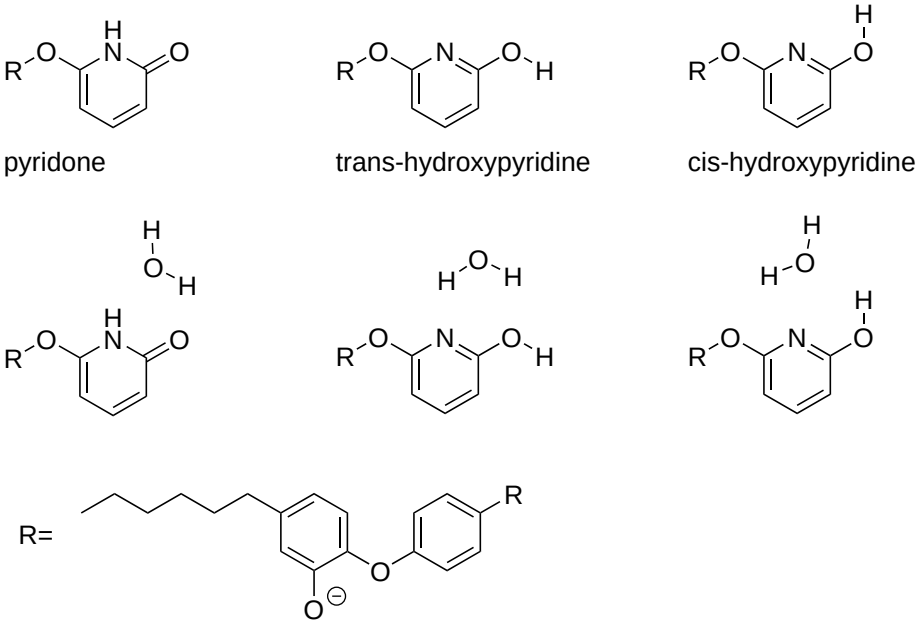


Figure 10.4: Structure of SKTS1 in pyridone or hydroxypyridine form. In case of hydroxypyridine trans- or cis-conformation was considered. Additionally, inclusion of explicit water molecule to pyridone, trans- or cis-hydroxypyridine is depicted for clarification.

As known from literature a hydroxypyridine ring is energetically favored in vacuum, whereas a pyridone is lower in energy in aqueous solution which has been shown experimentally or by QM calculations [246], [247], [248], [249]. To investigate the energetic difference of SKTS1 in its pyridone or hydroxypyridine form, firstly the methodology of the QM calculations has to be evaluated. In a first step only the pyridone or hydroxypyridine ring have been calculated in absence of the diphenylether scaffold to obtain qualitative comparability to results from literature. In contrast to the work of Galvao et al. the calculations were simplified and instead of the Gibbs energies only the difference in inner energy U was evaluated. After structure optimization employing B3LYP-D3 and the def-TZVP basis set in *vacuo* or with COSMO (ϵ=78,39), single point energy calculation utilizing different DFT or ab initio methods have been performed on top of these structures with the aug-cc-pVDZ basis set.
Table 10.1 shows that indeed the environment influences the preference of the tautomeric state. Except for the results obtained with B3LYP-D3 and SCS-CC2, cis-hydroxypyridine is favored by 1-2 kcal/mol whereas the inclusion of COSMO reveals that pyridone is lower in

energy by ≈ 2-6 kcal/mol depending on the used methodology. This is in indeed qualitatively concordant to the literature and therefore the same methodologies were applied to SKTS1. As shown in Figure 10.4 not only trans-hydroxypyridine but also cis-hydroxypyridine have been considered as well. Additionally, the system was extended by one explicit water molecule to further investigate environmental influence on the tautomeric state. The optimization and single point energy calculation procedure have been performed in the same manner as described before. Differences in structure due to optimization in vacuum or COSMO can be measured by a RMSD value of 0.24 Å (pyridone), 0.26 Å (trans-hydroxypyridine) or 0.26 Å (cis-hydroxypyridine). In terms of the inhibitor SKTS1, QM calculations in vacuum or in COSMO reveal that all methods (except for energies obtained with MP2/aug-cc-pVDZ for the cis-hydroxypyridine) predict that the pyridone form of SKTS1 is the energetically lower configuration compared to hydroxypyridine (see Table 10.2 and Table 10.3). Differences in terms of the presence or absence of COSMO regarding the optimization influence the obtained single point energy. By applying COSMO in the optimization step, energies obtained from the different single point calculations show a slight increase of energy by 0.5 kcal/mol or less for SKTS1's hydroxypyridine compared to its pyridone form. In general terms theory predicts that the pyridone form of SKTS1 is energetically more stable regardless of the used method or environment.

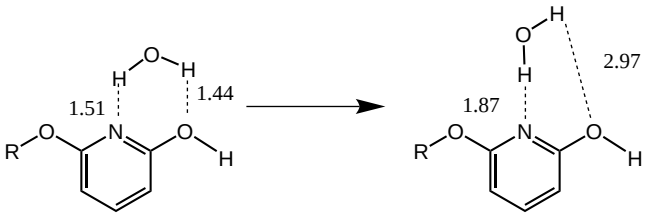


Figure 10.5: Interaction between SKTS1 in its hydroxypyridine form with explicit water molecule before (**left**) and after (**right**) optimization. All values in Å.

Concerning the cis-hydroxypyridine form DFT methods predict an increase of energy difference of 0.4 kcal/mol up to ≈ 2-3 kcal/mol if one employs the COSMO approach. The usage of ab initio methods leads to a decrease of about 0.5 kcal/mol up to ≈ 1 kcal/mol, thus a destabilization of the pyridone form. If one takes an explicit water molecule into account, the energetic difference between both tautomeric forms decreases by ≈ 1 kcal/mol but still describing the SKTS1's pyridone form to be more stable. For the trans-hydroxypyridine form of SKTS1 a stabilization of about 4 kcal/mol occurs if one employs implicit solvation for all used methods. In contrast to the cis-form, the inclusion of an explicit water molecule leads

to an increase and therefore destabilization of the trans-hydroxypyridine form which can be explained by Figure 10.5. The optimized structure shows an increased distance between the hydroxyl group of the ring and the hydrogen of the water and therefore an unfavorable interaction contrary to the results of cis-hydroxypyridine.
If one compares both forms for hydroxypyridine, one can see that the cis-form is about 5 kcal/mol lower than the trans-from in case of vacuum energies, while these differences decrease to $\approx$ 1 kcal/mol by inclusion of COSMO. These findings are also observable in MD simulation which will be shown in the next section. Also, another observation can be made: The usage of COSMO and inclusion of one explicit water molecule decreases the stability of SKTS1's pyridone tautomer whereas this trend is contrary to the simple pyridone ring. Thus, binding to InhA might be accompanied by increased constraints applied to the preferred tautomeric state of the inhibitor itself and may account for a less stable enzyme-inhibitor complex and shorter residence time.

Table 10.1: Single point energy difference of cis-hydroxypyiridine in relation to pyridone single point energy obtained from B3LYP-D3/def-TZVP optimized structures in vacuum and COSMO (ϵ=78,39). All calculations were performed using aug-cc-pVDZ basis sets. All energy values are shown in kcal/mol.

environment	**vacuum**	**COSMO**	**vacuum**	**COSMO**
B3LYP-D3	0.6*	5.8*	0.3**	6.4**
BHLYP-D3	-0.9*	4.7*	-1.3**	5.3**
MP2	-2.3*	2.3*	-2.3**	3.1**
CC2	-0.2*	4.2*	-0.1**	5.1**
SCS-MP2	-0.8*	3.7*	-1.0**	4.3**
SCS-CC2	0.8*	5.1*	0.8**	5.8**

* Single point calculation of B3LYP-D3/def-TZVP optimized structures in vacuum. ** Single point calculation of B3LYP-D3/def-TZVP optimized structures with COSMO.

Table 10.2: Single point energy difference of SKTS1's cis-hydroxypyridine in relation to SKTS1's pyridone single point energy obtained from B3LYP-D3/def-TZVP optimized structures in vacuum and COSMO (ϵ=78,39). All calculations were performed with aug-cc-pVDZ basis sets. All energy values are given in kcal/mol.

environment	vacuum	COSMO	vacuum	COSMO	vacuum/ H2O	COSMO/ H2O
B3LYP-D3	3.0*	3.1*	3.1**	3.3**	1.8 ***	2.4 ***
BHLYP-D3	1.8*	2.0*	1.9**	2.3**	0.9 ***	1.7 ***
MP2	-0.6*	-0.9*	-0.3**	-0.6**	-1.2 ***	-1.0 ***
CC2	1.3*	0.8*	1.7**	1.2**	0.5 ***	0.6 ***
SCS-MP2	0.8*	0.5*	1.0**	0.7**	0.1 ***	0.4 ***
SCS-CC2	2.3*	1.8*	2.6**	2.1**	1.5 ***	1.6 ***

* Single point calculation of B3LYP-D3/def-TZVP optimized structures in vacuum. ** Single point calculation of B3LYP-D3/def-TZVP optimized structures with COSMO. *** Single point calculation of B3LYP-D3/def-TZVP optimized structures with COSMO including explicit solvent molecule.

Table 10.3: Single point energy difference of SKTS1's trans-hydroxypyridine in relation to SKTS1's pyridone single point energy obtained from B3LYP-D3/def-TZVP optimized structures in vacuum and COSMO (ϵ=78,39). All calculations were performed with aug-cc-pVDZ basis sets. All energy values are presented in kcal/mol.

environment	vacuum	COSMO	vacuum	COSMO	vacuum/ H2O	COSMO/ H2O
B3LYP-D3	8.3*	4.4*	8.5**	4.6**	11.9 ***	4.6 ***
BHLYP-D3	7.3*	3.3*	7.6**	3.6**	11.4 ***	3.9 ***
MP2	4.7*	0.3*	5.2**	0.8**	9.1 ***	1.4 ***
CC2	6.6*	2.2*	7.1**	2.6**	10.9 ***	3.2 ***
SCS-MP2	6.1*	1.7*	6.5**	2.0**	9.9 ***	2.2 ***
SCS-CC2	7.6*	3.1*	8.1**	3.4**	11.4 ***	3.7 ***

* Single point calculation of B3LYP-D3/def-TZVP optimized structures in vacuum. ** Single point calculation of B3LYP-D3/def-TZVP optimized structures with COSMO. *** Single point calculation of B3LYP-D3/def-TZVP optimized structures with COSMO including explicit solvent molecule.

10.3 MD Simulation of SKTS1 in pyridone and hydroxypyridine form in complex with FabI or InhA

As the interest lies on the tautomeric state of SKTS1 in FabI or InhA molecular dynamic simulations have been performed. Aim of these simulations is to evaluate the difference in in-

teractions of the diphenylether inhibitor in each tautomeric form for both proteins compared to the X-ray structure. If the mean RMSD value of the pyridone unit is lower than that of the hydroxypyridine unit (or the other way around in case of InhA), one can assume that this would go along with the interaction pattern obtained from the crystal structure. As the focus was laid on the tautomeric forms, the RMSD of the pyridone or hydroxypyridone was solely calculated to decrease the interference of the whole structure. Additionally, the variation of distances during the MD simulation for interactions with the protein environment has been investigated to underpin results from the RMSD analysis.

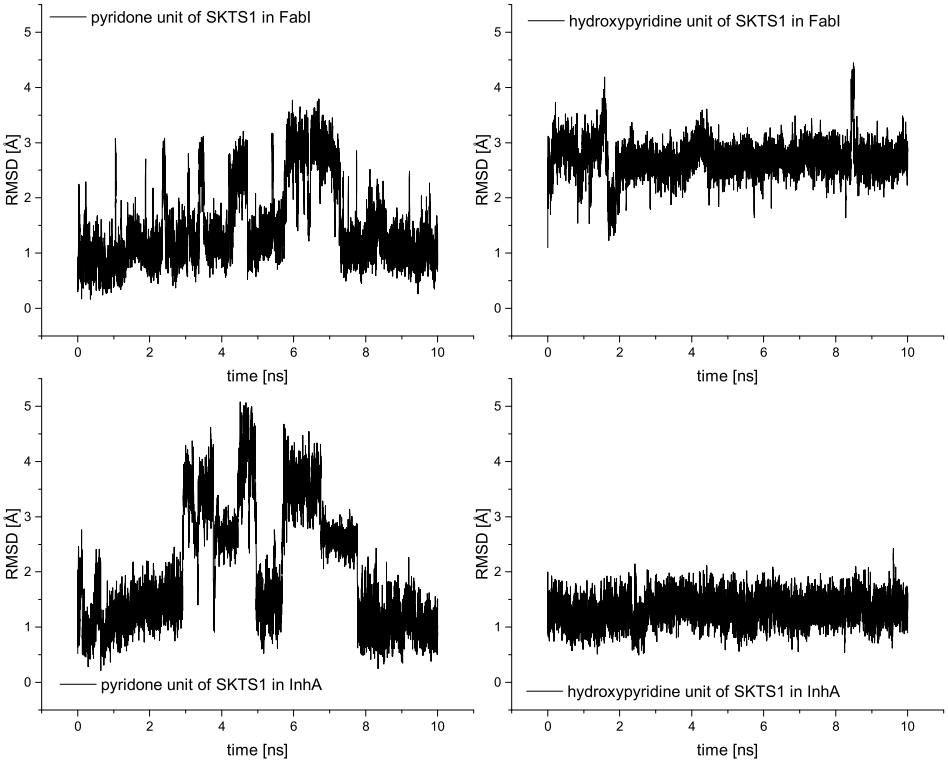


Figure 10.6: RMSD plot of pydridone or hydroxypyridine unit with ether-oxygen of SKTS1 in complex with FabI or InhA during the 10 ns production run. SKTS1 in FabI shows a mean of 1.47 Å (pyridone) or 2.71 Å (hydroxypyridine) whereas in InhA it has one of 2.05 Å (pyridone) or 1.32 Å (hydroxypyridine). Mean and standard devations are calculated from 10,000 data points.

As in the section 5.1 the proteins FabI or InhA were described by the Amber14 force field ff14SB while for SKTS1 (in its pyridone and hydroxypyridine form) and the respective cofactors ($NADP^+$ in FabI and NAD^+ in InhA) the generalized AMBER force field (GAFF) was employed. The preparation of the inhibitor-protein complex in terms of minimization and gradual heating were done in the same manner as described in section 5.1. The duration for each production run was 10 ns. Each tautomer of SKTS1 was simulated once in FabI or InhA resulting in a total simulation time of 40 ns.
In case of FabI it is expected that SKTS1 shows the most stable interaction with the protein environment in its pyridone form. In terms of the RMSD values obtained during the 10 ns simulation time, one can see that the pyridone unit shows a mean of 1.47 Å while the hydroxypyridine unit has a mean value of 2.71 Å (Figure 10.6) which is in line with experimental findings. The difference in RMSD values is small if one looks at the distances between the pyridone or hydroxypyridine ring and the surrounding amino acids.

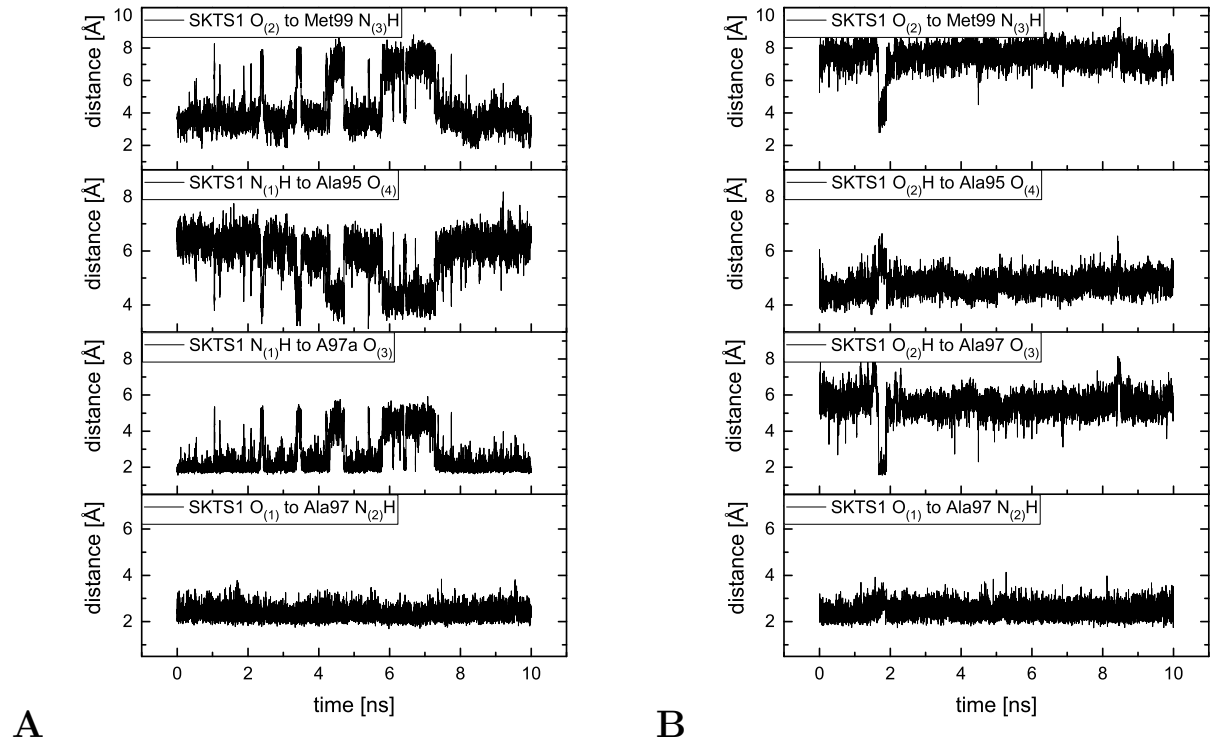


Figure 10.7: The variation in distance as a function of simulation time for SKTS1 in pyridone (**A**) or hydroxypyridine form (**B**) in FabI during 10 ns production run.

As one can see in Figure 10.7 the pyridone form of SKTS1 shows distances to Ala97

backbone NH of ≈ 2.5 Å or to Ala97's carbonyl oxygen with a distances of about 2 Å. Occasionally one can observe an increase of distance to Ala97's backbone carbonyl oxygen up to 4.5 Å due to a water-mediated interaction and in consequence a decrease to Ala95 by a rotation of the pyridone ring in the simulation time of ≈ 6 – 7.5 ns. Furthermore, a favorable water-mediated interaction to Met99 can be observed which also weakens during the just mentioned simulation time. Interestingly all these measured distances are in accordance to the RMSD plot. An increase in RMSD value is in relation to a weakening of interactions with Ala97 or Met99.

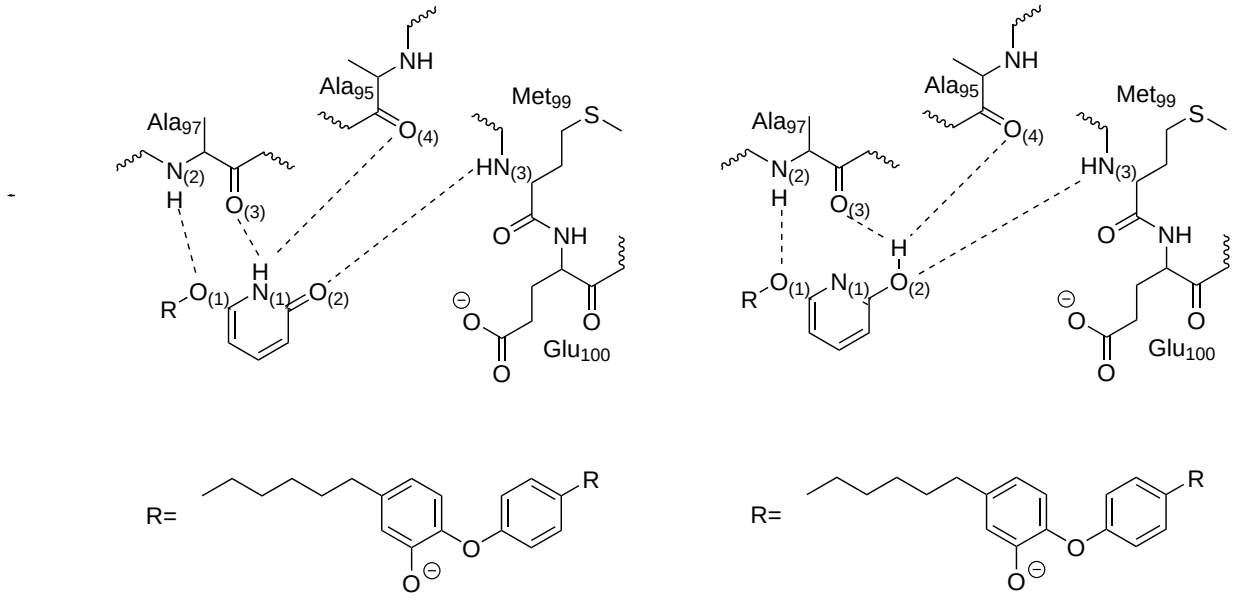


Figure 10.8: Interactions of pyridone (**left**) and hydroxypyridine (**right**) of SKTS1 in FabI depicted with dashed lines, corresponding to distance plots of Figure 10.7. Location of amino acids and SKTS1 serves for clarification and does not reflect actual position in MD simulations.

For the hydroxypyridone form SKTS1 shows quite the opposite behaviour. Interactions between Ala97's backbone carbonyl oxygen or to Met99 are only observable for a few hundreds of picoseconds while the hydroxypyridine ring is mostly rotated and thus directed towards Ala95. This kind of movement corresponds well to the increased RMSD values and therefore a different orientation as in the crystal structure can be observed. Only the interaction of SKTS1's ether group and A98 backbone NH are comparable for both tautomeric forms.

Contrary to the MD simulation of SKTS1 in FabI, the mean RMSD value of the hydroxypyridine unit of SKTS1 in InhA is lower (1.32 Å) than the calculated one for the pyridone unit (2.05 Å). This supports the assumption that a favorable binding of SKTS1 is enabled if it is in the hydroxypyridine form. Especially, one can observe fewer fluctuations with respect to the RMSD values compared to the pyridone tautomer of SKTS1 in FabI. This is understandable if one looks at the distance plots (Figure 10.7) of SKTS1 in InhA. As one can see for the hydroxypyridine form the interaction with Met98's backbone NH or carbonyl oxygen is quite pronounced. No fluctutations and thus no reorientation of the hydroxypyridine unit was observed during the 10 ns simulation time.

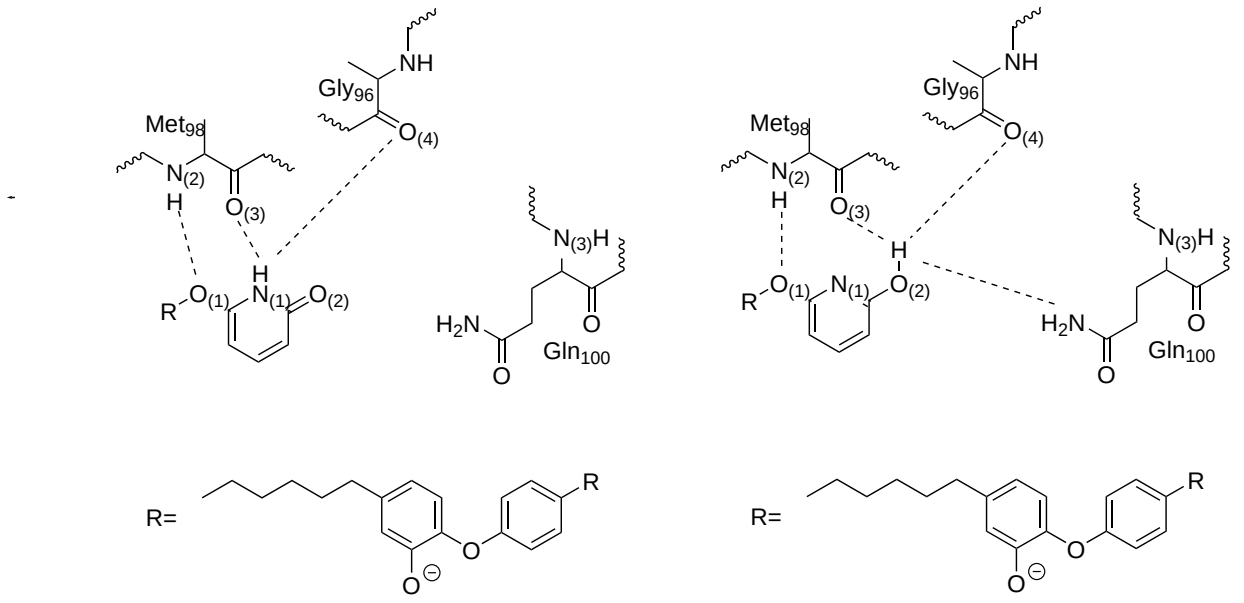


Figure 10.9: Interactions of pyridone (**left**) and hydroxypyridine (**right**) of SKTS1 in InhA depicted with dashed lines, corresponding to distance plots of Figure 10.10. Location of amino acids and SKTS1 serves for clarification and does not reflect actual position in MD simulations.

For the pyridone form in InhA one can observe a similar behavior to the hydroxypyridone form in FabI. During ≈ 3.5 to 8 ns simulation time the distance to Gly96 decreases due to a direct or water-mediated interaction which leads to a reorientation of the ring as observed for SKTS1's hydroxypyridine form in FabI to Ala97. This movement in distance also correlates with its RMSD values, as it is visualized within the RMSD plots given in Figure 10.6. Ad-

ditionally, the interaction of SKTS1's ether oxygen with Met98's backbone NH shows larger fluctuations although the distance itself is closer during the time course of simulation.

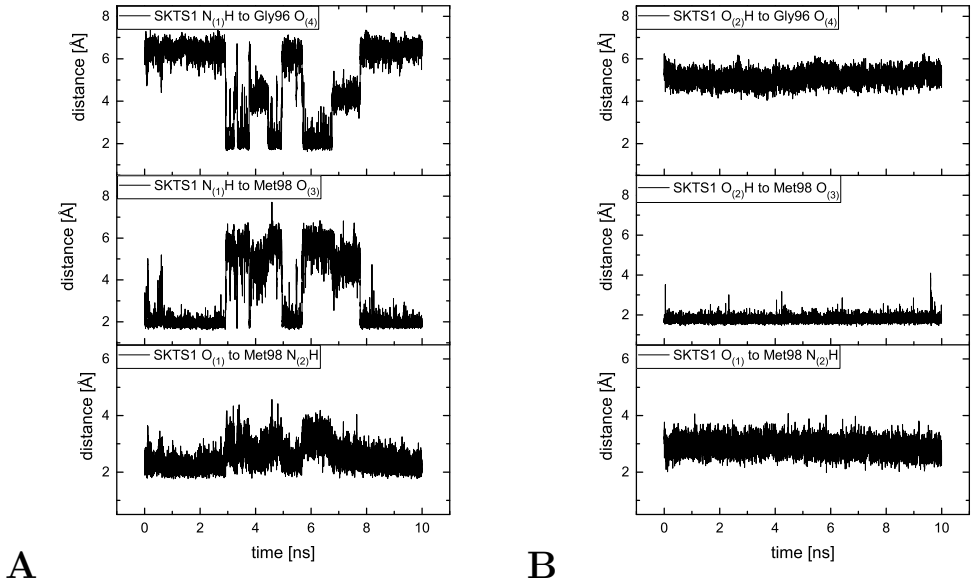
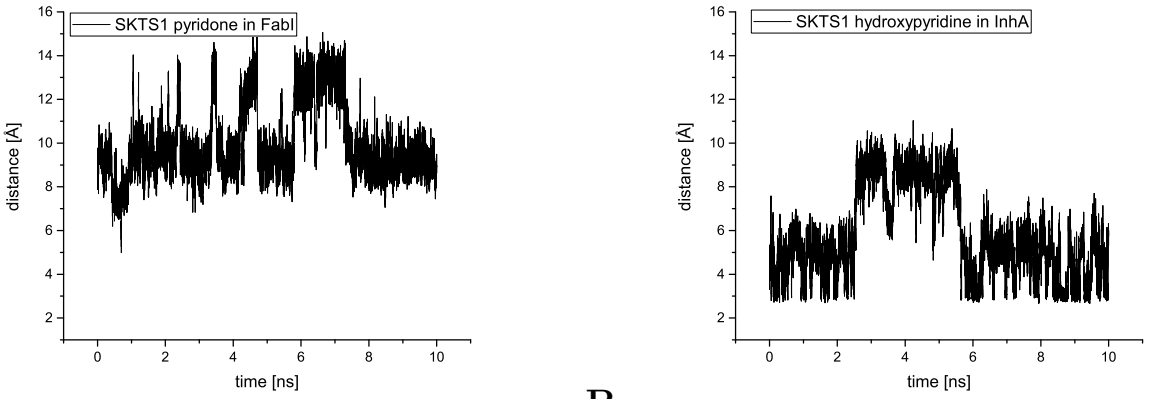


Figure 10.10: The variation in distance as a function of simulation time for SKTS1 in pyridone (**A**) or hydroxypyridine form (**B**) in InhA during 10 ns production run.

As it is now clarified in which tautomeric form SKTS1 is present in the respective protein, the next question regarding the interaction to Glu100 (FabI) or Gln100 (InhA) has been tackled. The distance plot of the MD simulation of SKTS1 in FabI (pyridone form) shows distances of ≈ 9 Å leading to the assumption that no interaction to Glu100 is existing (Figure 10.11 left). Furthermore, an increase of distance is observable up to 10 – 13 Å during 6-7 ns. On the other hand, SKTS1 in InhA (hydroxypyridine form) shows a water-mediated interaction to Gln100 which is interrupted during 2-6 ns (Figure 10.11-A). Indeed, one can assume that an interaction towards this amino acid is possible, although, one has to be aware of its strength as the distance increases.



A **B**

Figure 10.11: The variation in distance as a function of simulation time for SKTS1 in pyridone form and Glu100 in FabI (**A**) and for SKTS1 in hydroxypyridine form and Gln100 in InhA (**B**), during 10 ns production run.

As mentioned in the previous section an analysis in regard of the conformation of hydroxypyridine in FabI or InhA has been performed. Aim of it was to figure out if the cis or trans form is favored during the MD simulation. By measuring the dihedral of SKTS1 in its hydroxypyridine form during the MD simulation (red marked atoms in Figure 10.12) one obtains a plot which shows that the diphenylether inhibitor in InhA is solely present as cis-hydroxypyridine (dihedral of -10 to 40°) whereas rarely the trans-form (± 175°) can be observed in FabI (see dihedral plot in Figure 10.12).

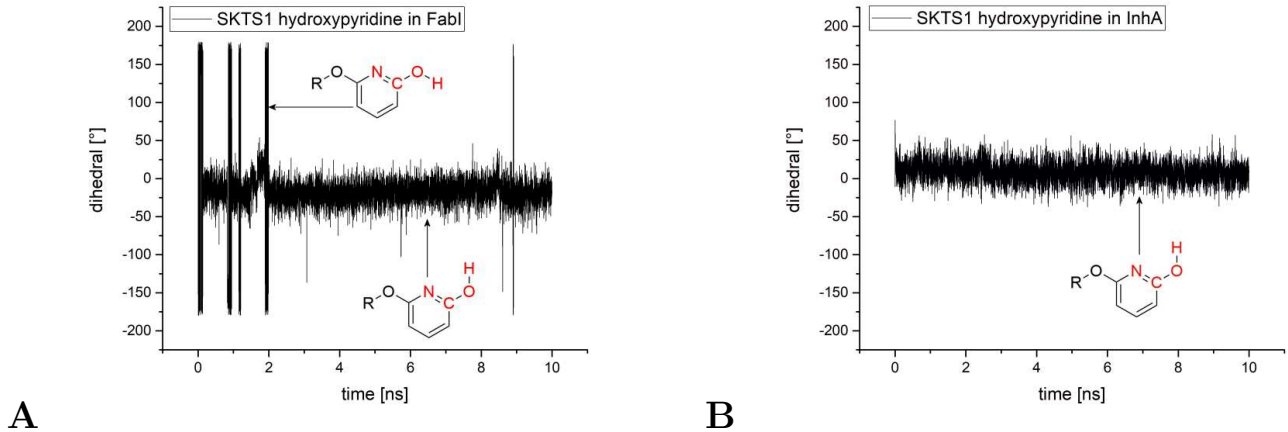


Figure 10.12: Dihedral plot of red marked atoms of SKTS1 in hydroxypyridine form in complex with FabI or InhA during a 10 ns MD simulation. In case of FabI occasionally dihedrals of $\approx \pm 175°$ can be seen, leading to trans-hydroxylpyridine. Dihedrals in the range of $\pm 50°$ result in the structure of cis-hydroxypyridine which mostly occur in FabI and solely in InhA.

10.4 MD Simulation of SKTS1 derivatives in InhA

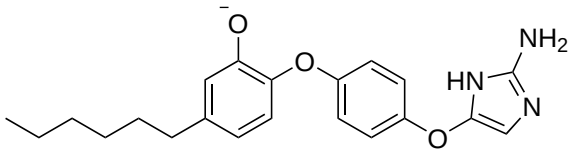


Figure 10.13: Structure of SKTS1 derivative on the basis of a diphenylether scaffold by substituting the pyridone-ring by a 2-aminoimidazol.

To extend the study in respect to diphenylether inhibitors the pyridone ring has been replaced by a 2-aminoimidazole group (depicted in Figure 10.14). Based on the crystal structure of the initial inhibitior in complex with InhA, SKT1 has been manually modified to obtain the derivative. This structure has also been tested in InhA by MD simulations to investigate its interaction with the protein environment and to make an assumption on its suitability as possible inhibitor. If one compares the measured distance to Met98 backbone NH with SKTS1, one can see that both structures show the same strength of interaction. Also, the imidazole NH group of the derivative reveals the same interaction towards Met98 backbone carbonyl-oxygen as the hydroxyl-group of SKTS1's hydroxypyridine ring.

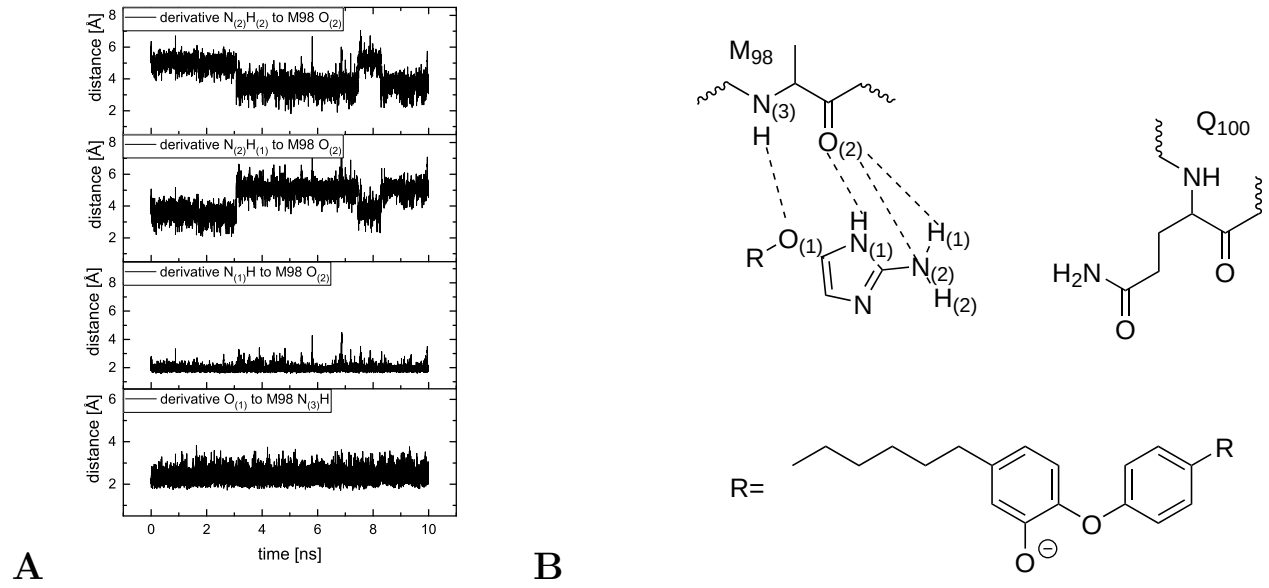


Figure 10.14: The variation in distance as a function of simulation time for 2-aminoimidazol derivative in InhA during 10 ns production run (**A**) and its corresponding interactions represented in dashed lines (**B**). Location of amino acids and 2-aminoimidazol derivative serves for clarification and does not reflect actual position in MD simulations.

Furthermore, the hydrogens of the amino group of the derivative form an additional interaction with Met98 backbone carbonyl-oxygen which appears occasionally water-mediated. Due to the rotation of the amine one can observe a switch with respect to its hydrogens. The observed increase in distance for one hydrogen results in a decrease of the other but still maintaining a favorable interaction to the carboxyl group.

10.5 Conclusions

In this chapter the tautomeric state of the diphenyl ether inhibitor SKTS1 was investigated which has been crystallized in the enoyl-acyl carrier proteins FabI of S. *aureus* and InhA of *Mycobacterium tuberculosis* (Figure 10.2). The crystal structures indicate that SKTS1 is in a pyridone form in FabI while it is in a hydroxypyridine form in InhA (Figure 10.3). Furthermore, interactions with Glu100 in FabI or Gln100 in InhA have been observed in the

X-ray structure. Hence, this study verified their relevance in terms of the inhibitor's binding. Finally, a 2-aminoimidazole derivative of SKTS1 was investigated in InhA and its interaction and stability in this protein was probed. The advantage of this derivative is the lack of tautomerization and thus this structure might offer a similar or more favorable binding behavior compared to SKTS1. In order to gain insight on these questions, QM calculations as well as MD simulations have been applied.

Firstly, QM calculations of the simple pyridone or hydroxypyridine ring were found in line with the literature that pyridone is more stable in polar solution while hydroxypyridine is more stable in vacuum. Contrary to that, computations predict that the pyridone form of SKTS1 is always lower in energy independent of the influence of the solvent. Still, one interesting trend can be observed. Solvent effects decrease the stability of the pyridone form of SKTS1 which is in contrast to the simple pyridone ring. Additionally, QM calculations revealed a favorable state of SKTS1's cis-hydroxypyridine compared to its trans-form which was also confirmed by MD simulation.

In terms of the tautomerization of SKTS1 in FabI or InhA, MD simulations of both tautomeric states in the respective protein have been conducted and results are in a good agreement with the interaction pattern obtained from the crystal structure. The question related to the most stable binding mode is underpinned by the findings in this study. In FabI the RMSD values of the pyridone tautomer show indeed a smaller mean RMSD value compared to the hydroxypyridine form. In the case of InhA the MD simulations also validate that SKTS1's hydroxypyridine form is related to be the best possible binding mode. Fluctuations in terms of the RMSD value are relative to interactions to the protein environment and are in accordance to the in- or decrease of distance to surrounding amino acids. Another conclusion concerning SKTS1's pyridone ring in FabI can be made. Due to its flexibility further interactions by rotation occur and therefore lead to fluctuations of RMSD values in the course of simulation time. This kind of behavior could not be observed for SKTS1 in InhA.

The expected interaction of SKTS1 to Glu100 in FabI could not be supported as the distance between the amino acid and the inhibitor was too large to assume an interaction. Opposite to this finding one can observe a water-based interaction between SKTS1 to Gln100 in InhA, although, a weakening in interaction occurred during the course of simulation.

The MD simulation of the 2-aminoimidazol derivative show promising results in respect to its interaction to the protein environment in InhA and are comparable to SKTS1 in its hydroxypyridine form. Especially the additional interaction due the amino-group might strengthen the interaction and therefore increase the duration of binding. Further theoretical investigations in terms of free binding energy or docking approaches and also inhibition experiments

are of need to validate the MD simulations. Still, these findings are encouraging and might offer an additional possibility for the development of an inhibitor for InhA which should be further investigated.

Chapter 11

The first steps in the development of a covalent inhibitor for FadA5 - a thiolase of mycobacterium tuberculose

11.1 Introduction

Tuberculosis (TB) exists for millenia and still represents a serious health problem which can occur all over the world. It belongs to the top ten causes of death in a global context and is the leading cause of death as an infectious agent for the past years [250]. Hence, the cases of deaths are even higher compared to HIV/AIDS ([251]. 6.3 millions of new cases of TB have been reported for 2016 and thus represent 61% of the estimated 10.4 million. New cases which occurred in 2016 are located in WHO South-East Asia region (45%), in WHO African Region (25%) and in WHO Western Pacific Region (17%). A smaller amount can be found in WHO Eastern Mediterranean Region (7%), in WHO European Region (3%) and in WHO Region of the Americas (3%). Globally, new treatment outcome data reveal a success rate off 83% [250]. In contrast to that, the 129.689 cases in which drug-resistant TB is involved show on global scale a success rate of only 54%. The treatment of HIV-associated TB was only in 78% successful (2015), while a significant smaller proportion of 30% showed a positive outcome in treatment (2014) [250]. New forms of diagnostic, drugs and medication plans are steadily developed, although, the process is a slow one.

Tb is an infectious disease which is caused by bacillus *Mycobacterium tuberculosis*. In most cases one has to deal with pulmonary TB which attacks the lungs. The disease is spread by e.g. coughing, whereby people who are infected with pulmonary TB distribute the airborne

bacteria. If organs other than the lung are affected, e.g. lymph nodes, skin, abdomen, genitourinary one refers to extra-pulmonary TB [252]. The development of TB disease occurs only in 5-1 % of the estimated 1.7 billion people who are infected with M. *tuberculosis*. People with diabetes who are infected with HIV or are affected by risk factors as alcohol, smoking or under-nutrition show a higher probability to develop TB disease.
The distribution occurs by inhalation of aerosol bacteria and leads to the infection cycle (Figure 11.1). The role of the bacteria's metabolism and its influence to the host lipid is not fully solved yet [253]. It is known that foamy macrophages which contain lipid droplets are concentrated in infected mouse lungs and are also concentrated in lung granulomas of infected humans [254], [255]. MTB is located close to lipid bodies in foamy macrophages and the bacteria accumulate presumably also in the intracellular lipid bodies [256]. It is possible that due to the high amount of cholesterol esters in foamy macrophages a binding to MTB is caused [257], [258]. In case of *mycobacterium leprae* it is known that one can observe an accumulation of cholesterol esters in infected macrophages [259].
Pandey et al. were able to show that MTB's cholesterol metabolism was of necessity and a disturbed cholesterol transport could hamper the virulence in the mouse model [260]. Dubnau et al. investigated genes which are over exprimed in human macrophages after infection by MTB and identified FadA5 which is up-regulated [261]. This lipidmetabolizing thiolase FadA5 is necessary to utilize cholesterol as carbon source and is ultimately responsible for the virulence of MTB in chronic lung infections of mice [253]. FadA5 plays a role in the β-oxidation of the cholesterol side chain and the thiolysis of acetoacetyl-coenzyme A (CoA) and consequently is needed in the production of androesterones (Figure 11.2). Hence it is suggested that the resulting products due to the metabolism of cholesterol in M. *tuberculosis* cause the persistence [253].

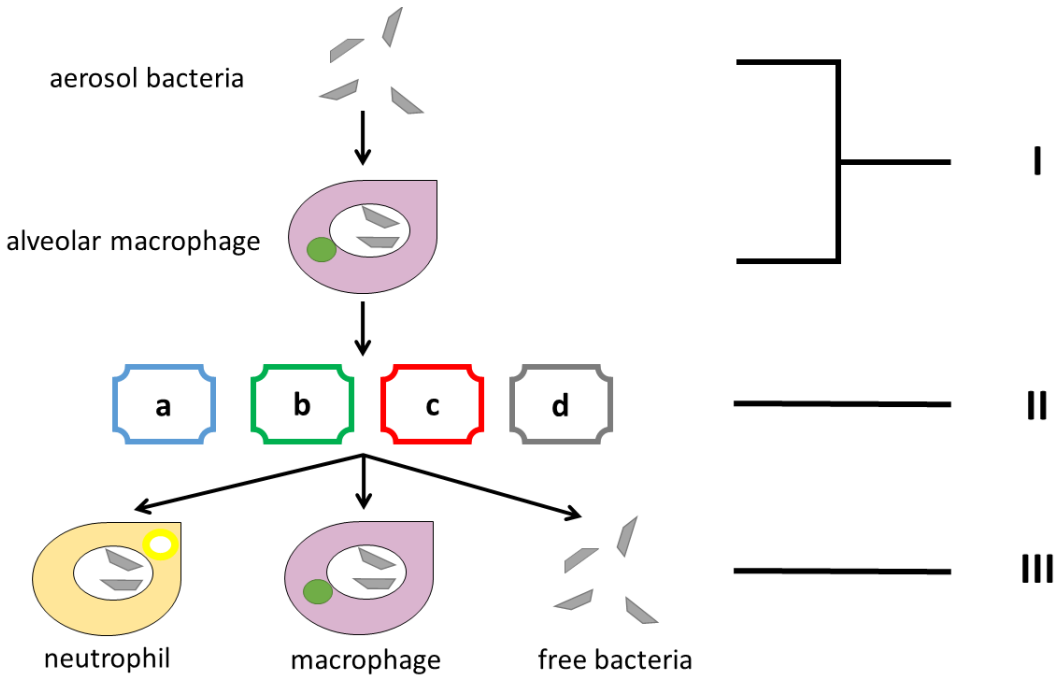


Figure 11.1: Infection and transmission of MTB adapted from [262]. I – The inhalation of bacteria leads to the absorption by alveolar macrophages in the lung. II – MTB causes proinflammatory reactions, which promotes the formation of macrophage-centric granuloma. They show a range of heterogenous forms during the latent or chronic phase during the infection in humans. III – Additionally, it is suggested that migration to the airways by infected neutrophis or macrophages play an important role in distribution of the bacilli [262].

3-ketoacyl-CoA thiolases are degradative thiolases which catalyze the reaction of 3-ketoacyl-CoA to a shortened acyl-CoA species and an acetyl-CoA [263], [264], [7]. This reaction was proposed as a two-step "ping-pong" mechanism by Modies et al. and is depicted in Figure 11.3 [263], [265]. Thiolases contain a conserved active sites with a catalytic dyade which consists a acid/base cysteine and a histidine [266]. Studies of Nesbitt et al. revealed that FadA5 is a degradative thiolase and were able to show the side chain degradation in cholesterol [253].The degradation is the favored process but still the carbon-carbon cleavage is reversible [267].

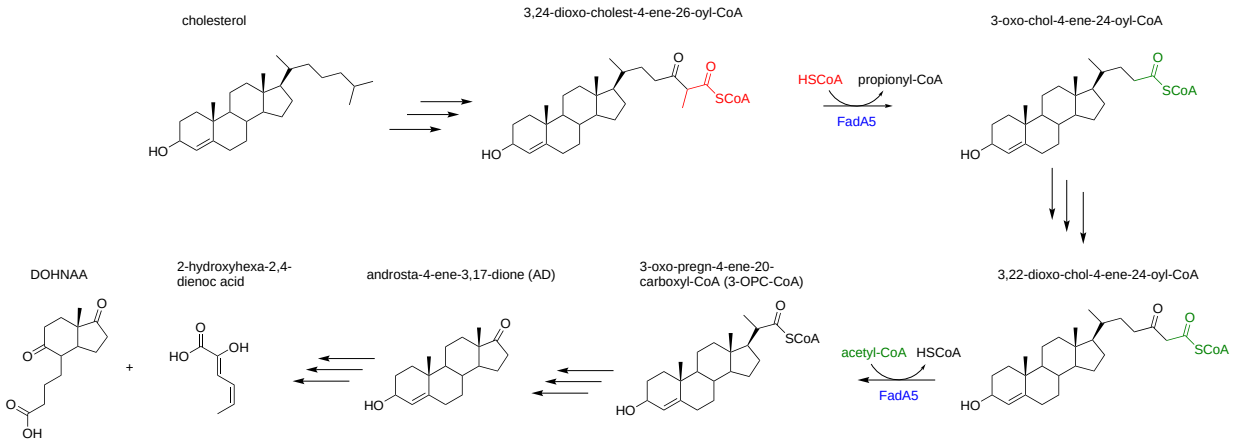


Figure 11.2: C-C cleavage of ketothioester to propionyl-CoA (red) or acetyl-CoA (green) by FadA5 (blue) into steroid products adapted from [253]. DOHNAA refers to 3-[(3aS,4S,7aS)-7a-methyl-1,5-dioxo-2,3,3a,4,6,7-hexahydroinden-4-yl]propanoic acid.

Schaefer et al. showed that the production of 3-OPC-CoA and Ac-Coa by FadA5 is promoted after addition of 3,22-dioxo-chol-4-ene-24-oyl-CoA [267]. Investigations revealed that 3,22-dioxo-chol-4-ene-24-oyl-CoA is 400-fold more specific compared to AcAc-CoA which is related to the steroid structure. Thus, a steroid-binding pocked was identified which leads to the favorable result and might be a potential target. Still, the mechanism of steroid binding remains unexplained. Two possibilities are discussed by Schaefer. On the one hand steroid binding might occur along the CoA-binding site and the catalytic triade.

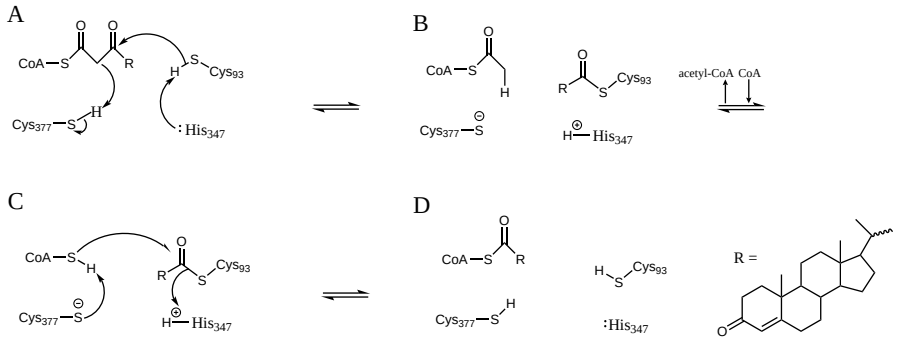


Figure 11.3: Proposed thiolase reaction mechanism as degradation (A to D) or biosynthesis (D to A) adapted from [265]. **A** – Proton abstraction of Cys93 and nucleophilic attack at ß-ketoacyl-CoA. **B** – Formation of acyl-cysteine and acetyl-CoA. **C** – Exchange of acetyl-CoA by CoA, which attacks acyl-CoA to obtain two-carbon shortened acyl-CoA and free Cys93. If methyl group is at the α-carbon of ß-ketoacyl-CoA one obtains propionyl-CoA.

Otherwise a reorientation of the steroid binding loops might be conceivable. Nevertheless,

these results indicate a possible novel target for the treatment of chronic TB. Topic of this chapter is the development of a possible covalent inhibitor originating from a non-covalent structure. Due to the crystallization of OPC and CoA in FadA5 by the group of Kisker from the University of Würzburg the way was paved for computational investigations and the design of possible novel inhibitors which will be explained in the following chapters.

11.2 General approach of designing a covalent inhibitor

The first consideration to design a new inhibitor is the active site which one wants to target. As mentioned in section 3.5.1 the covalent inhibition is an at least two-step mechanism. First of all, the formation of a non-covalent complex is needed and therefore, one has to find an appropriate binding-site. The postulated mechanism in terms of degradation starts with the abstraction of a proton of Cys93 and its covalent attack with 3,22-dioxo-chol-4-ene-24-oyl-CoA (Figure 11.3). Afterwards the proton of Cys377 is abstracted by the resulting intermediate. The answer to address FadA5 non-covalently could be found in the first step of the degradation reaction. First of all, a non-covalent binding site for the coenzyme A as well as for the steroid is in FadA5 present [267]. Additionally, the catalytic dyade consisting of Cys93 and His347 can be used for the covalent formation. This leads to the suggestion that state A of the degradation might serve as starting point for covalent modification of FadA5 (Figure 11.3).

The next question is related to the topic of how to design the scaffold of this new inhibitor. It is of necessity that it shows an adequate non-covalent binding to the active site as well as an appropriate warhead for covalent modification. Three different possibilities can tackle this problem. First of all, one can focus on the binding site of the coenzyme A (Figure 11.4-A). CoA should able to enter and leave FadA5 rapidly to enable an efficient degradation (Figure 11.3-B/C). Hence, one can place the warhead at the coenzyme A's thiol group, which is directed to Cys93 (Figure 11.6-A). But drawbacks occur for this approach as the synthesis of structures of scaffold A is too complex. Additionally, one has to consider the selectivity of possible inhibitor candidate. CoA occurs in many biological reactions in human as in the metabolism of fats or of carbon and proteins. This could lead to many unintended side reactions. Another problem is the identification of an adequate docking procedure due to its flexibility. These factors complicate the evaluation of the findings.

Scaffold B employs the steroid moiety to interact with the binding site. Schaefer et al. were able to crystallize the non-covalent complex of the steroid OPC (3-oxo-preg-4-ene-20-

carboxylic acid) and CoA in FadA5 (Figure 11.6-B). Therefore, this scaffold contains a steroid (exemplary progesterone) with a warhead linked to C21 or C22. One advantage is the range of steroids which are available for purchase. This facilitates experimental work by decreasing the steps of synthesis. Due to the rigidity of the steroid, docking procedures are easier to perform and therefore, it is easier to develop a re-docking and afterwards docking protocol. To mimic the substrate 3,22-dioxo-chol-4-ene-24-oyl-CoA, which is involved in the first step of degradation (Figure 11.3-A), scaffold C was devised. This scaffold contains the CoA part as well as steroid structure and the di-keto group would be exchanged by the warhead. Nevertheless, the synthesis of scaffold C is of increased complexity due to the CoA moiety and the connection of the warhead in between.

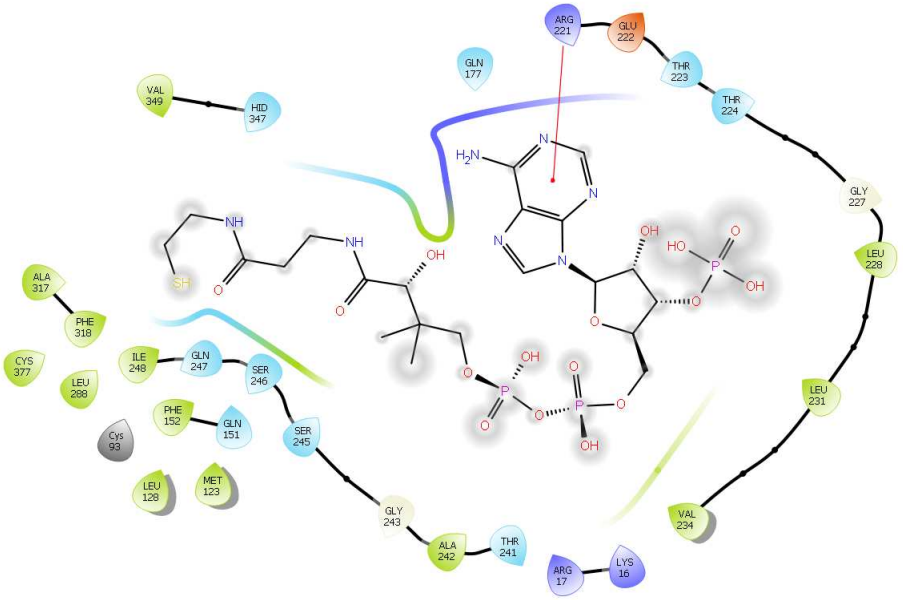


Figure 11.4: Interaction of coenzyme A with the protein environment and crystal water molecules of FadA5 in the X-ray structure.

Considering the aspects on experimental sides and also including the possibility of unwanted side effects of CoA, Scaffold B was pursued for further development of a covalent inhibitor. One important assumption has been made while designing these scaffolds. It is of importance that the abstraction of the proton of Cys93 by His347 occurs, to provide a reactive thiolate. This enables the formation of a covalent bond with the electrophilic warhead.

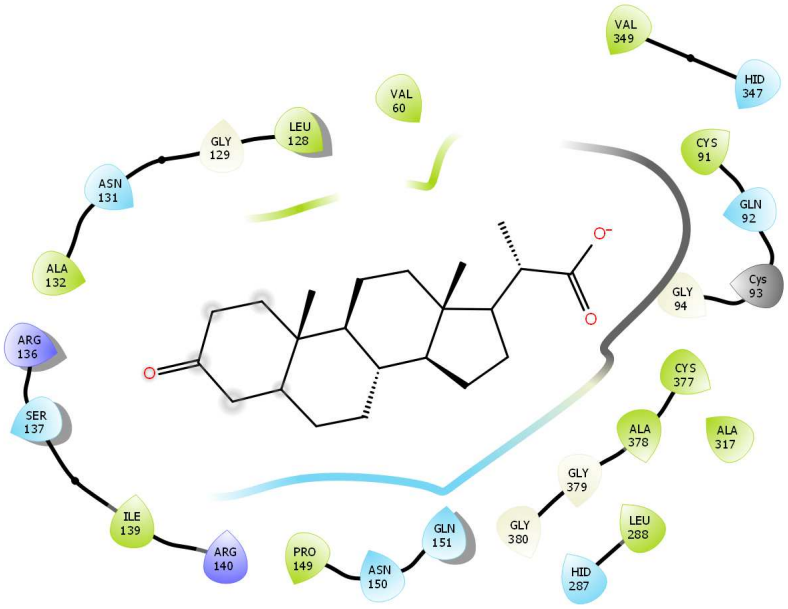


Figure 11.5: Interaction of OPC with the protein environment and crystal water molecules of FadA5 in the X-ray structure.

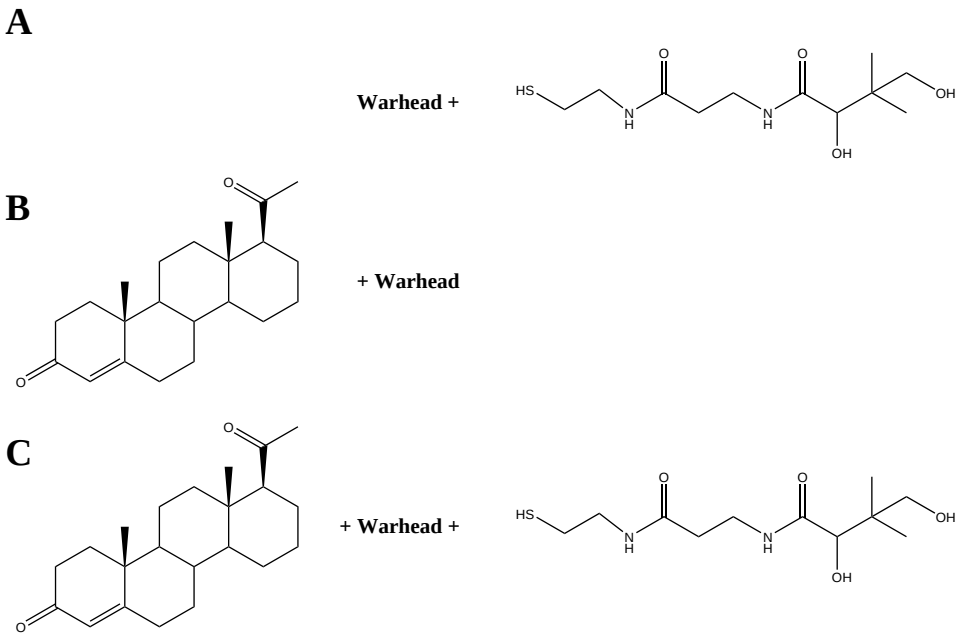


Figure 11.6: Scheme for general design of a covalent inhibitor for FadA5. **A** - Insertion of a warhead on thiol of co-enzyme A (scaffold A). **B** - Insertion of a warhead on the D-ring of a steroid, exemplary progesterone (scaffold B). **C** - Insertion of warhead in between steroid, exemplary progesterone, and co-enzyme A (scaffold C).

11.3 MD simulation of OPC and CoA in FadA5 variant C93S and apoprotein

To investigate the stability of OPC and CoA in FadA5, MD simulations have been performed. The pdb structure of mutant C93S of FadA5 in complex with OPC and CoA was used as a starting point (pdb code 4UBT). The pdb file contained four chains of the mutant, while chain D did not contain OPC. As no huge differences were seen in the active site, chain A has been chosen for further simulations.

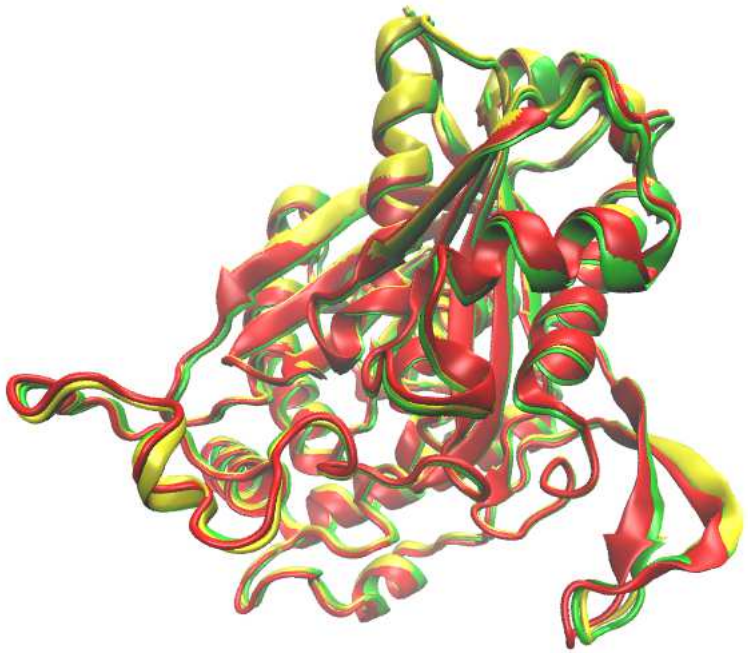


Figure 11.7: Comparision of monomers A (**red**), B (**green**) and C (**yellow**). Comparison of active site residues Cys93, His347, Cys377 and OPC showed that there is no significant difference in their position.

The ligand-protein complex has been prepared as described in section 5.1 as well the applied protocol for minimization and gradual heating. The production run was performed at 300 K without restraints and with duration of 5 ns.

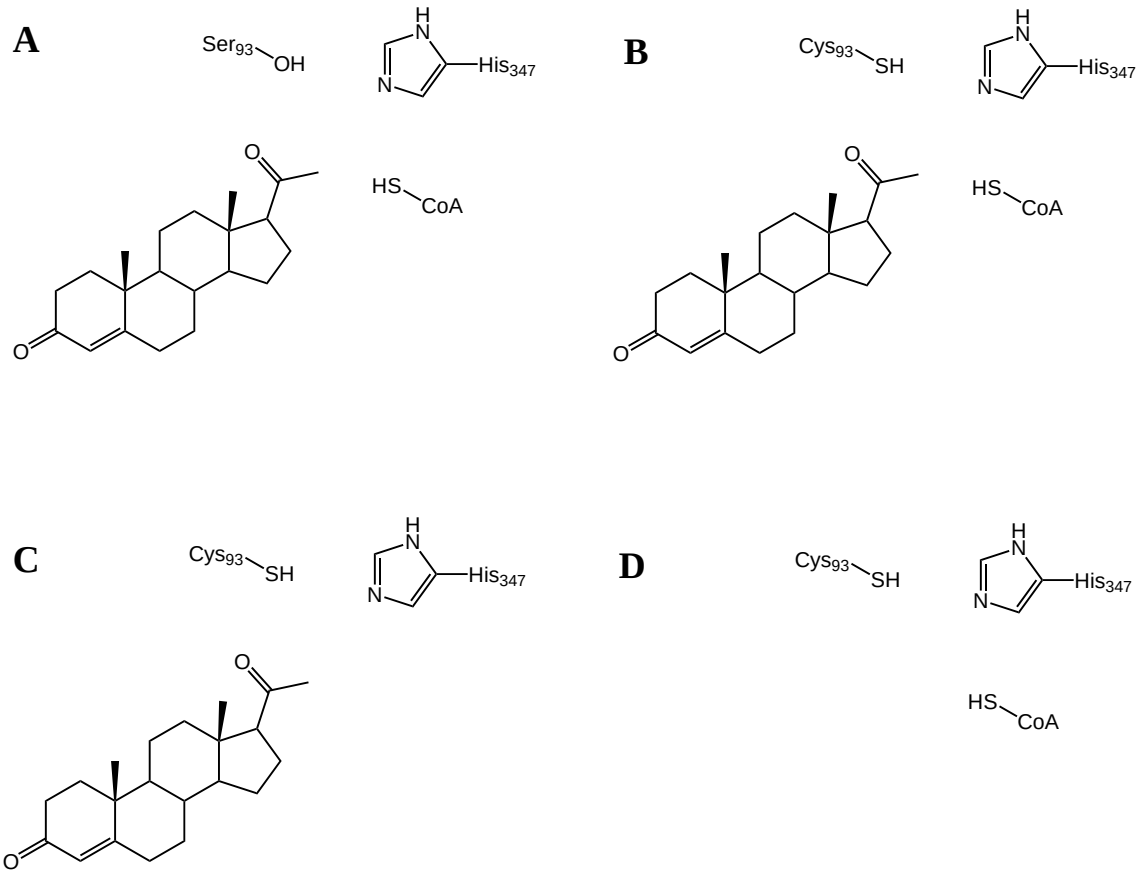


Figure 11.8: **A** – C93S variant of FadA5 in complex with OPC and CoA (complex I). **B** – Apoprotein of FadA5 in complex with OPC and CoA (complex II). **C** – Apoprotein of FadA5 in complex with OPC and in absence of CoA (complex III). **D** – Apopotein of FadA5 in complex with CoA and in absence of OPC (complex IV).

Four different simulations have been performed (Figure 11.8). First of all, the variant C93S containing OPC and CoA have been simulated to serve as a reference system (complex I). Afterwards Ser93 was changed to Cys93 to obtain the wild type. In this occasion the system was simulated in presence of OPC and CoA (complex II), absence of CoA (complex III) and absence of OPC (complex IV) to probe the effect of the co-enzyme on the steroid (Figure 11.8). RMSD values of OPC and CoA show a complex dependence in their behavior (Figure 11.9). For complex I the steroid show small fluctuations. The RMSD values of the co-enzyme remain at 1.5 Å for 3 ns but increase to 2.5 Å due to the motion of the 3-phosphorylated adenosine part. The modification of the C93S variant to the wild type (complex II) leads to fluctuations of CoA as in complex I. This time they are related to the movement of the di-phosphate of CoA. One can observe an increase in mean for the calculated

RMSD values from 1.76 Å to 2.15 Å whereas the standard deviation is comparable to complex I (see Table 11.1). In case of the steroid the RMSD values show a lower calculated mean while the standard deviation is equivalent for both complexes. In terms of the RMSD plot of CoA in complex IV one can note that the value is about 1-2 Å for 4.5 ns and increases up to 3 Å similar to complex I or II. For OPC in complex III (absence of CoA) one can observe an increase of the mean RMSD value to 1.87 Å and its standard deviation. As the binding pocket of the co-enzyme became unoccupied a motion of OPC was enabled. Still, one can assume that the binding site of the steroid moiety ensures sufficient stabilization for OPC and in consequence declines the necessity of CoA since the ligand remained in the binding site. Hence, scaffold B might serve as possible draft for designing a covalent inhibitor. It is of importance to mention that an additional assumption was made: Due to the need of the release of Ac-CoA and the binding of CoA in the second to third step of degradation (Figure 11.3) this exchange has to occur frequently. The following investigations consider that CoA is absent and the inhibitor binds at that time period. According to that guess the following docking and MD simulation of scaffold B have been performed in absence of the coenzyme.

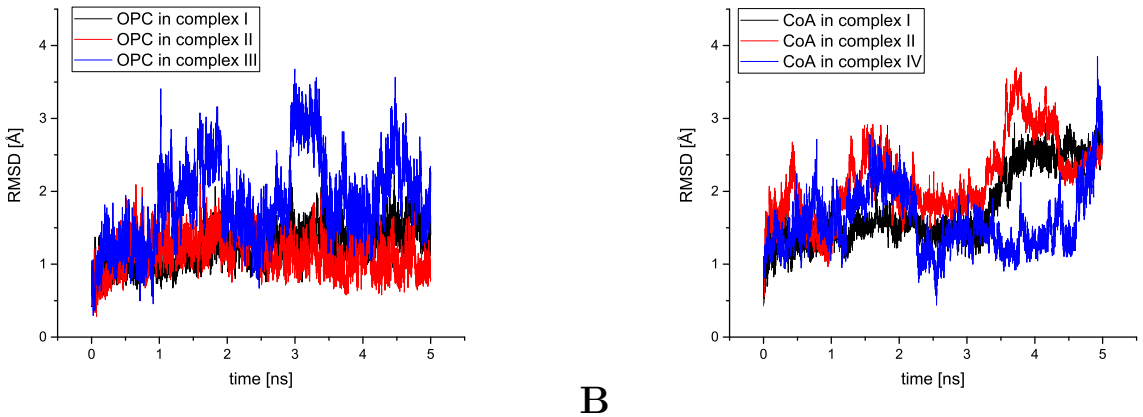


Figure 11.9: **A** – RMSD values as a function of time for OPC in complex I (black curve), complex II (red curve), complex III (blue curve). **B** - RMSD values as a function of time for CoA in complex I (black curve), complex II (red curve), complex IV (blue curve).

Table 11.1: Calculated mean and standard deviation of RMSD values for OPC and CoA during 5 ns production run correspond to Figure 11.9. All values in Å.

system	mean	standard deviation
OPC in complex I	1.24	0.27
OPC in complex II	1.16	0.27
OPC in complex III	1.87	0.61
CoA in complex I	1.76	0.52
CoA in complex II	2.15	0.54
CoA in complex IV	1.60	0.45

11.4 Proposed synthesizable structures for docking

For scaffold B, a range of different structures are possible due to the variety of warheads and steroids and its combination. The choice of warheads was limited to acrylamide, chloromethylketone, cyanamide, carbamate and bromodihydroisoxazole (Figure 11.10). Progesterone, pregnenolone, prasterone and estrone which are comparable structures to OPC served as steroid recognition unit (Figure 11.11). In all investigated inhibitors containing prasterone or estrone (except for the carbamat warhead) the steroid was also modified by exchanging the hydroxyl group of the A-ring with a hydroxymethylether. To ensure the realization of these structures, a synthesis plan was developed by the Holzgrabe group from the University of Würzburg. Because of the combination of different warheads and steroids 36 synthesizable inhibitors were suggested (Figure 11.12 - Figure 11.15).

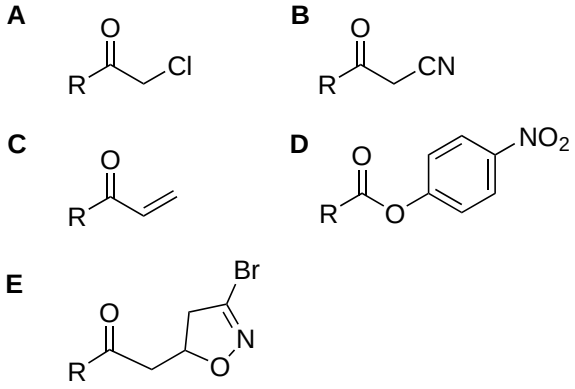


Figure 11.10: Different warheads as **A** – acrylamide, **B** – chloromethylketone, **C** – cyanamide, **D** – carbamate or **E** – bromodihydroisoxazol connected to the steroids after synthesis.

If one takes a closer look at the structures, one can notice that solely the (3-methoxy)estrone

or (3-methoxy)prasterone scaffold contains bromodihydroisoxazole (EstId, EstIId, PraId, PraIId) as a warhead. This is due the circumstance that those molecules (except for EstIe) were initially suggested and probed *via* docking simulations. Findings of these simulations lead to development of further structures represented in Figure 11.14 and Figure 11.15. Details on this matter will be discussed in Section 11.6.2.

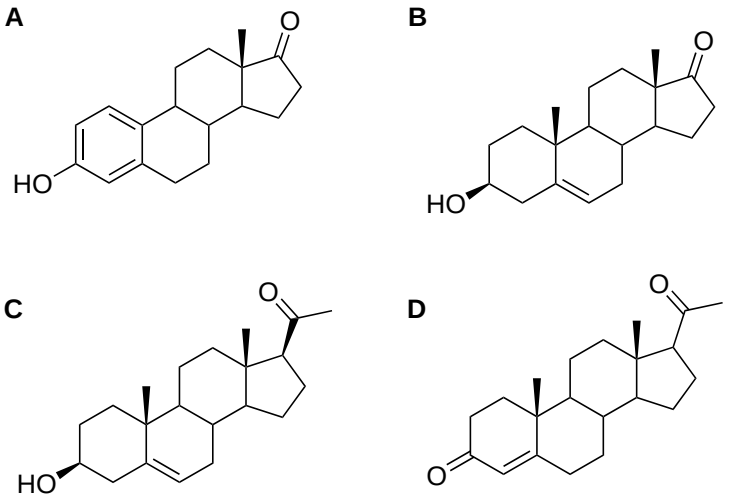


Figure 11.11: Different steroids as **A** – progesterone, **B** – pregnenolone, **C** – prasterone or **D** – estrone which were used for scaffold B.

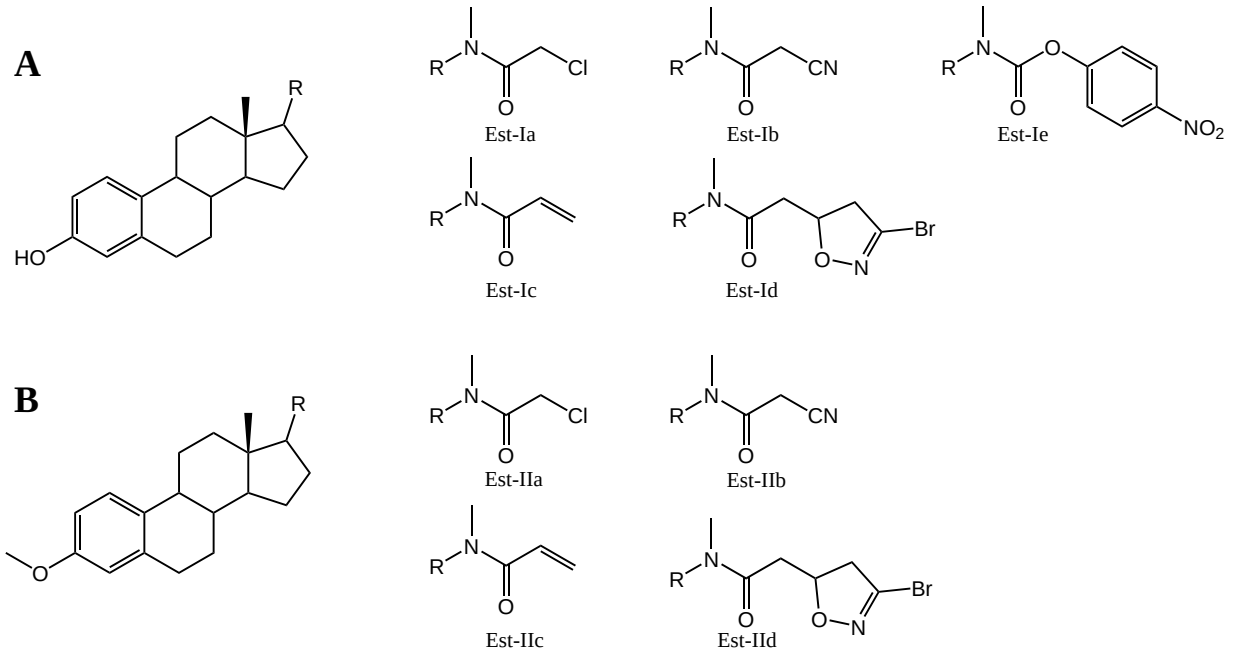


Figure 11.12: **A** – Estrone steroid structure in combination with selected warheads connected at the D-ring. **B** – 3-methoxy-estrone steroid structure in combination with selected warheads linked to the D-Ring.

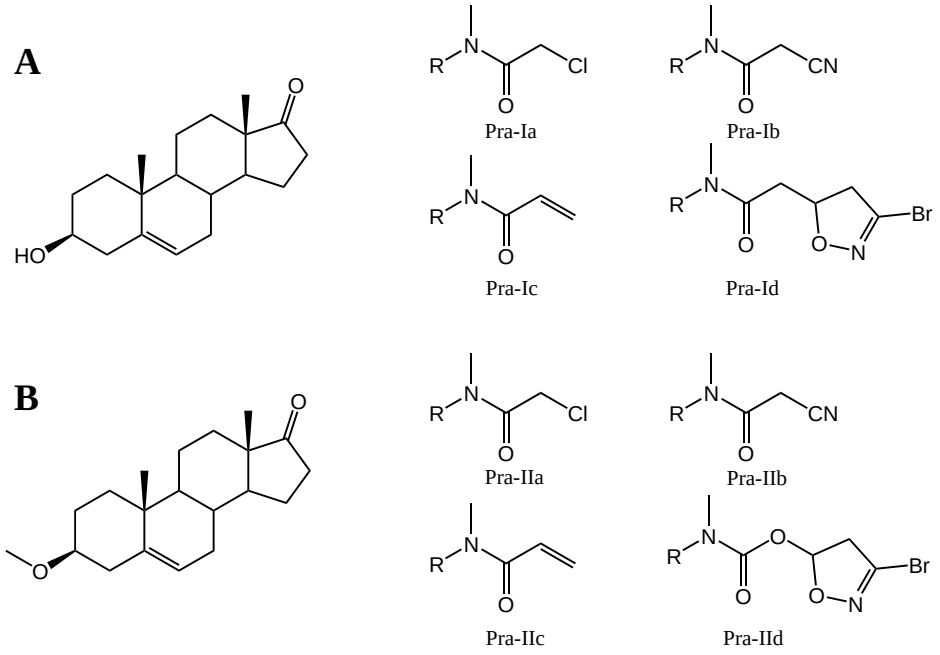


Figure 11.13: **A** – Prasterone steroid structure in combination with selected warheads connected to the D-ring. **B** – 3-methoxy-prasterone steroid skeleton in combination with selected warheads linked to the D-ring.

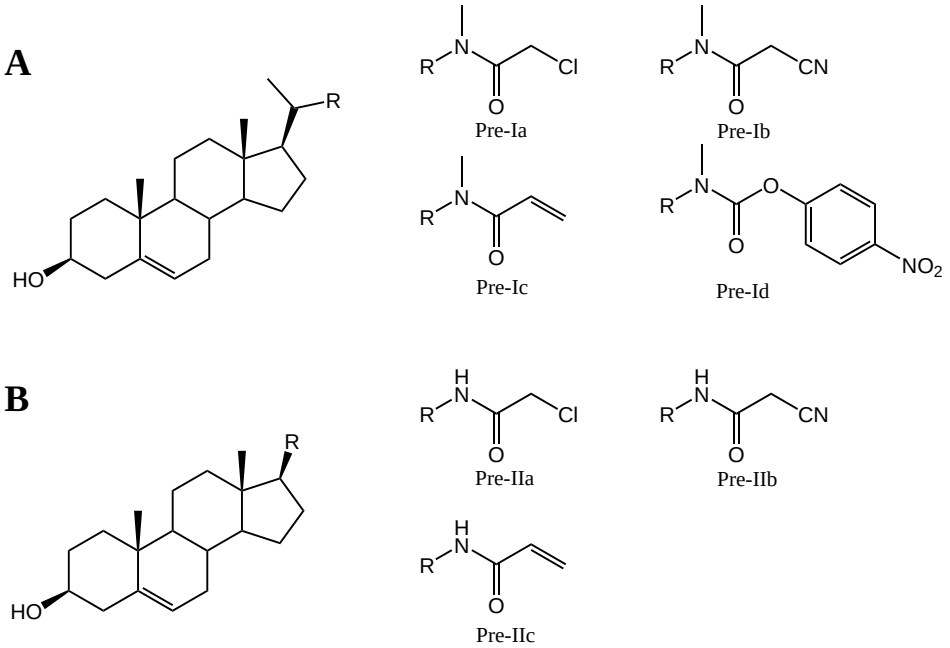


Figure 11.14: **A** – Pregnenolone steroid structure in combination with selected warheads connected to the D-ring. **B** – 3-methoxy-prasterone steroid structure in combination with selected warheads linked to the D-ring.

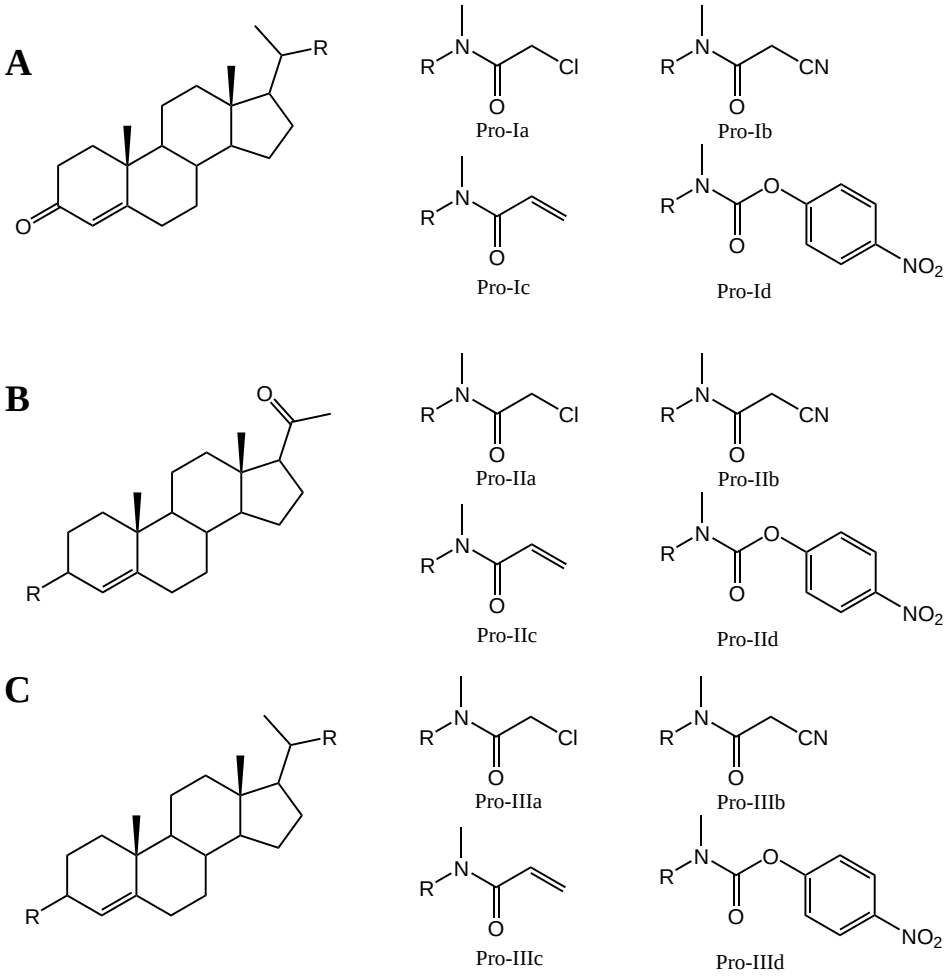


Figure 11.15: Progesterone steroid structure in combination with selected warheads connected to the D-ring (**A**), A-ring (**B**) or both (**C**).

11.5 QM calculation of the reaction energy of selected warheads

The second step of inhibition is the formation of the covalent bond between warhead and targeted amino acid. As mentioned before Cys93 appears to be appropiate for a nucleophilic attack since it is involved in the reaction with 3,22-dioxo-chol-4-ene-24-oyl-CoA and is also part of the catalytic dyade (Figure 11.3). Therefore, an evaluation of the suitability of the selected warheads is needed. QM computations have been performed to estimate the reaction energy employing the Turbomole V6.6 program package. DFT and wave function-based methods have been compared with CCSD(T) to evaluate the best methodology describing

(reaction)energy. For sake of completeness, environmental effects were included by using implicit solvation models.

In a first step structures were optimized on B3LYP-D3 level and the basis set def-TZVP. As a model for implicit solvation the COSMO approach was employed with an ϵ-value of 78 to mimic the molecules interaction within a water surrounding. The cysteine residue was represented as methanethiol to save computational efforts. In case of chloromethylketone and bromodihydroisoxazole, a methanethiolate was used. The warhead was exteneded by a methyl group to obtain the same structure used for QM/MM calculations which will be discussed later on.

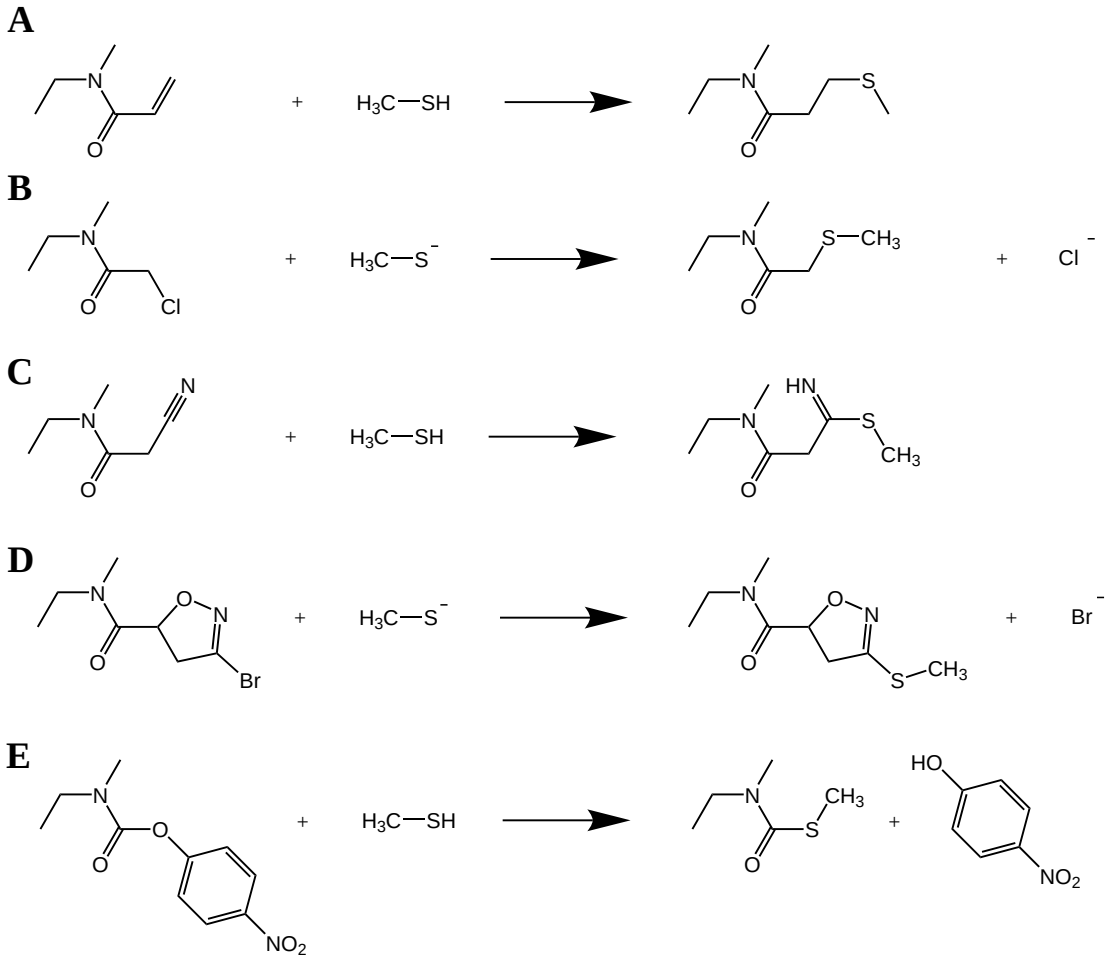


Figure 11.16: Calculated reaction of selected warheads as acrylamide (**A**), chloromethylketone (**B**), cyanamide (**C**), bromodihydroisoxazole (**D**) and carbamate (**E**) with methanethiol or methanethiolate.

11.5.1 Benchmark calculation in vacuum

To calculate the reaction energy, DFT methods as B3LYP and BHLYP with Grimme's dispersion correction or ab initio approaches as MP2, CC2 or its spin component scaled variations were used to obtain the relative energetic difference between reactants and products. CCSD(T) was selected as benchmark reference since it represents the gold standard for ground state closed shell systems. The application of aug-cc-pVDZ results, since negatively charged atoms and molecules are involved in the reaction. As mentioned in a previous chapter, no implementation of CCSD(T) in combination with the COSMO approach was available in Turbomole V6.6. Hence, all energies had to be calculated in vacuum first, to obtain comparable results to CCSD(T).
The difference in reaction energy is summerized for the investigated warheads in Table 11.2: Bromodihydroisoxazole shows the highest reaction energy, followed by chloromethylketone. The difference to acrylamide and cyanamide results due to the fact that both reactions lead to an abstraction of a halogen and thus most likely to an irreversible inhibition (as a reattack is unlikely). In case of acrylamide the reaction energy is about 5 kcal/mol lower compared to chloromethylketone and might lead to an irreversible inhibition, too. The lowest reaction energies of about -9 kcal/mol or -3 kcal/mol can be observed for cynamide and carbamate warheads which are in the range of reversible covalent inhibition.

Table 11.2: Reaction energies obtained by QM calculation in vacuum. Aug-cc-pVDZ has been used as basis set except for * marked method. In that case the def-TZVP basis set was employed. All energies shown are in kcal/mol.

aug-cc-pVDZ// vacuum	**acrylamide**	**chloro-methyl-ketone**	**cyanamide**	**bromodi-hydro-isoxazole**	**carbamate**
CCSD(T)	-25.9	-31.4	-8.7	-44.6	-2.7
B3LYP-D3*	-22.9	-30.7	-9.4	-46.0	-0.3
BHLYP-D3	-25.7	-32.7	-9.4	-48.9	-2.8[1]
MP2	-28.4	-32.3	-8.5[1]	-45.8	-4.0
CC2	-28.5	-32.3	-9.8	-45.8	-1.6
SCS-MP2	-25.9[1]	-31.6	-4.7	-45.3[1]	-2.5
SCS-CC2	-25.7	-31.5[1]	-5.2	-45.2[1]	-1.4

[1]Reaction energies with the lowest deviation compared to CCSD(T) calculated energies.

Table 11.3 shows the deviation of calculated reaction energies of the selected warheads.

Table 11.3: Absolute deviation of reaction energies compared to CCSD(T) energies. Aug-cc-pVDZ was used as basis set except for * marked methods which employed the def-TZVP basis set. All deviations shown are in kcal/mol.

deviation	acrylamide	chloro-methyl-ketone	cyanamide	bromodi-hydro-isoxazole	carbamate	MAE
B3LYP-D3*	3.0[2]	0.7	0.7	1.4	2.4	1.6
BHLYP-D3	0.2	1.3	0.6	4.2[2]	0,1	1.3
MP2	2.5	0.9	0.2[1]	1.1	1,3	1.2
CC2	2.6	0.9	1.1	1.2	1,2	1.4
SCS-MP2	0.0[1]	0.2	4.0[2]	0.6[1]	0,2	1.0
SCS-CC2	0.2	0.1[1]	3.5[2]	0.6[1]	1,4	1.2

[1] Lowest deviation compared to CCSD(T) calculated reaction energies. [2] Deviations which are above 3 kcal/mol compared to CCSD(T) calculated reaction energies.

For acrylamide, chloromethylketone and bromodihydroisoxazole, best results were obtained by employing the spin-component scaled post-HF methods. DFT methods mostly perform with a comparable accuracy as MP2 or CC2 except for BHLYP which show similar results as SCS-CC2 for acrylamide or even better ones for carbamate. Worst results were obtained in case of B3LYP for acrylamide and BHLYP for bromodihydroisoxazole as they show deviations of 3.0 and 4.2 kcal/mol. Solely the computed reaction energies of cyanamide are different. DFT methods outperform SCS-MP2 and SCS-CC2 as deviation of 3.5 and 4.0 kcal/mol occur. Also CC2 is slightly worse than utilized DFT methods resulting in a deviation of 1.1 kcal/mol. The best result can be achieved by computing the reaction energies with MP2. In case of the MAE (mean averaged error) MP2 and SCS-MP2 or SCS-CC2 perform best, although CC2 and applied DFT methods show reasonable results, too (except for above mentioned occurences).

11.5.2 Influence of the environment on reaction energies

After determining the best method for each warhead in vacuum, the COSMO approach (ϵ=78) has been applied to evaluate environmental influences regarding the reaction energy. One can observe that the utilization of implicit solvation decreases the reaction energy of bromodihydroisoxazole slightly by 1 kcal/mol (Table 11.4). In case of acrylamide and cyanamide an upshift between 1.3 and 1.6 kcal/mol or 2.7 and 3.3 kcal/mol occurs in terms of the reaction energy. For carbamate it ranges at the same level (-2.8 to -3.2 kcal/mol). The

highest impact can be seen for chloromethylketone after including environmental effects. By employing the COSMO approach, the reaction energy decreases by about 6 kcal/mol. This difference between bromodihydroisoxazle and chloromethylketone is interesting because one would suggest that both of them would show the same trend.

Table 11.4: Reaction energies obtained by QM calculation in COSMO. Aug-cc-pVDZ has been used as basis set except for * marked method which employed the def-TZVP basis set was employed. All energies shown are in kcal/mol.

aug-cc-pVDZ// COSMO	**acrylamide**	**chloro-methyl-ketone**	**cyanamide**	**bromodi-hydro-isoxazole**	**carbamate**
B3LYP-D3*	-21.6	-36.9	-6.2	-46.6	-5.9
BHLYP-D3	-24.2	-38.8	-6.1	-49.4	-3.6[1]
MP2	-27.0	-38.8	-5.8[1]	-46.3	-5.7
CC2	-27.2	-38.8	-7.0	-46.4	-6.8
SCS-MP2	-24.5[1]	-38.0	-2.0	-45.7[1]	-4.5[1]
SCS-CC2	-24.4	-37.9[1]	-2.5	-45.8[1]	-5.3

[1]Reaction energies with the lowest deviation compared to CCSD(T) calculated energies.

Table 11.5: Absolute deviation of reaction energies compared to vacuum energies. Aug-cc-pVDZ has been used as basis set except for * marked method. In that case the def-TZVP basis set was employed. All deviations shown are in kcal/mol.

deviation vaccum	**acrylamide**	**chloro-methyl-ketone**	**cyanamide**	**bromodi-hydro-isoxazole**	**carbamate**
B3LYP-D3*	1.3	-6.2	3.2	-0.9	-3.1
BHLYP-D3	1.5	-6.1	3.3	-0.8	-3.2
MP2	1.4	-6.5	2.7	-1.1	-2.9
CC2	1.3	-6.5	2.8	-1.2	-2.8
SCS-MP2	1.4	-6.4	2.7	-1.0	-2.9
SCS-CC2	1.3	-6.4	2.7	-1.1	-2.8

Therefore, one can take a closer look on each structure and their energies. Table 11.6 shows the differences in energy obtained in vacuum or COSMO of each structure for the reaction of chloromethylketone with methanethiol. Especially the last two rows are of interest. They show the difference between the deviation of vacuum and COSMO energy in

products and educts for neutral or charged molecules (and atoms). In case of the neutral structures (educt and product of chloromethylketone) difference in deviation are of about 2 kcal/mol. The most significant differences occur for negatively charged atoms or molecules of the reaction. Especially chlorine's absolute energy is about 5 kcal/mol lower compared to methanethiolate if solvent effects are included. This might by reasoned due to the stabilization of its negative charge. This downshift of energy leads to an increased reaction energy of chloromethylketone if COSMO is applied.

Table 11.6: Absolute deviation of each structure calculated with COSMO compared to vacuum energy. Differences between deviations of neutral structures (educt = chloromethylketone and product) and negatively charged atoms/molecules (Cl^- and CH_3S^-) are shown. Aug-cc-pVDZ has been used as basis set. All energies shown are in kcal/mol.

method	**educt**	**CH3S$^-$**	**product**	**Cl$^-$**	**reaction energy**	**Δ(educt/ product)**	**Δ(Cl-/ CH3S$^-$)**
B3LYP	9.2	66.1	10.9	70.7	6.2	1.7	4.5
BHLYP	9.8	67.3	11.5	71.7	6.1	1.7	4.4
MP2	8.8	66.7	10.5	71.4	6.5	1.7	4.7
CC2	8.6	66.5	10.3	71.3	6.5	1.7	4.8
SCS-MP2	8.8	66.7	10.5	71.4	6.4	1.7	4.7
SCS-CC2	8.7	66.6	10.3	71.3	6.4	1.7	4.7

The same analysis was conducted for bromodihydroisoxazole and the computed energies in vacuum and in COSMO were compared (Table 11.7). In this occasion neutral molecules involved in the reaction show similar deviations as for chloromethylketone. In contrast bromide and methanethiol show the same deviation after applying implicit solvation. Hence, the decrease in reaction energy is explained due to the influence of the neutral molecules. The stabilization of the negatively charged bromide is therefore about the same as methanethiol which does not result to a significant influence of the reaction energy.

Table 11.7: Absolute deviation of each structure calculated with COSMO compared to vacuum energy. Differences between deviations of neutral structures (educt = bromodihydroisoxazole and product) and negatively charged atoms/molecules (Br- and CH3S-) are shown. Aug-cc-pVDZ has been used as basis set. All energies shown are in kcal/mol.

method	**educt**	**CH3S-**	**product**	**Br-**	**reaction energy**	**Δ(educt/ product)**	**Δ(Br^-/ $CH3S^-$)**
B3LYP	10.9	66.1	12.0	65.9	0.9	1.1	0.2
BHLYP	11.5	67.3	12.8	66.8	0.8	1.3	0.5
MP2	10.4	66.7	11.5	66.6	1.1	1.1	0.1
CC2	10.2	66.5	11.3	66.6	1.2	1.1	0.1
SCS-MP2	10.4	66.7	11.5	66.7	1.0	1.1	0.1
SCS-CC2	10.3	66.6	11.4	66.6	1.1	1.1	0.0

11.6 Docking procedure of different inhibitors of type scaffold B

11.6.1 Redocking of OPC in C93S variant of FadA5

To evaluate the suitability of the suggested inhibitors, docking approaches have been employed. In the beginning it is of importance to perform a re-docking of OPC to verify docking parameters which therefore can be used for the proposed ligands. In consequence, OPC was removed from the protein ligand complex and was re-docked to the FadA5 C93S variant. Afterwards, the differences in structure regarding the X-ray structure of OPC were quantified by calculating the RMSD. If many poses with a low RMSD value are obtained in the docking, one can assume that used parameters are applicable for the structures of the type scaffold B. Furthermore, the ranking of the poses was rescored for additional refinement of the docking results.

The docking procedure was carried out in cooperation with the Sotriffer group, University of Würzburg, by employing the program package GOLDv5.2.2. First of all, the protein was prepared in the following manner: The ligand, co-factor and crystal water were removed except for two water molecules close to Cys93 and Cys377 since they show interactions with OPC and the mentioned amino acids (Figure 11.5). As the protein was used from the X-ray structure, no hydrogen atoms were available. Hence, hydrogens were assigned by utilizing the MOE software package at a pH of 6.7 at which the C93S variant was crystallized. Additionally, His347 was protonated at the delta position as this is needed for the proton transfer depicted in Figure 11.3. The ligand was prepared independently by the MOE software pack-

age. Hydrogens were also assigned at a pH of 6.7 resulting to a deprotonated carbon acid which is understandable as its calculated pKa-value is about 4.6 obtained from the Chemdraw software package. After minimization with a convergence criterion of 0.05 kcal/mol, the ligand was used for docking. The co-factor A was not considered, as the focus of interest lies at the scaffold B which is expected to react in absence of CoA. The number of GA runs was set to 50 to obtain a sufficient number of poses. Additionally, GA settings have been set to very flexible which increases the computational time but might also increase the quality of the docking results. In case of the algorithm default settings for the parameters of GOLD have been employed. To avoid bias of the results no constraints have been applied. One important aspect has been varied: the specification of the approximate radius r of the binding site. First of all, a certain atom has to be selected as center of the active site. Because Cys93's sulfur atom is the center of proposed reaction it was chosen. Three different r-values have been used: 10, 20 and 30 Å. To evaluate the docked structures, the CHEMPLP fitness function was applied as it is recommended by the GOLD developer. Afterwards the best ten poses ranked by CHEMPLP have been rescored by DSX with CSD pair-potential by the python script of Dr. Michael Hein and Maximilian Kuhn. Another important thing is to mention: As shown in section 11.3, MD simulations revealed that OPC remained in the active site even in absence of the cofactor. For this reason, CoA was not included in the docking process.

As one can see in Table 11.8, a r-value of 10 Å results in only two poses (in the top ten) which show a RMSD value below 0.7 Å. Especially the CHEMPLP scoring function does not give the right order of ranking. The second structure was ranked as worst pose although it has a RMSD value of 0.62 Å. In this case the DSX scoring leads to the right order of ranking. The best results were obtained by utilizing a r-value of 20 Å. All poses show a RMSD value below 0.70 Å. In terms of order, the DSX scoring moves the poses with a RMSD value of 0.55 and 0.31 Å to the second and third place which are ranked as worst regarding the CHEMPLP scoring. In case of $r = 30$ Å resulting poses are comparable to findings which employed a r-value of 10 Å. Only two poses show a RMSD value below 0.8 Å. Concerning rescoring, DSX show worse results than CHEMPLP. As CHEMPLP correctly rank the poses with a low RMSD value to the first places DSX solely scores the one with the RMSD value of 0.50 Å well. The second structure with a RMSD value of 0.78 Å is ranked in fourth place. The results of the re-docking show indeed that it is possible to obtain reliable poses by applying a r-value of 20 Å. In conclusion those parameters can be used for the docking of the selected inhibitor.

Table 11.8: Docking results of OPC in FadA5 C93S variant by varying r to 10 Å, 20 Å and 30 Å. The poses were scored by CHEMPLP and rescored by DSX. Order of poses on the basis of the DSX ranking. Negative DSX values are shown for clarity. RMSD values are calculated in relation to X-ray structure of OPC in FadA5 C93S variant. All RMSD values are given in Å.

	r = **10 Å**			r = **20 Å**		
Ranking	**RMSD**	**-DSX**	**CHEMPLP**	**RMSD**	**-DSX**	**CHEMPLP**
1	0.65	94.423	60.42	0.64	99.487	64.49
2	0.62	93.076	59.1	0.55	99.303	64.23
3	7.53	84.868	59.11	0.31	99.093	63.9
4	7.53	82.27	59.82	0.64	97.403	65.18
5	7.52	81.664	60.05	0.70	97.257	64.81
6	7.55	80.589	59.49	0.60	96.932	64.55
7	7.54	79.191	60.02	0.69	96.874	64.96
8	7.54	78.335	59.93	0.62	95.926	65.34
9	7.53	77.418	59.93	0.64	95.906	64.94
10	7.55	76.985	59.76	0.70	93.546	64.78

	r = **30 Å**		
Ranking	**RMSD**	**-DSX**	**CHEMPLP**
1	0.50	94.896	62.05
2	7.48	94.476	59.7
3	7.48	94.258	59.62
4	0.78	94.068	62.26
5	7.47	94.027	59.57
6	7.50	92.954	59.59
7	7.49	92.06	59.59
8	7.50	91.429	59.67
9	7.49	91.102	59.59
10	7.10	72.159	61.21

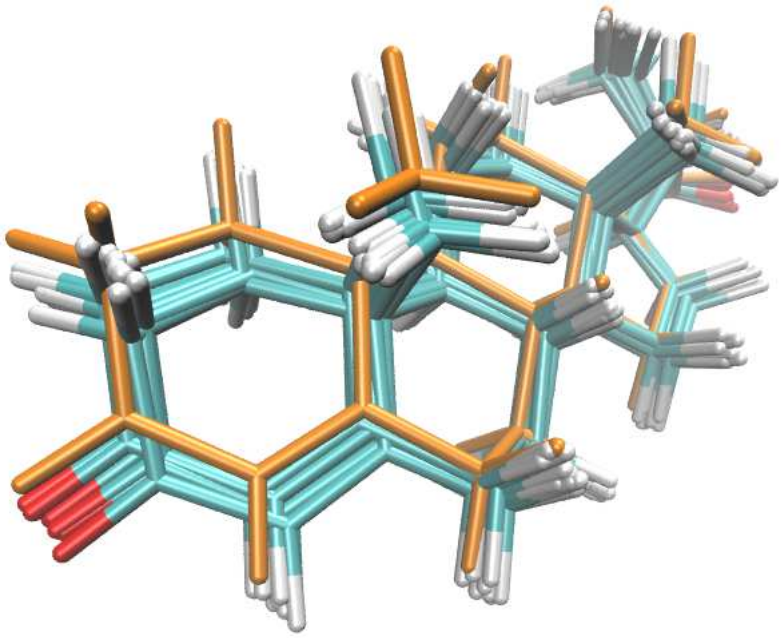


Figure 11.17: Top ten docked poses of OPC in FadA5 in absence of CoA by applying a r value of 20 Å. OPC structure of X-ray is orange colored in comparison to docked structures for clarity.

11.6.2 Docking of possible inhibitors of type scaffold B

After establishing a reliable docking procedure which performs well in redocking OPC in FadA5 C93S variant, this docking protocol was employed on the inhibitors of type scaffold B (Figure 11.12 to Figure 11.15). Just like in the re-docking, CoA was not considered and the docking was solely performed for the inhibitors. As the focus was laid on the apoprotein's modification, the active site contained a protonated Cys93 and doubly protonated His347. The first suggested structures were built on an estrone or prasterone scaffold and furthermore, also included a bromodihydroisoxazole as a warhead (Figure 11.12 and Figure 11.13). This time the RMSD value was not of interest as different steroid scaffolds were used. Instead, the investigation focused on the distance between the nucleophilic sulfur and the electrophilic warhead of the respective structure. This is quite obvious as the formation of a covalent bond is dependent on the distance. Therefore, the scope lies on the identification of possible ligands whose warheads are close to Cys93's sulfur (< 4 Å). These structures were further probed *via* MD simulations. Solely the best results are presented in Table 11.9 and Table 11.10, other results will be discussed in the following.

The docking simulations of prasterone revealed that this scaffold is unable to modify FadA5's active site. In terms of distance, these ligands (Pra-Ia to Pra-Id) showed distances of 13 to 15 Å due to a "flipped" steroid scaffold compared to OPC (exemplary shown for Est-1d in Figure 11.18-C). Also the modification of the hydroxyl-group (structures from type Pra-IIX) show distances in the same range. By exchanging the steroid scaffold to estrone one obtains different poses, comparable to OPC in the X-ray structure. As one can see in Table 11.9, Est-Ia has solely one pose which has an appropriate distance of 4.0 Å, while the other poses also show the behavior of a "flipped" steroid scaffold (Figure 11.18-C). Still, in terms of DSX and PLP ranking it was scored better than poses from rank 2 – 10. Hence, it was considered for MD simulations. The docking Est-Ib resulted in solely one pose with an adequate distance of 3.4 Å. Interestingly, PLP ranking predicts that pose as the best one whereas DSX delivers the opposite result. To gain further insight, MD simulations of this pose were conducted. The modification of estrone (Est-IIb) lead to solely two flipped orientations. Although one obtains eight poses with the right directionality towards the catalytic center, the distance ranges from 4.8 to 6.9 Å. Because of these docking results Est-IIb was not considered for MD simulations.

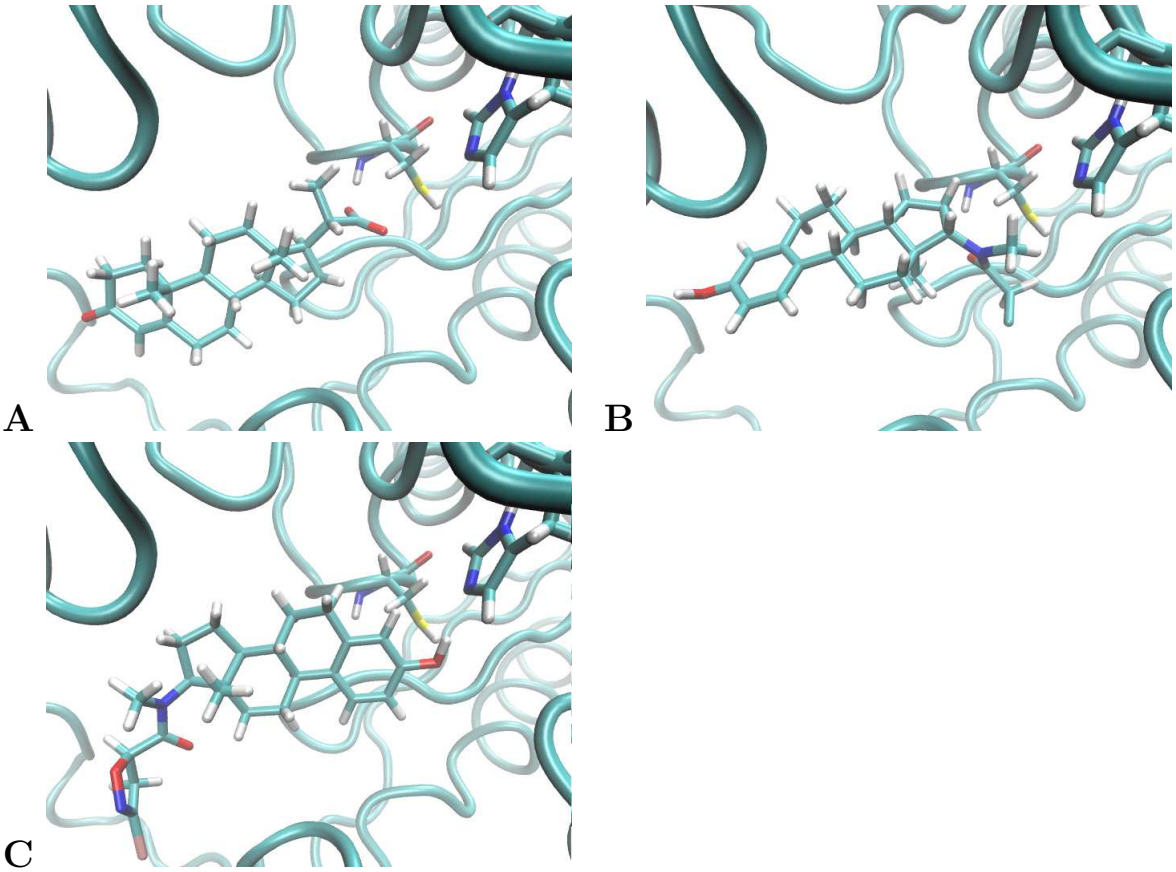


Figure 11.18: **A** – The crystal structure of OPC in FadA5. **B** – Docked structure of Est-Ia in FadA5. **C** – Docked structure of Est-1d in FadA5.

In case of ligands employing bromodihydroisoxazole as warhead, docking simulations revealed that in almost all cases the warhead was directed opposite to the catalytic center, i.e it showed a "flipped" steroid scaffold compared to OPC leading to distances of ≈ 14 Å. If one compares the size of bromohydroisoxazole to acrylamide or chloromethylketone, the warhead appears to be too large to fit into the binding pocket. In consequence, it was excluded as warhead in the following design of possible inhibitors.

Prasterone as steroid scaffold is in combination with the selected warheads unsuitable to obtain acceptable docking results. Therefore, pregnenolone and progesterone were included as steroid structures. In case of progesterone the warhead could be linked to the A- and D-ring (or at both locations) and thus lead to structures depicted in Figure 11.15. The design of structures Pro-IIIa to Pro-IIId was pursued to overcome the flipped pose of the warhead observed by the previous docking studies. At least one warhead should be directed to the active site if they are linked at both positions ("sandwich-structures"). Bromodihydroisoxazole was neglected as possible warhead for pregnenolone and progesterone, for above mentioned

Table 11.9: Docking results of Est-Ia, Est-Ib and Est-Ic in FadA5 apoprotein. The poses were scored by CHEMPLP and rescored by DSX. Order of poses on the basis of DSX ranking. Negative DSX values are shown for clarity. Distance was measured between electrophilic carbon center of warhead and Cys93 sulfur. The distance is given in Å.

		Est-Ia			**Est-Ib**	
Ranking	**distance**	**-DSX**	**PLP**	**distance**	**-DSX**	**PLP**
1	4.0	96.55	65.03	14.6	94.63	62.21
2	12.3	93.52	61.01	14.6	94.46	62.17
3	12.3	93.22	60.98	14.6	94.21	62.23
4	12.3	93.17	60.87	14.6	93.36	62.32
5	12.3	92.96	60.88	14.5	93.17	62.16
6	12.3	92.89	61.12	14.7	92.79	62.13
7	12.4	92.88	60.88	14.6	92.53	62.35
8	12.4	92.21	61.00	14.6	92.33	62.20
9	12.4	91.68	60.78	14.6	91.57	62.16
10	12.4	89.55	60.79	3.4	76.72	64.01

		Est-Ic	
Ranking	**distance**	**-DSX**	**PLP**
1	3.4	102.77	60.00
2	3.7	102.44	59.80
3	13.6	85.87	59.56
4	13.5	85.79	59.53
5	13.6	85.52	59.53
6	13.6	85.34	59.54
7	13.6	85.20	59.54
8	13.6	84.38	59.71
9	13.6	84.23	59.44
10	3.4	82.56	59.97

reasons. Distances obtained by the docking simulation of inhibitor Est-Ie are larger than 5 Å. In all poses it is directed towards the active site but because of the size of the nitrobenzene it could not approach the cysteine closer. Structures containing a pregnenolone (Figure 11.14) showed the same behavior as ligands based on the prasterone scaffold. All docked inhibitors pointed towards the solvent and revealed distances above 14 Å.

Inhibitors composed of pregnenolone resulted in different findings. Distances of Pro-Ia and also Pro-IIa were > 15 Å due to flipped poses. If both rings are modified (Pro-IIIa), one obtains nine poses of chloromethylketone linked to the A ring directed to the active site with a distance of ≈ 7 Å. In reverse, the distance of the warhead located at the D ring was > 14

Table 11.10: Docking results of Est-IIa and Est-IIc in FadA5 apoprotein. The poses were scored by CHEMPLP and rescored by DSX. Order of poses on the basis of DSX ranking. Negative DSX values are shown for clarity. Distance was measured between electrophilic carbon center of warhead and Cys93 sulfur. The distance is given in Å

		Est-IIa			**Est-IIc**	
Ranking	**distance**	**-DSX**	**PLP**	**distance**	**-DSX**	**PLP**
1	4.0	97.79	64.47	3.5	109.61	58.38
2	4.0	97.54	65.3	16.6	94.10	58.51
3	4.1	96.84	64.79	3.6	93.58	58.19
4	6.2	96.66	63.80	3.6	92.89	58.28
5	4.0	96.29	64.82	16.6	92.72	58.40
6	4.1	95.77	64.55	7.0	91.43	58.93
7	3.9	95.75	65.28	6.9	90.57	58.62
8	6.2	95.72	63.72	6.8	90.18	58.57
9	3.8	95.62	65.50	7.1	88.25	59.59
10	3.3	80.76	65.85	7.0	87.37	58.96

Å. Although the distance for covalent modification is insufficient, one still can see that the concept of "sandwich-structures" result in the right positioning. These conclusions can also be drawn for the other structures. Docked ligands with cyanamide attached to the D-ring (Pro-Ib to Pro-IIIb) show solely one pose directed to the active site with a distance of 6 Å. Pro-IIIb has seven poses directed to the binding pocket with distance of about 8 Å. This can also be observed for Pro-Ic to Pro-IIIc. Only two poses of the singly modified rings are directed towards the active site with a distance of 6 to 7 Å.

A slightly different observation was obtained for the structures combining the pregnenolone scaffold and the carbamate. For Pro-Ia the best four poses show distances of about 5 Å, for Pro-IIa all of the best ten poses show distances of 7 Å. Docked carbamate structures with a singly modified ring are positioned in the right orientation. In contrast to the other "sandwich-structures", Pro-IIId shows eight poses where the carbamate at the D-ring is placed right. Still these distances are inadequate for covalent modification (5.5 – 7.5 Å).

Nevertheless, some conclusions can be drawn by comapring the investigated structures. Prasterone and pregnenolone are not suitable as a steroid scaffold as all docked poses showed "flipped" orientations of the warheads. Although the idea of a "sandwich-ligand" overcomes this problem, the distances for covalent modification were too large. Therefore, structures in Table 11.9 and Table 11.10 were used for further MD simulations.

11.7 MD simulations of docked structures of type scaffold B

The docking studies resulted in five structures with poses in which the warhead is directed towards the active site and also shows an appropriate distance (< 4 Å) for covalent modification. Still one important question needs to be answered: Can these distances be maintained during a certain amount of time? This question is of importance because the longer a close distance can be maintained the likelier a covalent modification occurs. To shed light in this matter and also to verify the docking results MD simulations have been performed and distances according to Figure 11.19 have been measured.
In contrast to the simulations described in section 11.3, the system was prepared in absence of the coenzyme and furthermore two different protonation states were considered (Figure 11.19). As depcited in Figure 11.3, the mechanism starts with the abstraction of the proton from Cys93 to His347. Subsquently, one has the neutral state at the beginning, where Cys93 still harbors the proton (protonation state A) and the zwitterionic state in which His347 is doubly protonated and Cys93 is negatively charged (protonation state B). Hence, the influence of the protonation state regarding the distance of the warhead to Cys93 was of interest. The conducted MD simulations are analog to the simulations of OPC in FadA5 regarding parameterization, minimizing, heating and production runs. To probe the reproducibility, the MD simulations have been executed from the same starting point thrice for each protonation state with duration of 5 ns leading to a total simulation time of 150 ns for the ligands.

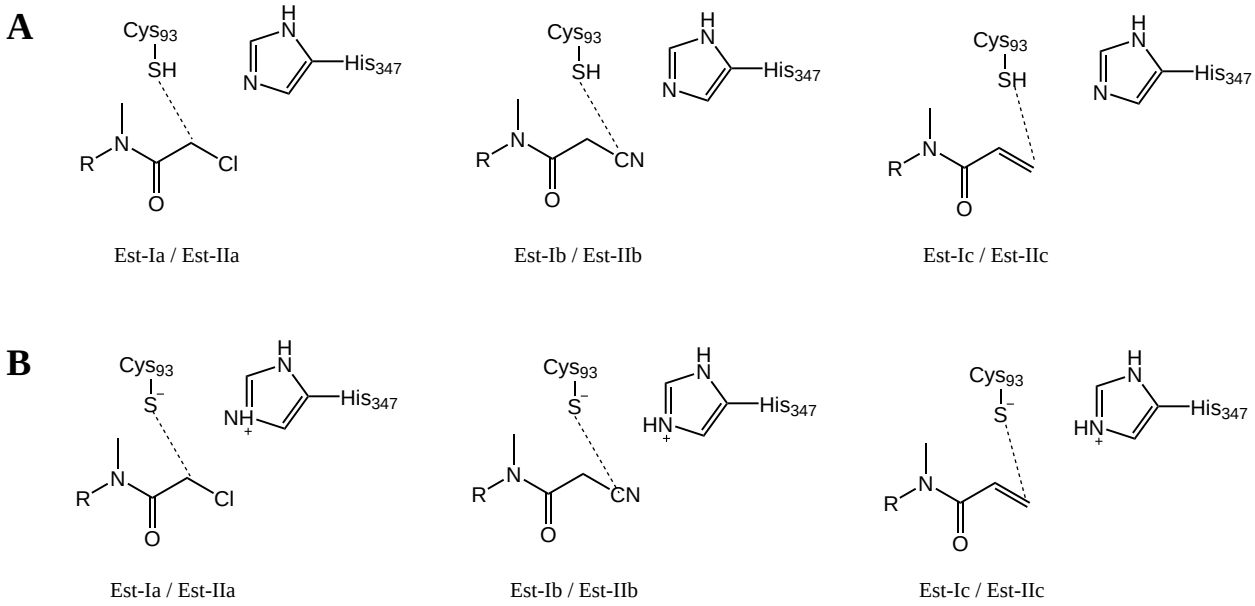


Figure 11.19: Distance between electrophilic carbon of warhead and nucleophilic sulfur of Cys93 depicted with dashed lines in (**A**) neutral protonation state of Cys93/His347 dyad or (**B**) in zwitterionic state. Location of warhead and amino acids serves for clarification and does not reflect actual position in docking or MD simulations. R corresponds to steroid scaffold as shown in Figure 11.12.

Table 11.11: Mean and standard deviation of distances between electrophilic carbon of ligand and nucleophilic sulfur of Cys93 during a 5 ns production run corresponding to Figure 11.20 - Figure 11.24. MD runs marked with * correspond to protonation state A, MD runs marked with ** to protonation state B. All values are given in Å.

		Est-Ia	**Est-Ib**	**Est-Ic**	**Est-IIa**	**Est-IIc**
MD run 1*	**mean / SD**	4.2 / 0.5	4.3 / 0.5	5.6 / 1.9	4.2 / 0.5	5.9 / 1.0
MD run 2*	**mean / SD**	3.7 / 0.4	5.6 / 1.3	4.8 / 0.7	5.1 / 0.6	5.0 / 0.6
MD run 3*	**mean / SD**	3.9 / 0.4	6.7 / 1.2	4.8 / 0.7	4.3 / 0.5	6.1 / 1.3
MD run 1**	**mean / SD**	5.3 / 2.3	5.2 / 1.3	5.7 / 0.7	6.1 / 2.5	5.9 / 1.2
MD run 2**	**mean / SD**	4.0 / 0 .7	8.4 / 2.1	5.7 / 1.1	4.3 / 0.8	6.0 / 1.3
MD run 3**	**mean / SD**	4.0 / 0.5	4.4 / 0.6	8.6 / 2.2	4.6 / 0.9	6.0 / 0.9
MD 1 – 3*	**mean**	3.9	5.5	5.1	4.5	5.7
MD 1 – 3**	**mean**	4.4	6.0	6.7	5.0	6.0

The best results were obtained for Est-Ia. In terms of the protonation state A, every MD simulation showed a mean of distance of 3.9 Å. By considering the zwitterionic state one obtains a larger mean value of 4.4 Å. This is caused by the first MD simulation in pro-

tonation state B as an increase in distance occurs to 10 Å after 4 ns. Consequently, the mean increases for simulation in this protonation state (Figure 11.20-B). Contrary to this simulation, the second and third production run show a mean of 4.0 Å, a sufficient distance for covalent modification. As this distance is maintained for all simulations during the time course, a small standard deviation in the range of 0.4 to 0.7 Å is obtained (except for the first MD run in protonation state B). Thus, one can expect a nucleophilic attack by Cys93's sulfur.

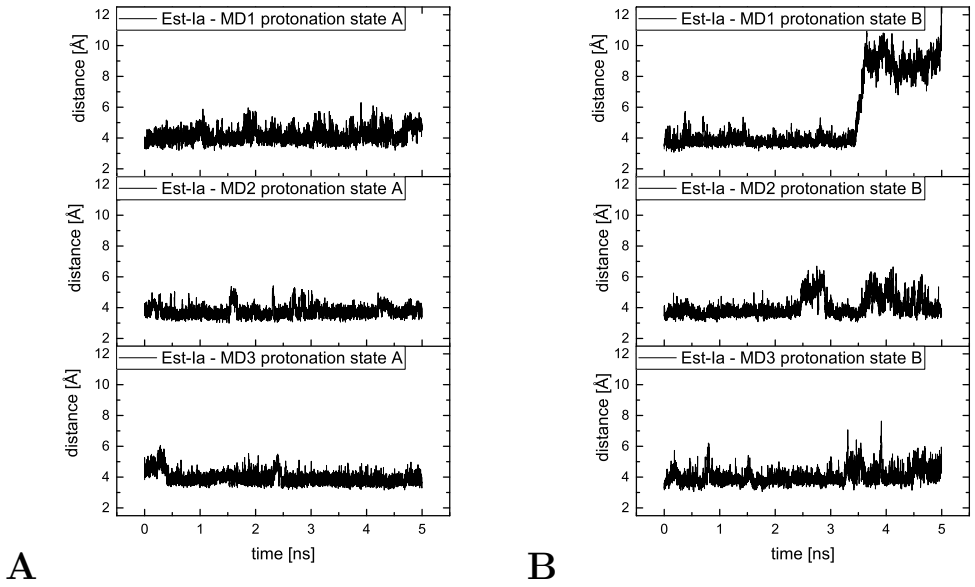


Figure 11.20: The variation in distance as a function of simulation time for inhibitor Est-Ia corresponding to Figure 11.19 during three 5 ns production runs. **A** – Simulation of docked structure Est-Ia in protonation state A. **B** – Simulation of docked structure Est-Ia in protonation state B.

This is not the case for the remaining structures. Although the first MD run of Est-Ib in the neutral state showed a mean in distance of 4.3 Å (due to a fluctuation at 3 ns), this observation could not be reproduced for the second and also third simulation, leading to a mean of 5.5 Å for all three simulations. Regarding the zwitterionic state, the mean of all three simulations increases to 6.0 Å especially due to the rise in distance to 10 Å in the second MD simulation (Figure 11.21). Comparable results were obtained for Est-Ic. The calculated mean of distance in the neutral state amounts 5.1 Å and even a sperate analysis of the simulations, values ranges from 4.8 to 5.6 Å. In consequence, a nucleophilic attack is unlikely. One might argue that there are certain periods of time in which a reaction can occur

but from a statistical point of view this occurrence is less likely compared to Est-Ia. The same conclusion can be made for Est-IIa. In the first and third MD simulation of protonation state A the distance of about 4.0 Å can be maintained for a certain amount of time, still the mean of these simulations are 4.2 or 4.3 Å. The trend in the zwitterionic state is just like before. The mean of simulations in the zwitterionic state are slightly higher compared to state A resulting in a mean of 5.0 Å. The only difference to Est-Ia is the exchanged hydroxylgroup of the estrone. This modification leads to a slight reorientation of the steroid due to lipohilic interactions to Ile139 which occasionally occur and therefore worsen the results. The analysis of Est-IIc affirms that this inhibitor is also not suitable for covalent modification by Cys93's sulfur. Both mean distances of protonation state A and B are too large that one cannot expect a nucleophilic attack (Figure 11.24).

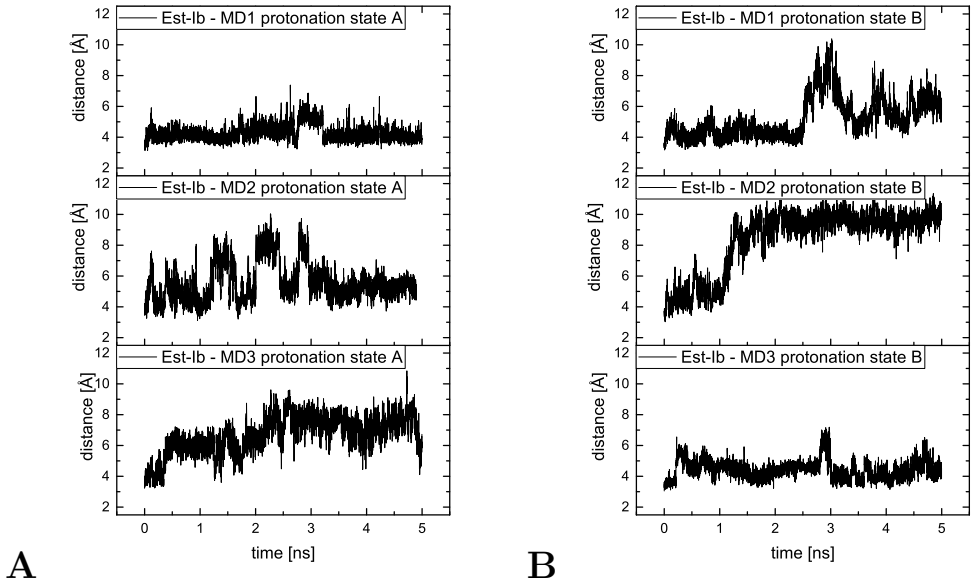


Figure 11.21: The variation in distance as a function of simulation time for inhibitor Est-Ib corresponding to Figure 11.19 during three 5 ns production runs. **A** – Simulation of docked structure Est-Ib in protonation state A. **B** – Simulation of docked structure Est-Ib in protonation state B.

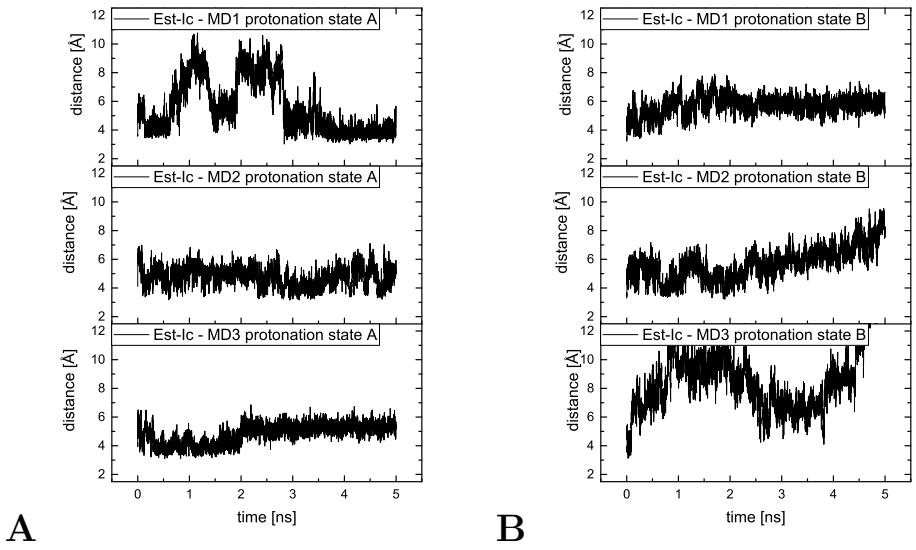


Figure 11.22: The variation in distance as a function of simulation time for inhibitor Est-Ic corresponding to Figure 11.19 during three 5 ns production runs. **A** – Simulation of docked structure Est-Ic in protonation state A. **B** – Simulation of docked structure Est-Ic in protonation state B.

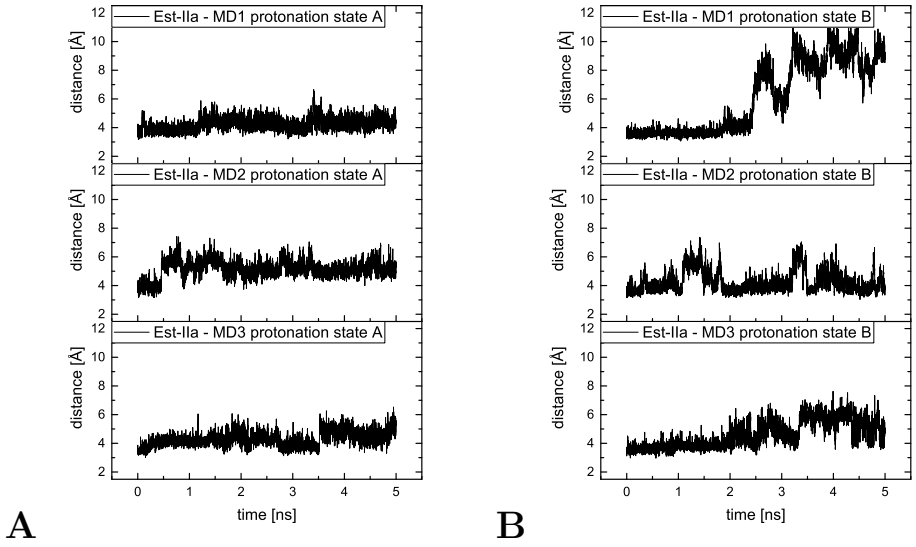


Figure 11.23: The variation in distance as a function of simulation time for inhibitor Est-IIa corresponding to Figure 11.19 during three 5 ns production runs. **A** – Simulation of docked structure Est-IIa in protonation state A. **B** – Simulation of docked structure Est-IIa in protonation state B.

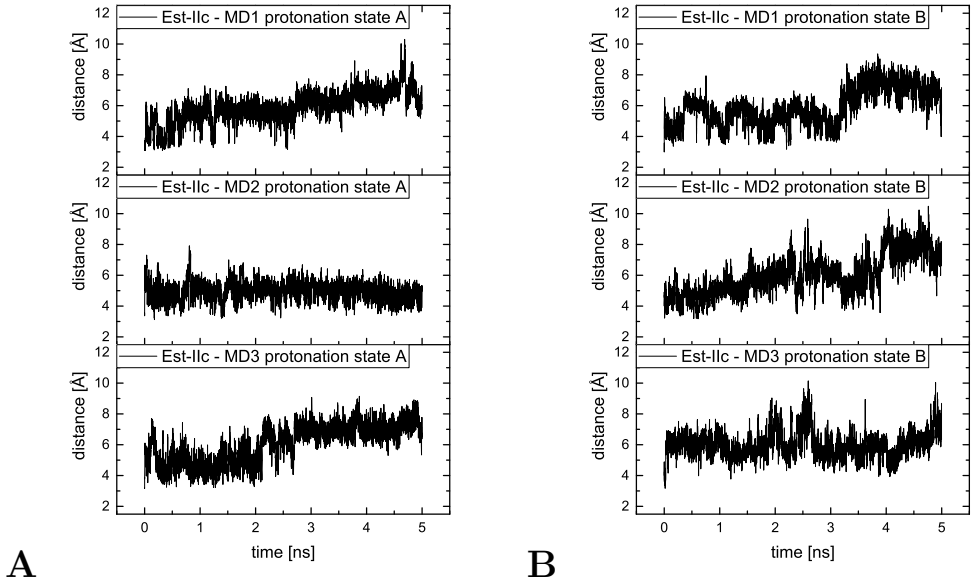


Figure 11.24: The variation in distance as a function of simulation time for inhibitor Est-IIc corresponding to Figure 11.19 during three 5 ns production runs. **A** – Simulation of docked structure Est-IIb in protonation state A. **B** – Simulation of docked structure Est-IIc in protonation state B.

11.8 Estimation of reaction energy and barrier

The results of docking and MD simulations lead to the assumption that solely Est-Ia appears to be a suitable candidate for inhibition. Still, the question of the reaction energy and barrier has to be considered. In case of the QM calculations only reaction energies of the warheads were calculated. To account the influence of the steroid further QM calculations in solution have been performed. Additionally, the reaction energy of Est-Ia was estimated by nudged-elastic band (NEB) calculations employing the CAST program package. To accelerate calculations a semi-empirical approach was utilizied and also benchmarked beforehand. In a last step a relaxed scan on QM/MM level was conducted to investigate the influence of the protein environment with respect to the reaction barrier and energy of Est-Ia. Also Est-Ic was included in the investigations which will be discussed in the following.

11.8.1 Benchmark – Influence of steroid skeleton

The warheads were extended by estrone since Est-Ia showed the best results regarding docking and MD simulations. For reasons of comparability to the previous calculations, all warheads including estrone were tested (Figure 11.25). Those structures were optimized on B3LYP-D3 level using the def-TZVP basis set. As before the COSMO approach with a dielectric value of 78 was employed. Afterwards single point calculations have been carried out. The results of the reaction energies are presented in Table 11.12.

Table 11.12: Reaction energies obtained by QM calculation in COSMO. Energies marked with * are energies obtained with the best methodology compared to the CCSD(T). Aug-cc-pVDZ has been used as basis set. All values are given in kcal/mol.

aug-cc-pVDZ// COSMO	**Est-Ia**	**Est-Ib**	**Est-Ic**	**Est-Id**	**Est-Ie**
B3LYP-D3	-37.2	-5.1	-20.2	-47.1	-5.3
BHLYP-D3	-39.0	-4.9	-22.6	-49.9	-2.8
MP2	-39.1	-4.7*	-25.5	-47.3	-5.1
CC2	-39.2	-6.2	-25.8	-47.4	-6.4
SCS-MP2	-38.3	-1.1	-22.9*	-46.6*	-3.9
SCS-CC2	-38.2*	-1.7	-22.9	-46.6*	-4.8

Table 11.13: deviation of reaction energies including estrone skeleton compared to reaction energies of warheads in COSMO. Deviations marked with * are deviations obtained with the best methodology compared to the CCSD(T). All values are given in kcal/mol.

deviation to warhead	**Est-Ia**	**Est-Ib**	**Est-Ic**	**Est-Id**	**Est-Ie**
B3LYP-D3	0.3	1.1	1.4	0.5	0.5
BHLYP-D3	0.2	1.2	1.6	0.5	0.7
MP2	0.3	1.1*	1.5	1.0	0.5
CC2	0.4	0.8	1.4	1.0	0.4
SCS-MP2	0.2	0.9	1.5*	0.9*	0.6
SCS-CC2	0.3*	0.8	1.5	0.9*	0.5

The deviations compared to the warheads show low deviations for Est-Ia and Est-Id as they vary between 0.2 to 1.0 kcal/mol (Table 11.13). For Est-Ib they are around 1 kcal/mol and for acrylamide about 1.5 kcal/mol. For Est-Ie the difference in reaction energy is also

below 0.7 kcal/mol and therefore, one can conclude that the steroid does not have a significant effect on the reaction energy.

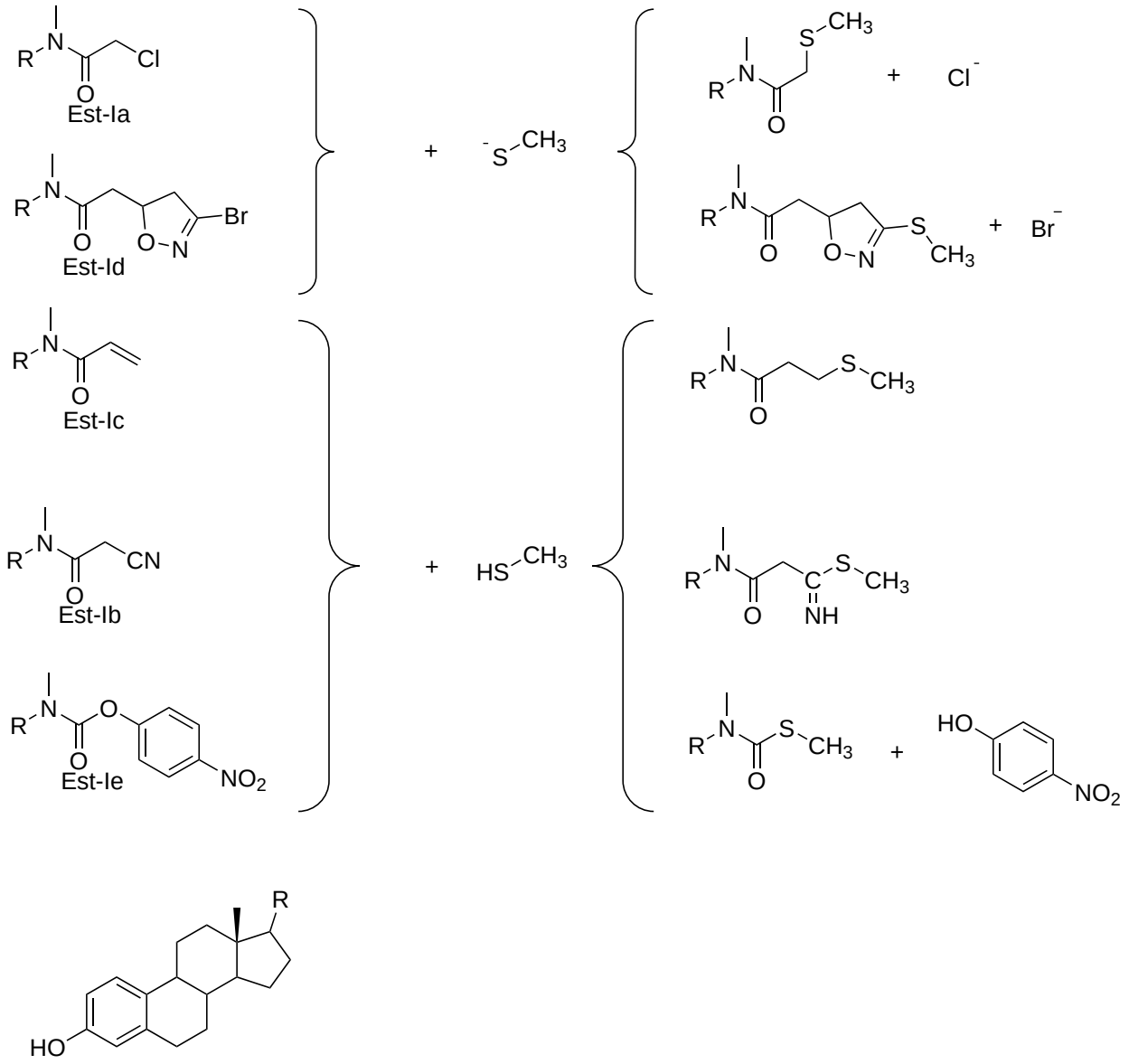


Figure 11.25: Calculated reaction of Est-Ia, Est-Ib, Est-Ic, Est-Id and Est-Ie in reaction with methanethiol or methanethiolate.

11.8.2 Semiempirical approach and NEB calculations

As above mentioned the NEB approach was selected to calculate the reaction barrier. To decrease computational costs, the semi-empirical method PM6 including Grimme's dispersion

correction D3 was employed for geometry optimization. To evaluate the suitability of this methodology reactants and products shown in Figure 11.25 were optimized on PM6-D3 level applying the COSMO approach with an ϵ-value of 78. These calculations employed the CAST program package, which utilized the MOPAC 2016 energy interface. The RMSD values in Table 11.14 are all below 0.25 Å and no significant, structural differences occur. These findings indicate the suitability of PM6-D3 regarding structure optimization. Afterwards single point calculations have been conducted using the PM6-D3 optimized structures to evaluate differences in reaction energy which might occur due to the change of methodology for structure optimization (Table 11.15).

Table 11.14: Root-mean-square deviation (RMSD) of PM6-D3//COSMO optimized structures compared to B3LYP-D3/def-TZVP//COSMO structures.

structure	**RMSD [Å]**
Est-Ic	0.18
CH_3S-Est-Ic	0.07
Est-Ia	0.13
CH_3S-Est-Ia	0.14
Est-Ib	0.06
CH_3S-Est-Ib	0.08
Est-Id	0.25
CH_3S-Est-Id	0.08
Est-Ie	0.12
CH_3S-Est-Ie	0.18
p-hydroxynitrobenzene	0.03
CH_3SH	0.02
CH_3S-	0.03

The differences in reaction energy compared to the B3LYP-D3 optimized structures are shown in Table 11.16. For Est-Ia and Est-Id their deviations are of about 1 to 1.5 kcal/mol. Est-Ie shows the lowest deviations of < 0.5 kcal/mol for the used DFT functionals, while ab initio methods are larger than 1.3 kcal/mol. In case of Est-Ib and Est-Ic they have a larger impact of ≈ 3 to 3.5 kcal/mol. Interestingly, the reaction energy for Est-Ic increases whereas Est-Ib's decreases. Still, the trends of inhibition are comparable to previous findings.

Table 11.15: Reaction energies obtained by QM calculation in COSMO. Energies marked with * are energies obtained with the best methodology compared to the CCSD(T). Aug-cc-pVDZ has been used as basis set. All energies shown are in kcal/mol.

aug-cc-pVDZ// COSMO	**Est-Ia**	**Est-Ib**	**Est-Ic**	**Est-Id**	**Est-Ie**
B3LYP-D3	-36.3	-2.7	-22.9	-48.6	-4,8
BHLYP-D3	-37.8	-2.1	-25.6	-51.0	-2.8
MP2	-38.0	-2.0*	-29.0	-48.3	-3.8
CC2	-38.2	-3.7	-29.3	-48.4	-4.4
SCS-MP2	-37.3	1.8	-26.2*	-47.8*	-2.6
SCS-CC2	-37.2*	1.1	-26.2	-47.8*	-3.0

Table 11.16: Absolute deviation of reaction energies compared to COSMO energies of B3LYP-D3/def-TZVP//COSMO optimized structures. Deviations marked with * are deviations obtained with the best methodology compared to the CCSD(T). All deviations shown are in kcal/mol.

deviation PM6-D3 optimization	**Est-Ia**	**Est-Ib**	**Est-Ic**	**Est-Id**	**Est-Ie**
B3LYP-D3	0.8	2.5	2.8	1.5	0,5
BHLYP-D3	1.1	2.7	3.0	1.1	0,1
MP2	1.1	2.7*	3.5	1.0	1,4
CC2	1.0	2.5	3.5	0.9	2,0
SCS-MP2	1.0	2.9	3.3*	1.2*	1,3
SCS-CC2	0.9*	2.8	3.3	1.2*	1,8

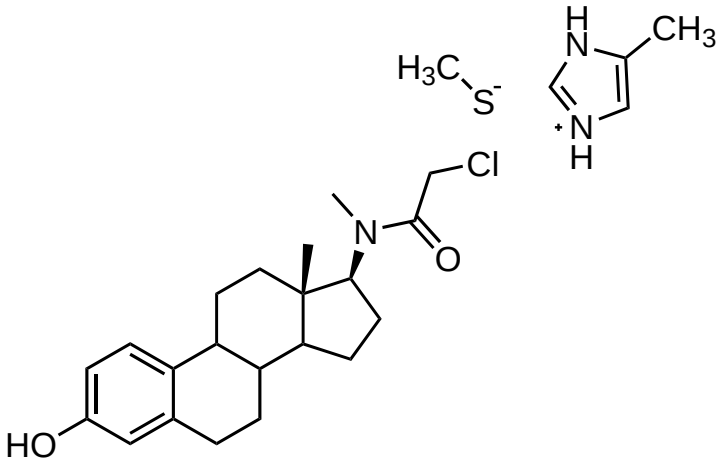


Figure 11.26: Extracted structures for optimization and NEB calculations. In case of Cys93 and His347 solely the side chain has been used.

The next step contained the calculation of the energy barrier of Est-Ia. Four structures including Est-Ia and the side chains of Cys93 and His347 (Figure 11.26) of the third MD run of protonation state B were employed for NEB calculations calculations. First of all, the structures of these frames have been optimized using PM6-D3//COSMO. Afterwards the covalent structure was modeled using Avogadro in the following way: The distance between electrophilic carbon of Est-Ia and sulfur was decreased to 1.9 Å and a constrained optimization between these atoms was performed. Subsequently an optimization without constrains was conducted to obtain a relaxed covalent structure. Thus, one aquires the starting- and endpoint for the NEB calculation. For each structure 20 NEB images were generated resulting in four different reaction paths (Figure 11.27). On each of these images single point calculations have been performed on SCS-CC2/aug-cc-pVDZ//COSMO level which performed best in benchmark calculations regarding the warhead chloromethylketone.

Table 11.17: Calculated reaction barrier and reaction energy of Est-Ia on SCS-CC2/aug-cc-pVDZ level employing COSMO. Single point calculations have been conducted on structures obtained by the NEB run on PM6-D3 level. Four different structures have been employed from the third MD run of protonation state B of Est-Ia.

aug-cc-pVDZ//COSMO	**TS**	**product**
Snapshot 1 (1 ns)	13.0	-28.3
Snapshot 2 (2 ns)	10.9	-20.6
Snapshot 3 (4 ns)	11.8	-24.8
Snapshot 4 (5 ns)	8.6	-20.7

The computed heights of the barriers of the reaction are ≈ 9 to 13 kcal/mol). The reaction energy decreases to -20 to -28 kcal/mol which implies that the product is about 10 kcal/mol less stable compared to previous results. Still, these calculations indicate that the reaction energy and barrier for covalent modification by Est-Ia is appropriate.

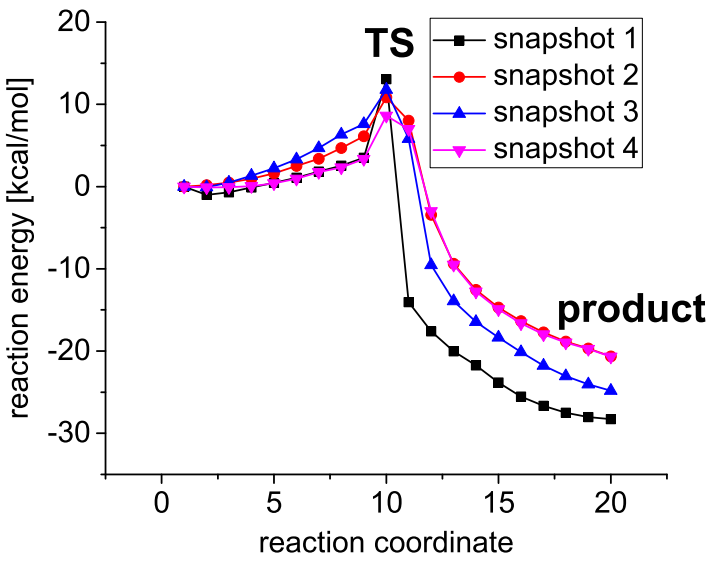


Figure 11.27: Calculated reaction path by a NEB calculation on PM6-D3 level.

11.9 QM/MM calculations of Est-Ia and Est-Ic

The final step contains the evaluation of Est-Ia's suitability to inhib FadA5. Thus, QM/MM calculations aimed to describe the reaction of Est-Ia in enzyme. As a starting point, a structure of the third MD run of Est-Ia (protonation state B) has been utilized for further calculations. The first step was the geometry optimization of Est-Ia, Cys93 and His347. The red colored QM region depicted in Figure 11.28 were optimized on B3LYP-D3 level with the def-TZVP basis set. The covalent bond cut between QM and MM part of Est-Ia had to be done as in Figure 11.28, as two C-C bonds could not be saturated. This explains why the warheads were extended by a methyl group in previous QM benchmark calculations. The AMBER14 force field was employed to describe the MM part. After optimization, a relaxed scan has been conducted by decreasing the distance between Cys93's sulfur and the electrophilic carbon of Est-Ia.

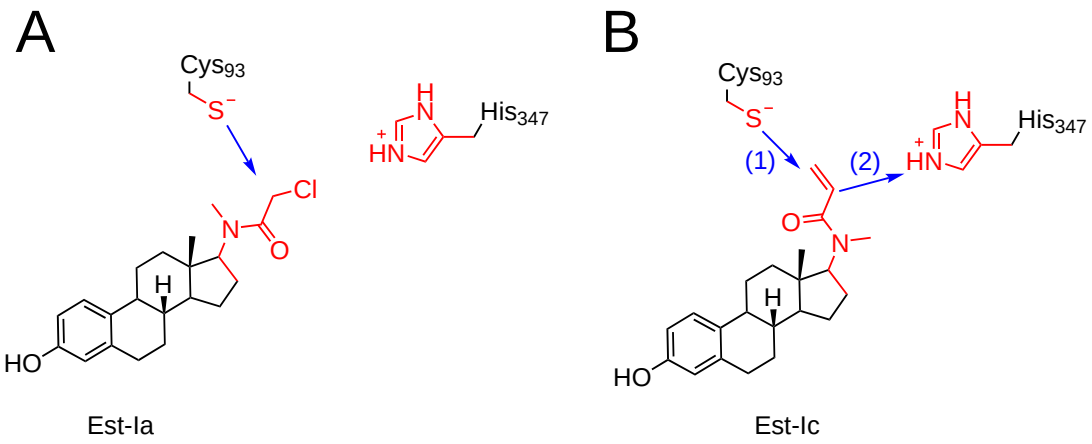


Figure 11.28: Red coloured lines indicate atoms, which has been described in the QM part of QM/MM calculation. **A** - Blue arrow indicates coordinate between nucleophilic sulfur of Cys93 and electrophilic carbon of Est-Ia, which has been decreased during the relaxed scan. **B** – First reaction coordinate (1) is the decrease in distance between nucleophilic sulfur of Cys93 and electrophilic carbon of Est-Ic. Afterwards the distance between carbon and hydrogen of His347 has been decreased (2). Location of Est-Ia, Est-Ic and amino acids serves for clarification and does not reflect actual position in QM/MM calculation.

The resulting reaction barrier of 28.8 kcal/mol is about 15 to 20 kcal/mol higher compared to the ones obtained by NEB calculation (Table 11.18). It is to be expected that this energetic hurdle is hard to overcome. Furthermore, the reaction energy of 0.1 kcal/mol indicate that the formation of a covalent bond between Est-Ia and Cys93 is not to be expected. This is in contrast with previous results. In case of reaction energies, the nucleophilic attack in vacuo is -31.5 for SCS-CC2/aug-cc-pVDZ. The application of COSMO leads to reaction energies of $\approx$ 38 kcal/mol. For NEB calculations the reaction energy is in a range of $\approx$ 20 to 28 kcal/mol. One of the major differences to previous calculations is the stabilization of the chlorine by implicit solvation. As mentioned in section 11.5.2 the absolute deviation between chlorine in vacuum and COSMO is larger compared to neutral species. To overcome this problem, two explicit water molecules were included to the QM region to stabilize the chlorine. By doing so one obtains a decrease of 12.7 kcal/mol in respect to the transition state and a reaction energy of -4.0 kcal/mol. In terms of kinetics, one can expect that a reaction barrier of 16.1 kcal/mol can be overcome if water molecules stabilize the chlorine. Still, the reaction energy of -4.0 kcal/mol is in contrast to previous findings. This leads to the question if the QM or QM/MM approach leads to false results.

To tackle this problem, the reaction energy and barrier of Est-Ic was calculated, applying

Table 11.18: Reaction energy and barrier on QM/MM level obtained by relaxed scan. Geometry optimization of structures were conducted on B3LYP-D3 level and a def-TZVP basis set. Energy evaluation was performed by single point calculation of initial structure (non-covalent complex), transition state (TS) and end of reaction (covalent complex). For Est-Ia SCS-CC2 with the aug-cc-pVDZ basis set was employed, for Est-Ic the SCS-MP2 method with the aug-cc-pVDZ basis set. All values are given in kcal/mol.

	Est-Ia SCS-CC2/ aug-cc-pVDZ	Est-Ia + 2 H2O SCS-CC2/ aug-cc-pVDZ	Est-Ic SCS-MP2/ aug-cc-pVDZ
non-covalent complex	0.0	0.0	0.0
TS	28.8	16.1	16.8
covalent complex	0.1	-4.0	-37.9

the same QM/MM method. Compared to Est-Ib or Est-Ie, the reaction energies are in a range of ≈ -23 to -26 kcal/mol and therefore lower. Additionally, it harbors poses in the docking, in which the warhead is directed to the active site. In terms of MD simulations one could see that the distance between Cys93's sulfur and Est-Ic electrophilic carbon does increase and a nucleophilic attack is not likely. To overcome this fact a harmonic distant restraint was applied on Cys93's sulfur and the electrophilic carbon to obtain a relaxed structure in which Est-Ic's warhead can be covalently modified. Of course one has to keep in mind that this is an artificial state which has been forced to calculate the reaction energy and barrier. The QM/MM system was prepared as before but without the inclusion of water molecules to the QM region. The relaxed scan was devided in two parts. Firstly, the distance between Cys93's sulfur and Est-Ic electrophilic carbon was reduced to 1.9 Å. Afterwards, it was kept fixed and the distance between His347's hydrogen was moved to Est-Ic (Figure 11.28). Instead of SCS-CC2 the spin component scaled variant of MP2 with the aug-cc-pVDZ basis set was applied, as this methodology was the best in comparison with CCSD(T). The resulting reaction barrier of 16.8 kcal/mol is comparable with the one obtained of Est-Ia including two water molecules. Regarding the reaction energy of -37.9 kcal/mol it is much lower compared to Est-Ia or previous calculation of the warhead or Est-Ic in vacuo or COSMO. These findings indicate that differences of Est-Ia (in terms of reaction energy) do not occur because of the application of QM/MM methods, but rather because of the enzyme's environment.

11.10 Conclusion

In this chapter first steps have been taken for the development of a covalent inhibitor originating from a non-covalent ligand for FadA5 (a thiolase of *Mycobacterium tuberculosis*). The group of Kisker from the University of Würzburg crystallized the steroid OPC (3-oxo-preg-4-ene-20-carboxylic acid) and CoA in non-covalent complex with the C93S mutant of FadA5 and identified a steroid binding pocket. This X-ray structure served as a starting point for the following computational investigations.
To modify FadA5 covalently, the active site dyad consisting of Cys93 and His347 appear to be suitable and therefore three different inhibitor scaffolds are considered which could be expanded by a warhead to result in a covalent inhibitor:

- **Scaffold A**: A warhead could be attached to the thiol group of CoA (Figure 11.6-A).
- **Scaffold B**: One can think of linking the warhead to a steroid scaffold (Figure 11.6-B).
- **Scaffold C**: The warhead could be placed between the steroid and the coenzyme. (Figure 11.6-C).

As CoA occurs in many biological reactions in human, problems in respect to the selectivity can arise if scaffold A is used. Furthermore, the synthesis might be very complex as in the case of scaffold C. Hence, both scaffolds were neglected. Scaffold B offers the advantage of a range of purchasable steroids which facilitates experimental work by decreasing the steps of synthesis and was further pursued for investigation. In the course of this study, 36 synthesizable structure related to scaffold B were investigated. These possible inhibitors occur due to the combination of five different warheads adressing cysteine (acrylamide, chloromethylketone, cyanamide, carbamate and bromodihydroisoxazole) and six different steroids (progesterone, pregnenolone, prasterone, 3-methoxy-prasterone, estrone and 3-methoxy-estrone) as depicted in Figure 11.10 to Figure 11.15.
At the start, MD simulations have been conducted for the C93S FadA5 mutant in complex with OPC and CoA or for the FadA5 wild type in complex with OPC, CoA or with both (Figure 11.8). These simulations are needed to probe the stability of OPC in wild type as well as in absence of CoA. Results showed that the RMSD of OPC does not increase strongly in wild type but in absence of CoA. Still, OPC remains in the steroid binding pocket so that one can conclude that an attached warhead can also be directed towards the active site for covalent modification.
Afterwards, the covalent modification of methanethiol (representing the cysteine side chain)

by above mentioned warheads have been calculated on QM level to provide insight on the reactivity of each warhead and additionally, to determine the most accurate methodology for selected warheads. In most cases spin-component scaled MP2 or CC2 offer the best results in comparison to CCSD(T). In terms of cyanamide they show the largest error of about 4 kcal/mol. Furthermore, the application of implicit solvation revealed different results depending on the investigated warhead. In terms of acrylamide and cyanamide one obtains a decrease in reaction energy. For bromodihydroisoxazole a small increase occurs, while the reaction energy of the carbamate warhead increases by a value of 3 kcal/mol. The largest influence of the solvent can be observed for chloromethylketone as a downshift of 6 kcal/mol is obtained after employing COSMO. This can be explained by the stabilization of the chlorine due to the implicit solvation.

As the warheads appear suitable for covalent modification the next step was to establish a suitable redocking protocol for OPC which can be transferred to the docking of proposed inhibitors. This procedure is needed to obtain meaningful poses which can be used for MD simulations to probe the stability of the inhibitors in complex with FadA5. In this matter the specification of the approximate radius r of the binding site was crucial. Best results in the redocking were obtained if a r-value of 20 Å was applied leading to high-ranked poses of OPC with RMSD values below 1 Å compared to the X-ray structure.

This setup was used for the docking of the proposed inhibitors. In contrast to OPC, good poses of the inhibitors were referred to those which showed a small distance between the electrophilic part of the warhead and the nucleophilic sulfur of Cys93. This is understandable, as the probability for covalent modification is enabled if a sufficient small distance is given. The docking simulations of the inhibitors revealed different findings. Bromodihydroisoxazole appears to be unsuitable as warhead for this scaffold, as it is too huge to fit in the active site resulting in poses which are directed to the solvent. This behavior was observed for many poses of the remaining inhibitors and thus the idea came up to design "sandwich-structures" which are modified by two warheads at the A- or D-ring. Although the docking poses of these "sandwich-structures" did not show sufficient distances for covalent modification, all poses showed that at least one warhead is always directed towards the active site. Best results were obtained for five inhibitors which consist of estrone or 3-methoxy-estrone (Est-Ia, Est-Ib, Est-Ic, Est-IIa and Est-IIc, see Figure 11.12). The poses of these inhibitors were used for MD simulations to probe if the small distance between warhead and sulfur can be maintained in the course of simulation time. The MD simulations revealed that only Est-Ia fulfilled this requirement.

The next question which needed to be answered was the covalent reaction of Est-Ia itself. In

order to do so, preceding calculations had to be conducted. As the first QM benchmark only focused on the warhead, calculations have been extended by including the steroid scaffold estrone. This is because docking and MD simulations revealed Est-Ia to be the best candidate. Still, all warheads were included for reasons of comparability to previous calculations. These computations showed that differences in reaction energy are largest by a value of 1.5 and therefore indicate that the attachment of the steroid does not affect the reactivity of the warheads significantly.

Afterwards, investigations on semi-empirical and QM/MM level have been conducted to gain insight in regard to the covalent modification. Reaction energies as well as the reaction barriers of Est-Ia obtained by NEB calculations on semi-empirical level are contrary to the findings on QM/MM level. In terms of the reaction barrier QM/MM calculations reveal that it is about 15 kcal/mol higher compared to semi-empirical results. The inclusion of two explicit water molecules in the QM region decreased the barrier; hence solvents are relevant for the kinetics of Est-Ia. Also, the reaction energy is overestimated in case of semi-empirical results as they are of -20 to -28 kcal/mol while they are much smaller for QM/MM computations (0.1 kcal/mol in absence of explicit water molecules or -4.0 kcal/mol in presence of two explicit water molecules). This shows that the stabilization of the chlorine molecule is of importance. In terms of the QM or semi-empirical calculations implicit solvation was applied thus offering a perfect stabilization of the chlorine which is not the case for Est-Ia in the QM/MM calculation of the protein-inhibitor complex.

To gain further insight on this effect the covalent modification of FadA5 by Est-Ic was also calculated on QM/MM level. In contrast to Est-Ia, calculations of Est-Ic revealed an appropriate reaction barrier of 16.8 kcal/mol and especially a reaction energy of -38 kcal/mol. Additionally, results show that acrylamide is not prone to the solvent as chloromethylketone. However, dynamics show that fluctuations in distance are too large for a nucleophilic attack for Est-Ic although acrylamide appears to be a fitting warhead.

The last step of this study was conducted by the Tonge group from the Stony Brook University. They tested experimentally the inhibition of Est-Ia, Est-Ic and Est-Ie in FadA5 but did not see any meaningful inhibition. After two hours of incubation all tested compounds showed $< 50\%$ irreversible inhibition. Additionally, no competitive inhibition was observed for these structures. This is in accordance with the calculated results. For Est-Ie all docked poses showed, that the warhead itself is not close enough for covalent modification and therefore would not react. For Est-Ic the reaction energy and barrier seems favorable, but the distance for a nucleophilic attack of Cys93 sulfur cannot be maintained during the MD simulations. This might explain the experimental results. For Est-Ia one might expect some

inhibition, although the reaction energy is not very large.
These experimental findings reveal that research is still necessary for the development of a covalent inhibitor. As mentioned above, no competitive inhibition was observed in the experiments and a way to overcome this problem might be the change of the inhibitor scaffold. Instead of solely focusing on the steroid scaffold it might be essential to include the CoA moiety as described in Figure 11.6 although the complexity of synthesis increases. This shows that further research is needed in regard of the development of a covalent inhibitor for FadA5.

Chapter 12

Theoretical investigations on the modulation of the product profile of *Bacillus megaterium levansucrase* by ene-reactions

12.1 Introduction

The importance of carbohydrates in living organisms is shown in many forms. On the one hand, they are of necessity for the structural integrity e.g. as cellulose in cell wands of plants or as chitin in fungi or bugs [268]. On the other hand, they are essential for energy storage in form of glycogen in humans. Carbohydrates are classified by their size. If their polymerization degree (DP) is > 9, they are named polysaccharides, if they consist of 3 to 9 units, they are called oligosaccharides. In terms of smaller units, one refers to mono-, di- and trisaccharides which are compromised of e.g. glucose, galactose, fructose or xylolose. The smallest building block of oligo- or polysaccharides are called monosaccharides. They are organized by number of atoms, carbonyl groups and chirality. Because of the number of hydroxyl groups, one obtains a variety of possibility of connectivity and by implication a series of complex structures.

In cosmetics, bacterial polysaccharaides are used as they mostly do not show toxicity and posess appropriate physicochemical properties for the formulation of cosmetics as well as good tolerability [269]. To obtain these products, microbacterial fructosyl transferases (EC 2.4.1.99) are used; polymerases which catalyze the biosynthesis of fructans [270]. Two known

fructan sucrases are levansucrase and inolsucrases. The synthesis of levan occurs by ß-(2,6)-linked fructosyl residues by levansucrase (EC 2.4.1.10) [271] (Figure 12.1). Inulines are mainly produced by inolosucrases (EC.2.4.1.9) [272], [273].

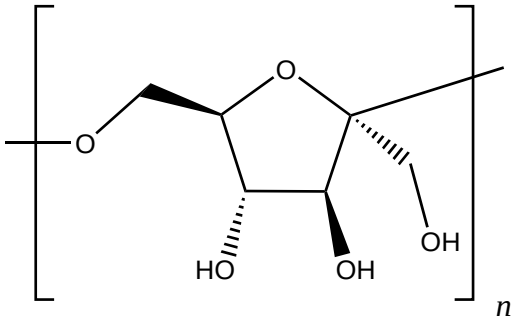


Figure 12.1: Chemical structure of mono-unit of levan.

The heterologusly exprimed levansucrase of *bacillus megaterium* (Bm-LS) of the glycoside hydrolase family GH68 is known to catalyze the synthesis of a number of fructo oligosaccharides. The range of the carbohydrate size ranges from 2 to 20 fructose units [274]. Strube et al. have performed an extensive mutation study of Bm-Ls, to obtain insight on the elongation of the carbohydrate polymer and to evaluate relevant interactions in the process of polymerization [275]. The active site of Bm-LS consists three catalytic amino acids Asp95, Asp257 and Glu352. At this Asp95 serves as catalytic nucleophile, while Glu352 acts as acid/base catalysator [274]. The interaction of the 3- and 4-OH group of the fructosyl unit in the active site is attributed to Asp257, which additionally stabilizes the transition state [272], [273]. By the mutation of selected amino acids and the analysis of the synthesized carbohydrates of Bm-LS variants, it was possible to draw conclusions on the functionality of respective residues (Figure 12.2). While Asn252 shows an interaction motif for tri-/tetrasaccharides, Lys373 shows this behavior for penta-/hexasaccharides and Tyr247 for nona/decasaccharides. In terms of Arg370, this residue is important for the stabilization of the carbohydrate in the progress of elongation. Therefore, it was possible to predict the direction of Bm-LS polymerization progress which appears in a clockwise motion. These insights in respect to the biosynthesis of levan in Bm-LS are of importance for further investigations.
In the past many attempts of molecular engineering of enzymes showed improvements concerning activity, specifity and stability [276]. E.g. chemical modification of proteins offers a way for accessing tailored bio-catalysts for different applications. In this manner tyrosine or tryptophane residues are suitable for conjugation as their occurrence is less than usually con-

jugated residues as cysteine, threonine, serine and lysine and often are buried when located at the surface [277], [278], [279], [280], [281], [282], [283]. 25 tyrosine residues can be found in Bm-LS although only residues 41, 196, 247, 281 and 446 are the most solvent exposed ones [284]. The group of Seibel from the University of Würzburg conducted investigations on the site-selective chemo modification of these tyrosine residues in Bm-LS. In this respect different scientific issues arose and were additionally investigated by computations which will be discussed in this section.
In the beginning the bioconjugation of tyrosine residues of wild type Bm-LS have been conducted by introducing modifier 1 (7-((6-aminohexyl)oxy)-2-methyl-2,3-dihydrophthalazine-1,4-dione with an azine group). The mass spectrometry detected two modified tyrosine residues although the exact residue could not be determined. In consequence, computations were necessary to shed light into this matter.

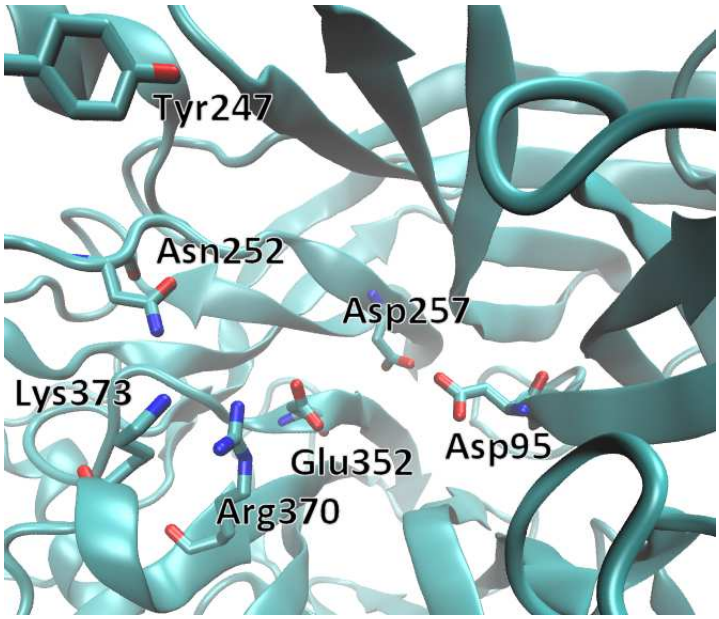


Figure 12.2: Active site residues of levansucrase from Bm-LS, which interacts with the elongation process of the fructooligosacharides.

Afterwards, possible key residues have been mutated experimentally to obtain new accessible tyrosine residues for bioconjugation. In these cases, modifier 1 or modifier 2 (2-methyl-6-(pent-4-yn-1-yloxy)-2,3-dihydrophthalazine-1,4-dione with an alkyne group) have been investigated *via* MD simulations and have been compared with experimental findings regarding the change of product spectra.
Lastly, the group of Seibel showed that different levansucrase variants were able to synthesize

polysaccharides due to the chemical modification of modifier 2 or 2-AzGlc. As mentioned before the products of wild type Bm-LS are oligosaccharides with a DP of 2-20 and by mutation and bioconjugation polysaccharides were obtainable. Thus, investigations of these covalent modifier-protein complexes were also included in the computations (Figure 12.3).

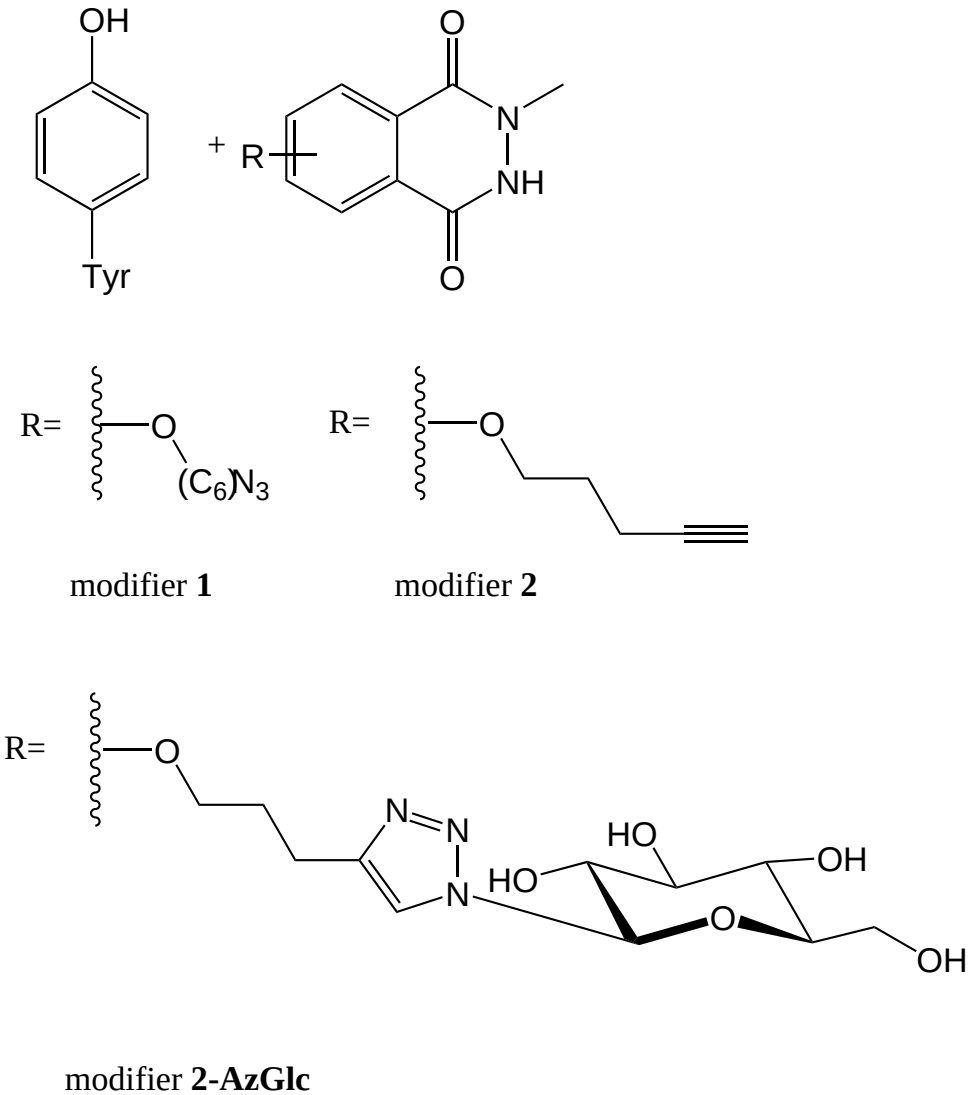


Figure 12.3: Tyrosine targeting structures as 4-phenyl-3H-1,2,4-triazole-3,5(4H)-dione (PTAD) = modifier 1, 6-(6-azidohexyl)-2-methyl-2,3-dihydrophthalazine-1,4-dione = modifier 1 methyl-6-(pent-4-yn-1-yloxy)-2,3-dihypdrophthalazine-1,4-dione) = modifier 2 and methyl-6-(pent-4-yn-1-yloxy)-2,3-dihydrophthalazine-1,4-dione) modified with „click reaction“ with 1-azido-1-deoxy-ß-D-glucopyranoside = modifier 2-AzGlc.

12.2 General Information on MD simulation

One focus of investigation was the determination of the tyrosine residues which were modified in levansucrase wild type. The actual position of modifier 1 has been investigated by covalently connecting it to different tyrosine residues (Tyr196, Tyr247 and Tyr446) of Bm-LS. Afterwards Molecular Dynamic simulations with duration of 10 ns for each investigated covalent complex have been conducted. As a starting point, the crystal structure of the D257A Bm-LS mutant has been used (pdb code 3OM2) and mutated to the wild type by changing Ala257 to Asp257 (Figure 12.4). Modifier 1 was afterwards attached on each of the above-mentioned tyrosine residues, employing the Avogadro program package. If possible, the modifier was bound to C3 or C5 of the tyrosine side chain. In the course of this procedure, the nearby protein environment (5 Å radius of the corresponding tyrosine residue) was considered to avoid atom clashes and unlikely binding modes were neglected. As mentioned before mass spectrometry determined two modified tyrosine residues. Therefore, the simulation considered single and twofold modifications of tyrosine residues by modifier 1. Additional aspects have to be considered in the simulations because the structure of the modifier is present as a stereoisomer. In consequence, the 1-azido-5-methoxypentane group can be present at position 6 or 7 of the phthalazinedione scaffold which also has been taken into account for most cases (Figure 12.5).
Bm-LS was described by application of the Amber14 force field ff14SB whereas modifier 1, 2 or 2-AzGlc were parameterized by GAFF. The same protocol for minimization and gradual heating was used described in section 5.1.

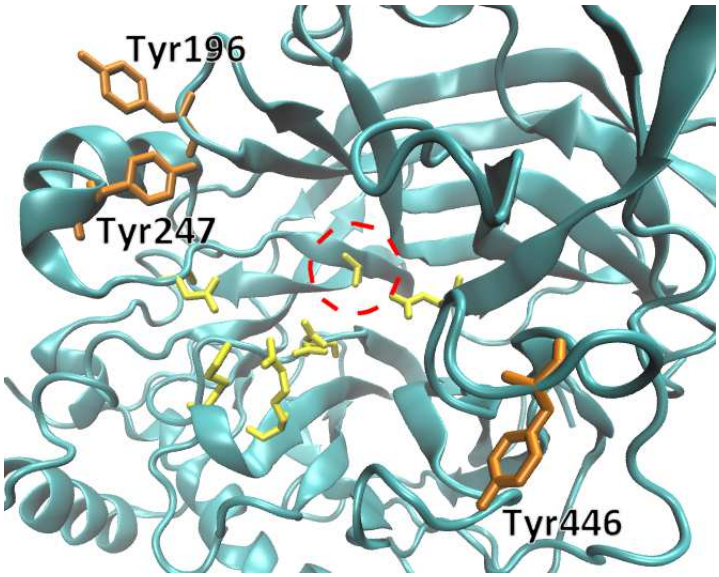


Figure 12.4: Modification of D257A mutant of *Bacillus megaterium* levansucrase to wild type protein (dotted circle shows position of Ala257/Asp257). Possible tyrosine residues which can be modified by modifier 1 are depicted in orange licorice style. Amino acids of the active site are presented in yellow licorice style.

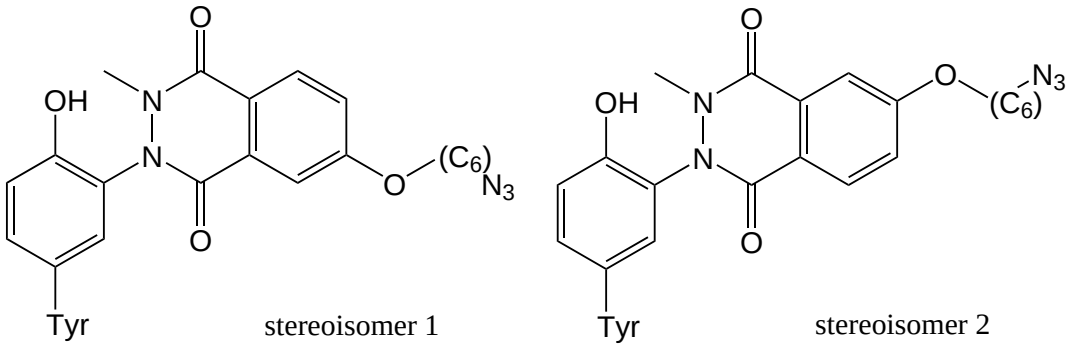


Figure 12.5: Stereoisomers of modifier 1. Stereoisomers of modifier 2 and 2-AzGlc are obtained by exchanging the 1-azido-6-methoxyhexane

12.3 Covalent modification of wild type Bm-LS with modifier 1

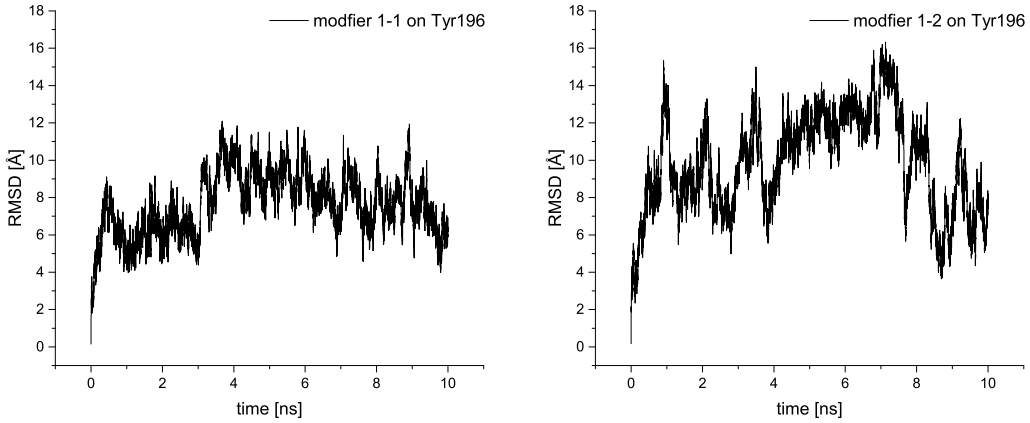


Figure 12.6: Course of RMSD values versus time plots of modifier 2-1 or 2-2 covalently bound to Tyr196 in B. megaterium levansacucrase during the 10 ns production run.

First of all, the single modification of Tyr196, Tyr247 or Tyr446 was investigated to gain insight on the protein ligand interaction. By attaching modifier 1-1 (stereoisomer 1) or 1-2 (stereoisomer 2) on Tyr196 one cannot observe any movement towards the active site. In contrast the modifier orients itself towards the solvent. Figure 12.6 shows in terms of the calculated RMSD values larger deviations of stereoisomer 2 than for 1. Still, these variations do not change the fact that the single modification of Tyr196 does not harbour any effect on the polymerization reactions in Bm-LS and thus not affecting the product profile of Bm-LS as no interaction of modifier 1 to any amino acid of the active side can be observed (see Figure 12.7).

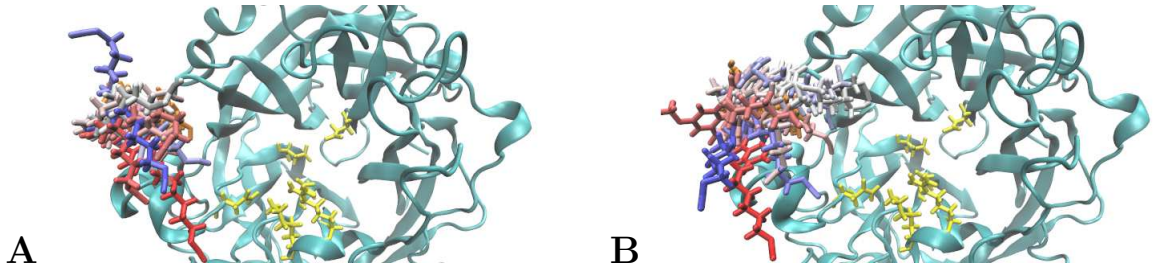
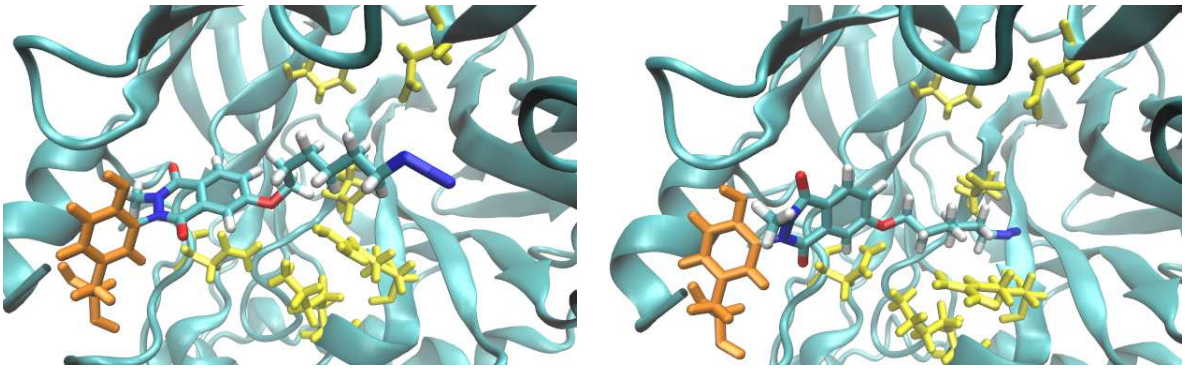


Figure 12.7: **A** – MD simulation of modifier 1-1 covalently bound to Tyr196. **B** - MD simulation of modifier 1-2 covalently bound to Tyr196. Every ns of the temporal progress presented from red-coloured (start of simulation) to blue-coloured modifier 1 (end of simulation). Initial frame of active site residues is shown for clarification in yellow licorice presentation.

For the modification of Tyr247 two different possibilities of the linkage of modifier 1 are enabled. In Figure 12.8 one can see that the ligand can be attached directing towards the solvent or in direction of the active site. By including both stereoisomers of the structure, one has to consider four different simulations in for the covalent modification of Tyr247. A glance at the variation of RMSD values in Figure 12.9 shows that the stereoisomerism has a subordinate role in these four complexes. In terms of modifier 1-2 the increase of RMSD occurs after a certain amount of time (after $\approx$ 4 ns for pose I or after $\approx$ 6 ns for pose II) whereas for modifier 1-1 this is observable in an earlier stage of the molecular dynamics simulation. Of more importance is the pose of the ligand. In case of pose I the increase of RMSD value goes along with the movement of the ligand leading away from the active site. By attaching modifier 1 in pose II to Tyr247 the increase of RMSD values is lower. Figure 12.10-B and Figure 12.10-D show that despite of the stereoisomerism modifier 1 is pointing towards the active site. Thus, it is likely that this pose leads to interactions with the polymerization process of the polysaccharide and therefore has an effect on the product profile. Still, one has to consider that a certain amount of flexibility in terms of modifier 1 is available as fluctuations in the RMSD plots are in a range of 8 Å.



pose I pose II

Figure 12.8: Possible conformations (pose I or pose II) of modifier 1 covalently bound to Tyr247 depicted in orange licorice presentation close to amino acids of the active site shown in yellow licorice presentation.

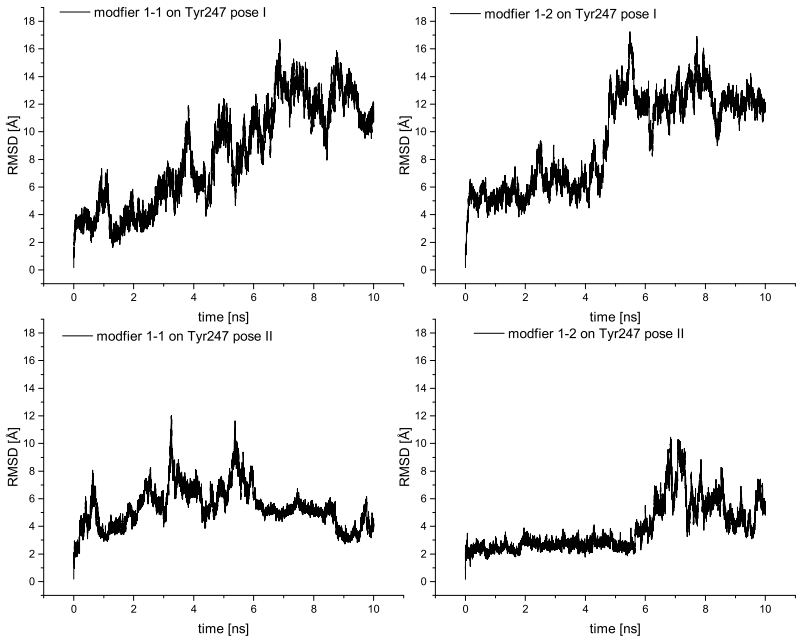


Figure 12.9: Course of RMSD values versus time plots of modifier 1-1 or 1-2 in pose I or pose II covalently bound to Tyr247 in B. megaterium levansucrase during the 10 ns production run.

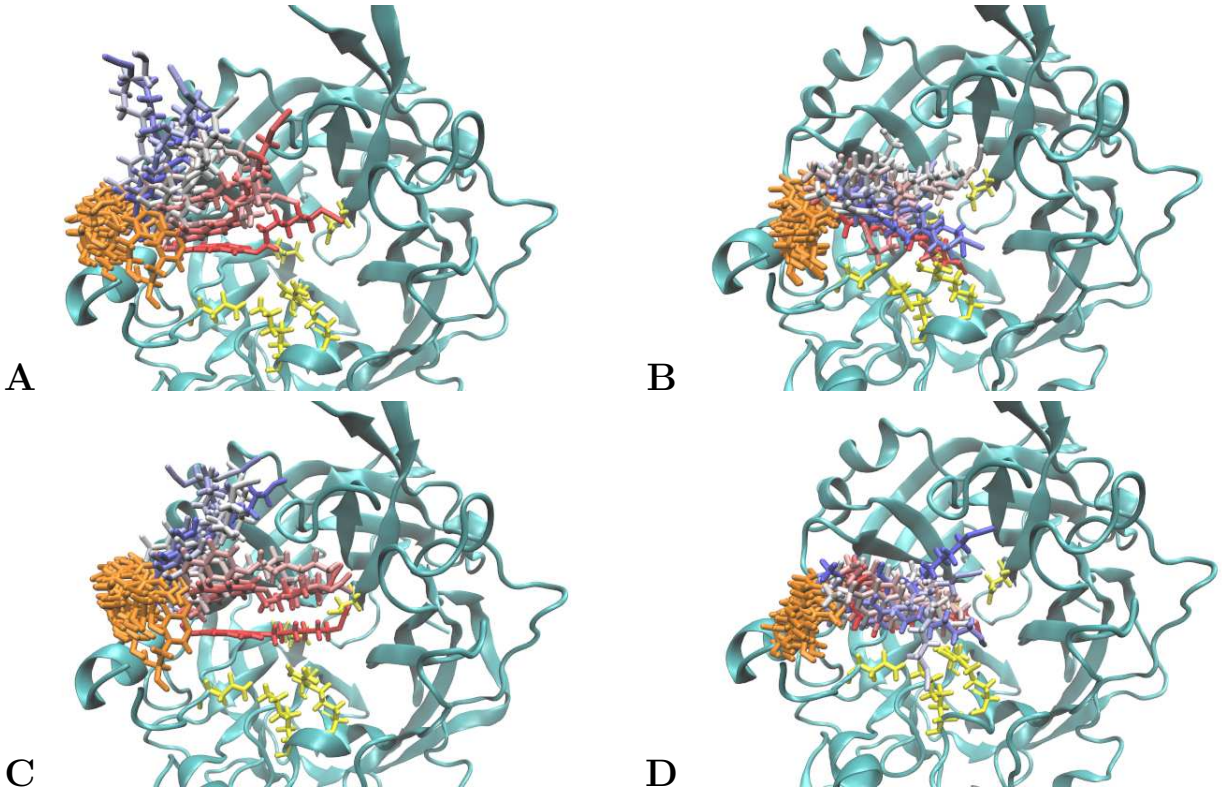


Figure 12.10: **A** – MD simulation of modifier 1-1 in pose I bound to Tyr247. **B** – MD simulation of modifier 1-1 in pose II bound to Tyr247. **C** – MD simulation of modifier 1-2 in pose I bound to Tyr247. **D** – MD simulation of modifier 1-2 in pose II bound to Tyr247. Every ns of the temporal progress presented from red-coloured (start of simulation) to blue-coloured modifier 1 (end of simulation). Initial frame of active site residues is shown for clarification in yellow licorice presentation.

The last simulation was of the modified Tyr446 residue. Similar to the covalent modification of Tyr196 one can only observe a movement of the ligand into the opposite direction of the active site. The RMSD plots of both stereoisomers show fluctuations in the same range and also lead to the conclusion that the stereoisomerism is not of importance in this case. Therefore, it is likely that the single modification of Tyr446 also has no effect on the product profile of Bm-LS.

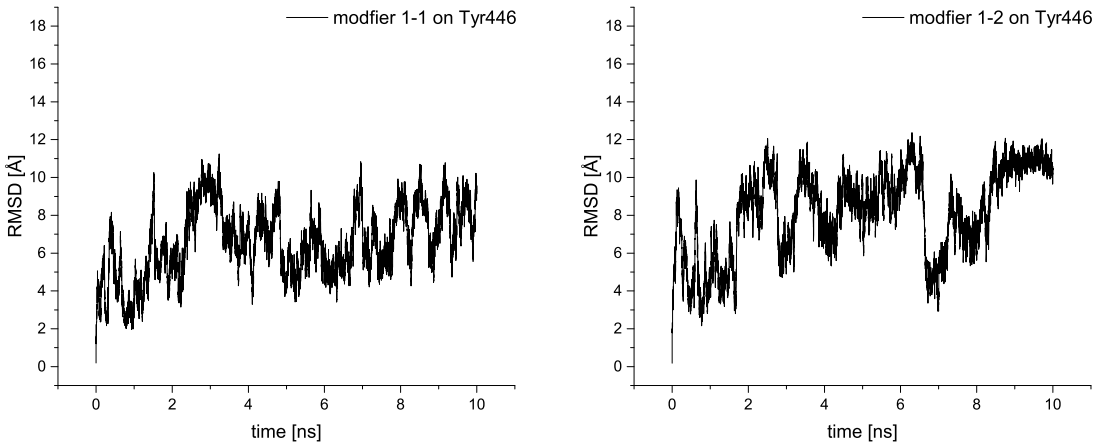


Figure 12.11: Course of RMSD values versus time plots of modifier 1-1 or 1-2 covalently bound to Tyr446 in B. megaterium levansucrase during the 10 ns production run.

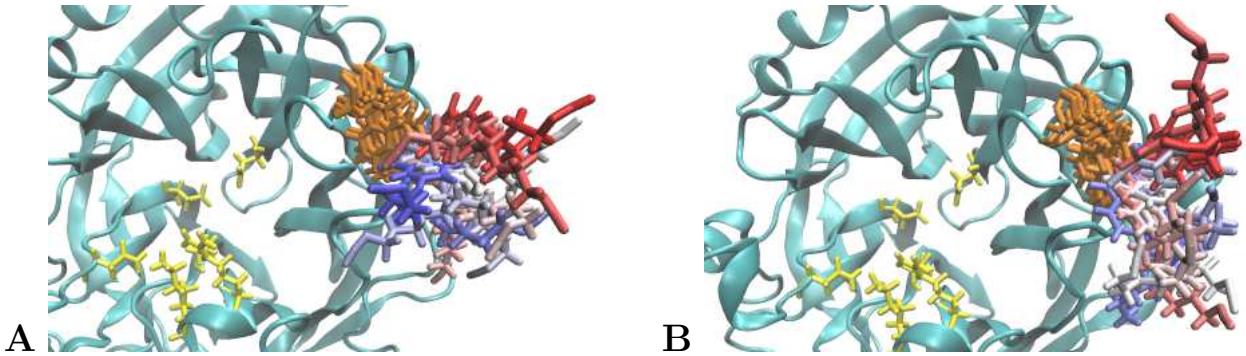


Figure 12.12: **A** – MD simulation of modifier 1-1 covalently bound to Tyr446. **B** - MD simulation of modifier 1-2 covalently bound to Tyr446. Every ns of the temporal progress presented from red-coloured (start of simulation) to blue-coloured modifier 1 (end of simulation). Initial frame of active site residues is shown for clarification in yellow licorice presentation.

The next step was the simultaneous modification of two tyrosine residues by modifier 1 to describe the simulations more correctly in relation to experimental findings. The previous MD simulations revealed that the single modification of Tyr196 or Tyr446 does not appear to have any influence on the active site thus to a modified production profile. Therefore, a double modification of both tyrosine residues at the same time was not conducted.
In the following, a twofold modification of Tyr196 and Tyr247 has been performed. Contrary to the previous simulations solely stereoisomer 1 has been considered as the stereoisomerism does not have a significant impact on the interaction with the active site in the case of the

modified Tyr247. More of interest is the pose of the covalent ligand on Tyr247. Thus, only two simulations resulted in different poses of modifier 1-1 on Tyr247. Interestingly one can observe a motion in both cases in regard of modifier 2-1 on Tyr196 as it moves to the direction of the second modifier 1-1 on Tyr247. The reason for this change of position is a π-π interaction between the phthalazinedione scaffold of modifier 1-1 on Tyr196 and Tyr247's phenyl ring resulting in a parallel alignment (see Figure 12.14). In contrast to the single modification of Tyr247, one can observe that modifier 1-1 in pose I shows an increase up to 8 Å in RMSD value (instead of a steady to 14 Å) and a relatively stable course after 4 ns (black curve in Figure 12.13). This stabilization is caused by the above mentioned π-π interaction and additionally leads to a redirection of modifier 1-1 (on Tyr247) closer to the active site. The second modifier shows a large increase of RMSD after $\approx$ 500 ps which goes along with the reorientation towards the second modifier. In the second case of the simultaneous modification of Tyr196 and Tyr247 in pose II the same movement of 2-1 on Tyr196 also occurs due to the π-π interaction. In the course of RMSD values during the simulation fluctuations of 1-1 on Tyr196 are larger than before as the π-stacking at first evolves after 2 ns and requires a certain amount of time to consolidate. In contrast to the single modification of Tyr247 one can observe a continuous increase of RMSD of the covalent ligand on this amino acid due to the flexibility of the 1-azido-5-methoxypentane group. Nevertheless, it is directed towards the active site and should also lead to an interaction with the polymerization process.

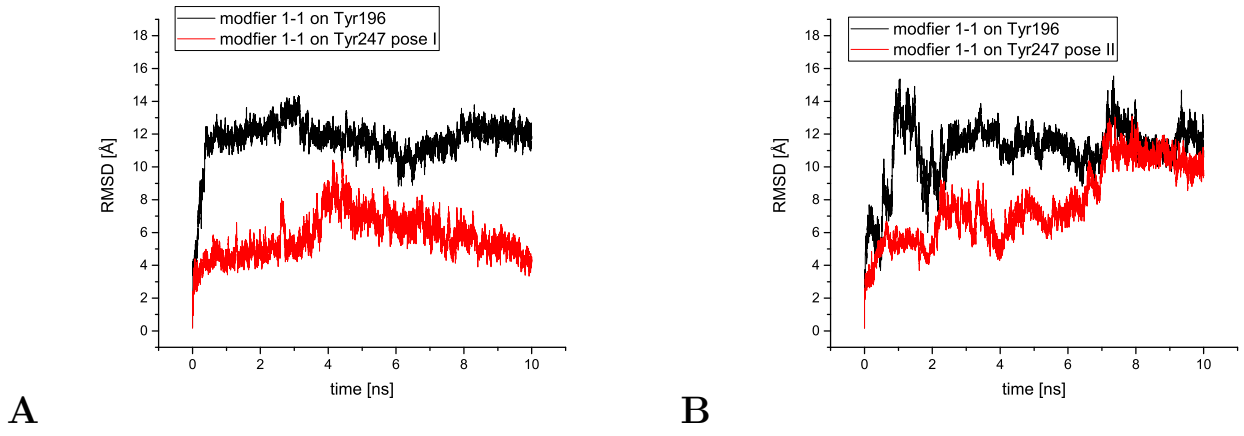


Figure 12.13: Course of RMSD values versus time plots of modifier 1-1 covalently bound to **A** - Tyr196 (black curve) and Tyr247 in pose I (red curve) or **B** - Tyr196 (black curve) and Tyr247 in pose II (red curve) in Bm-LS during the 10 ns production run.

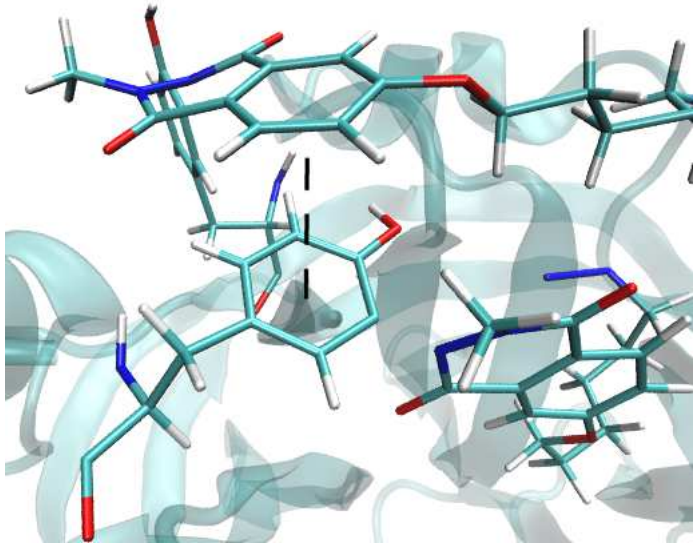


Figure 12.14: Observed π-π interaction between phthalazinedione scaffold of modifier 1-1 covalently bound to Tyr196 and phenyl ring of Tyr247 indicated by red dotted lines during 10 ns production run.

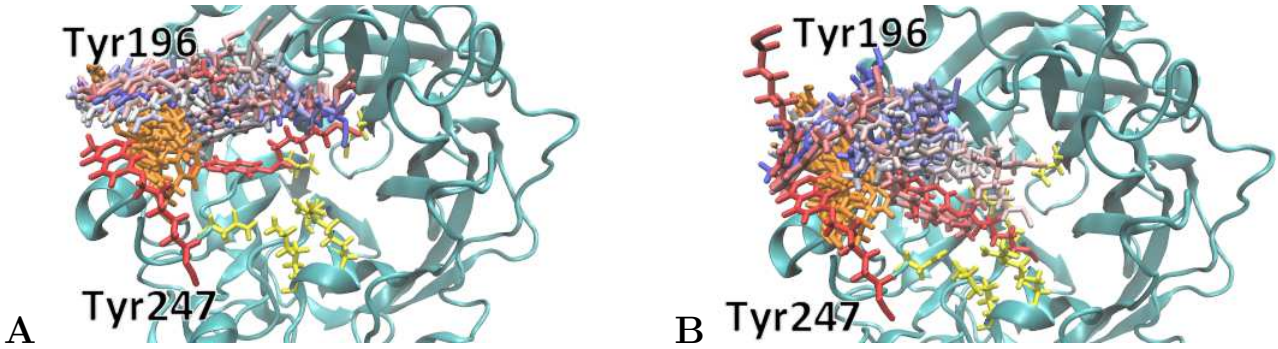


Figure 12.15: **A** - MD simulation of modifier 1-1 covalently bound to Tyr196 and Tyr247 in pose I. **B** - MD simulation of modifier 1-1 covalently bound to Tyr196 and Tyr247 in pose II. Every ns of the temporal progress presented from red-coloured (start of simulation) to blue-coloured modifier 1 (end of simulation). Modified tyrosine residues are depicted in orange licorice presentation, initial frame of active site residues in yellow licorice presentation for clarification.

In contrast to these findings, a twofold linkage of 2-1 on Tyr247 and Tyr446 does not lead to any interaction. As before, only the stereoisomer 1 was considered and two simulations with the difference of the position of 1-1 on Tyr247 were performed. In terms of the modification of Tyr446 both cases show RMSD values in the range of 2.5 to 10 Å similar to the single

modification of this residue (black curve in Figure 12.16). As it is shown in Figure 12.17 the modifier is directed to the solvent and does not move towards the active site during the simulation time. For Tyr247 one can obtain further insights. At first it was expected that the modification in pose I would not interefere with the active site if Bm-LS is singly modified. This is not the case for this simulation. In contrast to the previous assumption the ligand in pose I does move towards the active site and could interact with it. As for pose II modifier 1-1 shows a large motion out of its initial position which is accompanied by an increase of RMSD to 20 Å (red curve in Figure 12.16). This occurency is unlikely attributed to the concurrent covalent binding of the second ligand on Tyr446 as both are > 25 Å apart. Hence, one can conclude that 1-1 on Tyr247 has a larger flexibility as expected and is capable of motions into and also out of the active site despite of the pose.

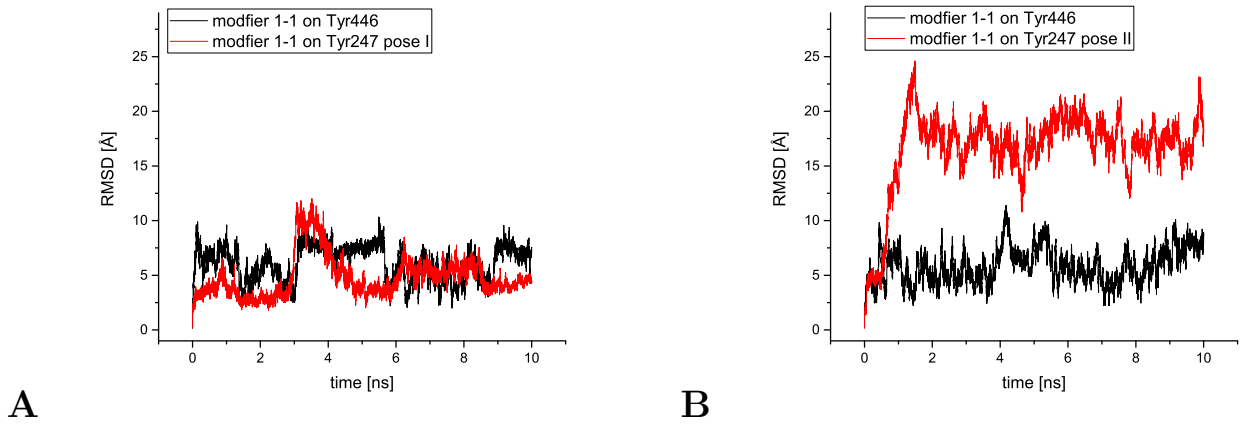


Figure 12.16: Course of RMSD values versus time plots of modifier 1-1 covalently bound to A – Tyr446 (black curve) and Tyr247 in pose I (red curve) or B – Tyr446 (black curve) and Tyr247 in pose II (red curve) in B. megaterium levansucrase during the 10 ns production run.

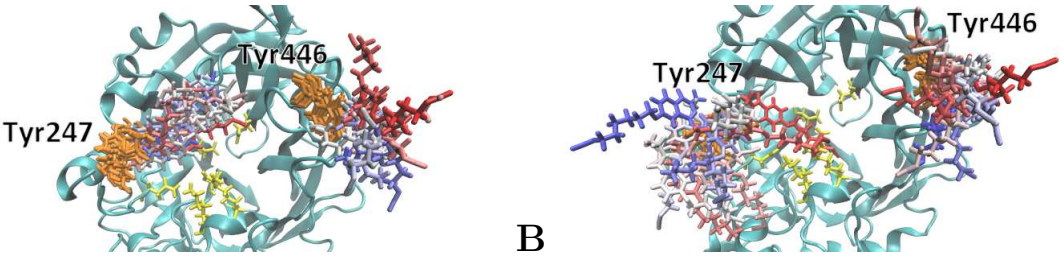


Figure 12.17: **A** - MD simulation of modifier 1-1 covalently bound to Tyr446 and Tyr247 in pose I. **B** - MD simulation of modifier 1-1 covalently bound to Tyr446 and Tyr247 in pose II. Every ns of the temporal progress presented from red-coloured (start of simulation) to blue-coloured modifier 1 (end of simulation). Modified tyrosine residues are depicted in orange licorice presentation, initial frame of active site residues in yellow licorice presentation for clarification.

The first conclusion which can be drawn by the simulations is that the covalent modification of Tyr247 with modifier 1-1 is likely to be responsible for the altered product spectrum of the carbohydrate polymer. Despite the initial assumption of the importance of the pose of ligand-linkage the simulations revealed for both cases a motion towards the active site due to the flexibility of the ligand which results in a movement opposite directed to the active site in one case. In terms of Tyr196, no modification of the product profile is to be expected if a second ligand is absent. In combination with 1-1 on Tyr247 a motion to the other ligand and thus a stabilization due to π-stacking occurs. It appears that 1-1 on Tyr196 has a supportive role. The covalent binding of modifier 1-1 on Tyr446 is unlikely as no movement towards the active site occurs, regardless of the presence or absence of a second ligand. Thus, it is to be expected that Tyr196 and Tyr247 are the modified residues which were detected by mass spectrometry. Experimental mutations of Bm-LS wild type to the Y196F/Y247F double variant incubated with modifier 2 confirmed the conclusions from the MD simulations as no change in product profile was observed.

12.4 Covalent modification of triple mutated Bm-LS

As the determination of the modified tyrosine residues by modifier 1 was successful, the scope of computations was extended. The interest of investigation lies on the controlled modification of the beginning of the elongation process in Bm-LS. Therefore, three novel tyrosine residues were inserted by triple mutation at possible key positions. At the same time, the modification of Tyr196 and Tyr247 was prevented by mutating them to phenylalanine. Thus, three different triple mutations of Bm-LS have been investigated: Y196F/Y247F/N126Y,

Y196F/Y247F/Q381Y, and Y196F/Y247F/F445Y (Figure 12.18). Additionally, modifier 2 was included into the simulations for the variant Y196F/Y247F/N126Y. The preparation of the MD simulation was conducted as for the wild type employing the same force fields for description of the protein and ligand. Moreover, the protocol of system minimization and equilibration as well as the production and duration are the same as mentioned in section 5.1.

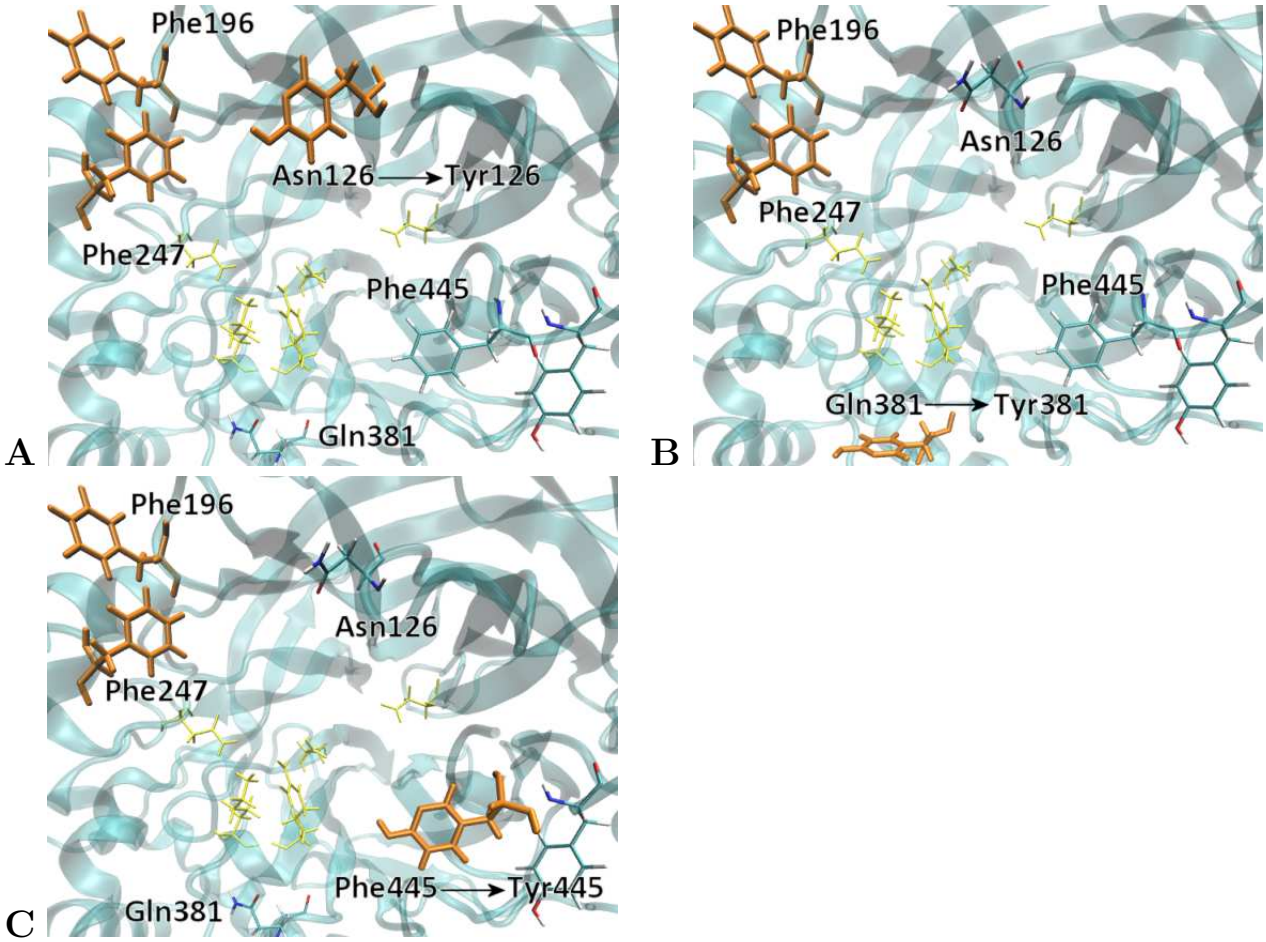


Figure 12.18: Triple mutants of B. megaterium levansucrase. **A** – Y196F, Y247F and N126Y triple mutation. **B** – Y196F, Y247F and Q381Y triple mutation. **C** – Y196F, Y247F and P445Y triple mutation. Modified residues are presented in orange licorice presentation, initial frame of active site residues in yellow licorice presentation serves for clarification. Modified amino acids are depicted in orange licorice presentation. Initial frame of active site residues is shown in yellow licorice presentation and serves for clarification.

Three different cases have been considered for the Y196F/Y247F/N126Y variant of Bm-LS. The linkage of modifier 1 on N126Y was created on the C-3 atom (pose I) of the benzyl-

ring as well as on the C-5 atom (pose II). For the latter both stereoisomers were considered. Initial frames of the MD simulations of all three cases are depicted in Figure 12.19.

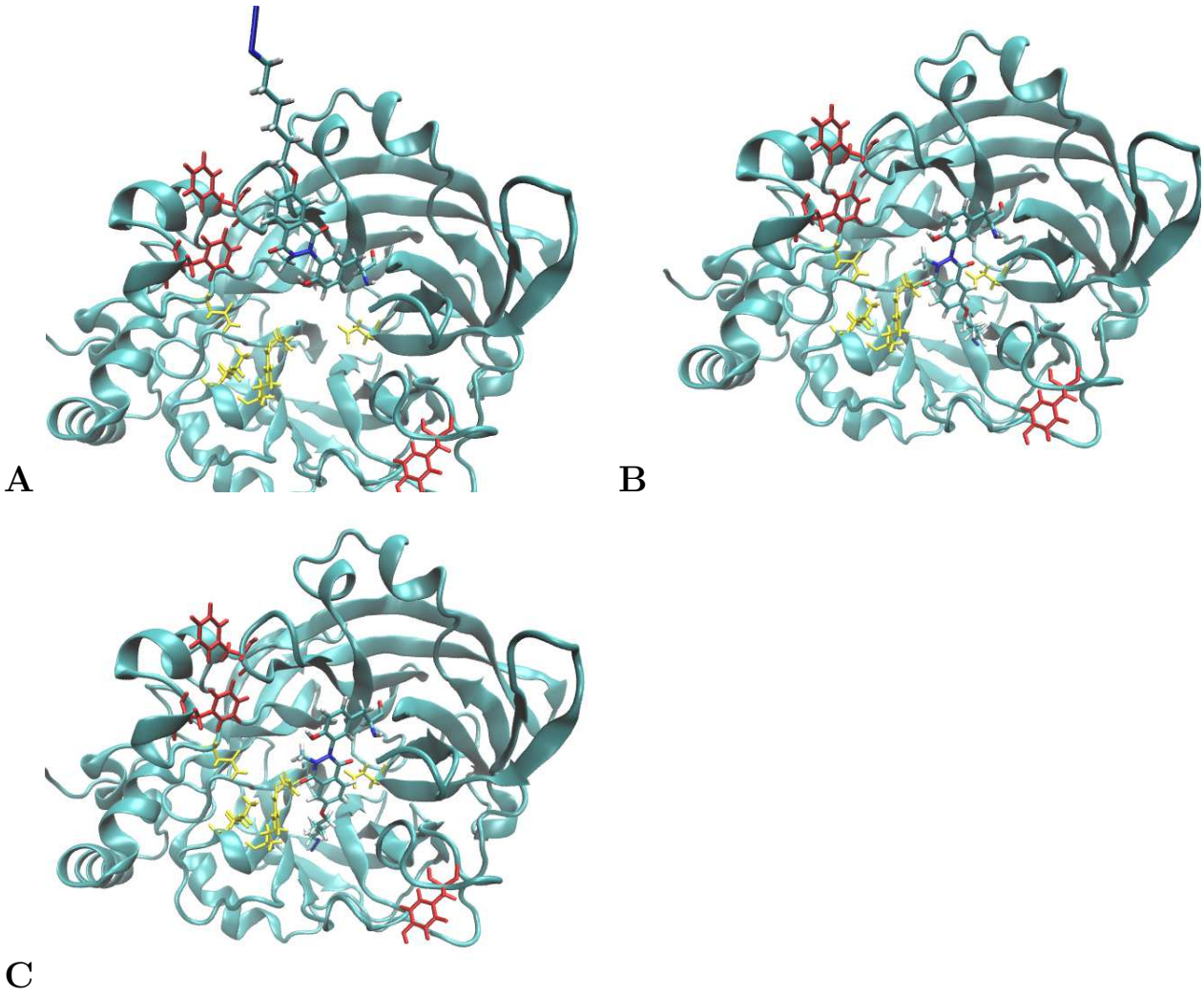


Figure 12.19: **A** – Initial frame of modifier 2-1 covalently bound to N126Y in pose I directed towards water shell. **B** - Initial frame of modifier 1-1 covalently bound to N126Y in pose II. **C** - Initial frame of modifier 1-2 covalently bound to N126Y in pose II. Active site residues are depicted in yellow licorice presentation, mutated tyrosine residues in red.

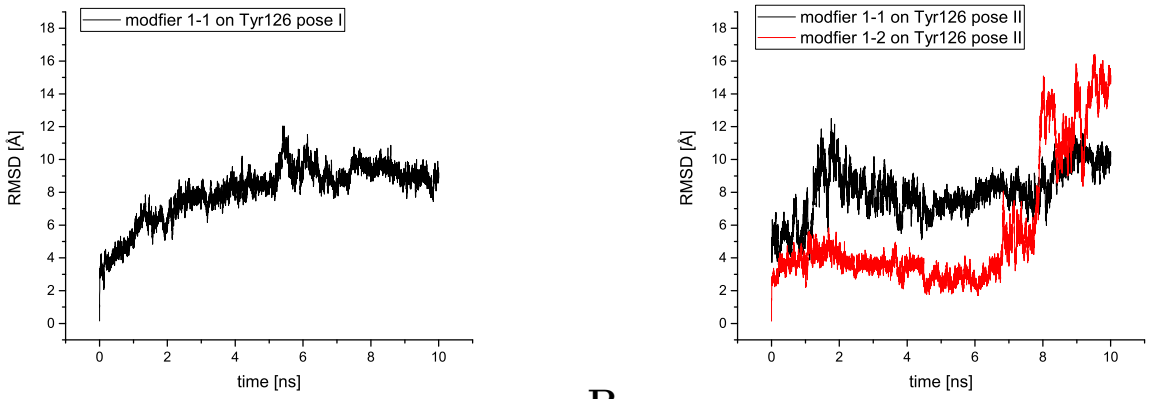


Figure 12.20: Course of RMSD values versus time plots of **A** – modifier 1-1 covalently bound to N126Y in pose I or of **B** – modifier 1-1 covalently bound to N126Y in pose II (black curve) or modifier 1-2 (red curve) covalently bound to NY126 in pose II during the 10 ns production run of the levansucrase Y196F/Y247F/N126Y variant.

The initial position of modifier 1-1 attached on N126Y in pose I is directed to the solvent. After 600 ps a motion towards the protein occurs leading to a position maintained over the course of simulation time (see Figure 12.21-A). This goes in hand with the increase in RMSD which mainly is caused due the first movement and afterwards fluctuations occur due to the flexibility of the 1-azido-5-methoxypentane group (Figure 12.20-A). The movement of modifier 1-1 is induced by the mutated tyrosine residue. As the N126Y mutation was manually inserted, the tyrosine residue was not adapted to the protein environment due to its enlarged size. Therefore, a movement of the amino acid was observable and in consequence of the covalent ligand. The attachment of 1-1 in pose II also shows a similar movement of the mutated residue leading to an increase of the ligand's RMSD in the first 2 ns. Afterwards, the modifier is directed towards the active site. Contrary to the Tyr247 modification of wild type Bm-LS, 1-1 is directed to the starting point of polymer elongation (Figure 12.21-B). In terms of stereoisomer 2 (in pose II) it is directed in the same manner for the first 8 ns. Unexpectedly one can observe a movement of 1-2 to the same position as 1-1 in pose I after 8 ns (Figure 12.21-C). This can clearly be comprehended by a glance at the RMSD plot which also shows a RMSD between 2 and 4 Å and a sudden increase to 14 Å.

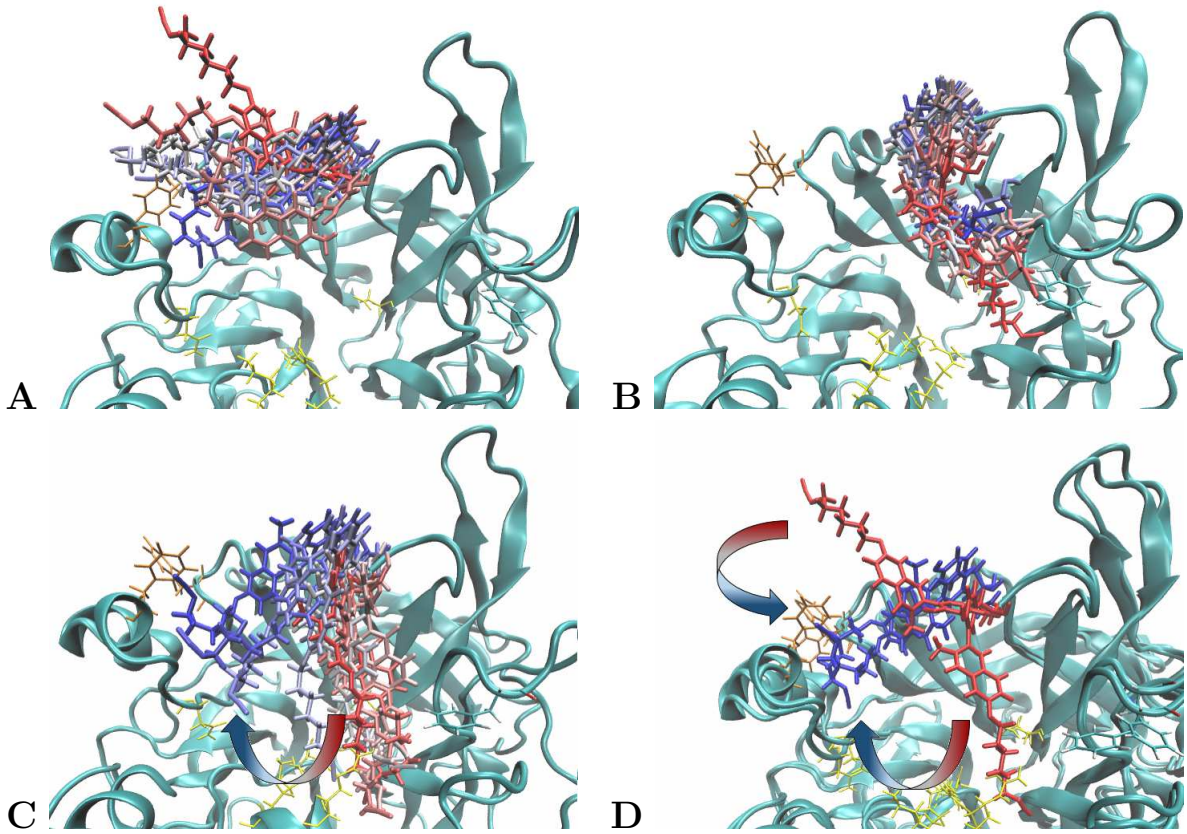


Figure 12.21: **A** - MD simulation of modifier 1-1 covalently bound to N126Y in pose I. **B** - MD simulation of modifier 1-1 covalently bound to N126Y in pose II **C** - MD simulation of modifier 1-2 covalently bound to N126Y in pose II. **D** - Initial and last frame of 1-1 (in pose I) and 1-2 (in pose II). Every ns of the temporal progress presented from red-coloured (start of simulation) to blue-coloured modifier 1 (end of simulation). Drawn arrows indicate movement of modifier during 10 ns production run. Mutated residues are depicted in orange licorice presentation, initial frame of active site residues in yellow licorice presentation for clarification.

In respect to the covalent modification of Bm-LS variant Y196F/Y247F/Q381Y only one pose of the modifier arises. As depicted in Figure 12.22 solely the linkage on the C-3 atom of the tyrosine residue was feasible due to steric hindrance of the protein environment. In case of this variant only stereoisomer 1 was considered. During the 10 ns simulation one cannot notice any indication of interaction with active site amino acids but on the contrary it remains stable on the surface after 2 ns (Figure 12.23).

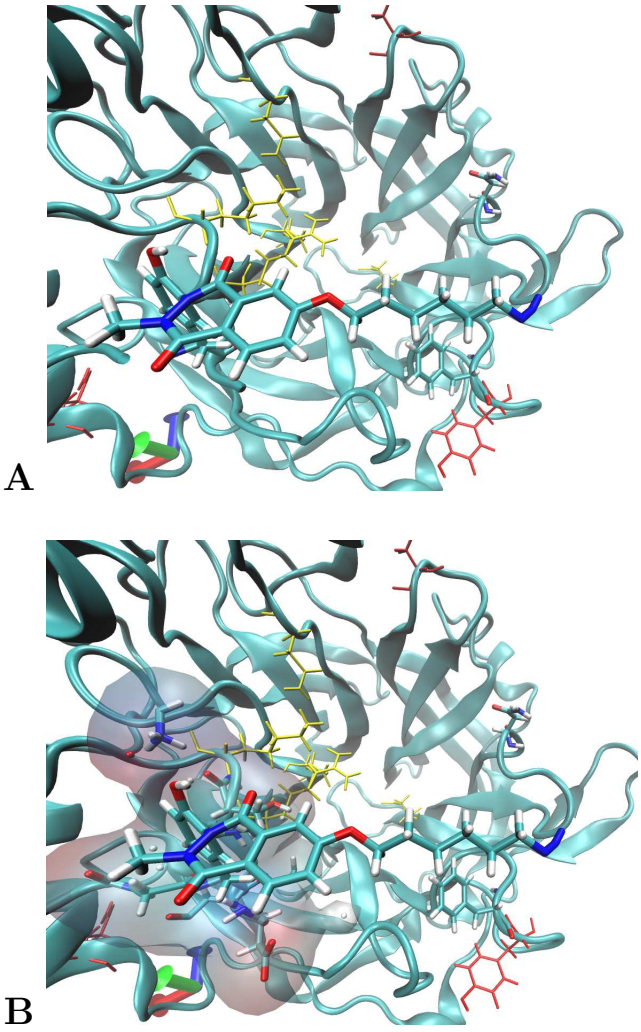


Figure 12.22: **A** – Initial frame of modifier 1-1 covalently bound to Q381Y directed towards the water shell. **B** – Quick surf presentation of amino acids close to Q381Y. Active site residues are depicted in yellow licorice presentation.

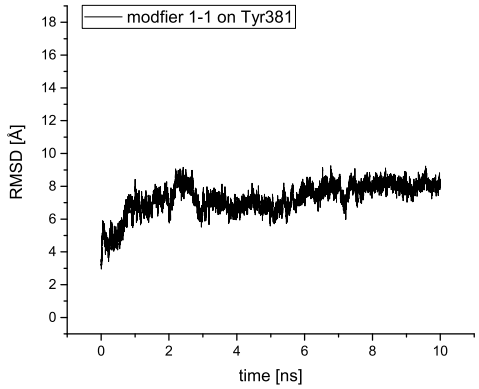


Figure 12.23: Course of RMSD values versus time plots of modifier 1-1 covalently bound to Q381Y during the 10 ns production run of the levansucrase Y196F/Y247F/Q381Y variant.

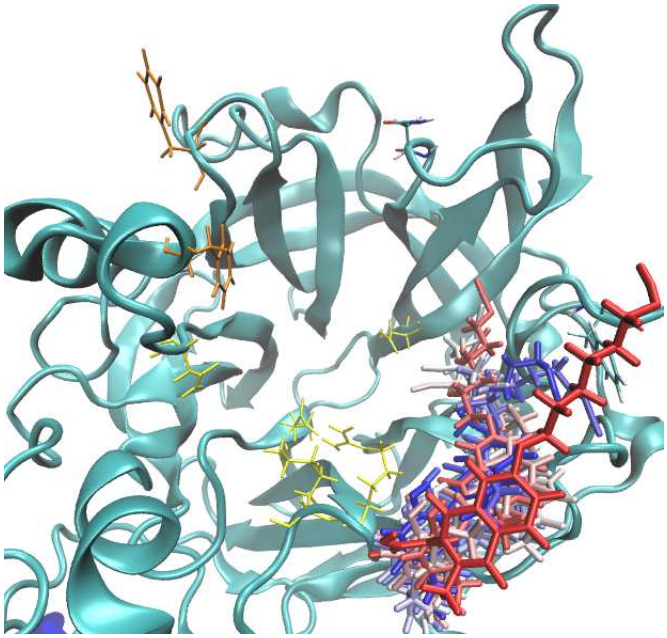


Figure 12.24: MD simulation of modifier 1-1 covalently bound to Q381Y. Every ns of the temporal progress presented from red-coloured (start of simulation) to blue-coloured modifier 1 (end of simulation). Mutated residues are depicted in orange licorice presentation, initial frame of active site residues in yellow licorice presentation for clarification.

The last investigated variant Y196F/Y247F/F445Y enabled modification on the C-3 as well as on the C-5 atom of F445Y leading to the poses I and II (Figure 12.25). As in the previous case, only stereoisomer 1 was considered in the simulations. For the MD simulation

of 1-1 in pose I, the obtained RMSD values are small in relation to pose II (Figure 12.26-A). In this case the ligand stays steady directed towards the starting point of polymerization. This is explainable as no movement of Tyr446 occurs and in consequence no motion in regard of the modifier. For pose II an increase of 1-1 up to 14 Å is observable after 500 ps pointing out to the solvent. In contrast to the other pose Tyr445 forms a T-stacked interaction with Tyr446 thus 'pulling' the ligand further away from the active site.

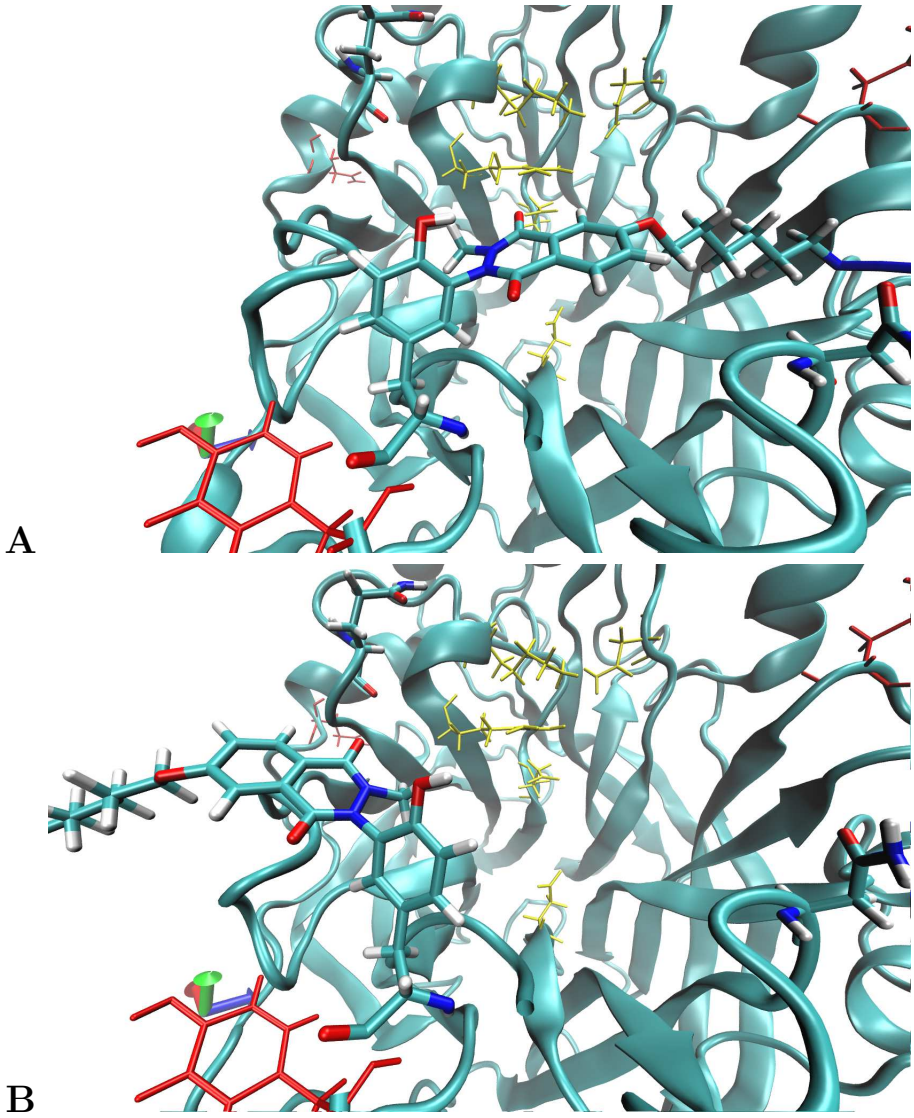


Figure 12.25: **A** – Initial frame of modifier 1-1 covalently bound to P445Y in pose I. **B** - Initial frame of modifier 1-1 covalently bound to P445Y in pose II. Active site residues are depicted in yellow licorice presentation, mutated tyrosine residues in red.

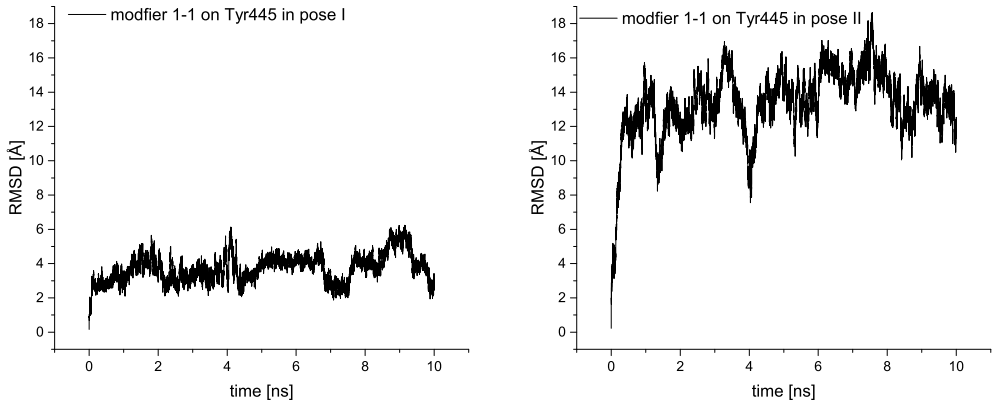


Figure 12.26: Course of RMSD values versus time plots of **A** – modifier 1-1 covalently bound to P445Y in pose I or of **B** – modifier 1-1 covalently bound to P445Y in pose II during the 10 ns production run of the levansucrase Y196F/Y247F/F445Y variant.

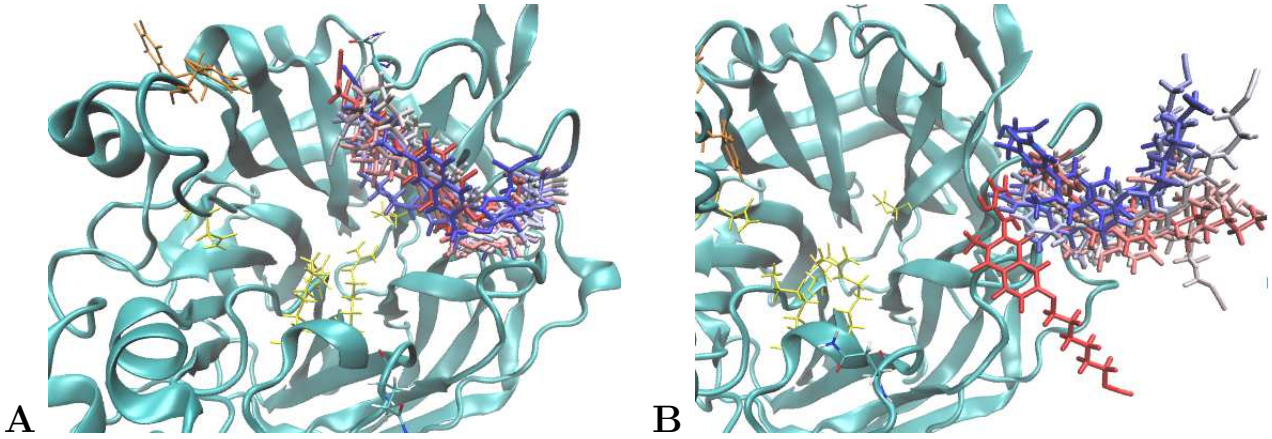


Figure 12.27: **A** - MD simulation of modifier 1-1 covalently bound to P445Y in pose I. **B** - MD simulation of modifier 1-1 covalently bound to P445Y in pose II. Every ns of the temporal progress presented from red-coloured (start of simulation) to blue-coloured modifier 2 (end of simulation). Mutated residues are depicted in orange licorice presentation, initial frame of active site residues in yellow licorice presentation for clarification.

By summing up the present results one might expect that the modification of the triple variant Y196F/Y247F/N126Y and also Y196F/Y247F/F445Y of Bm-LS might change the product profile. But experimental data lead to a slightly different conclusion. It was shown that only the triple mutation Y196F/Y247F/N126Y promotes the synthesis of small products. This leads to the conclusion that pose II is more likely in case of the Y196F/Y247F/F445Y variant whereas the simulations of Y196F/Y247F/N126Y are in line with the experiments.

To gain further insights on the mechanism itself sucrose and 6-kestose, output materials of the elongation have been inserted into the active site and simulated in combination with a covalent modified N126Y.

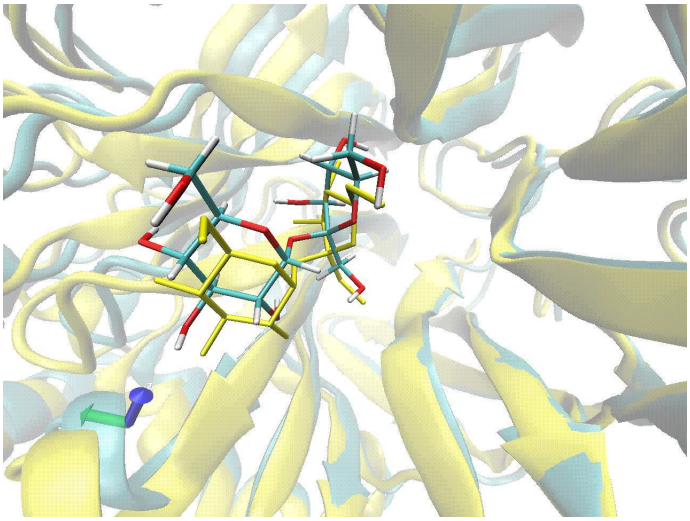


Figure 12.28: Alignment of pdb structures 1PT2 to 3OM2 after manual insertion of sucrose depicted in yellow licorice presentation (from 1PT2) or coloured by atom name (inserted in 3OM2).

By aligning the protein structure of Bm-LS in complex with sucrose (pdb code 1PT2) and of the triple variant it was possible to transfer the position of sucrose to the mutated protein-modifier complex of BM-LS (Figure 12.28). The initial simulation contained sucrose in absence of modifier 1 to estimate its interaction with the active site and afterwards, the ligand was included in all variations as considered before to investigate the interplay of these complexes. Firstly, the absence of modifier 1 shows an increase in RMSD values until 4 ns simulation time due to a slight rotation of the glucose unit. Afterwards the fluctuations in RMSD are of $\approx$ 0.5 Å thus one can expect a stable non-covalent complex. By inserting 1-1 in pose I the RMSD value as well as its fluctuations increases for the disaccharide. Interestingly the same movement of 1-1 towards the active site can be observed as in absence of sucrose. At the same time interactions between ligand and the disaccharide occur explaining its increased RMSD value. Modifier 1-1 in pose II also stays directed towards the starting point of polymerization as in the previous simulation. Especially in presence of sucrose a stable conformation is maintained which also restraints the movement of sucrose which is reflected in the small fluctuations and RMSD values in the plot (Figure 12.29). The last complex contained stereoisomer 2 in pose II in combination with sucrose. In this

connection a transport of sucrose out of the active site occurs evoked by the ligands movement which was also observed in absence of the sugar. Contrary to the previous simulation one can observe a return of the modifier to its initial position after "accompanying" sucrose towards the solvent.

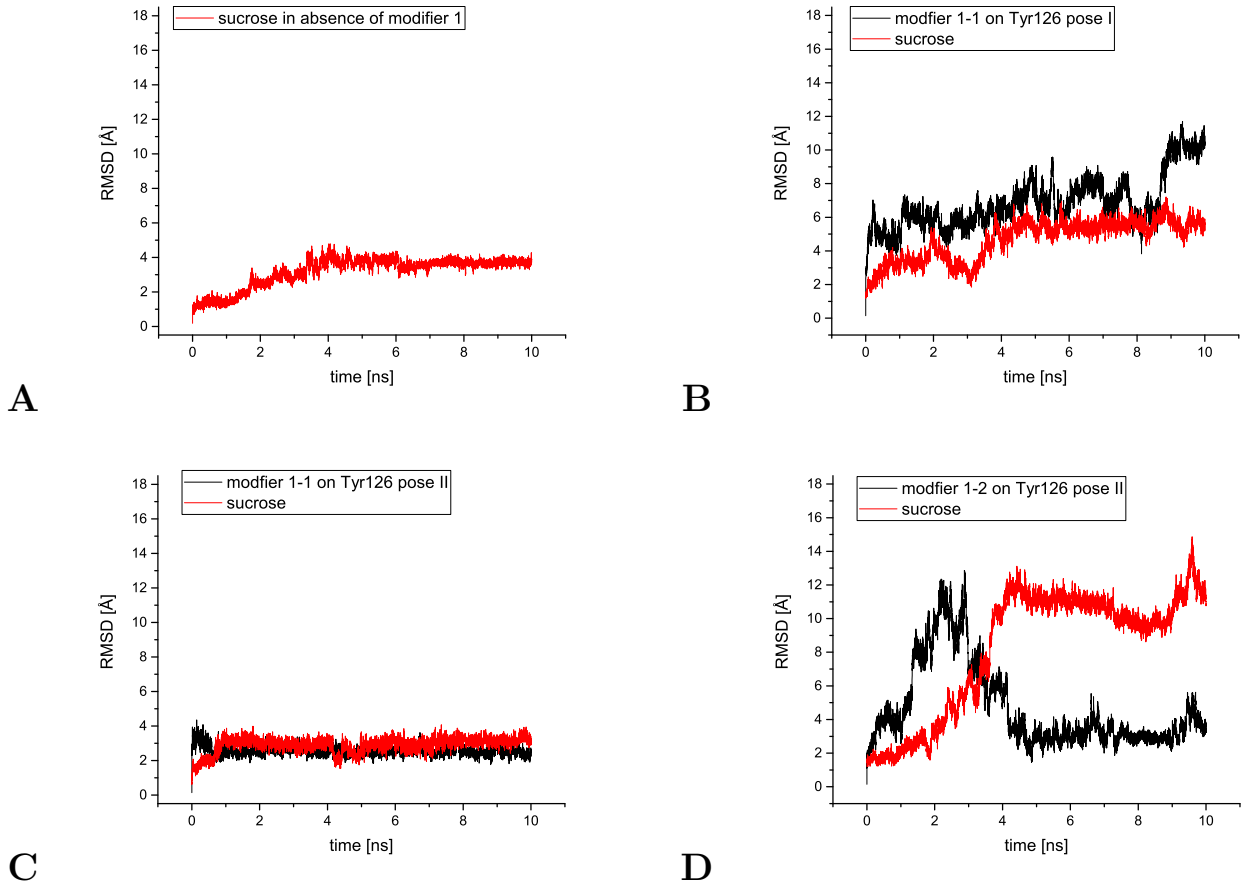


Figure 12.29: Course of RMSD values versus time plots of **A** – sucrose in absence of modifier 2, **B** – sucrose in presence of modifier 1-1 covalently bound to N126Y in pose I, **C** - sucrose in presence of modifier 1-1 covalently bound to N126Y in pose II or of **D** – modifier 1-2 covalently bound to N126Y in pose II during the 10 ns production run of the levansucrase Y196F/Y247F/N126Y variant.

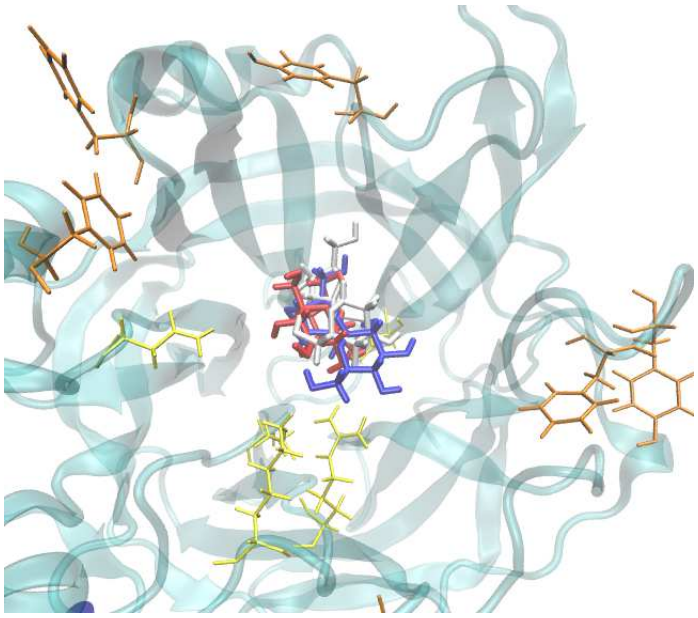


Figure 12.30: MD simulation of sucrose with initial structure (red), after 5 ns (white) or 10 ns (blue) depicted in licorice representation. Initial frame of active site residues are depicted in yellow licorice presentation for clarification.

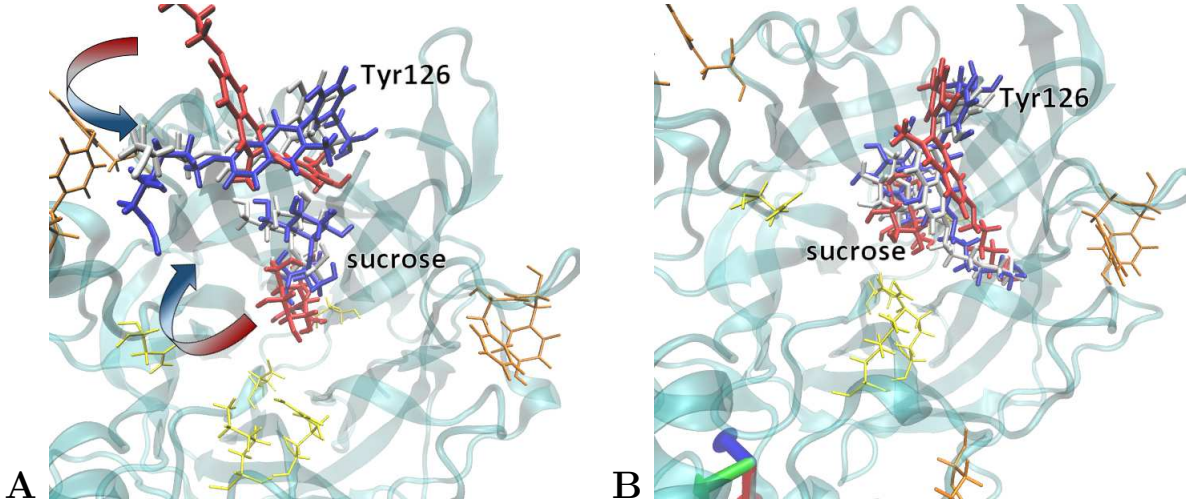


Figure 12.31: **A** - MD simulation of modifier 1-1 covalently bound to N126Y in pose I or **B** - MD simulation of modifier 1-1 covalently bound to N126Y in pose II in presence of sucrose depicted as initial structure (red), after 5 ns (white) or 10 ns (blue) in licorice representation. Drawn arrows indicate movement of modifier and sucrose during 10 ns production run. Initial frame of active site residues is depicted in yellow licorice presentation for clarification.

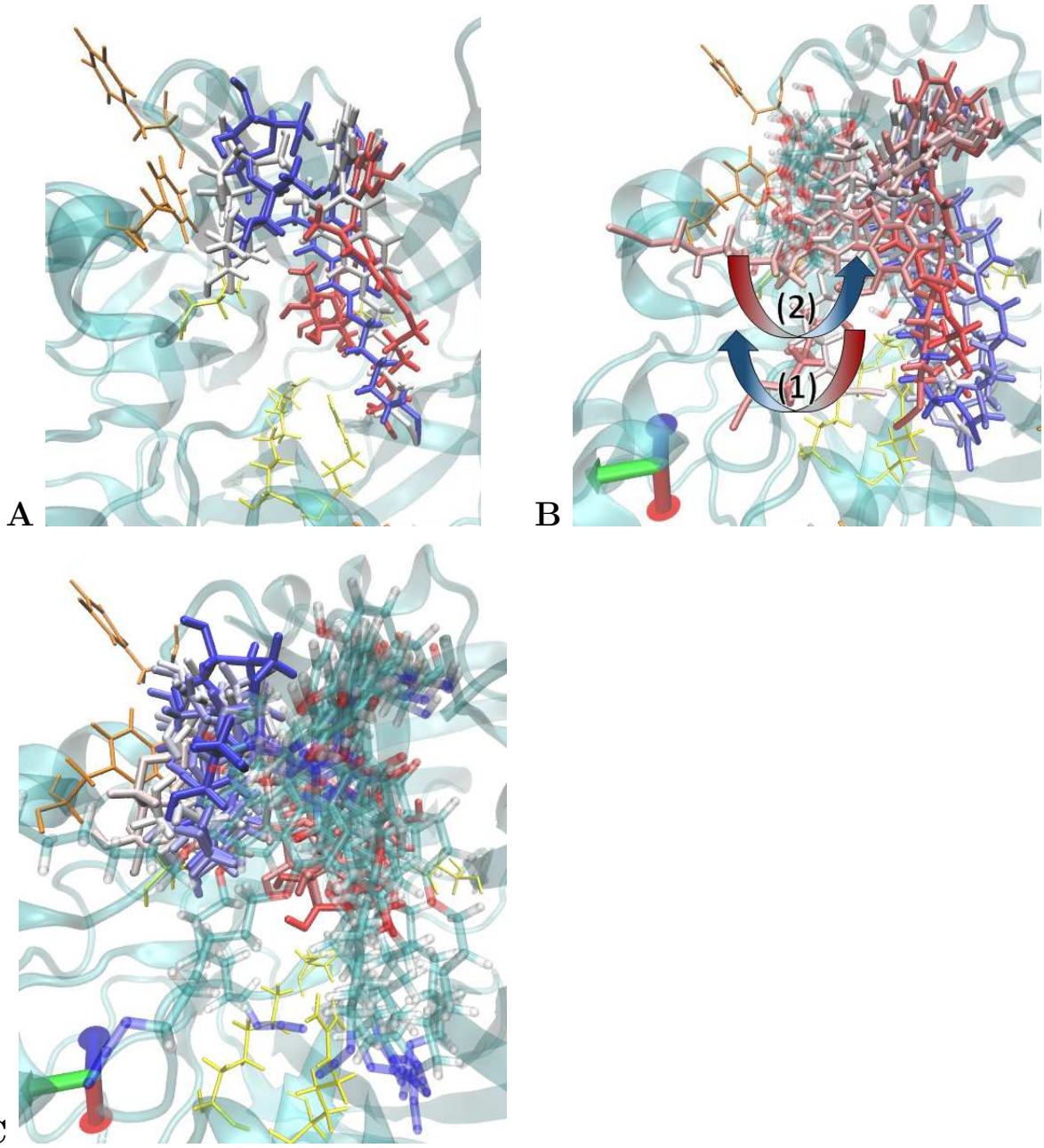
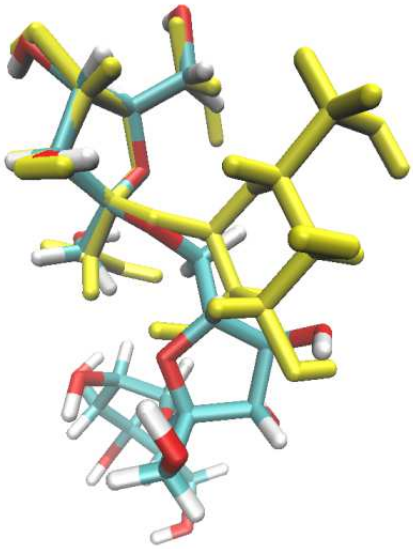


Figure 12.32: **A** - MD simulation of modifier 1-2 covalently bound to N126Y in pose II depicted as initial structure (red), after 5 ns (white) or 10 ns (blue) in licorice representation. **B** – For- and backward movement of modifier 1-1 during 10 ns production run indicated as arrow 1 and 2. **C** – Movement of sucrose out of active site during 10 ns production run is shown. Initial frame of active site residues is depicted in yellow licorice presentation for clarification.



A

Figure 12.33: Alignment of 6-kestose to sucrose in 3OM2. Sucrose is shown in yellow licorice presentation, 6-kestose in 3OM2 coloured by name.

The last simulations contained the consideration of the second smallest output product 6-kestose. The fructose unit of 6-kestose has been aligned with the same unit of sucrose in complex with the triple mutant of Bm-LS which was previously created (Figure 12.33). As the fructose unit of the oligosaccharide is the starting point of polymerization it was chosen for alignment instead of considering the glucose unit. This becomes clear if one thinks about the previous simulation of sucrose in absence of the modifier. The change in RMSD value mainly occurs by the rotation of the glucose and not fructose unit.

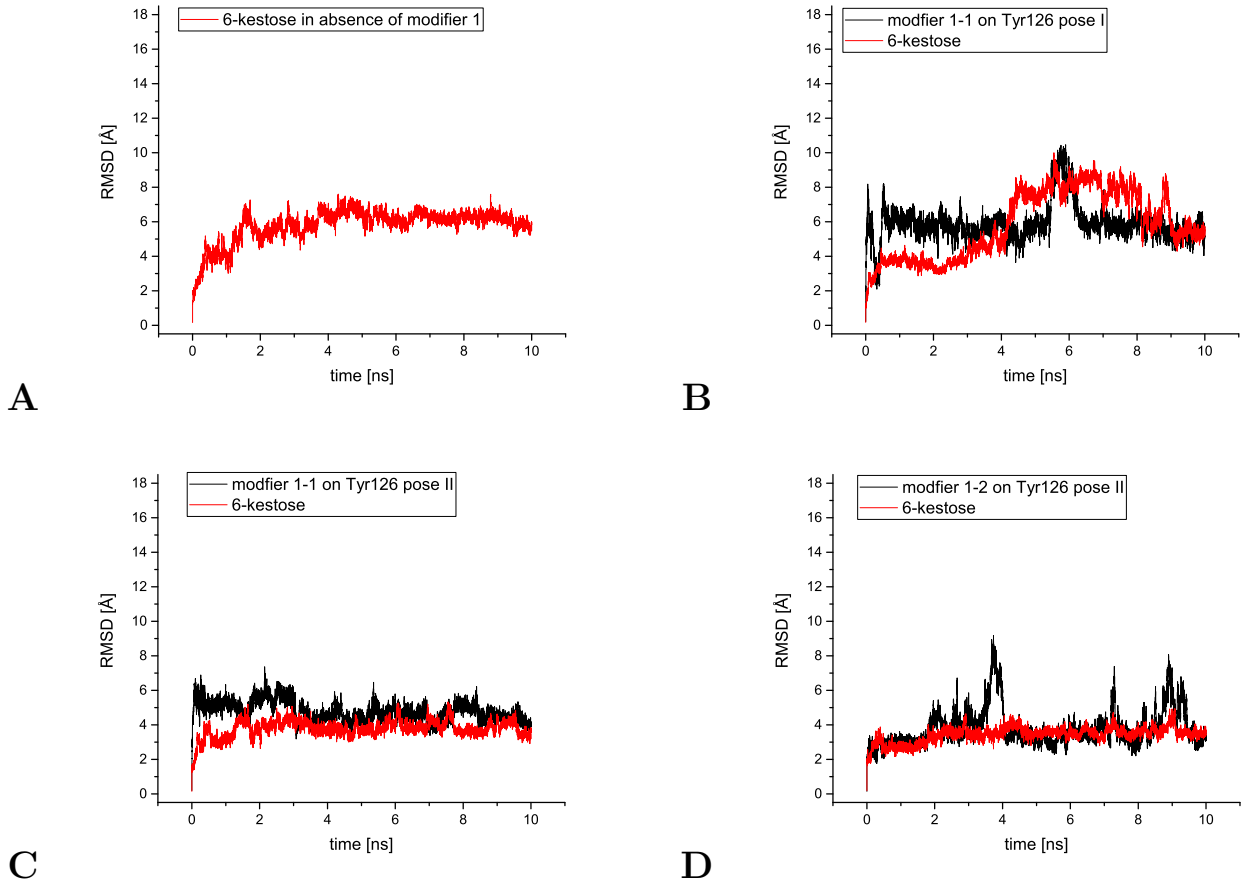


Figure 12.34: Course of RMSD values versus time plots of **A** – 6-kestose in absence of modifier 1, **B** – 6-kestose in presence of modifier 1-1 covalently bound to N126Y in pose I, **C** – 6-kestose in presence of modifier 1-1 covalently bound to N126Y in pose II or of **D** – modifier 1-2 covalently bound to N126Y in pose II during the 10 ns production run of the levansucrase Y196F/Y247F/N126Y variant.

If one solely focuses on the 6-kestose Bm-LS complex one can observe an equilibration in structure. The stabilization of the non-covalent complex occurs after 2 ns due interactions between the glucose unit of 6-kestose and the N126Y mutant. This leads to a raise of RMSD value to 6 Å implied in Figure 12.34-A. As in the case of sucrose only small fluctuations of the initial fructose unit are observable. The inclusion of 2-1 in pose I also reveals a similar motion as in the case of sucrose. A motion of 6-kestose and the covalent ligand occurs as an interaction between both establishes which leads to an increase of RMSD value of the trisaccharide after 4 ns. In contrast to the sugar the modifier is not strongly affected by this interaction as it only shows a single twist of the phthalazinedione scaffold at 6 ns accompanied by a slight raise in RMSD but a motion back to its equilibrated position. The simulations

of modifier 1 as stereoisomer 1 or 2 in pose II reveal comparable results. In both cases 6-kestose is tightly placed in its position as the ligand is directed towards it. This interaction additionally stabilizes the modifier itself as the course of RMSD reveals lower values compared to the absence of the trisaccharide. Fluctuations which occur for stereoisomer 2 are related to the motion of the 1-azido-5-methoxypentane group.

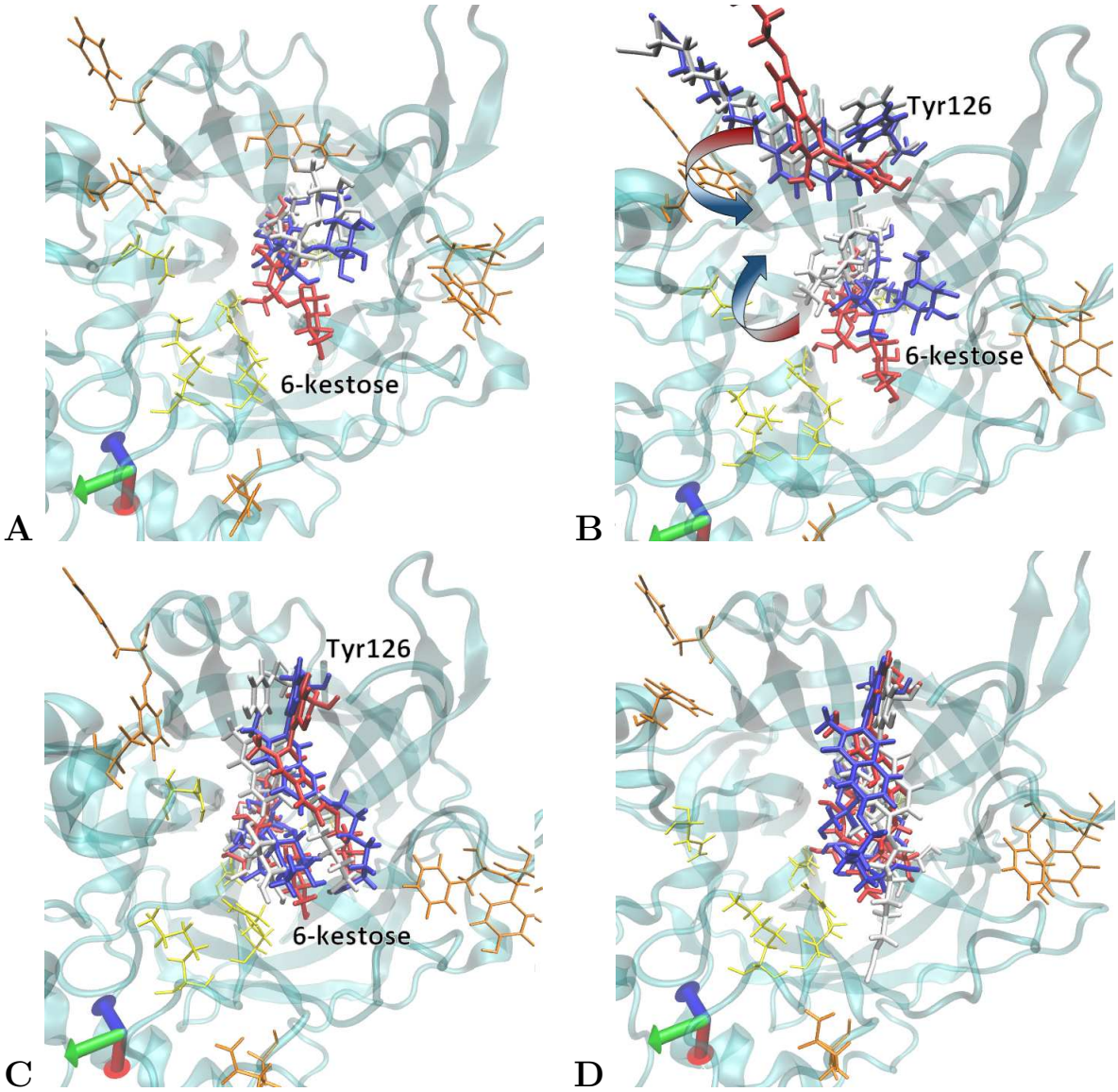
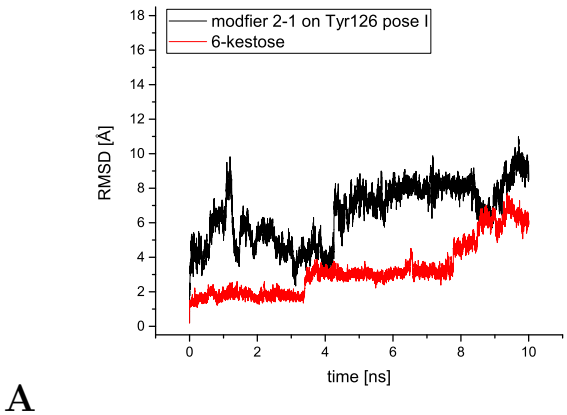


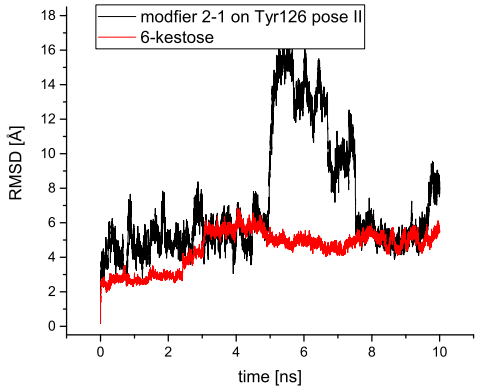
Figure 12.35: **A** – MD simulation of 6-kestose in absence of modifier 2. **B** - MD simulation of modifier 1-1 covalently bound to N126Y in pose I, **C** - of modifier 1-1 covalently bound to N126Y in pose II or **D** – of modifier 1-2 covalently bound to N126Y in pose II in presence of sucrose. Temporal progress of simulation time is depicted as initial structure (red), after 5 ns (white) or 10 ns (blue) in licorice representation. Drawn arrows indicate movement of modifier and 6-kestose during 10 ns production run. Initial frame of active site residues is depicted in yellow licorice presentation for clarification.

In a next step MD simulations have been conducted to investigate dynamics of the Y196F/Y247F/N126Y variant of Bm-LS in covalent complex with modifier 2 and in presence of 6-kestose. The initial positions of modifier 1 have been reused for modifier 2 to obtain

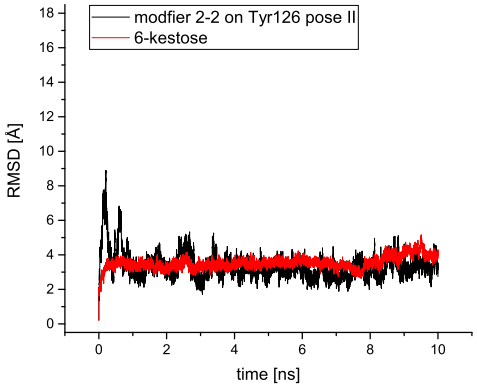
comparability. In case of modifier 2-1 in pose I a similar behaviour is observable as for modifier 1-1. Firstly, a movement of modifier 2-1 towards the active site can be seen and consequently a rise to 8 Å in RMSD value is observable (Figure 12.36-A). After 8 ns the trisaccharide draw closer to the modifier and forms interactions with 2-1 which leads to an increase in RMSD from 3 Å to 6 Å. In case of modifier 2-1 in pose II a similar movement of the modifier is observable as for modifier 1-2 in pose II in presence of sucrose. Firstly, a motion of 2-1 towards the same protein region occurs with a large increase in RMSD to a value of 16 Å which does not affect 6-kestose as no rise in RMSD is observable (Figure 12.36-B). Afterwards 2-1 returns to its initial position ater 7 ns and stays directed towards the active site.



A



B



C

Figure 12.36: Course of RMSD values versus time plots of **A** – 6-kestose in presence of modifier 2-1 covalently bound to N126Y in pose I, **B** – 6-kestose in presence of modifier 2-1 covalently bound to N126Y in pose II or of C – modifier 2-2 covalently bound to N126Y in pose II during the 10 ns production run of the levansucrase Y196F/Y247F/N126Y variant.

Lastly, dynamics of 2-2 in pose II reveal a stable conformation of the ligand as well as of 6-kestose. Firstly, a slight motion of 2-2 is observable for the first nanosecond but the modifier quickly relaxes and stays directed to the active site.

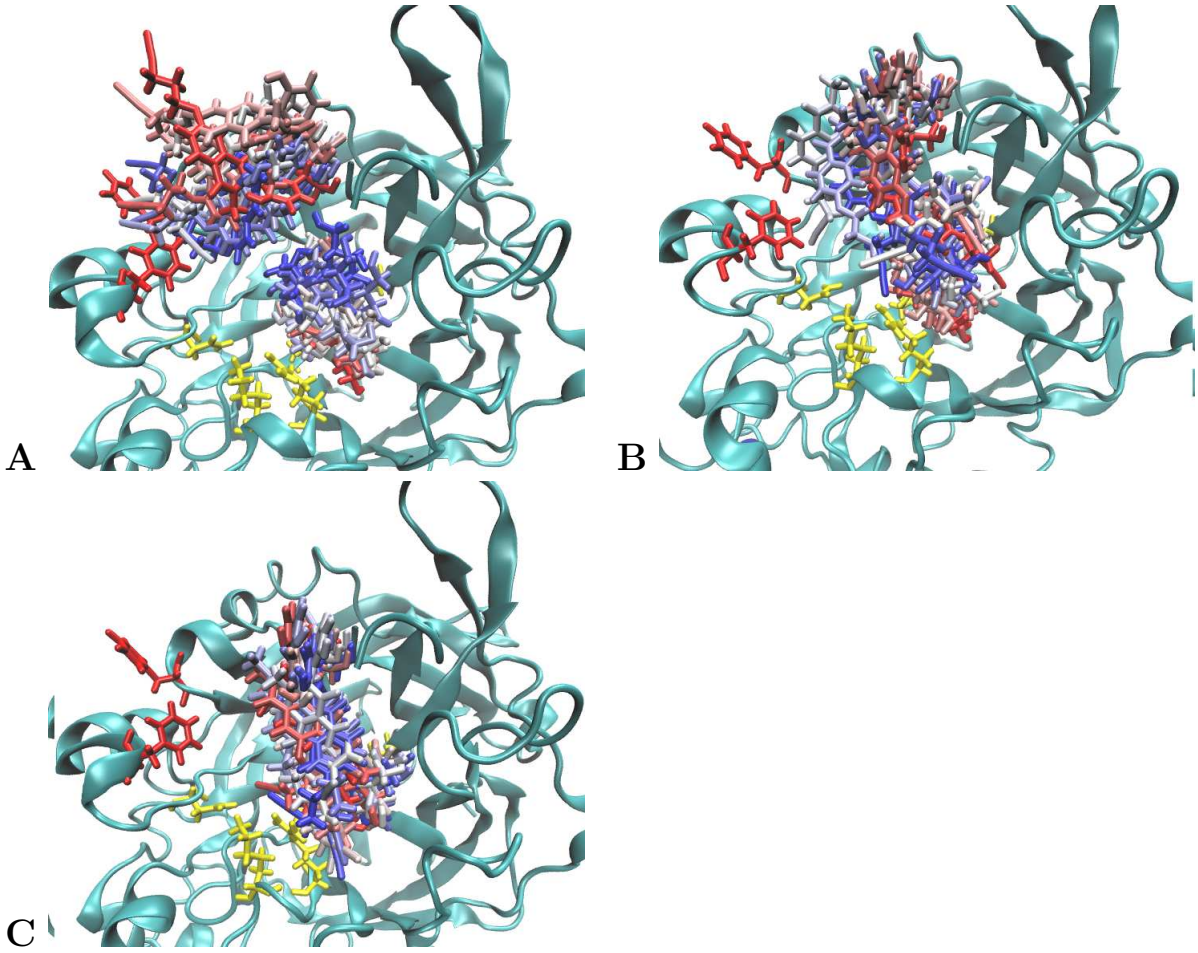


Figure 12.37: Snapshots of production run in Y196F/Y247F/N126Y variant of Bm-LS. **A** – Modifier 2-1 covalently bound to N126Y in pose I **B** - Modifier 2-1 covalently bound to N126Y in pose II **C** – Modifier 2-2 covalently bound to N126Y in pose II in presence of sucrose. Every ns of the temporal progress presented from red-coloured (start of simulation) to blue-coloured modifier 2 (end of simulation). Initial frame of active site residues is depicted in yellow licorice presentation for clarification.

12.5 Polymer elongation *via* modifier 2 and 2-AzGlc of different Bm-LS variants

Experimental data revealed a change in product profile to polysaccharides for three different cases: on the one hand for the Bm-LS variant Y247F and the covalent modification of Tyr196 with 2-AzGlc and on the other for the Y196F/Y247F/S125Y mutant with modifier 2 or 2-AzGlc covalently bound to S125Y. In order to reveal interactions of the protein-modifier complex MD simulations with duration of 10 ns have been conducted with the same protocol as mentioned in section 5.1.

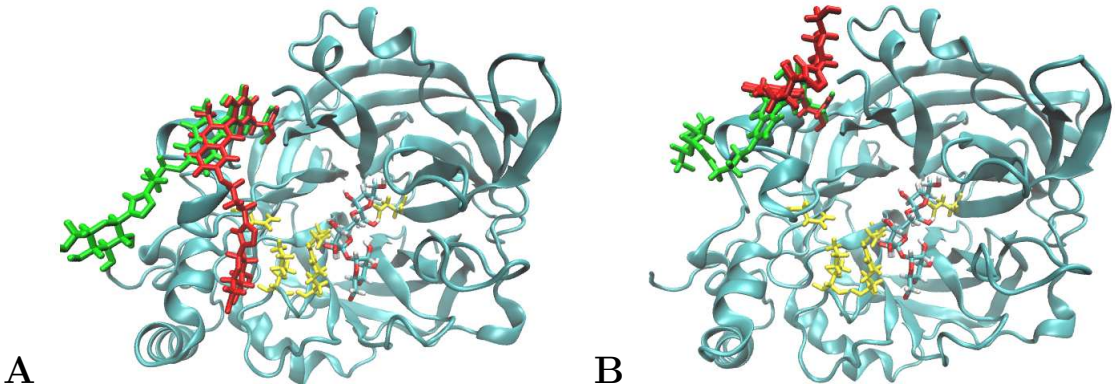
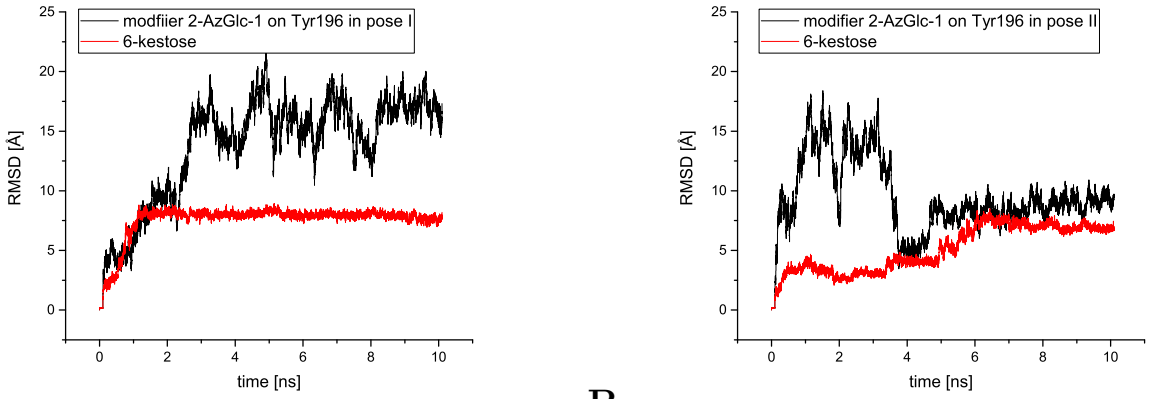


Figure 12.38: **A** – Initial position of modifier 2-AzGlc-1 in pose I (red) or in pose II (green). **B** – Position of modifier 2-AzGlc-1 in pose I (red) or in pose II (green) after 10 ns production run in levansucrase Y247F variant. Initial frame of active site residues is depicted in yellow licorice presentation for clarification.

Firstly 2-AzGlc has been attached as stereoisomer 1 in two different poses on Tyr196 in the Y247F variant as depicted in Figure 12.38-A. Afterwards a 10 ns production run has been performed in presence of 6-kestose. For 2-AzGlc stereoisomer 1 an increase to 15 Å in RMSD value is observable after 2 ns which goes along with a decrease in distance to Lys194 backbone oxygen (Figure 12.39-A and Figure 12.41-B). This interaction was maintained for 2 ns and was re-established after 8 ns. Additionally, intramolecular interactions of 2-AzGlc are observable after 8 ns leading to a stabilization of RMSD values. For 6-kestose a motion in the first ns is observable leading to a similar position as in absence of any modifer. Still, no interactions between the trisaccharide and the modifier was observable as well as a motion of the modifier towards the active site.



A **B**

Figure 12.39: Course of RMSD values versus time plots of **A** – 6-kestose in presence of modifier 2-AzGlc-1 covalently bound to Tyr196 in pose I, **B** – 6-kestose in presence of modifier 2-AzGlc-1 covalently bound to Tyr196 in pose II during the 10 ns production run of the Bm-LS variant Y247F.

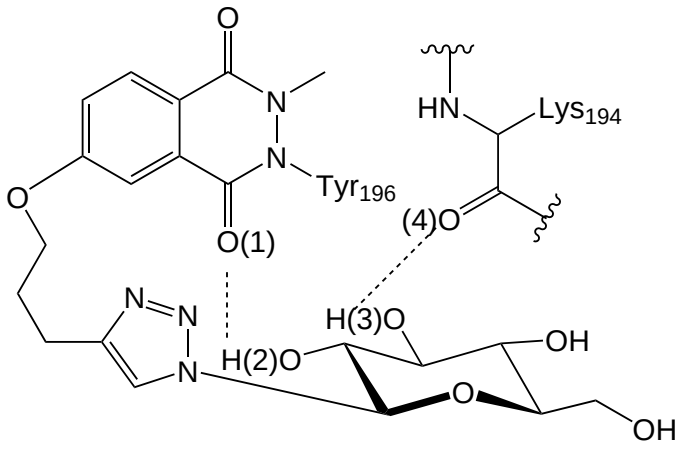
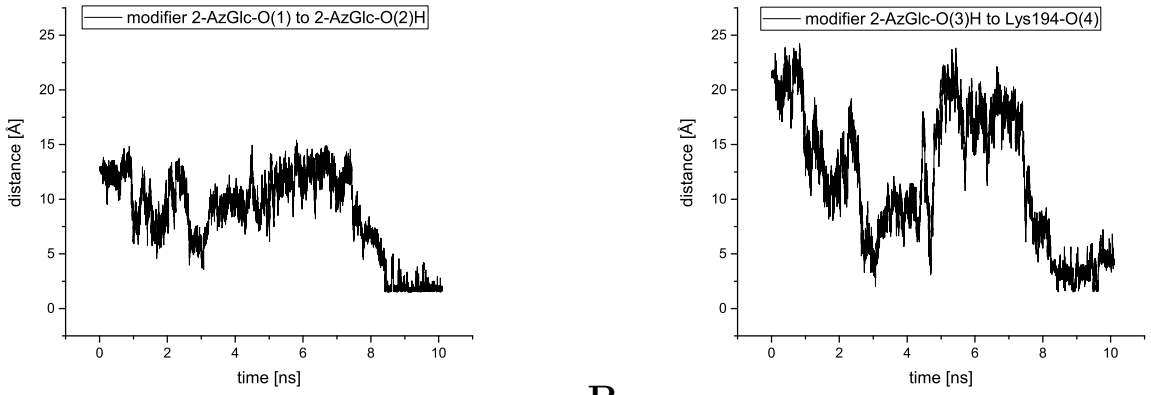


Figure 12.40: Interaction of covalent protein-inhibitor complex of 2-AzGlc-1 in pose I and Bm-LS variant Y247F depicted with dashed lines. Location of amino acids and 2-AzGlc serves for clarification and does not reflect actual position in MD simulations.



A **B**

Figure 12.41: The variation in distance as a function of simulation time for modifier 2-AzGlc-1 (pose I) in Bm-LS variant Y247F corresponding to Figure 12.40 during a 10 ns production run. **A** – Distance between 2-AzGlc-1-O(1) to 2-AzGlc-1-O(2)H in pose I. **B** – Distance between Lys194-O(4) to 2-AzGlc-1-O(3) in pose I.

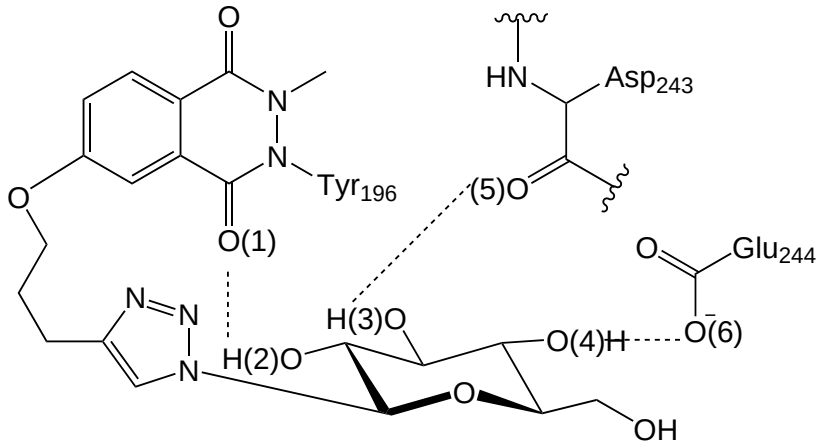
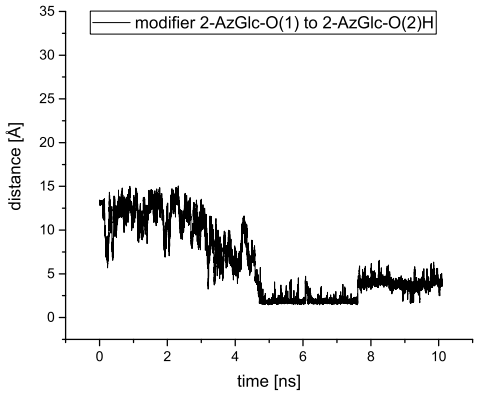


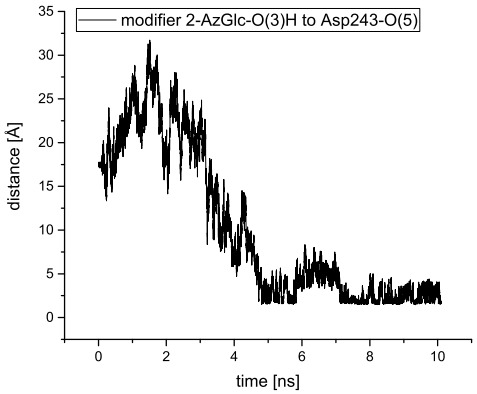
Figure 12.42: Interaction of covalent protein-inhibitor complex of 2-AzGlc-1 in pose II and Bm-LS variant Y247F depicted with dashed lines. Location of amino acids and 2-AzGlc serves for clarification and does not reflect actual position in MD simulations.

For 2-AzGlc in pose II similar observations have been made. In terms of RMSD values an increase to 15 Å can be seen after 1 ns which drops to 10 Å after 5 ns. This course is correlated with arising interactions of the modifier and Asp243 backbone oxygen and Glu244 carboxylate group. As before, intramolecular hydrogen bonds are established for 2-1AzGlc. All these points lead to the stabilization in RMSD value thus less fluctuations after 5 ns. 6-kestose shows a steady increase in RMSD value to 10 Å and stabilizes after 6 ns. As in

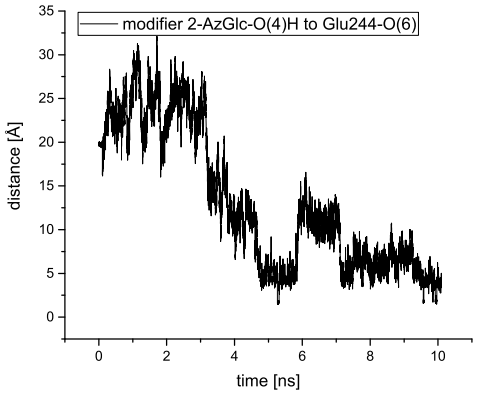
the previous case, no interactions between the trisaccharide and 2-AzGlc was seen as well as motions of the modifier into direction of the active site.



A



B



C

Figure 12.43: The variation in distance as a function of simulation time for modifier 2-AzGlc-1 (pose II) in Bm-LS variant Y247F corresponding to Figure 12.42 during a 10 ns production run. **A** – Distance between 2-AzGlc-1-O(1) to 2-AzGlc-1-O(2)H in pose II. **B** – Distance between Asp243-O(5) to 2-AzGlc-1-O(3) in pose II. **C** - Distance between Glu244-O(6) to 2-AzGlc-1-O(3) in pose II.

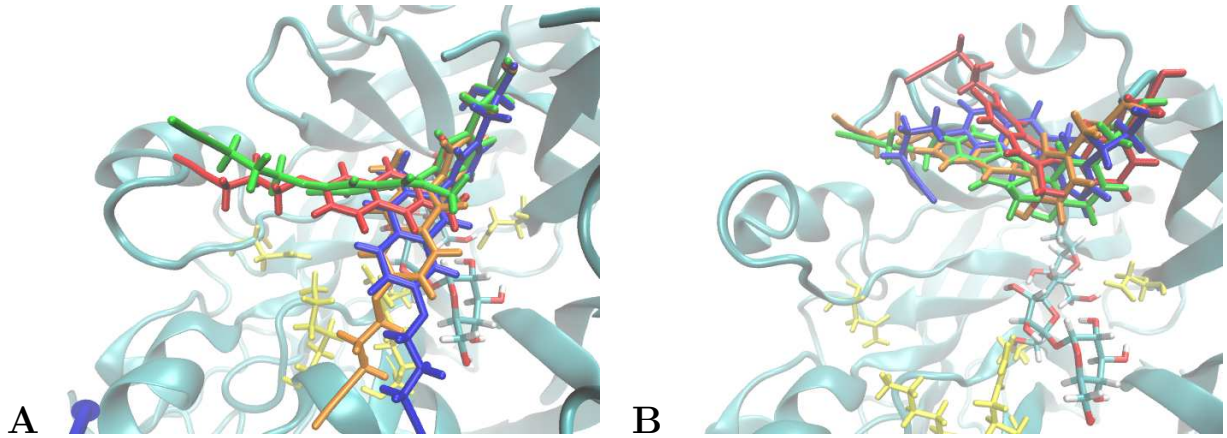


Figure 12.44: **A** – Initial position of modifier 2-1 in pose I (red) or in pose II (blue) or of modifier 2-2 in pose I (green) or in pose II (orange). **B** – Position of modifier 2-1 in pose I (red) or in pose II (blue) or of modifier 2-2 in pose I (green) or in pose II (orange) after 10 ns production run. Initial frame of active site residues is depicted in yellow licorice presentation for clarification.

As experimental findings revealed the production of polysaccharides for the Y196F/Y247F /S125Y variant of Bm-LS in complex with modifier 2 and 2-AzGlc further simulations have been conducted. For modifier 2 both stereoisomers and two different poses have been included in the computations (Figure 12.44-A). If one takes a look at the RMSD values, a higher increase is observable for the attachement to S125Y in pose II (Figure 12.45). This becomes clear as the modifier shows a larger motion during the simulation compared to pose I. Interestingly, all four variations of attaching modifier 2 lead to a similar position after 10 ns (Figure 12.44-B). Differences can be seen in case of the stereoisomers: lower fluctuations in RMSD are observable for stereoisomer 2 which is also correlated with the formation of stable direct hydrogen bonds with Glu170's amide group (Figure 12.47).

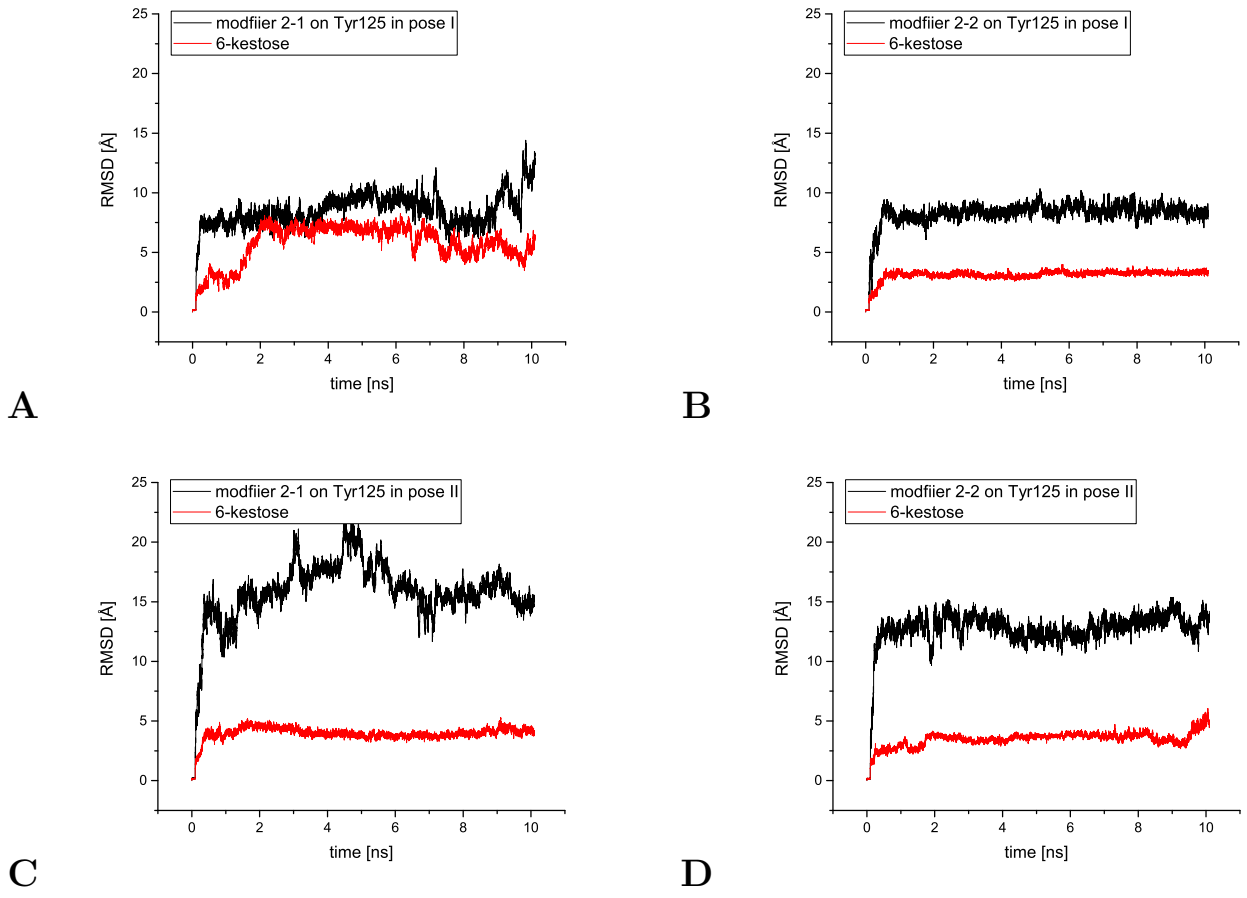


Figure 12.45: Course of RMSD values versus time plots of **A** – 6-kestose in presence of modifier 2-1 covalently bound to Tyr126 in pose I, **B** – 6-kestose in presence of modifier 2-2 covalently bound to Tyr126 in pose I, **C** - 6-kestose in presence of modifier 2-1 covalently bound to Tyr126 in pose II, **D** - 6-kestose in presence of modifier 2-2 covalently bound to Tyr126 in pose II during the 10 ns production run of Bm-LS variant Y196F/Y247F/S125Y.

In case of 6-kestose interactions between modifier 2-1 in pose I and the trisaccharide are established leading to the increase in RMSD value. This was not seen for the three other cases. Additionally, modifier 2 is not directed towards the active site in the course of simulation time as modifier 1.

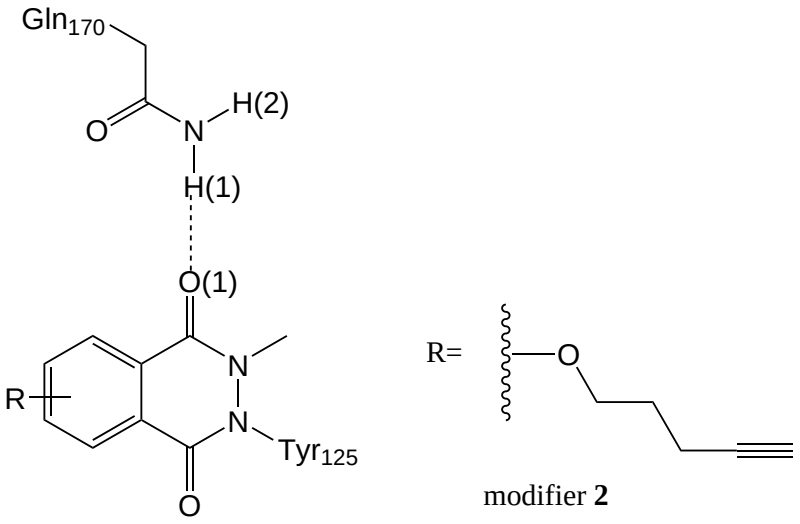


Figure 12.46: Interaction of covalent protein-inhibitor complex of 2-1 and Bm-LS variant Y196F/Y247F/S125Y depicted with dashed lines. Location of amino acids and 2 serves for clarification and does not reflect actual position in MD simulations.

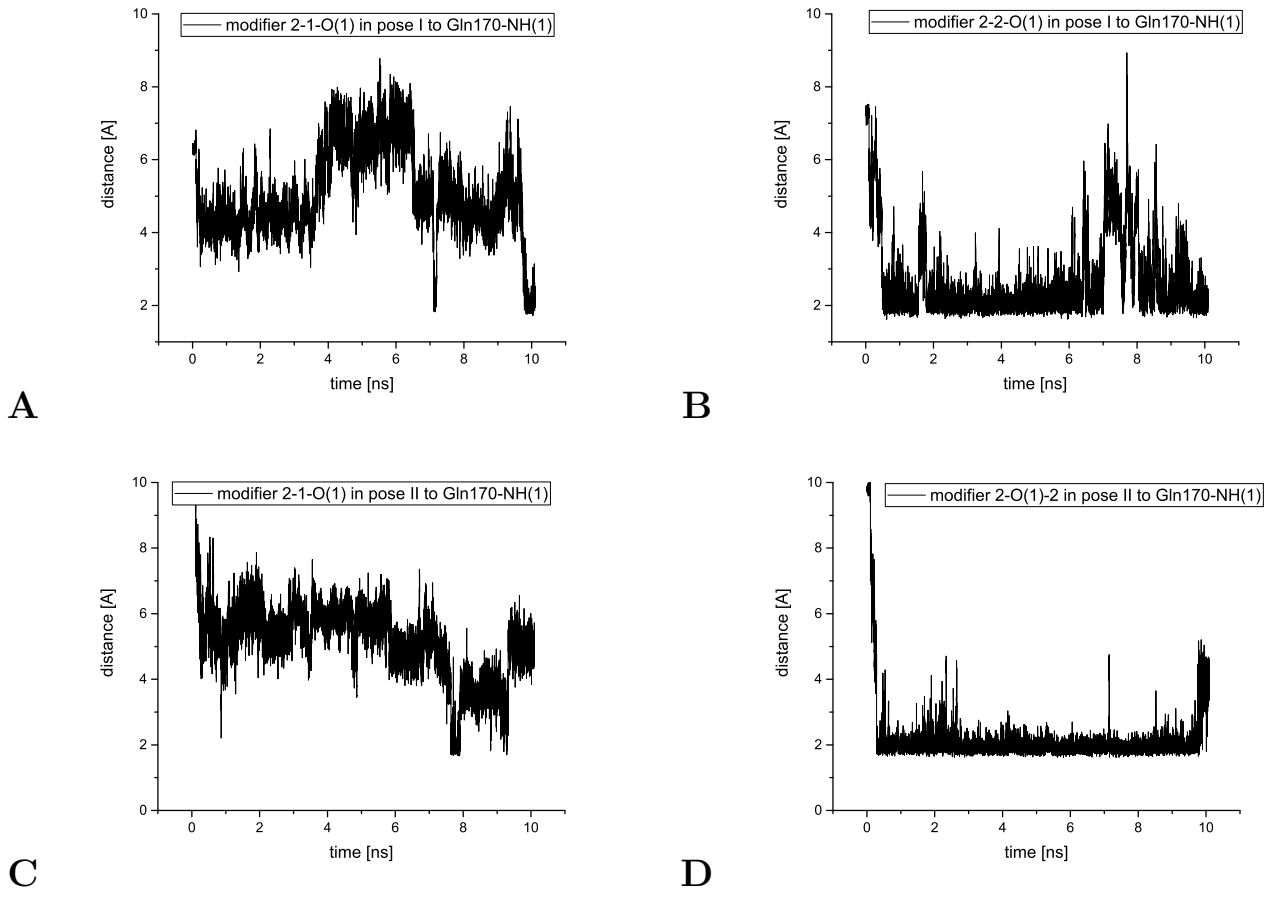


Figure 12.47: The variation in distance as a function of simulation time for modifier 2-AzGlc-1 (pose I) in Bm-LS Y196F/Y247F/S125Y variant corresponding to Figure 12.5 during a 10 ns production run. Distance between **A** – 2-AzGlc-1-O(1) in pose I, **B** – 2-AzGlc-2-O(1) in pose I, **C** – 2-AzGlc-1-O(1) in pose II or **D** – 2-AzGlc-2-O(1) in pose II and Gln170-NH(1).

Lastly, dynamics of 2-AzGlc in the Y196F/Y247F/S125Y variant of Bm-LS have been investigated. Due to steric hindrance of the protein environment it was not possible to attach 2-AzGlc-2 in pose II to Tyr125 thus only three different simulations have been performed. As modifier 2 the extended 2-AzGlc show similar motions towards the same region of Bm-LS independent from the starting point (Figure 12.48-B). Consequently, the rise in RMSD for pose II is much larger compared to simulations of 2-AzGlc in pose I as it has to move farther away from its initial position. In contrast to modifier 2 no interactions to 6-kestose can be observed resulting in similar trends in RMSD for the trisaccharide. For 2-AzGlc-1 in pose I interactions between modifier and backbone oxygen of Lys194 are established between 2 and 6 ns. Afterwards hydrogen bonds between Asp251 are formed (Figure 12.51). For stereoisomer

2 in the same pose interactions between Asp251 are established and maintained in the course of simulation time (Figure 12.52). Contrary to 2-AzGlc-1 in pose I simulations in pose II reveal at first an interaction with Asn252 between 2 and 4 ns (Figure 12.53-A). The next observed formation of hydrogen bonds occurred at 8 ns between Lys194 and the modifier (Figure 12.53-B). Interestingly, no intramolecular interactions of modifier 2 or 2-AzGlc have been established in the simulations in the Y196F/Y247F/S125Y variant of Bm-LS.

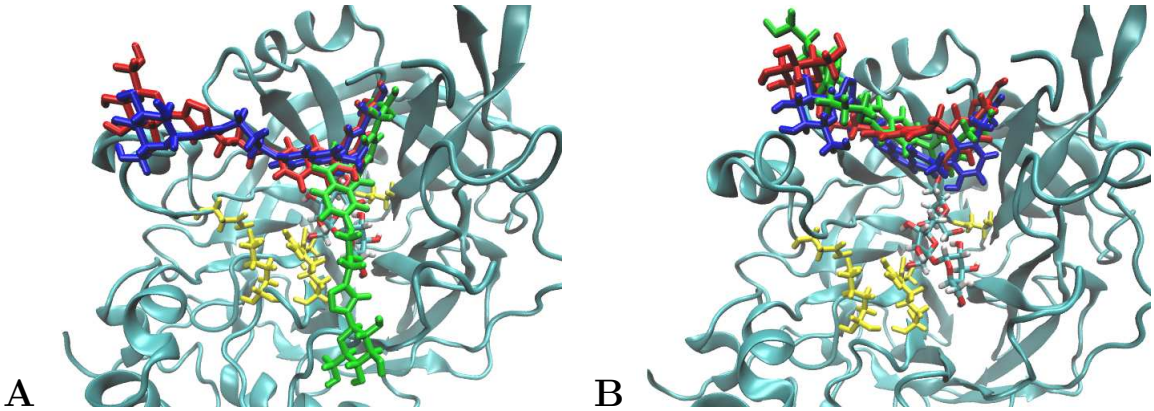


Figure 12.48: **A** – Initial position of modifier 2-AzGlc-1 in pose I (red) or in pose II (blue) or of modifier 2-AzGlc-2 in pose I (green). **B** – Position of modifier 2-AzGlc-1 in pose I (red) or in pose II (blue) or of modifier 2-AzGlc-2 in pose I (green) after 10 ns production run. Initial frame of active site residues is depicted in yellow licorice presentation for clarification.

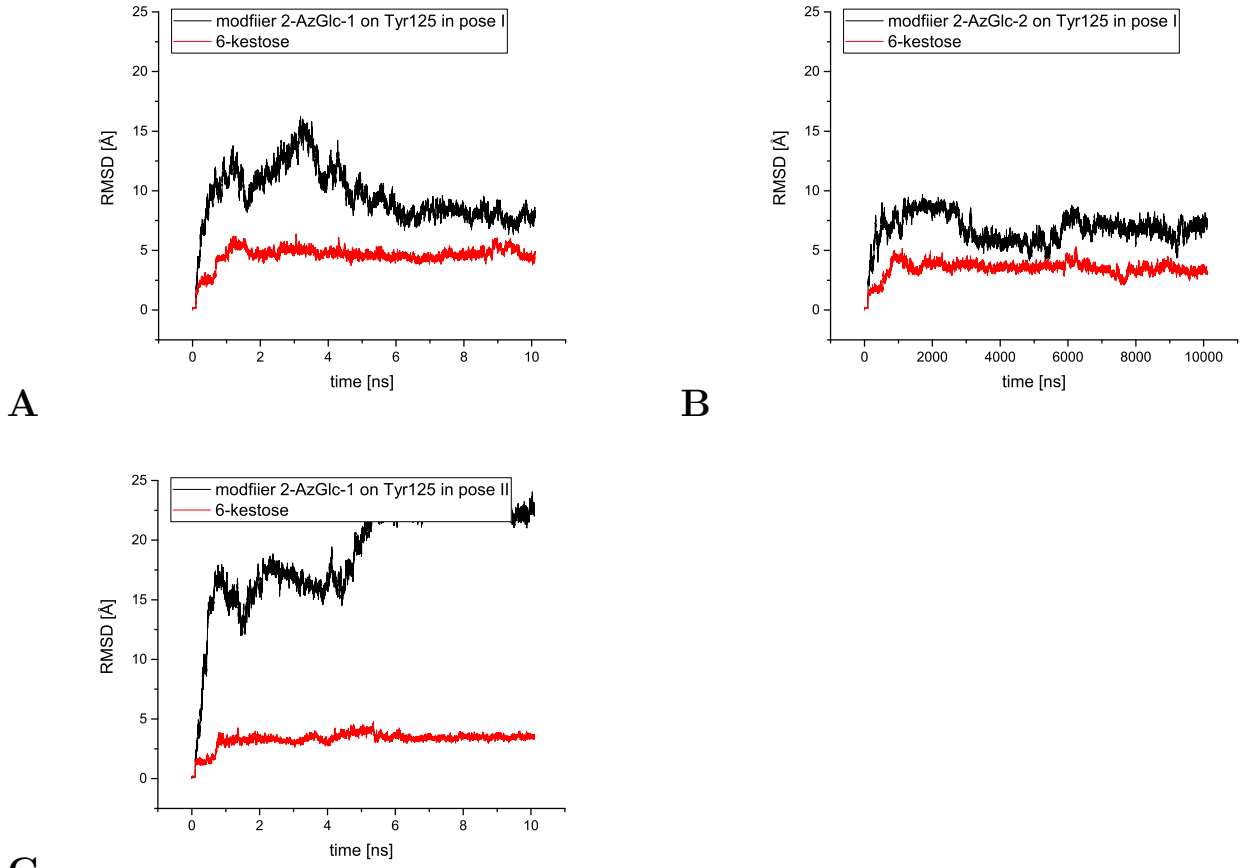


Figure 12.49: Course of RMSD values versus time plots of **A** – 6-kestose in presence of modifier 2-AzGlc-1 covalently bound to Tyr126 in pose I, **B** – 6-kestose in presence of modifier 2-AzGlc-2 covalently bound to Tyr126 in pose I, **C** - 6-kestose in presence of modifier 2-AzGlc-1 covalently bound to Tyr126 in pose II during the 10 ns production run of Bm-LS variant Y196F/Y247F/S125Y.

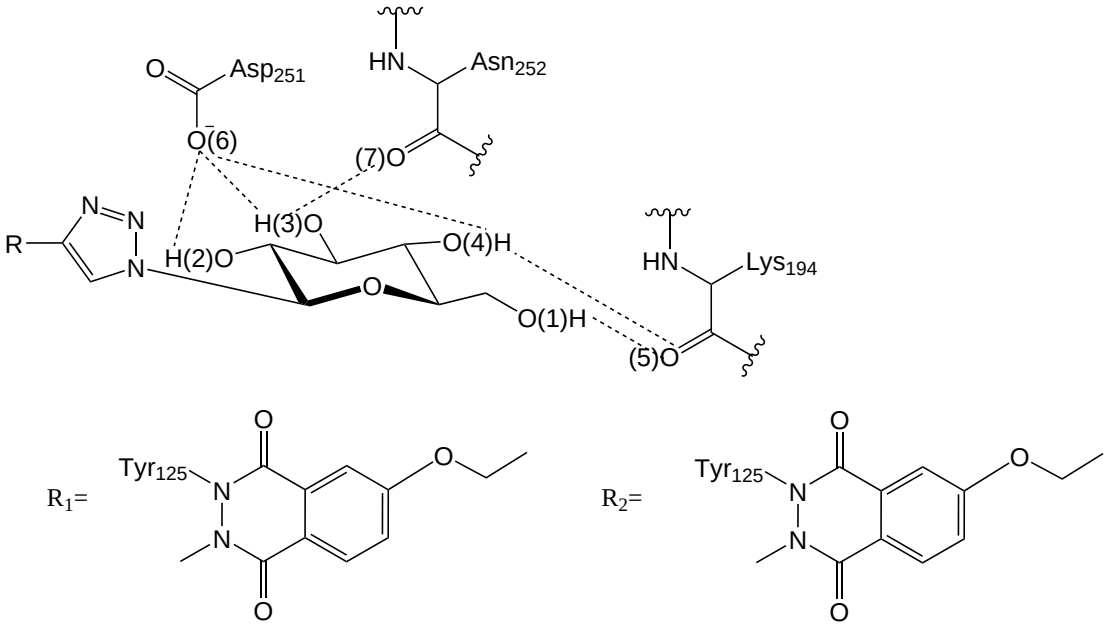


Figure 12.50: Interaction of covalent protein-inhibitor complex of 2-AzGlc-1 or 2-AzGlc-2 and Bm-LS variant Y196F/Y247F/S125Y depicted with dashed lines. Location of amino acids and 2-AzGlc serves for clarification and does not reflect actual position in MD simulations.

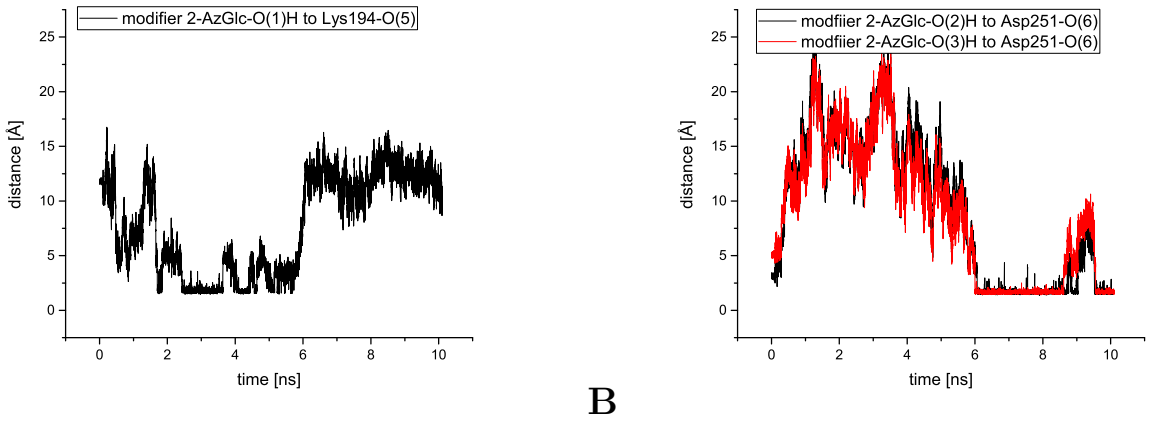


Figure 12.51: The variation in distance as a function of simulation time for modifier 2-AzGlc-1 (pose I) in Bm-LS Y196F/Y247F/S125Y variant corresponding to Figure 12.50 during a 10 ns production run. Distance between **A** – 2-AzGlc-1-O(1)H in pose I and Lys194-O(5) or between **B** – 2-AzGlc-1-O(2)H or 2-AzGlc-1-O(3)H in pose I and Asp251-O(6).

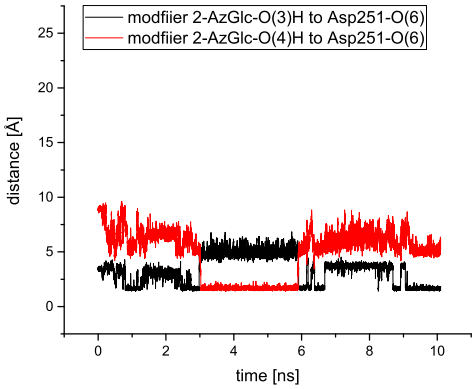
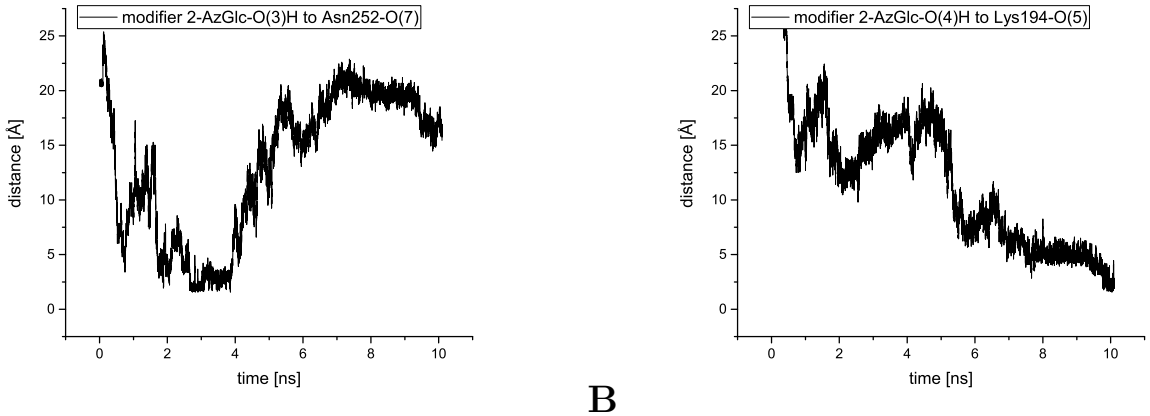


Figure 12.52: The variation in distance as a function of simulation time for modifier 2-AzGlc-2 (pose I) in Bm-LS Y196F/Y247F/S125Y variant corresponding to Figure 12.50 during a 10 ns production run. Distance between 2-AzGlc-2-O(3)H or 2-AzGlc-2-O(4)H in pose I and Asp251-O(6).



A **B**

Figure 12.53: The variation in distance as a function of simulation time for modifier 2-AzGlc-1 (pose II) in Bm-LS Y196F/Y247F/S125Y variant corresponding to Figure 12.50 during a 10 ns production run. Distance between **A** – 2-AzGlc-1-O(3) in pose II and Asn252-O(7) or between **B** – 2-AzGlc-1-O(2)H in pose II and Lys194-O(5).

12.6 Conclusion and comparison to experimental findings

Investigations on the covalent modification of *Bacillus megaterium levansucrase* (Bm-LS) wild type or triple variant by modifier 1, 2 or 2-AzGlc and its effects on the product profile have been conducted in this chapter (Figure 12.5). Bm-LS polymerizes mono-units of levan and usually produces oligosaccharides with a polymerization degree of 2-20. The introduction of modifier 1 lead to a covalent modification of tyrosine residues and thus to an altered product profile with smaller oligosaccharides. Mass spectrometry verified that two tyrosines were varied but could not determine the exact ones. Hence, MD simulations were conducted by linking modifier 1 (if possible both stereoisomers) to Tyr196, Tyr247 or Tyr446 and dynamics were investigated. If a motion towards the active site is detectable one can suggest that a disturbance of the polymerization occurs. In consequence one can conclude which tyrosine is likely modified.

The first aspect was the determination of the modified tyrosine residues in Bm-LS wild type. Initial simulations which include a single covalent attachment of modifier 1-1 (modifier 1 stereoisomer 1) showed that for Tyr196 or Tyr446 no interaction to important active site amino acids can be observed. These findings were independent from the stereoisomer. Only the linkage of the ligand to Tyr247 revealed a possible disturbance of the elongation process. Investigations have further been extended to the combined modification of two tyrosine residues according to the findings of the mass spectrometry. In the process, the twofold modification of Tyr196 and Tyr446 was neglected as individual simulations did not lead to active site interactions. In terms of the concurrent linkage of modifier 1 on Tyr196 and Tyr247 a π-π interaction between the phthalazinedione scaffold of modifier 1-1 on Tyr196 and Tyr247's phenyl ring was noticeable leading to a change of orientation of the ligand on Tyr196 towards Tyr247 (Figure 12.14). Additionally, a stabilization of the second modifier occurred as fluctuations in the RMSD decreased. Simulations of the simultaneous covalent binding of 1-1 on Tyr247 and Tyr446 strengthen the assumption of the subordinate role of Tyr446 as it was not directed towards the active site. Yet it became clear that the modifier shows a certain amount of flexibility as, despite of the pose, motions towards or reverse to the active site are possible. These findings go along with the experimental analysis of products obtained by the Y196F/Y247F mutant. No altered oligosaccharide/levan production was obtained by chemical modification of this variant *via* modifier 1. This indicates that the different product profile occurs due the reaction of the modifier and those two tyrosine residues.

Secondly, investigations on the triple variants of Bm-LS have been carried out. Triple mutations as Y196F/Y247F/N126Y, Y196F/Y247F/Q381Y and Y196F/Y247F/F445Y were conducted and also simulated. The tyrosine residues 196 and 247 have been replaced by phenylalanine to avoid modification. Aim of these changes is to gain control on the polymerization process by inserting tyrosine residues at critical positions *via* mutations which should be solvent exposed and accessible for covalent modification. Experimental results concluded indeed a change of product profile which occurs because of the Y196F/Y247F/N126Y variant. This leads to the heightening in synthesis of small polymers. In terms of the MD simulation of the Y196F/Y247F/F445Y mutant the results occur ambiguous. Depending on the pose of modifier 1 one might or might not expect a disturbance in elongation. On the experimental side, this event was not confirmed. For the case of the Y196F/Y247F/Q381Y mutant no interaction with the active site was observed during the simulation as seen in experiments. Theoretical investigations on Y196F/Y247F/N126Y lead to two possible explanations regarding the changed product profile. Simulations revealed that modifier 1 is directed to the region of elongation which could lead to disturbances. To further study this Bm-LS variant simulations have been extended by including sucrose or 6-kestose source materials of the elongation process. In consequence new findings arose as a motion was visible leading to a transport of sucrose out of the active site. Additionally, a stabilization of the modifier and sucrose or 6-kestose can occur, which is leading to a blockade and thus to a disturbance of the elongation process. Both occurrences are in line with experimental results and might explain the increase of small products of the Y196F/Y247F/N126Y variant. Furthermore, investigations have been extended by conducting simulations of modifier 2 in the Y196F/Y247F/N126Y variant in presence of 6-kestose. In case of the isomer 2 in pose II one can see a stable conformation and a occupancy of the active site by modifier 2. In contrast to that isomer 1 shows motions receding and also approaching towards active site residues. These findings are also in line with experimental results but in contrast to modifier 1 a larger flexibility is observed.

Lastly, dynamics of modifier 2 and 2-AzGlc have been investigated for Y247F or Y196F/Y247F /S125Y variants of Bm-LS. The group of Seibel revealed that these variants of Bm-LS in covalent complex with modifier 2 or 2-AzGlc lead to a production of polysaccharides. To gain further insight on the polymerization MD simulations have been conducted.

In all simulations no interactions with active site residues have been observed which goes along with experimental findings. If the modifier would occupy the active site one can expect the production of shorter oligosaccharides. This is not the case as polysaccharides were produced by the mutants of Bm-LS. Additionally, the research gave insight in terms of

protein-modifier interactions. In case of the Y247F variant polar intermolecular interactions of 2-AzGlc and Lys194 are established, while for the Y196F/Y247F/S125Y variant Glu170 (for modifier 2) or Lys194 and Asp251 (for modifier 2-AzGlc) hydrogen bonds are formed. Interestingly, modifier 2 as well as 2-AzGlc are moving to the same region of protein in the Y196F/Y247F/S125Y despite of the initial position. These findings suggest that the modifier might act as cap due to increased steric hindrance and thus could prevent an early depart of the size-increasing oligosaccharides. Still it has to be mentioned that these results only indicate a possible elongation path. The mechanism is nevertheless not fully understood and further research is of necessity.

Chapter 13

Environmental influence on the proton transfer in cysteine/histidine dyads

13.1 Introduction

Protonation states or the understanding of the proton transfer in enzymes or biological systems is in nearly all theoretical work of relevance. The proton transfer within the enzyme is deterministic for a series of processes e.g. peptide cleavage reactions or enabling covalent modification of inhibitors [285]. Another often investigated problem is the determination of the pKa value of respective protein residues which is, due to the number of atoms, not a trivial matter. The pKa predictor PROPKA offers one possibility to solve this problem [286], [287]. Still, the application of this software can be problematic as it is based on empirical parameters for atomic charge and bond geometries thus might lead to unreliable results [288]. Another option to obtain pKa values is via pH-titration MD (pHtMD) [289]. The procedure is comparable to classical wet-lab titration, but a huge amount of simulation time is necessary.

A lot of studies, which deal with proton transfer in enzymes, are applying QM/MM methodologies, whereas the QM part is calculated with density functional methods [285]. Due to the size of enzymes the choice of QM/MM methodologies is evident, as the description of a whole protein is still not feasible with pure QM. Therefore, the reaction itself is calculated with QM approaches, while the environment is described with simpler methods (force field). The relevance of such investigations can be explained as follows: changes in the pH value is concurrent involved with a change of the protonation state of amino acids as cysteine, lysine, histidine etc. Thus the abstraction or addition of a proton at these amino acids leads to a

change of the charge distribution within the whole protein and can result in conformational changes. Therefore, binding properties of ligands or the activity of the protein itself might be affected [289].

In literature plenty of QM/MM studies concerning the proton transfer in enzymes can be found. Vidossich et al. investigated the water-mediated proton transfer and its role in the peroxidase catalysis of horseradish peroxidase *via* MD simulations and QM/MM calculations. Another theoretical study focused on the reaction mechanism of the peptide-bond formation in the ribosome and revealed additionally to the direct transfer a proton-shuttle mechanism via the ribose-2'-OH group.

Among the clarification of enzyme specific mechanisms the understanding of neutral or zwitterionic protonation states in cysteine/histidine dyads is an integrial part and in case of cysteine proteases still subject of controversy [290]. Mladenovic et al. showed the relevance of active side histidine which enables the water-mediated proton transfer to epoxide- or aziridine-based inhibitors *via* a relay system and therefore is responsible for the activity of these inhibitors. In case of the arginine deiminase catalysis the protonation state of the cysteine nucleophile was determined by comparison with experimental findings [291]. For KasA, which is involved in the biosynthesis of the cell wall of *Mycobacterium tuberculosis*, Lee et al. revealed that the protonation state of the catalytic Cys/His dyade can switch easily between both protonation states [292]. Of course these are just a few examples which can be found in literature in regard to the determination of the proton transfer or the protonation state in enzymes.

Clearly, it is of necessity to probe the reliability of applied quantum chemical method for the description of the protonation state. One can find a lot of benchmarking studies on this matter, although they are mostly performed under gas phase conditions [293], [294], [295], [296], [297]. Further analysis on this matter has been conducted by Paasche et al. In his benchmark paper, the accuracy of different quantum chemical methods has been probed for the cysteine-histidine proton transfer [285]. Investigations have been expanded from the gas phase by mimicking the enzyme environment with continuum models or the explicit inclusion via QM/MM methodologies. These investigations focused solely on the proton transfer starting from the neutral state and do not include the backwards reaction.

In this chapter similar computational investigations have been conducted to describe the proton transfer of the catyltic cys/his dyade in FadA5 or in cruzain (T. *cruzi*) *via* MD simulations and QM/MM calculations. In contrast to the work of Paasche et al. both protonation states have been considered as starting points for the proton transfer to detect possible discrepancies in the description. In this matter several post-HF and DFT functionals were

additionally benchmarked against CCSD(T) to probe the suitability of applied methodologies. Furthermore, investigations have been extended to analyze the relaxation process of the protein due to the proton transfer and will be described in the following sections.

13.2 System preparation and MD simulations

As a prerequisite it was necessary to perform MD simulations to obtain equilibrated structures of the apoprotein of FadA5 and cruzain for QM/MM calculations. In case of FadA5 chain B of the apoprotein structure (pdb code 4UBW) was used as starting point. In contrast to chain A, all residues were located in the X-ray structure of chain B. As no apoprotein structure of cruzain was available a different solution was needed. Therefore, a structure of a non-covalent inhibitor bound to cruzain has been chosen (pdb code 1ME4) because of its high X-ray structure resolution of 1.2 Å. Afterwards the inhibitor was removed to obtain the artificial apoprotein structure of cruzain.
Just like in section 5.1 the apoprotein complexes of FadA5 and cruzain were described by the AMBER14 force field ff14SB. In regard to minimization and equilibration the same protocol was conducted as described in section 5.1. In terms of the proton transfer two different protonation states are conceivable: (protonation state I) a neutral catalytic dyade consisting of a protonated cysteine and a singly protonated histidine or (protonation state II) a zwitterionic catalytic dyade with a deprotonated, negatively charged cysteine and a doubly protonated positively charged histidine (Figure 13.1). As no proton transfer can take place during a classical MM simulation both states have to be considered in the MD simulations to obtain two different starting points for the QM/MM proton transfer calculations. By doing so it is observable that the protein environment relaxes to the protonation state of the catalytic dyade and these effects are evaluated in the next sections. The duration of each simulation was 10 ns leading to a total simulation time of 40 ns by including both protonation states of the catalytic dyade of FadA5 and cruzain.
In contrast to the other sections a detailed investigation on the MD simulations was not performed as they were only needed to obtain a relaxed structure as starting point for further QM/MM calculations.

13.3 QM/MM calculations and benchmark

Firstly, a relaxed structure of FadA5 and cruzain in both protonation states has been extracted from the MD simulation which served as starting point for the QM/MM calculations. It was paid attention that all structures contained a water molecule located between active site cysteine and histidine which acts as proton donator and acceptor. By doing so one achieves comparability for the investigated reaction.
The initial step started with a local optimization by employing QM/MM methodology utilizing the ChemShell 3.6.0 software package and an electrostatic embedding with the charge shift method. The QM region (depicted in red in Figure 13.1) has been optimized by employing B3LYP-D3 and the def-TZVP basis set. The rest of the system was treated by the AMBER14 force field. To decrease computational costs only molecules in a radius of 10 Å around the active site cysteine residue were optimized, while the rest of the system was fixed. Afterwards a relaxed scan on QM/MM level has been conducted resulting in four different potential energy surfaces (Figure 13.2 and Figure 13.3). In all cases the same methodology has been used as mentioned before. Interestingly, one can observe a similar trend for both proteins: the stability of the protonation state is determined by the initial protonation state of the used snapshot. If the relaxed scan starts from the optimized neutral catalytic state (protonation state I) it is energetically lower than the zwitterionic state after water-mediated proton transfer and vice versa. In case of FadA5 the height of barrier is 9.7 kcal/mol while the zwitterionic state is higher by 8.1 kcal/mol compared to the initial neutral starting point. These findings are similar to the QM/MM computations of cruzain whereby the obtained values are lower: the barrier is of 7.9 kcal/mol while the energy difference after the proton transfer is 3.1 kcal/mol. Contrary to these findings are the results from the zwitterionic starting point. It is observable that this state is favored if the snapshot of the MD simulation in the zwitterionic one is used. For FadA5 the barrier increases to 17.0 kcal/mol and the neutral state is described as less stable compared to the zwitterionic state with difference in energy of 13.1 kcal/mol. Again, this trend can be seen for cruzain with a barrier of 8.7 kcal/mol and an energetic difference of 4.3 kcal/mol after proton transfer.

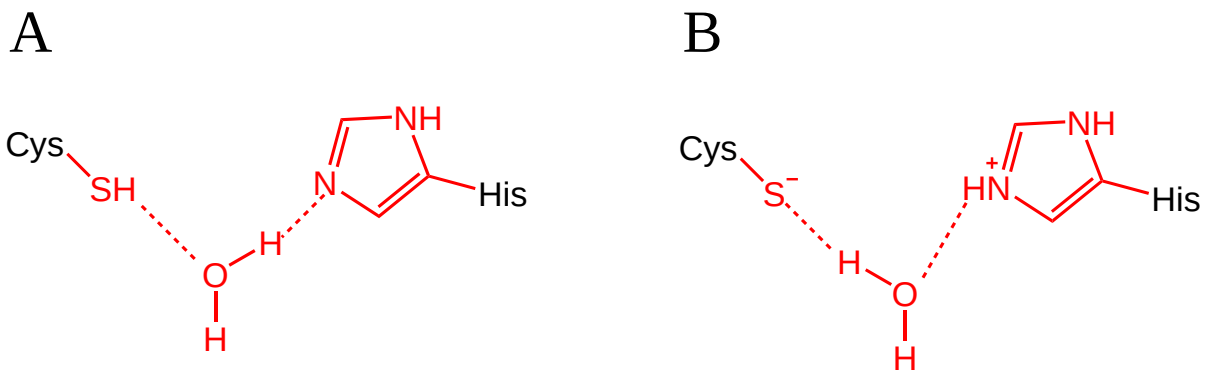


Figure 13.1: Red coloured lines indicate atoms, which has been used in the QM part of a QM/MM relaxed scan of the proton transfer in FadA5 or cruzain. **A** – Starting point from a neutral catalytic dyade of cysteine and histidine (protonation state I). **B** – Starting point from a zwitterionic catalytic dyade of cysteine and histidine (protonation state II). Dashed lines indicate the coordinates which were varied during relaxed scan. Location of amino acids and water molecule serves for clarification and does not reflect actual position in QM/MM calculation.

These findings indicate a large influence of the surrounding on the stability of the protonation state which is observable for both proteins. As negatively charged sulfur is present in the QM region, the necessity of augmented basis sets is obvious for a better description of the transition as well as the zwitterionic state. Thus, a benchmark has been performed by conducting QM/MM single point energy calculations on the transition state as well as for both protonation states employing different wave function-based or DFT approaches summarized in Table 13.1 and Table 13.2 while the description of the MM part remained the same. The basis set was increased to aug-cc-pVDZ, in case of SCS-MP2 or SCS-CC2 additionally to aug-ccpVTZ. The results have been compared to the CCSD(T) methodology which is the gold standard for ground state energy calculations.
If one takes a look at Table 13.1 or Table 13.2, it is observable that the results obtained by SCS-MP2 or SCS-CC2 in combination with aug-cc-pVDZ agree best with the benchmark values obtained with CCSD(T)/aug-cc-pVDZ. Slightly worse results are achieved if the basis set is increased to aug-cc-pVTZ except for SCS-MP2 in cruzain starting from protonation state I. This might be explainable as only aug-cc-pVDZ was employed for CCSD(T) calculations. It is possible that an increase of the basis set to aug-cc-pVTZ in case of CCSD(T) might result in comparable results but was not performed due to computational costs. Additionally, one can see that the usage of a larger basis set decreases the barrier height while the "final" state after proton transfer is destabilized. These effects are still not large and do not affect

the trend in prediction. Computations also show that the usage of the spin-component scaled electron correlation method constantly improves the results. In terms of the DFT functionals the same observation as before can be made.

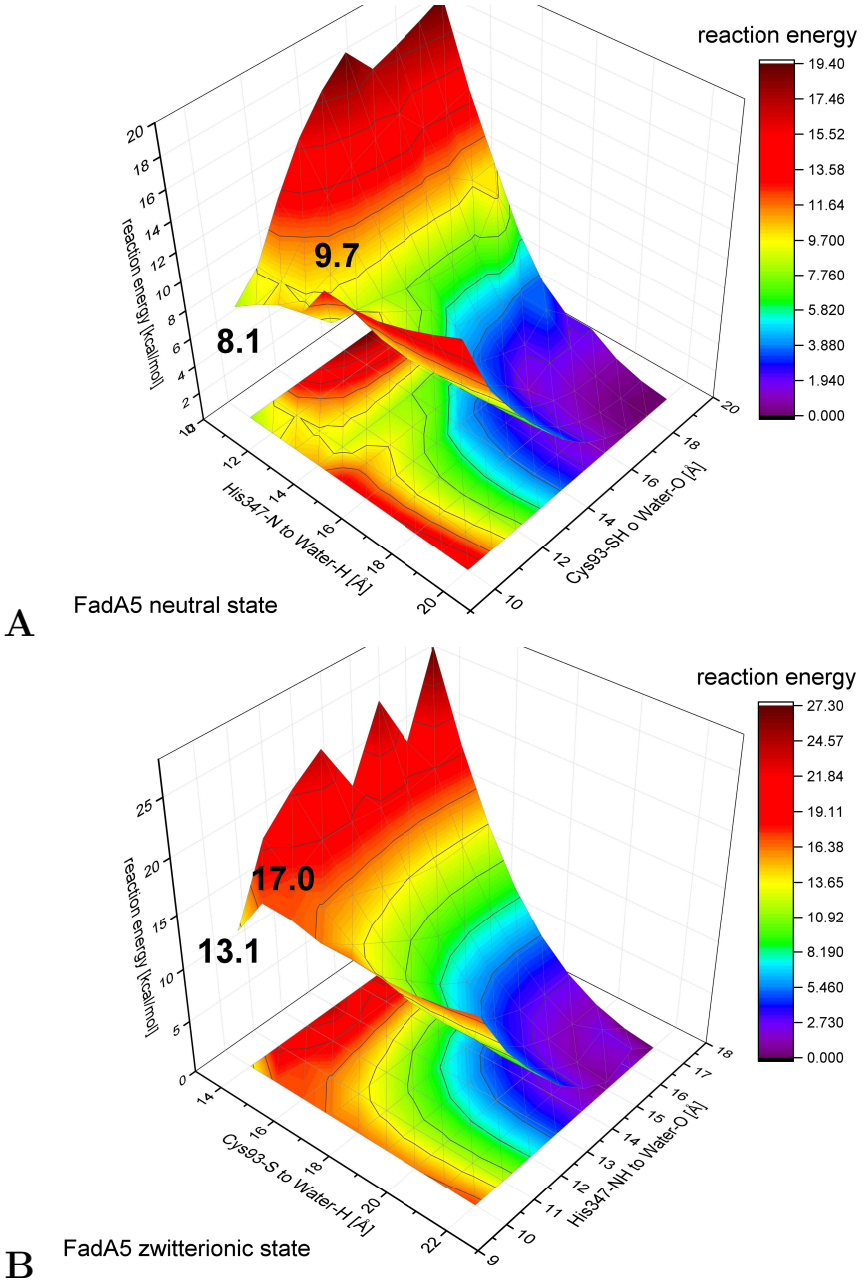


Figure 13.2: Relaxed scan of proton transfer in FadA5 depicted in Figure 13.1. **A** – Starting from neutral state of cysteine and histidine catalytic dyade. **B** – Starting from zwitterionic state of cysteine and histidine catalytic dyade. Energies are obtained by B3LYP-D3/def-TZVP optimization and are in relation to the starting point of the relaxed scan. All energy values are given in kcal/mol.

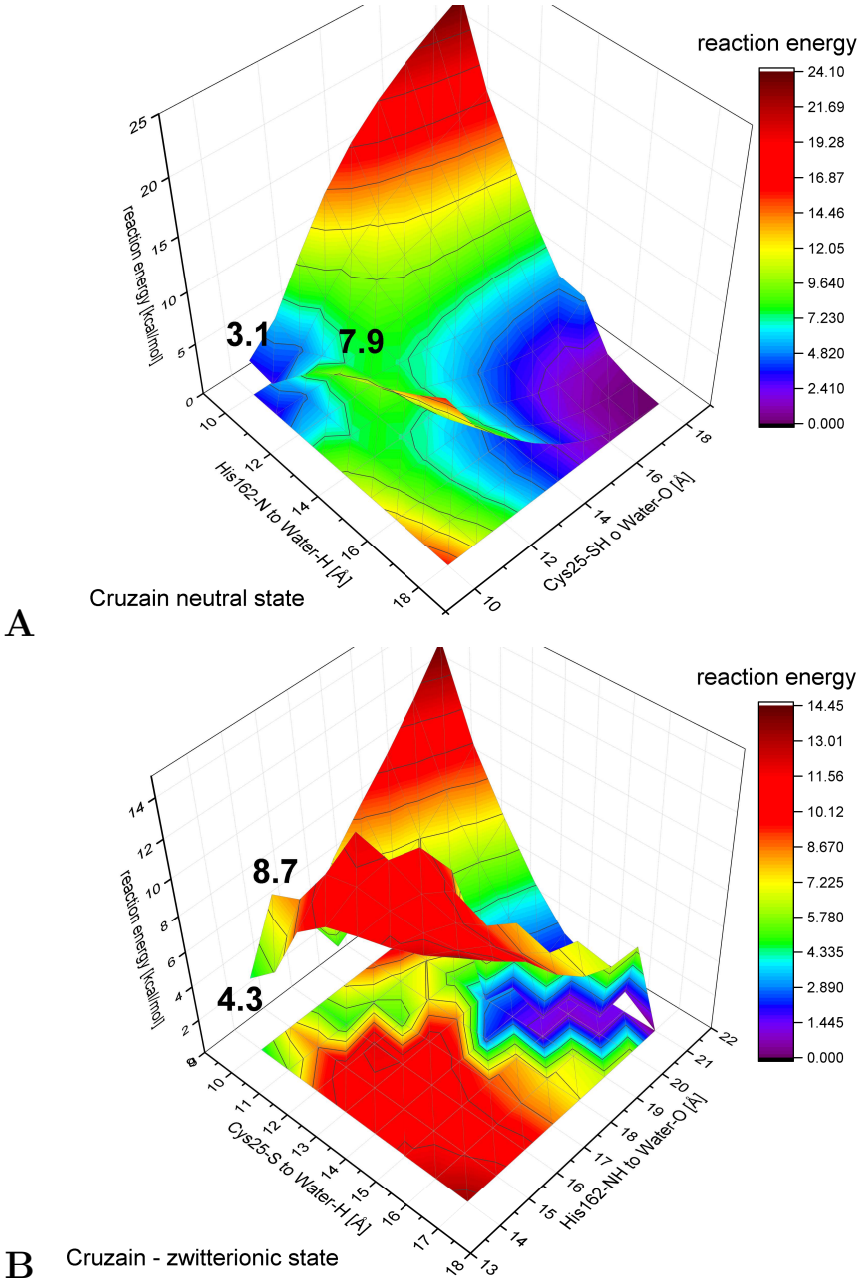


Figure 13.3: Relaxed scan of proton transfer in cruzain depicted in Figure 13.1. **A** – Starting from neutral state of cysteine and histidine catalytic dyade. **B** – Starting from zwitterionic state of cysteine and histidine catalytic dyade. Energies are obtained by B3LYP-D3/def-TZVP optimization and are in relation to the starting point of the relaxed scan. All energy values are given in kcal/mol.

Table 13.1: QMMM transition state and reaction energies (ΔE_{reac}) of water-mediated proton transfer in FadA5 employing different ab initio or DFT calculations compared to CCSD(T)/aug-cc-pVDZ calculations from neutral or zwitterionic starting point. For * marked methods the aug-cc-pVTZ basis was used, otherwise the aug-cc-pVDZ basis set. All energy values are shown in kcal/mol.

neutral catalytic dyade starting point		
HIDCYS	**TS**	**zwitterion.**
CCSD(T)	12.5	9.2
HF	20.2	10.8
MP2	10.2	8.6
CC2	9.9	8.4
SCS-MP2	12.8	9.7
SCS-CC2	12.5	9.6
SCS-MP2*	12.4	10.1
SCS-CC2*	12.2	10.1
LDA	-2.5	-0.4
PBE	4.8	4.9
BP86	4.7	4.7
BLYP	7.5	6.1
B3LYP	8.8	6.4
B3LYP_G	9.3	6.9
BHLYP	12.2	7.7

zwitterionic catalytic dyade starting point		
HIPCYM	**TS**	**neutral**
CCSD(T)	19.1	14.1
HF	30.2	22.0
MP2	16.9	13.6
CC2	16.2	12.9
SCS-MP2	19.9	14.6
SCS-CC2	19.4	14.1
SCS-MP2*	18.9	14.7
SCS-CC2*	18.5	14.2
LDA	7.3	11.9
PBE	12.5	12.7
BP86	12.6	12.9
BLYP	15.2	13.3
B3LYP	17.1	14.8
B3LYP_G	17.6	15.1
BHLYP	21.0	17.2

Table 13.2: QMMM transition state and reaction energies (ΔE_{reac}) of water-mediated proton transfer in cruzain employing different ab initio or DFT calculations compared to CCSD(T)/aug-cc-pVDZ calculations from neutral or zwitterionic starting point. For * marked methods the aug-cc-pVTZ basis was used, otherwise the aug-cc-pVDZ basis set. All energy values are given in kcal/mol.

neutral catalytic dyade starting point		
HIPCYM	**TS**	**neutral**
CCSD(T)	10.6	3.7
HF	18.5	3.1
MP2	8.2	3.0
CC2	7.9	3.2
SCS-MP2	9.9	3.6
SCS-CC2	9.7	3.8
SCS-MP2*	10.4	3.9
SCS-CC2*	10.2	4.1
LDA	-3.3	-2.2
PBE	2.8	1.1
BP86	2.7	0.9
BLYP	5.3	1.8
B3LYP	7.0	1.7
B3LYP_G	7.3	2.2
BHLYP	10.6	2.3

zwitterionic catalytic dyade starting point		
HIDCYS	**TS**	**zwitterion.**
CCSD(T)	10.4	4.0
HF	17.5	7.5
MP2	8.8	4.2
CC2	8.5	3.9
SCS-MP2	10.5	4.1
SCS-CC2	10.2	3.9
SCS-MP2*	10.2	4.4
SCS-CC2*	9.9	4.2
LDA	2.1	7.6
PBE	5.9	5.6
BP86	6.0	5.6
BLYP	7.8	5.0
B3LYP	9.0	5.8
B3LYP_G	8.8	5.6
BHLYP	11.2	6.5

Except for LDA, the initial protonation state is always lower in energy compared to the one after proton transfer. Interestingly, BHLYP agrees better with the CCSD(T) value than MP2 or CC2 in case of the neutral starting point and B3LYP or B3LYP_G for the zwitterionic starting point in FadA5. Still, in all other cases DFT approaches are always worse compared to spin-component scaled methods as their MAE is larger. If one takes the MAE values of the transition state and the proton transfer reaction energies into account, it is observable that the lowest MAE values are achieved if the spin component scaled methodologies is employed (Figure 13.4). Pure MP2 or CC2 results are about 1 kcal/mol higher in MAE value. As to be expected Hartree Fock leads to the worst results in terms of the applied ab initio methods. For DFT results one can clearly see that Jacobs Ladder is reproduced: the accuracy increases at the same time as the complexity of the applied functional. LDA which is the simplest approximation leads to a MAE value of 9 kcal/mol while the MAE value of GGA methods as PBE, BP86 or BLYP are below 5 kcal/mol [298]. The hybrid functional B3LYP (in the Turbomole and Gaussian implementation) and also BHLYP perform similar to MP2 or CC2 and have a MAE value below 2.2 kcal/mol.
To gain further insight, MAE values of the reaction barrier or of the proton transfer reaction energy have been calculated (Figure 13.5 and Figure 13.6). Except for the spin-component scaled methods larger MAE values are obtained for the reaction barrier compared to the reaction energy. Especially HF, LDA and the GGA methods PBE and PB86 have in this respect MAE values in the range of 7 to 12 kcal/mol. Comparable results regarding the MAE values are only obtained if spin-component scaled methods are applied as their MAE values are below 0.5 kcal/mol for the reaction energy as well as for the reaction barrier.

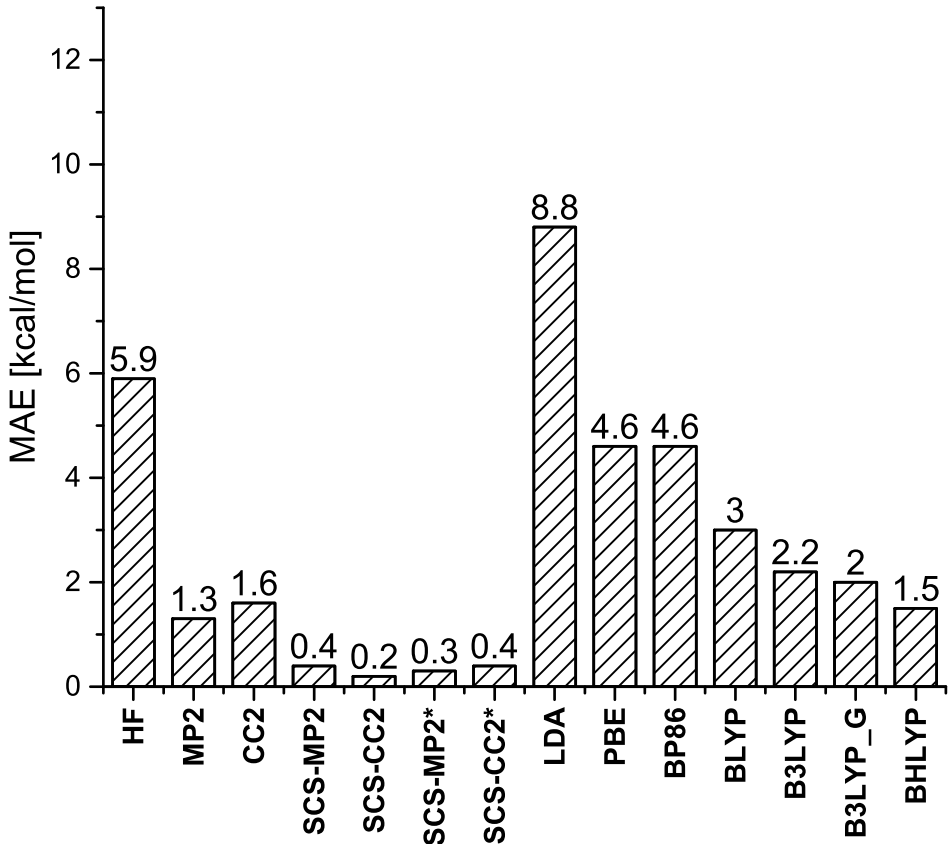


Figure 13.4: Mean average error with respect to CCSD(T)/aug-cc-pVDZ averaged of all proton transfer reaction energies and reaction barrier (neutral and zwitterionic) in FadA5 and cruzain. For * marked methods the aug-cc-pVTZ basis was used, otherwise the aug-cc-pVDZ basis set. All MAE values are presented in kcal/mol.

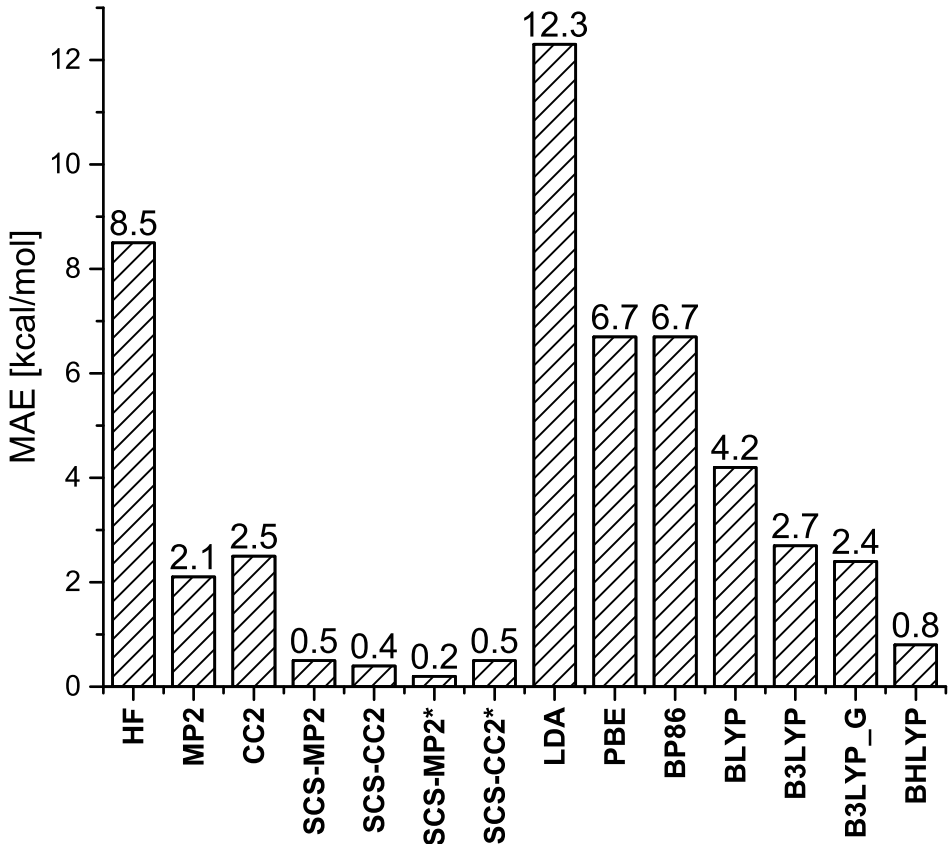


Figure 13.5: Mean average error with respect to CCSD(T)/aug-cc-pVDZ averaged of all reaction barrier (neutral and zwitterionic) in FadA5 and cruzain. For * marked methods the aug-cc-pVTZ basis was used, otherwise the aug-cc-pVDZ basis set. All MAE values are presented in kcal/mol.

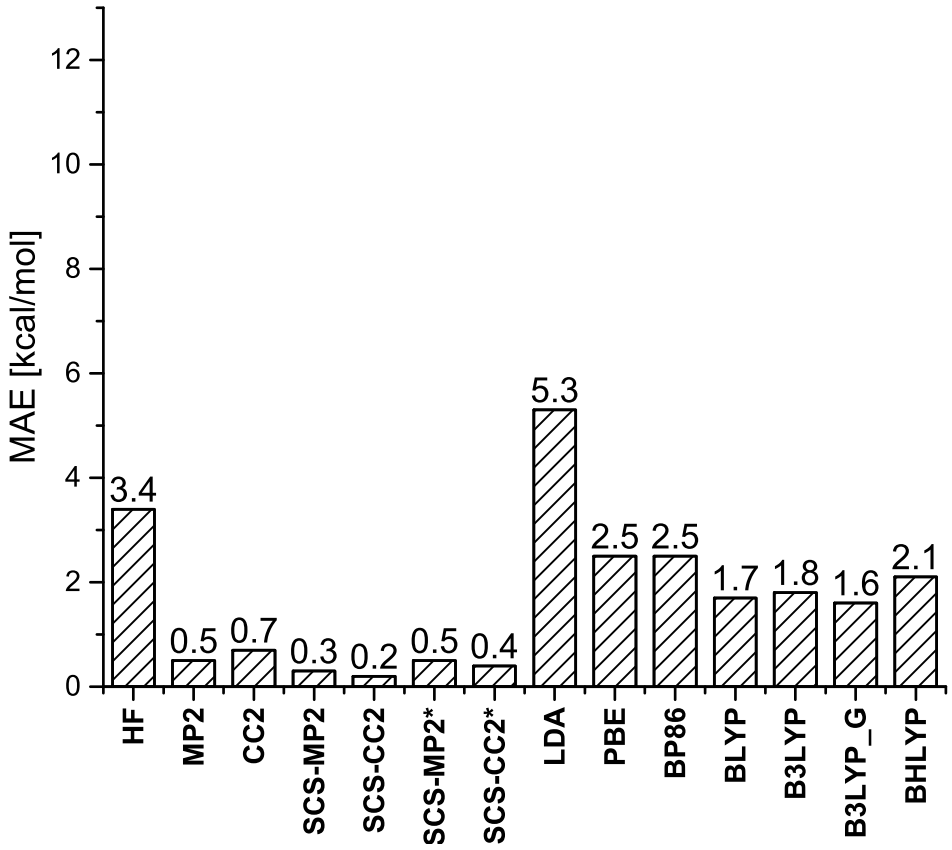


Figure 13.6: Mean average error with respect to CCSD(T)/aug-cc-pVDZ averaged of all proton transfer reaction energies (neutral and zwitterionic) in FadA5 and cruzain. For * marked methods the aug-cc-pVTZ basis was used, otherwise the aug-cc-pVDZ basis set. All MAE values are presented in kcal/mol.

To prove the validity of CCSD(T) it is necessary to analyze the D_1 diagnostics. As one can see in Table 13.3 the CCSD D_1 diagnostics show values below 0.06 and thus the description of the QM part via the coupled cluster approach can be considered as valid as one can see that multiple configurations in the wave functions' description is not necessarily needed.

Table 13.3: Summary of CCSD D_1 diagnostic of neutral, transition and zwitterionic state of proton transfer in FadA5 or cruzain. All calculations were performed with the CCSD(T)/aug-cc-pDVZ approach.

FadA5	**starting point protonation state I**	**neutral**	**transition state**	**zwitterionic**
	CCSD D_1 diagnostic	0.0405	0.0440	0.0495
	starting point protonation state II	**zwitterionic**	**transition state**	**neutral**
	CCSD D_1 diagnostic	0.0499	0.0406	0.0402
cruzain	**starting point protonation state I**	**neutral**	**transition state**	**zwitterionic**
	CCSD D_1 diagnostic	0.0428	0.0456	0.0505
	starting point protonation state II	**zwitterionic**	**transition state**	**neutral**
	CCSD D_1 diagnostic	0.0541	0.0447	0.0447

13.4 Change of protein environment and its influence on proton transfer

Results of the previous section indicate a large influence of the protein environment on the prevailing protonation state. As results of the benchmark calculations also show that this observation can be made independent of the applied methodology (except for LDA) further investigations have been conducted. One can imagine two different protein environments A and B resulting from the snapshot of the MD simulation in protonation state I (neutral starting point) or protonation state II (zwitterionic starting point). If one starts from the neutral state in environment A, this state is clearly preferred as the snapshot is extracted from a MD simulation in which the protein has adapted to the protonation state. Now the idea is to insert the QM region of the neutral, transition and zwitterionic state of environment A into the "wrong" environment B which is adapted to the zwitterionic protonation state II. By doing so it is to be expected that the zwitterionic state is now stabilized by environment B

and should be lower in energy compared to the neutral state. This change of environment can also be done by inserting above mentioned QM regions of environment B into environment A.

In a first step the optimized QM structures of the neutral, transition and zwitterionic state including active site cysteine, histidine residue as well as the water molecule involved in the proton transfer have been extracted from environment A (or B) and were inserted into environment B (or A). This is exemplary shown for the neutral state in environment A and B in Figure 13.7. To ensure a connection between the residues, backbone atoms of cysteine and histidine have also been considered in the exchange. Afterwards a QM/MM optimization on B3LYP/def-TZVP level including dispersion correction has been performed for all three states with the same QM region and applied constraints as in environment A (or B) depicted in Figure 13.1. The MM region was described with the AMBER14 force field. Afterwards QM/MM single point energy calculations were conducted on SCS-MP2 or SCS-CC2 level with an aug-cc-pVDZ basis set as these methods show the best results in the benchmark.

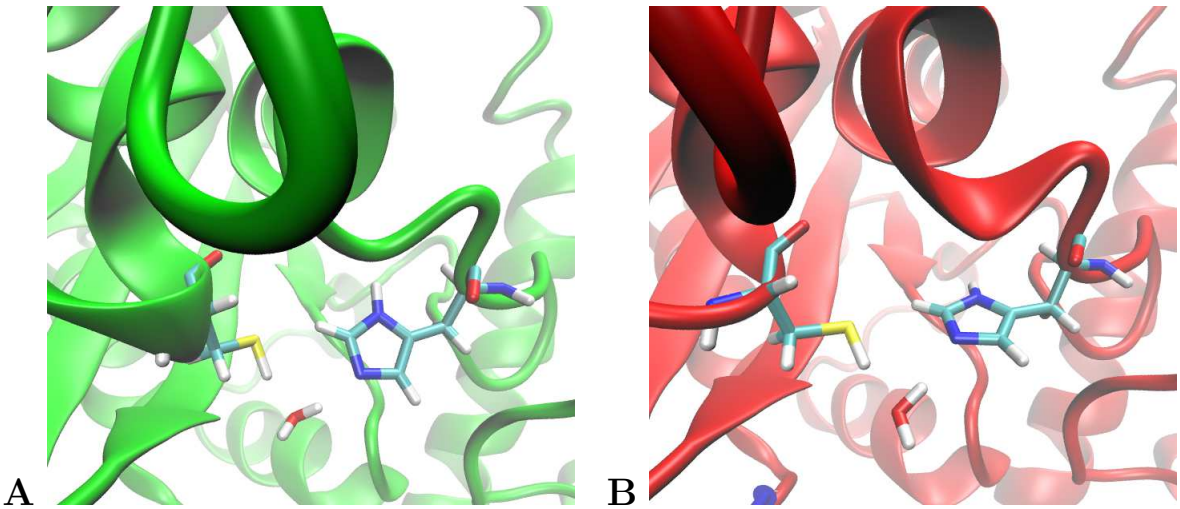


Figure 13.7: **A** – QM/MM optimized structure of neutral catalytic dyade of cysteine and histidine in cruzain environment A (green). **B** – QM/MM optimized structure of neutral catalytic dyade of cysteine and histidine inserted in environment B (red) of cruzain and reoptimized. The QM/MM optimization employed B3LYP-D3/def-TZVP and the AMBER14 force field.

As initially expected one can see a decrease in reaction energy of the proton transfer after the exchange (Table 13.4 and Figure 13.8). In case of FadA5 the zwitterionic state decreases to the same level as the neutral state (if one starts from the protonation state I) while the height of barrier also lowers by $\sim$ 2 kcal/mol. In the other case the QM region of the neutral

state in the "wrong" environment A (extracted from environment B) is 20 kcal/mol more stable in comparison to the zwitterionic state. Interestingly the previous transition state is lower than the initial neutral state by 14 kcal/mol. In case of cruzain comparable results are obtained. In both cases the height of barrier decreases by approximately 3 to 5 kcal/mol while the endpoint of the proton transfer is lower by 1 or 2 kcal/mol. These findings strengthen the assumption of the influence of the protein environment regarding the preferred protonation state.

Table 13.4: QM/MM reaction energies (ΔE_{reac}) of water-mediated proton transfer in FadA5 or cruzain employing different ab initio method. All calculations used the aug-cc-pVDZ basis set. All energy values in kcal/mol.

FadA5	**SCS-MP2**[1]	**SCS-CC2**[1]	**SCS-MP2**[2]	**SCS-CC2**[2]
TS	12.8	12.5	10.1	9.6
zwitterion.	9.7	9.6	-0.2	-0.2
	SCS-MP2[3]	**SCS-CC2**[3]	**SCS-MP2**[4]	**SCS-CC2**[4]
TS	19.9	19.4	-13.7	-13.9
neutral	14.6	14.1	-19.9	-20.1

cruzain	**SCS-MP2**[1]	**SCS-CC2**[1]	**SCS-MP2**[2]	SCS-CC2[2]
TS	9.9	9.7	6.7	6.5
zwitterion.	3.6	3.8	-1.2	-0.9
	SCS-MP2[3]	**SCS-CC2**[3]	**SCS-MP2**[4]	**SCS-CC2**[4]
TS	10.5	10.2	5.7	5.4
neutral	4.1	3.9	-2.2	-2.4

[1]Calculated energy of transition state and reaction energy from neutral starting point in environment I. [2]Calculated energy of transition state and reaction energy from neutral starting point in environment II. [3]Calculated energy of transition state and reaction energy from zwitterionic starting point in environment II. [4]Calculated energy of transition state and reaction energy from zwitterionic starting point in environment I.

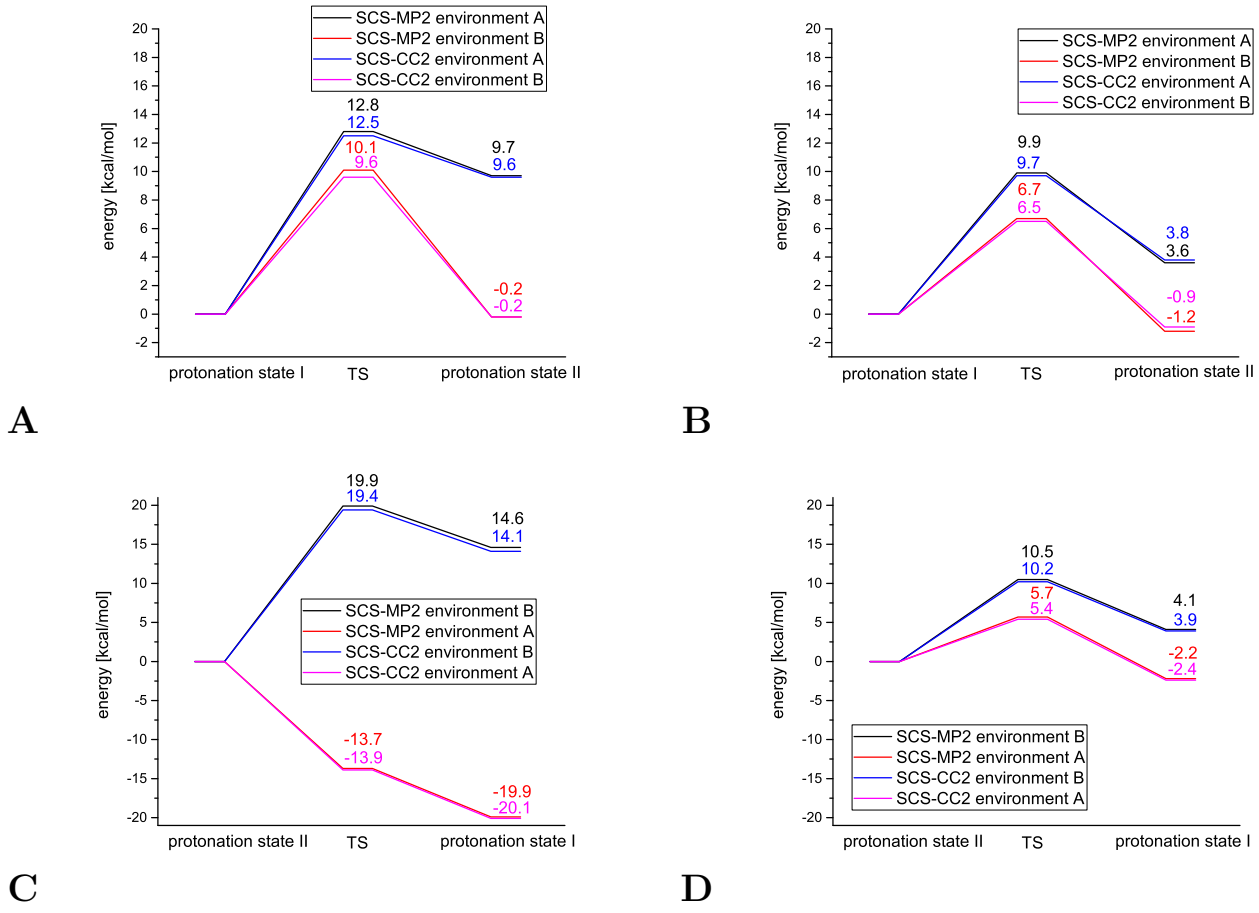


Figure 13.8: Reaction energies of proton transfer obtained by QM/MM single point calculations employing SCS-MP2 or SCS-CC2 and a aug-cc-pVDZ basis set starting from protonation state I in both environment A and B in **A** – FadA5 or **B** – cruzain or from protonation state II in both environment in **C** – FadA5 or **D** – cruzain. Depicted values correspond to Table 13.4. All values in kcal/mol.

13.5 Dynamics of protein adaptation to the protonation state

Lastly, the investigations focused on the dynamics of the protein adaptation to the protonation state. It is imaginable that the proton transfer itself is a much faster reaction as the rearrangement of the protein. If one imagines a water dimer, hydrogen bonds can be formed or broken due to thermal fluctuations alone, thus the shifting of protons along hydrogen bonds can occur spontaneously at 300 K [299], [300]. In consequence the question arises

how much time is needed for the adaptation process of the protein. Therefore classical MD simulations have been performed using the QM/MM optimized structure of the neutral state in environment A and B (or zwitterionic state in both environments) to gain further insight on the time scale of environmental relaxation. The difference to previous simulations is the usage of a harmonical restraint by a value of 2000 kcal mol^{-1} $Å^{-2}$ to prevent large motions of the QM region. The reason why both environments were considered is due to two circumstances. Firstly, both snapshots of the respective protonation state starts from a QM/MM optimized structure. Therefore it is understandable that a MD simulation is needed to equilibrate the MM region. Secondly, it is of interest how much time is required for a relaxation of the "wrong" environment in comparison to the "right" one. To achieve a certain amount of comparability and validity 2-dimensional RMSD plots were calculated which include the motion of both MD simulations. In concrete terms 200 snapshots of the MD simulation of protonation state I (FadA5 and cruzain) in the wrong environment B have been compared with 200 snapshots of the MD simulation in the right environment A which also has been done for protonation state II (Figure 13.9). To decrease the computational costs only the protein environment in a radius of 10 Å close to the active cysteine was considered.

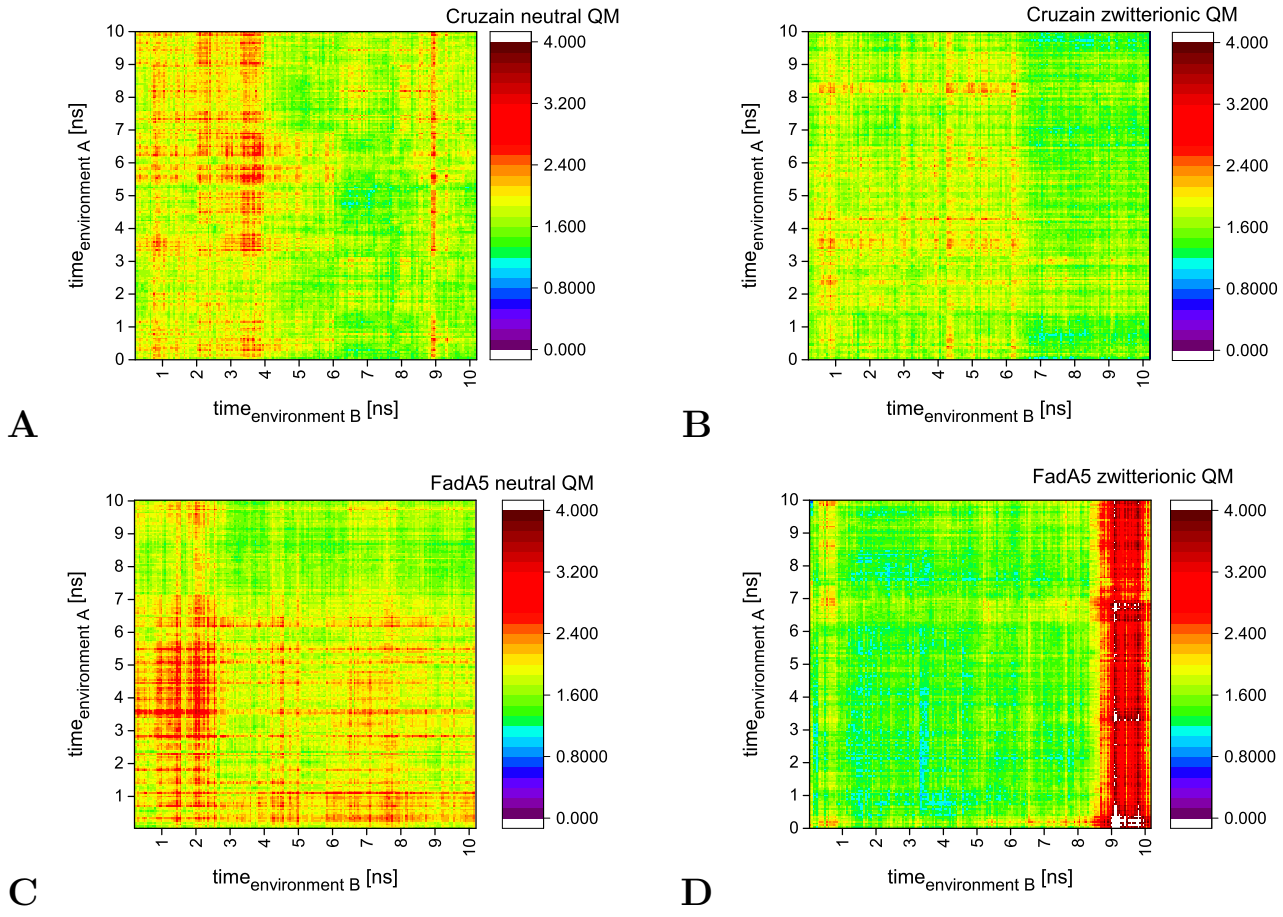


Figure 13.9: 2d-RMSD plot of MD simulation of QM part inserted into different MM region (x-axis) compared to MD simulation of QM part in original MM region (y-axis). RMSD was calculated for 10 Å protein around Cys25 for cruzain or Cys93 for FadA5. 200 snapshots have been considered for each MD simulation. **A** - Neutral QM region of cruzain in environment A and B. **B** – Zwitterionic QM region of cruzain in environment A and B. **C** - Neutral QM region of FadA5 in environment A and B. **D** – Zwitterionic QM region of FadA5 in environment A and B.

To simplify the 2d RMSD plots the mean RMSD values have been calculated by using the following equation:

$$meanRMSD = \sum_{b=1}^{n} \frac{RMSD_t^b}{n} \tag{13.1}$$

The mean RMSD is the averaged summation of all RMSD values ($RMSD_t^b$) of the MD simulation in the "right" environment compared to a snapshot in the artifical environment (Figure 13.10).

For cruzain in protonation state I (neutral dyad) it is observable that after 4 ns a relaxation of the environment occurs leading to a decrease in mean RMSD value from ≈ 2.3 Å to 1.8 Å and after 7 ns to 1.5 Å. For the zwitterionic protonation state a reduction in mean RMSD value takes place after 6 ns from ≈ 2 Å to 1.5 Å. In contrast to protonation state I snapshots in the wrong environment during 7 – 10 ns show RMSD values below 1 Å compared to the snapshots in the simulation time of 9-10 ns in the right environment (Figure 13.9-B). Additionally, no increase in mean RMSD value can be seen after 7 ns. For FadA5 in protonation state I a diminishing in mean RMSD value can be observed after 2 ns. If one compares the 2d RMSD plot of FadA5 in the neutral state (Figure 13.9-C) to cruzain in in the zwitterionic state (Figure 13.9-B), it becomes clear that in case of FadA5 only snapshots after 7 ns in environment A (right environment) are similar to structures in environment B leading to low RMSD values in the upper part of the 2d-RMSD plot. Lastly, the dynamics of the zwitterionic state of FadA5 in both environments have been analyzed. In terms of the mean RMSD value a decrease below 1.5 Å can be observed after 1 ns and a rise to 4 Å after 8 ns.

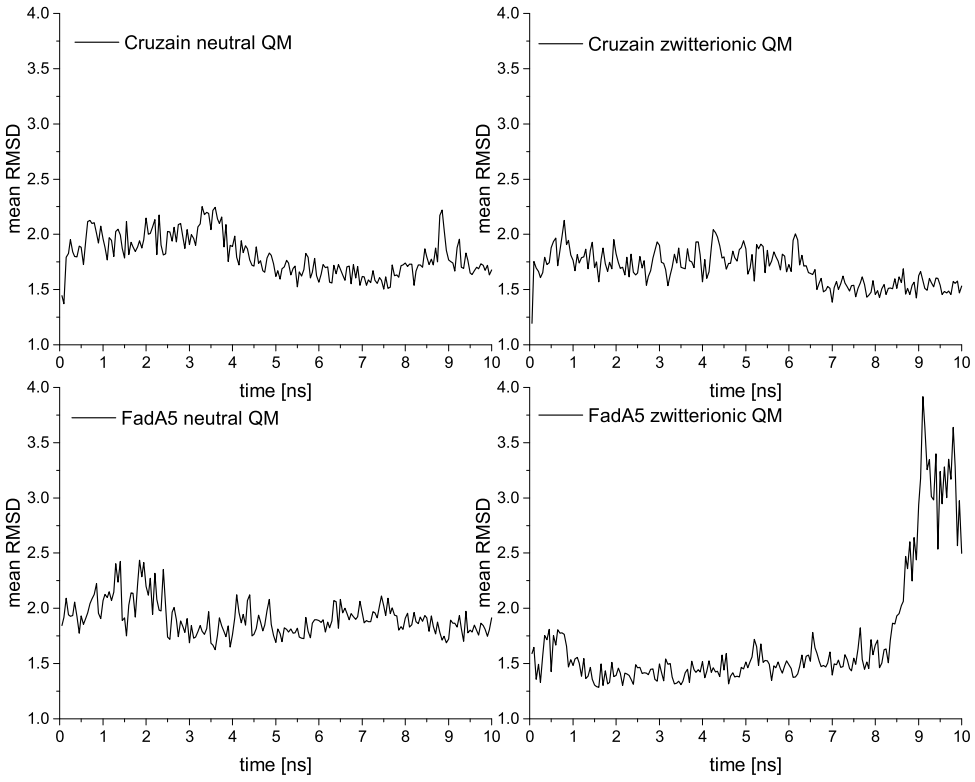


Figure 13.10: Mean RMSD values as a function of time considering protein environment in a radius of 10 Å close to Cys25 for cruzain or Cys93 for FadA5. 200 snapshots have been considered for each MD simulation. All values are given in Å.

13.6 Conclusion

Investigations in respect to the proton transfer in cysteine histidine dyads were conducted in this chapter to gain further insight environmental effects (i.e. enzyme environment). In this study MD simulations and QM or QM/MM calculations were performed on FadA5 and cruzain which served as test systems. Why are these investigations of interest? Cysteine histidine dyads are often located in the active site of e.g. cysteine proteases and can be adressed by covalent inhibitors. Still, the modification can only be initiated if the cysteine is deprotonated to act as a nucleophile. Therefore, it is of importance to understand the mechanism of the proton transfer in enzyme as well as the influence of the surroundings and

its effects on the reaction.
Initially, MD simulations have been conducted to obtain a relaxed structure for the proton transfer calculated on QM/MM level. In this context the neutral and zwitterionic protonation states have been considered: a singly protonated cysteine and histidine (protonation state I) or a deprotonated cysteine and doubly protonated histidine (protonation state II). It was observable that the preferred protonation state is determined by the used snapshot. If it is extracted by a MD simulation in the neutral state (Figure 13.1-A), it is favored compared to the zwitterionic state (Figure 13.1-B) and vice versa. This was shown for all four cases by a relaxed scan utilizing QM/MM methodologies on B3LYP/def-TZVP level. To benchmark these results, starting- and endpoint as well as the transition state of the proton transfer have been benchmarked with different wave function-based or DFT approaches. Additionally, the basis set was increased to aug-cc-pVDZ to achieve a better description of negative charges. In comparison to CCSD(T) spin component scaled methods perform best and are always superior to DFT approaches. Those results reveal barriers in a range of $\approx$ 10 to 20 kcal/mol while the state after proton transfer was less stable by values of $\approx$ 4 to 15 kcal/mol. If one compares FadA5 and cruzain, the energy values obtained from the latter are always lower in terms of the height of barrier or the proton transfer reaction energy.
To gain further insight on the influence of the protein environment on the proton transfer the QM region of starting- and endpoint and also the transition state (from the neutral state) were inserted in the "wrong" environment, i.e. in the environment of the zwitterionic protonation state and vice versa. By doing so, a decrease in the height of barrier and furthermore, a stabilization of the state after proton transfer is observable. Especially in case of protonation state II (zwitterionic) of FadA5 no barrier was present and the neutral state after proton transfer is about 20 kcal/mol more stable. These results show a large influence of the protein surrounding in respect to the protonation states. This means that the change in environment leads to the opposite description concerning the preference of the protonation state.
Furthermore, dynamics of the protein environment relaxation regarding the protonation states have been investigated. It is understandable that the proton transfer itself is a fast process. But how long does the protein environment need to adapt to the change of protonation state? This question was tackled by classic MD simulations explained in the following: The simulations included constrained atoms of the active site and considered both environments to allow an adaptation of the environment while fixing atoms of the QM region. By doing so, a degree of comparability was ensured. It was observable that the relaxation of the environment requires at least 1 ns but can also have a duration of 6 ns. Still, it has to be mentioned that these relaxed structures mostly reveal RMSD values of 1.6 Å in the

2d-RMSD plots. Therefore, it is hard to say how much time is exactly needed for a relaxation of the environment to the respective protonation state. As mentioned above, however, the surrounding requires at least 1 ns for relaxation towards the respective protonation state and therefore revealing a qualitative answer.

Chapter 14

Acknowledgment

Zu allererst möchte ich meinen Dank meinem Doktorvater Prof. Dr. Bernd Engels aussprechen, der mir die Möglichkeit eröffnete in diesem spannenden Themengebiet zu forschen und mich in jeder Hinsicht unterstützt hat. Er räumte mir viele Freiheiten in der Forschung ein, sodass ich auch eigenen Ideen nachgehen konnte und mich so weiterentwickeln konnte. Auch seien hierbei seine Gartenfeiern zu erwähnen, die mir die rheinländische Kultur näherbrachten.
Ebenfalls gebührt der Dank dem wunderbaren Arbeitskreis, der durch hitzige und witzige Diskussionen (welche auch über die theoretische Chemie hinausgingen) ein angenehmes Arbeitsklima schaffte. Zuerst muss Ursula Rüppel genannt werden, die stets mit Rat und Tat zur Seite stand und immer ein offenes Ohr hatte.
Der Dank geht sowohl an den „alten Kern“ aber auch an die jetzigen Generation. Vor allem sollen Dr. Daniel Bellinger, Anna Heilos, Eileen Welz, Dustin Kaiser, Sara Wirsing, Dr. Johannes Becker und Daniel Weber genannt werden.
Vor allem dem besten Freund Dr. Daniel Bellinger sollen weitere Zeilen gewidmet werden. Danke für den initialen Gedanken in der Theoretischen Chemie zu promovieren, für das Korrekturlesen, für die Diskussionen im Studium, als Kollegen im Büro oder auch nach Feierabend in der „Reue“ bei 3-3.
Der Dank geht auch an Sarah Braunmiller, welche sich ebenfalls die Zeit nahm meine Arbeit Korrektur zu lesen.
Ebenfalls bedanke ich mich bei Prof. Dr. Volker Engel für seine Bereitschaft Zweitgutachter dieser Arbeit zu sein. Auch möchte ich mich recht herzlich bei den Kooperationspartner bedanken. Hierbei sind Prof. Dr. Kisker mit Dr. Sandra Eltschkner, Prof. Dr. Seibel mit Dr. Maria-Elena Ortiz und Julia Ertl, Prof. Dr. Ute Hellmich und Annika Wagner, Prof. Dr. Daniel Rauh, Prof. Dr. Holzgrabe und Antonio Ferraro, Prof. Dr. Sotriffer mit Dr.

Yogesh Narkhede und Dr. Maximilian Kuhn und Prof. Dr. Schirmeister zu nennen.
Zu guter Letzt möchte ich mich bei meiner Familie bedanken, die mich stets bedingungslos unterstützt haben, bei meinem Vater Le Van Yen, der mich schon früh für Computer begeistert hat, bei meiner Mutter Le Pham Thi Thanh, die mich mit Carepaketen versorgt hatte, bei meinen Brüdern Le Thanh Lac Tran, Le Yen Dinh Nguyen und Dr. Le Pham Hai Dang und bei meiner Schwägerin Miriam Le. Insbesondere die zwei Kleinen sollen genannt werden, meine Nichte Margaretha Tuong Vi Le und mein Neffe Elia Van Nhan Le, die mit nur einem Lächeln alle Sorgen vergessen lassen.

Chapter 15

References

[1] Dubey, *Advanced Biotechnology*, S Chand Co Ltd, **2014**.

[2] E. Buchner, *Berichte der Deutschen Chemischen Gesellschaft* **1897**, *30*, 117–124.

[3] J. H. Norhtrop, *Biological Reviews* **1935**, *10*, 263–282.

[4] L. . Johnson, D. C. Phillips, *Nature* **1965**, *206*, 761–763.

[5] J. M. Berg, J. L. Tymoczko, L. Stryer, *Stryer Biochemie*, Springer Spektrum, **2014**.

[6] S. Masamune, M. A. J. Palmer, R. Gamboni, S. Thompson, J. T. Davis, S. F. Williams, O. P. Peoples, A. J. Sinskey, C. T. Walsh, *Journal of the American Chemical Society* **1989**, *111*, 1879–1881.

[7] H. F. Gilbert, B. J. Lennox, C. D. Mossman, W. C. Carle, *Journal of Biological Chemistry* **1981**, *256*, 7371–7377.

[8] M. Mathieu, Y. Modis, J. P. Zeelen, C. K. Engel, R. A. Abagyan, A. Ahlberg, B. Rasmussen, V. S. Lamzin, W. H. Kunau, R. K. Wierenga, *Journal of Molecular Biology* **1997**, *273*, 714–728.

[9] V. Germain, E. L. Rylott, T. R. Larson, S. M. Sherson, N. Bechtold, J. P. Carde, J. H. Bryce, I. A. Graham, S. M. Smith, *The Plant Journal: For Cell and Molecular Biology* **2001**, *28*, 1–12.

[10] A. Birkenmaier, H. M. Möller, B. Philipp, *FEMS Microbiology Letters* **2011**, *318*, 123–130.

[11] L. Sellés Vidal, C. L. Kelly, P. M. Mordaka, J. T. Heap, *Biochimica et Biophysika acta* **2018**, *1866*, 327–347.

[12] E. J. Toone, *Advances in Enzymology and Related Areas of Molecular Biology: Protein Evolution*, Wiley-Interscience, **2006**.

[13] N. C. Price, L. Stevens, *Fundamentals of Enzymology: The Cell and Molecular Biology of Catalytic Proteins*, USA: Oxford University Press, **1999**.

[14] J. W. DePierre, L. Ernster, *Annual Review of Biochemistry* **1977**, *46*, 201–262.

[15] M. Kappor, J. Gopalakrishnapai, N. Surolia, A. Surolia, *Biochemical Journal* **2004**, *381*, 735–741.

[16] L. Michaelis, M. L. Menten, *Biochemische Zeitschrift* **1913**, *49*, 333–369.

[17] X. Morelli, R. Bourgeas, P. Roche, *Current Opinion in Chemical Biology* **2011**, *15*, 475–481.

[18] A. U. Rehmann, S. Saud, N. Ahmad, A. Wadood, R. Hamid, *Current Trends in Biomedical Engineering Biosciences* **2017**, *4*.

[19] J. M. Strelow, *SLAS Discovery* **2017**, *22*, 3–20.

[20] D. Leung, G. Abbenante, D. P. Fairlie, *Journal of Medicinal Chemistry* **2000**, *43*, 305–341.

[21] A. Tuley, W. Fast, *Biochemistry* **2018**, *57*, 3326–3337.

[22] M. H. Johansson, *Mini Revies in Medicinal Chemistry* **2012**, *12*, 1330–1344.

[23] M. Groll, C. R. Berkers, H. L. Ploegh, H. Ovaa, *Structure* **2006**, *14*, 451–456.

[24] S. A. Thompson, P. R. Andres, R. P. Hanzlik, *Journal of Medicinal Chemistry* **1986**, *29*, 104–111.

[25] S. Amslinger, *ChemMedChem* **2010**, *5*, 351–356.

[26] F. Jensen, *Introduction to Computational Chemistry Second Edition*, Wiley, **2006**.

[27] T. Schlick, *Reviews in Computational Chemistry* **1992**, *3*.

[28] M. L. McKee, M. Page, *Reviews in Computational Chemistry* **1993**, *4*, 35–65.

[29] E. London, *Zeitschrift für Physik* **1930**, *63*, 245–279.

[30] M. Karplus, G. A. Petsko, *Nature* **1990**, *347*, 631–639.

[31] H. F. Stillinger, A. Rahman, *The Journal of Chemical Physics* **1974**, *60*, 1545–1557.

[32] V. N. Maiorov, G. M. Crippen, *Journal of Molecular Biology* **1994**, *235*, 625–634.

[33] B. J. McConkey, V. Sobolev, M. Edelman, *Current Science* **2002**, *83*, 845–856.

[34] I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, T. E. Ferrin, *Journal of Molecular Biology* **1982**, *161*, 269–288.

[35] E. Fischer, *Berichte der Deutschen Chemischen Gesellschaft* **1894**, *27*, 2985–2993.

[36] D. E. Koshland Jr, *Angewandte Chemie International Edition English* **1994**, *33*, 2375–2378.

[37] N. Moitessier, P. Englebienne, D. Lee, J. Lawandi, C. R. Corbeil, *British Journal of Pharmacology* **2008**, *153*, 7–26.

[38] A. T. Brint, P. Willett, *Journal of Chemical Information and Computer Sciences* **1987**, *27*, 152–158.

[39] R. Norel, D. Fischer, H. J. Wolfson, R. Nussinov, *Protein engineering* **1994**, *7*, 39–46.

[40] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, *Journal of Molecular Biology* **1996**, *261*, 470–489.

[41] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson, *Journal of Computational Chemistry* **1999**, *19*, 1639–1662.

[42] G. Jones, P. Willett, R. C. Glen, A. R. Leach, R. Taylor, *Journal of Molecular Biology* **1997**, *267*, 727–748.

[43] C. M. Oshiro, I. D. Kuntz, J. S. Dixon, *Journal of Computer-Aided Molecular Design* **1995**, *9*, 113–130.

[44] O. Korb, T. Stützle, T. E. Exner, *Journal of Chemical Information and Modeling* **2009**, *49*, 84–96.

[45] J. Srinivasan, T. E. Cheatham, P. Cieplak, P. A. Kollmann, D. A. Case, *Journal of the American Chemical Society* **1998**, *120*, 9401–9409.

[46] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, T. E. Cheatham 3rd, *Accounts of Chemical Research* **2000**, *33*, 889–897.

[47] S. Genheden, U. Ryde, *Expert Opinion on Drug Discovery* **2015**, *10*, 449–461.

[48] J. B. Taylor, D. J. Triggle, *Comprehensive Medicinal Chemistry II Volume 4*, **2007**.

[49] B. Honig, K. Sharp, A. S. Yang, *The Journal of Physical Chemistry* **1993**, *97*, 1101–1109.

[50] J. Warwicker, H. C. Watson, *Journal of Molecular Biology* **1982**, *157*, 671–679.

[51] K. A. Sharp, B. Honig, *Annual Reviews of Biophysics and Biophysical Chemistry* **1990**, *19*, 301–332.

[52] W. C. Still, A. Tempczyk, R. C. Hawley, T. Hendrickson, *Journal of the American Chemical Society* **1990**, *112*, 6127–6129.

[53] M. Shirts, D. L. Mobley, J. D. Chodera, *Annual Reports in Computational Chemistry* **2007**, *3*, 41–59.

[54] A. Szabo, N. S. Ostlund, *Modern Quantum Chemistry Introduction to Advanced Electronic Structure Theory*, Dover, **1982**.

[55] C. Møller, M. S. Plesset, *Physical Review* **1934**, *46*, 618–622.

[56] G. D. Purvis III, R. J. Bartlett, *The Journal of Chemical Physics* **1982**, *76*, 1910–1918.

[57] J. Noga, R. J. Bartlett, *The Journal of Chemical Physics* **1987**, *86*, 7041–7050.

[58] G. E. Scuseria, T. J. Lee, *The Journal of Chemical Physics* **1990**, *93*, 5851–5855.

[59] K. Raghavachari, G. W. Trucks, J. A. Pople, M. Head-Gordon, *Chemical Physics Letters* **1989**, *157*, 479–483.

[60] T. Helgaker, P. Jorgensen, J. Olsen, *Molecular Electronic-Structure Theory*, **2000**.

[61] O. Christiansen, H. Koch, P. Jørgensen, *Chemical Physical Letters* **1995**, *243*, 409–418.

[62] T. B. Adler, G. Knizia, H. J. Werner, *The Journal of Chemical Physics* **2007**, *127*.

[63] T. B. Adler, H. J. Werner, *The Journal of Chemical Physics* **2011**, *135*.

[64] S. Grimme, *The Journal of Chemical Physics* **2003**, *118*, 9095–9102.

[65] A. Hellweg, S. A. Grün, C. Hättig, *Physical Chemistry Chemical Physics* **2008**, *10*, 4119–4127.

[66] P. Hohenberg, W. Kohn, *Physical Review* **1964**, *136*, 864–871.

[67] R. G. Parr, W. Yang, *Density Functional Theory*, Oxford University Press, **1989**.

[68] E. J. Baerends, O. V. Gritsenko, *The Journal of Physical Chemistry A* **1997**, *101*, 5383–5403.

[69] W. Kohn, L. J. Sham, *Physical Reviews* **1965**, *140*, A1133–A1138.

[70] J. P. Perdew, K. Schmidt, *AIP Conference Proceedings* **2001**, *577*.

[71] T. Tsuneda, *Density Functional Theory in Quantum Chemistry*, **2014**.

[72] A. D. Becke, *Physical Review A. General Physics.* **1988**, *38*, 3098–3100.

[73] N. Mardirossian, M. Head-Gordon, *Molecular Physics* **2017**, *115*, 2315–2372.

[74] A. D. Becke, *The Journal of Chemical Physics* **1993**, *98*, 1372–1377.

[75] A. D. Becke, *The Journal of Chemical Physics* **1993**, *98*, 5648–5652.

[76] C. Lee, W. Yang, R. G. Parr, *Physical Review B* **1988**, *37*, 785–789.

[77] Y. Zhao, D. G. Truhlar, *The Journal of Chemical Physics* **2006**, *125*, 194101.

[78] Y. Zhao, D. G. Truhlar, *Theoretical Chemicstry Accounts* **2007**, *120*, 215–241.

[79] Y. Zhho, D. G. Truhlar, *Journal of Chemical Theory and Computation* **2008**, *4*, 1849–1868.

[80] R. Peverati, D. G. Truhlar, *The Journal of Physical Chemistry Letters* **2011**, *2*, 2810–2817.

[81] S. Grimme, *Journal of Computational Chemistry* **2004**, *25*, 1463–1473.

[82] S. Grimme, J. Antony, T. Schwabe, C. Mück-Lichtenfeld, *Organic Biomolecular Chemistry* **2007**, *5*, 741–758.

[83] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *The Journal of Chemical Physics* **2010**, *132*, 154104.

[84] M. J. Allen, D. J. Tozer, *The Journal of Chemical Physics* **2002**, *117*, 11113.

[85] U. Zimmerli, M. Parrinello, P. Koumoutsakos, *The Journal of Chemical Physics* **2004**, *120*, 2693–2699.

[86] S. Tsuzuki, *The Journal of Chemical Physics* **2001**, *114*, 3949–3957.

[87] L. Goerigk, S. Grimme, *Physical Chemistry Chemical Physics* **2011**, *13*, 6670–6688.

[88] T. Risthaus, S. Grimme, *Journal of Chemical Theory and Computation* **2013**, *9*, 1580–1591.

[89] A. Kovács, J. C. Dobrowolski, S. Ostrowski, J. E. Rode, *International Journal of Quantum Chemistry* **2017**, *117*.

[90] L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, S. Grimme, *Physical Chemistry Chemical Physics* **2017**, *19*, 32184–32215.

[91] J. Tomasi, B. Mennucci, R. Cammi, *Chemical Reviews* **2005**, *105*, 2999–3094.

[92] A. Klamt, G. Schüürmann, *Journal of the Chemical Society, Perkin Transactions 2* **1993**, *5*, 799–805.

[93] M. Levitt, A. Warshel, *Nature* **1975**, *253*, 694–698.

[94] M. J. Field, P. A. Bash, M. Karplus, *Journal of Computational Chemistry* **1990**, *11*, 700–733.

[95] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, *Journal of Computational Chemistry* **1983**, *4*, 187–217.

[96] H. M. Senn, W. Thiel, *Angewandte Chemie (International Edition in English)* **2009**, *48*, 1198–1229.

[97] U. Ryde, *Journal of Computer-Aided Molecular Design* **1996**, *10*, 153–164.

[98] P. Sherwood, A. H. de Vries, S. J. Collins, S. P. Greatbanks, N. A. Burton, M. A. Vincent, I. H. Hillier, *Faraday Discussions* **1997**, *106*, 79–92.

[99] D. A. Case, V. Babin, J. T. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, T. E. Cheatham, T. A. Darden, R. E. Duke, H. Gohlke, A. W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T. S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, Walker, J. Wang, R. M. Wolf, X. Wu, P. A. Kollman, *Amber 14*, University of California, San Francisco, **2014**.

[100] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, *The Journal of Chemical Theory and Computation* **2015**, *11*, 3696–3713.

[101] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *Journal of Computational Chemistry* **2004**, *25*, 1157–1174.

[102] J. Wang, W. Wang, P. A. Kollman, D. A. Case, *Journal of Molecular Graphics and Modelling* **2006**, *25*, 247–260.

[103] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, *Gaussian 03, Revision C.02*, Gaussian, Inc., Wallingford, CT, 2004.

[104] R. Ditchfield, W. J. Hehre, J. A. Pople, *Journal of Chemical Physics* **1971**, *54*, 724–728.

[105] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, *The Journal of Chemical Physics* **1983**, *79*, 926–935.

[106] R. Ahlrichs, M. Bär, M. Häser, H. Horn, C. Kölmel, *Chemical Physics Letters* **1989**, *162*, 165–169.

[107] O. Treutler, R. Ahlrichs, *The Journal of Chemical Physics* **1995**, *102*, 346–354.

[108] M. Von Arnim, R. Ahlrichs, *The Journal of Chemical Physics* **1998**, *16*, 1746–1757.

[109] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, . Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, *Gaussian09 Revision E.01*, Gaussian Inc. Wallingford CT 2009.

[110] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, G. R. Hutchison, *Journal of Cheminformatics* **2012**, *4:17*.

[111] C. Grebner, J. Becker, D. Weber, D. Bellinger, M. Tafipolski, C. Brückner, B. Engels, *Journal of Computational Chemistry* **2014**, *35*, 1801–1807.

[112] J. J. P. Stewart, *Stewart Computational chemistry, web: http://openmopac.net*.

[113] *ChemShell, a Computational Chemistry Shell, see www.chemshell.org*.

[114] P. Sherwood, A. H. de Vries, M. F. Guest, G. Schreckenbach, C. A. Catlow, S. A. French, A. A. Sokol, S. T. Bromley, W. Thiel, A. J. Turner, S. Billeter, F. Terstegen, S. Thiel, J. Kendrick, S. C. Rogers, J. Casci, M. Watson, F. King, E. Karlsen, M. Sjøvoll, A. Fahmi, A. Schäfer, C. Lennartz, *Journal of Molecular Structure: THEOCHEM* **2003**, *632*, 1 – 28.

[115] J. Köstner, J. M. Carr, T. W. Keal, W. Thiel, A. Wander, P. Sherwood, *The Journal of Physical Chemistry A* **2009**, *113*, 11856–11865.

[116] *Origin (OriginLab, Northampton, MA).*

[117] W. Humphrey, A. Dalke, K. Schulten, *Journal of Molecular Graphics* **1996**, *14*, 33–38.

[118] O. C. Olson, J. A. Joyce, *Nature reviews. Cancer.* **2015**, *15*, 712–729.

[119] N. D. Rawlings, F. R. Morton, A. J. Barrett, *Nucleic Acids Research* **2006**, *34*, D270–D272.

[120] A. K. Clark, P. K. Yip, J. Grist, C. Gentry, A. A. Staniland, F. Marchand, M. Dehvari, G. Wotherspoon, J. Winter, J. Ullah, S. Bevan, M. Malcangio, *Proceedings of the National Academy of Scineces of the United States of America* **2007**, *104*, 10655–10660.

[121] M. Yin, J. Soikkeli, T. Jahkola, S. Virolainen, O. Saksela, E. Hölttä, *The American Journal of Pathology.* **2012**, *181*, 2202–2216.

[122] G. Devanathan, J. L. Turnbull, E. Ziomek, E. O. Purisima, R. Ménard, T. Sulea, *Biochemical and Biophysical Research Communications.* **2005**, *329*, 445–452.

[123] D. K. Nägler, R. Zhang, W. Tam, T. Sulea, E. O. Purisima, R. Ménard, *Biochemistry* **1999**, *38*, 12648–12654.

[124] M. M. Mohamed, B. F. Sloane, *Nature reviews. Cancer* **2006**, *6*, 764–775.

[125] B. F. Sloane, J. R. Dunn, K. V. Honn, *Science* **1981**, *212*, 1151–1153.

[126] I. Berdowska, *Clinica Chimica Acta; International Journal Of* $C^{l}inicalChemistry\boldsymbol{2004}, 342, 41 - -69.$

[127] C. Jedeszko, B. F. Sloane, *Biological Chemistry* **2004**, *385*, 1017–1027.

[128] N. Guinec, V. Dalet-Fumeron, M. Pagano, *Biological Chemistry Hoppe-Seyler* **1993**, *374*, 1135–1146.

[129] G. Bruchard, M. Mignot, V. Derangère, F. Chalmin, A. Chevriaux, F. Végran, W. Boireau, B. Simon, B. Ryffel, J. L. Connat, J. Kanellopoulos, F. Martin, C. Rébé, L. Apetoh, F. Ghiringhelli, *Nature medicine* **2013**, *19*, 57–64.

[130] T. Shree, O. C. Olson, B. T. Elie, J. C. Kester, A. L. Garfall, K. Simpson, K. M. Bell-McGuinn, E. C. Zabor, E. Brogi, J. A. Joyce, *Genes Development* **2011**, *25*, 2465–2479.

[131] M. Sajid, J. H. McKerrow, *Molecular and Biochemical Parasitology* **2002**, *120*, 1–21.

[132] G. Simmons, D. N. Gosalia, A. J. Rennekamp, J. D. Reeves, S. L. Diamond, P. Bates, *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 11876–11881.

[133] S. L. Martin, K. L. Moffitt, A. McDowell, C. Greenan, R. J. Bright-Thomas, A. M. Jones, A. K. Webb, J. S. Elborn, *Pediatric Pulmonology.* **2010**, *45*, 860–868.

[134] S. P. Lutgens, K. B. Cleutjens, M. J. Daemen, S. Heeneman, *FASEB journal: official publication of the Federation of American Societies for Experimental Biology.* **2007**, *21*, 3029–3041.

[135] R. A. Nixon, A. M. Cataldo, P. M. Mathews, *Neurochemical Research* **2000**, *25*, 1161–1172.

[136] Y. Yasuda, J. Kaleta, D. Brömme, *Advanced Drug Delivery Reviews.* **2005**, *57*, 973–993.

[137] J. T. Palmer, D. Rasnick, J. L. Klaus, D. Bromme, *Journal of Medicinal Chemistry* **1995**, *38*, 3193–3196.

[138] R. Ettari, E. Nizi, M. E. D. F. Francesco, G. Dude MA, Pradel, R. Vičík, T. Schirmeister, N. Micale, S. Grasso, M. Zappalà, *Journal of Medicinal Chemistry* **2008**, *51*, 988–996.

[139] R. Ettari, C. Bonaccorso, N. Micale, C. Heindl, T. Schirmeister, M. L. Calabro, S. Grasso, M. Zappalà, *ChemMedChem* **2011**, *6*, 1228–1237.

[140] S. Previti, R. Ettari, S. Cosconati, G. Amendola, K. Chouchene, A. Wagner, U. A. Hellmich, K. Ulrich, R. L. Krauth-Siegel, P. R. Wich, I. Schmid, T. Schirmeister, J. Gut, P. J. Rosenthal, S. Grasso, M. Zappalà, *Journal of Medicinal Chemistry* **2017**, *60*, 6911–6923.

[141] A. Breuning, B. Degel, F. Schulz, C. Büchold, M. Stempka, U. Machon, S. Heppner, C. Gelhaus, M. Leippe, M. Leyh, C. Kisker, J. Rath, A. Stich, J. Gut, P. J. Rosenthal, C. Schmuck, T. Schirmeister, *Journal of Medicinal Chemistry* **2010**, *53*, 1951–1963.

[142] C. Machon, U.and Büchold, M. Stempka, T. Schirmeister, C. Gelhaus, M. Leippe, J. Gut, P. J. Rosenthal, C. Kisker, M. Leyh, C. Schmuck, *Journal of Medicinal Chemistry* **2009**, *52*, 5662–5672.

[143] D. A. Shannon, R. Banerjee, E. R. Webster, D. W. Bak, C. Wang, E. Weerapana, *Journal of the Amercian Chemical Society* **2014**, *136*, 3330–3333.

[144] L. Leesnitzer, D. J. Parks, R. K. Bledsoe, J. E. Cobb, J. L. Collins, T. G. Consler, R. G. Davis, E. A. Hull-Ryde, J. M. Lenhard, L. Patel, K. D. Plunket, J. L. Shenk, J. B. Stimmel, C. Therapontos, T. M. Willson, S. G. Blanchard, *Biochemistry* **2002**, *41*, 6640–6650.

[145] R. Vicik, M. Busemann, C. Gelhaus, N. Stiefl, J. Scheiber, W. Schmitz, F. Schulz, M. Mladenovic, B. Engels, M. Leippe, K. Baumann, T. Schirmeister, *ChemMedChem* **2006**, *1*, 1126–1141.

[146] R. L. Lundblad, *Chemical Reagents for Protein Modification*, CRC Press, **2014**.

[147] J. D. Gregory, *Journal of American Chemical Society* **1955**, *77*, 3922–3923.

[148] K. A. Schäfer, C. Huber, R. Ahlrichs, *The Journal of Chemical Physics* **1994**, *100*, 5829–5835.

[149] R. A. Kendall, T. H. Dunning, R. J. Harrison, *The Journal of Chemical Physics* **1992**, *96*, 6796–6806.

[150] T. Yanai, D. P. Tew, N. C. Handy, *Chemical Physics Letters* **2004**, *393*, 51–57.

[151] J.-D. Chai, M. Head-Gordon, *The Journal of CHemical Physics* **2008**, *128*, 084106.

[152] R. Brun, J. Blum, F. Chappuis, C. Burri, *Lancet* **2010**, *375*, 148–159.

[153] K. Stuart, R. Brun, S. Croft, A. Fairlamb, R. E. Gürtler, J. McKerrow, S. Reed, R. Tarleton, *The Journal of Clinical Investigation* **2008**, *118*, 1301–1310.

[154] E. Gluenz, M. L. Povelones, P. T. Englund, K. Gull, *Moelcuar and Cellular Biology* **2011**, *31*, 1012–1021.

[155] J. Melchers, M. Diechtierow, K. Fehér, *The Journal of Biological Chemistry* **2008**, *283*, 30401–30411.

[156] A. H. Farilamb, A. Cerami, *Annual Review of Microbiology* **1992**, *46*, 695–729.

[157] R. L. Kraut-Siegel, H. Bauer, R. H. Schirmer, *Angewandte Chemie Internation Edition* **2005**, *44*, 690–715.

[158] M. Dormeyer, N. Reckenfelderbäumer, H. Ludemann, R. L. Krauth-Siegel, *The Journal of Biological Chemistry* **2001**, *276*, 10602–10606.

[159] F. Fueller, B. Jehle, K. Putzker, J. D. Lewis, R. L. Kraut-Siegel, *The Journal of Biological Chemistry* **2012**, *287*, 8792–8802.

[160] D. M. Spencer, T. J. Wandless, S. L. Schreiber, G. R. Crabtree, *Science* **1993**, *262*, 1019–1024.

[161] V. M. Rivera, T. Clackson, S. Natesan, R. Pollock, J. F. Amara, T. Keenan, S. R. Magari, T. Phillips, N. L. Courage, F. Cerasoli Jr, D. A. Holt, M. Gilman, V. M. Rivera, T. Clackson, S. Natesan, R. Pollock, J. F. Amara, T. Keenan, S. R. Magari, T. Phillips, N. L. Courage, F. Cerasoli Jr, D. A. Holt, *Nature Medicine* **1996**, *2*, 1028–1032.

[162] M. A. Farrar, J. Alberol-Ila, R. M. Perlmutter, *Nature* **1996**, *383*, 178–181.

[163] T. Miyamoto, R. DeRose, A. Suarez, M. Wittwer, B. Ernst, E. Constable, M. J. Wolfgang, C. Mukherjee, D. J. Meyers, T. Inoue, *Nature Chemical Biology* **2012**, *8*, 465–470.

[164] S. Gehrig, T. Efferth, *International Journal of Molecular Medicine* **2008**, *22*, 411–419.

[165] WHO **2006**, *Fact sheet No. 259*.

[166] H. B. Tanowitz, M. Wittner, C. Werner, L. M. Weiss, L. V. Kirchhoff, C. Bacchi, *Encyclopedia of Microbiology. Lederberg J (ed).* **2000**, *4*, 725–741.

[167] V. Hemmige, H. Tanowitz, A. Sethi, *International Journal of Dermatology* **2012**, *51*, 501–508.

[168] A. M. Sartori, M. N. Sotto, L. M. Braz, C. Oliveira Júnior Oda, R. A. Patzina, A. A. Barone, M. A. Shikanai-Yasuda, *Transactions of the Royal Society of Tropical Medicine and Hygiene* **1999**, *93*, 631–632.

[169] N. Añez, H. Carrasco, H. Parada, G. Crisante, A. Rojas, N. Gonzalez, J. L. Ramirez, P. Guevara, C. Rivero, R. Borges, J. V. Scorza, *The American Journal of Tropical Medicine and Hygiene* **1999**, *60*, 215–222.

[170] C. W. Expert **2002**, *Brazil: World Health Organization*.

[171] A. Biolo, A. L. Ribeiro, N. Clausell, *Progress in cardiovascular diseases.* **2010**, *52*, 300–316.

[172] J. C. Engel, P. S. Doyle, I. Hsieh, J. H. McKerrow, *The Journal of Experimental Medicine* **1998**, *188*, 725–734.

[173] L. V. Kirchhoff, *Infectious Disease Clinics of North America* **1993**, *7*, 487–502.

[174] P. S. Doyle, Y. M. Zhou, J. C. Engel, J. H. McKerrow, *Antimicrobial Agents and Chemotherapy* **2007**, *51*, 3932–3939.

[175] S. Barr, K. Warner, B. Kornreic, J. Piscitelli, A. Wolfe, L. Benet, J. H. McKerrow, *Antimicrobial Agents and Chemotherapy* **2005**, *49*, 5160–5161.

[176] C. Bryant, I. D. Kerr, M. Debnath, K. K. Ang, J. Ratnam, R. S. Ferreira, P. Jaishankar, D. Zhao, M. R. Arkin, J. H. McKerrow, L. S. Brinen, A. R. Renslo, *Bioorganic Medicinal Chemistry Letters* **2009**, *19*, 6218–6221.

[177] *Molecular Operating Environment (MOE)*, **2018**, Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite 910, Montreal, QC, Canada, H3A 2R7.

[178] H. Gohlke, M. Hendlich, G. Klebe, *Journal of Molecular Biology* **2000**, *295*, 337–356.

[179] S. P. Staal, *Proceedings of the National Academy of Sciences of the United States of America* **1987**, *84*, 5034–5037.

[180] A. Bellacosa, J. R. Testa, S. P. Staal, P. N. Tsichlis, *Science* **1991**, *254*, 247–277.

[181] P. F. Jones, T. Jakubowicz, F. J. Pitossi, F. Maurer, B. A. Hemmings, *Proceedings of the National Academy of Sciences of the United States of America* **1991**, *88*, 4171–4175.

[182] J. R. Coffer, P. J. Woodgett, *European Journal of Biochemistry* **1991**, *201*, 475–481.

[183] B. D. Manning, A. Toker, *Cell* **2017**, *169*, 381–405.

[184] L. C. Cantley, *Harvey Lectures* **2004**, 103–122.

[185] P. S. Mundi, J. Sachdev, C. McCourt, K. Kalinsky, *British Journal of Clinical Pharmacology* **2016**, *82*, 943–956.

[186] I. Vivanco, C. L. Sawyers, *Nature Reviews. Cancer* **2002**, *2*, 489–501.

[187] J. A. Fresno Vara, E. Casado, J. de Castro, P. Cejas, C. Belda-Iniesta, M. González-Barón, *Cancer treatment reviews* **2004**, *30*, 193–204.

[188] B. Vanhaesebroeck, J. Guillermet-Guibert, M. Graupera, B. Bilanges, *Nature reviews. Molecular Cell Biology* **2010**, *11*, 329–341.

[189] B. J. Mayer, R. Ren, K. L. Clark, D. Baltimore, *Cell* **1993**, *73*, 629–30.

[190] R. J. Haslam, H. B. Koide, B. A. Hemmings, *Nature* **1993**, *363*, 309–310.

[191] A. Musacchio, T. Gibson, J. Rice, P.and Thompson, M. Saraste, *Trends in Biochemical Sciences* **1993**, *18*, 343–348.

[192] T. J. Gibson, M. Hyvönen, A. Musacchio, M. Saraste, E. Birney, *Trends in Biochemical Sciences* **1994**, *19*, 349–353.

[193] E. Ingley, B. A. Hemmings, *Journal of Cellular Biochemistry* **1994**, *56*, 436–443.

[194] T. Pawson, *Nature* **1995**, *373*, 573–580.

[195] M. Saraste, M. Hyvönen, *Current Opinion in Structural Biology* **1995**, *5*, 403–408.

[196] T. F. Franke, D. R. Kaplan, L. C. Cantley, A. Toker, *Science* **1997**, *275*, 665–668.

[197] A. Guilherme, J. K. Klarlund, G. Krystal, M. P. Czech, *The Journal of Biological Chemistry* **1996**, *271*, 29533–29536.

[198] D. R. Alessi, M. Andjelkovic, B. Caudwell, P. Cron, N. Morrice, P. Cohen, B. A. Hemmings, *The EMBO Journal* **1996**, *15*, 6541–6551.

[199] D. R. Alessi, S. R. James, C. P. Downes, A. B. Holmes, P. Gaffney, C. B. Reese, P. Cohen, *Current Biology* **1997**, *7*, 261–269.

[200] D. Stokoe, L. R. Stephens, T. Copeland, P. R. Gaffney, C. . B. Reese, G. F. Painter, A. B. Holmes, F. McCormick, P. T. Hawkins, *Science* **1997**, *277*, 567–570.

[201] J. Yang, P. Cron, V. M. Good, V. Thompson, B. A. Hemmings, D. Barford, *Nature Structural Biology* **2002**, *9*, 940–944.

[202] M. Andjelković, T. Jakubowicz, P. Cron, X. F. Ming, J. W. Han, B. A. Hemmings, *Proceedings of the National Academy of Sciences of the United States of America* **1996**, *93*, 5699–5704.

[203] T. Gao, F. Furnari, A. C. Newton, *Molecular Cell* **2005**, *18*, 13–24.

[204] J. Brognard, E. Sierecki, T. Gao, A. C. Newton, *Molecular Cell* **2007**, *25*, 917–931.

[205] P. J. Parker, F. B. Caudwell, P. Cohen, *European Journal of Biochemistry* **1983**, *130*, 227–234.

[206] D. B. Rylatt, A. Aitken, T. Bilham, G. D. Condon, N. Embi, P. Cohen, *European Journal of Biochemistry* **1980**, *107*, 529–537.

[207] K. E. van der Vos, P. J. Coffer, *Antioxidants Redox Signaling* **2011**, *14*, 579–592.

[208] A. E. Webb, A. Brunet, *Trends in Biochemical Sciences* **2014**, *39*, 159–169.

[209] K. Inoki, Y. Li, T. Zhu, J. Wu, K. L. Guan, *Nature Cell Biology* **2002**, *4*, 648–657.

[210] B. D. Manning, A. R. Tee, M. N. Logsdon, J. Blenis, L. C. Cantley, *Molecular Cell* **2002**, *10*, 151–162.

[211] C. J. Potter, L. G. Pedraza, T. Xu, *Nature Cell biology* **2002**, *4*, 658–665.

[212] R. A. Saxton, D. M. Sabatini, *Cell* **2017**, *168*, 960–976.

[213] A. S. Clark, K. West, S. Streicher, P. A. Dennis, *Molecular Cancer Therapeutics* **2002**, *1*, 707–717.

[214] M. Rudolph, T. Anzeneder, A. Schulz, G. Beckmann, A. T. Byrne, M. Jeffers, C. Pena, O. Politz, K. Köchert, R. Vonk, J. Reischl, *BMC Cancer* **2016**, *16*, 1–12.

[215] M. Chen, C. P. Pratt, M. E. Zeeman, N. Schultz, B. S. Taylor, A. O'Neill, M. Castillo-Martin, D. G. Nowak, A. Naguib, D. M. Grace, J. Murn, N. Navin, G. S. Atwal, C. Sander, W. L. Gerald, C. Cordon-Cardo, A. C. Newton, B. S. Carver, L. C. Trotman, *Cancer Cell* **2011**, *20*, 173–186.

[216] J. Zhang, P. L. Yang, N. S. Gray, *Nature Reviews. Cancer* **2009**, *9*, 28–39.

[217] J. Lin, D. Sampath, M. A. Nannini, B. B. Lee, M. Degtyarev, J. Oeh, H. Savage, Z. Guan, R. Hong, R. Kassees, L. B. Lee, T. Risom, S. Gross, B. M. Liederer, H. Koeppen, N. J. Skelton, J. J. Wallin, M. Belvin, E. Punnoose, L. S. Friedman, K. Lin, *Clincal Cancer Research* **2013**, *19*, 1760–1772.

[218] R. L. B. Costa, H. S. Han, W. J. Gradishar, *Breast Cancer Research and Treatment* **2018**, *169*, 397–406.

[219] M. Dumble, M. C. Crouthamel, S. Y. Zhang, M. Schaber, D. Levy, K. Robell, Q. Liu, D. J. Figueroa, E. A. Minthorn, M. A. Seefeld, M. B. Rouse, S. K. Rabindran, D. A. Heerding, R. Kumar, *PloS one* **2014**, *9*, e100880.

[220] A. Spencer, S. S. Yoon, S. J. Harrison, S. R. Morris, D. A. Smith, R. A. Brigandi, J. Gauvin, R. Kumar, J. B. Opalinska, C. Chen, *Blood* **2014**, *124*, 2190–2195.

[221] A. W. Tolcher, A. Patnaik, K. P. Papadopoulos, D. W. Rasco, C. R. Becerra, A. J. Allred, K. Orford, G. Aktan, G. Ferron-Brady, N. Ibrahim, J. Gauvin, M. Motwani, M. Cornfeld, *Cancer Chemotherapy and Pharmacology* **2015**, *75*, 183–189.

[222] Z. Fang, C. Grütter, D. Rauh, *ACS Chemical Biology* **2013**, *8*, 58–70.

[223] J. R. Simard, M. Getlik, C. Grütter, R. Schneider, S. Wulfert, D. Rauh, *Journal of the American Chemical Society* **2010**, *132*, 4152–4160.

[224] V. Calleja, D. Alcor, M. Laguerre, J. Park, B. Vojnovic, B. A. Hemmings, J. Downward, P. J. Parker, B. Larijani, *PLoS biology* **2007**, *5*, e95.

[225] V. Calleja, M. Laguerre, P. J. Parker, B. Larijani, *PLoS Biology* **2009**, *7*, e17.

[226] S. F. Barnett, D. Defeo-Jones, S. Fu, P. J. Hancock, K. M. Haskell, R. E. Jones, J. A. Kahana, A. M. Kral, K. Leander, L. L. Lee, J. Malinowski, E. M. McAvoy, D. D. Nahas, R. G. Robinson, H. E. Huber, *The Biochemical Journal* **2005**, *385*, 399–408.

[227] J. Weisner, R. Gontla, L. van der Westhuizen, S. Oeck, J. Ketzer, P. Janning, A. Richters, T. Mühlenberg, Z. Fang, A. Taher, V. Jendrossek, S. C. Pelly, S. Bauer, W. A. van Otterlo, D. Rauh, *Angewandte Chemie (International ed. in English)* **2015**, *54*, 10313–10316.

[228] A. Fiser, R. K. Do, A. Sali, *Protein Sciencce* **2000**, *9*, 1753–1773.

[229] A. Sali, T. L. Blundell, *Journal of Molecular Biology* **1993**, *234*, 779–815.

[230] M. A. Martí-Renom, A. C. Stuart, A. Fiser, R. Sánchez, F. Melo, A. Sali, *Annual Review of Biophysics and Biomolecular Structures* **2000**, *29*, 291–325.

[231] R. I. Aminov, *Frontiers in Microbiology* **2010**, *1*, 1–5.

[232] E. P. Abraham, E. Chain, *Nature* **1940**, *146*, 837.

[233] J. Davies, D. Davies, *Microbiology and Molecular Biology Reviews* **2010**, *74*, 417–433.

[234] B. K. Tiwary, K. Pradhan, A. K. Nanda, R. Chakraborty, *Journal of Chemical Biology Therapeutics* **2015**, *1*, 104.

[235] S. B. Levy, B. Marshall, *Nature Medicine* **2004**, *10*, 122–129.

[236] F. R. DeLeo, H. F. Chambers, *The Journal of Clinical Investigation* **2009**, *119*, 2464–2474.

[237] S. B. Zaman, M. A. Hussain, R. Nye, V. Mehta, K. T. Mamun, N. Hossain, *Cureus* **2017**, *9*, 1403.

[238] D. M. Sievert, J. T. Rudrik, J. B. Patel, L. C. McDonald, M. J. Wilkins, J. C. Hageman, *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* **2008**, *46*, 668–674.

[239] G. W. H. Organization: **2014**.

[240] H. M. Lode, R. Stahlmann, *Arzneimitteltherapie* **2015**, *33*, 47–53.

[241] R. A. Copeland, D. L. Pompliano, T. D. Meek, *Nature Reviews Drug Discovery* **2006**, *5*, 730–739.

[242] K. P. Cusack, Y. Wang, M. Z. Hoemann, J. Marjanovic, R. G. Heym, A. Vasudevan, *Bioorganic Medicinal Chemistry Letters* **2015**, *25*, 2019–2027.

[243] H. Lu, K. England, C. am Ende, J. J. Truglio, S. Luckner, B. G. Reddy, N. L. Marlenee, S. E. Knudson, D. L. Knudson, R. A. Bowen, C. Kisker, R. A. Slayden, P. J. Tonge, *ACS Chemical Biology* **2009**, *4*, 221–231.

[244] D. J. Payne, P. V. Warren, D. J. Holmes, Y. Ji, J. T. Lonsdale, *Drug Discov Today* **2001**, *6*, 537–544.

[245] R. P. Massengo-Tiassé, J. E. Cronan, *Cellular and molecular life sciences: CMLS* **2009**, *66*, 1507–1517.

[246] T. L. P. Galvão, I. M. Rocha, M. D. M. C. Ribeiro da Silva, M. A. V. Ribeiro da Silva, *The Journal of Phycial Chemistry A* **2013**, *117*, 12668–12674.

[247] T. L. P. Galvão, I. M. Rocha, M. D. Ribeiro da Silva, M. A. Ribeiro da Silva, *The Journal of Physical Chemistry. A* **2014**, *118*, 3360–3366.

[248] C. López, R. M. Claramunt, I. Alkorta, J. Elguero, *Spectroscopy* **2000**, *14*, 121–126.

[249] N. Tsuchida, S. Yamabe, *The Journal of Phycial Chemistry A* **2005**, *109*, 1974–1980.

[250] WHO.

[251] E. W. Tiemersma, M. J. van der Werf, M. W. Borgdorff, B. G. Williams, N. J. Nagelkerke, *PLoS One* **2011**, *6*, e17601.

[252] J. Y. Lee, *Tuberculosis Respiratory Diseases* **2015**, *78*, 47–55.

[253] N. M. Nesbitt, X. Yang, P. Fontán, I. Kolesnikova, I. Smith, N. S. Sampson, E. Dubnau, *Infection and Immunity* **2010**, *78*, 275–282.

[254] P. J. Cardona, R. Llatjós, S. Gordillo, J. Díaz, I. Ojanguren, A. Ariza, V. Ausina, *Scandinavian Journal of Immunology* **2000**, *52*, 156–163.

[255] D. G. Russel, *Nature Reviews. Microbiology* **2007**, *5*, 39–47.

[256] P. Peyron, J. Vaubourgeix, Y. Poquet, F. Levillain, C. Botanch, F. Bardou, M. Daffé, J. F. Emile, B. Marchou, P. J. Cardona, C. de Chastellier, F. Altare, *PLoS Pathogens* **2008**, *4*, e1000204.

[257] F. R. Maxfield, D. Wüstner, *The Journal of Clinical Investigation* **2002**, *110*, 891–898.

[258] D. Wüstner, M. Mondal, I. Tabas, F. R. Maxfield, *Traffic (Copenhagen, Denmark)* **2005**, *6*, 396–412.

[259] I. G. Kurup, P. R. Mahadevan, *Journal of Biosciences* **1982**, *4*, 307–316.

[260] A. Pandey, C. M. Sassetti, *Proceedings of the National Academy of Sciences of the United States of America.* **2008**, *105*, 4376–4380.

[261] E. Dubnau, P. Fontán, R. Manganelli, S. Soares-Appel, I. Smith, *Infection and Immunity* **2002**, *70*.

[262] D. G. Russel, *Current Opinion in Microbiology* **2013**, *16*, 78–84.

[263] Y. Modis, R. K. Wierenga, *Journal of Molecular Biology* **2000**, *297*, 1171–1182.

[264] H. F. Gilbert, *Biochemistry* **1981**, *20*, 5643–5649.

[265] Y. Modis, R. K. Wierenga, *Structure* **1999**, *7*, 1279–1290.

[266] A. M. Haapalainen, G. Meriläinen, R. K. Wierenga, *Trends in Biochemical Sciences* **2006**, *31*, 64–71.

[267] C. M. Schaefer, R. Lu, N. M. Nesbitt, J. Schiebel, N. S. Sampson, C. Kisker, *Structure* **2015**, *23*, 21–33.

[268] R. A. Dwerk, *Chem. Rev.* **1996**, 683–720.

[269] F. Freitas, V. D. Alves, M. A. M. Reis, *Springer, Cham* **2014**.

[270] M. L. Velázquez-Hernández, V. M. Baizabal-Aguirre, A. Bravo-Patiño, M. Cajero-Juárez, M. P. Chávez-Moctezuma, J. J. Valdez-Alarcón, *J Appl Microbiol* **2009**, *106*.

[271] R. Chambert, G.-T. G., *Eur. J. Biochem* **1976**, 493–508.

[272] L. K. Ozimek, S. Kralj, M. J. van der Maarel, L. Dijkhuizen, *Microbiology* **2006**, 1187–96.

[273] L. K. Dijkhuizen, G. J. W. Ozimek, M. J. E. C. Euverink, L. van der Maarel, *FEBS Letters* **2005**, *579*, 1124–1128.

[274] A. Homann, R. Biedendieck, S. Götze, D. Jahn, J. Seibel, *Biochemical Journal* **2007**, *407 Pt(2)*, 189–198.

[275] C. P. Strube, A. Homann, M. Gamer, D. Jahn, J. Seibel, D. W. Heinz, *The journal of biological chemistry* **2011**, 17593–600.

[276] M. T. Reetz, *Journal of American Chemical Society* **2013**, 12480–12496.

[277] C. D. Spicer, B. G. Davis, *Nature Communication* **2014**.

[278] J. M. Antos, J. M. McFarland, A. T. Iavarone, M. B. Francis, *Journal of American Chemical Society* **2009**, *131*, 6301–6308.

[279] K. L. Seim, A. C. Obermeyer, M. B. Francis, *Journal of American Chemical Society* **2011**, *133*, 16970–16976.

[280] H. Ban, J. Gavrilyuk, C. Barbas III, *Journal of American Chemical Society* **2010**, *132*, 1523–1525.

[281] S. Meunier, E. Strable, M. G. Finn, *Chemical Biology* **2004**, *11*, 319–326.

[282] S. D. Tilley, M. B. Francis, *Journal of American Chemical Society* **2006**, *128*, 1080–1081.

[283] N. S. Joshi, L. R. Whitaker, M. B. Francis, *Journal of American Chemical Society* **2004**, *126*.

[284] M. E. Ortiz-Soto, J. Ertl, J. Mut, J. Adelmann, T. A. Le, J. Shan, J. Teßmar, A. Schlosser, B. Engels, J. Seibel, *Chemical Science* **2018**, *9*, 5312–5321.

[285] A. Paasche, T. Schirmeister, B. Engels, *Journal of Chemical Theory and Computation* **2013**, *9*, 1765–1777.

[286] H. Li, A. D. Robertson, J. H. Jensen, *Proteins* **2005**, *61*, 704–721.

[287] D. Bas, D. Rogers, J. H. Jensen, *Proteins* **2008**, *73*, 765–783.

[288] T. Matsui, T. Baba, K. Kamiya, Y. Shigeta, *Physical Chemistry Chemical Physics* **2012**, *14*, 4181–4187.

[289] E. Socher, H. Sticht, *Scientific Reports* **2016**, *6*, Article Number: 22523.

[290] H. H. Otto, T. Schirmeister, *Chemical Reviews* **1997**, *97*, 133–172.

[291] Z. Ke, H. Guo, D. Xie, S. Wang, Y. Zhang, *The Journal of Physical Chemistry B* **2011**, *115*, 3725–3733.

[292] W. Lee, S. R. Luckner, C. Kisker, P. J. Tonge, B. Engels, *Biochemistry* **2011**, *50*, 5743–5756.

[293] K. Range, D. Riccardi, Q. Cui, M. Elstner, D. M. York, *Physical Chemistry Chemical Physics* **2005**, *7*, 3070–3079.

[294] D. A. Anick, *The Journal of Physical Chemistry A* **2003**, *107*, 1348–1358.

[295] S. Sadhukhan, D. Muñoz, C. Adamo, G. E. Scuseria, *Chemical Physics Letters* **1999**, *306*, 83–87.

[296] J. A. Frey, A. Müller, M. Loasada, S. Leutwyler, *The Journal of Physical Chemistry B* **2007**, *111*, 3534–3542.

[297] N. F. Brás, M. A. S. Perez, P. A. Fernandes, J. S. Pedro, M. J. Ramos, *Journal of Chemical Theory and Computation* **2011**, *7*, 3898–3908.

[298] J. P. Perdew, *Physical Reviews B* **1986**, *33*, 8822–8824.

[299] F. Stillinger, *Science* **1980**, *209*.

[300] I. Ohmine, S. Saito, *Accounts of Chemical Research* **1999**, *32*, 741–749.

Chapter 16

Appendix

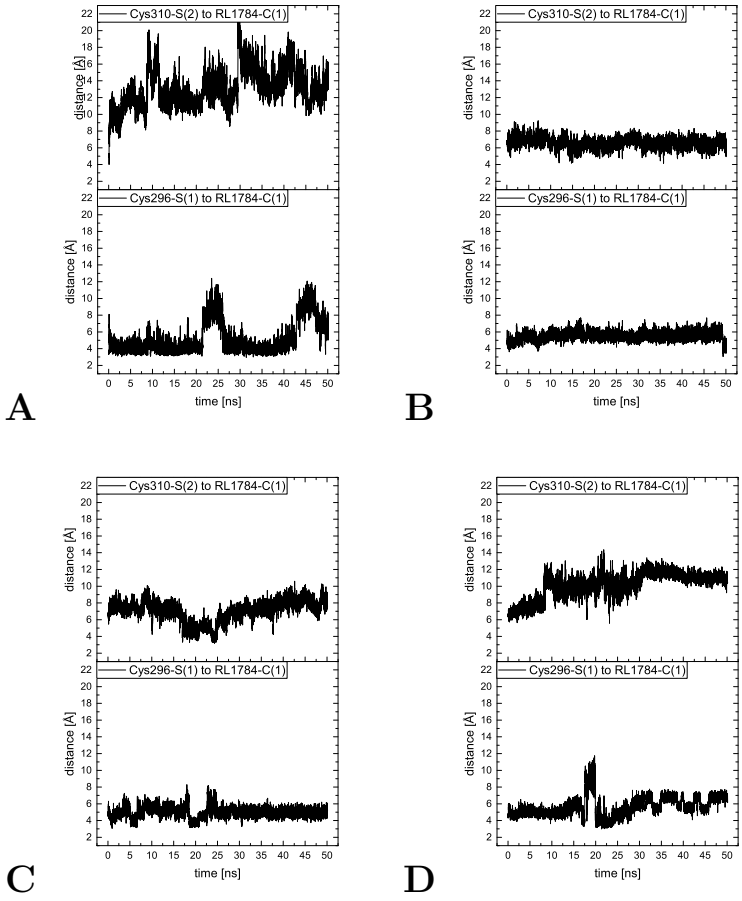


Figure 16.1: The variation in distance as a function of simulation time for RL1784 with Cys296 or Cys310 in structure Ic (**A**), in structure IIa (**B**), in structure IIb (**C**) and in structure IIc (**D**), corresponding to Figure 9.8.

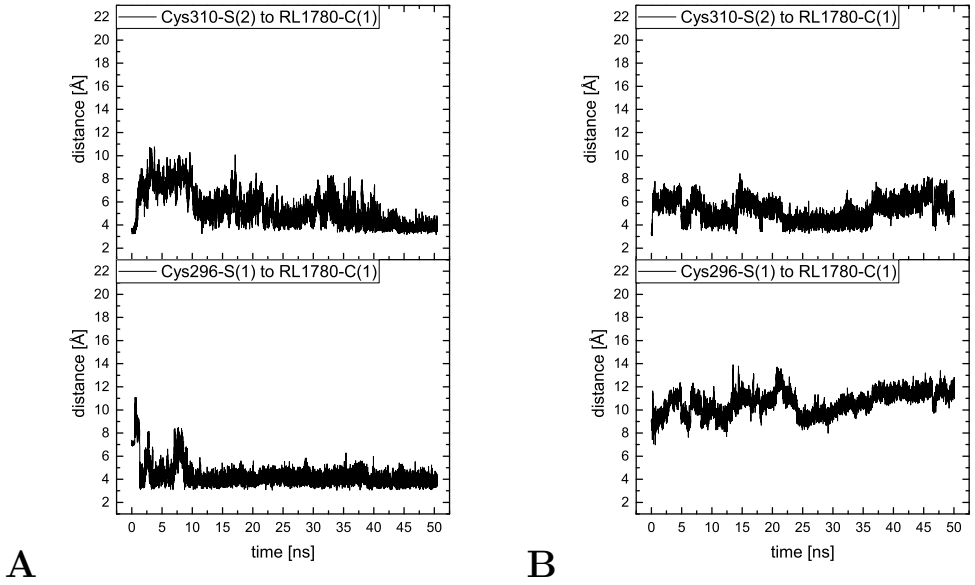


Figure 16.2: The variation in distance as a function of simulation time for RL1780 with Cys296 or Cys310 in structure Ib (**A**) and in structure IIc (**B**) corresponding to Figure 9.8.

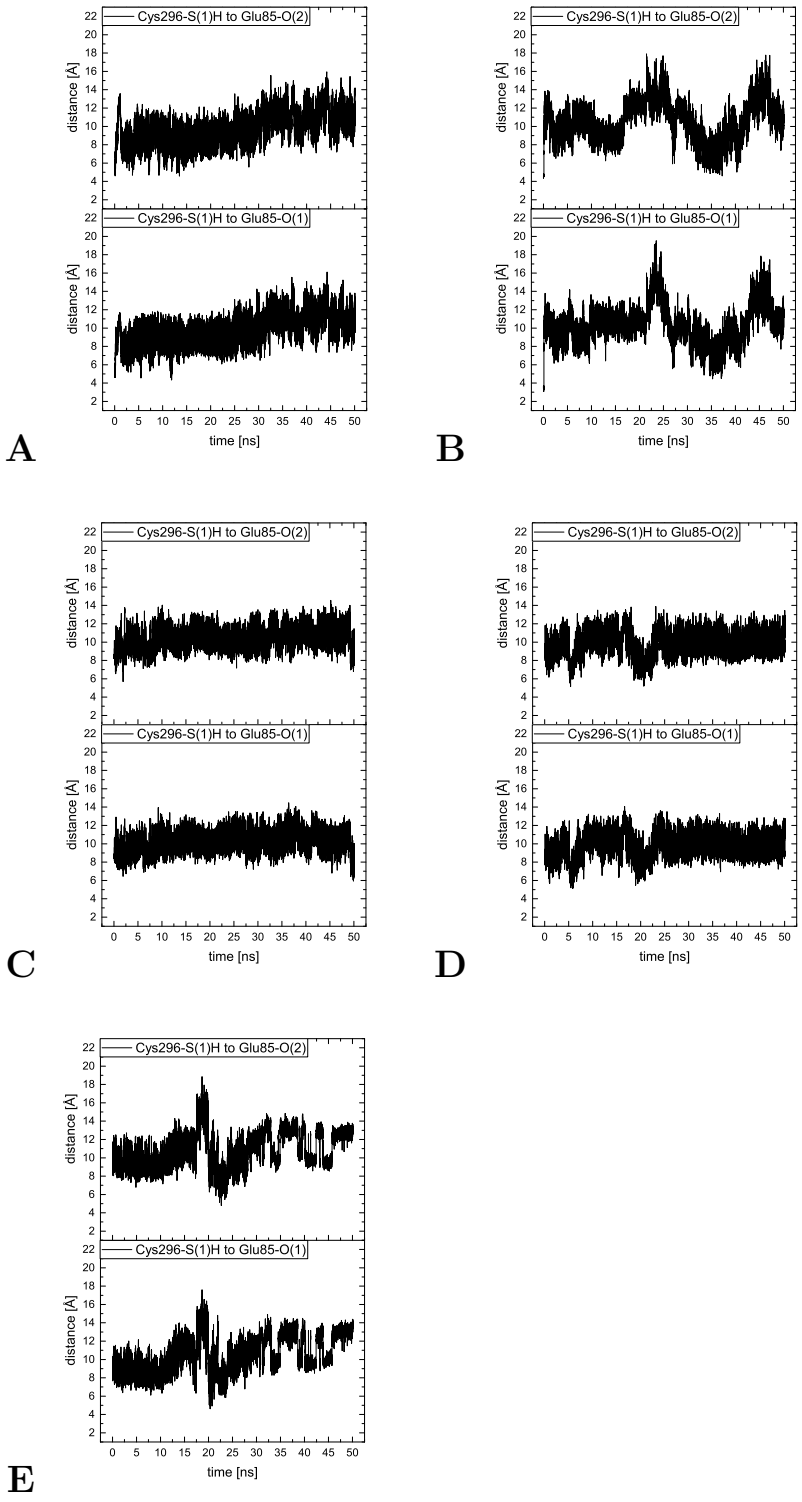


Figure 16.3: The variation in distance as a function of simulation time for Glu85 with Cys296 in presence of RL1784 in structure Ia (**A**), in structure Ic (**B**), in structure IIa (**C**), in structure IIb (**D**) and in structure IIc (**E**), corresponding to Figure 9.12.

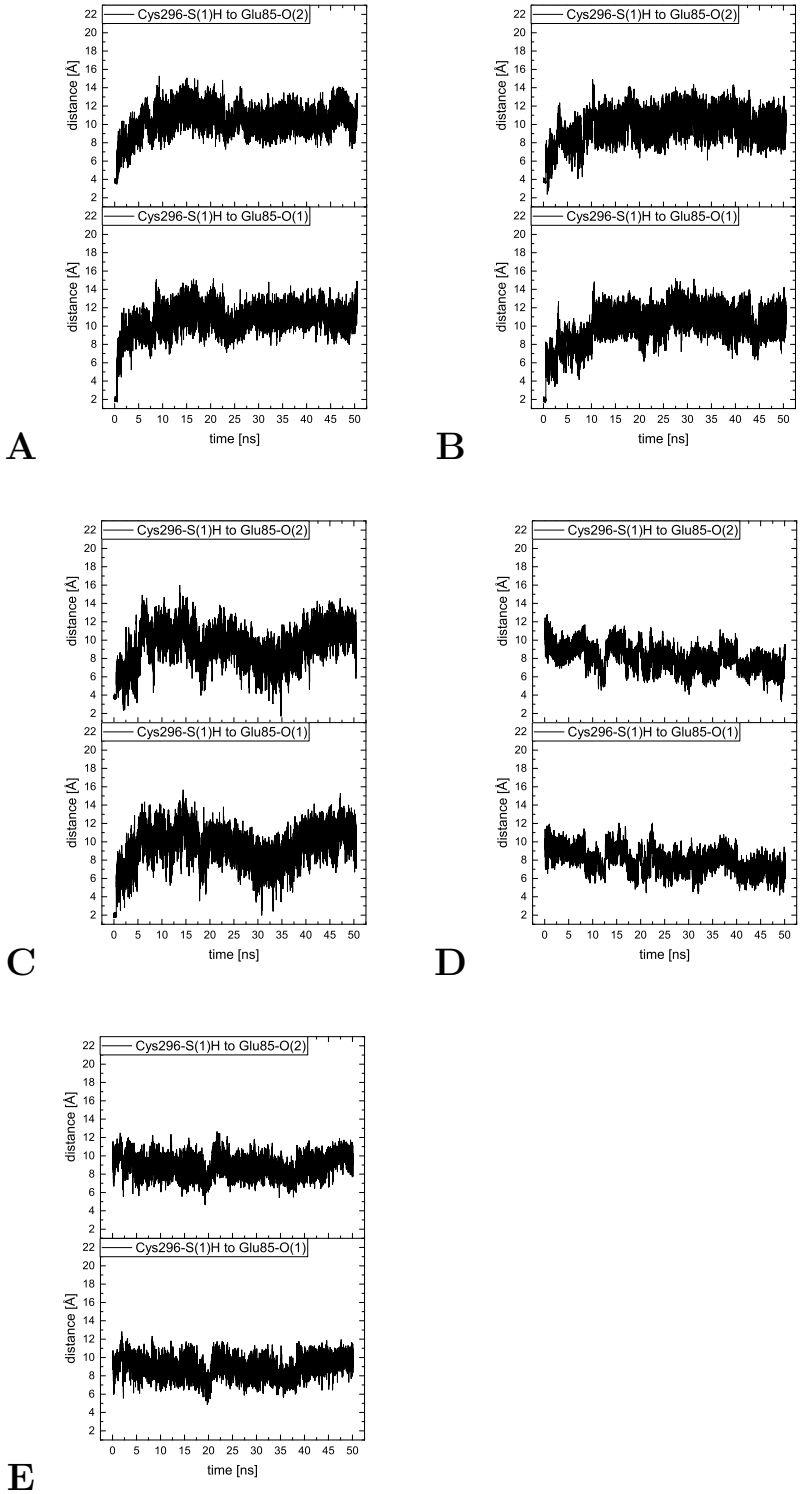


Figure 16.4: The variation in distance as a function of simulation time b Glu85 with Cys296 in presence of RL1780 in structure Ia (**A**), in structure Ib (**B**), in structure Ic (**C**), in structure IIa (**D**) and in structure IIc (**E**), corresponding to Figure 9.12.

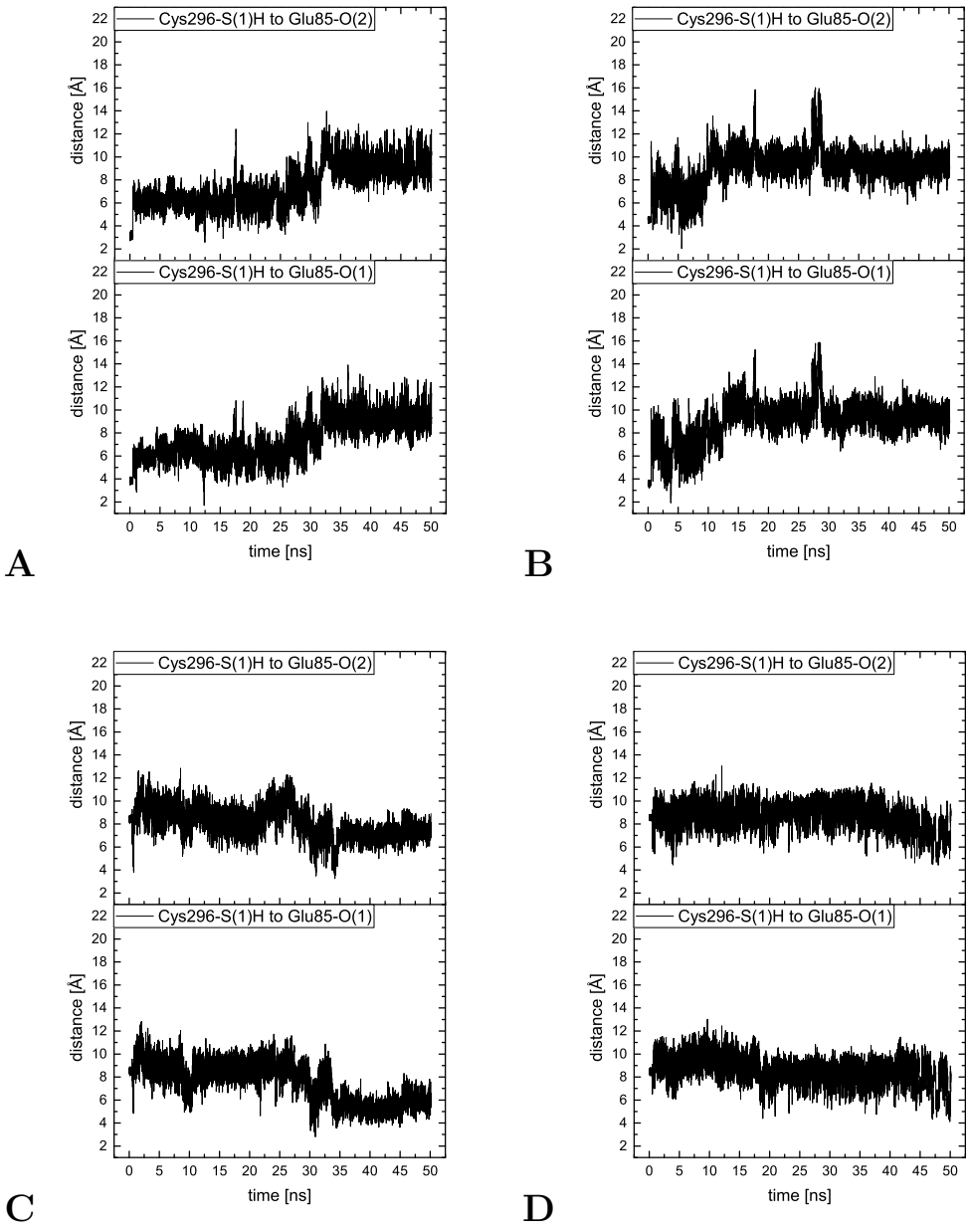


Figure 16.5: The variation in distance as a function of simulation time for Glu85 with Cys296 in presence of RL2132 in structure Ia (**A**), in structure Ic (**B**), in structure IIIb (**C**) and in structure IIIc (**D**), corresponding to Figure 9.12.

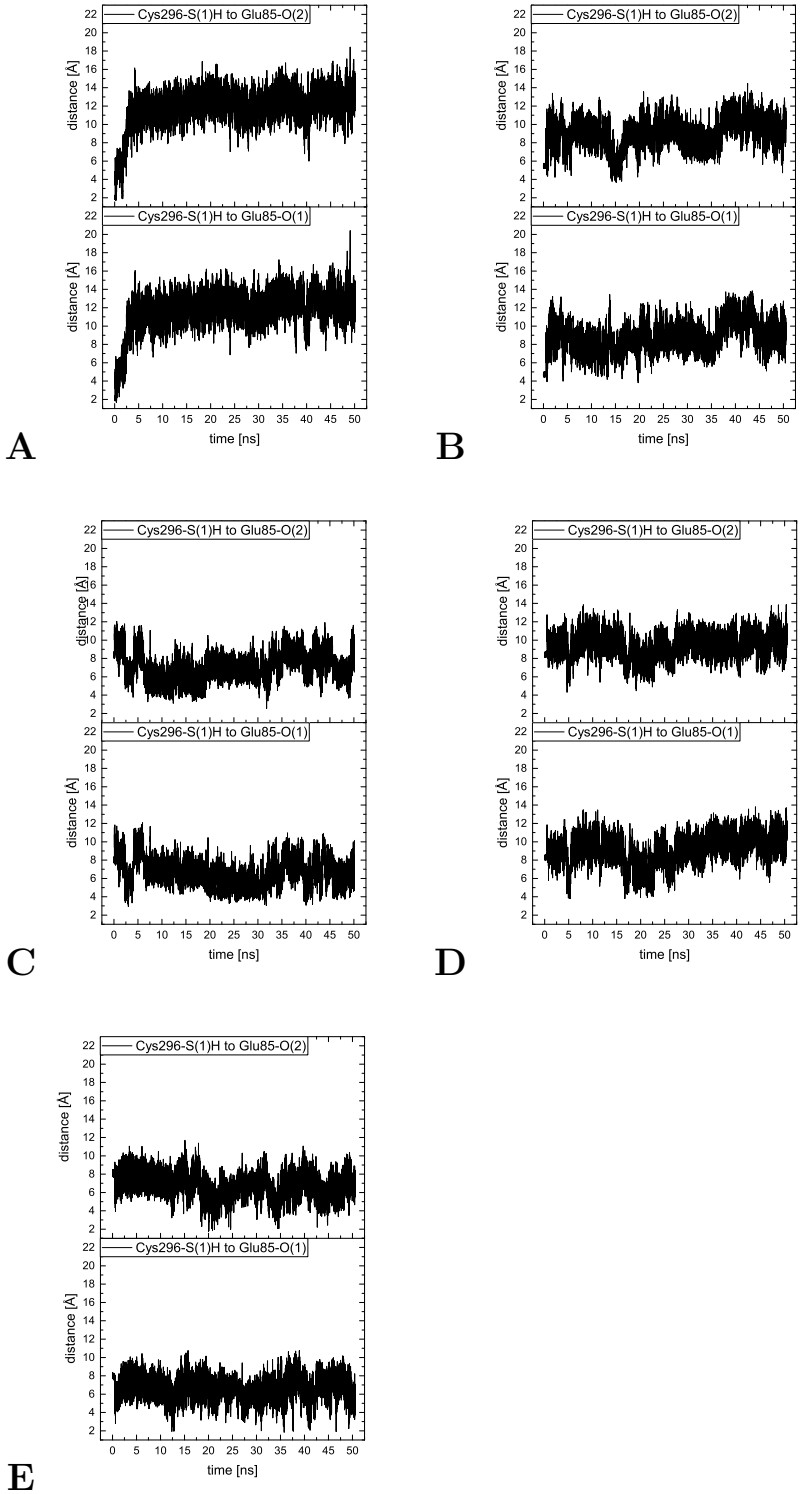


Figure 16.6: The variation in distance as a function of simulation time for Glu85 with Cys296 in absence of RL1784, RL1780 or RL2132 in structure Ia (**A**), in structure Ic (**B**), in structure IIa (**C**), in structure IIb (**D**) and in structure IIc (**E**), corresponding to Figure 9.12.

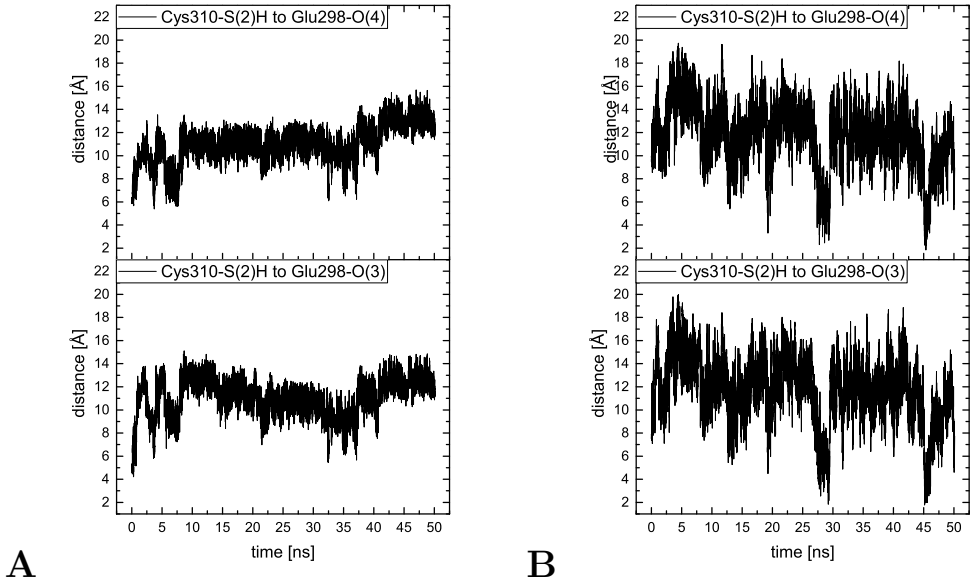


Figure 16.7: The variation in distance as a function of simulation time for Glu298 with Cys310 in presence of RL1784 in structure Ia (**A**) and in structure Ic (**B**) corresponding to Figure 9.17.

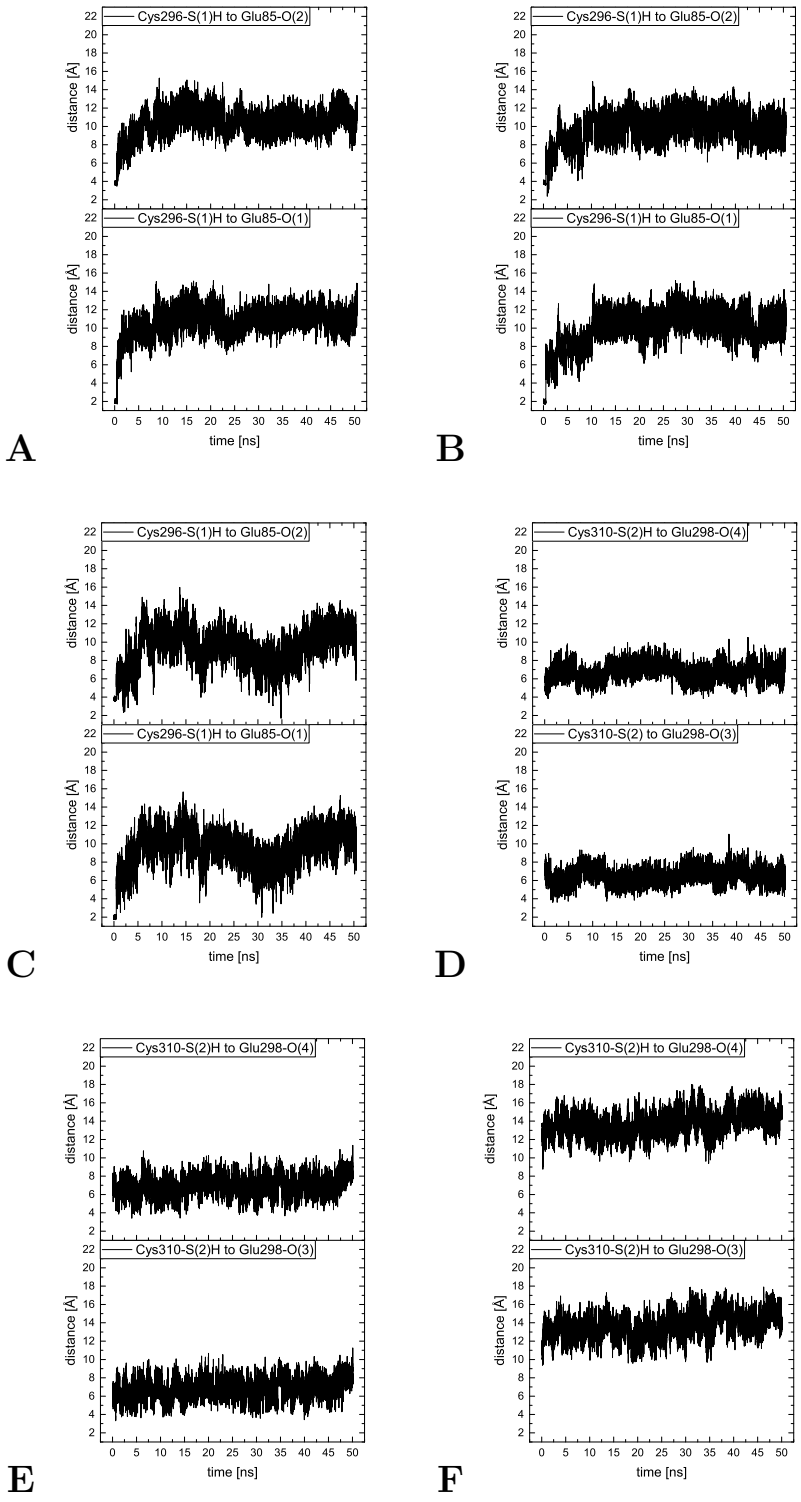


Figure 16.8: The variation in distance as a function of simulation time for Glu298 with Cys310 in presence of RL1780 in structure Ia (**A**), in structure Ib (**B**), in structure Ic (**C**), in structure IIa (**D**), in structure IIb (**E**) and in structure IIc (F), corresponding to Figure 9.17.

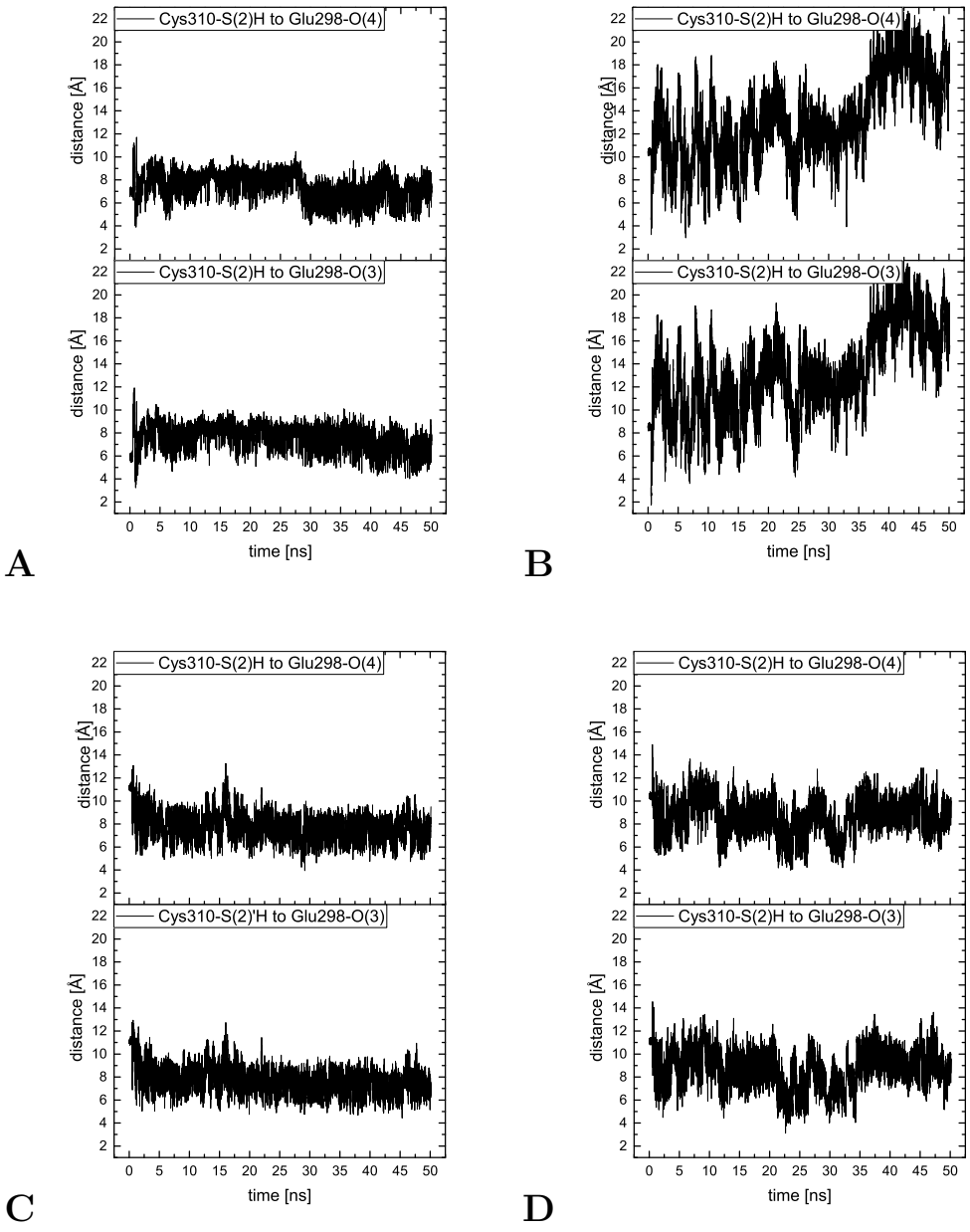


Figure 16.9: The variation in distance as a function of simulation time for Glu298 with Cys310 in presence of RL2132 in structure Ia (**A**), in structure Ic (**B**), in structure IIa (**C**) and in structure IIb (**D**), corresponding to Figure 9.17.

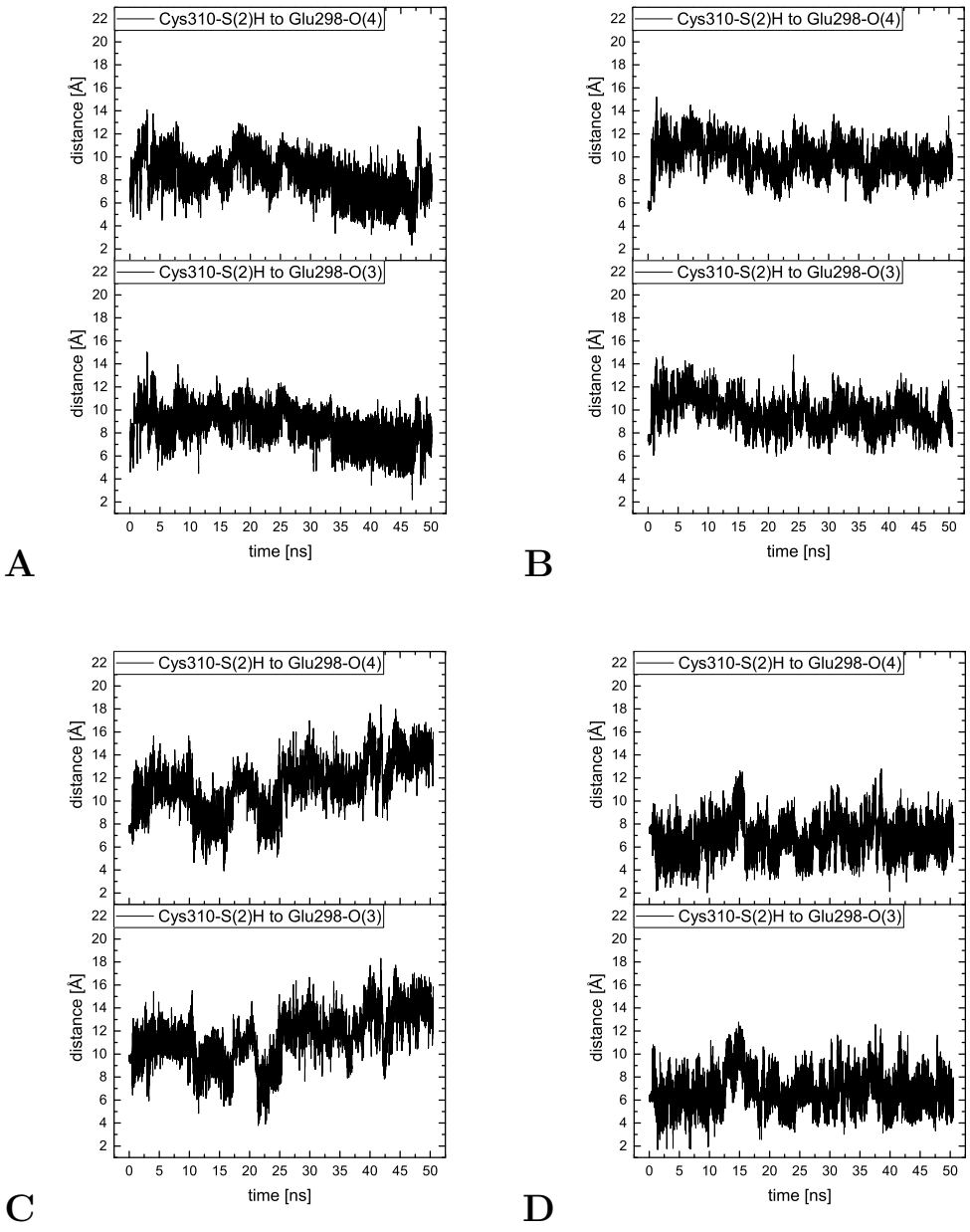


Figure 16.10: The variation in distance as a function of simulation time for Glu298 with Cys310 in absence of RL1784, RL1780 or RL2132 in structure Ia (**A**), in structure Ib (**B**), in structure Ic (**C**) and in structure IIc (**D**), corresponding to Figure 9.17.

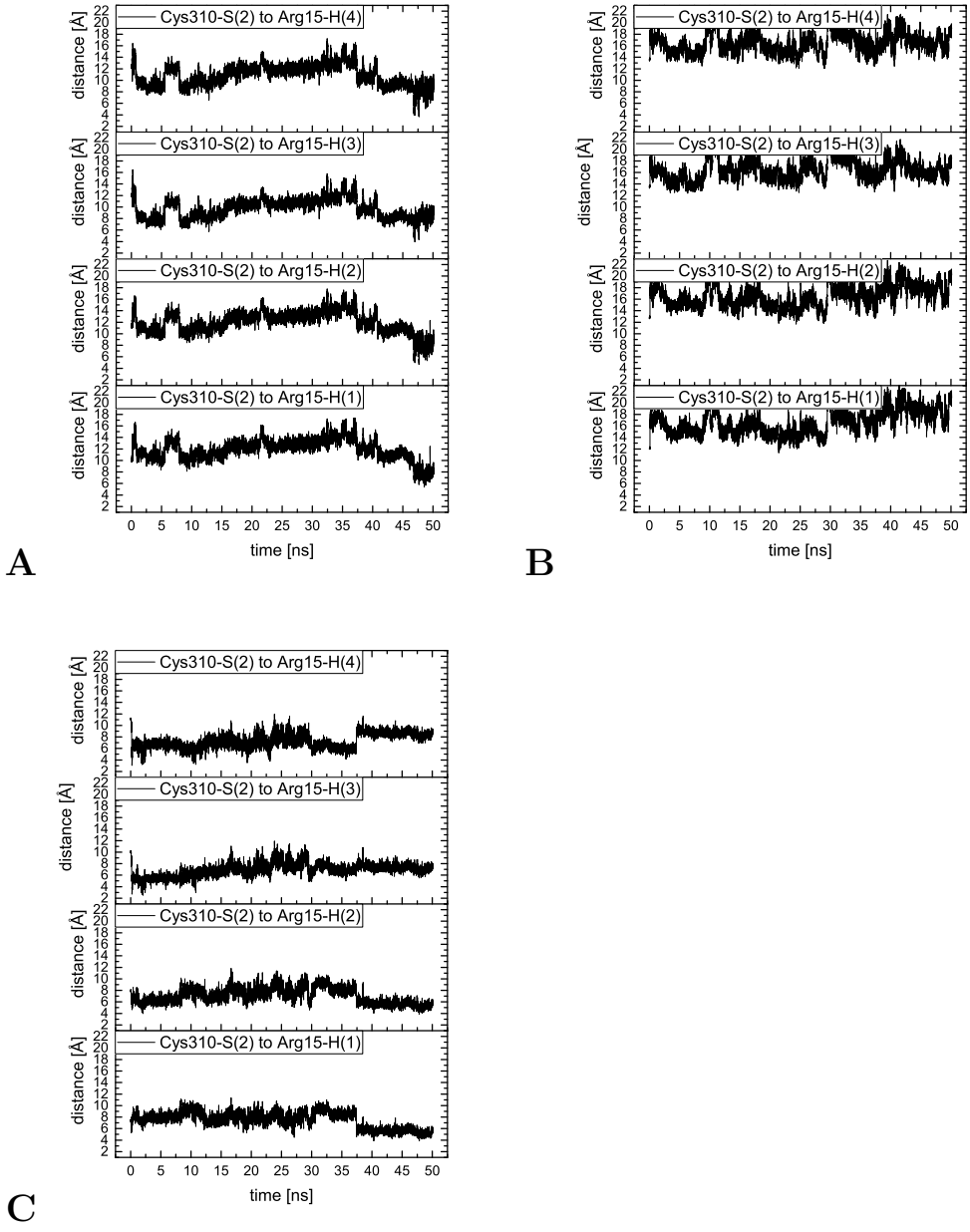


Figure 16.11: The variation in distance as a function of simulation time for Arg15 with Cys310 in presence of RL1784 in structure Ia (**A**), in structure Ic (**B**) and in structure IIc (**C**), corresponding to Figure 9.21.

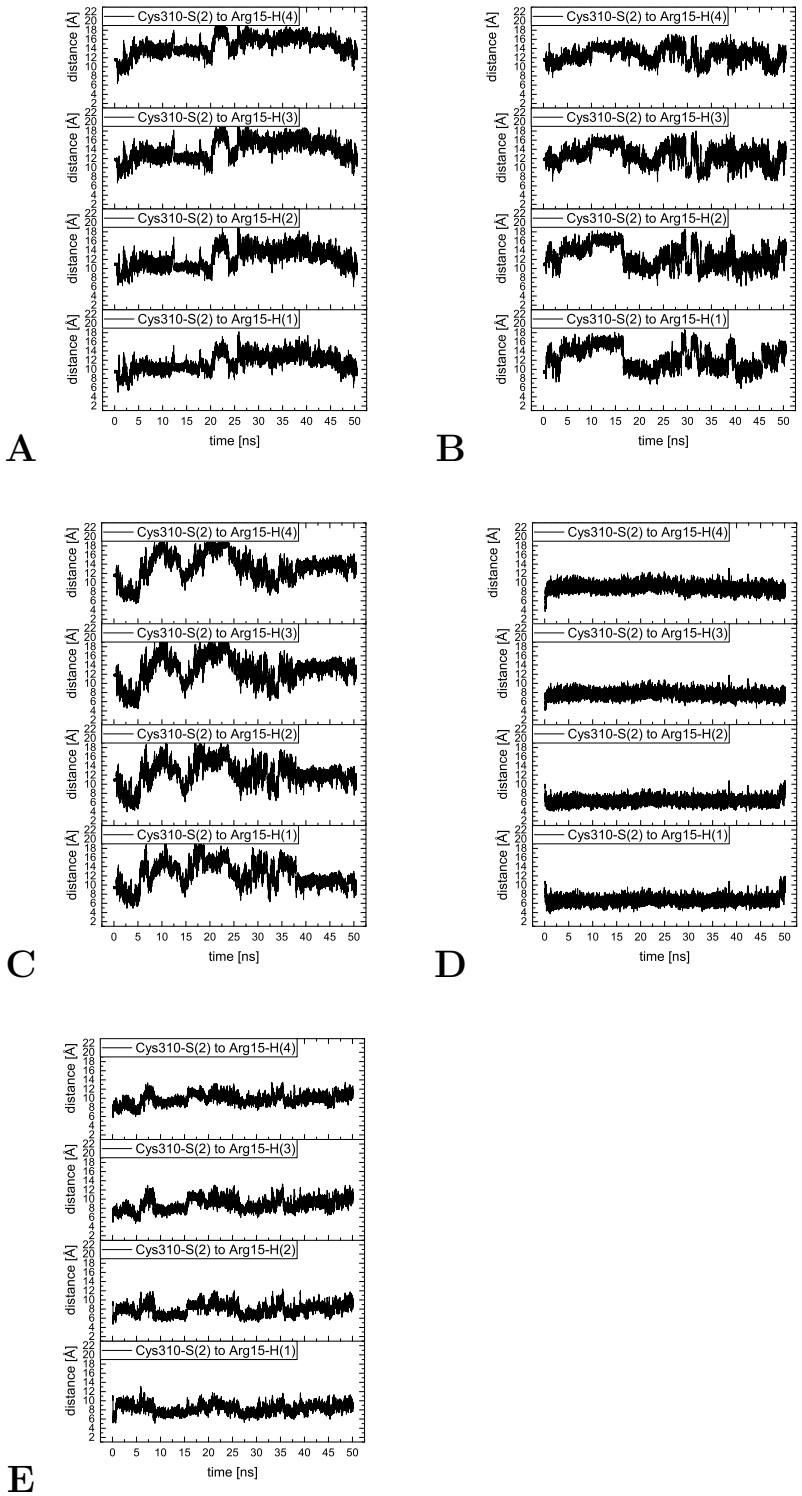


Figure 16.12: The variation in distance as a function of simulation time for Arg15 with Cys310 in presence of RL1780 in structure Ia (**A**), in structure Ib (**B**), in structure Ic (**C**), in structure IIa (**D**) and in structure IIb (**E**), corresponding to Figure 9.21.

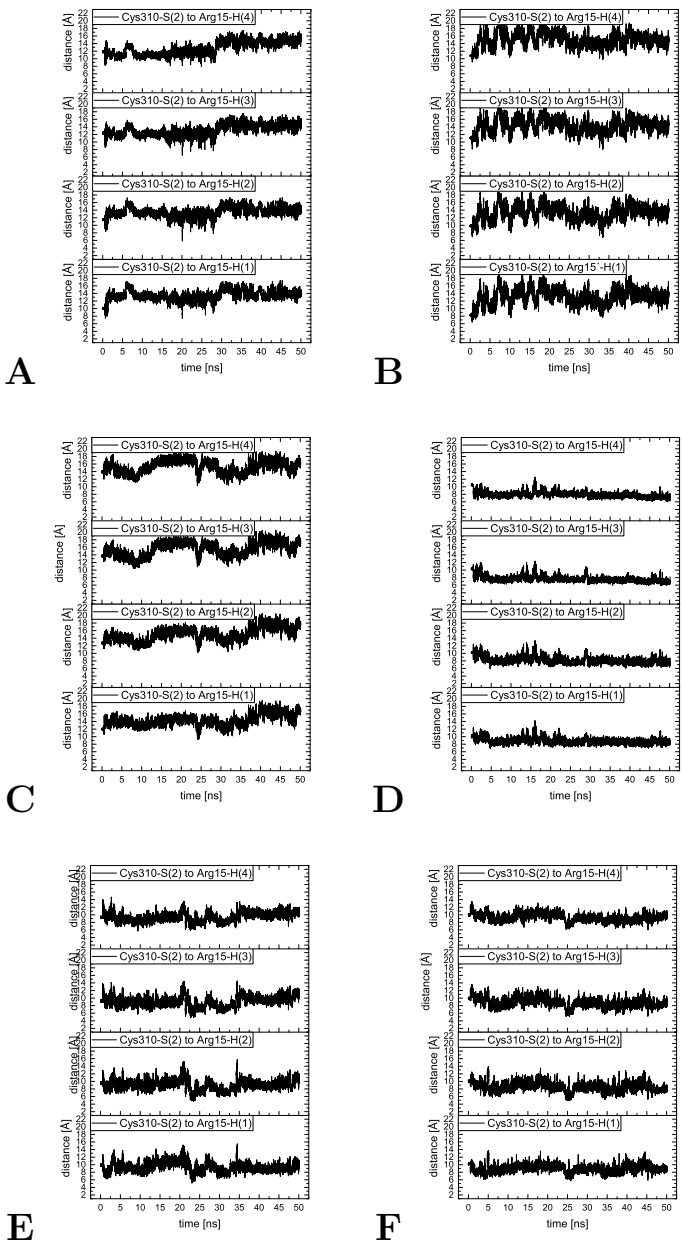


Figure 16.13: The variation in distance as a function of simulation time for Arg15 with Cys310 in presence of RL2132 in structure Ia (**A**), in structure Ib (**B**), in structure Ic (**C**), in structure IIIa (**D**), in structure IIIb (**E**) and in structure IIIc (**F**), corresponding to Figure 9.21.

Chapter 17

Abbreviations

ADP	Adenosine diphosphate
AIDS	Acquired immunodefiency syndroms
Ala	Alanine
AP	Alchemical Perturbation
Arg	Arginine
Asn	Asparagine
Asp	Aspartic Acid
ATP	Adenosine Triphosphate
B3LYP	Becke, 3-parameter, Lee-Yang functional
B3LYP-D	B3LYP with Grimme's dispersion correction
B3LYP_G	B3LYP implemented in Gaussian
Bm-LS	*Bacillus megaterium levansucrase*
CAI	Covalent allosteric inhibitor
CAST	Conformational Analysis and Search Tool
CoA	Coenzyme A
CCSD(T)	Coupled Cluster with single and double and perturbative triple excitation
CHARMM	Chemistry at Harvard macromolecular mechanics
CID	Chemical induced dimerization
CNS	Central nervous system
CPI	Cysteine protease inhibitor
CTSC	Cathepsin C
Cys	Cysteine
DNA	Desoxyribonucleic acid
DFT	Density functional theory
DP	Degree of polymerization
DPE	Diphenylether
EC	Enzyme commission
ECM	Extracellular matrix
ENR	Enoyl-(acyl-carrier-protein)
FAS	Fatty acid biosynthesis

FF	Forcefield
GA	Genetic algorithm
GAFF	Generalized AMBER force field
GB	Generalized Born
GGA	Generalized gradient approximation
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
GPCR	G protein-coupled receptors
GSK3	Glycogen synthase kinase 3
HAT	Human African Trypanosomiasis
HF	Hartree Fock
His	Histidine
HIV	Human immunodeficiency virus
IC	Incremental construction
Ile	Isoleucine
LDA	Local density approximation
Leu	Leucine
Lys	Lysine
MA	Matching algorithms
MAE	Mean averaged error
MD	Molecular dynamics
MDR	multidrug-resistant
Met	Methionine
MM	Molecular mechanics
MP2	Møller-Plesset perturbation theory 2
MRSA	Methicillin-resistant *Staphylococcus Aureus*
MTB	*Mycobacterium tuberculosis*
NADPH	Nicotinamide adenine dinucleotide phosphate
NADH	Nicotinamde adenine dinucleotide

NEB	Nugded elastic band
NMR	Nuclear magnetic resonance
NTD	Neglected tropical disease
PB	Poisson-Boltzmann
PDB	Protein data bank
PDK1	Phosphoinositide-dependent protein kinase 1
PH	Pleckstrin homology
Phe	Phenylalanine
pHtMD	pH-titration MD
PKB	Protein kinase B
Pro	Proline
Px	Peroxidase
QM	Quantum Mechanics
RNA	Ribonucleic acid
RMSD	Root mean square deviation
RTK	Receptor tyrosine kinases
SAS	Solvent accessible surface
SASA	Solvent accessible surface area
SCF	Self consistent field
SCS	Spin component scaling
Ser	Serine
SD	standard deviation
T.b.	*Trypanosoma brucei*
TB	Tuberculosis
Thr	Threonine
TPX	Tryparedoxin
TR	Trypanothione reductase
Trp	Tryptophan
Tyr	Tyrosine
UDP	Uridine diphosphate
Val	Valine
WHO	World Health Organization